On the number of matroids compared to the number of sparse paving matroids

Rudi Pendavingh Jorn van der Pol[*]

Department of Mathematics and Computer Science
Eindhoven University of Technology
Eindhoven, The Netherlands

{R.A.Pendavingh, J.G.v.d.Pol}@tue.nl

Submitted: Dec 10, 2014; Accepted: May 26, 2015; Published: Jun 15, 2015
Mathematics Subject Classifications: 05B35, 05A16

Abstract

It has been conjectured that sparse paving matroids will eventually predominate in any asymptotic enumeration of matroids, i.e. that $\lim_{n\to\infty} s_n/m_n = 1$, where m_n denotes the number of matroids on n elements, and s_n the number of sparse paving matroids. In this paper, we show that

$$\lim_{n\to\infty} \frac{\log s_n}{\log m_n} = 1.$$

We prove this by arguing that each matroid on n elements has a faithful description consisting of a stable set of a Johnson graph together with a (by comparison) vanishing amount of other information, and using that stable sets in these Johnson graphs correspond one-to-one to sparse paving matroids on n elements.

As a consequence of our result, we find that for all $\beta > \sqrt{\frac{\ln 2}{2}} = 0.5887\cdots$, asymptotically almost all matroids on n elements have rank in the range $n/2 \pm \beta\sqrt{n}$.

Keywords: Matroid theory, asymptotic enumeration

1 Introduction

After matroids up to 8 elements were enumerated by Blackburn, Crapo, and Higgs [BCH73], it was noted that a substantial fraction of the matroids were paving matroids, that is, matroids M whose circuits are all of cardinality at least $r(M)$. Crapo and Rota

[*]This research has been supported by the Netherlands Organisation for Scientific Research (NWO) grant 613.001.211.

speculated that perhaps ‘paving matroids will predominate in any enumeration of matroids’ [CR70, p. 3.17]. Mayhew, Newman, Welsh and Whittle [MNWW11, Conjecture 1.6] make the more precise conjecture in that the asymptotic fraction of matroids on n elements that are paving tends to 1 as n tends to infinity. Their conjecture is equivalent to the seemingly stronger statement that

$$\lim_{n\to\infty} s_n/m_n = 1. \tag{1}$$

Here m_n denotes the number of matroids on a fixed ground set of n elements, and s_n is the number of sparse paving matroids (a matroid is sparse paving if both it and its dual are paving).

Sparse paving matroids seem benign objects compared to matroids in general. For example, it is straightforward that asymptotically almost all sparse paving matroids are highly connected. The predominance of sparse paving matroids as in (1) would thus immediately imply the predominance of k-connected matroids, which is conjectured [MNWW11, Conjecture 1.5] but remains an open problem. Similarly, it is relatively straightforward that asymptotically all sparse paving matroids have a fixed uniform matroid $U_{a,b}$ as a minor, but the analogous statement for general matroids is open. Further examples along these lines are easy to find, whence the interest in conjecture (1).

Combining the lower bound of Graham and Sloane [GS80, Theorem 1] (as pointed out in [MW13]) on s_n and the upper bound of Bansal, Pendavingh and van der Pol [BPvdP14] on m_n, we have

$$\frac{1}{n}\binom{n}{\lfloor n/2\rfloor} \leqslant \log s_n \leqslant \log m_n \leqslant \frac{2+o(1)}{n}\binom{n}{\lfloor n/2\rfloor} \qquad \text{as } n\to\infty. \tag{2}$$

These bounds do not suffice to prove (1), but merely imply

$$\log s_n \leqslant \log m_n \leqslant (2+o(1))\log s_n \text{ as } n\to\infty,$$

or equivalently, that $s_n \leqslant m_n \leqslant s_n^{2+o(1)}$.

The sparse paving matroids showed a somewhat less benign side when we attempted to narrow the gap between the upper and the lower bound in (2). Being unable to improve either bound, we devised a way to directly compare the number of matroids to the number of sparse paving matroids. This enabled us to prove the main result of this paper, that

$$\log m_n \leqslant (1+o(1))\log s_n \text{ as } n\to\infty,$$

or equivalently, $m_n = s_n^{1+o(1)}$. Our method is closely related to the one used in [BPvdP14] to prove the upper bound on m_n, which itself is an adaptation of a method to bound s_n.

We will briefly outline the method and describe how it differs from earlier work. Key to our method is an algorithm for producing a compressed description of any given matroid M on E of rank r. The compression algorithm considers the set of bases of M as a subset of all the r-subsets of E, which are the vertices the Johnson graph $J(E,r)$. We obtain a compact description of the matroid by starting from the full set of vertices A of the

Johnson graph $G = J(E, r)$ and iteratively taking away neighborhoods of vertices from A while describing the set of bases among these neighbourhoods. As long as there are vertices of high degree in $G[A]$ to pick, the rate at which we need to add information into our matroid description compares favourably to the decrease in the size of A. We argued that while A is large there will be such vertices of high degree, and by the time A contains no more than a certain α-fraction of the vertices, the total amount of information stored so far will still be relatively modest. In our previous paper, we completed the description of the matroid by adding $\alpha\binom{n}{r}$ bits to describe the subset of bases among the remaining vertices of A, and this is what ultimately dominated the length of the matroid description we obtained. Hence, the cost of describing the bases among the final set A was the bottleneck for producing a tighter upper bound on m_n.

Previously, the presence of a vertex of high degree in $G[A]$ was necessary to show that the set of bases in its neighbourhood can be described using a relatively small amount of information. In the present paper, we show that in the final stage a small maximum degree is advantageous as well. In particular, if $G[A]$ does not contain vertices of high degree, then most of the neighbours of $X \in A$ lie outside A, so that for most of these neighbours it is known whether they are a basis of the matroid or not from the matroid description so far. Exploiting this information and matroid structure, we find that the $\alpha\binom{n}{r}$ bits we used before can be replaced by a certain stable set $T \subseteq A$ to obtain a faithful description of the matroid. Since our matroid description now consists of a stable set in the Johnson graph together with some contained amount of further information, it becomes possible to compare the number of matroids directly to the number of stable sets in the Johnson graph. As stable sets in $J(E, r)$ are in 1-1 correspondence to sparse paving matroids, this implies our main result.

A further result in this paper is that for $\beta > \sqrt{\frac{\ln 2}{2}} = 0.5887\cdots$ asymptotically all matroids on n elements have a rank between $n/2 - \beta\sqrt{n}$ and $n/2 + \beta\sqrt{n}$. This is related to a second conjecture from [MNWW11, Conjecture 1.10], that asymptotically all matroids on n elements have a rank between $(n-1)/2$ and $(n+1)/2$.

After giving preliminaries on graphs and matroids in Section 2, we present both results in Section 3. Finally, we discuss several remaining open problems related to our main results in Section 4.

2 Preliminaries

2.1 Graphs and stable sets

We only consider loopless, undirected graphs in this paper. If G is any graph, then we write $\Delta(G)$ for the maximum degree in G. Further, for any $A \subseteq V(G)$, we write $G[A]$ for the subgraph of G induced by the vertices in A. A set of vertices $A \subseteq V(G)$ is *stable* if $G[A]$ spans no edges.[1] We write $i(G)$ for the number of stable sets in G.

[1] Note that what we call a *stable set* is often called an *independent set* in graph theory. As independent set has a different meaning in matroid theory, this serves to avoid confusion.

In [BPvdP14], Nikhil Bansal and the current authors proved the following result on the number of stable sets in regular graphs. In it, the smallest eigenvalue of a graph refers to the smallest eigenvalue of the adjacency matrix of that graph.

Theorem 1. *Let G be a d-regular graph on N vertices with smallest eigenvalue $-\lambda$, with $d > 0$. Then $i(G) \leqslant \sum_{s=0}^{\lceil \sigma N \rceil} \binom{N}{s} 2^{\alpha N}$, where $\alpha = \frac{\lambda}{d+\lambda}$ and $\sigma = \frac{\ln(d+1)}{d+\lambda}$.*

The quantity αN is known as the *Hoffman bound*, which is an upper bound on the cardinality of a stable set in G.

2.2 The Johnson graph

Let E be a finite set, and let $0 < r < |E|$. The Johnson graph $J(E, r)$ is the graph with vertex set

$$\binom{E}{r} := \{X \subseteq E : \ |X| = r\},$$

in which any two vertices are adjacent if and only if they have $r - 1$ elements in common; equivalently, the vertices X and Y are adjacent whenever $|X \triangle Y| = 2$.

Let $[n] := \{1, \ldots, n\}$. We abbreviate $J(n, r) := J([n], r)$. Clearly $J(E, r) \cong J(n, r)$ when $|E| = n$. Note that $J(n, r) \cong J(n, n - r)$; the function $X \mapsto [n] \setminus X$ provides an explicit isomorphism.

The spectrum of the Johnson graph is known: if $r \leqslant n/2$, the eigenvalues of $J(n, r)$ are

$$(r - i)(n - r - i) - i, \qquad i = 0, 1, \ldots, r. \tag{3}$$

Hence, the smallest eigenvalue of $J(n, r)$ is $-r$, and using Theorem 1 it was derived in [BPvdP14] that

$$\log i(J(n, r)) \leqslant \frac{2}{n}\binom{n}{\lfloor n/2 \rfloor}(1 + o(1)) \qquad \text{as } n \to \infty. \tag{4}$$

A corresponding lower bound is obtained from a construction by Graham and Sloane [GS80, Theorem 1], who show that

$$\log i(J(n, r)) \geqslant \frac{1}{n}\binom{n}{r}. \tag{5}$$

If $X \in \binom{E}{r}$, then we will use the graph-theoretic term *neighbourhood* to denote the set

$$N(X) := \{X - x + y : x \in X, y \in E \setminus X\}.$$

Note that $N(X)$ is precisely the neighbourhood of X, when X is seen as a vertex in the Johnson graph $J(E, r)$. The neighbourhood of X has the structure of a Cartesian graph product of K_r and K_{n-r}. In particular, we can distinguish 'rows'

$$R_X(x) := \{X - x + y : y \in E \setminus X\}, \qquad\qquad x \in X \tag{6}$$

and 'columns'

$$C_X(y) := \{X - x + y : x \in X\}, \qquad\qquad y \in E \setminus X, \tag{7}$$

that induce cliques in the neighbourhood of X, see Figure 1.

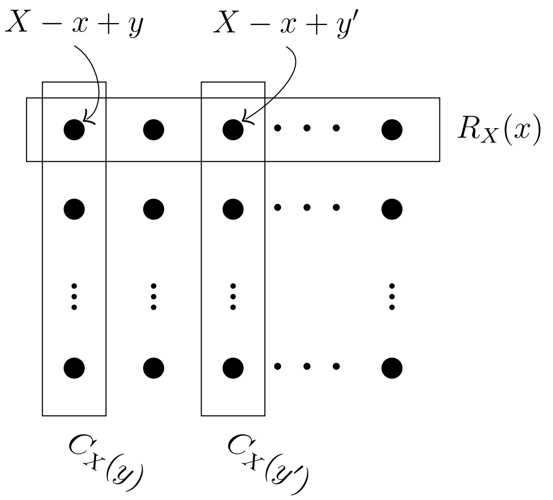

Figure 1: The neighbourhood of X in the Johnson graph. The rows and columns, indexed by $x \in X$ resp. $y \in E \setminus X$, form cliques.

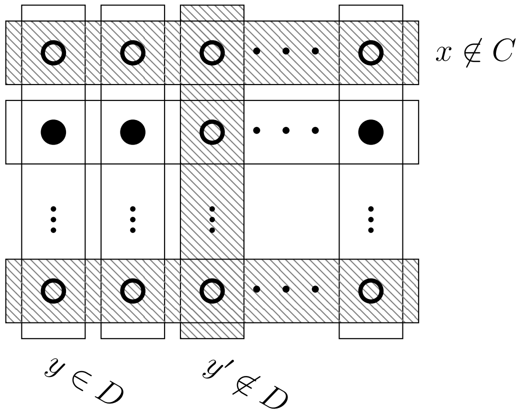

Figure 2: If $r(X) = r - 1$, then the bases in $N(X)$ (represented here by solid vertices) can be recovered by removing all rows corresponding to $x \notin C$ and columns corresponding to $y \notin D$.

2.3 Matroids

A matroid is a pair $M = (E, \mathcal{B})$ such that E is a finite set and $\mathcal{B}$ is a non-empty set of subsets of E satisfying the *basis-exchange axiom*

$$\text{for all } B, B' \in \mathcal{B} \text{ and } e \in B \setminus B' \text{ there exists an } f \in B' \setminus B \text{ such that } B - e + f \in \mathcal{B}. \quad (8)$$

We assume familiarity with the definitions of circuit, flat, rank, dual etc., for which we refer to Oxley's book [Oxl11].

While we mostly use the notation of [Oxl11], the following is nonstandard. We write $\mathbb{M}_{n,r}$ for the set of all matroids of rank r with ground set $[n]$, and $\mathbb{M}_n$ for the set of matroids with ground set $[n]$, and we put $m_{n,r} := \#\mathbb{M}_{n,r}$ and $m_n := \#\mathbb{M}_n$.

A matroid M is said to be *paving* if each circuit of M has cardinality at least $r(M)$, and M is *sparse paving* if both M and its dual are paving. We define

$$s_{n,r} := \#\{M \in \mathbb{M}_{n,r} : M \text{ is sparse paving}\}, \ s_n := \#\{M \in \mathbb{M}_n : M \text{ is sparse paving}\}.$$

The following lemma is essentially established by Piff and Welsh [PW71]. A proof can be found in [BPvdP14, Lemma 8].

Lemma 2. *Let $\mathcal{B} \subseteq \binom{E}{r}$, then $\mathcal{B}$ is the collection bases of a sparse paving matroid if and only if $\binom{E}{r} \setminus \mathcal{B}$ is a stable set in $J(E, r)$.*

Hence $s_{n,r} = i(J(n, r))$, which is bounded by (4).

2.4 The local structure of matroids in the Johnson graph

The following matroid lemma is elementary, but it has a central role in this paper.

Lemma 3. *Let $M = (E, \mathcal{B})$ be a matroid of rank r, and let $X \in \binom{E}{r}$ be such that $r_M(X) = r - 1$. There is a unique circuit C of M so that $C \subseteq X$ and a unique cocircuit D of M so that $D \cap X = \emptyset$.*

The lemma implies that if $r(X) = r - 1$, then the set of bases of M in the neighbourhood of X has a very simple structure that is completely determined by the circuit C and the cocircuit D:

$$N(X) \cap \mathcal{B} = \{X - x + y : x \in C, y \in D\}, \tag{9}$$

see Figure 2. It is this simplicity which enables us to make faithful descriptions of matroids which are nearly as concise as a description of a stable set in the Johnson graph. In the present paper, we use (9) in two ways.

The first use was already implicit in our previous paper [BPvdP14], where we used *local covers* as a short certificate for (in-)dependence in the neighbourhood of a non-basis. We review the definition, and quote the lemma which we will again use in our present argument.

If Y is a set in a matroid M, and $F \in \mathcal{F}(M)$ is a flat such that $|F \cap Y| > r_M(F)$, then Y is necessarily dependent, and $(F, r_M(F))$ serves as a certificate for the dependence of Y. In this case, we say that $(F, r_M(F))$ *covers* Y. Note that M can be reconstructed from its groundset, rank, and a collection of (flat,rank)-pairs such that each non-basis of M is covered by at least one flat in the collection.[2]

A *local cover* at $X \in \binom{E}{r}$ is a subset $\mathcal{Z}_X \subseteq \{(F, r_M(F)) : F \in \mathcal{F}(M)\}$ with the property that each $Y \in N(X) \cup \{X\}$ is either independent, of covered by some $(F, r_M(F)) \in \mathcal{Z}_X$. If X is a non-basis, then the local cover can be surprisingly small, as the following lemma shows. A proof can be found in [BPvdP14, Lemma 20], and also follows from Lemma 3.

Lemma 4. *Let M be a rank-r matroid on groundset E, and let $X \in \binom{E}{r}$ be dependent in M. Then there exists a local cover $\mathcal{Z}_X$ with $|\mathcal{Z}_X| \leqslant 2$.*

The second use of (9) is new to this paper. The very restricted structure of $N(X) \cap \mathcal{B}$ in the neighbourhood of a dependent set X will allow us to recover the partition $(N(X) \setminus \mathcal{B}, N(X) \cap \mathcal{B})$ from partial information $(K \setminus \mathcal{B}, K \cap \mathcal{B})$ for certain $K \subseteq N(X)$, so that a faithful matroid encoding can be even more sparse if we rely on a decoder which can infer from (9).

2.5 Binomial coefficients

We will use the following standard bounds on the sum of binomial coefficients

$$\left(\frac{n}{k}\right)^k \leqslant \sum_{i=0}^{k} \binom{n}{i} \leqslant \left(\frac{en}{k}\right)^k, \tag{10}$$

[2] NB: In [BPvdP14], just the flat F (rather than the pair $(F, r_M(F))$) was used as a certificate for dependence. However, as the rank of F is necessary to reconstruct the matroid, we choose to use the pair here.

a proof of which can be found in [Juk11, Proposition 1.4]. The following estimate of the central binomial coefficient is a consequence of Stirling's approximation:

$$\binom{n}{\lfloor n/2 \rfloor} = \Theta\left(\frac{2^n}{\sqrt{n}}\right) \qquad \text{as } n \to \infty. \tag{11}$$

We will also need the following result on binomial coefficients.

Lemma 5. *If* $k < \lfloor n/2 \rfloor$*, then* $\binom{n}{\lfloor n/2 \rfloor - k} \leqslant \binom{n}{\lfloor n/2 \rfloor} \exp\left(-\frac{k^2}{\lceil n/2 \rceil + k}\right)$.

Proof. Using the identity $\binom{n}{t-1} = \frac{t}{n+1-t}\binom{n}{t}$ repeatedly, we find

$$\begin{aligned}\binom{n}{\lfloor n/2 \rfloor - k} &= \binom{n}{\lfloor n/2 \rfloor} \prod_{i=0}^{k-1} \frac{\lfloor n/2 \rfloor - i}{n + 1 - (\lfloor n/2 \rfloor - i)} \\ &\leqslant \binom{n}{\lfloor n/2 \rfloor} \prod_{i=0}^{k-1} \left(1 - \frac{k}{\lceil n/2 \rceil + k + 1 - i}\right) \leqslant \binom{n}{\lfloor n/2 \rfloor} \left(1 - \frac{k}{\lceil n/2 \rceil + k + 1}\right)^k.\end{aligned}$$

The lemma now follows from the inequality $1 - x \leqslant \mathrm{e}^{-x}$. □

3 The number of matroids

3.1 A procedure for encoding non-bases

The bound in Theorem 1 is based on a procedure to construct a concise description of stable sets in a regular graph. Bounding the number of such concise descriptions immediately gives an upper bound on the number of stable sets in a regular graph. The procedure was adapted from a procedure described by Alon, Balogh, Morris, and Samotij in [ABMS14], who cite Kleitman and Winston [KW82] as the original source. A detailed account of the procedure for counting stable sets can be found in the recent survey paper by Samotij [Sam14].

In [BPvdP14], this idea was combined with local covers to construct a concise description of matroids. In particular, it was shown that any matroid can be described by a stable set (in the Johnson graph) of non-bases S, a collection of flats covering all non-bases in $S \cup N(S)$, and a (relatively short) list of all the non-bases that are not yet covered. By bounding the number of possibilities for each of these sets, one obtains an upper bound on the number of matroids.

The bound that was obtained in [BPvdP14] is dominated by the list of non-bases that are not yet covered. This seems to be wasteful: if we can take into account more information about this list, we may be able to obtain stronger bounds. In the current section, we extend the encoding procedure to capture more information about this list of non-bases.

We will analyse the number of matroids on a fixed ground set $E = [n]$ of rank r. In what follows, we will assume that $r \leqslant n/2$, and we will use $G = J(E, r)$ as a shorthand notation.

Procedure 1 The procedure for encoding matroids

Input: Matroid $M = (E, \mathcal{B})$ of rank r on n elements, $r \leqslant n/2$
Output: $(S, \mathcal{Z}, A, T)$

```
Set $A \leftarrow V(G)$, $S \leftarrow \emptyset$, $\mathcal{Z} \leftarrow \emptyset$          ▷ $G = J(E, r)$
while $|A| > \alpha_{n,r}N$ or $\Delta(G[A]) \geqslant r$ do          ▷ $\alpha_{n,r}N$ is the Hoffman bound
    Pick the first vertex $X$ in the canonical ordering of $A$
    if $X$ is dependent in $M$ then
        Set $S \leftarrow S \cup \{X\}$, $A \leftarrow A \setminus (\{X\} \cup N(X))$
        Set $\mathcal{Z} \leftarrow \mathcal{Z} \cup \mathcal{Z}_X$          ▷ $\mathcal{Z}_X$ defined in Lemma 4
    else
        Set $A \leftarrow A \setminus \{X\}$
    end if
end while

Set $A' \leftarrow \{X \in A \setminus \mathcal{B} : \exists e \in X, f \in E \setminus X$ such that
        $R_X(e) \cap A = C_X(f) \cap A = \emptyset, \quad R_X(e) \cap \mathcal{B}, C_X(f) \cap \mathcal{B} \neq \emptyset\}$
Set $T \leftarrow$ maximal stable set in $G[A']$
```

We will further fix a linear ordering $\leqslant_G$ on the vertices of G. By the *canonical ordering* on $A \subseteq V(G)$, we refer to the following procedure to order the set A linearly. Let v be the vertex with maximum degree in $G[A]$; if there are multiple such v, then we choose the one that is minimal with respect to $\leqslant_G$. Call v the first vertex in the canonical ordering, and apply iteratively to $A \setminus \{v\}$.

Throughout the main loop of the procedure, two disjoint vertex sets are maintained: a set S of selected vertices (which will grow during execution), and a set A of available vertices (which will shrink). The main loop runs as long as $|A|$ is sufficiently large, or $G[A]$ contains a vertex of sufficiently high degree.

3.2 Analysis of the procedure

Throughout this section, we write

$$N = \binom{n}{r}, \quad \text{resp.} \quad d = r(n-r), \tag{12}$$

for the number of vertices, resp. degree of the Johnson graph $J(n, r)$. We will also abbreviate

$$\alpha_{n,r} = \frac{d}{d+\lambda} = \frac{1}{n-r+1}, \tag{13}$$

and

$$\sigma_{n,r} = \frac{\ln(d+1)}{d+\lambda} = \frac{\ln(r(n-r)+1)}{r(n-r+1)}. \tag{14}$$

The quantities $\alpha_{n,r}$ and $\sigma_{n,r}$ will play the same role as α and σ in the statement of Theorem 1. In particular, $\alpha_{n,r}N$ is the Hoffman bound.

Claim 6. *Upon termination of the procedure, we have* $|S| \leqslant \left(\sigma_{n,r} + \frac{1}{r+1}\alpha_{n,r}\right)\binom{n}{r}$*,* $|A| \leqslant \alpha_{n,r}\binom{n}{r}$*, and* $|\mathcal{Z}| \leqslant 2|S|$*.*

For future reference, we record

$$\tilde{\sigma}_{n,r} = \sigma_{n,r} + \frac{1}{r+1}\alpha_{n,r}. \tag{15}$$

Proof. Note that the sets S, $\mathcal{Z}$ and A only change during execution of the while loop.

In each traversal, $|A|$ decreases, and the procedure does not stop before $|A| \leqslant \alpha_{n,r}N$, thus proving the bound on $|A|$.

As A only gets smaller in each traversal, execution of the while loop falls apart into two stages: during the first stage, $|A| > \alpha_{n,r}N$, while during the second stage, $|A| \leqslant \alpha_{n,r}N$ and $\Delta(G[A]) \geqslant r$.

The first stage was analysed in [BPvdP14, Lemma 16], where it was shown that during this stage at most $\sigma_{n,r}N$ vertices are added to S. At the start of the second stage, A contains at most $\alpha_{n,r}N$ vertices. Throughout this stage, each element that is added to S has degree at least r in A, as they are the first vertex in the canonical ordering on A. So each time a vertex is added to S during the second stage, at least $r+1$ vertices are removed from A. Hence, during the second stage, at most $\frac{1}{r+1}\alpha_{n,r}N$ vertices are added to S. Combining the bounds on the number of elements added to S during both stages, we obtain the bound on $|S|$.

The set $\mathcal{Z}$ is only extended when a vertex is added to S. Each time this happens, at most two new flats are introduced to $\mathcal{Z}$ by Lemma 4, so $|\mathcal{Z}| \leqslant 2|S|$. □

The following claim is obvious.

Claim 7. *Upon termination of Stage 2, all vertices in* A *have at most* $r-1$ *neighbours in* A*.*

Claim 8. $S \cup T$ *is a stable set in* $J(n,r)$*.*

Proof. First, S is a stable set, as each time a vertex is added to S, its neighbours are deleted from A, and hence never will be considered again. By the same argument, no element in A has a neighbour in S. By construction, the set T is stable, and as $T \subseteq A$, it follows that $S \cup T$ is stable. □

Claim 9. *Upon termination of the procedure, the triple* (S,T,A) *is completely determined by* $S \cup T$*.*

Proof. Let $(S,\mathcal{Z},A,T)$ be the output of the procedure when it is run on input $M = ([n],\mathcal{B})$ of rank r. As $S \cup T$ is a stable set in $J(n,r)$, the set $\binom{[n]}{r} \setminus (S \cup T)$ is the set of bases of a sparse paving matroid M'. Let $(S',\mathcal{Z}',A',T')$ be the result of running the procedure on input M'. We claim that $S' = S$ and $A' = A$. This implies the lemma, as $T = (S \cup T) \setminus S'$.

The claim follows, as the order in which r-sets are considered in the main loop is deterministic, and depends only on the choice that is made in each traversal. These choices are the same in both instances. By construction of M', we have X dependent (when encoding M) if and only if $X \in S$, which is equivalent to saying that X is dependent (when encoding M'). The final equivalence follows, since vertices in T come after vertices in S in the canonical ordering in each traversal. □

Let K be the set of non-bases of a matroid. Note that throughout the procedure,

$$S \subseteq K \subseteq S \cup N(S) \cup A \tag{16}$$

is maintained as an invariant.

If the encoding procedure indeed constructs a concise description of matroids, it should be possible to reconstruct K from the output of the procedure. Non-bases in $S \cup N(S)$ are easily recognised, as they are covered by $\mathcal{Z}$. On the other hand, recognising non-bases in A is a bit more involved.

The following claim is the engine of the corresponding decoding procedure. It roughly states that if $X \in A$, then X being dependent is completely determined by $(S, \mathcal{Z}, T)$.

Claim 10. *Let M be a matroid without loops and coloops, and let $(S, \mathcal{Z}, A, T)$ be the output of the procedure on input M. Let $X \in A$, then X is dependent in M if and only if*

(i) there exists $x \in X$ such that $R_X(x)$ is disjoint from A and all $Y \in R_X(x)$ are dependent; or

(ii) there exists $y \in E \setminus X$ such that $C_X(y)$ is disjoint from A and all $Y \in C_X(y)$ are dependent; or

(iii) $X \in T$; or

(iv) X has a neighbour $Y = X - x + y$ in T, and there are $e \in Y$, $f \in E \setminus Y$ with the property that both $R_Y(e)$ and $C_Y(f)$ are disjoint from A, both contain a basis, and at least one of $Y - e + x \in R_Y(e)$ and $Y - y + f \in C_Y(f)$ is dependent.

Proof. To prove sufficiency, suppose that $X \in A$. If (i) holds, then X must be dependent, for if X would be independent and each $Y \in R_X(x)$ is dependent, then x is a coloop, which contradicts our assumption on M. Similarly, if (ii) holds, then X must be dependent, for if X would be independent and each $Y \in C_X(y)$ would be dependent, then y is a loop, again contradicting our assumption on M.

If (iii) holds, then X must be dependent, as by construction T contains only non-bases.

Finally, suppose that (iv) holds. Let $Y = X - x + y$ be an element of T neighbouring X satisfying the properties mentioned in (iv). As Y is dependent, and has an independent neighbour, it must have rank $r - 1$. It follows that there is a unique circuit C contained in Y, and a unique cocircuit D disjoint from Y. As $C_Y(f)$ contains an independent set, we have

$$C = \{g \in Y : Y - g + f \text{ is independent}\},$$

and since $R_Y(e)$ contains an independent set, we have

$$D = \{h \in E \setminus Y : Y - e + h \text{ is independent}\}.$$

Note that $X = Y - y + x$ is independent if and only if C is not contained in X (so $y \in C$, or equivalently $Y - y + f$ is independent), and X is not disjoint from D (so $x \in D$, or equivalently $Y - e + x$ is independent). Taking the contrapositive, we find that X is dependent if and only if at least one of $Y - y + f$ or $Y - e + x$ is dependent.

It remains to prove necessity. Let us assume that $X \in A$ is dependent. If neither (i) nor (ii) holds, then $X \in A'$. As T is a maximal stable set in $G[A']$, we have $A' \subseteq T \cup N(T)$, so either $X \in T$, or X has a neighbour in T. In the former case, we have (iii), and we are done. So assume that X has a neighbour in T. Call this neighbour $Y = X - x + y$. As $T \subseteq A'$, there must exist $e \in Y$ and $f \in E \setminus Y$, so that $R_Y(e)$ and $C_Y(f)$ are disjoint from A, and both contain an independent set of the matroid. Now we use again that X is dependent if and only if at least one of $Y - y + f$ or $Y - e + x$ are dependent. □

The following lemma shows that the procedure can be used to construct an alternative description for a matroid, provided that the matroid has no loops or coloops.

Lemma 11. *Let $M = ([n], \mathcal{B})$ be a matroid of rank r that does not have any loops or coloops. Let $(S, \mathcal{Z}, A, T)$ be the output of the procedure on input M. Then M can be reconstructed from n, r, and $(S \cup T, \mathcal{Z})$.*

Proof. First, it follows from Claim 9 that the pair $(S \cup T, \mathcal{Z})$ actually contains the more detailed information $(S, \mathcal{Z}, A, T)$. The matroid M can be reconstructed from n, r, and $(S, \mathcal{Z}, A, T)$ if for each $X \in \binom{[n]}{r}$ it can be decided whether X is dependent or independent, based on the available information alone.

Let $K = \binom{[n]}{r} \setminus \mathcal{B}$ be the set of non-bases in M. Recall that throughout the procedure, (16) is maintained as an invariant. It can be verified from (S, A) if X is in $S \cup N(S) \cup A$. If it is not, then X must be independent. So we can suppose that $X \in S \cup N(S) \cup A$.

If $X \in S \cup N(S)$, then X is dependent if and only if it is covered by some flat in $\mathcal{Z}$.

Hence, for each r-set that is not in A, we can reconstruct whether it is dependent from $(S, \mathcal{Z})$ alone. It remains to identify the non-bases in A. By Claim 10, we only need to verify, for each $X \in A$, if at least one of (i)–(iv) holds. Verification of each of these items depends only on $(S, \mathcal{Z}, T)$:

(i) By Claim 9, the set A is completely determined by S, so for each neighbour of X, checking if it belongs to A can be done if only S is available. If Y is a neighbour of X that is not in A, then either it is not in $S \cup N(S)$, in which case Y must be independent – or it is in $S \cup N(S)$, in which case dependency of Y depends only on $(S, \mathcal{Z})$.

(ii) Similar.

(iii) $X \in T$ obviously depends only on T.

(iv) For each neighbour Y of X, $Y \in T$ depends only on T. By Claim 9, for each neighbour of Y, determining whether it is in A depends only on S. If the neighbour is not contained in A, then we can deduce from $(S, \mathcal{Z})$ whether it is independent or not, by the previous part of this proof. □

3.3 Bounding the number of matroids

Let us write m'_n (resp. $m'_{n,r}$) for the number of matroids (resp. rank-r matroids) on groundset $[n]$ that do not contain loops or coloops.

Lemma 12. $m'_{n,r} \leqslant s_{n,r} \sum_{k=0}^{2\lceil\tilde{\sigma}_{n,r}N\rceil} \binom{2^n(n+1)}{k}$.

Proof. In view of Lemma 11, a matroid $M \in \mathbb{M}_{n,r}$ without loops or coloops can be described by a pair $(U, \mathcal{Z})$, in which U is a stable set of $J(n,r)$, and $\mathcal{Z} \subseteq \{(F, r_M(F)) : F \in \mathcal{F}(M)\}$. By Claim 6, we can assume that $|\mathcal{Z}| \leqslant 2\lceil\tilde{\sigma}_{n,r}N\rceil$.

Note that $m'_{n,r}$ is at most the number of pairs $(U, \mathcal{Z})$, and a bound on the number of such pairs follows immediately from the following two observations. First, as U is a stable set in $J(n,r)$, it can be chosen in at most $i(J(n,r)) = s_{n,r}$ ways. Second, as $\mathcal{Z}$ is a subset of $2^{[n]} \times \{0,1,\ldots,n\}$ of cardinality at most $2\lceil\tilde{\sigma}_{n,r}N\rceil$, it can be chosen in at most $\sum_{k=0}^{2\lceil\tilde{\sigma}_{n,r}N\rceil} \binom{2^n(n+1)}{k}$ ways. □

The sum of binomial coefficients in the statement of Lemma 12 can be bounded by an application of (10),

$$\sum_{k=0}^{2\lceil\tilde{\sigma}_{n,r}N\rceil} \binom{2^n(n+1)}{k} \leqslant \left(\frac{\mathrm{e}2^n(n+1)}{2\left\lceil\tilde{\sigma}_{n,r}\binom{n}{r}\right\rceil}\right)^{2\left\lceil\tilde{\sigma}_{n,r}\binom{n}{r}\right\rceil} \leqslant \left(\frac{\mathrm{e}2^n(n+1)^3}{4\binom{n}{r}}\right)^{2\left\lceil\tilde{\sigma}_{n,r}\binom{n}{r}\right\rceil}. \tag{17}$$

The next lemma gives a bound that is uniform in r.

Lemma 13. *There exists $c > 0$ such that for sufficiently large n, and $0 \leqslant r \leqslant n/2$,*

$$\sum_{k=0}^{2\lceil\tilde{\sigma}_{n,r}N\rceil} \binom{2^n(n+1)}{k} \leqslant \exp_2\left(c\frac{\log^2 n}{n^2}\binom{n}{\lfloor n/2\rfloor}\right).$$

Proof. We may assume that $n \geqslant 4$. Recall that $\tilde{\sigma}_{n,r} = \sigma_{n,r} + \frac{1}{r+1}\alpha_{n,r}$. It follows from [BPvdP14, Lemma 19] that $\sigma_{n,r}\binom{n}{r} \leqslant \frac{8\ln n}{n^2}\binom{n}{\lfloor n/2\rfloor}$. It is easily verified that $\frac{1}{r+1}\alpha_{n,r}\binom{n}{r}$ is increasing in $r = 0,1,\ldots,\lfloor n/2\rfloor$, so

$$\frac{1}{r+1}\alpha_{n,r}\binom{n}{r} \leqslant \frac{4}{n^2}\binom{n}{\lfloor n/2\rfloor} \leqslant \frac{3\ln n}{n^2}\binom{n}{\lfloor n/2\rfloor}.$$

Combining these bounds, we obtain $\tilde{\sigma}_{n,r}\binom{n}{r} \leqslant \frac{11\ln n}{n^2}\binom{n}{\lfloor n/2\rfloor}$. An application of (10) gives

$$\max_{0\leqslant r\leqslant n/2} \sum_{k=0}^{2\lceil\tilde{\sigma}_{n,r}N\rceil} \binom{2^n(n+1)}{k} \leqslant \exp_2\left(\frac{22\ln n}{n^2}\binom{n}{\lfloor n/2\rfloor}\log\left(\frac{\mathrm{e}2^n(n+1)}{\frac{22\ln n}{n^2}\binom{n}{\lfloor n/2\rfloor}}\right)(1+o(1))\right).$$

The lemma follows as the expression in the logarithm is $n^{O(1)}$ by (11). □

Theorem 14. $\log m'_n \leqslant (1+o(1))\log s_n$ *as $n \to \infty$.*

Proof. Upon combining the upper bound on $m'_{n,r}$ with the fact that $s_{n,r} \leqslant s_n$, we find by Lemma 13 that

$$m'_{n,r} \leqslant s_n \exp_2\left(c\frac{\log^2 n}{n^2}\binom{n}{\lfloor n/2\rfloor}\right),$$

at least for $r \leqslant n/2$. By duality, this same bound holds for $m'_{n,n-r}$. As $m'_n = \sum_r m'_{n,r}$, we obtain

$$\log m'_n \leqslant \log(s_n) + \log(n+1) + c\frac{\log^2 n}{n^2}\binom{n}{\lfloor n/2\rfloor}.$$

The theorem now follows since $\log s_n \geqslant \frac{1}{n}\binom{n}{\lfloor n/2\rfloor}$. □

It was shown in [MNWW11, Theorem 2.3] that almost every matroid has no loops or coloops, i.e. that

$$m'_n = m_n(1 - o(1)) \qquad \text{as } n \to \infty. \tag{18}$$

This immediately implies the following corollary.

Corollary 15. *$\log m_n = (1 + o(1)) \log s_n$ as $n \to \infty$.*

3.4 The rank of a typical matroid

It was shown in [LOSW13, Corollary 2.3] that for all $\varepsilon > 0$, the fraction of matroids having rank r in the range $(1/2 - \varepsilon)n < r < (1/2 + \varepsilon)n$ tends to 1. In this section, we prove a slightly stronger statement.

Theorem 16. *For all $\beta > \sqrt{\frac{\ln 2}{2}} = 0.5887\cdots$ the fraction of matroids with $n/2 - \beta\sqrt{n} < r < n/2 + \beta\sqrt{n}$ tends to 1 as $n \to \infty$.*

Proof. By duality, it suffices to show that $r(M) \leqslant n/2 - \beta\sqrt{n}$ for a vanishing fraction of matroids. In view of (18), it suffices to restrict our attention to matroids without loops or coloops. In fact, we will show that for β sufficiently large

$$\frac{1}{m'_n}\sum_{r=0}^{\lfloor n/2-\beta\sqrt{n}\rfloor} m'_{n,r} \to 0 \qquad \text{as } n \to \infty.$$

The term $m'_{n,r}$ can be bounded by combining Lemma 12 with the upper bound on $s_{n,r} = i(J(n,r))$ provided by Theorem 1. As $m'_n \geqslant s_{n,\lfloor n/2\rfloor} \geqslant 2^{\frac{1}{n}\binom{n}{\lfloor n/2\rfloor}}$, it follows that

$$\begin{aligned}\frac{1}{m'_n}\sum_{r=0}^{\lfloor n/2-\beta\sqrt{n}\rfloor} m'_{n,r} &\leqslant 2^{-\frac{1}{n}\binom{n}{\lfloor n/2\rfloor}}\sum_{r=0}^{\lfloor n/2-\beta\sqrt{n}\rfloor}\left[\sum_{s=0}^{\lceil\sigma_{n,r}\binom{n}{r}\rceil}\binom{\binom{n}{r}}{s}2^{\alpha_{n,r}\binom{n}{r}}\sum_{k=0}^{2\lceil\tilde{\sigma}_{n,r}\binom{n}{r}\rceil}\binom{2^n(n+1)}{k}\right]\\ &\leqslant 2^{-\frac{1}{n}\binom{n}{\lfloor n/2\rfloor}}\sum_{r=0}^{\lfloor n/2-\beta\sqrt{n}\rfloor}\left[2^{\alpha_{n,r}\binom{n}{r}}\left(\sum_{k=0}^{2\lceil\tilde{\sigma}_{n,r}\binom{n}{r}\rceil}\binom{2^n(n+1)}{k}\right)^2\right],\end{aligned}$$

which, by Lemma 13 and the inequality $\alpha_{n,r} \leqslant \frac{2}{n}$ is at most

$$\begin{aligned}2^{-\frac{1}{n}\binom{n}{\lfloor n/2\rfloor}} \sum_{r=0}^{\lfloor n/2-\beta\sqrt{n}\rfloor} &\exp_2\left(\frac{2}{n}\binom{n}{r} + 2c\frac{\log^2 n}{n^2}\binom{n}{\lfloor n/2\rfloor}\right)\right) \\ &\leqslant n\exp_2\left(-\frac{1}{n}\binom{n}{\lfloor n/2\rfloor}(1-o(1)) + \frac{2}{n}\binom{n}{\lfloor n/2-\beta\sqrt{n}\rfloor}\right)\right) \\ &\qquad\qquad \leqslant n\exp_2\left(-\frac{1}{n}\binom{n}{\lfloor n/2\rfloor}\left(1-2\mathrm{e}^{-2\beta^2}-o(1)\right)\right),\end{aligned}$$

where the final equality follows from Lemma 5 with $k = \beta\sqrt{n}$. For $\beta > \sqrt{\frac{\ln 2}{2}}$, the right-hand side tends to 0, thus concluding the proof. □

We like to remark at this point that Theorem 16 can be proved using the bound on $m_{n,r}$ that was derived in the proof of [BPvdP14, Theorem 3]. But since that paper does not have a separate lemma which we can refer to here, we make use of Lemma 12 of the present paper.

4 Some remaining problems

4.1 Counting the matroids without circuit-hyperplanes

It was conjectured in [BPvdP14, Conjecture 22] that

$$\lim_{n\to\infty} \frac{\#\{M \in \mathbb{M}_n : M \text{ has no circuit-hyperplanes }\}}{m_n} = 0 \tag{19}$$

We have tried to prove this conjecture by analysing the behaviour of (variants of) our compression algorithm on matroids without circuit-hyperplanes, so far without any success. We like to encourage the reader to give it another try, since it does feel as if we are very close. The key to proving a sufficient bound seems to be the behaviour of the algorithm when picking $T \subseteq A$. A more intelligent decoder may be able to reconstruct the matroid from a much sparser set T. Note that the asymptotic upper bounds on $|S|$ and $|\mathcal{Z}|$ do not get worse essentially if we insist that the algorithm continues its main loop while $\Delta(G[A]) \geqslant \epsilon r$, for any fixed $\epsilon > 0$.

4.2 The rank of a typical matroid

It was conjectured by Mayhew, Newman, Welsh, and Whittle in [MNWW11, Conjecture 1.10] that asymptotically almost all matroids have rank between $(n-1)/2$ and $(n+1)/2$. By Theorem 16, almost all matroids on n elements have rank between $n/2 - \beta\sqrt{n}$ and $n/2+\beta\sqrt{n}$ when $\beta > \sqrt{\frac{\ln 2}{2}}$. At the heart of the argument lies the fact that for sufficiently large k the ratio $m_{n,\lfloor n/2\rfloor-k}/m_{n,\lfloor n/2\rfloor}$ tends to 0, using that

$$\frac{m_{n,\lfloor n/2\rfloor-k}}{m_{n,\lfloor n/2\rfloor}} \leqslant \frac{s_{n,\lfloor n/2\rfloor-k}}{s_{n,\lfloor n/2\rfloor}} \cdot 2^{O\left(\frac{\log^2 n}{n^2}\binom{n}{\lfloor n/2\rfloor-k}\right)} \qquad \text{as } n \to \infty.$$

We see no better way to bound the factor $s_{n,n/2-k}/s_{n,n/2}$ than by combining the lower bound of Graham and Sloane and the upper bound from [BPvdP14], and we ask if a direct comparison would perhaps be possible. For example, if it could be shown that for some $c, c' > 0$ we have

$$c - c'\frac{\log^2 n}{n} \leqslant \frac{\log s_{n,r}}{\frac{1}{n}\binom{n}{r}} \leqslant c + c'\frac{\log^2 n}{n}$$

for all $r \approx n/2$, the above analysis would imply that asymptotically almost all matroids have rank between $n/2 - \beta' \log n$ and $n/2 + \beta' \log n$ for a sufficiently large constant β'.

It would be a good start if we could argue that asymptotically all *sparse paving* matroids on n elements have a rank between $(n-1)/2$ and $(n+1)/2$. Presently we cannot even prove the unimodality of $s_{n,r}$, i.e. that $s_{n,r} \leqslant s_{n,r'}$ for all $0 < r < r' \leqslant n/2$.

4.3 Comparing $\log s_{n,r}$ and $\log m_{n,r}$ for small r

By Theorem 16, most matroids have rank close to $n/2$. Consequently, in the derivation of our main result, Corollary 15, we are mainly interested in comparing $m_{n,r}$ to $s_{n,r}$ for $r \approx n/2$. In this regime, the upper bound

$$\log m_{n,r} \leqslant \log s_{n,r} + \log \sum_{k=0}^{2\lceil \tilde{\sigma}_{n,r} N\rceil} \binom{2^n(n+1)}{k} \leqslant \log s_{n,r} + 2\left\lceil \tilde{\sigma}_{n,r}\binom{n}{r}\right\rceil \log\left(\frac{\mathrm{e}2^n(n+1)^3}{4\binom{n}{r}}\right)$$

from Lemma 12 combined with (17) suffices to show that $(\log s_{n,r})/(\log m_{n,r}) \to 1$ as $n \to \infty$.

On the other hand, when r (or by duality $n-r$) is small, this upper bound will not suffice, as for $r \leqslant c \log n$ (where c is sufficiently small), we have

$$\log \sum_{k=0}^{2\lceil \tilde{\sigma}_{n,r} N\rceil} \binom{2^n(n+1)}{k} \geqslant \binom{n}{r},$$

which follows immediately from (10).

We note at this point that the best asymptotic lower bound on $s_{n,r}$ for fixed r is strictly better than the general bound of $\log s_{n,r} \geqslant \alpha(J(n,r)) \geqslant \binom{n}{r}/n$. The better bound follows from the work of Keevash [Kee14, Thm 6.8] who has recently proved an asymptotic estimate of the number of Steiner systems with parameters (n, r, q), where r and q are fixed:

$$\log S(n,r,q) = (1+o(1))\binom{q}{r}^{-1}\binom{n}{r}(q-r)\log n \qquad \text{as } n \to \infty.$$

Since each Steiner system with parameters $(n, r-1, r)$ uniquely determines a stable set of the Johnson graph $J(n,r)$, Keevash's estimate gives an asymptotic lower bound of

$$\log s_{n,r} \geqslant \log S(n, r-1, r) = (1+o(1))r^{-1}\binom{n}{r-1}\log n \qquad \text{as } n \to \infty. \tag{20}$$

The lower bound (20) is tight up to the $(1+o(1))$-factor. This follows from the trivial upper bound

$$\log s_{n,r} = \log i(J(n,r)) \leqslant \log \sum_{k=0}^{\alpha_{n,r}\binom{n}{r}} \binom{\binom{n}{r}}{k} = (1+o(1))\alpha_{n,r}\binom{n}{r}\log n \qquad \text{as } n \to \infty$$

and the identity $r^{-1}\binom{n}{r-1} = \alpha_{n,r}\binom{n}{r}$.

We speculate that for fixed r, the ratio $s_{n,r}/m_{n,r}$ tends to 0 as $n \to \infty$, while at the same time, $(\log s_{n,r})/(\log m_{n,r})$ tends to 1.

4.4 Comparing s_n and m_n

The matroid encoding procedure described in this paper returns a pair $(U, \mathcal{Z})$, where U is a stable set and $\mathcal{Z}$ is a partial flat cover. Note that we could have been more economical in constructing $\mathcal{Z}$. For example, when the encoding procedure is run on a sparse paving matroid, that matroid can be reconstructed from U alone. This suggests that in many cases $\mathcal{Z}$ can be pruned, while the original matroid can still be recovered.

The following information-theoretic perspective may be useful. Suppose that $(U, \mathcal{Z})$ is chosen uniformly at random from the set of all possible outputs of the encoding procedure when its input is some matroid in $\mathbb{M}_{n,r}$ without loops or coloops. By the chain rule for the entropy function $\mathcal{H}(\cdot)$,

$$\log m'_{n,r} = \mathcal{H}(U, \mathcal{Z}) = \mathcal{H}(U) + \mathcal{H}(\mathcal{Z} \mid U)\,.$$

The term $\mathcal{H}(U)$ is at most $\log s_{n,r}$. Currently, we naively bound $\mathcal{H}(\mathcal{Z} \mid U)$ by the logarithm of the number of possible partial flat covers, which does not take into account the mutual information between U and $\mathcal{Z}$. We expect that $\mathcal{H}(\mathcal{Z} \mid U)$ is much smaller than our naive bound, but presently do not see a more careful analysis of this quantity.

References

[ABMS14] Noga Alon, József Balogh, Robert Morris, and Wojciech Samotij. Counting sum-free sets in Abelian groups. *Isr. J. Math.*, 199(1):309–344, 2014.

[BCH73] John A. Blackburn, Henry H. Crapo, and Denis A. Higgs. A catalogue of combinatorial geometries. *Math. Comput.*, 27:155–166, 1973.

[BPvdP14] Nikhil Bansal, Rudi Pendavingh, and Jorn van der Pol. On the number of matroids. To appear in *Combinatorica*, 2014.

[CR70] Henry H. Crapo and Gian-Carlo Rota. *On the foundations of combinatorial theory: Combinatorial geometries.* The M.I.T. Press, Cambridge, Mass.-London, preliminary edition, 1970.

[GS80] R. L. Graham and N. J. A. Sloane. Lower bounds for constant weight codes. *IEEE Trans. Inform. Theory*, 26(1):37–43, 1980.

[Juk11] Stasys Jukna. *Extremal combinatorics: with applications in computer science*. Springer, 2011. Second edition.

[Kee14] Peter Keevash. The existence of designs. Preprint, 2014. arXiv:1401.3665

[KW82] Daniel J. Kleitman and Kenneth J. Winston. On the number of graphs without 4-cycles. *Discrete Math.*, 41(2):167–172, 1982.

[LOSW13] Lisa Lowrance, James Oxley, Charles Semple, and Dominic Welsh. On properties of almost all matroids. *Adv. Appl. Math.*, 50(1):115–124, 2013.

[MNWW11] Dillon Mayhew, Mike Newman, Dominic Welsh, and Geoff Whittle. On the asymptotic proportion of connected matroids. *European J. Combin.*, 32(6):882–890, 2011.

[MW13] Dillon Mayhew and Dominic Welsh. On the number of sparse paving matroids. *Adv. Appl. Math.*, 50(1):125–131, 2013.

[Oxl11] James Oxley. *Matroid theory*, volume 21 of *Oxford Graduate Texts in Mathematics*. Oxford University Press, Oxford, second edition, 2011.

[PW71] M. J. Piff and D. J. A. Welsh. The number of combinatorial geometries. *Bull. London Math. Soc.*, 3:55–56, 1971.

[Sam14] Wojciech Samotij. Counting independent sets in graphs. Preprint, 2014. arXiv:1412.0940v1