Randomness in classes of matroids

Will Critchlow

VICTORIA UNIVERSITY OF WELLINGTON
Te Whare Wānanga o te Ūpoko o te Ika a Māui

School of Mathematics, Statistics
and Operations Research
Te Kura Mātai Tatauranga, Rangahau Pūnaha

A thesis

submitted to the Victoria University of Wellington

in fulfilment of the requirements for the degree of

Doctor of Philosophy

in Mathematics.
Victoria University of Wellington

2018

Abstract

This thesis is inspired by the observation that we have no good random model for matroids. That stands in contrast to graphs, which admit a number of elegant random models. As a result we have relatively little understanding of the properties of a "typical" matroid. Acknowledging the difficulty of the general case, we attempt to gain a grasp on randomness in some chosen classes of matroids.

Firstly, we consider sparse paving matroids, which are conjectured to dominate the class of matroids (that is to say, asymptotically almost all matroids would be sparse paving). If this conjecture were true, then many results on properties of the random sparse paving matroid would also hold for the random matroid. Sparse paving matroids have limited richness of structure, making counting arguments in particular more feasible than for general matroids. This enables us to prove a number of asymptotic results, particularly with regards to minors.

Secondly, we look at Graham-Sloane matroids, a special subset of sparse paving matroids with even more limited structure - in fact Graham-Sloane matroids on a labelled groundset can be randomly generated by a process as simple as independently including certain bases with probability $\frac{1}{2}$. Notably, counting Graham-Sloane matroids has provided the best known lower bound on the total number of matroids, to log-log level. Despite the vast size of the class we are able to prove severe restrictions on what minors of Graham-Sloane matroids can look like.

Finally we consider transversal matroids, which arise naturally in the context of other mathematical objects - in particular partial transversals of set systems and partial matchings in bipartite graphs. Although transversal matroids are not in one-to-one correspondence with bipartite graphs, we shall link the two closely

enough to gain some useful results through exploiting the properties of random bipartite graphs. Returning to the theme of matroid minors, we prove that the class of transversal matroids of given rank is defined by finitely many excluded series-minors. Lastly we consider some other topics, including the axiomatisability of transversal matroids and related classes.

Acknowledgements

First and foremost I would like to thank my supervisor Dillon. Your advice, support and oversight has made a vital contribution to this thesis, and I appreciate that you gave me so much freedom to work on problems that I found interesting.

I'd like to thank my VUW officemates past and present - Andrew, Emma, Susan, Gio, and Meenu - for being so kind, fun and forgiving. I'd also like to thank all the graduate students at Vic, as well as my other friends here, for helping to make the past three years a very rewarding experience outside the office. If you are a non-matroid theorist who has bothered to read this, this probably includes you - but especially Bex, Jasmine, Matt and Del. Thank you for your willingness to provide enjoyable distractions whenever you could.

Finally, I'd like to thank Mum, Dad, Lorna, Auntie Jill and Jane for all their support, and for putting up with my prolonged absence. I miss you guys!

Contents

Chapter 1

Introduction

The initial inspiration for this thesis lay in the observation that, in contrast to graph theory's many well understood models for a random graph, there exist no useful models for a random matroid, and matroid theory is correspondingly poor in asymptotic-type results. By this we mean results of the kind "asymptotically almost all matroids have property $\mathcal{P}$".

Definition 1.0.1. Let $\mathcal{M}$ be a minor-closed class of matroids and $\mathcal{M}_n$ the members of $\mathcal{M}$ on n elements. A property $\mathcal{P}$ is said to hold for *asymptotically almost all* matroids in $\mathcal{M}_n$ if the following holds:

$$\lim_{n \to \infty} \frac{\#\{M \in \mathcal{M}_n : M \text{ has property } \mathcal{P}\}}{\#\mathcal{M}_n} = 1$$

Some progress on asymptotic questions has been made recently, especially by Pendavingh and van der Pol who showed (amongst other things) that asymptotically almost all matroids are k-connected [23], are not a truncation [23], contain a large uniform minor [23], and contain minors isomorphic to each of a small collection of matroids [24].

Although much of the work of this thesis is aimed at establishing asymptotic results, it has grown to include a wider survey of three particular classes of matroids, each of which was believed more accessible to random models than the class of matroids at large. Within these classes we attempt to find asymptotic results as well as answer other questions about the nature of the classes.

In Chapter 2, we encounter sparse paving matroids. In terms of asymptotic results, this may well turn out to be the most important subclass of matroids, due to the following conjecture:

Conjecture 1.0.2 (Mayhew, Newman, Whittle, Welsh [17])**.** *Asymptotically almost all matroids are sparse paving.*

This conjecture, which is weakly supported by computations of the numbers of matroids and sparse paving matroids on up to 10 elements [1] [19], would imply that statements of the kind described earlier could be implied simply by a proof that "asymptotically almost all sparse paving matroids have property $\mathcal{P}$". We believe such proofs may be easier than their more general counterparts, as the structure of sparse paving matroids has a more straightforward characterisation than that of general matroids. In particular, their only non-trivial circuits are all rank r, and these uniquely define the matroid. This means that a rank-r sparse paving matroid on n elements corresponds to an r-uniform hypergraph on n vertices. Moreover, there is a (less obvious) correspondence between these circuits and vertices of the well-studied Johnson graph. Thus we can hope to draw on existing graph theoretical results and techniques when studying sparse paving matroids.

The results in Chapter 2 are largely partial answers to the question "which matroids are minors of asymptotically almost all sparse paving matroids".

This builds upon work by Pendavingh and van der Pol [24] and we are able to resolve the question for some specific matroids as well as classes of matroids. Restricting to matroids of fixed rank, even more extensive results are possible.

Chapter 3 focuses on a subclass of sparse paving matroids, the Graham-Sloane matroids. These arise from a particular construction which Graham and Sloane [6] developed based on an idea of Knuth [11]. Graham Sloane matroids are noteworthy for providing (at least as far as *log-log* level) the best known lower bound on m_n, the number of matroids on n elements (that said, we show that Graham-Sloane matroids are heavily outnumbered by sparse paving matroids at an absolute level).

Much of our interest in Graham-Sloane matroids is derived from the fact that, although in some sense a very large subclass of sparse paving matroids, they are quite different from other matroids in that class. Graham-Sloane matroids are distinguished by their highly symmetric construction which places substantial restrictions on which substructures can appear. In particular we shall show that (in some sense) almost all Graham-Sloane matroids are non-minors of almost all other Graham-Sloane matroids. Moreover, at least at the fixed rank case, asymptotically almost all sparse paving matroids are not a minor of any Graham-Sloane matroid.

Finally, Chapter 4 explores transversal matroids. These are the matroids which arise from partial transversals of set systems; equivalently, from partial matchings in bipartite graphs. Although each matroid may have more than one "presentation" in the world of bipartite graphs, we can place some controls on this in cases of fixed rank and / or principal transversal matroids.

This means that we can exploit properties of random bipartite graphs to give asymptotic bounds on the nature of these matroids.

As well as asymptotic questions, we explore some more general aspects of transversal matroids, in particular taking an interest in the behaviour of minimally non-transversal matroids. This includes a proof that the number of minimally non-transversal matroids of fixed rank is bounded. Minimally non-transversal matroids are also used to demonstrate the non-axiomatisability of transversal matroids in simple logical language, work which builds on an equivalent result for representable matroids proved by Mayhew, Newman and Whittle [20].

The three chapters are somewhat independent from one another, although the chapter on Graham-Sloane matroids reprises some material from the chapter on sparse paving matroids and we also find sparse paving matroids to be a convenient source of examples in the chapter on transversal matroids. Nonetheless there are common themes running through the work. Large parts of all three chapters relate to the original motivation of finding asymptotic results. Another theme is that of utilising correspondences between classes of matroids and classes of graphs. A third is questions of minor inclusion and excluded or forbidden minors, which arise in all three chapters. Throughout all chapters, much of the work is combinatorial in nature and we shall frequently express problems in probabilistic terms.

The majority of work contained in this thesis is original; non-original results are quoted with attribution and without proof. The only exception to this is a few preliminary lemmas that were so fundamental and intuitive that finding an attribution seemed impossible; in these cases we have generally included our own improvised proof for completeness. Of the non-original

work presented, the most important sources are Pendavingh and van der Pol [24] [23] [25], followed by Mayhew, Newman, Welsh and Whittle [17] [20]. Those authors are also the source of considerable inspiration for the work in this thesis. We also borrowed some material and substantial guidance from Bonin's survey of transversal matroids [2].

1.0.1 Preliminaries and Notation

Familiarity with the fundamentals of matroid theory as outlined in Oxley [21] is assumed, and as far as possible we follow the notation and terminology defined therein. We also assume familiarity with basic graph theory and probability. Several simple but potentially non-obvious results on probability are detailed in Section 4.1.1 and are primarily used in our work on transversal matroids.

1.0.1.1 Special notation

In line with other literature, we write $\mathbb{M}_n$ to represent the class of (labelled) matroids of size n - which we shall always assume to have groundset $[n] = \{1, \ldots, n\}$ - and $\mathbb{M}_{n,r}$ the class of matroids of size n and rank r. We shall also use $m_n = |\mathbb{M}_n|$ and $m_{n,r} = |\mathbb{M}_{n,r}|$

Similar notation is used for subclasses of matroids. For example, we use $\mathbb{S}_n$ to denote the class of sparse paving matroids of size n and $\mathbb{S}_{n,r}$ the class of sparse paving matroids of size n and rank r. Likewise $s_n = |\mathbb{S}_n|$ and $s_{n,r} = |\mathbb{S}_{n,r}|$. Notation for other classes of matroids will be introduced as it is needed.

We write $[n]$ for $\{1, 2, \ldots, n\}$ and $[n]^{(r)}$ for the r-subsets of $[n]$ (and in general, $X^{(r)}$ for the r-subsets of a set X). We also occasionally use the notation

$[n, m] := \{n, n+1, \dots, m\}$

In graph theoretical contexts, we use $\Gamma(X)$ to denote the closed neighbourhood of a vertex x, and $\Gamma(X)$ to denote the closed neighbourhood of a vertex set X (the more standard $N(x), N(X)$ is avoided due to similarities with some matroid notation). *Closed* in this context means simply that we include $x \in \Gamma(x)$ and $X \subseteq \Gamma(X)$.

We use $\sqcup$ to indicate disjoint union.

1.0.1.2   Special terminology and usages

A good deal of this thesis deals with asymptotic results - here we shall give a formal, general definition of asymptoticness. Let $\mathcal{M}$ denote a class of matroids and $\mathcal{M}_n$ be the restriction of $\mathcal{M}$ to matroids on groundset $[n]$. We say that a property $\mathcal{P}$ holds for *asymptotically almost all* matroids in $\mathcal{M}$ if

$$\lim_{n \to \infty} \frac{\#\{M \in \mathcal{M}_n : M \text{ has property } \mathcal{P}\}}{\#\mathcal{M}_n} = 1$$

We note that asymptotic results in this thesis are split between results that we shall loosely describe as being "general" and results that we describe as "fixed rank". The latter describes cases where $\mathcal{M}$ is restricted to matroids of some fixed rank r. Results of this type are less powerful than general results, but often easier to obtain. At the end of Chapter 2 we shall devote some time to discussing the merits of this type of result and the challenge in extending such results to the general case.

A collection of matroids on n elements is *labelled* if we distinguish between matroids based on the labelling of their elements, and *unlabelled* if we do

not. In the latter case matroids that were isomorphic up to labelling would be considered identical.

To some extent we shall rely on context to make it obvious whether we are dealing with labelled or unlabelled matroids. Individual matroids and finite collections of matroids are almost always unlabelled, that is to say we only care about their structure. On the other hand when we deal with classes and subclasses of matroids these always consist of labelled matroids. Note however that we shall only be considering classes which are closed under isomorphism.

We use the term *non-basis* to describe a dependent set of cardinality equal to the rank of its matroid. In the case of sparse paving matroids, non-bases are equivalent to *circuit-hyperplanes* (but we shall mostly avoid using the latter term). Note that a matroid is uniquely defined by its groundset and non-bases (this is equivalent to the more familiar fact that a matroid is uniquely defined by its groundset and bases). We shall often use $\mathcal{L}(M)$ to denote the collection of non-bases of M.

When discussing what graph theorists typically call *independent sets* of graphs, we shall use the alternative term *stable set*. This avoids confusion with the matroid theoretical definition of *independent set* and is consistent with other matroid theoretical publications.

Definition 1.0.3 (Stable set)**.** Let G be a graph and $V(G)$ be its vertex set. A set $A \subseteq V(G)$ is *stable* if no pair of elements of A are joined by an edge of G.

Finally, another recurring theme of this thesis is matroid minors. Unless otherwise stated we are interested only in unlabelled minors, so by an "N-minor" we really mean a minor isomorphic to N. We shall use the notation

$M \succeq N$ to mean "M contains a minor isomorphic to N". We make a distinction between the commonly (and often equivalently) used terms *forbidden minor* and *excluded minor.*

Definition 1.0.4 (Forbidden minor)**.** Let $\mathcal{M}$ be a class of matroids. A matroid N is said to be a *forbidden minor* of $\mathcal{M}$ if and only if N is not a minor of any matroid in $\mathcal{M}$.

Definition 1.0.5 (Minor-closed class)**.** A class of matroids $\mathcal{M}$ is *minor-closed* if for every matroid $M \in \mathcal{M}$ and every minor N of M, N is also in $\mathcal{M}$. A matroid N is said to be an *excluded minor* of $\mathcal{M}$ if N is minimally non-$\mathcal{M}$ - that is to say, N is not a member of $\mathcal{M}$, but every proper minor of N is.

We shall also use the term *excluded series-minors,* defined as above with "series-minor" everywhere replacing "minor".

Chapter 2

Sparse paving matroids

2.1 Introduction and definitions

2.1.1 Motivation and background

As mentioned in the introduction, a key motivation for working with sparse paving matroids is the conjecture of Mayhew, Newman, Welsh and Whittle [17] that asymptotically almost every matroid is sparse paving. This conjecture formalises speculation that has existed at least since 1970 when Crapo and Rota [3] suggested "paving matroids may predominate in any asymptotic enumeration of matroids" (note this would imply sparse paving matroids predominate). The conjecture has been cited in a number of subsequent research papers typically expressing some degree of confidence in the conjecture. Much of the credibility of the conjecture is based on the catalogue of matroids of small size. Blackburn, Crapo and Higgs [1] initially catalogued matroids on up to 8 elements, in which we can observe a significant increase in the proportion of paving matroids between $n = 7$ and $n = 8$; even at this

stage the case is substantially strengthened by the results of Mayhew and Royle [19] who used a computer program to determine the number of sparse paving matroids in $\mathbb{M}_n$, for n up to 9 (with partial results for $n = 10$). Some of these results are shown in Table 2.1.

Table 2.1: Comparison of m_n, the number of matroids on n elements, with s_n, the number of sparse paving elements on n elements. As n becomes larger it appears sparse paving matroids start to dominate.

n	s_n	m_n	% sparse paving
5	6	38	15.8%
6	13	98	13.3%
7	35	306	11.4%
8	342	1724	19.8%
9	226864	383172	59.2%

Further motivation comes from a second conjecture of Mayhew, Newman, Welsh and Whittle [17].

Conjecture 2.1.1. *Let N be a sparse paving matroid. Asymptotically almost all matroids have a minor isomorphic to N.*

This conjecture, if combined with a positive resolution of Conjecture 1.0.2, would essentially answer the question of which matroids are minors of asymptotically almost all matroids. Specifically:

Conjecture 2.1.2. *Let N be a matroid. Then N is a minor of asymptotically almost all matroids if, and only if, N is sparse paving.*

If N is not sparse paving, it cannot be a minor of any sparse paving matroid (since that class is minor-closed) and so would not be a minor of asymptoti-

cally almost all matroids (in fact, it would be a non-minor of asymptotically almost all matroids).

Previous progress [24] [23] has directly tackled the question of which matroids are minors of asymptotically almost all matroids. We shall instead aim for intermediate results, by using counting arguments to construct proofs that certain matroids and classes of matroids are minors of asymptotically almost all sparse paving matroids. These are shown in Figure 2.1 and include:

- Uniform matroids

- Sparse paving matroids of rank r whose non-bases all intersect in a single set of $r - 2$ elements.

- Sparse paving matroids whose non-bases are pairwise disjoint

- W^3

If Conjecture 1.0.2 were to be proven, then these results would automatically extend to asymptotically almost all matroids.

Figure 2.1: Some matroids which we show are contained as minors in asymptotically almost all sparse paving matroids. Leftmost, matroids with all non-bases pairwise disjoint; centre, matroids in which all non-bases intersect in some $(r - 1)$-set; rightmost, the rank-3 whirl, W^3.

In the final part of the chapter we turn our attention to fixed-rank sparse paving matroids, and use a different (Ramsey-theoretical) approach to prove

an analogous fixed-rank result for a larger class of matroids, that in which every non-basis contains an element unique to that non-basis. Note that amongst fixed-rank matroids, sparse paving matroids do not predominate (see Pendavingh and van der Pol [25]) so to extend these results to all matroids, we would first need to extend to sparse paving matroids of general rank (a problem which we shall discuss later).

2.1.1.1   *Counting matroids in minor-closed classes* (Pendavingh and van der Pol's work)

The most important previous work in this area is the paper of Pendavingh and van der Pol [24] which shows that the following matroids are minors of asymptotically almost all matroids: $U_{2,k}$ (for $k \geq 2$), $U_{3,6}$, P_6, Q_6, R_6. These are shown in Figure 2.2. Later work by the same authors [23] proved that the same result holds for all uniform matroids.
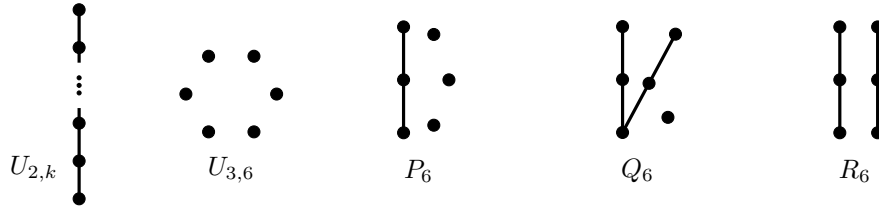

Figure 2.2: Matroids which Pendavingh and van der Pol showed to be minors of asymptotically almost all matroids. The general $U_{t,k}$ is not shown; that result was proven later and using different methods.

Notably the above list includes all the rank-3 matroids on 6 elements except one, the rank-3 whirl W^3, which Pendavingh and van der Pol were unable to resolve. Although we do not answer the question of whether W^3 is contained in asymptotically almost all matroids, it is amongst the matroids which we shall show to be a minor of asymptotically almost all sparse paving

matroids.

The methods in Pendavingh and van der Pol's paper were largely not reused by this thesis. In particular they relied on the notion of *cover complextiy*. This is the minimal size of *flat cover* - a collection of flats identifying the non-bases of a matroid. By bounding the number of matroids with low cover complexity it was possible to show that the number of matroids without an N-minor was asymptotically small, where N is one of the matroids shown in Figure 2.2.

The final section of the paper included some results relating to Graham-Sloane matroids, which Pendavingh and van der Pol suggest might provide some evidence against Conjecture 2.1.2. We shall discuss this part of the paper in Chapter 3 and in this case we shall be able to make use of the same methods as Pendavingh and van der Pol to greatly extend their results. However, in doing so we also demonstrate that Graham-Sloane matroids, whilst very interesting in their own right, are perhaps less relevant to Conjecture 2.1.2 (and other asymptotic results for sparse paving matroids) than previously thought.

Before we introduce any of our own work we shall thoroughly review the definition of sparse paving matroids, and in particular the important correspondence between the non-bases of sparse paving matroids and stable sets in the Johnson Graph.

2.1.2 Preliminaries for sparse paving matroids

We recall first the definition of a sparse paving matroid.

Definition 2.1.3 (Paving, Sparse paving matroid). A matroid is called *paving* if all its circuits are of rank at least $r(M)$. A matroid is *sparse paving* if both it and its dual are paving.

For the purposes of this work the above is not the most useful definition of a sparse paving matroid. In fact we shall demonstrate an equivalence that enables us to think of sparse paving matroids in a different way, as structures of circuit-hyperplanes.

First we note that the circuit-hyperplanes of a sparse paving matroid are precisely the *non-bases* of M (the sets of size $r(M)$ not contained in the set $\mathcal{B}(M)$ of bases of M). In general we shall refer to non-bases rather than circuit-hyperplanes. We use $\mathcal{L}(M)$ to denote the collection of non-bases of M.

The following two lemmas show that sparse paving matroids can be alternatively defined through a very simple condition on the intersection of non-bases. This equivalent definition was realised by Knuth [11]. However the proofs offered here are our own.

Lemma 2.1.4. *Given any two non-bases C, D of a sparse paving matroid M, we have $|C \cap D| \leq r - 2$.*

In other words, the non-bases cannot intersect in $r - 1$ elements. The result is a consequence of the augmentation axiom of matroid theory (**I3**):

I3:for I_1, I_2 independent in M, with $|I_2| < |I_1|$, there is always some $e \in I_1 \setminus I_2$ such that $I_2 \cup \{e\}$ is independent in M.

Proof. We prove the contrapositive: assume M has two non-bases C, D with

$C = X \cup \{c\}, D = X \cup \{d\}$, for some $X \in [n]^{(r-1)}$. Note that since M is paving, X must be independent. We proceed to show M cannot be sparse paving. If M is paving then there must be some element $e \in E \setminus X$ such that $X \cup \{e\}$ is independent (and therefore a basis) in M. This is true because otherwise we can take I_2 to be X and I_1 to be any basis of M, and we have a contradiction of **I3**. So such an e exists, and some basis $B = X \cup \{e\}$. But now consider the set $Z = E \setminus (X \cup \{c, d\})$ in the dual matroid M^*. Since M^* is paving then Z must be independent, since it has cardinality $r(M^*) - 1$. But then we can take $I_2 = Z, I_1 = B^c$ and we again have a contradiction of **I3**: $B^c \setminus Z = \{c, d\}$ but $Z \cup \{c\}$ and $Z \cup \{d\}$ are both non-bases of M^*. So M^* cannot be paving and M cannot be sparse paving. $\qquad \square$

More important still is the following lemma: that such a construction of non-bases always defines a valid sparse paving matroids.

Lemma 2.1.5. *Given a set E, let $\mathcal{L}$ be a collection of subsets of E such that for any $L_1, L_2 \in \mathcal{L}$ we have $|L_1| = |L_2|$ and $|L_1 \cap L_2| \leq |L_1| - 2 = |L_2| - 2$. Then $\mathcal{L}$ defines the non-bases of a sparse paving matroid.*

Proof. It is easiest here to work with the axiomatisation of matroids by their bases. This defines a matroid to be a pair $(E, \mathcal{B})$, where $\mathcal{B}$ is a non-empty collection of subsets of E meeting the following conditions (**B2**):

B2: for distinct $B_1, B_2 \in \mathcal{B}$, and $x \in B_1 - B_2$, we have some $y \in B_2 - B_1$ such that $((B_1 - x) \cup y) \in \mathcal{B}$

Let the sets in $\mathcal{L}$ have cardinality r. We claim that $\mathcal{B} = E^{(r)} \setminus \mathcal{L}$ satisfies the above. Suppose it doesn't. Then there is some $B_1, B_2 \in \mathcal{B}, x \in B_1 - B_2$

so that for any $y \in B_2 - B_1, ((B_1 - x) \cup y) \in \mathcal{L}$. Now we cannot have $|B_2 - B_1| = 1$, since then for any $y \in B_2 - B_1, ((B_1 - x) \cup y) = B_2$. So we can pick two elements y_1, y_2 from $B_2 - B_1$. But now we have $((B_1 - x) \cup y_1), ((B_1 - x) \cup y_2) \in \mathcal{L}$ - and this contradicts the condition that any two members of $\mathcal{L}$ must differ in more than one element. $\qquad\square$

We are now at the point where we may think of sparse paving matroids as defined simply by a collection of subsets of $E^{(r)}$ whose pairwise intersections have cardinality at most $r - 2$. We can now note that these collections are in direct correspondence with stable sets of the *Johnson Graph* $J(n, r)$, which we shall now introduce.

Definition 2.1.6 (Johnson Graph)**.** The Johnson Graph $J(n, r)$ is the graph whose vertices are subsets of $[n]$ of size r, with two vertices S_1, S_2 joined if and only if $|S_1 \cap S_2| = r - 1$.

It follows from Lemmas 2.1.4 and 2.1.5 that stable sets of vertices in $J(n, r)$ correspond to the collections of non-bases of sparse paving matroids on groundset $[n]$. This link was first noted in a paper of Piff and Welsh [26].

2.2   Maximum extensions

The most important result of this section is an asymptotic lower bound on the number of non-bases in a sparse paving matroid of fixed rank. As well as being an interesting result in its own right, this bound also proves of great use in the following section, where we consider questions of minor inclusion. An analogous result for the general case already exists courtesy of Pendavingh and van der Pol [23].

To tackle this problem we find it convenient to work in the world of stable sets in the Johnson Graph. Our strategy will be to define an extension on stable sets of the Johnson Graph and show that each stable set has some unique maximal stable extension. Since we have lower bounds on the cardinality of a maximal stable set, we can apply well-known combinatorial lemmas to achieve lower bounds on the size of asymptotically almost all stable sets of $J(n, r)$.

We consider stable sets in the Johnson graph $J(n, r)$. Let $\mathcal{I}(n, r)$ be the collection of stable sets in $J(n, r)$, which corresponds to sparse paving matroids of groundset $[n]$ and rank r. We use v_X to mean the vertex in the Johnson graph representing the set X.

Definition 2.2.1 (Extension, Maximal extension)**.** Given a stable set I, let J be an *extension* of I if J is stable, and I is contained in J. Such an extension is *maximal* if J is maximal with respect to its stableness.

Definition 2.2.2 (Mother)**.** We call an extension J of a stable set I the *mother* of I if $|J|$ is maximum with respect to all possible extensions and, amongst all sets meeting this criterion, J appears first under a lexicographical ordering of stable sets. We use $m(I)$ to denote the mother of I.

Note that the nature of the ordering on stable sets does not matter, with a lexicographical ordering being chosen purely for familiarity. The important point is that we have some consistent "tie-breaker" on stable sets of equal size, in order that every stable set have a unique mother.

The following two theorems give basic lower bounds on the sizes of maximum and maximal extensions.Recall that for X a collection of vertices in $J(n, r)$, we shall use $\Gamma(X)$ to denote the closed neighbourhood of X.

Theorem 2.2.3. *Let I be a stable set in $J(n, r)$ and let $U = V(J(n, r)) \backslash \Gamma(I)$. Then for any maximum-sized extension J of I, we have $|J| \geq |I| + \frac{|U|}{n}$*

Since we are dealing with maximum-sized, rather than maximal, extensions, the proof simply requires us to construct such an extension.

Proof. Let the vertices in our Johnson graph be labelled by elements of $[n]^{(r)}$. For each vertex $v \in J(n, r)$, let X_v be the r-set represented by v. Similarly for each r-set $X \in [n]^{(r)}$, let v_X be the vertex representing X in $J(n, r)$.

For $i = 0, \ldots, n-1$, let $U_i = \{u \in U : \sum_{x \in X_u} x \equiv i \mod n\}$ - that is to say, all the vertices representing r-sets whose elements sum to i. Clearly the U_i form a partition of U, and so for some i we must have $|U_i| > \frac{|U|}{n}$. But now we claim $I \cup U_i$ is a stable set. Clearly it is true that no member of U_i is adjacent to an member of I, and no two members of I are adjacent to each other - these both follow from the definition of the problem. So it remains to show that no two members of U_i are adjacent. But adjacent vertices of $J(n, r)$ represent r-sets differing only be the replacement of a single element. Say then that this is the case for some $u, v \in U_i$, i.e. that $X_v = X_u \setminus e \cup f$. We know $e \neq f$; without loss of generality we say $e < f$. Then we have $0 \leq e < f \leq n-1$, and so $1 \leq f - e \leq n-1$. But also

$$f - e \cong \sum_{x \in X_v} x - \sum_{x \in X_u} x \equiv i - i \equiv 0 \mod n$$

contradicting $1 \leq f - e \leq n-1$. So we cannot have u and v contained in the same part U_i. $\qquad \square$

The construction used in this proof is actually a generalisation of the construction of Graham-Sloane matroids, which we shall discuss in depth later. In particular, where $I = \emptyset$, the above argument becomes simply the Graham-Sloane construction that gives us the current best lower bound on $m_{n,r}$.

Theorem 2.2.4. *Let I be a stable set in $J(n,r)$ with $|I| = k, |m(I)| = k + t$ Then for any maximal extension J of I, we have $|J| \geq k + \frac{t}{r}$*

Proof. Let $U = m(I) \setminus I$. U is obviously a stable set and $|U| = t$. We claim that at most r vertices of U can be in the neighbourhood of a given vertex $v \in J(n,r)$. This follows fairly easily: v corresponds to a set of r elements, and all of its neighbours represent r-sets formed by replacing a single element in X_v. This provides a natural partitioning of $\Gamma(v)$ into cliques:

$$\Gamma(v) = v \cup \bigcup_{x \in v} \Gamma_x(v)$$

where $\Gamma_x(v)$ consists of all vertices u such that $X_u = X_v \setminus x \cup y : y \notin X_v$. But $\Gamma_x(v)$ is a clique, so its members are pairwise adjacent and no stable set in $\Gamma(v)$ can have cardinality greater than r).

This leads to a proof of our lemma as follows: given any vertex $v \in J(n,r)$, we must have $|\Gamma(v) \cap U| \leq r$, else the above is contradicted. But now assume we have a maximal extension J of I, with $|J| < k + \frac{t}{r}$. But now we have

$$\Gamma(J) \cap U \leq r|J \setminus I| < r\frac{t}{r} = t$$

So there is at least one vertex of U not contained in the neighbourhood of J, contradicting the claimed maximality of J. $\qquad\square$

We now turn towards finding an asymptotic lower bound for the sizes of a stable set in $J(n,r)$. To do so we shall first identify a lower bound on the size of maximal stable sets, and then show that asymptotically almost all stable sets are at least almost half the size of their maximum-sized extensions.

Lemma 2.2.5. *For any fixed $\epsilon > 0, r \in \mathbb{N}^+$ all stable sets of $J(n,r)$ that are maximal with respect to extension must have cardinality at least $\frac{1}{rn}\binom{n}{r}$.*

Proof. As before, assume the vertices in our Johnson graph are labelled by elements of $[n]^{(r)}$, with X_v being the r-set represented by v, and v_X being the vertex representing X in $J(n,r)$.

Suppose I is maximal with respect to extension. Then $\Gamma(I) = V(J(n,r))$. But for any $v \in J(n,r)$ the neighbourhood of v contains only $r(n-r)+1$ vertices. And $|\Gamma(I)| \leq \sum_{v \in I} |\Gamma(v)| = (r(n-r)+1)|I|$. So $|I| \geq \frac{1}{r(n-r)+1}\binom{n}{r} \geq \frac{1}{rn}\binom{n}{r}$. $\qquad\square$

Before we can extend this result to an asymptotic lower bound on stable sets, we must introduce the concept of a *shadow* and prove a related lemma.

Definition 2.2.6 (Shadow)**.** For $\mathcal{A} \subseteq X^{(r)}$, the *shadow* $\partial\mathcal{A}$ of $\mathcal{A}$ is the set system $\{B \in X^{(r-1)} : B \cup \{i\} \in \mathcal{A}, \text{ for some } i \notin B\}$.

Lemma 2.2.7 (Local LYM)**.** *For any set E and set family $\mathcal{A} \subseteq E^{(r)}$,*

$$\frac{|\mathcal{A}|}{\binom{n}{r}} \leq \frac{|\partial\mathcal{A}|}{\binom{n}{r-1}}$$

This is so-named for its relation to the famous LYM identity (independently

proven by Lubell, Yamamoto and Meshalkin).

Proof. Each member of $\mathcal{A}$ corresponds to r members of $\partial\mathcal{A}$. Each member of $\partial\mathcal{A}$ corresponds to at most $n + 1 - r$ members of $\mathcal{A}$. So

$$\frac{|\mathcal{A}|}{|\partial\mathcal{A}|} \leq \frac{n+1-r}{r} = \frac{\binom{n}{r}}{\binom{n}{r-1}}$$

and the result follows via some trivial rearranging. $\qquad\square$

Finally, we need a result that formalises the fact that most subsets of a maximal stable set are about half its size. The following lemma expresses a well-known property of large sets.

Lemma 2.2.8. *Let $\delta, \epsilon > 0$. Let X be a set of cardinality n and Y be drawn uniformly at random from the subsets of X. Then there exists $N := N(\epsilon, \delta)$ such that if $n > N$, we have*

$$\Pr(|Y| \in ((0.5 - \delta)n, (0.5 + \delta)n)) > 1 - \epsilon$$

There are many proofs, most somewhat dull and involving consideration of binomial coefficients. For example:

Proof. Taken as given that binomial coefficients are unimodal peaking at $n/2$, we can see that the number of subsets of X whose cardinality falls outside the given interval is less than $(1 - 2\delta)n\binom{n}{\lfloor(0.5-\delta)n\rfloor}$. In turn,

$$n\binom{n}{\lfloor(0.5-\delta)n\rfloor} \leq n\binom{n}{\lfloor n/2 \rfloor}\left(\frac{\lfloor n/2\rfloor(\lfloor n/2\rfloor - 1)\cdots\lfloor(0.5-\delta)n\rfloor}{\lfloor n/2\rfloor(\lfloor n/2\rfloor + 1)\cdots\lfloor(0.5+\delta)n\rfloor}\right)$$

$$\leq n \binom{n}{\lfloor n/2 \rfloor} \left(\frac{0.5 + \delta/2}{0.5 - \delta/2} \right)^{n\delta/2} = o\left(\binom{n}{\lfloor n/2 \rfloor} \right)$$

Clearly the number of subsets of X with cardinality inside our interval is at least $\binom{n}{\lfloor n/2 \rfloor}$, so we are done. $\qquad\square$

We can now finally prove an asymptotic lower bound on the cardinality of stable sets:

Theorem 2.2.9. *Asymptotically almost all stable sets in the Johnson Graph $J(n,r)$ have cardinality at least $(1-o(1))\frac{1}{2rn} \binom{n}{r}$. Equivalently, asymptotically almost all sparse paving matroids of rank r have at least $(1-o(1))\frac{1}{2rn} \binom{n}{r}$ non-bases.*

The result still requires a little effort to obtain. Intuitively, it might seem obvious - most subsets of a maximal stable set are about half its size - but we have the complication that any given set has many maximal extensions. This means we are required to assign each stable set I a unique maximal extension - the *mother* $m(I)$ as defined in Definition 2.2.2.

Proof of Theorem 2.2.9. Assume we are drawing I uniformly at random from the stable sets of $J(n,r)$.

Define $\mathcal{Y}$ to be the collection of maximal stable sets of $J(n,r)$, and let $\epsilon > 0$. For any stable set J of $J(n,r)$, let

$$p_J := \Pr\left(|I| < (1-\epsilon)\frac{1}{2rn} \binom{n}{r} \Big| m(I) = J \right)$$

We have

$$\Pr\left(|I| < (1-\epsilon)\frac{1}{2rn}\binom{n}{r}\right) = \sum_{J \in \mathcal{Y}} p_J \Pr\left(m(I) = J\right)$$

We shall show that as $n \to \infty$, p_J becomes vanishingly small for any choice of J. Firstly, Bayes's Theorem gives us

$$p_J = \frac{\Pr(|I| < (1-\epsilon)\frac{1}{2rn}\binom{n}{r})\Pr(m(I) = J : |I| < (1-\epsilon)\frac{1}{2rn}\binom{n}{r})}{\Pr(m(I) = J)}$$

Now we are close to applying Local LYM. Whereas in the previous proof we constructed a family of sets each represened by a single vertex in the Johnson graph, now we construct a family each of whose sets comprises r vertices in the Johnson Graph. Specifically, fix a maximal stable set $J \subseteq V(J(n,r))$ and, for $0 \le k \le |J|$, define the family $\mathcal{A}_k$ by

$$\mathcal{A}_k = \{J' \in J^{(k)} : m(J') \ne J\}$$

We observe that $\partial \mathcal{A}_k \subseteq \mathcal{A}_{k-1}$. To see this, suppose $J' \in \mathcal{A}_k$. Then $m(J') > |J|$, under the ordering we are using to define the mother (see Definition 2.2.2). But now suppose $J'' \subseteq J'$. Clearly any extension of J' is also an extension of J'', so $|m(J'')| \ge |m(J')| > |J|)$. Applying this reasoning to members of $\partial \mathcal{A}_k$ tells us they must also be members of $\mathcal{A}_{k-1}$.

Also, by LYM we have $\frac{|\mathcal{A}_k|}{\binom{|J|}{k}} \le \frac{|\partial \mathcal{A}_k|}{\binom{|J|}{k-1}}$ and so combining these two observations we obtain that

$$\frac{|\mathcal{A}_k|}{\binom{|J|}{k}} \le \frac{|\partial \mathcal{A}_k|}{\binom{|J|}{k-1}} \le \frac{|\mathcal{A}_{k-1}|}{\binom{|J|}{k-1}}$$

By recursion of the same argument, it is seen that for any $l < k$, we have

$$\frac{|\mathcal{A}_k|}{\binom{|J|}{l}} \leq \frac{|\mathcal{A}_l|}{\binom{|J|}{l}}.$$

This means that for any cardinality α, and considering I drawn uniformly at random from the subsets of J, we have

$$\frac{\Pr(m(I) = J : |I| < \alpha)}{\Pr(m(I) = J)} \leq 1$$

Since we know the cardinality of J is at least $\frac{1}{rn}\binom{n}{r}$, and this function is unbounded, we may apply Lemma 2.2.8 and (for example by setting $\alpha :=$ $(1 - \epsilon)\frac{1}{2rn}\binom{n}{r}$) we can see for sufficiently large n we must have

$$p_J = \Pr\left(|I| < (1 - \epsilon)\frac{1}{2rn}\binom{n}{r}\Big| m(I) = J\right) \leq \epsilon$$

And since this applies for all choices of J, we see

$$\Pr\left(|I| < (1 - \epsilon)\frac{1}{2rn}\binom{n}{r}\right) = \sum_{J \in \mathcal{Y}} \Pr\left(m(I) = J\right).p_J \leq \epsilon \sum_{J \in \mathcal{Y}} \Pr\left(m(I) = J\right) = \epsilon$$

$\square$

Although this result is sufficient for the purposes of our paper, it would be nice to obtain a uniform bound across all values of r. Formally:

Conjecture 2.2.10. *There exists a constant c such that, for any value of r, when drawing M uniformly at random from $\mathbb{S}_{n,r}$, asymptotically almost certainly M has at least $\frac{c}{n}\binom{n}{r}$ non-bases.*

We lack any compelling evidence for this conjecture, but it is encouraging that, as implied by the proof of Theorem 5.3 in Pendavingh and van der Pol's paper [23], a result of this type holds when r is not fixed but rather $r = \alpha n$,

for some constant $0 < \alpha < 1$. Specifically, if drawing M uniformly at random from $\mathbb{S}_{n,\alpha n}$, asymptotically almost certainly M has at least $\frac{1}{5n}\binom{n}{r}$ non-bases. However this case is fundamentally different to the fixed rank case, in which $r/n \to 0$ as $n \to \infty$.

We can see that one possible proof route runs into trouble. It would be nice to say that the expected number of non-bases in a randomly chosen rank-r, n element matroid M' is greater than the expected number of non-bases in a randomly chosen contraction M/A of a randomly chosen matroid M of much higher rank down to n elements and rank r. And there is some intuitive case for this: in the latter case, an r-set $X \subset E(M/A)$ cannot be a non-basis if $(X \cup A)$ intersects in all but on element with a non-basis of the original M - this "blocks" a lot of potential non-bases in M/A. However in some cases (e.g. where M is Graham-Sloane) we may find that the only non-bases not blocked are those from a single large stable set on $E(M) \setminus A$, which forces the contracted matroid into an efficient structure of non-bases within a single large set - whereas in the case of M', having a free choice of non-bases increases the chances that an inefficient structure will be adopted! If most large matroids do have an efficient, Graham-Sloane type structure, then the conjecture may turn out to be false.

2.3 Minors included in asymptotically almost all sparse paving matroids

In [24], Pendavingh and van der Pol showed that asymptotically almost all matroids contain an N-minor, for N one of the matroids $U_{2,k}, U_{3,6}, P_6, Q_6$ or R_6. In [23] they proved the same for all uniform matroids.

Here we consider a related question.

Question 2.3.1. For which sparse paving matroids N can we say that asymptotically almost all sparse paving matroids contain an N-minor?

Interest in this problem is largely motivated by the conjecture of Mayhew, Newman, Welsh and Whittle [17] that asymptotically almost all matroids are sparse paving. If true, the conjecture would imply that if asymptotically almost all sparse paving matroids have an N-minor, then so must asymptotically almost all matroids.

We note that the methods of Pendavingh and van der Pol in [23] would seem to imply that the result holds for N uniform, although this is not noted. We shall prove the uniform case by a different method, which also show that a number of other matroids are contained as minors in asymptotically almost all sparse paving matroids. As listed previously, these include:

- Sparse paving matroids of rank r whose non-bases all intersect in a single set of $r - 2$ elements.

- Sparse paving matroids whose non-bases are pairwise disjoint

- W^3

We note that the result could surely be extended to many more matroids with simple structures of non-bases, but here we shall only explicitly prove the four above cases.

Additionally, we show some matroids to be contained as minors in asymptotically almost all sparse paving matroids of fixed rank r (where obviously r must be at least the rank of the target minor). This is a weaker statement

but we can prove it for more target minors. Specifically, the statement holds for all sparse paving matroids whose non-bases all contain at least one *loose element* - meaning an element not contained in any other non-bases. We also note one more case where this result holds, but that will be proved in Chapter 3, since it forms part of an important result about Graham-Sloane matroids.

A very useful part of the apparatus used is the fact that asymptotically almost all sparse paving matroids $M \in \mathbb{M}_n$ have at least $(1 + o(1)) \frac{1}{5n} \binom{n}{r(M)}$ non-bases. This result was essentially shown by Pendavingh and van der Pol [23] in their recent paper (included as Theorem 2.3.4 in this thesis).

The following result is also fundamental.

Lemma 2.3.2. *For all $0 < \delta \leq 1$, asymptotically almost all n-element sparse paving matroids have rank r in the interval $I = ((0.5 - \delta)n, (0.5 + \delta)n)$*

A similar result already exists in the general matroid case, and is included in [13].

Proof. Let t_n be the number of sparse paving matroids on n with rank $r \notin I$. The number of rank-r sparse paving matroids is bounded above by the number of subsets of $[n]^{(r)}$ giving

$$t_n < \sum_{r \notin I} 2^{\binom{n}{r}} \leq n \cdot 2^{\binom{n}{\lfloor (0.5 - \delta)n \rfloor}}$$

And we know from the class of Graham-Sloane matroids [6] that

$$s_n > 2^{\frac{1}{n} \binom{n}{\lfloor n/2 \rfloor}}$$

The latter clearly outgrows the former - for example, by considering binomial coefficients as in the proof of Lemma 2.2.8 we get

$$\frac{\log t_n}{\log s_n} = (1+o(1))\frac{\binom{n}{\lfloor(0.5-\delta)n\rfloor}}{\frac{1}{n}\binom{n}{\lfloor n/2\rfloor}} = (1+o(1))n\left(\frac{(\lfloor n/2\rfloor)!(\lceil n/2\rceil)!}{(\lfloor(0.5-\delta)n\rfloor)!(\lceil(0.5+\delta)n\rceil)!}\right)$$

$$= (1+o(1))n\left(\frac{\lfloor n/2\rfloor-1}{\lceil n/2\rceil+1}\right)\left(\frac{\lfloor n/2\rfloor-2}{\lceil n/2\rceil+2}\right)\cdots\left(\frac{\lfloor(0.5-\delta)n\rfloor}{\lceil(0.5+\delta)n\rceil}\right)$$

$$\leq (1+o(1))n\left(\frac{0.5-\delta/2}{0.5+\delta/2}\right)^{\delta n/2}$$

which vanishes as $n \to \infty$. $\square$

As observed earlier (Lemmas 2.1.4 and 2.1.5), the collections of non-bases of sparse paving matroids of size n and rank r are in direct correspondence with the stable sets of the Johnson graph $J(n,r)$. This was originally shown in a paper of Piff and Welsh [26].

We now work towards an asymptotic lower bound on the number of non-bases in a sparse paving matroid.

Let $i(G)$ denote the total number of stable sets in the graph G.

Lemma 2.3.3. $\log i(J(n,r)) > \binom{n}{r}/n$

This result is given by the construction of Graham and Sloane used in [6].

By $i(G,m)$ we shall mean the number of stable sets of G with size no greater than m.

The following result was stated by Pendavingh and van der Pol as a lemma

(Lemma 5.2 of [23]), in the course of showing a related theorem for the general matroid case, but is directly useful to us as we deal with sparse paving matroids.

Theorem 2.3.4 (Pendavingh, van der Pol). *There exists $\epsilon > 0$ such that*
$$\log i\left(J(n,r), \frac{1}{5n}\binom{n}{r}\right) \leq (1 - \epsilon)\log s_n$$

Knowing this, we can make the following statement about sparse paving matroids.

Corollary 2.3.5. *Asymptotically almost all sparse paving matroids $M \in \mathbb{S}_n$ have at least $\frac{1}{5n}\binom{n}{r(M)}$ non-bases.*

Proof. Note that $i\left(J(n,r), \frac{1}{5n}\binom{n}{r}\right)$ counts the matroids in $\mathbb{S}_{n,r}$ which have no more than $\frac{1}{5n}\binom{n}{r(M)}$ non-bases.

Following from Theorem 2.3.4 we get

$$\log \sum_{r=1}^{n} i\left(J(n,r), \frac{1}{5n}\binom{n}{r}\right) \leq (1 - \epsilon)\log s_n + \log n$$

So $\sum_{r=1}^{n} i\left(J(n,r), \frac{1}{5n}\binom{n}{r}\right) = o(s_n)$. $\qquad\qquad \square$

2.3.1 A counting approach to the minor inclusion question

Theorem 2.3.10 will provide a condition under which a matroid N will be contained as a minor in asymptotically almost all sparse paving matroids. Roughly, we will show that if the collection of non-bases of a sparse paving matroid contains many separate substructures isomorphic to the collection

of non-bases of N, then almost certainly at least one such instance forms an N-minor (i.e. by containing no additional non-bases). A consequence of this is that we only need to worry about replicating the non-bases of N, and can ignore the bases (or as they might be better considered in this situation, the non-non-bases!)

We need to make precise what we mean by "separate substructures".

Definition 2.3.6 (Line-structure)**.** A *line-structure* $\mathcal{L}$ of rank r is defined to be a collection $\mathcal{L}$ of equicardinal sets ("lines") obeying the rule that for distinct $L_1, L_2 \in \mathcal{L}$ we have $|L_1 \cap L_2| \leq |L_1| - 2 = |L_2| - 2$. We say a line-structure has *rank* r if its members have cardinality r. By a *substructure* of a line-structure we mean simply a subset of its lines.

Isomorphism between line-structures is in fact defined by isomorphism between hypergraphs (where each line-structure can be viewed as a hypergraph on the union of its elements, with lines as hyperedges). The definition we give is equivalent, albeit expressed somewhat differently, and in terms of line-structures.

Definition 2.3.7. Line-structures $\mathcal{L}_1$ and $\mathcal{L}_2$ are *isomorphic* if there exists a bijection $\Phi : \mathcal{L}_1 \to \mathcal{L}_2$ such that if $\mathcal{L}'$ is some subset of $\mathcal{L}_1$, then

$$\left| \bigcap_{L \in \mathcal{L}'} L \right| = \left| \bigcap_{L \in \mathcal{L}'} \Phi(L) \right|$$

We say a line-structure $\mathcal{L}_1$ *contains a copy of* $\mathcal{L}_2$ (and write $\mathcal{L}_2 \subseteq \mathcal{L}_1$) if a substructure of the lines of $\mathcal{L}_1$ is isomorphic to $\mathcal{L}_2$. Note this can only occur when $\mathcal{L}_1$ and $\mathcal{L}_2$ are of equal rank.

The relation to sparse paving matroids is hopefully obvious: the collection

of non-bases of a sparse paving matroid M forms a line-structure, which we denote by $\mathcal{L}(M)$. Equally, by Lemma 2.1.5, every line-structure $\mathcal{L}$ defines a sparse paving matroid whose non-bases are isomorphic to $\mathcal{L}$, with groundset exactly the union of those non-bases. A sparse paving matroid and its line-structure are depicted in Figure 2.3.

Definition 2.3.8 (Element-disjoint)**.** line-structures $\mathcal{L}_1$ and $\mathcal{L}_2$ are *element-disjoint* if no element appears in both. In other words, the unions of the elements in their lines are disjoint:

$$\left(\bigcup_{L \in \mathcal{L}_1} L \right) \cap \left(\bigcup_{L \in \mathcal{L}_2} L \right) = \emptyset$$

As already noted, we shall be using "contains a copy" in the sense of line substructures; in other words, by "$\mathcal{L}(M/A)$ contains a copy of $\mathcal{L}$" we mean that $\mathcal{L}(M/A)$ contains a substructure $\mathcal{L}'$ isomorphic to $\mathcal{L}$. We note that this is **not** equivalent to M/A containing $\mathcal{L}$ as a minor, since $\mathcal{L}(M/A)$ might contain additional lines on the elements of $\mathcal{L}'$ (equivalently, M/A contains additional non-bases).

Definition 2.3.9 (Abundant)**.** We say that a line-structure $\mathcal{L}$ of rank r is *abundant* in sparse paving matroids if, for any $m \in \mathbb{N}$, the following holds: drawing M uniformly at random from $\mathbb{S}_n$, asymptotically almost certainly there exists a set $A \subseteq [n]$, with $|A| = r(M) - r$, such that $\mathcal{L}(M/A)$ contains m element-disjoint copies of $\mathcal{L}$.

Theorem 2.3.10. *Let N be a fixed sparse paving matroid of rank r. If $\mathcal{L}(N)$ is abundant in sparse paving matroids, then asymptotically almost all sparse*
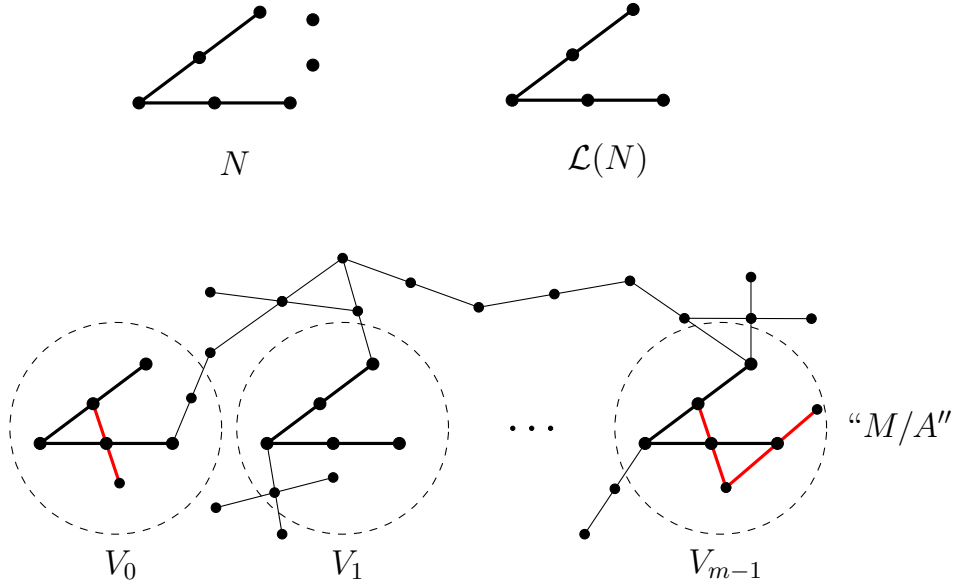
Figure 2.3: A rank-3 matroid N on 7 elements and its line-structure $\mathcal{L}(N)$. Below is a sketch depicting a hypothetical line-structure (not realistic) of a large rank-3 matroid whose non-bases contain m copies of $\mathcal{L}(N)$. This matroid represents a possible M/A (as described in the proof of Theorem 2.3.11), where we have identified disjoint sets $V_0, \cdots V_{m-1}$ such that $|V_i| = 7 = |E(N)|$ and $\mathcal{L}((M/A)|V_i)$ contains a copy of $\mathcal{L}(N)$. If any $\mathcal{L}((M/A)|V_i)$ contains no lines other than the lines in its $\mathcal{L}(N)$-copy, then we have $M = T_i(M)$ (as defined in proof of Theorem 2.3.11) and $(M/A)|V_i = (T_i(M)/A)|V_i \cong N$. In other words, M contains an N-minor. In the above picture this is indeed the case for $i = 1$: the only lines wholly contained within V_1 are the two lines of the $\mathcal{L}(N)$-copy.

paving matroids contain N as a minor.

It is easier to prove directly the following theorem, of which Theorem 2.3.10 is an immediate consequence.

Let N be a fixed sparse paving matroid of rank r_N and let $\mathbb{S}_{n,m,N}$ denote the collection of matroids $M \in \mathbb{S}_n$ for which there exists $A \subset [n], |A| = r(M) - r_N$ such that $\mathcal{L}(M/A)$ contains m element-disjoint copies of $\mathcal{L}(N)$.

Theorem 2.3.11. *Let N be a fixed sparse paving matroid and let M be a matroid drawn uniformly from $\mathbb{S}_{n,m,N}$.*

For any $\epsilon > 0$, there exists $m(\epsilon)$ such that if $m > m(\epsilon)$, then, with probability at least $(1 - \epsilon)$, M will contain N as a minor.

Proof. Assume without loss of generality that $\epsilon < 1$.

We aim to assign to M a series of parent matroids $T_0, \ldots, T_{m-1} \in \mathbb{S}_{n,m,N}$, so that every choice of M has precisely one choice of parent series. We shall show that every parent contains an N-minor, and that asymptotically almost certainly M is equal to one of its parents.

Let r_N denote the rank of N, $n_N := |E(N)|$ the number of elements of N and $r_d = r(M) - r_N$.

Fix a total ordering $+(n, r_d)$ on $[n]^{(r_d)}$. Write $A >^+ B$ if A is above B in this ordering.

Now take some $A \in [n]^{(r_d)}$ and consider the matroid $M \setminus A$. Imagine taking a collection

$$U = \{(U_0, \mathcal{L}_0), \ldots, (U_{m-1}, \mathcal{L}_{m-1})\}$$

where the U_i denote a collection of pairwise disjoint n_N subsets of $[n] \setminus A$ and the $\mathcal{L}_i$ denote m pairwise element-disjoint copies of the line-structure $\mathcal{L}(N)$ on $[n]^{(r_N)}$ such that, for each pair $(U_i, \mathcal{L}_i)$, the lines of $\mathcal{L}_i$ are contained in $U_i^{(r_N)}$. We define $\mathcal{U}_{n,m,N}$ to be the universe of all possible collections of this type. Note that if $M \in \mathbb{S}_{n,m,N}$, then there exists A such that $\mathcal{L}(M/A)$ contains a collection $U \in \mathcal{U}_{n,m,N}$.

Fix some total ordering $*(n, m, \mathcal{L}(N))$ on the set $\mathcal{U}_{n,m,\mathcal{L}(N)}$. For ease we say that $U >^* W$ if U is above W in this ordering.

The reason for all this careful preparation will be to ensure each matroid has a single well defined series of parents, which means we avoid duplicate counting in the later steps of our proof. Note we can be as careless as we like over the actual orderings: any total ordering will do (although it is probably most intuitive to imagine that the orderings are of a lexicographical nature).

Now, we know there exists some $A \in [n]^{(r_d)}$ such that $\mathcal{L}(M/A)$ contains m disjoint copies of $\mathcal{L}(N)$. We can assume without loss of generality that A be the maximal such set under our ordering relation $>^+$. Now (assuming that $n(M) - r_d > mn_N$), we also have some $V \in \mathcal{U}_{n,m,N}$ contained in M/A. Let V be the maximal such collection under $>^*$ and denote its members (the n_N-subsets) as $(V_0, \mathcal{L}_0), \ldots, (V_{m-1}, \mathcal{L}_{m-1})$, ordered lexicographically.

The goal is to show that asymptotically almost certainly there is some $0 \le i \le m-1$ such that $\mathcal{L}((M/A)|V_i) = \mathcal{L}_i$ - if this is the case, then $((M/A)|V_i) \cong N$, forming an N-minor in M. The sketch in Figure 2.3 loosely illustrates this strategy; the proof that now follows requires a little more rigour than is included there!

With some care we define

$$\Sigma_i = \{A \cup S : S \in V_i^{(r_N)}\}$$

Each Σ_i represent a collection of possible non-bases and provide a simple condition for M to contain an N-minor - we simply require one of the Σ_i to contain no non-bases of M other than those of form $A \cup l : l \in \mathcal{L}_i$.

We now can set the *i-parent* $T_i(M)$ of M to be the matroid formed from M by relaxing all non-bases in Σ_i except those which are part of $\mathcal{L}_i$ (in Figure 2.3, these non-bases - after contraction by A - are the red lines within V_i).

The *i*-parent of a matroid has some significant properties. Let $T = T_i(M)$ for some choice of M. Then:

Claim 2.3.11.1. $T_i(M) \succeq N$

Proof. $(T_i(M)/A)|V_i \cong N$. $\qquad \square$

Now let $\mathcal{W}_i(T) = \{M' \in \mathbb{S}_{n,m,N} : T_i(M') = T\}$. Recall that $s_{n,r}$ denotes the number of rank-r sparse paving matroids on n elements.

Claim 2.3.11.2. $T_i(M) \in \mathcal{W}_i(T_i(M)))$; i.e. $T_i(T_i(M)) = T_i(M)$

Proof. The relaxation of the non-bases strictly does not affect the primacy of A or V in their respective orderings. $\qquad \square$

Claim 2.3.11.3. $|\mathcal{W}_i(T_i(M))| \leq s_{n_N, r_N}$

Proof. Every matroid in $\mathcal{W}_i(T_i(M))$ can be constructed from $T_i(M)$ by adding non-bases strictly within V_i. Clearly the number of ways to do this is bounded above by the number of rank-r sparse paving matroids we can construct on groundset V_i. $\qquad \square$

We shall show that, asymptotically almost certainly, $M = T_i(M)$ for some $0 \leq i \leq m - 1$, and therefore M contains an N-minor.

Although the remainder of our proof is strictly speaking a counting argument, it perhaps is more naturally understood in the language of probability, which is how we shall proceed. Recall that we are considering M drawn uniformly at random from $\mathbb{S}_{n,m,N}$.

Let X_i be the event that $M = T_i(M)$ and $\bar{X}_i$ the complement; let $Z_k = \bar{X}_0 \cup \cdots \cup \bar{X}_k$. We claim that for all $i \in \{0, \ldots, m-1\}$ the following holds:

$$\Pr(Z_i) \leq \left(\frac{s_{n_N, r_N} - 1}{s_{n_N, r_N}} \right)^{i+1} \tag{2.1}$$

This is sufficient to prove our theorem: for any $\epsilon > 0$ we could now achieve $\Pr(M \not\succeq N) < \epsilon$ simply by setting $m(\epsilon) \geq -\log \epsilon / (\log(s_{n_N} - 1) - \log(s_{n_N}))$.

We use the general identity

$$\Pr(A \cap B) = \Pr(A|B)\Pr(B)$$

which tells us that

$$\Pr(Z_i) = \Pr(\bar{X}_i \cap Z_{i-1}) = \Pr(\bar{X}_i|Z_{i-1})\Pr(Z_{i-1})$$

Repeated application gives

$$\Pr(Z_i) = \prod_{0 \leq j \leq i} \Pr(\bar{X}_j | Z_{j-1})$$

Therefore we are done if we can show for all i that

$$\Pr(\bar{X}_i | Z_{i-1}) \leq \frac{s_{n_N, r_N} - 1}{s_{n_N, r_N}}$$

Equivalently we can (and will) show that $\Pr(X_i | Z_i) \geq \frac{1}{s_{n_N, r_N}}$.

Let $\mathcal{S}_i := \{M \in \mathbb{S}_{n,m,N} : M \neq T_j(M) \text{ for } j < i\}$, i.e those matroids for which Z_{i-1} holds. Note that $M \in \mathcal{S}_i \iff T_i(M) \in \mathcal{S}_i$.

Consider some matroid $T \in \mathcal{S}_i$. If T is an i-parent, we have already seen that $|\mathcal{W}_i(T)| \leq s_{n_N, r_N}$. Moreover $T \in \mathcal{W}_i(T)$, so we have

$$\Pr(X_i | T_i(M) = T) \geq \frac{1}{s_{n_N, r_N}}$$

Since this holds for any T that is an i-parent, we have

$$\Pr(X_i | Z_{i-1}) = \sum_{\substack{T \in \mathcal{S}_i \\ T \text{ an } i \text{ parent}}} \Pr(X_i | T_i(M) = T) \Pr(T_i(M) = T)$$

$$\geq \frac{1}{s_{n_N, r_N}} \sum_{\substack{T \in \mathcal{S}_i \\ T \text{ an } i \text{ parent}}} \Pr(T_i(M) = T) = \frac{1}{s_{n_N, r_N}}$$

We may cancel the sum of probabilities because each matroid has a well-defined i-parent, meaning it is counted precisely once in the above summation. $\square$

Note that the above method can be easily adapted to provide an equivalent

result restricted to matroids of any given rank $r \geq r_N$, by requiring that we contract by a set A of given size. In particular, if we do not contract at all, we get the following result:

Theorem 2.3.12. *Let N be a fixed (sparse) paving matroid. Let M be a matroid drawn randomly from those matroids in $\mathbb{S}_{n,r_N}$ which contain m element-disjoint copies of L.*

For any $\epsilon > 0$, there exists $m(\epsilon)$ such that if $m > m(\epsilon)$, then M will contain N as a minor with probability at least $1 - \epsilon$.

We state this because it proves useful to be able to quote the theorem in this form when pursuing results for matroids of fixed rank in Section 2.3.2.

2.3.1.1 Abundant minors

Recall our earlier definition of *abundance*:

Definition 2.3.13 (Abundant). A rank-r line-structure $\mathcal{L}$ is *abundant* in sparse paving matroids if for any $m \in \mathbb{N}$, for asymptotically almost all n-element sparse paving matroids M there exists a set $A \in [n]^{(r(M)-r)}$ such that $\mathcal{L}(M/A)$ contains m element-disjoint copies of $\mathcal{L}$.

Definition 2.3.14 (Extremal density). The *extremal density* in $\mathbb{S}_{n,r}$ of a line-structure L of rank r is

$$\mathrm{ex}(n, \mathcal{L}) = \max\left\{ p : \exists M, |\mathcal{L}(M)| \leq p\left(\frac{1}{n}\binom{n}{r}\right), \mathcal{L}(M) \not\supseteq \mathcal{L} \right\}$$

taken over matroids $M \in \mathbb{S}_{n,r}$.

In simpler terms, it is the maximal density of non-bases a sparse paving matroid can have before its non-bases must contain a copy of $\mathcal{L}$.

We wish to show that certain conditions of extremal density can imply abundance. That enables us to say that any sparse paving matroid whose line-structure meets those conditions will be abundant, and therefore a minor of asymptotically almost all sparse paving matroids.

Theorem 2.3.15. *Let $\mathcal{L}$ be a line-structure of rank $r_{\mathcal{L}}$ on $n_{\mathcal{L}}$ elements, and k any integer. Let $\mathcal{L}^k$ denote the line-structure consisting of k element-disjoint copies of $\mathcal{L}$. Then*

$$\lim_{n \to \infty} \mathrm{ex}(n, \mathcal{L}^k) = \lim_{n \to \infty} \mathrm{ex}(n, \mathcal{L})$$

Proof. Our proof is inductive. Suppose that

$$\lim_{n \to \infty} \mathrm{ex}(n, \mathcal{L}^{k-1}) = \lim_{n \to \infty} \mathrm{ex}(n, \mathcal{L}) = p$$

Clearly $\mathrm{ex}(n, \mathcal{L}^k) \geq \mathrm{ex}(n, \mathcal{L})$, so we need to show that for any $\epsilon > 0$, there exists $n(\epsilon)$ such that $n > n(\epsilon) \implies \mathrm{ex}(n, \mathcal{L}^k) < p + \epsilon$. Let M be an n-element sparse-paving matroid of rank $r_{\mathcal{L}}$ with density of non-bases (the fraction of all $r_{\mathcal{L}}$-sets of M which are non-bases) being $\frac{p+\epsilon}{n}$, and suppose that n is large enough that $\mathrm{ex}(n, \mathcal{L}^{k-1}), \mathrm{ex}(n, \mathcal{L}) < p + \epsilon/2$. So $\mathcal{L}(M)$ must contain some copy of $\mathcal{L}$. But the number of non-bases intersecting with a line in our copy of $\mathcal{L}$ is at most $n_{\mathcal{L}} \left(\frac{1}{n - r_{\mathcal{L}}} \right) \binom{n}{r_{\mathcal{L}} - 1}$ (see argument in Lemma 3.1.4). For large enough n, this is less than $\frac{\epsilon}{2n} \binom{n}{r_{\mathcal{L}}}$, and so even relaxing all these non-bases, we are left with a density of non-bases greater than $\mathrm{ex}(n, \mathcal{L}^{k-1})$. Hence we must also contain some copy of $\mathcal{L}^{k-1}$ disjoint to our copy of $\mathcal{L}$, and the union of these is a copy of $\mathcal{L}^k$. $\square$

Lemma 2.3.16 (Abundance Lemma). *Let N be a sparse paving matroid. Suppose the following holds: $\exists \epsilon > 0, \exists n_0 \in \mathbb{N}$, such that for any $n > n_0$ we have $\mathrm{ex}((n, \mathcal{L}(N))) < \frac{1}{10} - \epsilon$. Then N is abundant in the sparse paving matroids.*

Proof. Firstly, by Lemma 2.3.15 we can assume without loss of generality that for $n > n_0$ we have $\mathrm{ex}((n, (\mathcal{L}(N))^m)) < \frac{1}{10} - \epsilon$.

We know from Corollary 2.3.5 that asymptotically almost all $M \in \mathbb{S}_{n,r}$ have

$$|\mathcal{L}(M)| > \left(\frac{1}{5n} \binom{n}{r} \right)$$

and that also for any $\delta > 0$ asymptotically almost always $\left(\frac{1}{2} - \delta \right) n \leq r(M) \leq \left(\frac{1}{2} + \delta \right) n$ (Lemma 2.3.2)

Consider only sparse paving matroids M that meet the above two conditions. Contracting M by $r - r_N$ randomly chosen elements, the expected density of non-bases in our new matroid is equal to the density of non-bases in M, so there must be some set A of $r - r_N$ elements such that

$$|\mathcal{L}(M/A)| > \left(\frac{1}{5n} \binom{n'}{r_N} \right)$$

where $n' = n + r_N - r$.

Now set $\delta > 5\epsilon/2$ and $n > \frac{2n_0}{1 - \delta}$, so $n' > n_0$.

$$|\mathcal{L}(M/A)| > \left(\frac{1}{10} - \epsilon \right) \left(\frac{1}{n'} \binom{n'}{r_N} \right) > \mathrm{ex}((n', \mathcal{L}(N))$$

And M/A has to contain a copy of $(\mathcal{L}(N))^m$, and hence m copies of $\mathcal{L}(N)$. $\quad \square$

We can now identify various matroids which have line-structures abundant in the sparse paving matroids, and hence (by Theorem 2.3.10) are contained as a minor in asymptotically almost all sparse paving matroids.

Firstly, using the fact the $\mathcal{L}(U_{t,k})$ is empty and so trivially abundant in the sparse paving matroids:

Theorem 2.3.17. *For any positive integers $t \leq k$, the uniform matroid $U_{t,k}$ is contained as a minor in asymptotically almost all sparse paving matroids.*

For other cases we shall present a proof of abundance; the minor-inclusion is asymptotically almost all sparse paving matroids follows as a trivial corollary thanks to Theorem 2.3.10.

Theorem 2.3.18. *Let N be a sparse paving matroid of rank r in which all non-bases are pairwise disjoint. Then $\mathcal{L}(N)$ is abundant in the sparse paving matroids.*

Proof. Clearly any matroid with non-zero density of non-bases contains a single non-basis. Apply Lemma 2.3.15. $\square$

Corollary 2.3.19. *Let N be any sparse paving matroid of rank r in which all non-bases are pairwise disjoint. Asymptotically almost all sparse paving matroids contain N as a minor.*

Theorem 2.3.20. *Let N be a sparse paving matroid of rank r in which all non-bases meet in a set of size $r-2$. $\mathcal{L}(N)$ is abundant in the sparse paving matroids.*

Proof. Suppose N is a sparse paving matroid on n_N elements with rank r_N and all its lines intersect in one $(r_N - 2)$-subset of $[n_N]$. Let M be an n-element sparse paving matroid of rank r_N with density of non-bases being $\frac{\epsilon}{n}$, for some $\epsilon > 0$. Now given any $(r - 2)$-subset of $[n]$, the expected number of non-bases of M which contain that set is $\frac{\epsilon}{n}\binom{(n+2-r_N)}{2}$, so there must be some $(r_N - 2)$-subset A such that

$$|\{L \in \mathcal{L}(M) : L \supseteq A\}| \geq \frac{\epsilon}{n}\binom{n+2-r_N}{2} \geq \frac{\epsilon}{n}\binom{n/4}{2}$$

For large enough n this number will be greater than $|\mathcal{L}(N)|$, and hence $\mathrm{ex}((n, \mathcal{L}(N))) < \epsilon$. $\square$

Corollary 2.3.21. *Let N be any sparse paving matroid of rank r in which all non-bases meet in a set of size $r - 2$. Asymptotically almost all sparse paving matroids contain N as a minor.*

We next consider the rank 3 whirl, W^3.

Theorem 2.3.22. *There exists a matroid N, containing W^3 as a minor, such that $\mathcal{L}(N)$ is abundant in the sparse paving matroids.*

Proof. Let N be the rank-12 matroid formed by 9 free extensions on W^3. Clearly $N \succeq W^3$. We shall show that for $\epsilon = \frac{1}{2}\left(\frac{1}{10} - \frac{1}{11}\right)$, and large enough n, $\mathrm{ex}((n, \mathcal{L}(N))) < \frac{1}{10} - \epsilon$. Consider a matroid M in $\mathbb{S}(n, 12)$ with at least density $(\frac{1}{10} - \epsilon)$ of non-bases.

Imagine choosing at random a set of 10 elements $\{a'_1, a'_2, \ldots, a'_{10}\}$ from our groundset. Let Y be the random variable denoting the number of non-bases which contain the chosen set. We note that

$$\mathrm{E}(Y) \geq \left(\frac{1}{10} - \epsilon\right)\left(\frac{n-10}{2}\right)$$

Now consider separately drawing at random a non-basis $L = \{a_1, \dots, a_{12}\} \in \mathcal{L}(M)$, and randomly discarding one element (a_{12}, say) to form the set $L^- = \{a_1, \dots, a_{11}\}$. For each $i \in \{1, \dots, 11\}$, let X_i be the derived variable given by the number of non-bases of M containing $L^- \setminus a_i$ (a set of size 10).

Now for any choice of i we have $\mathrm{E}(X_i) = \sqrt{\mathrm{E}(Y^2)}$: This is because we are essentially performing the same count as before, except that each 10-element set is selected with probability proportional to the number of non-bases containing it. So

$$\mathrm{E}(X_i) = \sqrt{\mathrm{E}(Y^2)} \geq \mathrm{E}(Y) \geq \left(\frac{1}{10} - \epsilon\right)\left(\frac{n-10}{2}\right)$$

with the first inequality due to Jensen's inequality. But then, assuming large enough n,

$$\mathrm{E}(\sum_{i=1}^{11} X_i) > \left(1 + \frac{1}{20}\right)\left(\frac{n-10}{2}\right) > \frac{n}{2} + 11$$

(of course 11 is completely arbitrary here, we could equally say greater than $\frac{n}{2} + c$ for any positive constant c. We have written 11 because this accounts for the 11 times our original C will have been counted.)

So there is some choice of $L = \{a_1, \dots, a_12\}$ and L^- (which we shall again assume without loss of generality to be $\{a_1, \dots, a_11\}$) such that

$$\sum_{a_i \in L^-} X_i > n/2 + 11$$

This means then that there are more than $n/2$ distinct ways to construct a non-basis other than L itself. But each of these involves adding two elements, so by the pigeonhole principle we must have some element $b \notin L$ such that b is added more than once - that is to say, M contains more than one non-basis of form

$$C - \{a_{12}, a_i\} \cup \{b, y\}, i \in \{1, \ldots, 11\}, y \notin C$$

Note that the choices of deleted a_i must be different, else the non-bases would intersect in 11 elements. Say without loss of generality we have non-bases $L' = \{a_1, \ldots, a_{10}, b, c\}$ and $L'' = \{a_1, \ldots, a_9, a_{11}, b, d\}$. So the line-structure $\mathcal{L} = \{L, L', L''\}$ has extremal density less than $\frac{1}{10} - \epsilon$ and is abundant in the sparse paving matroids. But $\mathcal{L} \cong \mathcal{L}(N)$ so we are done. $\square$

Corollary 2.3.23. *Asymptotically almost all sparse paving matroids contain* W^3 *as a minor.*

Figure 2.4: Some classes of matroids contained as minors in asymptotically almost all sparse paving matroids: leftmost, matroids with all non-bases pairwise disjoint; centre, matroids in which all non-bases intersect in some $(r - 1)$-set; rightmost, the rank-3 whirl, W^3.

We could continue in this vein, addressing target minors one by one, and doubtless there are many which would fall to arguments like the above (albeit perhaps increasingly complicated arguments!) For now we shall leave that area open - however in Conjecture 2.3.29 we shall give some indication of which cases we believe may be accessible.

2.3.2 A Ramsey-theoretical approach for the fixed-rank case

We have already noted that sparse paving matroids can be identified with hypergraphs whose edges are the non-bases of the matroid. Moreover the independent (or dependent) sets of any given rank can be viewed as a hypergraph. This opens up varied possibilities for using Ramsey-theoretical techniques as a means to answering minor-inclusion questions.

The relevant result of Ramsey theory is the general Ramsey Theorem for hypergraphs.

Definition 2.3.24 (Hypergraph, r-uniform hypergraph)**.** An r*-uniform hypergraph* consists of a pair $(V, \mathcal{E})$, where V is a set (the *vertices*) and $\mathcal{E} \subseteq V^{(r)}$ (the *hyperedges*).

Definition 2.3.25 (Complete hypergraph)**.** The *complete hypergraph $K_n^{(r)}$* is an r-uniform hypergraph n vertices containing every possible hyperedge. In other words, $|V| = n$ and $\mathcal{E} = V^{(r)}$.

Theorem 2.3.26 (Ramsey)**.** *Let G be a r-uniform hypergraph on n vertices with edges k-coloured - that is, each hyperedge is labelled with one of k colours*

$c_1, c_2, \ldots, c_k$. *Then for any positive integers* $t_1, t_2, \ldots, t_k$ *there exists some number* $n_0(t_1, t_2, \ldots, t_k)$ *such that if* $n > n_0$ *then for some* $i \in \{1, 2, \ldots, k\}$, G *contains a* c_i*-coloured* $K_{t_i}^{(r)}$*-subhypergraph.*

2.3.2.1   Uniform matroids

In the previous section we proved at length that various matroids, including all uniform matroids, are contained in asymptotically almost all sparse paving matroids. However the result for uniform matroids can also be achieved almost immediately using Ramsey theory.

Theorem 2.3.27. *For any positive integers* $t \leq k$, *asymptotically almost all sparse paving matroids contain* $U_{t,k}$ *as a minor.*

Proof. Let $M \in \mathbb{S}_n$ be a randomly chosen sparse paving matroid. Note that asymptotically almost certainly $t < r(M) < 3n/4$ (due to Lemma 2.3.2) and we may consider only these cases. We shall show that for all values of r in this range, matroids in $\mathbb{S}_{n,r}$ must contain a $U_{t,k}$ minor. For $M \in \mathbb{S}_{n,r}$ consider M/A, where $A = \{1, 2, \ldots, (r - t)\}$. Now M/A has at least $n/4$ elements. Now we are done if, on some set of k elements, M/A contains no non-bases. But now the set $\mathcal{L}(M)$ of non-bases of M/A is a hypergraph, and hence by Ramsey if $n/4$ is sufficiently large (specifically, $n > 4R_r(r + 1, k)$) then $\mathcal{L}(M)$ must contain either a clique $K_{r+1}^{(r)}$ or stable set $E_k^{(r)}$. The former case is impossible, as the elements of the clique would form a hyperplane of cardinality $r + 1$ in M/A, contradicting sparse pavingness. So there is some set B of k elements on which $\mathcal{L}(M)$ is empty. Then $(M/A)|B \cong U_{t,k}$. $\square$

The next definition enables us to describe (and prove) a minor-inclusion result for a substantial class of matroids. This results represents the broadest class

for which we shall prove a result of asymptotically almost certain minor inclusion (albeit in the limited case of fixed-rank sparse paving matroids). It is also notable that the proof requires several unusual applications of Ramsey-theoretical methods.

2.3.2.2   Loose elements and tied non-bases

Definition 2.3.28 (Loose element, Tied non-basis)**.** Let M be a sparse paving matroid. An element x in a non-basis L of M is called *loose* if no other non-basis of M contains x. Moreover, a non-basis L is called *tied* if it contains no loose elements.

Due to the freedom offered by a loose element, we suggest that the highest barrier to proving Conjecture 2.1.2 may be the case where the target minor has no loose elements. In particular, we conjecture that the following relaxation of Conjecture 2.1.2 may be more accessible.

Conjecture 2.3.29. *Let N be a sparse paving matroid. Suppose that every minor of N has a loose element in at least one non-basis. Then asymptotically almost all sparse paving matroids have an N-minor.*

The main theorem of this chapter is weaker still, and only applies to sparse paving matroids of fixed rank, but still represents a significant advance on past results.

Theorem 2.3.30. *Let N be a rank-r sparse paving matroid with no tied non-bases. Then asymptotically almost all sparse paving matroids of rank r contain N as a minor.*

The proof will use Ramsey theory to establish that all sparse paving matroids
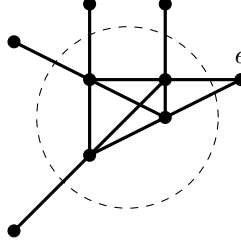
Figure 2.5: A 4-element fort in a matroid with rank $r = 3$. The fort is those elements within the dashed circle. Every collection of $r - 1$ elements within the fort forms a non-basis with an r-th element outside the fort. These r-th elements are not necessarily distinct, as illustrated by the element e which appears in two lines. In a moat, there are no non-bases of this kind.

of sufficient size and rank r must fall into at least one of two categories, within each of which asymptotically almost all matroids contain N as a minor. One category is those matroids which contain (as a subhypergraph of the non-bases) sufficiently many copies of $\mathcal{L}(N)$, which we show by demonstrating the inclusion of a structure which we shall call a *fort*. The other is matroids some contraction of which contains sufficiently many copies of a structure which we shall call a *moat* (with some additional requirements). We shall first define those structures.

Definition 2.3.31 (Fort)**.** Let M be a sparse paving matroid of rank r. A collection of elements X forms a *fort* of M if for every $A \in X^{(r-1)}, \exists L \in \mathcal{L}(M) : L = A \cup e, e \notin X$.

An example of a fort is shown in Figure 2.5.

Definition 2.3.32 (Moat)**.** Let M be a sparse paving matroid of rank r. A collection of elements X forms a *moat* of M if for every $A \in X^{(r-1)}, \nexists L \in \mathcal{L}(M) : L = A \cup e, e \notin X$.

So forts and moats are in some sense inverse to one another - in one case,

we have maximised the number of non-bases intersecting with X in precisely $r - 1$ elements; in the other, we have minimised it. In either case, the structure provides a somewhat controlled space in which to demonstrate the asymptotically almost certain existence of minors. For moats, we have created a sort of vacuum free from outside interference, within which we can show that there is a non-vanishing chance of an $\mathcal{L}(N)$ copy occurring. In the case of forts, we want the lines of the fort themselves to contain an $\mathcal{L}(N)$ copy; in each line, the single element outside the fort corresponds to the loose element in a non-basis. The main challenge here is showing that we don't have any unwanted intersections between those elements.

Our proof will proceed by demonstrating that given it contains a sufficient number of element-disjoint moats of sufficient size, a matroid will asymptotically almost certainly contain an N-minor; we then show something similar for forts. Finally we shall apply a Ramsey argument to prove that a sufficiently large matroid will contain a sufficient collection of either moats or forts.

We'll first address the case of moats. The usefulness of moats is that if we can fix some moat and everything outside the moat, then we can allow any legitimate structure of non-bases within the moat (i.e. those contained entirely within the moat X) and the whole structure will remain sparse paving. However in this instance we will want to make further restrictions on the interior of the moat.

Definition 2.3.33 (Empty moat)**.** Let M and N be sparse paving matroids of rank r. We say a moat X of M is *empty* if there are no non-bases in the interior of the moat, that is to say X contains no non-bases of M. We also shall say X is *sub-N* if the non-bases of M contained in X are isomorphic to

a subhypergraph of the line structure $\mathcal{L}(N)$.

We note that an empty moat X is sub-N for any choice of N.

Lemma 2.3.34. *Let N be a sparse paving matroid with n_N elements and rank r. We let $\mathbb{S}_{n,r,m,N}$ denote the members of $\mathbb{S}_{n,r}$ which contain at least m pairwise disjoint sub-N moats of size n_N. Let $\epsilon > 0$. Then there exists an integer $m(\epsilon)$ such that for large enough n the following holds: let $m \geq m(\epsilon)$ and M be drawn randomly from $\mathbb{S}_{n,r,m,N}$. With probability at least $(1-\epsilon)$, M contains N as a minor.*

The proof of this is very similar to Theorem 2.3.11. The main difference is that we are now demanding that our moats contain a substructure of $\mathcal{L}(N)$, whereas in Theorem 2.3.11 we required superstructures of $\mathcal{L}(N)$.

This proof relies on $r(M)$ being equal to $r(N)$. If $r(M) \geq r(N)$ then we would be required to work within M/A, for some set A. But in that environment we cannot freely add non-bases within a moat, since there might be non-bases in M (before contraction by A) that "block" such a move. It turns out that, in the wider context of proving Theorem 2.3.30, this is the only reason we require a fixed rank.

Proof. Assume without loss of generality that $\epsilon < 1$.

We aim to assign to M a series of parent matroids $T_0, \ldots, T_{m-1} \in \mathbb{S}_{n,m,N}$, so that every choice of M has precisely one choice of parents. We show then that for any series of parents, at most proportion ϵ of their shared children are free of N-minors. The same will then be true for the union of offspring of all series of parents, which is of course equal to the entirety of $\mathbb{S}_{n,r,m,N}$.

Imagine taking a collection $U = \{(U_0, \mathcal{L}_0), \cdots, (U_{m-1}, \mathcal{L}_{m-1})\}$ where the U_i denote a collection of pairwise disjoint n_N subsets of $[n]$ and the $\mathcal{L}_i$ denote m

copies of the line-structure $\mathcal{L}(N)$ on $[n]^{(r_N)}$ such that, for each pair $(U_i, \mathcal{L}_i)$, the lines of $\mathcal{L}_i$ are strictly contained in $U_i^{(r)}$. We define $\mathcal{U}_{n,m,N}$ to be the universe of all possible collections of this type.

Let $*(n, r, m, N)$ be a total ordering on $\mathcal{U}_{n,r,m,N}$ - as before, any total ordering will do, although some kind of lexicographical ordering is most intuitive. Say that $V >^* W$ if V is above W in this ordering.

Now, we have at least one collection of m element-disjoint sub-N moats in M - call them $V_0, \ldots, V_{m-1}$ - which means that we also have a collection $\mathcal{L}_0, \ldots, \mathcal{L}_{m-1}$, where each $\mathcal{L}_i$ denotes an $\mathcal{L}(N)$ copy within V_i, such that $\mathcal{L}(M|V_i) \subseteq \mathcal{L}_i$. Let's assume that $V = \{(V_0, \mathcal{L}_0), \ldots, (V_{m-1}, \mathcal{L}_{m-1})\}$ is the maximal such collection under our ordering $>^*$.

Define the *i-parent* $T_i(M)$ of M to be the matroid whose line-structure is $\mathcal{L}(M) \cup \mathcal{L}_i$, such that $\mathcal{L}(T_i(M)|V_i) \cong \mathcal{L}_i$, and $\mathcal{L}(T_i(M)|([n] \setminus V_i)) \cong \mathcal{L}(M|([n] \setminus V_i))$. In other words, $T_i(M)$ is identical to M except on V_i, where its line-structure is now equal to L_i.

Clearly $T_i(M) \succeq N$. Moreover, if $\mathcal{W}_i(T) = \{M' \in \mathbb{S}_{n,m,N} : T_i(M') = T\}$, then $T_i(M) \in \mathcal{W}_i(T_i(M))$, and $|\mathcal{W}_i(T_i(M))| \leq 2^{|\mathcal{L}(N)|}$.

Let X_i be the event that $M = T_i(M)$ and $\bar{X}_i$ be the converse. Further, let $Z_k = \bar{X}_0 \cup \cdots \cup \bar{X}_k$. Note that in the event M is N-free, Z_{m-1} is implied.

We aim to show that:

$$\Pr(Z_{m-1}) \leq \left(\frac{2^{|\mathcal{L}(N)|} - 1}{2^{|\mathcal{L}(N)|}} \right)^m \qquad (2.2)$$

Having shown this we will be done: for any $\epsilon > 0$ we can achieve $\Pr(M \not\succeq N) < \epsilon$. We simply need $m > -\log \epsilon / (\log(2^{|\mathcal{L}(N)|} - 1) - \log(2^{|\mathcal{L}(N)|}))$ and this

provides our value for $m(\epsilon)$.

To prove (2.2) we continue as in the proof of Theorem 2.3.11: it suffices to show that, for all $0 \leq k \leq m - 1$,

$$\Pr(X_k | Z_{k-1}) \geq \frac{1}{2^{|\mathcal{L}(N)|}}$$

Let $\mathcal{S}_i := \{M \in \mathbb{S}_{n,r,m,N} : M \neq T_j(M) \text{ for } j < i\}$, i.e those matroids for which Z_{i-1} holds. Note that $M \in \mathcal{S}_i \iff T_i(M) \in \mathcal{S}_i$.

Consider some matroid $T \in \mathcal{S}_i$. Since $T \in \mathcal{W}_i(T)$ and $|\mathcal{W}_i(T)| \leq 2^{|\mathcal{L}(N)|}$, we have

$$\Pr(X_i | T_i(M) = T) \geq \frac{1}{2^{|\mathcal{L}(N)|}}$$

Since this holds for any T that is an i-parent, we have

$$\Pr(X_i | Z_{i-1}) = \sum_{T \in \mathcal{S}_i} \Pr(X_i | T_i(M) = T)\Pr(T_i(M) = T)$$

$$\geq \frac{1}{2^{|\mathcal{L}(N)|}} \sum_{\substack{T \in \mathcal{S}_i \\ T \text{ an } i \text{ parent}}} \Pr(T_i(M) = T) = \frac{1}{2^{|\mathcal{L}(N)|}}$$

where again we may only cancel the sum of probabilities because our careful definition of a unique i-parent ensures that each matroid is counted precisely once in the summation. $\qquad\square$

Note that though we have chosen only to prove the result for sparse paving matroids, we believe the same result could be proven for general matroids in essentially the same way.

Now we want to deal with forts. Recall our earlier definition: let M be a sparse paving matroid of rank r. A collection of elements X forms a *fort* of M if for every $A \in X^{(r-1)}, \exists L \in \mathcal{L}(M) : L = A \cup e, e \notin X$.

For any fort X we shall use the notation $\mathcal{L}_X^-$ to denote the collection of non-bases of M which contain $r-1$ elements of X. Note that we may also apply this notation to any subset $A \subseteq X$, since A is itself a fort. We want to show the following.

Lemma 2.3.35. *Suppose X is a fort of size n in a sparse paving matroid M of rank r. Then for any $m \geq r$, there exists $n(m, r-1)$ such that $n > n(m, r-1)$ implies that there exists $X' \subset X, |X| \geq m$, such that no two elements of $\mathcal{L}_{X'}^-$ intersect outside of X'.*

That is to say that given an arbitrarily large fort in a sparse paving matroid M of rank r, we can find an arbitrarily large subset X' of that fort on which each $(r-1)$-sets of X' can be matched with a unique element of $M \backslash X'$ to form a non-basis of M. These unique elements of $M \setminus X'$ form the loose elements in a copy of $\mathcal{L}(N)$; since every $(r-1)$-subset of X' is paired with such an element, we can readily choose $|\mathcal{L}(N)|$ of them isomorphic to the non-bases of N minus their loose elements, and so have a copy of $\mathcal{L}(N)$.

First we need to prove a result that is in some sense anti-Ramsey, that is to say we want to demonstrate the existence of *polychromatic* cliques in a hypergraph, given sufficient size and one simple condition.

Definition 2.3.36 (Polychromatic)**.** Let G be an r-uniform hypergraph and $\mathcal{L}(G)$ denote the collection of hyperedges of G. Let $c : \mathcal{L}(G) \to \mathbb{N}$ be a colouring of the hyperedges of G. A subhypergraph H of G is *polychromatic* under c if for any edges $A, B \in \mathcal{L}(H)$, we have $A \neq B \implies c(A) \neq c(B)$.

In the following section we treat $X^{(r)}$ as the complete r-uniform hypergraph with vertex set X.

Lemma 2.3.37. *Let X be a set of n elements. Say we have a function $c : X^{(r)} \to \mathbb{N}$ such that $A, B \in X^{(r)}, |A \cap B| = r - 1 \implies c(A) \neq c(B)$. Then for any $m \geq r$, there exists $n(m, r)$ such that $n > n(m, r) \implies \exists X' \subset X, |X'| = m$, such that $(X')^{(r)}$ is polychromatic under c.*

We note that our condition is equivalent to specifying that c is a vertex-colouring of the Johnson graph $J(n, r)$. However it makes more sense for us to view c as a hyperedge-colouring of the r-uniform hypergraph.

Proof. We prove this by (transfinite) induction on m and r, ordering on r first. Thus we say $(k, s) < (m, r)$ if either $s < r$, or $s = r$ and $k < m$. We take as our inductive hypothesis that $n(k, s)$ exists for all $(k, s) < (m, r)$. We can see easily enough that $n(k, 1) = k$ and $n(1, r) = r$, which covers all base and limit cases in our ordering. Next we use our inductive hypothesis and proceed to demonstrate the existence of $n(m, r)$ in the case $m, r \neq 0$.

The proof will take an element $x \in X$ and demonstrate, for large enough n, the existence of nested sets $(X \setminus x) \supseteq X_1 \supseteq X_2 \supseteq X_3$ with $|X_3| = m - 1$ such that:

$$\text{for } A, B \in (X_1 \cup x)^{(r)} : x \in A, x \in B \implies c(A) \neq c(B)$$

$$\text{for } A, B \in (X_2 \cup x)^{(r)} : x \in A, x \notin B \implies c(A) \neq c(B)$$

$$\text{for } A, B \in (X_3 \cup x)^{(r)} : x \notin A, x \notin B \implies c(A) \neq c(B)$$

Taken together, and considering the nestedness of X_1, X_2 and X_3, the above conditions imply that for $A, B \in (X_3 \cup x)^{(r)}$ we have $c(A) \neq c(B)$, and therefore $X_3 \cup x$ satisfies our requirements.

First we choose some element $x_1 \in X$. We can induce a colouring c' : $(X \setminus x)^{(r-1)} \to \mathbb{N}$ as follows: for any $A' \in (X \setminus x)^{(r-1)}$, let $c'(A') = c(A' \cup x)$. Now, we invoke (by inductive hypothesis) the existence of $n(K, r-1)$ (for all K)to say that if $|X| \geq n(K, r-1) + 1$ then there must exist some set $X_1 \subseteq (X \setminus x)$ with $|X_1| \geq K$ such that for distinct $A', B' \in X_1^{(r-1)}$, we must have $c'(A') \neq c'(B')$ Therefore $c(A' \cup x) \neq c(B' \cup x)$.

Now in X_1 we can say that for distinct $A', B' \in X_1^{(r-1)}$, we have $c'(A') \neq c'(B')$. Moreover, if $B \in X_1^{(r)}$ and $A' \in X_1^{(r-1)}$ we have $A' \subset B \implies c'(A') \neq c(B)$ (since otherwise $c(B) = c(A' \cup x)$. But $|B \cap (A' \cup x)| = r - 1$, so the conditions of the proof would be contradicted). We claim the following:

Claim 2.3.37.1. For any K', given $|X_1|$ sufficiently large, we can find $X_2 \subseteq X_1$, with $|X_2| = K'$, such that for $B \in X_2^{(r)}$ and $A' \in X_2^{(r-1)}$ we have $c'(A') \neq c(B)$.

Proof of claim. Since every $(r-1)$-set of X_1 takes a different value under c', we have that for any $B \in X_1^{(r)}$, there can be at most 1 choice of $A' \in X_1^{(r-1)}$ such that $c(B) = c'(A')$. Moreover, if such an A' exists, then $|A' \setminus B| \neq 0$, i.e. there exists some element $e \in A' \setminus B$ (suppose not: then $c(A' \cup x) = c(B)$. But $|(A' \cup x) \cap B| = r - 1$, contradicting our conditions). Consider drawing some random $X_2 \in X_1^{(K')}$ (for understanding the probabilistic argument, it may be easier to imagine this process happens by randomly permuting the labels of elements X_1, and setting X_2 to the first K' elements). Now define the variable

$$Z := |\{B \in X_2^{(r)} : \exists A' \in X_2^{(r-1)}, c(B) = c'(A')\}|$$

Meanwhile, let Z_B be a variable taking value 1 if $\exists A' \in X_2^{(r-1)}, c(B) = c'(A')$, and value 0 otherwise. Now

$$Z = \sum_{B \in X_2^{(r)}} Z_B$$

But for any choice of B, the probability that Z_B takes value 1 is low: specifically, we know that if a valid choice of A' exists in $X_1^{(r-1)}$, then there is some $e \in A' \setminus B$, and then

$$\Pr(Z_B = 1) \le \Pr(e \in X_2) \le \frac{K'}{|X_1|}$$

$$\mathrm{E}(Z) \le \sum_{B \in X_2^{(r)}} \mathrm{E}(Z_B) = \sum_{B \in X_2^{(r)}} \Pr(Z_B = 1) \le \binom{K'}{r} \frac{K'}{|X_1|} \le \frac{(K')^{r+1}}{|X_1|}$$

If $|X_1| > (K')^{r+1}$ then $\mathrm{E}(Z) < 1$, so there exists some choice of X_2 such that

$$B \in X_2^{(r)}, A' \in X_2^{(r-1)} \implies c'(A') \ne c(B)$$

as required.   □

Now we are ready to complete the proof of Lemma 2.3.37. Using the existence of $n(m-1, r)$ we let $|X| \ge n((n(m-1,r))^{r+1}, r-1)) + 1$. Then choosing some x we can find X_1, X_2 using the process already described, with $X_2 \subseteq X_1 \subseteq (X \setminus x)$, such that $|X_1| \ge (n(m-1,r))^{r+1}$ and hence $|X_2| \ge n(m-1,r)$. Now, since $|X_2| \ge n(m-1,r)$ we can find $X_3 \subseteq X_2$ with $|X_3| \ge m-1$

such that for distinct $A, B \in X_2^{(r)}$ we must have $c(A) \neq c(B)$. But now consider the set $X_3 \cup x$. Clearly $|X_3 \cup x| = r$. We claim that for distinct $A, B \in (X_3 \cup x)^{(r)}$ we must have $c(A) \neq c(B)$. We have three cases to consider:

Firstly, the case where $x \in A, x \in B$. The claim is clearly satisfied for this case: $(A \setminus x)$ and $(B \setminus x)$ are both members of $X_3^{(r-1)}$ and $X_3 \subseteq X_1$, and so by the construction of X_1 we have $c(A) = c'(A \setminus x) \neq c'(B \setminus x) = c(B)$.

Secondly, the case where x is contained in only one of A and B. Say without loss of generality $x \in (A \setminus B)$. Then $A \setminus x$ is a member of $X_3^{(r-1)}$ and B is a member of $X_3^{(r)}$. But $X_3 \subseteq X_2$, and so by the construction of X_2 we have $c(A) = c'(A \setminus x) \neq c(B)$.

Finally, the case where $x \notin A, x \notin B$. Clearly then $A, B \in X_3^{(r)}$ and so by the construction of X_3 we have $c(A) \neq c(B)$.

So $n(m, r)$ exists and is bounded above by $n((n(m-1, r))^{r+1}, r-1)) + 1$. $\quad\square$

Finally, we can complete the proof of Lemma 2.3.35. Let's first restate the lemma:

Lemma 2.3.35. *Suppose X is a fort of size n in a sparse paving matroid M of rank r. Then for any $m \geq r$, there exists $n(m, r-1)$ such that $n > n(m, r-1)$ implies that there exists $X' \subset X, |X| \geq m$, such that no two elements of $\mathcal{L}_{X'}^-$ intersect outside of X'.*

Proof of Lemma 2.3.35. For any set $L^- \in X^{(r-1)}$ we have some unique ele-

ment e_{L^-} such that $L^- \cup e_{L^-}$ is a non-basis of M. Define $c : X^{(r-1)} \to \mathbb{N}$ by $c(L^-) = e_{L^-}$. Lemma 2.3.37 guarantees the existence of $n(m, r-1)$ such that if $n > n(m, r-1)$ there will exist $X' \subseteq X$ with $|X'| = m$ and every $(r-1)$-subset of X' taking a different value under c. This X' satisfies our conditions. $\qquad \square$

Corollary 2.3.38. *Let N be a rank-r sparse paving matroid on n_N elements with no tied non-bases. Suppose X is a fort of size k in a sparse paving matroid $M \in \mathbb{S}_{n,r}$. For any $A \subseteq X$, Let $\mathcal{L}_A^-$ be the collection of non-bases of M which contain $r-1$ elements of A. There exists $K(N)$ such that $k \geq K(N)$ implies that $\mathcal{L}(M)$ contains a substructure $\mathcal{L}$ isomorphic to $\mathcal{L}(N)$, such that for every line $L \in \mathcal{L}$, we have $|L \cap X| = r - 1$.*

The intuition behind the proof has already been hinted at, but we shall prove it formally.

Proof. Apply Lemma 2.3.35 with $m = n_N - |\mathcal{L}(N)|$ and force $K \geq n(m, r-1)$. We obtain a set X' with $|X'| = n_N - |\mathcal{L}(N)|$ such that for any $(r-1)$-subset A of X' we have an element $f \in E(M)$, unique to A, such that $(A \cup f)$ is a non-basis of M.

Now, for each $L \in \mathcal{L}(N)$ we can choose some loose element e_L. Define

$$X'' = E(N) \setminus \left(\bigcup_{L \in \mathcal{L}(N)} e_L \right)$$

Now relabel the elements of X' identically to those of X''. For any $L \in \mathcal{L}(N)$, let A_L be the $(r-1)$-subset of X' with elements labelled by the elements of $L \setminus e_L$. There exists some element $f_L \in E(M)$ such that $(A_L \cup f_L)$ is a

non-basis of M. But now define $\mathcal{L} \subseteq \mathcal{L}(M)$ to be the collection of all such non-bases. Then $\mathcal{L} \cong \mathcal{L}(N)$ as required. $\square$

And finally:

Proof of Theorem 2.3.30. Recall this theorem stated that for any rank-r sparse paving matroid N with no tied non-bases, asymptotically almost all sparse paving matroids of rank r must contain N as a minor.

Let $n(m, r)$ be defined as in Lemma 2.3.35.

Fix $n_N = |E(N)|$. Let $M \in \mathbb{S}_{n,r}$ and allow n to grow arbitrarily large. We consider the $(r-1)$-uniform hypergraph on $[n]$ formed by including a hyperedge if and only if it is contained in a non-basis of M. Note firstly that a stable set in this hypergraph represents a moat. On the contrary, suppose we have a (sufficiently large) complete subhypergraph. Now consider the restriction of the r-uniform hypergraph $\mathcal{L}(M)$ to the elements of that subhypergraph, and apply Ramsey in a similar manner to Theorem 2.3.27 to find a stable set within that space. This then forms a fort which of course can be forced to be arbitrarily large by increasing the size of the complete subhypergraph.

So, by applying Ramsey to the $(r-1)$-uniform hypergraph defined above, given a sufficiently large n we are eventually able to find find either a moat of size $R_r(r+1, n_N)$ or a fort of size $n(n_N - |L(N)|, r)$. Specifically we require $n > n_0$, where n_0 is at most $R_{r-1}(R_r(r+1, n_N), R_r(r+1, n(n_N - |\mathcal{L}(N)|, r)))$.

Now we imagine that n is truly enormous, say $n = K n_0$. Then $[n]$ can be divided into intervals

$$[1, n_0], [n_0 + 1, 2n_0], \dots, [(K-1)n_0 + 1, Kn_0]$$

We note that by Ramsey each of these intervals must contain either a fort of cardinality $n(n_N - |\mathcal{L}(N)|, r)$ or a moat of cardinality $R_r(r+1, n_N)$. In the former case we have a copy of $\mathcal{L}(N)$, by Corollary 2.3.38. In the latter case we must have (again by the argument of Theorem 2.3.27) some U_{r,n_N} minor within our moat - and this then forms an *empty* moat (and hence also an sub-N moat) on those n_N elements! So by making K large enough we can force that M contains either arbitrarily many copies of $\mathcal{L}(N)$, or contains arbitrarily many sub-N n_N-moats (moreover being element-disjoint, in either case). Then by applying either Theorem 2.3.12 or Lemma 2.3.34 respectively we can see that asymptotically almost certainly M will contain N as a minor. $\square$

Corollary 2.3.39. *Let N be a sparse paving matroid with no tied non-bases, and $r \geq r_N$. Then asymptotically almost all sparse paving matroids of rank r contain N as a minor.*

Proof. It suffices to observe that N is contained as a minor of a matroid of rank r which also has no tied non-bases. To see this, simply form N' by adding the set of elements $A = \{a_1, a_2, \dots, a_{r-r_N}\}$ to the groundset of N, and define $\mathcal{L}(N') = \{L \cup A : L \in \mathcal{L}(N)\}$ $\square$

So far all our results, in both the general rank and fixed rank case, have involved minors with no tied non-bases. In the closing section of the next chapter we shall prove that a particular family of matroids are minors of asymptotically almost all sparse paving matroids of equal rank to them-

selves. The matroids in this family each have four non-bases, two of which are tied non-bases. This result has particularly significant consequences in relation to Graham-Sloane matroids, because matroids in this family cannot be a minor of any Graham-Sloane matroid. We shall briefly introduce the theorem here, and prove it at the end of the next chapter.

Theorem 2.3.40. *Let N be a sparse paving matroid with 4 non-bases of form $X \cup \{a, b\}, X \cup \{c, d\}, Y \cup \{a', b\}, Y \cup \{c, d\}$, where $|X| = |Y| = r(N) - 2$ and $|X \cap Y| \leq r(N) - 4$. Then asymptotically almost all sparse paving matroids of rank $r(N)$ contain N as a minor.*
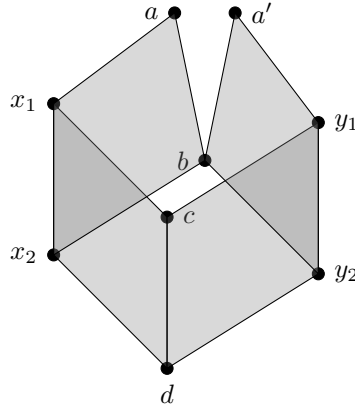
One such matroid is illustrated in Figure 2.6.

Figure 2.6: An example of the matroids described in Theorem 2.3.40

2.3.3 Extending fixed rank results to general rank results?

Imagine $\mathcal{M}$ is a minor-closed class of matroids. Let $\mathcal{M}_{n,r} := \mathcal{M} \cap \mathbb{M}_{n,r}$. We make the following conjecture:

Conjecture 2.3.41. *Fix $r > 0$, allow $n \to \infty$. Suppose N is a minor of asymptotically almost all matroids in $\mathcal{M}_{n,r}$. Then N is a minor of asymptotically almost all matroids in $\mathcal{M}$.*

This conjecture, if true, would enable us to extend our fixed-rank results from Section 2.3.2 to more general results for all sparse paving matroids. Although we cannot find much strong evidence either way on this conjecture, there are some relevant existing results. These are discussed in Pendavingh and van der Pol's guest post at the *Matroid Union* blog [22]. In particular, they show the following two theorems.

Theorem 2.3.42. *[Pendavingh, van der Pol] Suppose $\mathcal{M}$ is a class of matroids closed under deletion. Then for any $n > r$*

$$\frac{\log\left(1 + |\mathcal{M}_{n,r}|\right)}{\binom{n}{r}} \leq \frac{\log\left(1 + |\mathcal{M}_{n-1,r}|\right)}{\binom{n-1}{r}}$$

Repeated applicaton of (the statement dual to) Theorem 2.3.42 gives the following result:

Theorem 2.3.43 (Pendavingh, van der Pol)**.** *Suppose $\mathcal{M}$ is a class of matroids closed under deletion. Then for any $n > r > k$*

$$\frac{\log\left(1 + |\mathcal{M}_{n,r}|\right)}{\binom{n}{r}} \leq \frac{\log\left(1 + |\mathcal{M}_{n-k,r-k}|\right)}{\binom{n-k}{r-k}}$$

Why is this theorem useful? Essentially, the power of this result lies in the fact that that, for all r,

$$\log m_{n,r} \geq \log s_{n,r} \geq \frac{1}{n}\binom{n}{r}$$

therefore, if we could show for **any** r that $\log |\mathcal{M}_{n,r}| < \frac{1}{n}\binom{n}{r}$ then that would imply that asymptotically all members of $\mathbb{S}_n$ (and $\mathbb{M}_n$) were not in $\mathcal{M}$. Of course this effectively requires us to show that our minor-closed class (in this case the class of N-free sparse paving matroids, for some target minor N) is smaller that the class of sparse paving at a *log-level* which our methods do not achieve. (Pendavingh and van der Pol [23] have managed to achieve such a result for N-free matroids where N is a uniform matroid, providing a neat proof that asymptotically almost all sparse paving matroids, and asymptotically almost all matroids, contain N).

Although it is not of direct use, this result may appear to be a source of encouragement, as it suggests that at least sometimes general results are easily implied by fixed-rank results. However we recommend some caution: it is perhaps a little too easy to believe that something similar to the log-level result will hold at the absolute level. In situations like this intuition can turn out to be a poor guide.

For example, let $\mathcal{M}$ and $\mathcal{N}$ be minor-closed classes of matroids, with $\mathcal{N} \subset \mathcal{M}$. Then we might reasonably conjecture the following:

Conjecture 2.3.44 ((false))**.** *Let $\mathcal{M}$ and $\mathcal{N}$ be minor-closed classes of matroids, with $\mathcal{N} \subset \mathcal{M}$. Then for any $n \geq r \geq k$*

$$\frac{|\mathcal{N}_{n,r}|}{|\mathcal{M}_{n,r}|} \leq \frac{|\mathcal{N}_{n-k,r-k}|}{|\mathcal{M}_{n-k,r-k}|}$$

This conjecture, if it were true, would have been very useful to us - imagine N is a sparse paving matroid which is contained in asymptotically almost all rank-r sparse paving matroids, for even just one value of r. Then applying

Conjecture 2.3.44 with $\mathcal{M}$ as the class of sparse paving matroids, and $\mathcal{N}$ as the class of N-free matroids, would imply that N is contained in asymptotically almost all sparse paving matroids (and also in asymptotically all sparse paving matroids of rank s, for fixed $s > r$).

And the conjecture is pleasing to our intuition: after all, every minor of a matroid in $\mathcal{N}$ is also in $\mathcal{N}$, whereas a minor of a matroid in $\mathcal{M} \setminus \mathcal{N}$ is not necessarily in $\mathcal{M} \setminus \mathcal{N}$, so we might expect the former to dominate more as rank and size decrease. However that intuition is incorrect. The conjecture fails if we choose $\mathcal{M}$ to be the class of all matroids, and $\mathcal{N}$ to be the class of sparse paving matroids. Numerous counterexamples arise in Mayhew and Royle's [19] computations : for example, we can see that $|\mathcal{N}_{7,3}|/|\mathcal{M}_{7,3}| = 14/108 \approx 0.13$ whereas $|\mathcal{N}_{8,4}|/|\mathcal{M}_{8,4}| = 270/940 \approx 0.29$. In fact it seems likely that the conjecture is abjectly far from the truth - we suspect (Conjecture 1.0.2) that asymptotically almost all matroids are sparse paving, yet we know that asymptotically almost all matroids of rank r are **not** sparse paving when $r = 2$, and discussion in recent work of Pendavingh and van der Pol [25] includes a persuasive case that the same will be true for $r = 3$. From this we might guess, as they do, that the same will hold for any fixed r.

There is another approach we might consider for extending fixed-rank results to a general case. If we consider the finite class $\mathbb{S}_{n_0,r_0}$, can we always find N, R large enough that for $n > N, r > R$ a matroid in $\mathbb{S}_{n,r}$ will contain as minors at least cs_{n_0,r_0} members of $\mathbb{S}_{n_0,r_0}$? This would give us our desired result: if N is a minor of asymptotically almost all sparse paving matroids of rank r_0, we can find n_0 such that N is a minor of at least $(1 - c)s_{n_0,r_0}$ members of $\mathbb{S}_{n_0,r_0}$. Then when n, r sufficiently large, we would have that every matroid in $\mathbb{S}_{n,r}$ contained an N-minor.

Of course this is false, as easily demonstrated by the uniform matroid $U_{r,n}$ which will only contain one member of $\mathbb{S}_{n_0,r_0}$ as a minor - the uniform matroid U_{r_0,n_0}. But uniform matroids are a weird case, as are all sparse paving matroids with very few non-bases. What if we restrict to members of $\mathbb{S}_{n,r}$ which contain a reasonable number of non-bases, as almost all of them do? Unfortunately the answer is still no. The class of Graham-Sloane matroids, which we shall explore in the next chapter, includes matroids with a large number of non-bases, and yet asymptotically almost all sparse paving matroids of rank r_0 are forbidden minors in the class of Graham-Sloane matroids.

2.4 Other applications of counting methods

Although our focus has been very much on questions of minor inclusion, we believe the methods we have used might be adapted to tackle various other asymptotic questions. This section describes one such use. We shall show that the non-bases of a sparse paving asymptotically almost certainly exhibit a property which we shall call *blocking*.

Definition 2.4.1 (k-blocked)**.** We say a (sparse paving) matroid M is k-*blocked* if for any k-subset S of $E(M)$, there is a non-basis of M containing S as a subset.

The main focus of this section is Theorem 2.4.2 in which we prove asymptotically almost all sparse paving matroids are k-blocked, for arbitrary k.

Theorem 2.4.2. *Let M be a matroid drawn uniformly at random from sparse paving matroids on n elements. Then for any $k \in \mathbb{N}_{>0}$, M is k-blocked.*

Before starting a proof, we have one preliminary result to cover.

Lemma 2.4.3. *Let α be a positive real number. Then*

$$f(\alpha) = \alpha(1 - 1/\alpha)^{\alpha^2} \to 0 \ as \ \alpha \to \infty$$

Proof of Lemma. We consider first the function $g(x) = (1 - x)^{\frac{1}{x}}$

It is relatively easily checked that for $0 < x \le 1, g(x) \le \frac{1}{2}$. If not, then for some x we obtain $\left(\frac{1}{2}\right)^x + x < 1$. But letting $h(x) = \left(\frac{1}{2}\right)^x + x$, we get $h(0) = 1, h(1) = \frac{3}{2}$, and $h'(x) = 1 - \left(\frac{1}{2}\right)^x \log 2 \ge 1 - \log 2 > 0$, which shows this to be impossible.

We now see (by letting $x = 1/\alpha$) that for $\alpha > 1, (1 - 1/\alpha)^\alpha \le \frac{1}{2}$. So $f(\alpha) \le \alpha \left(\frac{1}{2}\right)^\alpha$ which clearly vanishes as $\alpha \to \infty$. $\qquad\square$

With that out of the way, we can proceed with an argument for proof of the main theorem.

Proof of Theorem 2.4.2. We take the strategy of considering a single k-set of elements (say $\{1, \dots, k\}$) and showing that asymptotically almost all sparse paving matroids on $[n]$ have a non-basis containing those elements. If we can show this to be true, and to approach the limit sufficiently quickly (i.e. quick enough to dominate the number of choices of k-set, which is $\binom{n}{k}$) then we will be done.

As in previous problems, our tactic is to fix as much as possible and reduce the problem to one of relatively simple counting. By Lemma 2.3.2 we may assume without loss of generality that we are operating in $\mathbb{S}_{n,r}$ and that $n/4 \le r \le 3n/4$, and prove that the proportion of non-k-blocked matroids vanishes uniformly across all values of r in this range.

Suppose we have a collection of sets $S_0, \ldots, S_l \in [n]^{(r)}$ such that $\forall 0 \le i \le l$, we have $S_i \supset [k]$, and for all $0 \le i < j \le l$ we have $|S_i \cap S_j| \le r - 3$.

Let Σ_i be the neighbourhood of S_i in $J(n, r)$:that is, $\Sigma_i = \{U \in [n]^{(r)} : U \cap S_i \ge r - 1\}$.

Observe that the Σ_i are disjoint; in other words, no element of Σ_j can be a neighbour of S_i, unless $j = i$.

For $0 \le i \le l$ we define $T_i(M)$ to be the matroid obtained from M by relaxing all non-bases in Σ_i. So

$$\mathcal{L}(T_i(M)) = \mathcal{L}(M) \setminus \Sigma_i$$

Let $\mathcal{W}_i(T) = \{N \in \mathbb{S}_{n,r} : T_i(N) = T\}$. Consider the matroid $T_i'(M)$ formed from $T_i(M)$ by adding S_i as a new non-basis. Clearly $T_i'(M) \in \mathcal{W}_i(T_i(M))$. Also $T_i'(M)$ contains a non-basis, S_i, with $[k]$ as a subset.

Claim 2.4.3.1. $\mathcal{W}_i(T_i(M))$ contains at most $(n - r + 1)^k$ matroids without a non-basis containing $[k]$

Proof. Consider that every matroid in $\mathcal{W}_i(T_i(M))$ consists of $T_i(M)$ tightened by the inclusion of some collection of non-bases from Σ_i. These non-bases are necessarily each of the form $S_i \setminus e \cup f$, where $e \in [k]$ and $f \notin S_i$. Clearly no choice of e can be repeated (else we get two non-bases intersecting in $r - 1$ elements, contradicting sparse pavingness) and for each choice of e we have at most $n - r$ choices of f (plus the possibility that no such non-basis is included). Since the same is true for all choices of e, the total choice is $(n - r + 1)^k$. $\square$

Let X_i be the event that some member of $\mathcal{L}(M) \cap \Sigma_i$ contains $[k]$. Let Y_i be the negation of this.

Recall M is drawn uniformly at random from $\mathbb{S}_{n,r}$. We first seek to show that:

$$\Pr\left(\bigcap_{i=0}^{l} Y_i\right) \leq \left(1 - n^{-k}\right)^{l+1}$$

It suffices to show that for any $0 \leq i \leq l$, we have $\Pr(X_i | Y_0 \cap \cdots \cap Y_{i-1}) \geq n^{-k}$.

Let $\mathcal{N}_i := \{N \in \mathbb{S}_{n,r} : \text{for } j < i, \nexists L \in \mathcal{L}(M) \cap \Sigma_j : [k] \subseteq L\}$, i.e. those matroids for which $Y_0, \ldots, Y_{i-1}$ hold. Note that $M \in \mathcal{N}_i \iff T_i(M) \in \mathcal{N}_i$.

For fixed T we have an upper bound of $(n-r+1)^k$ on the number of members of $\mathcal{W}_i(T)$ with no non-basis containing $[k]$, and a lower bound of 1 on the number of members of $\mathcal{W}_i(T)$ that do have a non-basis containing $[k]$. This gives us

$$\Pr(X_i | T_i(M) = T) \geq \frac{1}{(n-r+1)^k + 1} \geq n^{-k}$$

So we have shown the probability that no non-basis of M contains all the elements of $[k]$ is at most $(1 - n^{-k})^{l+1}$.

More generally we are required to consider all possible choices of k elements. So:

$$\Pr(M \text{ not } k\text{-blocked}) \leq \binom{n}{k}(1 - n^{-k})^{l+1} \leq n^k (1 - n^{-k})^{l+1}$$

But hang on! What we have now looks suspiciously related to the function in Lemma 2.4.3. In fact, we are done if we can show that we can choose $S_0, \ldots, S_l$ such that $l + 1 \geq n^{2k}$.

Contracting by $[k]$ we see this to be equivalent to finding a set $\mathcal{S}$ of at least n^{2k} vertices in $J(n-k, r-k)$, such that no two vertices share a common neighbour

(put another way, each pair of vertices represent $(r-k)$-sets intersecting in at most $(r-k)-3$ elements). If we can achieve this, then we can form S_i by adding the elements of $[k]$ to the i-th member of $\mathcal{S}$ (relabelling the elements of the latter so they fall in $[k+1,n]$).

As k diminishes in regard to r and n, we may again assume $n/4 \leq r \leq 3n/4$.

We form a collection of vertices in $J(n-k, r-k)$ (read sets in $[n-k]^{(r-k)}$) as follows: letting $n' = \lfloor (n-k)/2 \rfloor$ and $r' = \lfloor (r-k)/2 \rfloor$, we define:

$$\mathcal{S}_i^- = \{ S \in [n']^{(r')} : \sum_{s \in S} s = i \mod n' \}$$

(Note: this is related to the construction of Graham-Sloane matroids, which we meet in the next chapter.)

Any pair $\mathcal{S}_i^-, \mathcal{S}_j^- (i \neq j)$ must intersect in at most $r'-2$ elements (else the sums of their elements can not be congruent modulo n').

Now clearly the $\mathcal{S}_i^-$ form a partition of $[n']^{(r')}$ so we have some i with $|\mathcal{S}_i^-| \geq \frac{1}{n'} \binom{n'}{r'}$. Let's say without loss of generality that $\mathcal{S}_0^-$ is such a collection. Now we place an ordering on the sets of $\mathcal{S}_0^-$ and also an ordering on the distinct $(\lceil (r-k)/2 \rceil)$-sets of $[n'+1, n-k]$. Form a new collection $\mathcal{S}$ of $(r-k)$-sets in $[n-k]$ by uniting the i-th elements of the respective lists. Clearly there are at least $\frac{1}{n'} \binom{n'}{r'}$ elements in each list, so $|\mathcal{S}| \geq \frac{1}{n'} \binom{n'}{r'} \geq n^{2k}$, for large enough n and r. Finally any two members of $\mathcal{S}$ intersect in at most $r'-2$ elements on $[n']$ and also differ $[n'+1, n-k]$ - therefore intersecting in at most $(r-k)-3$ elements in total. $\qquad \square$

A more careful analysis could no doubt improve this result significantly (i.e. replacing k by a non-constant function) however we shall not include this as our main motivation was to illustrate the wider applications of the counting

approach developed earlier in the chapter.

Chapter 3

Graham-Sloane Matroids

3.1 Introduction

In 1974 Donald Knuth identified that a sparse paving matroid M with a large numbers of non-bases would give rise to a family of sparse paving matroids, obtained only by relaxation of non-bases of M, that would contain a very large number of matroids. Knuth [11] constructed such a matroid with enough non-bases to provide a new lower bound on m_n, the number of matroids on groundset $[n]$. Later Graham and Sloane [6] devised a construction that guaranteed a matroid with twice as many non-bases as Knuth's, and today that construction continues to provide the best general lower bound on m_n.

Graham and Sloane's construction was as follows

Definition 3.1.1 (Graham-Sloane matroid)**.** A matroid M on n elements is *Graham-Sloane* (or GS) if and only if there is some labelling $\sigma : E(M) \to [n]$ and $m \in \{0, 1, \ldots, n-1\}$ such that for every *non-basis L* of M,

$$\sum_{i \in L} \sigma(i) \equiv m \mod n$$

We note that GS matroids are closed under duals (easily checked) but not under minors. This latter fact is not so obvious, but can be checked by observing the two matroids in Figure 3.1. In that illustration, the sparse paving matroid M has 8 elements - which we label $1, 2 \cdots, 8$ - and its non-bases are precisely every trio of elements $\{a, b, c\} \subset \{1, 2, \ldots, 8\}$ such that $a + b + c \equiv 0 \mod 8$. However if we delete the element labelled "2" then the matroid M' that we obtain has 7 elements, 5 non-bases and one element e (that which was formerly labelled "6") which appears in only one non-basis (a "loose element"). This is not a GS matroid. To see this, allow any labelling of the vertices with $\{1, 2, \ldots, 7\}$. For each $m \in \{1, 2, \ldots, 7\}$ let $\mathcal{L}_{[7],m} = \{\{a, b, c\} \subset \{1, 2, \ldots, 7\} : a + b + c \equiv m \mod 7\}$. Now if M' is GS, $\mathcal{L}(M')$ must be a substructure of some $\mathcal{L}_{[7],m}$. But this is impossible: for each m, $\mathcal{L}_{[7],m}$ has only 5 lines, therefore we must have $\mathcal{L}(M') \cong \mathcal{L}_{[7],m}$ for some m. But no $\mathcal{L}_{[7],m}$ contains a loose element (easily checked by computer; in fact, the $M_{[7],m}$ are pairwise isomorphic).
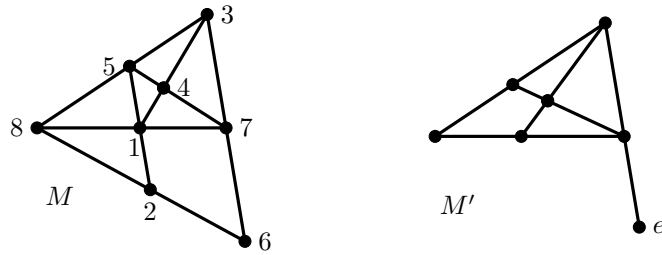

Figure 3.1: A Graham-Sloane matroid M and its non-GS minor $M' = M \setminus \{2\}$. Under the labelling shown, the elements in each line (non-basis) of M sum to $0 \mod 8$.

In this chapter we shall consider challenges of counting GS matroids, and

show that they vanish in the sparse paving matroids (this is not irreconcilable with them providing the best lower bound, since that bound is typically considered only at a log-log level).

We shall then consider the question of minor inclusion between GS matroids, inspired by the paper of Pendavingh and van der Pol [24] which identified $M(K_4)$ and the Vámos matroid as being forbidden minors in GS matroids of odd size (recall that we use "forbidden minor" without the implication of minimality that sometimes accompanies the term - so "M is forbidden in $\mathcal{M}$" means simply that no matroid in $\mathcal{M}$ has an M-minor). We shall exhibit a family of matroids whose p-th member is a forbidden minor in GS matroids except those of size divisible by p, for prime p. By showing however that such a matroid is contained in asymptotically almost every GS matroid on np elements, we are able to develop a surprising result - essentially stating that almost every GS matroid is forbidden as a minor in almost all other GS matroids.

In the final part of the chapter, we shall exhibit a class of sparse paving matroids which are forbidden minors of the entire class of GS matroids. In particular we shall then show that, for $r \geq 4$, asymptotically almost all sparse paving matroids of rank r contain one of these sparse paving matroids as a minor. This then implies that asymptotically almost all sparse paving matroids of rank r are forbidden minors of the class of GS matroids.

Constructing GS matroids from the Johnson graph

Given a stable set in the Johnson Graph $J(n, r)$, any subset of this set is also stable. In particular, if our stable set has size α, then its subsets are 2^α in number. Since each of these defines a (sparse paving) matroid, constructing

large stable sets in the Johnson graph is a highly effective method of generating large collections of matroids. This route was first pursued in 1974 by Knuth [11] who found a construction that guaranteed, for general n and r, a stable set of size at least $\frac{1}{2n}\binom{n}{r}$. This was later improved upon by Graham and Sloane [6] whose construction, which we shall describe, guarantees $\frac{1}{n}\binom{n}{r}$. In particular, where $r = \lfloor \frac{n}{2} \rfloor$, this generates a collection of matroids so big that it provides the best known general lower bound on m_n, the number of matroids on groundset $[n]$.

Before returning to the Graham-Sloane construction, we shall put it in the context of theoretical bounds on the size of stable sets in the Johnson Graph $J(n,r)$, and introduce another important source of large stable sets.

Definition 3.1.2 (Steiner system)**.** A *Steiner system $S(t, k, n)$* on the elements $[n]$ consists of a collection of k-element subsets of $[n]$, such that every t-element subset of $[n]$ is contained in precisely one member of the Steiner system.

Note that the existence of Steiner systems is only possible when certain conditions upon t, k and n are met, and even then is not guaranteed.

Definition 3.1.3 (Independence number)**.** The *independence number $\alpha(G)$* of a graph G is the cardinality of the largest stable set in G.

The following result is well-established in design theory.

Lemma 3.1.4. *For all n and $0 \le r \le n$ we have*

$$\alpha(J(n,r)) \le \frac{1}{n+1-r}\binom{n}{r}$$

Proof. Let I be a stable set of $J(n,r)$. Each set in $[n]^{(r-1)}$ can only be contained in one member of I, since any two r-sets containing the same $(r-1)$-set are necessarily adjacent in the Johnson Graph. Moreover, the number of $(r-1)$-sets contained in each r-set is r. This gives

$$\alpha(J(n,r)) \leq \frac{1}{r}\binom{n}{r-1} = \frac{r!(n-r)!}{r(r-1)!(n+1-r)!}\binom{n}{r} = \frac{1}{n+1-r}\binom{n}{r}$$

where the middle expression is arrived at by considering the expansions of $\binom{n}{r}$ and $\binom{n}{r-1}$. $\square$

Equality is achieved by the case where the vertices of the stable set correspond to the members of the Steiner system $S(r-1, r, n)$, when that exists.

We can now show that the Graham-Sloane construction finds a large stable set in $J(n,r)$ - thereby proving that it does indeed define a sparse paving matroid, and that ths matroid has a large number of non-bases.

Define a function $f : [n]^{(r)} \to [n]$ such that $f(X) = \sum_{x \in X} x \mod n$. Note that in every rank-r GS matroid on $[n]$, the non-bases must all take the same value under f.

Claim 3.1.4.1. If $f(X) = f(Y)$ then $X \not\sim Y$; that is to say X and Y are not adjacent in the Johnson Graph $J(n,r)$.

Proof. Suppose X, Y adjacent in the Johnson graph. Then there exist x, y with $x \neq y$ such that $X - Y = x, Y - X = y$. But then $G(X) - G(Y) = x - y \not\equiv 0 \mod n$, since $|x - y| < n$ and $x \neq y$. $\square$

Therefore, by partitioning the vertices of $J(n,r)$ according to the value they take under f, we will split the Johnson graph into n stable sets, at least one of which must have size greater than or equal to $\frac{1}{n}\binom{n}{r}$. And now every subset of the vertices in this part correspond to the non-bases of a distinct GS matroid. From this, setting $r = \lfloor \frac{n}{2} \rfloor$, we obtain a lower bound on $m_{n,r}$, the number of rank-r matroids on n elements:

$$\log m_{n,r} \geq \frac{1}{n}\binom{n}{\lfloor n/2 \rfloor}$$

We note that the Graham-Sloane method, whilst powerful and currently providing the best general lower bound on the size of the largest stable set in $J(n,r)$, is not optimal for all n, r. It is beaten by the construction where members of the stable set correspond to members of the Steiner system $S(r-1, r, n)$, whenever that exists. For fixed r, Keevash's [10] result on the existence of designs guarantees the existence of such systems for arbitrarily large n. However in the important case of $r = \lfloor n/2 \rfloor$, such systems are only known to exist for $n = 7$ and $n = 12$ (the first of these gives rise to the Fano plane) and we do not know if these systems exist for arbitrarily large n. If they were to exist, they would provide stable sets roughly twice the size of those found in the Graham-Sloane construction. This would be a significant result as it would give, at least for sporadic values of n, a lower bound on the number of matroids m_n that asymptotically approaches the current upper bound at a log-log level.

3.2 Counting Graham-Sloane Matroids

This section is preparation for later work in which we show the Graham-Sloane matroids to have remarkable minor-forbidding properties. The arguments later on will make use of the results proved in this section. Here we seek to get a grip on the problem of counting GS matroids across all possible orderings of elements. The motivation of this work is to facilitate the extension of results from the class of GS matroids defined on a particular labelling σ of $[n]$ to all GS matroids with groundset $[n]$.

We first want to define some notation:

Let $\mathbb{K}_n$ define the collection of all GS matroids on groundset $[n]$, and set $k_n := |\mathbb{K}_n|$.

Let $\sigma : [n] \to [n]$ be any permutation on n elements and m any integer in $[n]$. We define:

$$\mathbb{K}_{n,\sigma,m} := \{M \in \mathbb{K}_n : \text{ for all } L \in \mathcal{L}(M), \sum_{i \in L} \sigma(i) \equiv m \mod n\}$$

$$\mathbb{K}_{n,\sigma} := \bigcup_{m \in [n]} \mathbb{K}_{n,\sigma,m}$$

Finally define $k_{n,\sigma} := |\mathbb{K}_{n,\sigma}|$.

In particular we shall frequently refer to $\mathbb{K}_{n,\mathrm{id}}$ and $k_{n,\mathrm{id}}$, where id denotes the identity permutation.

Counting the set $\mathbb{K}_{n,\mathrm{id}}$ is straightforward. What about the set $\mathbb{K}_n$ of all GS matroids on groundset $[n]$? This becomes complicated because of the pos-

sibility that individual matroids may arise separately under many different permutations of $[n]$.

In order be able to count effectively in $\mathbb{K}_n$, we need to control the likelihood of matroids arising under more than one permutation. We first shall show that permutations on $[n]$ can be split into a number of equivalence classes that we shall call *addition-equivalent*. We write $\sigma_1 \cong \sigma_2$ when σ_1, σ_2 are in the same addition-equivalent class. We shall show that, for a randomly chosen $M \in \mathbb{K}_{n,\sigma}$, the following holds:

$$\Pr(\exists \sigma' : \sigma' \not\cong \sigma, M \in \mathbb{K}_{n,\sigma'}) \to 0 \text{ as } n \to \infty$$

The lemma we shall establish is of importance for later theorems in this thesis as it enables us to effortlessly extend results proved on $\mathbb{K}_{[n],\mathrm{id}}$ to $\mathbb{K}_{[n]}$.

We first need to define addition-equivalence for permutations:

Definition 3.2.1 (Addition-equivalent)**.** Permutations σ_1 and σ_2 on $[n]$ are *addition-equivalent* if, for any $X, Y \subseteq [n]$ with $|X| = |Y|$, the following holds:

$$\sum_{x \in X} \sigma_1(x) \equiv \sum_{y \in Y} \sigma_1(y) \mod n \text{ if and only if } \sum_{x \in X} \sigma_2(x) \equiv \sum_{y \in Y} \sigma_2(y) \mod n$$

It's easy to see that addition-equivalence defines an equivalence relation on permutations (reflexivity and symmetry are immediate; transitivity follows by composition of permutations). We also note the relevant implication of this definition: if σ_1 and σ_2 are addition equivalent, then $\mathbb{K}_{n,\sigma_1} = \mathbb{K}_{n,\sigma_2}$

Lemma 3.2.2. *Permutations σ and σ' on $[n]$ are addition-equivalent if and only if, for all $x \in [n]$, $\sigma'(x) = \alpha\sigma(x) + \beta$ where α is coprime to n, and $\beta \in \{0, \ldots, n-1\}$. (We are working modulo n, so $n \equiv 0$.)*

Note that if $\sigma(x) = \alpha x + \beta$ for α not coprime to n, then σ is addition-preserving, but not a permutation!

Proof. Without loss of generality we may assume $\sigma = \mathrm{id}$.

Firstly, suppose $\sigma'(x) = \alpha x + \beta$ for permissible α, β.

Take sets X, Y of size r with $\sum_{x \in X} \sigma(x) \equiv \sum_{y \in Y} \sigma(y) \mod n$.

$$\sum_{x \in X} \sigma'(x) \equiv \alpha \sum_{x \in X} x + r\beta \equiv \alpha \sum_{y \in Y} y + r\beta \equiv \sum_{y \in Y} \sigma'(y) \mod n$$

So σ' is addition-equivalent to id. Conversely, assume σ' is addition equivalent to id. Then for any $x_1, x_2, y_1, y_2 \in [n]$ such that $x_1 + x_2 = y_1 + y_2$, we have $\sigma'(x_1) + \sigma'(y_1) = \sigma'(x_2) + \sigma'(y_2)$. But in that case, for any x we have $\sigma'(x) + \sigma'(x+1) = \sigma'(x-1) + \sigma'(x+2)$ and $\sigma'(x+2) + \sigma'(x-2) = \sigma'(x+1) + \sigma'(x-1)$. Continuing, some straightforward algebraic manipulation gives us:

$$\sigma'(x) + \sigma'(x+1) + \sigma'(x+2) + \sigma'(x-2) = \sigma'(x-1) + \sigma'(x+2) + \sigma'(x+1) + \sigma'(x-1)$$

$$\sigma(x) + \sigma(x-2) = 2\sigma(x-1)$$

$$\sigma'(x) - \sigma'(x-1) = \sigma'(x-1) - \sigma'(x-2)$$

A simple induction then gives:

$$\sigma'(x) = (\sigma'(1) - \sigma'(0))x + \sigma'(0)$$

So if we set $\alpha := \sigma'(1) - \sigma'(0)$ and $\beta := \sigma'(0)$ we have what looks like the desired form for σ'. It only remains to check that $\alpha = \sigma'(1) - \sigma'(0)$ is coprime to n. But this must be the case, else there would be some $k < n$ such that $\alpha k \cong 0 \mod n$. And that would imply $\sigma'(k) = \sigma'(0)$, which cannot be the case since σ' is a permutation of $[n]$. $\qquad\square$

Corollary 3.2.3. *Let σ_1 and σ_2 be permutations of $[n]$ that are not addition equivalent. Then there exist $x_1, x_2, y_1, y_2 \in [n]$ such that $\sigma_1(x_1) + \sigma_1(x_2) = \sigma_1(y_1) + \sigma_1(y_2)$ but $\sigma_2(x_1) + \sigma_2(x_2) \neq \sigma_2(y_1) + \sigma_2(y_2)$.*

Proof. Suppose not, then we can argue as in the latter part of our proof of Lemma 3.2.2 in order to show σ_1 and σ_2 are addition equivalent. $\qquad\square$

Corollary 3.2.4. *The number of addition equivalent classes of permutations is at least $(n-2)!$*

Proof. There are $n!$ possible permutations and each addition equivalent class has size at most $n(n-1)$, as determined by the maximum number of choices of α, β. $\qquad\square$

Another requirement for our probabilistic argument is a result that binds asymptotically almost all GS matroids on $[n]$ to rank approximately $n/2$.

Lemma 3.2.5. *For any $0 < \delta$ and $\epsilon \leq 1$ there is some $N(\epsilon)$ such that, for any $n > N$ and $\sigma : [n] \to [n]$:*

$$|\{M \in \mathbb{K}_{n,\sigma} : (0.5 - \delta)n \leq r(M) \leq (0.5 + \delta)n\}| \geq (1 - \epsilon)k_{n,\sigma}$$

That is to say, asymptotically almost all matroids in $\mathbb{K}_{n,\sigma}$ have rank close to $n/2$.

Proof. Suppose $r < \frac{(1-\delta)n}{2}$. We can say that

$$\frac{\binom{n}{r}}{\binom{n}{\lfloor n/2 \rfloor}} = \frac{(\lfloor (0.5-\delta)n \rfloor + 1)(\lfloor (0.5-\delta)n \rfloor + 2) \cdots (\lfloor n/2 \rfloor)}{(\lceil n/2 \rceil + 1)(\lceil n/2 \rceil + 2) \cdots (\lceil (0.5+\delta)n \rceil)}$$

$$\leq \left(\frac{0.5 - \delta/2}{0.5 + \delta/2} \right)^{\delta n/2} = \left(\frac{1-\delta}{1+\delta} \right)^{\delta n/2}$$

Let $k_{n,\sigma,r}$ denote the number of rank-r matroids in $\mathbb{K}_{n,\sigma}$. A crude upper bound for this is $2^{\binom{n}{r}}$, which is the number of subsets of $[n]^{(r)}$. Now we can say

$$\log \left(\sum_{r<(0.5-\delta)n} k_{[n],\sigma,r} \right) \leq \log \left(\sum_{r<(0.5-\delta)n} 2^{\binom{n}{r}} \right)$$

$$\leq \log n + \binom{n}{\lfloor (0.5-\delta)n \rfloor} \leq \log n + \left(\frac{1-\delta}{1+\delta} \right)^{\delta n/2} \binom{n}{\lfloor n/2 \rfloor}$$

$$= (1 + o(1))n \left(\frac{1-\delta}{1+\delta} \right)^{\delta n/2} \log k_{n,\sigma,\lfloor n/2 \rfloor}$$

Clearly then for large enough n we have

$$\sum_{r<(0.5-\delta)n} k_{[n],\sigma,r} \leq \frac{\epsilon}{2} k_{n,\sigma,\lfloor n/2 \rfloor} \leq \frac{\epsilon}{2} k_{n,\sigma}$$

Since $\mathbb{K}_n$ is self-dual, we can obtain a similar result for $r > (0.5 + \delta)$ and these together imply the desired result. $\square$

Finally we need a lemma to express a somewhat obvious property of the Graham-Sloane construction. The construction partitions $J(n, r)$ into n stable sets, and it is intuitively clear that these stable sets have roughly equal size. For our purposes it suffices to simply put a crude lower bound on the size of these stable sets. Since we only use this result once, the bounds chosen are somewhat arbitrary.

Lemma 3.2.6. *Let $X \subseteq [n]$ such that $|X| > 9n/10 + 6$ and let $4n/10 + 3 < t < 6n/10$. Choose $m \in \{0, 1, \ldots, n-1\}$. Let*

$$Z := \{T \in X^{(t)} : \sum_{x \in T} x \equiv m \mod n\}$$

For large n, we must have $|Z| \geq 4^{4n/10}$.

Proof. In accordance with our needs, the proof is somewhat crude. For any $m \in \{0, 1, \ldots, n-1\}$, we know X contains at least $4n/10$ pairs of form $(x, m - x)$ (subtraction modulo n). Start to build $T \in X^{(t)}$ as follows: if t is even then choose a pair (x_1, x_2) of elements in X such that $x_2 = m - x_1$, and add this pair to T. If t odd then choose any x_0, x_1, x_2 so that $x_2 = (m - x_0) - x_1$ and add this triplet to T. Now in either case we are able to complete T by adding only pairs of form $(x, -x)$ - we have at least $4n/10$ such pairs to choose from and are choosing at least $2n/10$ pairs and at most $3n/10$ such pairs. So the number of choices for T is at least $\binom{4n/10}{n/10} \geq 4^{4n/10}$ $\square$

Now for the main lemma and proof.

Lemma 3.2.7. *Let σ be any permutation of $[n]$. Choose M uniformly at random from $\mathbb{K}_{n,\sigma}$.*

For any $0 < \epsilon \le 1$, there is some $N(\epsilon)$ such that if $n > N$:

$$\Pr(\exists \sigma' : \sigma' \not\cong \sigma, M \in \mathbb{K}_{n,\sigma'}) \le \epsilon$$

Proof. We may assume without loss of generality that $\sigma = \mathrm{id}$, so we are drawing M uniformly at random from $\mathbb{K}_{n,\mathrm{id}}$. We also impose the condition that $\frac{4n}{10} \le r(M) \le \frac{6n}{10}$. Let $m \in \{0, \dots, n-1\}$ be the unique choice so that $M \in \mathbb{K}_{n,\mathrm{id},m}$.

Suppose there is some σ' such that $\sigma' \not\cong \mathrm{id}$, but $M \in \mathbb{K}_{n,\sigma'}$. By Corollary 3.2.3 we can choose some $x_1, x_2, y_1, y_2 \in [n]$ such that $x_1 + y_1 = x_2 + y_2$ but $\sigma'(x_1) + \sigma'(y_1) \ne \sigma'(x_2) + \sigma'(y_2)$. Since $M \in \mathbb{K}_{n,\sigma'}$ we know that there cannot exist non-bases L_1, L_2 of M such that $L_1 \backslash L_2 = \{x_1, x_2\}$ and $L_2 \backslash L_1 = \{y_1, y_2\}$. We want to show that, within $\mathbb{K}_{n,\mathrm{id}}$, the chances of such a quartet x_1, x_2, y_1, y_2 existing are vanishingly small..

Consider the family F of ordered sets $\{x_1, x_2, y_1, y_2\} \in [n]^{(4)}$ such that $x_1 + x_2 \equiv y_1 + y_2 \mod n$. Clearly $|F| \le n^4$. Let $F' \subseteq F$ be the collection of sets $\{x_1, x_2, y_1, y_2\} \in F$ asuch that no pair L_1, L_2 of non-bases of M with $L_1 \setminus L_2 = \{x_1, x_2\}$ and $L_2 \setminus L_1 = \{y_1, y_2\}$. Now

$$\mathrm{E}(|F'|) = \sum_{X \in F} \Pr(X \in F') \le n^4 P$$

where P is simply the probability that, given a randomly chosen set $\{x_1, x_2, y_1, y_2\} \in F$, there will not exist non-bases L_1, L_2 of M with $L_1 \setminus L_2 = \{x_1, x_2\}$ and $L_2 \setminus L_1 = \{y_1, y_2\}$.

But for any given choice of $\{x_1, x_2, y_1, y_2\}$ the number of candidates for such a L_1, L_2 is determined by the choice of the intersection $U = L_1 \cap L_2$. Since $\frac{4n}{10} \le r(M) \le \frac{6n}{10}$ then we can apply Lemma 3.2.6 to see that we have at

least $4^{4n/10}$ choices for U such that the sum of elements in U is equal to $(m - x_1 - y_1)$ modulo n. In each case, consider the resulting pair of sets $U \cup \{x_1, x_2\}$ and $U \cup \{y_1, y_2\}$. These form a pair of non-bases with probability $\frac{1}{4}$. So the probability of no such pair of non-bases existing is $(1 - \frac{1}{4})^{4^{4n/10}}$. Now this probability shrinks much faster than n^{-4}, that is to say $\mathrm{E}(|F'|) \to 0$ as $n \to \infty$, so $\Pr(|F'| > 0) \to 0$ as $n \to \infty$.

Now given any $0 < \epsilon \le 1$ the chances of a matroid being generated under a different permutation can be made less than $\frac{\epsilon}{2}$ by choice of n And similarly Lemma 3.2.5 says that for large enough n at least $(1 - \frac{\epsilon}{2})k_{n,\mathrm{id}}$ of GS matroids on $[n]$ have rank $\frac{4n}{10} \le r \le \frac{6n}{10}$. Taking n large enough to satisfy both of the above gives the desired result. $\square$

Corollary 3.2.8. *For positive integers n,*

$$m_n > \mathbb{K}_n > (1 + o(1))(n-2)! 2^{\frac{1}{n}\binom{n}{\lfloor n/2 \rfloor}}$$

Proof. This follows from Lemma 3.2.7 and the lower bound on the number of addition preserving classes given in Corollary 3.2.4. $\square$

We are also now able to prove that asymptotically almost all sparse paving matroids are not GS.

Theorem 3.2.9. *Let M be chosen uniformly at random from $\mathbb{S}_n$. Asymptotically almost certainly M is not GS.*

Note this obviously implies the same is true when M is taken from $\mathbb{M}_n$.

Proof. Let $\mathbb{K}_n^+$ be defined as follows: $M \in \mathbb{K}_n^+$ if for some $m \in \{0, \ldots, n-1\}$

we have some permutation σ of $[n]$ and precisely one non-basis $L \in \mathcal{L}(M)$ such that

$$\sum_{e \in L} \sigma(e) \not\equiv m \mod n$$

and, for every other non-basis, the elements of that non-basis under σ sum to $m(\mod n)$. In other words, matroids in $\mathbb{K}_n^+$ are like GS matroids with a single rogue non-basis. Define $\mathbb{K}_{n,\sigma}^+$ and $\mathbb{K}_{n,\sigma,m}^+$ to be those matroids for which this condition holds with σ and m respectively.

In general all the results we have shown so far for $\mathbb{K}_n$ apply equally to $\mathbb{K}_n^+$ - we shall use some of these results without proof, since the proofs would be essentially identical to those already exhibited.

Let Φ define a many to many correspondence between $\mathbb{K}_n$ and $\mathbb{K}_n^+$ as follows: we say $(M, M') \in \Phi$ if and only if there exists σ, m such that $M \in \mathbb{K}_{n,\sigma,m}$ and $M' \in \mathbb{K}_{n,\sigma,m}^+$, and M' can be obtained from M by adding a single non-basis L and relaxing all non-bases of M that intersect with L in $r(M) - 1$ elements.

By a double counting argument we shall show that for any choice of σ we have, as $n \to \infty$,

$$\frac{k_{n,\sigma}}{|\mathbb{K}_{n,\sigma}^+|} \to 0$$

By Lemma 3.2.7 (and its obvious equivalent for $\mathbb{K}_n^+$) it is enough to show that, as $n \to \infty$,

$$\frac{k_n}{|\mathbb{K}_n^+|} \to 0$$

So, if $M \in \mathbb{K}_{n,\sigma,m}$ then how many choices of M' do we have such that

$(M, M') \in \Phi$? This is simply equal to the number of choices of a non-basis L whose elements do not sum to m under σ. We know (e.g. due to the upper bound on the number of matroids) that at most $\frac{2}{n}\binom{n}{r}$ choices of L can have elements summing to m, so the number of choices of L is $(1 + o(1))\binom{n}{r}$.

If $M' \in \mathbb{K}^+_{n,\sigma,m}$ then how many choices of M do we have such that $(M, M') \in \Phi$? It is clear which non-basis we have to relax, but we have a choice of which non-bases we add back in. There are r possible non-bases that could be added (delete an element of L, and replace it with the element that makes the elements of our new non-basis sum to m). We have a free choice of whether or not to include each of these non-bases, giving us a total choice of $2^r(M)$.

And by Lemma 3.2.5 (and its obvious equivalent for $\mathbb{K}^+_{n,\sigma}$) we can assume that $r(M) = r(M') \sim n/2$, so

$$\frac{k_{n,\sigma}}{|\mathbb{K}^+_{n,\sigma}|} = \frac{2^r}{(1 + o(1))\binom{n}{r}} \to 0$$

as $n \to \infty$. $\qquad\square$

Although this does show that GS matroids are heavily outnumbered by non-GS matroids (and indeed the bounds would vanish even more quickly if we considered matroids with a larger number of rogue non-bases) it seems that this line of argument is not by itself sufficient to provide an increase on the lower bound for m_n at a log-log level.

3.3 Minor inclusion & partially GS-forbidden minors

Definition 3.3.1 (Partially GS-forbidden). A matroid M is said to be *partially GS-forbidden* if there exists at least one integer n such that no matroid in $\mathbb{K}_n$ has an M-minor.

In [24], Pendavingh and van der Pol notably showed that $M(K_4)$ and the Vámos matroid were forbidden minors in GS matroids of odd size. They did this by showing that the relations between the non-bases of these matroids forced an equation in modular arithmetic that could only be solved if $2|n$. By extending this idea we shall later construct a sequence of GS matroids, the p-th member of which is a forbidden minor in GS matroids of size not divisible by p.

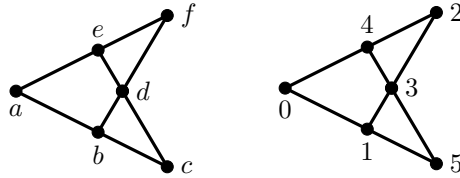We first outline the argument of Pendavingh and van der Pol, as used for $M(K_4)$.

Figure 3.2: $M(K_4)$. On the right hand side we have labelled the elements as $0, 1, \cdots 5$ such that for every $L \in \mathcal{L}(M(K_4))$ we have $\sum_{i \in L} i \equiv 0 \mod 6$

Theorem 3.3.2 (Pendavingh, van der Pol 2014). $M(K_4)$ *is not a minor of any GS matroid on an odd number of elements.*

We note that $M(K_4)$ certainly can be a minor of GS matroids on even numbers of elements - indeed it is itself a GS matroid, as can be seen from the

labelling on the right hand side of Figure 3.2.

Proof. Let M be a GS matroid in $\mathbb{K}_n$. Without loss of generality we may assume $M \in \mathbb{K}_{n,\mathrm{id}}$ and that for all $L \in \mathcal{L}(M)$, $\sum_{i \in L} i \equiv 0 \mod n$.

Suppose $M(K_4)$ is a minor of a M - then there exists disjoint sets $A, B \subseteq E(M)$ such that $M/A \setminus B \cong M(K_4)$. Let's suppose the elements of $M/A \setminus B$ take the labelling shown in the left hand part of Figure 3.2. Assuming without loss of generality that A is independent, we now have non-bases of M that are equal to $A \cup \{a, b, c\}, A \cup \{a, e, f\}, A \cup \{b, d, f\}, A \cup \{c, d, e\}$. The elements in each non-basis sum to $0 \mod n$, so we find that

$$a + b + c \equiv a + e + f \equiv b + d + f \equiv c + d + e \equiv -\sum_{i \in A} i \mod n$$

$$(a + b + c) + (a + e + f) - (b + d + f) - (c + d + e) \equiv 0 \mod n$$

$$2(a - d) \equiv 0 \mod n$$

If n were odd then 2 would be coprime to n, so we obtain $n | a - d$. But a and d are distinct elements of $[n]$, so $a - d \not\equiv 0 \mod n$, a contradiction. So n can only be even. $\square$

It turns out $M(K_4)$ is the smallest member of a much larger class, which we shall explore in the following section.

3.3.1   A class of partially GS-forbidden matroids

Let H_p be a sparse paving matroid of rank three defined on the set of elements $\{a, b\} \cup \{x_i, y_i : 0 \leq i < p\}$ by its set of non-bases:

$$\mathcal{L}(H_p) := \bigcup_{0 \leq i < p} \{\{a, x_i, y_i\}, \{b, x_i, y_{i+1}\}\}$$

where addition of indices is modulo p. An example of this configuration is given in Figure 3.3. It is simple to check - and easily seen from a line drawing - that H_p is indeed a sparse paving matroid.
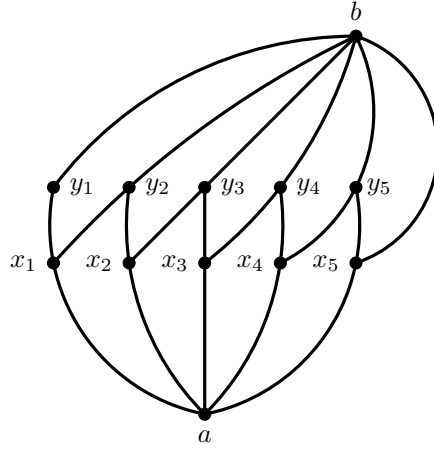

Figure 3.3: H_5

Although $H_2 \cong M(K_4)$ and is therefore GS we find that for $p > 2$, H_p is not GS. Suppose it were, then we may derive that

$$y_2 - y_1 \equiv y_3 - y_2 \equiv \cdots \equiv y_p - y_{p-1} \equiv y_1 - y_p \equiv a - b \mod 2p + 2$$

But then

$$p(a - b) \equiv \sum_{i=1}^{p} y_i - \sum_{i=1}^{p} y_i \equiv 0 \mod 2p + 2$$

which is impossible, since p does not divide $2p + 2$.

Let $\mathcal{H} = \{H_p : p \text{ prime}\}$. For primes $p > q$, $H_p \not\succeq H_q$, so $\mathcal{H}$ is an antichain under minor inclusion. Although this can be shown directly it will also be implied by (and not used in the proofs of!) Lemmas 3.3.3 and 3.3.6.

Lemma 3.3.3. *For any prime p, the matroid H_p cannot be contained as a minor in any GS matroid whose size is coprime to p.*

Proof. Say we have M a GS matroid on n elements. Working as in the Pendavingh / van der Pol proof for $M(K_4)$ we obtain that there is an equation

$$(a + x_0 + y_0) + \cdots + (a + x_{p-1} + y_{p-1}) \equiv (b + x_0 + y_1) + \cdots + (a + x_{p-1} + y_0) \mod n$$

which reduces to

$$pa + \sum_{0 \le i < p} (x_i + y_i) \equiv pb + \sum_{0 \le i < p} (x_i + y_i) \mod n$$

which in turn gives $p(a - b) \equiv 0 \mod n$, providing a contradiction to the uniqueness of a and b except in the cases where $p \mid n$. $\square$

3.3.2 Most Graham-Sloane matroids are partially GS-forbidden

Definition 3.3.4 (Density)**.** We say a set $X \subseteq \mathbb{N}$ has density d in the integers if

$$\lim_{n \to \infty} \frac{|X \cup [n]|}{n} = d$$

In this section we show that most (in some sense) Graham-Sloane matroids are forbidden as minors in the Graham-Sloane matroids of certain sizes. We achieve this by showing that with high probability they contain as a minor some member of our class $\mathcal{H}$. We shall present progressively stronger versions of this result, culminating in the following.

Theorem 3.3.5. *Define* $X_\epsilon \subseteq \mathbb{N}$ *as follows:* $n \in X_\epsilon$ *if and only if there exists some set* M *of density at least* $1 - \epsilon$ *in the integers, such that for any* $m \in M$, *at least* $(1 - \epsilon)k_n$ *matroids in* $\mathbb{K}_n$ *are forbidden minors of* $\mathbb{K}_m$.

Then X_ϵ *has density 1 in* $\mathbb{N}$.

Which roughly says that most Graham-Sloane matroids are forbidden as a minor in Graham-Sloane matroids of most sizes.

Lemma 3.3.6. *Fix a prime* p *and let* $n \to \infty$. *Consider GS matroids in* $\mathbb{K}_{np,\sigma}$ *where* σ *is any permutation of* $[np]$. *Uniformly across all choices of* σ, *asymptotically almost all GS matroids in* $\mathbb{K}_{np,\sigma}$ *contain* H_p *as a minor.*

Proof. Imagine drawing M uniformly at random from rank-r matroids in $\mathbb{K}_{np,\sigma,m}$. Without loss of generality we may assume that $np/4 \leq r(M) \leq 3np/4$ (by Lemma 3.2.5). As usual we can also assume that $\sigma = \text{id}$.

The proof relies on describing a set X_p of $2p + 2$ elements and showing that with high probability (uniform across all m, r) there is some set S such that $(M/S)|X_p \cong H_p$. We claim that

$$\Pr(M \succeq H_p) > 1 - \epsilon$$

Now, any r-set of $[np]$ whose elements sum to m is included in $\mathcal{L}(M)$ with independent probability $\frac{1}{2}$. Let $X_p = \{a, b, x_1, x_2, \ldots, x_p, y_1, y_2, \ldots, y_p\}$ where $a := n, b := 2n, x_i := in - 1, y_i := (p - i - 1)n + 1$. Note that $\forall 1 \leq i \leq p$, we have $a + x_i + y_i \equiv b + x_i + y_{i+1} \equiv 0 \mod np$, where the addition in the subscript is modulo p.

We seek a set S with

$$S \subset [np] \setminus X_p, |S| = r - 3, \sum_{s \in S} s \equiv m \mod np$$

Claim 3.3.6.1. For sufficiently large n, the number of choices of S is at least

$$\binom{np/2 - 2p - 5}{np/8 - 3}$$

Proof of claim. Firstly if $m \pmod{n} \notin \{-1, 0, 1\}$, include m in S. Else include $m - 3$ and 3 in S. Now we have either $r - 4$ or $r - 5$ spaces to fill in S, without loss of generality we may assume this number to be even (if not we can include np in S). This leaves at least $r - 6$ elements remaining to be chosen in S, from at least $np - 2p - 5$ elements in $[np]$. We complete S by including pairs of the form $(s, np - s)$: we have at least $\binom{np/2 - 2p - 5}{np/8 - 3}$ choices of these pairs. $\qquad\square$

For large enough n we have $\binom{np/2-2p-5}{np/8-3} > n$ and so we can find at least n sets satisfying our conditions for S, label these candidate sets $S_1, S_2, \ldots, S_n$.

Let Z_i be the event that $(M/S_i)|X_p \cong H_p$. But consider a set $X \in X_p^{(3)}$. If $\sum_{x \in X} \sigma(x) \equiv 0 \mod n$ then $S_i \cup X$ is a non-basis of M with probability $\frac{1}{2}$, independent of the other non-bases of M - otherwise $S_i \cup X$ cannot be a non-basis of M.

It is clear then that each possible choice of $(M/S_i)|X_p$ is equally likely and the number of possible choices is bounded above by $2^{\binom{2p+2}{3}}$. Moreover, one possible choice is to include as non-bases precisely the sets of form $\{a, x_i, y_i\}$ and $\{b, x_i, y_{i+1}\}$ - in which case $(M/S)|X_p \cong H_p$. So

$$\Pr(Z_i) \geq 2^{-\binom{2p+2}{3}}$$

Since the Z_i arise independently, we have that, as $n \to \infty$,

$$\Pr(Z_1 \cup Z_2 \cup \cdots \cup Z_n) \geq 1 - (1 - 2^{-\binom{2p+2}{3}})^n \to 1$$

$\square$

We can extend this result to all GS matroids:

Theorem 3.3.7. *Let $\mathcal{K}_p$ be the class of matroids not containing H_p as a minor. For any prime p and $0 < \epsilon \leq 1$ there exists $N(\epsilon)$ such that, for $n > N$, $|\mathcal{K}_p \cap \mathbb{K}_{np}| \leq \epsilon k_{np}$.*

Proof. Lemma 3.2.7 showed that for sufficient n almost all GS matroids in $\mathbb{K}_{np,\sigma}$ are not in $\mathbb{K}_{np,\sigma'}$ for any $\sigma' \neq \sigma$. So for any $0 < \epsilon \leq 1$, by choosing N large enough we can ensure that for any $n > N$, the following hold:

(i) For each permutation σ of $[np]$, at most $\frac{\epsilon}{2}k_{np,\sigma}$ matroids in $\mathbb{K}_{np,\sigma}$ do not have an H_p-minor.

(ii) For each permutation σ of $[np]$, at least $(1 - \frac{\epsilon}{2})k_{np,\sigma}$ GS matroids are unique to $\mathbb{K}_{np,\sigma}$.

So, taking $S([np])$ to be the set of all permutations of $[np]$, the number of matroids in $\mathbb{K}_{np}$ that have an H_p-minor is at least

$$\sum_{\sigma \in S([np])} (1 - \epsilon)k_{np,\sigma} = (1 - \epsilon) \sum_{\sigma \in S([np])} k_{np,\sigma} \leq (1 - \epsilon)k_{np}$$

$\square$

We can also achieve the following result.

Theorem 3.3.8. *Fix any $0 < \epsilon \leq 1$ and define $X_\epsilon \subseteq \mathbb{N}$ to be the set of $n \in \mathbb{N}$ such that the number of non-partially GS-forbidden matroids in $\mathbb{K}_n$ is at most ϵk_n.*

X_ϵ has density 1 in $\mathbb{N}$.

Put more informally, for almost every integer n, almost every GS matroid on $[n]$ is partially GS-forbidden.

Proof. Let ϵ be fixed and δ vary freely, $0 < \delta \leq 1$. Pick an initial sequence of the primes, $I_\delta = p_1, \ldots, p_s$ such that

$$\prod_{i=1}^{s} \frac{p_i - 1}{p_i} < \delta$$

That is to say, the density of multiples of these primes is at least $1 - \delta$. This is of course possible, since the limit of the density is 1 (every number is divisible by a prime number!). Let the set of these multiples be

$$X_\delta = \bigcup_{p \in I_\delta} \{n \in \mathbb{N} : p|n\}$$

Clearly X_δ has density at least $1 - \delta$ in the natural numbers.

For each prime $p_i, 1 \le i \le s$, we have by Theorem 3.3.7 some N_i such that for any $n > N_i$ with $p_i|n$, the number of non-partially GS-forbidden matroids on groundset $[n]$ is at most ϵk_n. Set $N(\delta, \epsilon)$ to be the largest such N_i. Now if $n > N(\delta, \epsilon), n \in X_\delta$ then $n \in X_\epsilon$. So X_ϵ has density 1 in X_δ, and hence density at least $1 - \delta$ in the integers. And since the value of δ was freely chosen from $(0, 1]$, the density of X_ϵ in the integers must be 1. $\qquad\square$

In fact, we can strengthen this further. This requires more definitions.

Definition 3.3.9 (*B*-smooth)**.** Let B be an integer. A positive integer is called *B-smooth* if none of its prime factors are greater than B.

Lemma 3.3.10. *For any integer B, the set of B-smooth integers has density 0 in the integers.*

This is a well-established result with many possible proofs. Here is one:

Proof. Suppose $\epsilon > 0$. Let $\mathcal{B}$ be the set of B-smooth integers. We shall show that the density of $\mathcal{B}$ must be less than ϵ. For this it suffices to show that for sufficiently large N, the density of $\mathcal{B}$ on $[N, 2N]$ is less than ϵ. Set $N > (B!)^{\log B - \log \epsilon}$. Then if we have $n \in \mathcal{B} \cap [N, 2N]$, we can see that there

must be some prime factor of n with multiplicity at least $\log B - \log \epsilon$. So

$$|\mathcal{B} \cap [N, 2N]| \leq \sum_{p=2}^{B} \frac{N}{p^{(\log B - \log \epsilon)}} \leq \frac{BN}{2^{(\log B - \log \epsilon)}} = \epsilon N$$

$\square$

This means that we can pick a new sequence of primes $p'_1, \ldots, p'_{s'}$,the density of whose multiples is at least $1 - \epsilon$, and such that $p'_1 > B$. Now we can strengthen our condition so that our matroids are not only partially GS-forbidden, but in fact forbidden as minors in GS matroids on all integers except some set of density less than ϵ: This is the result previously stated in Theorem 3.3.5, which we can now restate and prove.

Theorem 3.3.5. *Define $X_\epsilon \subseteq \mathbb{N}$ as follows: $n \in X_\epsilon$ if and only if there exists some set M of density at least $1 - \epsilon$ in the integers, such that for any $m \in M$, at least $(1 - \epsilon)k_n$ matroids in $\mathbb{K}_n$ are forbidden minors of $\mathbb{K}_m$.*

Then X_ϵ has density 1 in $\mathbb{N}$.

Proof. As in Theorem 3.3.8, but instead of picking an initial sequence of the primes, we pick an initial sequence of the set of primes greater than B, where $B = \frac{1}{\epsilon}$ $\square$

Casually we might phrase this as "For almost every choice of $m, n \in \mathbb{N}$, almost every GS matroid on groundset $[m]$ is a forbidden minor of the GS matroids on $[n]$".

3.4 Sparse paving matroids that are totally forbidden in GS matroids

In this section we establish that some sparse paving matroids are forbidden as minors in any GS matroids. Moreover, we shall show that there exist classes $\mathcal{N}_r$ of these matroids so that each member of $\mathcal{N}_r$ is a minor of asymptotically almost all sparse paving matroids of rank r. This means that asymptotically almost all sparse paving matroids of rank r are forbidden as minors in GS matroids.

Theorem 3.4.1. *There exists a sparse paving matroid which is not a minor of any GS matroid.*

Once again we use an argument similar to that used in Lemma 3.3.3 and originally in the paper of Pendavingh and van der Pol [24].

Proof. Let N be the sparse paving matroid on the groundset

$$\{a, a', b, c, d, x_1, x_2, y_1, y_2\}$$

with non-bases

$$\mathcal{L}(N) = \{\{a, b, x_1, x_2\}, \{a', b, y_1, y_2\}, \{c, d, x_1, x_2\}, \{c, d, y_1, y_2\}$$

We claim that N is not a minor of any GS matroid. A diagram of N is provided in Figure 3.4.

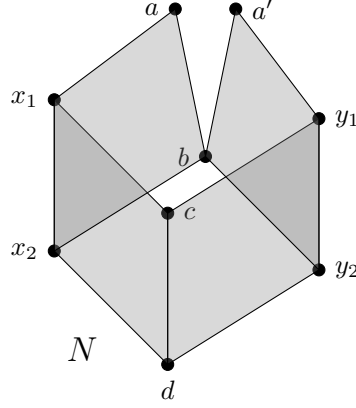The proof proceeds by contradiction. Assume the claim is false, so N is in

Figure 3.4: The matroid N described in the proof of Theorem 3.4.1

fact a minor of some GS matroid M. Without loss of generality we may assume that M is defined on groundset $[n]$. Then we can label each of $\{a, a', b, c, d, x_1, x_2, y_1, y_2\}$ with distinct values from $[n]$ such that the sum of elements in each non-basis of N is the same, modulo n. From the non-bases $\{c, d, x_1, x_2\}$ and $\{c, d, y_1, y_2\}$ we obtain that $x_1 + x_2 \equiv y_1 + y_2 \mod n$. Then from the non-bases $\{a, b, x_1, x_2\}$ and $\{a', b, y_1, y_2\}$ we obtain $x_1 + x_2 + a \equiv y_1 + y_2 + a' \mod n$. Combining these two equations we find that $a \equiv a' \mod n$, a contradiction, since a and a' must take distinct values in $[n]$. $\qquad \square$

3.4.1 Fixed rank sparse paving matroids

We have shown the existence of a single sparse paving matroid forbidden as a minor in all GS matroids. However we can extend this result by showing that asymptotically almost all sparse paving matroids of fixed rank r (for $r \geq 4$) contain a minor that is GS-forbidden - and hence are themselves GS forbidden.

Definition 3.4.2. For $r \geq 4$, let $\mathcal{N}_r$ be the collections of rank-r matroids

defined by the following description: M is a member of $\mathcal{N}_r$ if and only if its groundset can be written as

$$E(M) := \{a, a', b, c, d\} \sqcup (X \cup Y)$$

(where $|X| = |Y| = r - 2$ and $|X \cap Y| \leq r - 4$])

and its non-bases as

$$\mathcal{L}(M) = \{(X \cup \{a, b\}), (X \cup \{c, d\}), (Y \cup \{a', b\}), (Y \cup \{c, d\})\}$$

We note that the N defined in the proof of Theorem 3.4.1 is the sole member of $\mathcal{N}_4$.

By an argument similar to the proof of Theorem 3.4.1, a matroid in $\mathcal{N}_r$ cannot be a minor of any GS matroid. We will have shown that asymptotically almost all sparse paving matroids of rank r are GS-forbidden if we can show that asymptotically almost all of them contain an $\mathcal{N}_r$-minor. We prove that result now.

Theorem 3.4.3. *Let M be drawn uniformly at random from the collection $\mathbb{S}_{n,r}$ of rank-r sparse paving matroids on $[n]$. Asymptotically almost certainly M has a minor in $\mathcal{N}_r$.*

To prove Theorem 3.4.3 we shall need to consider a new class of structure, which we shall call *concertinas*. It will be easiest just to view these as a line-structure, although we should remember that every line-structure corresponds to the non-bases of a sparse paving matroid.

Definition 3.4.4 ((r,k)-concertina)**.** Imagine we have a sparse paving matroid M. Let $\mathcal{L}(M)$ be the collection of non-bases of M and a, b, c, d be elements not in the groundset of M. Define a line-structure

$$\mathcal{L} = \{\{a,b\} \cup L : L \in \mathcal{L}(M)\} \cup \{\{c,d\} \cup L : L \in \mathcal{L}(M)\}$$

We say a line-structure $\mathcal{L}$ which can be constructed in this way is a *concertina*. In particular we may call $\mathcal{L}$ an (r,k)-concertina where $r = r(\mathcal{L}) = r(M) + 2$, and $k = |\mathcal{L}(M)|$ (note that $|\mathcal{L}| = 2k$).

We shall call a, b, c, d the *roots* of the concertina. The *root pairs* are $\{a,b\}$ and $\{c,d\}$. The *hinge sets* of $\mathcal{L}$ are the non-bases of the original M.

An $\mathcal{N}_4$-matroid and a (4,3)-concertina (with disjoint hinge sets) can be viewed in Figure 3.5.

Note the relation between the structures: for example, if we fuse a and a' in our drawing of the $\mathcal{N}_4$-matroid (meaning precisely that we identify those two vertices in the line-structure) then we get a (4,2)-concertina.

As we work towards a proof of Theorem 3.4.3, we shall say that a line-structure $\mathcal{L}$ contains an $\mathcal{N}_r$-*substructure* if it has a substructure isomorphic to $\mathcal{L}(N)$, for some matroid $N \in \mathcal{N}_r$.

The following theorem, presented earlier as Theorem 2.2.9, enables us to assume that our sparse paving matroid M has a fairly large number of non-bases.

Theorem 3.4.5. *Let r be a fixed integer. Asymptotically almost all sparse paving matroids $M \in \mathbb{S}_{n,r}$ have at least $(1 - o(1)) \frac{1}{2rn} \binom{n}{r(M)}$ non-bases.*

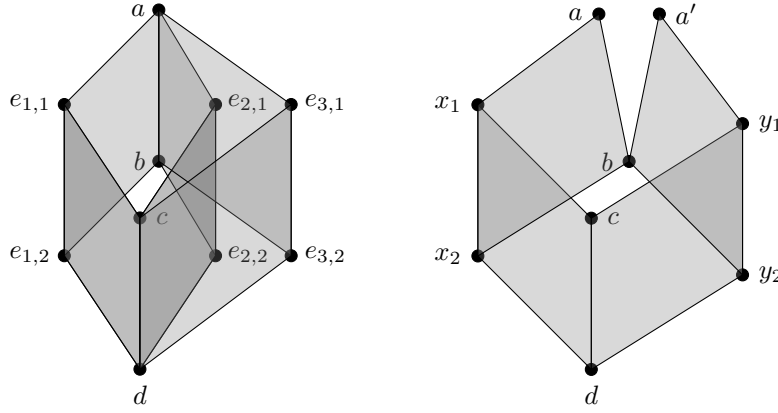Continuing to view the collections of non-bases of sparse paving matroids

Figure 3.5: A $(4,3)$-concertina (left) and an $\mathcal{N}_4$-matroid. The "lines" are depicted as faces here, and each contain 4 elements. Note that although the hinge sets shown here are disjoint, this is not necessarily the case in higher ranks, where they are only required to obey a sparse paving structure.

as line-structures, we shall show that when $M \in \mathbb{S}_{n,r}$ has sufficiently many non-bases, then $\mathcal{L}(M)$ contains either a large number of $\mathcal{N}_r$-substructures or an arbitrarily large number of arbitrarily large concertina substructures. We can then by a simple two-way counting argument show that occurrences of the former case are heavily outnumbered by the latter.

Of course substructures isomorphic to $\mathcal{L}(N)$ do not necessarily represent instances of N-minors, since the restriction to the elements of the substructure might contain other lines (non-bases). However we are able to make use of the following theorem (seen earlier as Theorem 2.3.10).

Theorem 3.4.6. *Let N be any sparse paving matroid. Suppose that for any integer m, for asymptotically almost every sparse paving matroid M of rank $r(N)$, we have $\mathcal{L}(M)$ contains at least m pairwise element-disjoint substructures isomorphic to $\mathcal{L}(N)$. Then N is a minor of asymptotically almost every sparse paving matroid.*

We have now covered all the relevant pre-existing results. The next lemma

represents the first stage in the proof proper of Theorem 3.4.3. We say that concertinas $\mathcal{L}_1$ and $\mathcal{L}_2$ are *root-disjoint* if they share no roots (they may still intersect in other elements).

Lemma 3.4.7. *For rank $r \geq 4$, for any positive integer m and constant $\alpha_1 > 0$, there exists $N \in \mathbb{N}$ and $\alpha_2 > 0$ such that the following holds:*

For all $n > N$, and all $M \in \mathbb{S}_{n,r}$ with at least $\alpha_1 \left(\frac{1}{n} \binom{n}{r} \right)$ non-bases, $\mathcal{L}(M)$ contains either at least m pairwise element-disjoint $\mathcal{N}_r$-substructures, or at least $\alpha_2 n$ root-disjoint $(r, \alpha_2 n^{r-3})$-concertinas.

Proof. We show that if $\mathcal{L}(M)$ does not contain m pairwise element-disjoint $\mathcal{N}_r$-substructures, then it must contain $\alpha_2 n$ or more root-disjoint $(r, \alpha_2 n^{r-3})$-concertinas (for some constant $\alpha_2(r, m, \alpha_1)$).

We can assume without loss of generality that $m = 1$ since, supposing some greater value of m offers a counterexample, then we can construct a counterexample with $m = 1$. This is because we could remove all $\mathcal{N}_r$-substructures, by deleting only $o(n)$ elements and $o(\frac{1}{n}\binom{n}{r})$ non-bases. Specifically, we need only delete every element in the (up to) $m-1$ pairwise element-disjoint $\mathcal{N}_r$-substructures, since if a copy was left after that, it too must have been pairwise element-disjoint with the others, contradicting the premise that $\mathcal{L}(M)$ contains fewer than m of the $\mathcal{N}_r$-substructures. Every time we delete the elements of an $\mathcal{N}_r$-substructure we remove at most $2r + 1$ elements, and each element can only be contained in at most $\frac{2}{n-1}\binom{n-1}{r-1}$ lines, so we remove at most $(m - 1)(2r + 1)n^{r-2} = o(\frac{1}{n}\binom{n}{r})$ lines.

Now, imagine we construct a bipartite graph G such that $V(G) = V_1 \sqcup V_2$

with V_1 representing pairs of elements of $E(M)$, and V_2 representing $(r-2)$-sets of $E(M)$. Two vertices in opposite parts are linked if and only if the combined elements they represent form a non-basis of M.

Imagine choosing $v := \{e_1, \ldots, e_{r-2}\}$ uniformly at random from the members of V_2. Let d_v denote the degree of v in G, and p_v denote the number of distinct 2-paths in G whose midpoint is v. Now, using our assumption that there are at least $\alpha_1 \left(\frac{1}{n} \binom{n}{r} \right)$ non-bases, we can see that

$$\mathrm{E}(d_v) \geq \frac{\alpha_1}{n} \binom{n-r}{2} = (1+o(1)) \frac{\alpha_1 n}{2}$$

Calculating $\mathrm{E}(p_v)$ is a little trickier - the total number of 2-paths, counted by their central vertex, is $\frac{1}{2} \sum_{x \in V_2} d_x (d_x - 1)$, so

$$\mathrm{E}(p_v) = \frac{1}{2} \mathrm{E}(d_v(d_v - 1)) = \frac{1}{2} \left(\mathrm{E}(d_v^2) - \mathrm{E}(d_v) \right)$$

$$\geq \frac{1}{2} \left((\mathrm{E}(d_v))^2 - \mathrm{E}(d_v) \right) = (1+o(1)) \frac{\alpha_1^2 n^2}{8}$$

where the first inequality is due to Jensen's inequality [9].

The total number of 2-paths with a midpoint in V_2 is at least

$$|V_2| \mathrm{E}(p_v) = (1+o(1)) \left(\binom{n}{r-2} \frac{\alpha_1^2 n^2}{8} \right) = (1+o(1)) \left(\frac{\alpha_1^2 n^r}{8(r-2)!} \right)$$

Now, imagine y and z are vertices of V_1 representing pairs of elements which intersect in one element, and suppose some other vertex x is the origin of 2-paths finishing at y and z respectively. Call these paths xuy and xvz and observe then that $x \cup u$, $x \cup v$, $u \cup y$ and $v \cup z$ are non-bases of M which together form an $\mathcal{N}_r$-substructure of $\mathcal{L}(M)$ (note u cannot equal v, else the

final two non-bases described intersect in $r-1$ elements, contradicting sparse pavingness; similarly u and v cannot intersect in more than $r-4$ elements, else the first two non-bases described would contradict sparse pavingness).

We can have at most $n/2$ pairs of elements before some pair intersect, so each vertex can have multiple 2-paths leading to at most $n/2$ other vertices. Also, if we have paths xuy and xvy, the vertices u and v cannot intersect in more than $r-4$ elements, since again this would contradict sparse pavingness. So each vertex can have at most $\frac{2}{n}\binom{n}{r-2}$ 2-paths leading to each other vertex, meaning each vertex is contained in at most $\binom{n}{r-2}$ 2-paths total. Now, the the total number of 2-paths is at least $(1+o(1))\frac{\alpha_1^2 n^r}{8}$ which for large enough N is greater than $\frac{\alpha_1^2 n^r}{9}$.

Now, let $c = \frac{\alpha_1^2}{9}$ be a constant. We already know that for large enough n the number of 2-paths with midpoints in V_2 is at least cn^r. So some vertex of V_1 must be contained in at least $2cn^{r-2}$ of them, and that vertex must have at least $4cn^{r-3}$ 2-paths with the same endpoint - which is to say, it must be a root of a $(r, 4cn^{r-3})$-concertina in $\mathcal{L}(M)$.

We require concertinas to be root-disjoint, but since removing two root pairs will remove at most $\binom{n}{r-2}$ 2-paths from consideration, for any constant c_2 we can remove $c_2 n$ pairs of root pairs whilst only removing $o(n^r)$ 2-paths from consideration.

So it suffices to set $\alpha_2 := 4c = \frac{4\alpha_1^2}{9(r-2)!}$. $\qquad\square$

Corollary 3.4.8. *For fixed $r \geq 4$ and $m \in \mathbb{N}$, let M be chosen uniformly at random from $\mathbb{S}_{n,r}$. Suppose that $\mathcal{L}(M)$ does not contain m pairwise element-disjoint $\mathcal{N}_r$-substructures. Then there exists $\alpha > 0$ such that asymptotically almost certainly $\mathcal{L}(M)$ contains at least αn pairwise root-disjoint $(r, \alpha n^{r-3})$-*

concertinas.

Proof. This follows straight away from Theorem 3.4.5 and Lemma 3.4.7. □

Theorem 3.4.3 is now within reach. The final lemma in our proof requires the introduction of more notation.

Definition 3.4.9. For $0 \leq k \leq m$, let $A(k, n, r, \alpha)$ denote the collection of those matroids $M \in S_{n,r}$ such that $\mathcal{L}(M)$ contains at least αn pairwise root-disjoint $(r, \alpha n^{r-3})$-concertinas. and precisely k pairwise element-disjoint $\mathcal{N}_r$-substructures (that is to say, we can find k pairwise element-disjoint $\mathcal{N}_r$-substructures, but not $k + 1$).

To complete the proof of Theorem 3.4.3 we now need only show the following:

Lemma 3.4.10. *Fix $\alpha > 0, r \geq 4$ and $k \in \mathbb{N}$. Let $A_k := A(k, n, r, \alpha)$*

Then, as $n \to \infty$, we have

$$\frac{|A_{k+1}|}{|A_k|} = \Omega(n^{0.5})$$

where we use Omega notation in the sense defined by Knuth, such that $f(x) = \Omega(g(x)) \iff g(x) = O(f(x))$.

In particular we note that this implies that for any fixed $m \in \mathbb{N}$ we have

$$\frac{\left| \bigcup_{0 \leq k < m} A_k \right|}{\left| A_m \right|} = o(1)$$

which in turn implies that asymptotically almost all rank-r sparse paving matroids contain m or more $\mathcal{N}_r$-substructures. Since there are finitely many

$\mathcal{N}_r$-matroids, and m can be made arbitrarily large, by a pigeonhole argument asymptotically almost all rank-r sparse paving matroids contain at least an arbitrarily large number m' of $\mathcal{L}(N)$-substructures, for some $N \in \mathcal{N}_r$. Theorem 3.4.3 then follows by a simple application of Theorem 3.4.6 and the fact that $\mathcal{N}_r$ matroids are GS-forbidden.

Proof of Lemma 3.4.10. Our strategy is to build a many-to-many correspondence between a non-vanishing subset of A_k and a vanishing subset of A_{k+1}. Counting arguments will then show the size of these two subsets to differ by at most a constant, thereby demonstrating that $|A_k|$ vanishes with respect to $|A_{k+1}|$. Recalling that our groundset is just $[n]$,

Let $B_k \subseteq A_k$ be the subset of matroids in A_k with the additional condition that one of the $(r, \alpha n^{r-3})$-concertinas must have a root labelled less than $n^{0.5}$.

Let $C_k \subseteq A_k$ be the subset of matroids in A_k with the additional condition that in *any* collection of k pairwise element-disjoint $\mathcal{N}_r$-substructures, one must contain an element labelled less than $n^{0.5}$.

It is readily seen that asymptotically almost all matroids in A_k are also in B_k. Imagine drawing M uniformly at random from A_k. Then M has at least $4\alpha n$ distinct roots of $(r, \alpha n^{r-3})$-concertinas, call these $e_1, \ldots, e_{\lfloor 4\alpha n \rfloor}$. For any given $1 \le i \le 4\alpha n$ we have $\Pr(e_i < n^{0.5}) = n^{0.5}/n = n^{-0.5}$. Now

$$\Pr(M \notin B_k) \le \Pr(e_1 > n^{0.5} \wedge \cdots \wedge e_{\lfloor 4\alpha n \rfloor} > n^{0.5}) \le (1 - n^{-0.5})^{\lfloor 4\alpha n \rfloor}$$

$$\le (1/e)^{\lfloor 4\alpha \sqrt{n} \rfloor} \le e^{1 - 4\alpha \sqrt{n}}$$

where we have used the inequality $(1 - x^{-1})^x \le e^{-1}$, for $x > 1$.

Conversely, asymptotically almost every member of A_k is not contained in C_k: the number of elements forced to appear in any collection of element-disjoint $\mathcal{N}_r$-substructures can be no greater than the number of vertices contained in one such collection, which is equal to $(2r+1)k$, so the chance that a randomly chosen matroid in A_k will also be in C_k is at most $(2r+1)kn^{-0.5}$.

In particular, the above two observations imply that, defining $D_k = B_k \setminus C_k$, asymptotically almost every member of A_k is in D_k.

We shall now construct a correspondence $\Phi \subseteq D_k \times C_{k+1}$ and by a double-counting argument demonstrate that $|D_k|/|C_{k+1}| = O(1)$.

We say that $(M, M') \in \Phi$ if and only if $\mathcal{L}(M')$ can be constructed from $\mathcal{L}(M)$ as follows:

Firstly, take the first (under some order, e.g. lexicographical) set of k pairwise element-disjoint $\mathcal{N}_r$-substructures in $\mathcal{L}(M)$ and ignore (temporarily relax) all lines intersecting any of the elements in these substructures. This has a negligible effect on the density of remaining lines, so we may still assume we have the ability to find an $(r, \alpha n^{r-3})$-concertina in $\mathcal{L}(M)$ such than some root is labelled less than $n^{0.5}$; take the smallest such root and label it b (we are aiming to build a $\mathcal{N}_r$-structure with labels as in Figure 3.5). Let (c, d) be the lexicographically first pair that form a root pair opposite b in some $(r, \alpha n^{r-3})$-concertina, and let a be the other root of that concertina (well defined, since if there are distinct elements a_1, a_2 with (a_1, b) and (a_2, b) both forming root pairs in (r, k)-concertinas (any k) with the root pair (c, d), then we would have a $\mathcal{N}_r$-substructure including the element b, contradicting $M \in D_k$). We may now adjust M to produce $M' \in C_{k+1}$ in the following way: firstly pick any element that is not a root of our concertina, and label this element a'. Now pick some line of our concertina that contains b but not a'. Label the hinge set as Y. There is one very specific, obscure condition that we must

place on our choice of Y: if there is exactly one hinge set Z so that $Z \cup \{a'\}$ is a subset of some line of of $\mathcal{L}(M)$, then we must not choose $Y = Z$! (The necessity of this condition will be made apparent later.)

We relax the line $(Y \cup \{a, b\})$ and insert a new line $(Y \cup \{a', b\})$, whilst relaxing any existing line containing any $r - 1$ elements from $Y \cup \{a', b\}$ (in order to preserve sparse pavingness). Preserving sparse pavingness only requires us to relax at most most $r - 1$ more lines, each containing $r - 1$ elements of $Y \cup \{a', b\}$ (note that there are up to r such lines, but that includes $(Y \cup \{a, b\})$, which we already relaxed). Finally, we restore the lines that we temporarily ignored at the start of the process. This is the end of the process; the resulting matroid is our M'.

Note that M' is certainly in C_{k+1} - we can now choose any other hinge set from our concertina, label this $X = \{x_1, \ldots, x_{r-2}\}$. Then $\mathcal{L}(M')$ contains an $\mathcal{N}_r$-substructure when restricting to $X \cup Y \cup \{a, a', b, c, d\}$ (arranged in a configuration similar to that shown in Figure 3.5).

Furthermore, we cannot have added more than one new *disjoint* $\mathcal{N}$-structure, since we only added one line - so any collection of $k+1$ element-disjoint copies of N' must include one containing the non-basis $(Y \cup \{a', b\})$ which of course includes the element b with its label less than $\log n$. Thus M' is in C_{k+1} and we have shown that Φ is well defined.

Consider $M \in D_k$. To how many elements in C_{k+1} does M correspond under Φ?

This is a simple counting argument: note that we had $O(n^{r-3})$ choices of the first line containing b (i.e. our choice of the hinge set Y), and $O(n)$ choices of a', meaning that M corresponds to $\Omega(n^{r-2})$ members of C_{k+1} under Φ.

To how many members of D_k does a matroid $M' \in C_{k+1}$ correspond under Φ? We have at most a constant $4(k+1)$ choices for the single line to be relaxed, and at most a constant r choices of "b" within that line. Similarly we have only $r-1$ choices for a', and this determines Y uniquely.

Moreover, c, d, and a are now uniquely determined. To see this take the first (under same ordering as before) set of k pairwise element-disjoint copies of $\mathcal{N}_r$-matroids in M excluding the element b, and as before temporarily relax lines intersecting any of these elements. We can now surely determine (c, d) to be the lexicographically first pair such that (c, d) is a root pair opposite b in a $(r, cn-1)$-concertina, **and** $Y \cup \{C, d\}$ is a non-basis of M' (again, if there were some other option, we could derive that M contained a $\mathcal{N}_r$-substructure including the element b, contradicting $M \in D_k$). And (again assuming M' actually is in correspondence with a member of D_k), our choice of a will also be clearly determined.

Reversing the construction necessarily requires us to make the above-mentioned relaxation of $(Y \cup \{a', b\})$ and the inclusion in its place of $(Y \cup a, b)$. We may also include a new line containing $Y \cup \{a', b\} \setminus \{y\}$ for all $y \in Y$. There are up to n choices for each of the $r-2$ lines we can include here. Finally, we might be able to add some new line containing $Y \cup \{a', b\} \setminus \{b\}$. However we contend that if we can, the number of choices for this is limited to a constant. To prove this, suppose that we can add such a line. Then there is some other hinge set of elements X such that $\mathcal{L}(M')$ (and $\mathcal{L}(M)$) contain lines $(X \cup \{c, d\}), (X \cup \{a, b\}), (X \cup \{a', e\})$ for some element e - since if this was not the case, we must have violated the obscure condition given for our choice of the hinge set Y. We now claim that we can only include a line of form $(Y \cup \{a', f\})$ if either $f = e$ or if f is necessarily contained in a line of any collection of $k+1$ pairwise element-disjoint $\mathcal{N}_r$ -substructures IN $\mathcal{L}(M')$.

Suppose not, and we can add some other non-basis $(Y \cup \{a', f\})$. Then in $\mathcal{L}(M)$ we have non-bases $(Y \cup \{c, d\}), (X \cup \{c, d\}), (X \cup \{a', e\}), (Y \cup \{a', f\})$ which form a $\mathcal{N}_r$-substructure. Moreover, this is element-disjoint to the existing k $\mathcal{N}_r$-substructures. Thus, except for $(2r + 1)k + 1$ choices of f, we arrive at a contradiction of the assumption that $M \in A_k$. So M' corresponds to at most cn^{r-2} members of D_k, for some constant c.

So under Φ we have that each member of D_k corresponds to $\Omega(n^{r-2})$ members of C_{k+1} and each member of C_{k+1} corresponds to (at most) $O(n^{r-2})$ members of D_k, so we obtain that $|C_{k+1}|/|D_k|) = \Omega(1)$. But also $|D_k|/|A_k| = \Omega(1)$, and $|A_{k+1}|/|C_{k+1}| = O(n^{0.5})$, so

$$\frac{|A_{k+1}|}{|A_k|} = \frac{|A_{k+1}|}{|C_{k+1}|} \times \frac{|C_k + 1|}{|D_k|} \times \frac{|D_k|}{|A_k|} = O(n^{0.5})$$

as required. $\square$

This completes the proof of Lemma 3.4.10, and (via an immediate application of Lemma 3.4.6) the proof of Theorem 3.4.3. Combining with the arguments in the proof of Theorem 3.4.1, we finally obtain our desired result:

Corollary 3.4.11. *Let $r \geq 4$. Asymptotically almost all sparse paving matroids of rank r are not a minor of any GS matroid.*

3.4.2   Other cases

We have shown that for all $r \geq 4$, asymptotically almost all sparse paving matroids of rank r are forbidden as a minor of GS matroids. What can we say for smaller values of r? For $r = 1$ or $r = 2$ it is easy to observe that all

sparse paving matroids are GS. The case for $r = 3$ is less clear

We know that not all rank 3 sparse paving matroids are GS and in fact at least some are not a minor of any GS matroid. For example, consider the *non-Fano* matroid F_7^- shown in Figure 3.6. This cannot be a minor of any GS matroid (suppose it is a minor of a matroid in $\mathbb{K}_n$; we obtain the equality $2a \equiv 2b \equiv 2c \equiv 2d \mod n$, which has no solutions for distinct a, b, c, d). However, the non-Fano matroid contains no loose elements and as noted in Chapter 2 we believe this may present a serious impediment to results of the kind "asymptotically almost all sparse paving matroids of rank r contain an F_7^--minor".
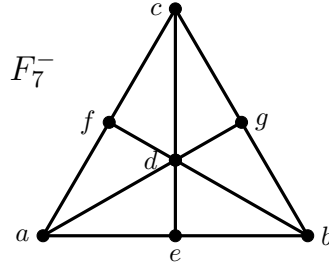

Figure 3.6: The non-Fano matroid F_7^-. Note that $F_7^- \setminus \{e\} \cong F_7^- \setminus \{f\} \cong F_7^- \setminus \{g\} \cong M(K_4)$; from these three minors respectively we can deduce that, if F_7^- were a minor of a GS matroid in $\mathbb{K}_n$, then $2a \equiv 2b \mod n$, and $2a \equiv 2c \mod n$, and $2a \equiv 2d \mod n$.

However it turns out we do have a rank-3 matroid which meet the stronger criteria (see Conjecture 2.3.29) that every minor has a loose element, and is GS-forbidden. This matroid is shown in Figure 3.7. Although its structure is rather more complicated that that of $\mathcal{N}_r$-matroids, we believe it is likely to yield to a similar counting argument, and hence conjecture with some confidence that asymptotically almost all rank-3 sparse paving matroids are GS-forbidden.

Conjecture 3.4.12. *Let M be drawn uniformly at random from* $\mathbb{S}_{n,3}$. *Asymptotically almost certainly M is GS-forbidden.*
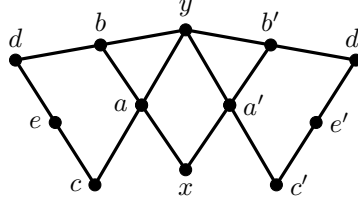

Figure 3.7: A rank-3 sparse paving matroid whose minors all contain a loose element, that is GS-forbidden. Suppose it is a minor of a GS matroid on $[n]$: by considering the non-bases (x, a, b) and (x, a', b') we can obtain that $a + b \cong a' + b' \mod n$. But then considering the 4 non-bases involving y, we get $(a + c) + (b + d) \cong (a' + c') + (b' + d') \mod n$. Cancelling, this gives $c + d \cong c' + d' \mod n$ which in turn implies that $e \cong e' \mod n$, an impossibility.

What about the general rank case?

Conjecture 3.4.13. *Asymptotically almost all sparse paving matroids are GS-forbidden*

We strongly suspect this to be true, as it would follow from Theorem 3.4.1 and a proof of the widely-believed conjecture of Mayhew, Newman, Welsh and Whittle [17] that every sparse paving matroid is a minor of asymptotically almost all sparse paving matroids. However this seems substantially harder to prove: for example, in our proof of Theorem 3.4.3, we are highly reliant on M being of the same rank as the target minors. Otherwise the correspondence defined by Φ could involve up to $r(M) - 1$ forced relaxations outside of our control.

Finally we note that results in the following chapter shall imply that asymptotically almost all transversal matroids of fixed rank, and asymptotically

almost all principal transversal matroids, are GS-forbidden. Since both these classes of matroids are yet to be defined, we shall save a formal statement of this theorem for later.

Chapter 4

Transversal Matroids

4.1 Introduction

Transversal matroids are interesting to us both for their non-matroidal "presentations" - which can take the appearance of set systems, bipartite graphs, or points in a simplex - and the opportunities these provide to obtain results through examining the non-matroidal objects. Let's first make clear what they are; we shall start with the most common definition, although it is the one we shall use the least.

Definition 4.1.1 (Partial transversal, transversal)**.** Let $\mathcal{A} = \{A_1, A_2, \dots, A_k\}$ be a set system. A set S of size r is a *partial transversal* of $\mathcal{A}$ if and only if there is some labelling $\{x_1, x_2, \dots, x_r\}$ of the elements of S such that for all $1 \leq i \leq r$, we have $x_i \in A_i$. (If $r = k$, we simply call X a *transversal* of $\mathcal{A}$).

Definition 4.1.2 (Transversal matroid)**.** A matroid M on groundset E is *transversal* if and only if there exists some set system $\mathcal{A}$ on E such that the

following holds: any set $S \subseteq E$ is independent in M if and only if S is a partial transversal of $\mathcal{A}$.

Equivalently, we can define transversal matroids through the language of partial matchings in bipartite graphs. Unfortunately, whilst each bipartite graph defines a single transversal matroid, the reverse is not true, so we cannot simulate a random transversal matroid by generating a random bipartite graph. However, there are two special cases in which we may effectively do this. The first is transversal matroids of fixed rank r. These each can be represented by (one or more) bipartite graphs with r elements on one side. We shall show that asymptotically almost all such bipartite graphs are the sole representative of a transversal matroid. this enables us to extend asymptotic results from bipartite graphs to rank-r transversal matroids.

The second special case is *principal transversal matroids*, also sometimes called *fundamental transversal matroids* or *simplicial matroids*. These are actually a fairly interesting class in their own right; unlike the class transversal matroids, the class of principal transversal matroids is self-dual and they hold an important place in the universe of *gammoids* (see Definition 4.1.3) : all gammoids are a minor of a principal transversal matroid. Strict gammoids are precisely those matroids which are contractions of principal transversal matroids whilst transversal matroids are precisely those matroids which are restrictions of principal transversal matroids, meaning principal transversal matroids sit within the intersection of those two classes.

Principal transversal matroids correspond precisely to those matroids with a presentation in a particular class of bipartite graphs; whilst a principal transversal matroid may have other presentations we shall show that almost all such *principal presentations* are the sole principal presentation of a transversal matroid. Furthermore a correspondence exists between prin-

cipal presentations of rank-r matroids on n elements, and bipartite graphs with parts of cardinality r and $n - r$. These two facts combined enable us to extend asymptotic results on bipartite graphs to principal transversal matroids.

Later in the chapter we drift away from randomness and consider a variety of other interesting questions related to transversal matroids. Transversal matroids are not minor-closed, but they are closed under taking *series minors* (matroids obtained by a sequence of deletions and *series contractions*, which we shall define in more detail). This enables us to consider excluded series-minors for the class of transversal matroids. There are known to be infinitely many excluded series-minors of transversal matroids, however we shall show that there are finitely many of any given rank. We shall also catalogue the excluded series-minors

Finally we consider axiomatisability of classes. Previously Mayhew, Newman and Whittle [20] have defined a subset of formulas MSOL called *M-logic* and showed that representable matroids are not axiomatisable in M-logic; this involved constructing representable and non-representable matroids that can be shown to be indistinguishable in M-logic. By a similar method we are able to show transversal matroids are not axiomatisable in M-logic; we also make the observation that the matroids constructed in the original paper are also (respectively) gammoids and non-gammoids, thereby showing gammoids are not axiomatisable in M-logic.

4.1.1 Preliminaries

Definition 4.1.3. A *gammoid* is a matroid M which arises in the following way: let there be a graph G with vertex set $V(G)$, and let S and T be subsets

(not necessarily disjoint) of $V(G)$. Set $E(M) := T$ and define a set $T' \subseteq T$ to be independent in M if there is a collection of $|T'|$ vertex-disjoint paths of G originating at vertices in S and ending at vertices in T'.

If there exists such a construction of M where $T = V(G)$, we say M is a *strict gammoid.*

Note that every gammoid is a restriction of a strict gammoid.

We make extensive use of some simple probabilistic results.

In particular, the well-known linearity of expectation:

Theorem 4.1.4 (Linearity of Expectation). *Given random variables $X_1, \dots, X_n$, we have*

$$\mathrm{E}(X_1 + \cdots + X_n) = \mathrm{E}(X_1) + \cdots + \mathrm{E}(X_n)$$

Also, the following simple result.

Lemma 4.1.5. *Let X be a random variable taking non-negative integer values. Then*

$$\Pr(X > 0) \leq \mathrm{E}(X)$$

Proof. Say $\Pr(X > 0) = p$. Then $\Pr(X \geq 1) = p$, so $\mathrm{E}(X) \geq p \cdot 1 = p$. $\qquad\square$

Typically such an X naturally appears as the sum of random variables $X_1, \dots, X_n$ each taking value 0 or 1. In this case, the lemma tells us that the probability of at least one event happening is less than the expected number

of events which happen.

Combining the above theorem and lemma gives the following.

Lemma 4.1.6. *Let $X_1, \ldots, X_n$ be random variables taking values 0 or 1. Let $X = \sum_{i=1}^{n} X_i$. Then*

$$\Pr(X > 0) \leq \sum_{i=1}^{n} \Pr(X_i = 1)$$

Proof. We first apply Lemma 4.1.5 to say $\Pr(X > 0) \leq \mathrm{E}(X)$. Applying linearity of expectation to the right hand side we have $\mathrm{E}(X) = \sum_{i=1}^{n} \mathrm{E}(X_i)$. Then since each X_i takes only values 0 and 1, we know that for each $1 \leq i \leq n$, we have $\mathrm{E}(X_i) = \Pr(X_i = 1)$. Substituting this into the original equation gives the desired result. $\square$

The above results will be used on several occasions and generally without quotation.

4.1.2   Transversal matroids and bipartite graphs

We show that our original definition of transversal matroids, in terms of partial transversals of set systems, can in fact be expressed in terms of *partial matchings* in bipartite graphs.

We say a bipartite graph G has parts X, Y to mean that $V(G) = X \sqcup Y$ with every edge of G containing a vertex from X and a vertex from Y.

Definition 4.1.7 (Partial matching, matching)**.** Imagine we have a bipartite graph G with parts X, Y. A set $A \subseteq X$ forms a *partial matching* with Y if there exists $B \subseteq Y$, with $|A| = |B|$ and there is a bijection $f : A \to B$

such that for any $v \in A$, there is an edge in G between v and $f(v)$. In other words, we can draw $|A|$ edges from A to B such that no two edges share a vertex. In the case where $B = Y$, we simply call it a *matching.*

We work from our original definition: A matroid M on groundset E is *transversal* if and only if there exists some set system $\mathcal{A}$ on E such that the following holds: any set $S \subseteq E$ is independent in M if and only if S is a partial transversal of $\mathcal{A}$.

Imagine a bipartite graph G with parts $X := E$ and $Y := \mathcal{A}$. That is to say, the vertices of X represent elements of E, and the vertices of Y represent sets in $\mathcal{A}$. For any $x \in X, y \in Y$, the edge xy exists if and only if $x \in y$. It therefore follows directly from the definition that a set $S \subseteq E$ is independent in M if and only if S is in a partial matching with a subset of Y. So, we can give an equivalent definition:

Definition 4.1.8 (Transversal matroid). A matroid M on groundset E is *transversal* if there exists such a bipartite graph $G = E \sqcup Y$ such that any set $S \subseteq E$ is independent in M if and only if S forms a partial matching with some subset of Y.

Looking at the other side:

Definition 4.1.9 (Presentation). Let M be a transversal matroid on groundset E. A bipartite graph G, one of whose parts is E, is said to be a *presentation* of M if and only if M arises from G in the manner described above. Such a presentation is *maximal* if we cannot create any bipartite graph $G' \supset G$, on the same elements and parts, such that G' is also a presentation of M.
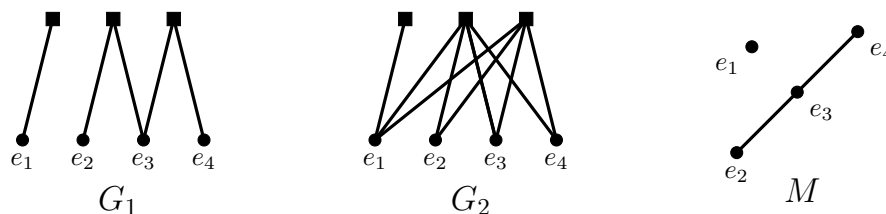
Figure 4.1: Two bipartite graphs, G_1 and G_2, which are each presentations of the same matroid M

Bonin's excellent survey [2] covers several basic facts and results about presentations. Foremost amongst them is the fundamental result that any set system (respectively, any bipartite graph) defines a transversal matroid.

Theorem 4.1.10 (Edmonds, Fulkerson [4])**.** *Let G be a bipartite graph on parts X, Y. Then G is a presentation of a transversal matroid on groundset X.*

Transversal matroids do not necessarily have only a single presentation, even up to isomorphism. For example, consider a rank-3 matroid M on 4 elements with a single non-basis. Figure 4.1 shows this matroid along with two very different bipartite presentations of the matroid. One of these graphs has only 5 edges, whilst the other has 9. In fact there are 10 distinct presentations up to isomorphism! Since such a small matroid has so many presentations, we might be forgiven for thinking that larger matroids have huge numbers of presentations. However for matroids of fixed rank, this is certainly not the case - matroids with more elements tend to have more dependencies, which greatly limit our ability to change presentations without changing matroids. In fact we shall show later that asymptotically almost all rank-r transversal matroids have only one bipartite presentation.

4.2 Maximal presentations and the fixed-rank case

Maximal presentations are of great interest due to the following result.

Theorem 4.2.1 (Mason [14])**.** *Let M be a rank-r matroid on groundset E. Consider bipartite graphs on vertex set comprising parts $X := E$ and $Y = \{y_1, \ldots, y_r\}$. Amongst these graphs there exists precisely one maximal presentation of M (up to labelling of Y).*

An accessible proof is given in Theorem 3.8 of Bonin's survey [2].

Unfortunately in general most presentations aren't maximal, nor is there any obvious way to randomly generate maximal presentations. However we shall find that for transversal matroids of fixed rank, the situation is much happier.

From this point onwards we shall, for a bipartite graph G, use $M_T(G)$ to denote the transversal matroid of which G is a presentation. We shall also use $\mathcal{E}(G)$ to describe the collection of edges of the graph G (not to be confused with somewhat similar notation used to denote the groundset of a matroid).

Lemma 4.2.2. *Suppose we have bipartite graphs G, G', G'' each on the vertex set $V = X \sqcup Y$ such that $\mathcal{E}(G) \subseteq \mathcal{E}(G') \subseteq \mathcal{E}(G'')$, and $M_T(G'') = M_T(G)$. Then $M_T(G') = M_T(G)$.*

Proof. We give a proof by contradiction. Since $\mathcal{E}(G) \subseteq \mathcal{E}(G')$, every matching in G is also a matching in G', and so every matching set in G is also a matching set in G'. This means that every independent set in $M_T(G)$

is an independent set in $M_T(G')$. And by a similar argument every independent set in $M_T(G')$ is an independent set in $M_T(G'')$. But now suppose $M_T(G') \neq M_T(G)$. This implies there must be some independent set in $M_T(G')$ (resp. matching set in G') that is not independent in $M_T(G)$ (resp. not matching in G). Then this set will also be independent in $m_T(G''$, so $M_T(G'') \neq M_T(G)$. $\qquad \square$

4.2.1 Random transversal matroids

Consider the random bipartite graph model $\mathcal{G}(n, m, p)$ in which we have parts X of cardinality n and Y of cardinality m, and for $x \in X, y \in Y$, the edge $\{x, y\}$ is included with probability p. We note that the case $p = \frac{1}{2}$ gives a uniform distribution over $\mathcal{G}(n, m)$, the class of (n, m)-part bipartite graphs.

We allow each graph drawn from this model to give rise to a transversal matroid, taking X to be the groundset of the matroid and the partial matchings in the graph to be the independent sets of the matroid. Let $M_T(G)$ be the transversal matroid derived from the bipartite graph G in this way.

We are interested in the case of transversal matroids of fixed rank r. Our aim is to show that asymptotically almost all presentations of rank-r presentations are maximal, enabling us to effectively treat a random fixed-rank transversal matroid as arising from a random bipartite graph.

First, a quick theorem.

Theorem 4.2.3. *Let G be a bipartite graph drawn from $\mathcal{G}(n, m)$, where m is fixed. As n grows, asymptotically almost certainly G will contain a matching of size m, meaning that $r(M_T(G)) = m$.*

Proof. Let X and Y be the parts of cardinality n and m respectively. Crudely, split X into parts $X_1, \cdots X_m$ each of size $\lfloor \frac{n}{m} \rfloor$ or $\lfloor \frac{n}{m} \rfloor + 1$. Let $Y = \{y_1, \cdots y_m\}$. We show that almost certainly there exists $x_1 \in X_i, \ldots, x_m \in X_m$ such that $\forall i \in \{1, \cdots m\}, x_i y_i$ is an edge. To do this we shall show that the probability of the negation is vanishing:

$$\Pr(\exists i \in \{1, \ldots, m\} : \forall x \in X_i, x \not\sim y_i)$$

$$\leq \mathrm{E}(|\{i : \forall x \in X_i, x \not\sim y_i\}|)$$

$$\leq m.\Pr(\forall x \in X_1, x \not\sim y_1)$$

$$\leq m.2^{-\lfloor \frac{n}{m} \rfloor}$$

(where we have assumed without loss of generality that $|X_1| = \lfloor \frac{n}{m} \rfloor$).

Clearly for all fixed m we have $m.2^{-\lfloor \frac{n}{m} \rfloor} \to 0$ as $n \to \infty$, and so we are done. $\qquad\square$

We also quote the following theorem; again Bonin's survey [2] has a proof.

Theorem 4.2.4. *Any transversal matroid of size n and rank r has a presentation in r sets; equivalently, in a bipartite graph with parts of size n, r respectively.*

Theorem 4.2.5. *Asymptotically almost all bipartite graphs in $\mathcal{G}(n, r)$ are a maximal presentation of a transversal matroid.*

Proof. We show that if drawing a matroid from the model $\mathcal{G}(n, r, \frac{1}{2})$, the probability that it is not maximal vanishes.

Let G be drawn from $\mathcal{G}(n, r, \frac{1}{2})$, and let X, Y be the parts of cardinality n, r respectively. We have

$$\Pr(G \text{ not maximal}) = \Pr(\exists e \in \mathcal{E}(G^c) : M_T(G \cup e) = M_T(G))$$

$$\leq \mathrm{E}(|\{y \in Y : \exists x \in X \setminus \Gamma(y), M_T(G \cup xy) = M_T(G)\}|)$$

$$\leq \sum_{y \in Y} \Pr(\exists x \in X \setminus \Gamma(y) : M_T(G \cup xy) = M_T(G))$$

Fix some $y \in Y$ define

$$p_y := \Pr(\exists x \in X \setminus \Gamma(y) : M_T(G \cup xy) = M_T(G))$$

Since the number of choices of y is a constant factor r, we are done if we can show that, for any choice of y, we have $p_y \to 0$ as $n \to \infty$.

Let $Z = X \setminus \Gamma(y)$. Now suppose we have disjoint sets $A, B \subset Z$ such that A and B each have a matching with $Y - y$. Then for any edge xy, $x \in Z$, we have $M_T(G \cup xy) \neq M_T(G)$ (since A and B are disjoint, at least one of them does not contain x; say without loss of generality that A does not). Then $A \cup x$ is independent in $M_T(G \cup xy)$ but not in $M_T(G)$. So if we can show that such sets A, B almost certainly exist, then $p_y \to 0$ and we are done.

Now asymptotically almost always we have $|\Gamma(y)| < 3n/4$, so we can ignore the case where $|Z| < n/4$. Now assuming $|Z| \geq n/4$, we can split Z arbitrarily into parts Z_A, Z_B with $|Z_A|, |Z_B| \geq \lfloor n/8 \rfloor$. We want to show that asymptotically almost certainly there exists $A \subseteq Z_A$, $B \subseteq Z_B$ such that A and B are in matching with $Y - y$. But since $n/8 \to \infty$ as $n \to \infty$, and

$|Y - y| = r - 1$ is fixed, Theorem 4.2.3 applies and we are done.   □

Theorems 4.2.4 and 4.2.5 give rise to the following corollary.

Corollary 4.2.6. *Asymptotically almost all transversal matroids of rank r have a single presentation in $\mathcal{G}(n, r)$, up to the labelling of the vertices in the bipartite graph.*

Proof. The number of n-element transversal matroids of rank r is asymptotically equal to the number of maximal presentations in $\mathcal{G}(n, r)$, whereas the number of transversal matroids with multiple presentations is bounded above by the number of non-maximal presentations in $\mathcal{G}(n, r)$.   □

We also note that asymptotically almost all transversal matroids have exactly $r!$ presentations when labelling is taken into account - these are the $r!$ possible labellings of the vertices in the non-groundset part of the bipartite graph (it is easily seen that the probability of automorphism under permutation of the labelling vanishes!).

Having proven Theorem 4.2.5, it is now easy to use probabilistic methods to prove various results for fixed rank transversal matroids.

Theorem 4.2.7. *Let N be a transversal matroid of rank r_0. For any fixed rank $r \geq r_0$, asymptotically almost all transversal matroid of rank r contain N as a minor.*

This is not a surprising result, since we are now effectively working in the world of random graphs, where subgraph and minor inclusion results are straightforward. We split the proof, first showing a lemma then following that with the rest of the argument. By an *induced subgraph* of G we mean a

graph obtained from G by vertex deletions, with no additional edge deletions.

Lemma 4.2.8. *Let H be a bipartite graph with r vertices in one part. Asymptotically almost every bipartite graph $G \in \mathcal{G}(n, r)$ contains an induced subgraph H' isomorphic to H, with the r-part of H' corresponding to the r-part of H.*

Proof. Let H contain r elements in one part and m in the other - call these parts H_B and H_A respectively. Now let G be a graph drawn uniformly at random from $\mathcal{G}(n, r)$ - call the n-part G_A and the r-part G_B. Now from G_A choose disjoint subsets $A_1, \ldots, A_{\lfloor n/m \rfloor}$ each of size m. Now take each A_i and consider the probability that $G(A_i \cup G_B)$ is isomorphic to H (with the vertices of G_B corresponding to the vertices of H_B). This probability is at least 2^{-rm}. But now the probability that G does not contain H as a subset (with G_B corresponding to H_B) is at most $(1 - 2^{-rm})^{n/m-1}$ which duly vanishes, since n is the only non-constant factor. $\square$

Proof of Theorem 4.2.7. We first show that there exists a matroid of rank r that contains N as a minor. This is simple: let G be a presentation of N. Now form G' by adding $r - r_0$ vertices to each part of G such that the new vertices are in a matching, but there are no edges between the new vertices and G. Now contracting $M_T(G')/\mathcal{E}(G' \backslash G) = N$, so $M_T(G') \succeq N$, and clearly $M_T(G')$ is of rank r.

Now take M to be our matroid of rank r containing N as a minor. Let H be a presentation of M. Asymptotically almost every bipartite graph in $\mathcal{G}(n, r)$ contains a subgraph H' isomorphic to H, with the r-part of H' corresponding to the r-part of H (by Lemma 4.2.8). But now contracting by the elements not contained as vertices in the n-part of H', we are left with

a matroid isomorphic to M'. So asymptotically almost all bipartite graphs in $\mathcal{G}(n, r)$ give rise to a transversal matroid containing M (and hence N) as a minor. Now by Theorem 4.2.5 the same is true for asymptotically almost every maximal presentation, and hence asymptotically almost all transversal matroids of rank r contain N as a minor. $\qquad\square$

4.2.2 Likelihood of matchings

For working with bipartite presentations of matroids, we shall want to consider the likelihood of the existence of a matching between two sets of vertices. We first familiarise ourselves with Hall's condition, which is necessary and sufficient for a bipartite graph to contain a matching.

Definition 4.2.9 (Hall's condition)**.** Let G be a bipartite graph on parts X, Y. A set $X' \subseteq X$ meets Hall's condition if, $\forall S \subseteq X', |\Gamma(S)| \geq |S|$.

Theorem 4.2.10 (Hall's marriage theorem [7])**.** *There is a matching from $X' \subseteq X$ to some $Y' \subseteq Y$ if and only if X' meets Hall's condition.*

In particular if $|X| = |Y|$ and Hall's condition holds for X itself, we have a *total matching* from X to Y (by which we mean a matching that spans all vertices in G). The following result is well-known and somewhat trivially proved.

Theorem 4.2.11. *Let X and Y be two sets of elements with $|X| = |Y| = n$. Form a bipartite graph G on $X \cup Y$ by independently including each edge $xy : x \in X, y \in Y$ with fixed probability $\frac{1}{2}$. Let p_n be the probability that G does not contain a total matching. Then $p_n \to 0$ as $n \to \infty$.*

However, we shall actually prove a more general result (this is useful later).

Theorem 4.2.12. *Fix $\delta > 0$ and let $\epsilon > 0$. let X and Y be two sets of elements with $|X| = n, |Y| = r$. Let G be drawn uniformly at random from all bipartite graphs on parts X, Y. Let $p_{n,r}$ be the probability that there exists some $X' \in X^{(r)}$ such that X' is not in matching with any subset of Y. There exists $n(\epsilon)$ such that, for $n \geq n(\epsilon)$ and $r \geq n/2 + \delta n$ we must have $p_{n,r} < \epsilon$.*

In other words, as $n \to \infty$, asymptotically almost all bipartite graphs with parts $|X|$ of cardinality n and $|Y|$ of cardinality at least $n/2 + \delta n$ have matchings from every $|Y|$-subset of X to Y. In the context of bipartite presentations, this actually says that for $r \geq n/2 + \delta n$, asymptotically almost all bipartite graphs $G \in \mathcal{G}(n, r)$ have $M_T(G) \cong U_{r,n}$.

Proof. We assume without loss of generality that $\delta < 0.1$

We can model drawing G uniformly at random by including the edges $\{xy : x \in X, y \in Y\}$ independently with probability $\frac{1}{2}$.

For such an X' to exist, by Hall's conditions, requires some $A \in X, B \in Y$ such that $1 \leq |A| \leq r$ and $|B| = r + 1 - |A|$ and, for all $x \in A, y \in B$ there is no edge xy.

Firstly we shall show that asymptotically almost certainly there are no such cases where $|A| \geq \delta n/2, |B| \geq \delta n/2$. This is easy enough to see; let Z be the number of such cases. Then

$$\mathrm{E}(Z) = \sum_{i=\frac{\delta n}{2}}^{r+1-\frac{\delta n}{2}} \binom{n}{i}\binom{r}{i-1}2^{-i(r+1-i)}$$

But for any $\delta n/2 \leq i \leq r + 1 - \delta n/2$ we have

$$\binom{n}{i}\binom{r}{i-1}2^{-i(r+1-i)} \leq n^n r^n 2^{-(\delta n/2)^2}$$

$$\implies \mathrm{E}(Z) \leq n(n^n r^n 2^{-(\delta n/2)^2} \to 0$$

For $|B| < \delta n/2$, we must consider a different measure of failure. If $|B| < \delta n/2$ then $|A| > n/2 + \delta n/2$. This means that we must have $|\Gamma(B)| < n/2 - \delta n/2$. We shall show that in fact there is asymptotically almost certainly no vertex $y \in Y$ with $|\Gamma(y)| < n/2 - \delta n/2$. Letting y_0 be any fixed vertex in Y,

$$\Pr(\exists y \in Y : |\Gamma(y)| < \frac{n}{2} - \frac{\delta n}{2}) \leq \mathrm{E}(|\{y \in Y : |\Gamma(y)| < \frac{n}{2} - \frac{\delta n}{2}\}|)$$

$$\leq r.\Pr(|\Gamma(y_0)| < \frac{n}{2} - \frac{\delta n}{2})$$

Now $|\Gamma(y_0)|$ may be seen as the sum of n independent trials taking value 1 or 0 - where each trial corresponds to the inclusion or non-inclusion of an edge from y_0 to a particular vertex in X. This makes it a binomial variable with parameters $n, 0.5)$. We now apply Hoeffding's inequality for binomial distributions

Theorem 4.2.13 (Hoeffding)**.** *Suppose $Z \sim \mathrm{Bin}(n, p)$. Then*

$$\Pr(Z < k) \leq \exp\left(-2\frac{(np-k)^2}{n}\right)$$

Applying Hoeffding with $p = \frac{1}{2}$ and $k = \frac{n}{2} - \frac{\delta n}{2}$ we find

$$Pr(\exists y \in Y : |\Gamma(y)| < k) \leq r.\exp\left(-2\frac{(\delta n/2)^2}{n}\right) \leq r.e^{-\delta^2 n/2} \to 0$$

Finally, the case where $|A| < \delta n/2$. In this case we find that $\Gamma(A) < \delta n/2$. Unsurprisingly this is implausible: we apply the same method as in the previous case to find

$$Pr(\exists x \in X : |\Gamma(x)| < \delta n/2) \leq n.\exp\left(-2\frac{(n/2 - \delta n/2)^2}{n}\right) = n.e^{-(1-\delta)^2 n/2} \to 0$$

$\square$

(The choice to use Hoeffding's inequality rather than the more obvious Law of large numbers was because the latter is not usually given with an explicit bound on the rate of convergence, and our cases required a probability of $o(r^{-1})$ and $o(n^{-1})$ respectively.)

When $n \geq 2r$, it is clear that no such result holds: there will almost certainly exist a vertex $y \in Y$ such that y has degree strictly less that $n/2$. But then $X \setminus \Gamma(y)$ does not meet Hall's condition.

4.3   Principal transversal matroids

As previously mentioned, we are interested in principal transversal matroids due to their important place in the universe of gammoids. Every gammoid is a minor of a principal transversal matroid; in fact we can make two stronger statements. Firstly, every gammoid is a restriction of a strict gammoid, and every strict gammoid is a contraction of a principal transversal matroids. Secondly, and dual to the first statement: every gammoid is a contraction

of a transversal matroid, and every transversal matroid is a restriction of a principal transversal matroid. A sketch of this worldview is show in Figure 4.2.
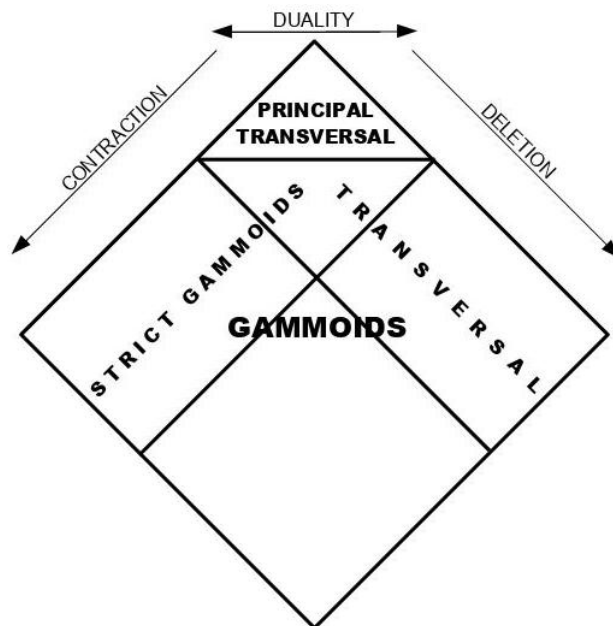

Figure 4.2: A view of the world of gammoids. Contraction and deletion may occur along the diagonal axes as indicated; duality is a reflection in the vertical axis. Size and rank are not shown. (Based on an original sketch by Ingleton.)

We begin this section by putting transversal matroids in a different context, from which principal transversal matroids might seem more naturally defined. Firstly we need some definitions:

Definition 4.3.1 (Cyclic flat)**.** A flat F of a matroid M is called *cyclic* if the matroid $M|F$ has no coloops. We use $\mathcal{Z}(M)$ to denote the collection of cyclic flats of M.

Note that we are talking about coloops of the matroid $M|F$; this is **not** equivalent to saying F contains no coloops of M.
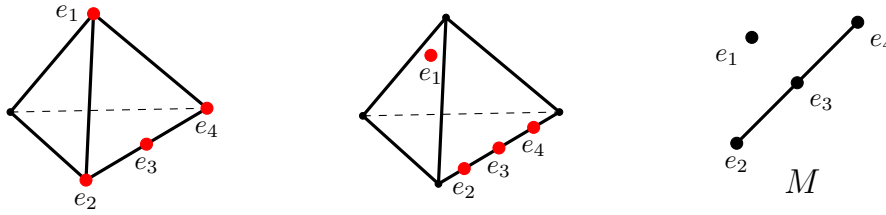
Figure 4.3: Two presentations of a matroid M on the simplex. This is actually the same matroid as seen in Figure 4.1, with the same two presentations in geometrical rather than graphical form. We see the left hand presentation is principal, but the right is not.

We can now offer yet another rephrasing of the definition of a transversal matroid.

Definition 4.3.2 (Transversal matroid)**.** A matroid is transversal if it can be represented geometrically as points on a simplex, such that each cyclic flat of the matroid corresponds to the points in a particular face (of any dimension) of the simplex.

Essentially, geometric and graphical representations are in direct correspondence - where an element is adjacent to a particular collection of vertices in a bipartite graphical presentation, in the corresponding geometrical representation it lies in the face containing precisely that collection of vertices.

Definition 4.3.3 (Principal transversal matroid, principal presentation)**.** A transversal matroid is *principal* if it has a geometrical representation on a simplex, such that at least one element lies at each vertex of the simplex. Such a representation is called a *principal presentation.*

Figure 4.3 shows two geometrical representations of a matroid on the simplex, one of which is principal.

Returning to the bipartite graphical context, this means a transversal ma-

troid is principal if it has a presentation such that each vertex in the non-groundset part of the graph is adjacent to at least one leaf.

Note that a principal transversal matroid may have non-principal presentations - in fact an example of this can be seen in Figure 4.1 where G_1 is a principal presentation of M, but G_2 (along with 8 other presentations of M) is not.

Principal transversal matroids have a number of nice properties.

Theorem 4.3.4 (Las Vergnas [12])**.** *The dual of a principal transversal matroid is necessarily principal transversal.*

We note the following obvious corollary. Let $\mathbb{P}_{n,r}$ denote the collection of rank-r principal transversal matroids on n elements.

Corollary 4.3.5. *For fixed n and all $0 \leq r \leq n$, $|\mathbb{P}_{n,r}| = |\mathbb{P}_{n,n-r}|$.*

A matroid can have more than one principal presentation. For example the matroid in Figures 4.1 and 4.3 clearly has multiple principal presentations, since the elements e_2, e_3, e_4 are interchangeable in the principal presentation. However we can show that any two distinct principal presentations cannot share the same set of leaves.

Theorem 4.3.6. *Every principal transversal matroid M has a unique principal presentation, up to labelling of the r-part of the graph and choice of a collection B of $r(M)$ leaves.*

Proof. We work in the graphical context. We only consider presentations in

which a given collection of r elements (without loss of generality say the first r elements) of our groundset are leaves with distinct neighbours (so they will form a matching with the r-part of our principal presentation graph). Let us call this collection B and label its members $l_1, \ldots, l_r$. Note that B is a basis of M.

Consider a principal presentation $G : V(G) = X \sqcup Y$, with X the groundset part of G, and let us label the vertices in Y as $y_1, \ldots, y_r$ such that for each $i \in \{1, \ldots, r\}$, $y_i \sim l_i$. Now let G' be a different principal presentation, with the same collection of leaves B, and let the vertex adjacent to l_i be labelled y_i'.

Given a vertex $x \in X$, let B_x be the collection of leaves in B adjacent to neighbours of x in G. Let B_x' be the collection of leaves adjacent to neighbours of x in G'.

Since G and G' are different, there is some vertex $x \in X \setminus B$ such that $\Gamma_G(x) \neq \Gamma_{G'}(x)$, and so $B_x \neq B_x'$.

Now consider the smallest circuit of $M_T(G)$ contained in $B \cup x$ - this is $B_x \cup x$. And the smallest circuit of $M_T(G')$ contained in $B \cup x$ is $B_x' \cup x$. So $\mathcal{C}(M_T(G)) \neq \mathcal{C}(M_T(G'))$ and thus $M_T(G) \neq M_T(G')$. $\qquad \square$

So we have just shown that different principal presentations of the same matroid cannot share a collection of r leaves in matching with the r-part. Our goal now is to show that, whilst principal transversal matroids may have multiple principal presentations with different leaf sets, asymptotically almost always they do not. Proving this requires a number of intermediate results.

Lemma 4.3.7. *Let G be drawn uniformly at random from principal presentations of rank-r matroids on $[n]$ elements with parts $X := [n]$ and Y. Suppose there exists $\epsilon > 0$ such that $r \geq (1 + \epsilon) \log(n - r)$ and $n - r \geq \epsilon n$. Then asymptotically almost certainly X only contains r leaves.*

Proof. This is a counting argument. Let $B(G)$ be the collection of leaves of G and $B_{min}(G) = \{l_1, \ldots, l_r\}$ be the first r leaves in X. Clearly every choice of G has a forced choice of $B_{min}(G)$, so by identifying G with $B_{min}(G)$ we can partition all principal presentations $G \in \mathcal{G}(n, r)$. Now, for any choice of $B \in X^{(r)}$ we have

$$\Pr(|B(G)| > r : B_{min}(G) = B) \leq \mathrm{E}((|B(G) \setminus B_{min}(G)|) : B_{min}(G) = B)$$

$$\leq (n - r) \cdot r \cdot 2^{-r}$$

since $r \cdot 2^{-r}$ is the probability that any given vertex in X is a leaf.

But $2^{-r} \leq ((n - r)r)^{-1-\epsilon}$, so the probability of more than r leaves is bounded by $(\epsilon n)^{-\epsilon}$ which vanishes as $n \to \infty$.

Since this is true for all choices of B, we have

$$\Pr(|B(G)| > r) = \sum_{B \in X^{(r)}} \Pr(|B(G)| > r : B_{min}(G) = B)\Pr(B_{min}(G) = B)$$

$$\leq (\epsilon n)^{-\epsilon} \sum_{B \in X^{(r)}} \Pr(B_{min}(G) = B) = (\epsilon n)^{-\epsilon} \to 0$$

$\square$

Theorem 4.3.8. *Fix any $\delta > 0$. Let G be chosen uniformly at random from all principal presentations of transversal matroids on $[n]$. Asymptotically almost certainly $(0.5 - \delta)n \leq r(M_T(G)) \leq (0.5 + \delta)n$. Moreover, G asymptotically almost certainly contains only $r(M_T(G))$ leaves.*

Proof. We shall assume without loss of generality that $\delta < 0.1$.

Let $N_r := N(n, r)$ be the number of principal presentations G such that $M_T(G)$ has rank r. Without loss of generality we may ignore permutations of the non-groundset part of the graph. Imagine first choosing r leaves, and then choosing any arrangement of edges on the remainder of the graph. Certainly every presentation is generated in this way, and by a similar argument as in Lemma 4.3.7 we have asymptotically almost all choices producing a unique presentation for suitable ranges of r - in particular when $(0.5 - \delta)n \leq r \leq (0.5 + \delta)n$. So we have

$$N_r \leq \binom{n}{r} 2^{(r(n-r))}$$

asymptotically approaching equality when $(0.5 - \delta)n \leq r \leq (0.5 + \delta)n$.

This function is a product of two functions, each symmetrically unimodal in r with peak at $n/2$, and so is itself symmetrically unimodal about that point. Letting I denote the interval $[(0.5 - \delta)n, (0.5 + \delta)n]$ we have

$$\sum_{r \notin I} N_r \leq (n - 2\delta) N_{\lfloor (0.5 - \delta)n \rfloor}$$

Now, for $\lfloor (0.5 - \delta)n \rfloor \leq r \leq \lfloor n/2 \rfloor$ we have

$$\frac{N_{r+1}}{N_r} = (1 + o(1))\frac{n-r}{r+1}(2^{n-2r})$$

So

$$\frac{N_{\lfloor(0.5-\delta)n\rfloor+1}}{N_{\lfloor(0.5-\delta)n\rfloor}} \geq \frac{\lceil(0.5+\delta)n\rceil}{\lfloor(0.5-\delta)n\rfloor+1}2^{2\delta n} \geq 2^{2\delta n}$$

$$\implies N_{\lfloor(0.5-\delta)n\rfloor+1} \geq 2^{2\delta n}N_{\lfloor(0.5-\delta)n\rfloor} \geq \frac{2^{2\delta n}}{n-2\delta}\sum_{r\notin I}N_r$$

.

Thus as $n \to \infty$, we have

$$\left(\frac{\sum_{r\notin I}N_r}{\sum_{r\in I}N_r}\right) \leq \left(\frac{\sum_{r\notin I}N_r}{N_{\lfloor(0.5-\delta)n\rfloor+1}}\right) \leq \frac{n-2\delta}{2^{2\delta n}} \to 0$$

in other words: asymptotically almost all principal presentations G on n elements give $r(M_T(G))$ in the range $[(0.5-\delta)n, (0.5+\delta)n]$. Since we can easily choose δ to satisfy the conditions of Lemma 4.3.7 we can also say that asymptotically almost certainly G only contains $r(M_T(G))$ leaves. $\square$

Theorem 4.3.9. *Let M be drawn uniformly at random from principal transversal matroids on $[n]$. Asymptotically almost certainly $(0.5 - \delta)n \leq r(M) \leq (0.5 + \delta)n$.*

This will require a preparatory lemma, which is a"Law of large numbers" type result showing that in a random bipartite graph, asymptotically almost certainly the size of the neighbourhoods of all small vertex-sets approach the expected value.

Lemma 4.3.10. *Fix $k \in \mathbb{N}$, $\delta > 0$ and let $\epsilon > 0$. Suppose we have a bipartite graph G chosen uniformly at random from $\mathcal{G}(n, r)$ with parts X of cardinality n and Y of cardinality r. There exists $n(\epsilon)$ such that when $n \geq n(\epsilon)$ and $n/2 \leq r \leq 2n$, we have*

$$\Pr(\exists X' \in X^{(k)} : \left||\Gamma(X')| - r(1 - 2^{-k})\right| \geq \delta r) < \epsilon$$

Proof. This is more or less identical to the arguments for the final cases in Theorem 4.2.12. The size of the neighbourhood of a given set is a binomial variable with parameters $r, 1 - 2^{-k}$ and the size of their complements is binomial with parameters $r, 2^{-k}$. So we apply Hoeffding to get

$$\Pr(\exists X' \in X^{(k)} : |\Gamma(X')| \leq r(1 - 2^{-k} - \delta)) \leq \binom{n}{k} exp(-2\delta^2 r) \to 0$$

and, symmetrically,

$$\Pr(\exists X' \in X^{(k)} : |\Gamma(X')| \geq r(1 - 2^{-k} - \delta)) \leq \binom{n}{k} exp(-2\delta^2 r) \to 0$$

Combining these asymptotic bounds gives the desired result. $\square$

Proof of Theorem 4.3.9. We shall assume without loss of generality that $\delta < 0.01$. Subject to that, we show the following claim:

Claim 4.3.10.1. Let $(0.5 - \delta)n \leq r \leq (0.5 + \delta)n$ and draw G uniformly at random from principal presentations of rank-r transversal matroids on $[n]$. Asymptotically almost certainly G is the only principal presentation of $M_T(G)$.

Given this claim, the number of principal transversal matroids with rank in

this range is asymptotically equal to the number of principal presentations with rank in this range. And the number of principal transversal matroids with rank outside this range is clearly bounded by the number of principal presentations with rank outside this range. And since by Theorem 4.3.8 asymptotically almost all principal presentations have rank in this range, so must asymptotically almost all principal transversal matroids.

We shall prove the claim firstly for the subset of cases where $r \leq n/2$. Let B be the leaf set of our presentation G (note by Lemma 4.3.7 and Theorem 4.3.8, we can assume there is only one such choice). Now by Theorem 4.3.6 we know this is the only principal presentation of $M_T(G)$ with this leaf set. What are the chances another leaf set exists?

Suppose as usual that G has parts X and Y. By Lemma 4.3.10 we can assume that every vertex in $x \in (X \setminus B)$ has neighbourhood of size between $0.4r$ and $0.6r$, and each pair $x_1, x_2 \in (X \setminus B)$ has $|\Gamma(x_1, x_2)| \geq 0.7r$. Similarly each $y \in Y$ has at least $0.4(n - r)$ neighbours, and each pair $y_1, y_2 \in Y$ has at least $0.7(n - r)$ neighbours, in $X \setminus B$. We may consciously neglect any presentations for which these conditions do not hold.

Now, we want to show G asymptotically almost certainly does not have an alternative principal presentation G' with leaf set B'. We first show that asymptotically almost certainly G does not have an alternative presentation G' with leaf set B' such that $|B' \cap B| \leq 0.35r$. Suppose it did, then take $x \in X \setminus (B \cup B')$. Let $B_x \subseteq B$ denote those leaves in G which share a neighbour with x. By assumption $0.4r \leq |B_x| \leq 0.6r$. But now $|B_x \setminus B'| \geq 0.05r > 2$, so we can find vertices $l_1, l_2 \in B_x \setminus B'$. And since $B_x \cup x$ is dependent in $M_T(G)$ we must also have $|\Gamma_{G'}(B_x \cup x)| \leq 0.6r$. But $\Gamma_{G'}(B_x \cup x) \supseteq \Gamma_{G'}(\{l_1, l_2\})$ and

by assumption we have $|\Gamma_{G'}(\{l_1, l_2\})| \geq 0.7r$, a contradiction.

So, we can assume $|B \cap B'| \geq 0.35r$. But now we also have $|X \setminus (B \cup B')| \geq 0.35r$. We claim that asymptotically almost certainly there exist $x \in X \setminus (B \cup B')$, $l \in B \setminus B'$ such that x and l share a neighbour in G. If $|B \setminus B'| \geq 2$ then we can find $y_1, y_2 \in Y$ each adjacent to a leaf in $B \setminus B'$. But now (again using our earlier assumptions about neighbourhood sizes) we have $|(X \setminus B \setminus \Gamma_G(\{y_1, y_2\})| \leq (n - r) - 0.7(n - r) = 0.3(n - r) \leq 0.35r$, so $\Gamma_G(\{y_1, y_2\}) \cap X \setminus (B \cup B') \neq \emptyset$ and we are done. The remaining case is when $B \setminus B'$ is a single element l. But now we already have $|X \setminus (B \cup B')| = n - r - 1$; letting y be the neighbour of l in G, y has at least $0.4(n - r)$ neighbours of which some must clearly be in $X \setminus (B \cup B')$.

So, we have some $x \in X \setminus (B \cup B'), l \in B \setminus B'$ sharing a neighbour in G. Now let $B_x \subset B$ be the leaves of G adjacent to neighbours of x. By assumption we have in G that $|\Gamma_G(B_x \cup x)| \leq 0.6r$. And clearly $B_x \cup x$ is a dependent set of $M_T(G) = M_T(G)$ and so in G' too we must have $|\Gamma_{G'}(B_x \cup x)| \leq 0.6r$. But $\Gamma_{G'}(B_x \cup x) \supseteq \Gamma_{G'}(\{l, x\})$ and since $l, x \in X \setminus B'$ we also have by assumption that $|\Gamma_{G'}(\{l, x\})| \geq 0.7r$, and a contradiction is reached.

We now note that having proved the result for $r \leq n/2$, it automatically holds for $r \geq n/2$. This is easy to see: by Corollary 4.3.5 the number of principal transversal matroids $|\mathbb{P}_{n,r}|$ is symmetric about $r = n/2$, and so is the number of principal presentations N_r. So

$$|\mathbb{P}_{n,r}| = (1 - o(1))N_r \implies |\mathbb{P}_{n,n-r}| = (1 - o(1))N_{n-r}$$

and thus we are done for $r > n/2$ also. $\square$

Combining Theorems 4.3.8 and 4.3.9 yields the following corollary.

Corollary 4.3.11. *Let M be drawn uniformly at random from principal transversal matroids on $[n]$. Asymptotically almost certainly M has only one principal presentation.*

Proof. Theorems 4.3.8 and 4.3.9 imply that asymptotically almost all principal presentations on n elements are unique. The number of transversal matroids with a unique principal presentation is by definition equal to the number of unique principal presentations, whereas the number of principle transversal matroids without a unique principal presentation is bounded above by (half) the number of non-unique principal presentations. Thus we can see that the proportion of principal transversal matroids without a unique principal presentation is bounded above by (half) the proportion of principal presentations that are not unique. Therefore it too is vanishing. $\qquad\square$

Theorem 4.3.12. *Let M be drawn uniformly at random from principal transversal matroids. Asymptotically almost certainly M contains a circuit of size $\lfloor r/2 \rfloor + 1$ or smaller.*

Proof. Corollary 4.3.11 tells us that it suffices to prove the result for a randomly chosen principal presentation. Let this presentation be G with parts X and Y and leaf set B as usual. For each vertex in $X \setminus L$, the degree of the vertex will be $\lfloor r/2 \rfloor$ or less with probability at least $\frac{1}{2}$. And so as $n \to \infty$ almost certainly there will be some $x \in X \setminus L$ with $d(x) \leq \lfloor r/2 \rfloor$. Now letting B_x be those leaves adjacent to neighbours of x, we see that $L_x \cup x$ is a circuit of size at most $\lfloor r/2 \rfloor + 1$. $\qquad\square$

We can now state a nice theorem for general transversal matroids.

Theorem 4.3.13. *Let N be a transversal matroid. Asymptotically almost all principal transversal matroids contain an N-minor.*

Proof. Without loss of generality we may assume a principal transversal matroid M has a unique principal presentation G with parts $X := E(M)$, Y and leaf set $B \subseteq X$ as in all previous examples, and we may also assume that $|Y| \geq n/3$ and $|X \setminus B| \geq n/3$. Let M be drawn uniformly at random from principal transversal matroids on $[n]$ meeting these assumptions. Due to symmetry we can assume without loss of generality that $B = [r(M)]$ - choosing M is now equivalent to generating a random bipartite graph on parts Y and $X \setminus B = [r(M) + 1, n]$.

Let H be a bipartite presentation of N. By Lemma 4.2.8 our random bipartite graph asymptotically almost certainly contains an induced subgraph isomorphic to H. Let the vertices of this subgraph be $X' \in X \setminus B$ and $Y' \subseteq Y$. Let $B' \subseteq B$ be the leaves in B adjacent to a vertex in Y'. Then $(M/(B \setminus B'))|X' \cong N$. $\square$

For the next corollary it may help to recall the universe of gammoids as shown in Figure 4.2.

Corollary 4.3.14. *Let N be a gammoid. Asymptotically almost all principal transversal matroids contain an N-minor*

Proof. We know every gammoid is a minor of a transversal matroid. Let N' be a transversal matroid with an N-minor. Asymptotically almost all principal transversal matroids contain an N'-minor, and thus an N-minor. $\square$

We can now tie the results of this chapter nicely to the work at the end of the previous chapter. Recall that we defined (Definition 3.4.2 a class $\mathcal{N}_r$, illustrated in 3.4, such that every $N \in \mathcal{N}_r$ was a forbidden minor of GS matroids. We claim that any such matroid N is also transversal. We shall prove this in the following section (Claim 4.4.3.1) as the simplest proof requires techniques yet to be introduced.

That enables us to say the following:

Theorem 4.3.15.   *(i) For $r \geq 4$, asymptotically almost every rank-r transversal matroid is a forbidden minor in GS matroids.*

(ii) Asymptotically almost every principal transversal matroid is a forbidden minor in GS matroids.

Proof. Let $N \in \mathcal{N}_4$. N is a forbidden minor of GS matroids, therefore so is any matroid containing an N-minor. But N is a transversal matroid of rank 4. Therefore:

  (i) By Theorem 4.2.7, for $r \geq 4$, asymptotically almost all rank-r transversal matroids contain an N-minor.

  (ii) By Theorem 4.3.13, asymptotically almost all principal transversal matroids have an N-minor.

                and hence, in either case, are GS-forbidden.   □

The latter case is also trivially implied by Theorem 4.3.12.

Finally, we can also link principal transversal matroids with fixed-rank transversal matroids.

Theorem 4.3.16. *Asymptotically almost every transversal matroid of rank r is principal.*

Proof. It suffices to show asymptotically almost every bipartite graph G in $\mathcal{G}(n,r)$ is a principal presentation. Let the r-part of G be $Y = \{y_1, \ldots, y_r\}$. Let p_i be the probability that there is no leaf adjacent to y_i. This is equal to $(1 - 2^{-r})^n$. And the probability that G is not a principal presentation is at most $\sum_{i=1}^{r} p_i = r(1 - 2^{-r})^n$, which clearly vanishes as n grows. $\square$

4.4 Minimal non-transversal matroids

In this section we shall classify the rank-3 matroids which are minimal non-transversal under taking series minors (i.e. every deletion or series contraction of the matroid is transversal). For brevity we shall refer to these as *series-minimal non-transversal* matroids.

Throughout the chapter we shall use the term *line* to mean a circuit of size 3.

4.4.1 A useful function for determining transversality

Before tackling this classification it will be helpful to introduce a new way of determining whether or not a matroid is transversal. The natural definitions of transversal matroids do not give rise to simple tests of transversality; instead we shall use a function described by Bonin [2] based on a related

function of Mason [15]. This function is defined recursively on subsets of the groundset, mapping them to an integer (not necessarily positive) value. For X a subset of $E(M)$, let the value of $\beta(X)$ be equal to the rank of M minus the rank of X minus the sum of the values taken by β on the cyclic flats (strictly) containing X. Formally,

Definition 4.4.1. Let M be a matroid, and for all $X \subseteq E(M)$ define recursively:

$$\beta(X) = r(M) - r(X) - \sum_{Y \in \mathcal{Z}(M): X \subset Y} \beta(Y)$$

where we recall that $\mathcal{Z}(M)$ is the collection of cyclic flats of M.

Since the β function is defined separately on each matroid, we may sometimes use $\beta_M(X)$ to mean the β function defined on M.

We note the following two lemmas associated with the β function. The first is a simple consequence of applying the β function to the rank-0 cyclic flat; the second is shown in, for example, Bonin's survey [2].

Lemma 4.4.2. *For any matroid M,* $\sum\limits_{Y \in \mathcal{Z}(M)} \beta(Y) = r(M)$.

Proof. Let Y_0 be the rank-0 cyclic flat. Then since every other cyclic flat contains Y_0, we have

$$\sum_{Y \in \mathcal{Z}(M)} \beta(Y) = \left(\sum_{Y \in \mathcal{Z}(M): Y \supset X_0} \beta(Y) \right) + \beta(Y_0)$$

$$= \left(\sum_{Y \in \mathcal{Z}(M): Y \supset Y_0} \beta(Y) \right) + \left(r(M) - r(Y_0) - \sum_{Y \in \mathcal{Z}(M): Y \supset Y_0} \beta(Y) \right)$$

$$= r(M) - r(Y_0) = r(M)$$

$$\square$$

Lemma 4.4.3. *A matroid M is transversal if and only if for all $X \subseteq E(M), \beta(X) \geq 0$.*

The second lemma is central to Mason [14] and Ingleton's [8] celebrated characterisation of transversal matroids.

In combination Lemmas 4.4.2 and 4.4.3 give us some easy checks for transversality. If we can find any subset of $E(M)$ which takes a negative value under β, we know M is not transversal. Similarly, since every circuit-hyperplane takes value 1, a transversal matroid M can have at most $r(M)$ circuit-hyperplanes.

As promised earlier, we can use the β function to demonstrate the transversality of matroids in the class $\mathcal{N}_r$, as defined at the end of the previous chapter (Definition 3.4.2) for all $r \geq 4$. This proof presents a nice illustration of the usefulness of the β function for quickly computing whether a matroid is transversal.

Claim 4.4.3.1. Let $N \in \mathcal{N}_r$. Then N is transversal.

Proof. N has non-bases $X \cup \{a,b\}, X \cup \{c,d\}, Y \cup \{a',b\}, Y \cup \{c,d\}$, where $|X \cap Y| \leq r(N) - 4$. Since N is sparse paving, the only cyclic flats are its non-bases, which each take value 1 under the β function. No three non-bases of N intersect, so every subset $X \in E(N)$ is either a subset of no cyclic flat, a subset of one cyclic flat, a subset of two cyclic flats, or the empty set. In each case we shall see that $\beta(X) \geq 0$. In the first case we have $\beta(X) =$

$r(N) - r(X) \geq 0$. In the second case we have $\sum_{Y \in \mathcal{Z}(N): X \subset Y} \beta(Y) = 1$ and $r(N) - r(X) \geq 1$, so $\beta(X) \geq 1$. In the third case we have $\sum_{Y \in \mathcal{Z}(N): X \subset Y} \beta(Y) = 2$ and $r(N) - r(X) \geq r(N) - |X| \geq 2$ by sparse pavingness of N, so $\beta(X) \geq 0$. In the final case we have $\sum_{Y \in \mathcal{Z}(N): X \subset Y} \beta(Y) = 4$ and $r(N) - r(X) = r(N) \geq 4$ by definition, so $\beta(X) \geq 0$. $\qquad\square$

4.4.2   Excluded series-minors of rank-3 transversal matroids

Classifying the series-minimal members of the class of non-transversal matroids involves considering a number of possible cases. We shall consider separately simple and non-simple matroids of rank 3.

The following lemma enables us to assume that circuits may only intersect in at most one element in the simple case.

Lemma 4.4.4. *Let M be a simple rank-3 series-minimal non-transversal matroid. For any rank-2 flat F of M, we have $|F| \leq 3$.*

Note that this condition is equivalent to saying M is co-paving.

Proof. Suppose not, and let F be a hyperplane in M of cardinality $|F| \geq 4$. Note that every 3-subset of F must be a circuit of M. Now let $M' = M - e$, for some $e \in F$. We know M' is transversal. Considering that every 3-subset of $F - e$ is also a 3-subset of F, it must be a circuit of M'. In a geometrical representation of M', all the elements of $F - e$ must therefore be arranged on a single "side". Since there can be at most one element at each of the two vertices in that side, and $F - e$ has at least 3 elements, there must be some $f \in F - e$ such that f is freely placed within that side.
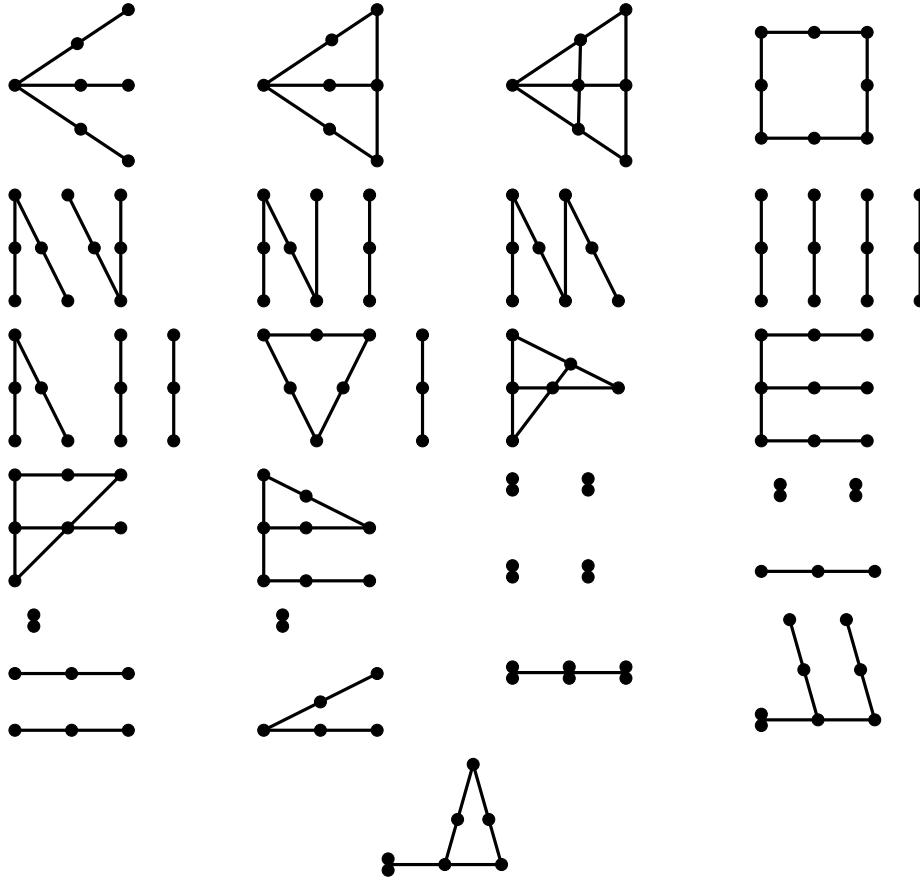
Figure 4.4: All the series-minimal non-transversal matroids of rank 3

Now, consider adding e to our geometrical representation. If we can do this, then M is transversal, which is a contradiction. So there is no place we can add e. But then removing f cannot make it possible to add e, so $M - f$ is non-transversal, contradicting minimality. □

4.4.2.1 Simple matroids

We shall break down the class of simple rank-3 matroids into three further subclasses. We shall use "lines" to refer to the circuits of size 3 (i.e. the non-bases).
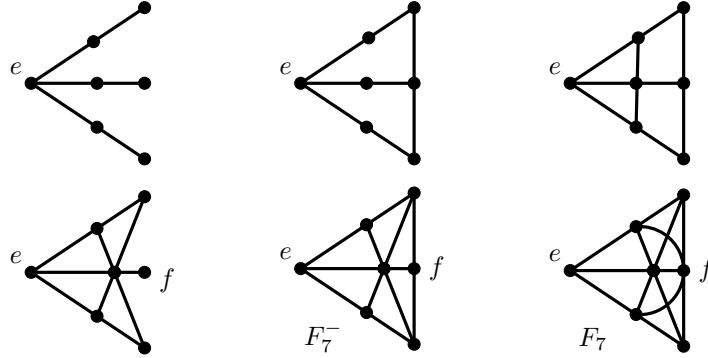
Figure 4.5: The possible minors restricting to the union of lines incident with e. Note that in each of the bottom row, the element f can be deleted to leave $M(K_4)$, which is itself series-minimal non-transversal. We have labelled the well-known Fano and Non-Fano matroids.

We first consider the class $\mathcal{S}'$ of those simple matroids in which a single element e is contained in at least three lines. Note this immediately ensures a matroid $M \in \mathcal{S}'$ is non-transversal. To see this, consider that each of the three lines are also circuit-hyperplanes and hence take value 1 under the β function. Now $\sum_{Y \in \mathcal{Z}(M):e \in Y} \beta(Y) \geq 3$, since the β function takes no negative values whilst M is transversal. But then $\beta(\{e\}) \leq r(M) - r(e) - 3 = 3 - 1 - 3 = -1$, so M cannot be transversal.

Considering the restriction to the 7 elements contained in those three lines, there are 6 possible arrangements, which are shown in Figure 4.5. Three of these matroids are minimal non-transversal; in the remaining cases, a single element (highlighted) may be removed, which leaves us with $M(K_4)$ (also non-transversal). Clearly if $M \in \mathcal{S}'$, M must contain one of these minimal arrangements (note that this is not quite an exact characterisation of $\mathcal{S}'$, since $M(K_4) \notin \mathcal{S}'$).

We now consider $\mathcal{S}''$ the class of simple matroids which are not in $\mathcal{S}'$ - i.e. in which no element is contained in 3 lines - and in which every line has at
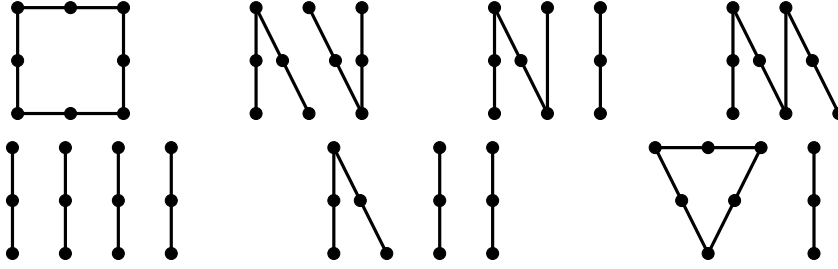
Figure 4.6: The series-minimal non-transversal members of S"

least one "loose" element (i.e. not contained in any other lines). We might reasonably consider drawing this class as a graph - each line can be viewed as an edge of the graph, with a loose element effectively being ignored (although we shall represent it as a point in the middle of the line), and the remaining two elements making up the vertices of the edge. Since we additionally have the condition that no element may be contained in three lines, $\mathcal{S}''$ corresponds to disjoint unions of paths and cycles in our graphical representation.

The following lemma characterises the series-minimal non-transversal members of $\mathcal{S}''$

Lemma 4.4.5. *$M \in \mathcal{S}''$ is series-minimal non-transversal if and only if M has exactly four lines.*

Proof. Clearly a matroid $M \in \mathcal{S}''$ is transversal if and only if it has three or fewer lines, since these can then be laid out on the three sides of a rank-3 simplex, whereas 4 lines cannot. Now if $M \in \mathcal{S}''$ has 4 lines, any deletion removes at least one line, so M is series-minimal. But if M has 5 or more lines, we may delete a loose element to get a proper series-minor of M which still has 4 or more lines. $\qquad \square$

These matroids are easily catalogued; there are 7 of them, shown in Figure 4.6.

The remaining simple matroids are characterised by having no elements in three lines, but at least one line C every element of which is contained in another line. We call such a line *tied* or say it has no *loose elements*. So there are at least 4 lines - consider the restriction to the elements spanned by these 4 lines; there are 7 possibilities, shown in Figure 4.7. Note that each of these must be non-transversal, as we have at least 4 circuit-hyperplanes. 4 of these turn out to be minimal non-transversal, and the other 3 reduce to cases already covered.

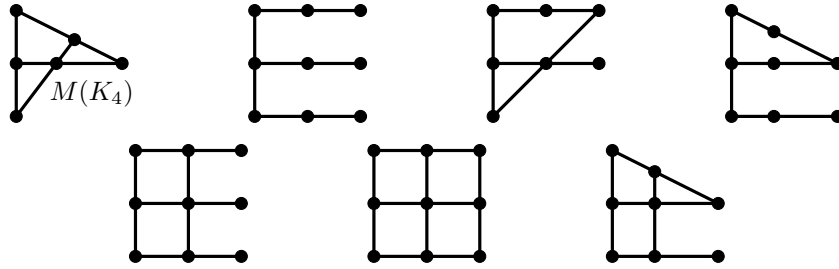Thus we have described 14 series-minimal non transversal simple matroids.

Figure 4.7: The 7 possibilities when restricting to the union of lines incident with a tied line. The 4 in the top row are series-minimal non-transversal.

4.4.2.2   Non-simple matroids

Next we consider non-simple matroids, i.e. matroids containing at least one parallel pair. First we shall consider the class $\mathcal{N}'$ of non-simple matroids where no member of a parallel pair is contained in a line. Consider that we can certainly fit three parallel pairs, or one parallel pair and one line, into our simplex. Therefore we will completely classify the series-minimal non-transversal minors of $\mathcal{N}'$ if we can show that all matroids of $\mathcal{N}'$ containing 4 parallel pairs, or two parallel pairs and a line, or a parallel pair and

two lines, are non-transversal. Again by considering the restriction to the elements spanned by the relevant parallel pairs and lines, we can show that all such matroids contains one of four minors (Figure 4.8), each of which is easily verified to be series-minimal non-transversal.
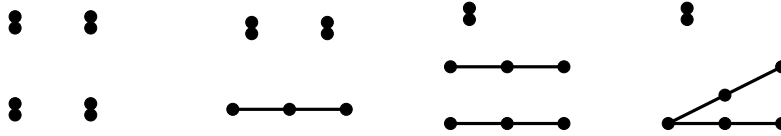

Figure 4.8: Minors which must be contained in non-transversal members of N'. These are all series-minimal non-transversal.

Finally we consider the remaining non-simple matroids, in which at least one parallel pair is contained in a line. Note there can be at most three lines, and these cannot intersect at a point, since the deletion of a single element from a parallel pair must leave a transversal matroid.

If we have three parallel pairs in a line, that alone forms a series-minimal non-transversal matroid.

If we have a line containing two parallel pairs, there cannot be a line disjoint to that, else we can reduce to one of the non-transversal cases in N'. Similarly we cannot have 4 parallel pairs. With those conditions in place there are still 13 possibilities, but it is easily verified that all these are either transversal or reduce to the case of three parallel pairs in a line (via a series contraction), so we have no series-minimal non-transversal matroids containing two parallel pairs in a line.

Finally we consider the case where no line contains more than one parallel element. It's easily see that there must be three lines, else our matroid is transversal. Note also that if we have two lines not containing a parallel element, they must both intersect with the line that does, else we can delete

a loose element from that line to reduce to a case from N'. These conditions reduce us to 8 cases, of which two are transversal, and the other six all reduce to one of two new series-minimal non-transversal cases.

Thus in tidying up the remaining cases we have found a final three series-minimal non-transversal matroids, which are shown in Figure 4.9.
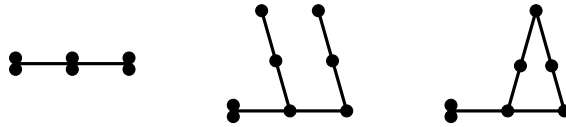

Figure 4.9: All the series-minimal non-transversal matroids in which at least one parallel pair is contained in a line

Thus we have completely catalogued the excluded series minors for non-transversal matroids of rank 3 - there are 21 of them. We shall now turn our attention to the general case of rank-r transversal matroids.

4.4.3   Excluded series-minors of rank-r transversal matroids

Theorem 4.4.6. *For any rank r, the class of transversal matroids has finitely many excluded series-minors of rank r.*

We shall make use of the following result, which is well established (see for example Bonin [2]). Transversal matroids may be defined as those matroids which can be represented on a simplex such that each cyclic flat of the matroid corresponds to a face of the simplex, and the following lemma is an immediate consequence of this.

Lemma 4.4.7. *A transversal matroid of rank r contains at most 2^r cyclic*

flats.

We also state and prove the following lemma. By elements *free in a cyclic flat*, we mean elements such that any set spanning that element must also span the cyclic flat - or to look at it another way, e is free in F if the only cyclic flats containing e are F and supersets of F.

Lemma 4.4.8. *Suppose M is a series-minimal non-transversal matroid. Any cyclic flat F of M can have no more than $r(F) + 1$ elements free within that cyclic flat.*

Proof. Suppose otherwise: there are free elements $x_1, \ldots, x_{r+2} \in F$, and consider $M \setminus x_1$, a transversal matroid. We shall show that $M \setminus x_1$ cannot be transversal by demonstrating that the deletion of x_1 cannot increase the β value of any subset of $E(M)$. Recall that the β function is defined recursively by

$$\beta(X) = r(M) - r(X) - \sum_{Y \in \mathcal{Z}(M): X \subset Y} \beta(Y)$$

We shall have to separately consider possible changes to $r(M)$, $r(X)$ and $\sum \beta(Y)$.

Firstly, since x_1 is contained in a cyclic flat, it cannot be a loop of M, and so $r(M \setminus x_1) = r(M)$.

Claim 4.4.8.1. Suppose $\{x_1 \cdots, x_{r+1}\} \subseteq X$, or $x_1 \notin X$. Then $r_M(X) = r_{M \setminus x_1}(X \setminus x_1)$.

Proof of claim. The latter case ($x_1 \notin X$) is trivial. For the former, suppose deleting x_1 reduces the rank of X - then any basis B of $M|X$ must contain

x_1, and any B cannot contain all of $x_2, \ldots, x_{r+2}$ since they along with x_1 form a dependent set. So there is some $i \in \{2, 3, \ldots, r+1\}$ such that $x_i \notin B$, and x_i must be in the closure of $B \setminus x_1$, else $(B \setminus x_1) \cup x_i$ would be independent of rank $r_M(X)$. But if x_i is spanned by $B \setminus x_1$ then (since x_i is free in F) so is every other element of F, including x_1, so B cannot be a basis. $\qquad\square$

Claim 4.4.8.2. Suppose X is a cyclic flat of M. Then X or $X \setminus x_1$ is a cyclic flat in $M \setminus x_1$.

Proof of claim. Since x_1 is free in F, all cyclic flats of M are either supersets of F, or do not contain x_1. Again the latter case is trivial; for the former case, $X \setminus x_1$ still contains $x_2, \ldots, x_{r+2}$. Now, none of these x_i can be a coloop in $(M \setminus x_1)|(X \setminus x_1)$ because $\{x_2, \ldots, x_{r+2}\} \setminus x_i$ is a basis thereof). And suppose there is some element $f \in F \setminus \{x_2, \ldots, x_{r+2}\}$ that becomes a coloop upon the deletion of x_1. Then we must have a circuit C_1 of $M|X$ with $f, x_1 \in C_1$. But also $\{x_1, \ldots, x_{r+1}\}$ is a circuit of $M|X$. We can now invoke one of the circuit axioms for matroids - the circuit exchange axiom **C3**. Let $\mathcal{C}(M)$ denote the collection of circuits of M.

C3(strong):

$$C_1, C_2 \in \mathcal{C}(M), e \in C_1 \cap C_2, f \in C_1 \setminus C_2 \implies \exists C_3 \in \mathcal{C}(M) : C_3 \subseteq ((C_1 \cup C_2) \setminus e), f \in C_3$$

Applying **C3** on $M|X$ with $C_2 := \{x_1, \ldots, x_{r+1}\}$ and $e := x_1$ we obtain a circuit in $M|X$ that contains f but not x_1. This then must also be a circuit in $(M \setminus x_1)|(X \setminus x_1)$, meaning f is still not a coloop. $\qquad\square$

Claim 4.4.8.3. Suppose $\{x_1 \cdots, x_{r+2}\} \subseteq X$, or $x_1 \notin X$. Then $\beta_M(X) = \beta_{M \setminus x_1}(X \setminus x_1)$.

Proof of claim. The proof is by top-down recursion. We have already established that $r(E(M)) = r(M)$ is unchanged. Suppose the statement holds for all sets of size greater than k meeting the conditions, and consider a set X of size k. Again we remind ourselves of the recursive definition of the β function.

$$\beta(X) = r(M) - r(X) - \sum_{Y \in \mathcal{Z}(M): X \subset Y} \beta(Y)$$

We have already shown in previous claims that $r(M)$ and $r(X)$ are unchanged. Now we also note that $\sum_{Y \in \mathcal{Z}(M): X \subset Y} \beta(Y)$ is unchanged: all cyclic flats containing X are still cyclic, and they are sets of size greater than k that meet the conditions, so by assumption their own β values are unchanged. Finally there cannot be some cyclic flat Y of M such that $X \not\subset Y$ but $(X \setminus x_1) \subset Y$ (this is clearly impossible when $x_1 \notin X$, so we must have $\{x_1 \cdots, x_{r+2}\} \subseteq X$, and then for example $x_2 \in Y$, but since x_2 is free in F we have $Y \supseteq F$, so $x_1 \in Y$. $\qquad\square$

Now, since the β function on $M \setminus x_1$ must never take a negative value, we see that any set $X \in E(M)$ with $\beta_M(X) < 0$ cannot have $\beta_M(X) = \beta_{M \setminus x_1}(X \setminus x_1)$, and so such an X must satisfy $x_1 \in X$. But by a similar argument we also have that $x_2 \in X$, and so on. Therefore we must have $\{x_1 \cdots, x_{r+2}\} \subseteq X$. But then our earlier claims tell us that $\beta_M(X) = \beta_{M \setminus x_1}(X \setminus x_1)$, so we have reached a contradiction. $\qquad\square$

Proof of Theorem 4.4.6. We shall demonstrate that series-minimal non-transversal matroids of rank r are bounded in size, and hence number. To do this we shall suppose M is series-minimal non-transversal and show that M containing a large number of elements would force either some deletion $M \setminus f$ to

contain too many cyclic flats (violating the condition of Lemma 4.4.7), or M
to contain too many free elements in a cyclic flat (violating Lemma 4.4.8).
In particular, we show that $|E(M)| \leq r(2^{2r} - 1)$.

Suppose M is a series-minimal non-transversal matroid of rank r, and con-
sider the matroid $M \setminus e$ formed by the deletion of an element e. By definition
M is not transversal, but $M \setminus e$ must be. Consider a presentation of $M \setminus e$
on the simplex, and define the *parent face* of an element f to be the unique
minimal face of the simplex in which f is contained in this presentation. Note
that this face represents a cyclic flat of $M \setminus f$ except in the case where the
face has dimension 1 (i.e. is a vertex of the simplex) and contains no elements
other than f. We can effectively ignore this case, because we are only inter-
ested in faces containing a large number of elements. Now, the elements of
$M \setminus e$ may be partitioned by parent face; necessarily (pigeonhole principle) a
large number of elements will share the same parent face, which means that
- if the number of elements in M is sufficiently large - $M \setminus e$ can be forced
to contain many free elements within a single cyclic flat. Specifically, if M
contains n elements, then some cyclic flat of $M \setminus e$ must contain at least
$(n - 1)/(2^r - 1)$ free elements. In particular if $n \geq r(2^{2r} - 1) + 1$, we have
that

$$(n - 1)/(2^r - 1) = \frac{1}{2^r - 1}(r(2^{2r} - 1)) = r(2^r + 1)$$

Now suppose we have some set Y of k free elements in a cyclic flat F of $M \setminus e$.
Either F or $F \cup e$ must be a cyclic flat in M (in either case the argument
is very similar: we shall prove the case for F), and since by Lemma 4.4.8
this is not to contain more than $r(F) + 1$ free elements, there must be at
least $k - r(F) - 1$ elements of Y each contained in a cyclic flat of M, that

is not a cyclic flat of $M \setminus e$. The only way such a cyclic flat could arise is if it contains e, so we have at least $k - r(F) - 1$ members of Y appearing in cyclic flats with e. However, if we have $r(F)$ members of Y appear in the same cyclic flat with e, then that cyclic flat spans F and so does not reduce the freedom of any element in F. So there must be a collection $\mathcal{L}$ of at least $\frac{k-r(F)-1}{r(F)-1}$ cyclic flats each containing e and some members of Y, and for any $G \in \mathcal{L}$, there are at least $k - r(F) + 1$ elements in $Y \setminus G$.

Now we want to count pairs (G, f) where $G \in \mathcal{L}$ and f is an element of Y that is "missed" by G, i.e. e $f \in Y \setminus G$. The number of pairs is at least

$$(k - r(F) + 1)\frac{k - r(F) - 1}{r(F) - 1}$$

and there are k elements in Y, so we can see that there must exist some element $f \in Y$ such that

$$|\{G \in \mathcal{L} : f \notin G\}| \geq \frac{(k - r(F) + 1)(k - r(F) - 1)}{k(r(F) - 1)} = \frac{(k - r(F))^2 - 1}{k(r(F) - 1)}$$

But now $M \setminus f$ must contain at least $\frac{(k-r(F))^2-1}{k(r(F)-1)}$ cyclic flats. In particular, if $k \geq 2^r(r-1) + 2r + 1$, then the number of cyclic flats in $M \setminus f$ will be strictly greater than 2^r, meaning that by Lemma 4.4.7 $M \setminus f$ cannot be transversal, thereby contradicting minimality of M.

If $F \cup e$ rather than F is a cyclic flat in M, the argument proceeds identically, save for specifying that $F \cup E$ should not be considered a member of $\mathcal{L}$.

Hence in either case we have a bound on the possible size, and therefore possible number, of series-minimal non-transversal matroids of rank r. $\square$

4.4.4   The antichain of excluded series-minors for transversal matroids

In the general transversal case, i.e. not restricting to a fixed rank, it is easily seen that we have an infinite number of forbidden minors. Specifically, for any $n \geq 3$, we can construct a series-minimal non-transversal matroid M_n of rank n as follows: take elements $E(M_n) = \{a_1, \ldots, a_n, b_1, \ldots, b_n\}$ and define the matroid by its circuits:

$$\mathcal{C}(M_n) = \{\{e_1, \ldots, e_n\} : e_i \in \{a_i, b_i\}\} \cup \{\{a_1, b_1\}, \ldots, \{a_n, b_n\}\}$$

This is in fact the graphic matroid with graphical representation consisting of an n-cycle with doubled-up edges. The class $\{M_n : n \in \mathbb{N}_{>2}\}$ forms an infinite antichain under inclusion of series minors.

Figure 4.10: M_3, a forbidden series minor for transversal matroids

Question 4.4.9. Let $\mathcal{A}$ be the antichain of excluded minors for the class of transversal matroids. Is $\mathcal{A}$ maximal?

This question is equivalent to saying "Is every transversal matroid a series minor of an excluded series minor of transversal matroids". The corresponding question for minors has already been answered in the affirmative for matroids representable over an infinite field (Mayhew, Newman and Whittle [18]) and for gammoids (Mayhew [16]). Both were proven using a similar constructive strategy in which by a series of steps the original target minor could be built up to contain an awkward collection of hyperplanes that "just" blocked them from the original class, thereby forming a minimal excluded minor.

The first steps in this strategy involve showing that the target minor is a mi-

nor of another matroid in the same class whose elements can be partitioned into 2 disjoint bases. Then we aim to show that the target minor is a minor of another matroid in the same class whose elements can be partitioned into 2 disjoint hyperplanes. We shall show that the first of these steps is easily proven for transversal matroids, but provide a counterexample to the second, casting considerable doubt on whether this kind of strategy is viable for transversal matroids.

In this section we shall make extensive use of bipartite presentations in our arguments, and so it is wise to regather our intuition in this arena, in particular regarding the effects of deletions and series contractions. Let G be a bipartite graph and $M_T(G)$ the transversal matroid of which G is a presentation. Let X and Y be the parts of G, with X corresponding to the groundset of $M_T(G)$.

The deletion operation is straightforward - a deletion of element x from M corresponds to deleting the same element (vertex) e from G. In the other direction, we may reverse the process (an *extension*) only by adding an element (vertex) x' to X, and linking it to some collection of vertices in Y to form a new graph G'. Then M is obtained from $M_T(G')$ by the deletion of x'.

Series contractions are rather more subtle. A series pair $\{x_1, x_2\}$ is parallel in the dual matroid, so they must link to the same vertices in G. Moreover, the complement of our series pair forms a hyperplane, so there must be one vertex in Y (call it y) that only links to x_1 and x_2. A series contraction corresponds to deleting y and either x_1 or x_2. The reverse (a *series extension*) can only be achieved by picking an element (vertex) x in X, and adding new

elements (vertices) x' to X and y to Y. Then we must link x to y, and x' to precisely the same vertices as x, to form our new graph G'. Then M is obtained from $M_T(G')$ by contracting the series pair (x, x').

Theorem 4.4.10. *Let M be a transversal matroid. Then M is a series minor of a transversal matroid M' such that the groundset of M' can be partitioned into two bases.*

Proof. Let B be a basis of M, and let $S = E(M) \setminus B$. Suppose M has size n and rank r; then let $S = \{s_1, \ldots, s_{n-r}\}$. Add elements $T = \{t_1, \ldots, t_{n-r}\}$ such that each t_i is a series extension of s_i. Call the resultant matroid M''. Now imagine a presentation of M'' as a bipartite graph G with parts $X = E(M)$ and Y. Clearly there are vertices $v_1, \ldots, v_r$ in Y that match with B and vertices $u_1, \ldots, u_{n-r}$ in Y which can be ordered such that u_i links only to the pair (s_i, t_i). Now form a new bipartite graph G' by adding to the groundset new vertices $w_1, \ldots, w_r$ such that each w_i links only to v_i. Finally we define $M' := M_T(G')$. Note that M' has rank n and that by construction one basis of M' is $E(M) = B \cup \{s_1, \ldots, s_{n-r}\}$.

We claim that M' satisfies our theorem. Firstly, M is a series minor of M'; that is clear since we can retrieve M by deleting $w_1, \ldots, w_r$ and contracting the series pairs $(s_1, t_1), \ldots, (s_{n-r}, t_{n-r})$. And $E(M')$ can be partitioned into sets $E(M)$ and $\{w_1, \ldots, w_r, t_1, \ldots, t_{n-r}\}$. We have already seen that the first of these is a basis - since each w_i matches to v_i, and each t_i matches to u_i, the latter is also a basis. $\qquad\square$

However we shall introduce a transversal matroid which provides a counterexample to the analogous claim for hyperplanes.

Theorem 4.4.11. *Let N be the 6 element matroid of rank 3 whose non-trivial circuits are precisely three disjoint parallel pairs (also the graphic matroid induced by a path of length 3 with doubled-up edges). Suppose M is a transversal matroid containing N as a series minor. Then the groundset of M cannot be partitioned into two hyperplanes.*
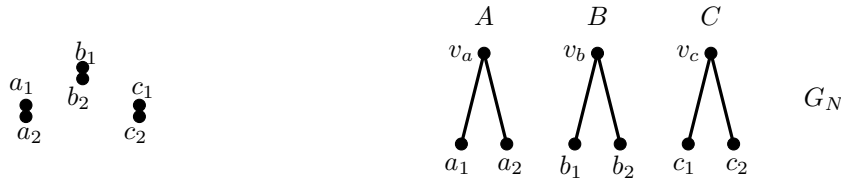

Figure 4.11: The matroid N. On the left, N drawn in conventional format. On the right, the only bipartite presentation of N (with components labelled for reference).

For the proof we shall largely consider the (maximal) bipartite presentations of N and M, which we shall denote G_N and G_M. In Figure 4.11 we can see G_N, the maximal bipartite presentation of N (in fact the only bipartite presentation: removing any edge creates a loop, and adding any edge relaxes a parallel pair). We've labelled the three components as A, B and C.

Proof. Suppose we have an M such that the groundset $E(M)$ partitions into two hyperplanes, H_1 and H_2. Let $G(M)$ be a bipartite presentation of M with parts F and $E = E(M)$. Since H_2 is a hyperplane of M, there must be some vertex $f_1 \in F$ such that every element (vertex) of H_1 links to f_1, and by a symmetrical argument there must also be some $f_2 \in F$ linked to every element (vertex) of H_2. In other words, there must exist vertices f_1 and f_2 which *cover* all of the elements (vertices) in E (meaning that the union of neighbourhoods $\Gamma(f_1) \cup \Gamma(f_2)$ is equal to E.

Recall that M is obtained from N by a sequence of single-element moves, each of which is either an extension (reversing a deletion), or a series extension (reversing a series contraction), and recall that we discussed earlier in this section what those moves look like in the context of bipartite presentations. We shall show that, however we extend the graph G_N to form G_M, there can be no pair of vertices which covers the 6 elements (vertices) of $E(N)$. Plainly there is no such pair in G_N, since any two vertices from the top part of the graph cover only 4 elements (vertices) of $E(N)$. Moreover, we note that any series extension does not involve adding any edges including elements (vertices) of $E(N)$, nor does it introduce new vertices that elements (vertices) of $E(N)$ might later be linked to. This means that if we did obtain an M in which two vertices of G_M covered all of $E(N)$, the same would hold true even if we reversed all the extension. So without loss of generality we may assume that M is obtained from N only by series extensions.

Every series extension introduces one new vertex to the top half of the graph (increasing the rank of our putative M by 1) and one to the bottom half (the groundset of our putative M). We claim that if the only moves we may make are to add series pairs, and to update the graph to a maximal presentation, then the three components of G_N (labelled A, B, C in Figure 4.11) cannot be connected. If this is the case, then clearly no pair of vertices can link to all three components, and so no pair of vertices can cover all the elements (vertices) of $E(N)$.

Consider that so long as the components remain pairwise disconnected, each series pair must be created using an existing element from only one of the components. Let's focus on what happens to a single component, A, whilst

keeping the others fixed. We'll use the notation $E(X)$ to denote $V(X) \cap E(M(X \cup B \cup C))$, i.e. the "bottom part" of X.

Claim 4.4.11.1. Let A' be formed from A by recursively forming series pairs from a new element and an existing element. Then $E(A')$ is a circuit of the matroid induced by the graph $G' = A' \cup B \cup C$.

Proof of claim. This can be shown with a simple inductive proof. Clearly in the base case $E(A)$ is a circuit of G_N. Suppose then that A'' is a rank $(r-1)$ circuit of $M(A'' \cup B \cup C)$ and we obtain A' from A' by adding a series pair formed of an existing element (vertex) $e \in E(A'')$, a new element (vertex) e' and a new vertex f in the upper part of our bipartite presentation. The component A' now consists of $r+1$ elements (vertices) in $E(A')$ and r vertices in the top part, so there is clearly no matching from the bottom part to the top, and hence it is dependent in the induced matroid. It remains to show that every r-element subset of the $E(A')$ is independent, i.e. has a matching to the top part. We can divide this into two cases, firstly where the subset contains e', and secondly where it contains everything except e'. In either case, the subset contains either e' or e together with $r-1$ elements from $E(A'')$. Since $E(A'')$ is a circuit, the $r-1$ elements from $E(A'')$ are in matching with the top part of A''. We can then complete the matching from $E(A')$ by adding the edge from either e' or e to the new vertex f. $\qquad\square$

Clearly adding series pairs does not by itself connect any pair of components, and it is now clear that updating to obtain a maximal presentation cannot do so either (although it will certainly add more edges within the components, in fact making each of them complete). For example, linking from the bottom part of A' to the top part of another component would relax the circuit $E(A')$ in the induced matroid. So the three components must remain separate, and

no pair of vertices can cover all six elements of $E(N)$.   □

Note that although this kind of argument presents an obstacle to reusing the methods of Mayhew, Newman and Whittle, it is not necessarily strong evidence of the result itself being false for transversal matroids, and it is possible that other angles of attack might prove more fruitful. For example, although our matroid N proved resistant to this approach, it is almost trivially a series minor of an excluded series minor for transversal elements (simply add two new elements in a parallel pair, as seen in the leftmost matroid of Figure 4.8).

4.4.5   Sparse paving excluded series minors of transversal matroids

It is relatively straightforward to characterise the sparse paving series-minimal non-transversal elements of rank r. Sparse paving matroids are entirely defined by a collection of non-bases of cardinality r, of which any pair can intersect in at most $r - 2$ elements. These non-bases are the only cyclic flats in a sparse paving matroid, which means calculating β values becomes fairly straightforward - if $\mathcal{L}(M)$ is the set of non-bases of a sparse paving matroid M, then for any $X \in E(M)$ we have

$$\beta(X) = r(M) - r(X) - |\{L \in \mathcal{L}(M) : X \subset L\}|$$

This means that a negative β value is obtained if and only if we have some $X \subset E(M)$ such that X is in the intersection of at least $r(M) - |X| + 1$ members of $\mathcal{L}(M)$. Note that X can be empty, in which case the condition is equivalent to having $r + 1$ non-bases. This situation makes sparse paving

matroids a convenient realm in which to answer questions about transversal matroids.

One question we might ask is what is the minimal value that β might take on a series-minimal non-transversal matroid. Since the minimum value of β on a matroid is always lower for non-transversal matroids than for transversal matroids, it might be viewed as some kind of measurement of how non-transversal a matroid M is. Naturally we might expect series-minimal non-transversal matroids to be only "a little bit" non-transversal, since any single deletion is enough to make them transversal. In other words, we might expect that for any series-minimal non-transversal M, the minimum value of $\beta_M(X)$ is close to zero - for example, in all the series-minimal non-transversal matroids of rank 3, we have $\min\limits_{X \subseteq E(M)} (\beta_M(X)) = -1$. In fact however we shall show the following result:

Lemma 4.4.12. *For any integer m, there exists a series-minimal non-transversal matroid M such that $\beta_M(\emptyset) = -m$*

Proof. We choose M to have rank $2m$ and groundset $\{1, \ldots, 6m\}$. Let M be sparse paving with non-bases precisely the sets of form $\{2k + 1, 2k + 2, \ldots, 2k + 2m\}$, where $0 \leq k \leq 3m$, and addition is modulo $6m$. There are $3m$ non-bases (remember the cyclic flats of a sparse paving matroid are precisely its non-bases), so

$$\beta_M(\emptyset) = r(M) - r(\emptyset) - \sum_{Y \in \mathcal{Z}(M)} \beta_M(Y) : Y \neq \emptyset$$

$$= 2m - 0 - 3m = -m$$

It remains to show that M is minimally non-transversal. M has no series

pairs, so we need only consider deletions, and the symmetry of M means that every deletion gives an isomorphic matroid. Without loss of generality, assume we delete the element labelled $6m$ giving a new matroid M^-. There remain $2m$ non-bases, so $\beta_{M^-}(\emptyset) = 0$. The only other sets which could take negative β value are those which are strict subsets of one or more non-basis of M^-. Let X be such a set. Note that since M^- is sparse paving, X must be independent, so $r(X) = |X|$. Each non-basis covers the elements in an interval of width $2m$ under our labelling, and that interval must include both the minimum labelled and maximum labelled member of X. So $|\{Y : Y \in \mathcal{Z}(M^-), X \subseteq Y\}| \leq m - \lceil \frac{|X|}{2} \rceil$, in turn giving

$$\beta_{M^-}(X) \geq r(M^-) - r(X) - m + \lceil \frac{|X|}{2} \rceil = m - \lfloor \frac{|X|}{2} \rfloor > 0$$

with the last inequality following from the fact that X is strictly contained in a non-basis, so $|X| < 2m$. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad\square$

4.5   Axiomatisability of transversal matroids

Another interesting (and little explored) topic relating to transversal matroids is that of axiomatisability. In the light of recent results for related classes [20] we shall try to obtain similar results for transversal matroids.

Conjecture 4.5.1. *The class of transversal matroids is not axiomatisable in MSOL (Monadic Second-Order Logic) for matroids.*

Although we believe this conjecture to be true, and have been unable to find a sentence that characterises transversal matroids, this seems to be difficult to prove. For example, methods involving EF games seem to run into great

difficulty.

We can however prove non-axiomatisability in the language of *M-logic*, as defined in the paper of Mayhew, Newman and Whittle [20]. In this paper it is shown that the class of representable matroids cannot be axiomatised by a finite set of sentences in *M*-logic.

Firstly, let's define MSOL and *M*-logic.

Monadic Second Order Logic for matroids, hereafter referred to as *MSOL*, is described in detail in Section 2 of [20]. The introduction we offer here is slightly less formal, with the aim only of providing a clear enough explanation for the reader to understand the subsequent work. So, in simplified terms, MSOL consists of the following:

Variables $x_1, x_2, \cdots ; X_1, X_2, \cdots$

Constants $E; \emptyset; 0, 1, 2, \cdots$

Functions $| \cdot |, \{\cdot\}, \bar{\cdot}, r(\cdot), +, \cup, \cap$

Relations $=, \in, \subseteq, \leq$

Logical symbols $\neg, \vee, \wedge, \exists, \forall$

Terms in MSOL are divided into 3 classes: firstly $\mathcal{E}$, which is the infinite set of variables $x_1, x_2, \cdots$ and corresponds to the groundset of a matroid. Secondly $\mathcal{S}$, which corresponds to subsets of the groundset, is the smallest collection of expressions satisfying:

(i) $E, \emptyset \in \mathcal{S}$

(ii) $X_1, X_2, \cdots \in \mathcal{S}$

(iii) $x_i \in \mathcal{E} \implies \{x_i\} \in \mathcal{S}$

(iv) $X, Y \in \mathcal{S} \implies \bar{X}, X \cup Y, X \cap Y \in \mathcal{S}$

Finally we have $\mathcal{N}$, representing non-negative integers, the smallest collection of expressions such that:

(i) $0, 1, \cdots \in \mathcal{N}$

(ii) $X \in \mathcal{S} \implies |X|, r(X) \in \mathcal{N}$

(iii) $p, q \in \mathcal{N} \implies p + q \in \mathcal{N}$

Any x_i, X_i that appear in a term are called *variables* of the term (a more formal definition is given in [20]). We can now define *atomic formulas, formulas* and *free variables.*

Atomic formulas are statements in the following four categories:

(i) '$x = y$' for $x, y \in \mathcal{E}$

(ii) '$X = Y$' or $X \subseteq Y$, for $X, Y \in \mathcal{S}$

(iii) '$p = q$' or $p < q$, for $p, q \in \mathcal{N}$

(iv) '$x \in X$', for $x \in \mathcal{E}$ and $X \in \mathcal{S}$

The variables of an atomic formula are simply the union of the variables in the terms which are used in the atomic formula. These variables are all considered *free variables* in the atomic formula, however for defining *formulas*

we shall need to distinguish between the set $\mathrm{Var}(P)$ of variables, and the set $\mathrm{Fr}(P)$ of free variables, of a formula P. A *formula* is an expression which can be arrived at by finite application of the following rules:

(i) every atomic formula P is a formula, with $\mathrm{Fr}(P) = \mathrm{Var}(P)$

(ii) If P is a formula, then '$\neg P$' is a formula, with the same variables and free variables as P

(iii) If P and Q are formulas and $\mathrm{Fr}(P) \cap (\mathrm{Var}(Q) \setminus \mathrm{Fr}(Q)) = \emptyset = \mathrm{Fr}(Q) \cap (\mathrm{Var}(P) \setminus \mathrm{Fr}(P))$, then '$P \wedge Q$' and '$P \vee Q$' are formulas, with $\mathrm{Var}(P \wedge Q) = \mathrm{Var}(P \vee Q) = \mathrm{Var}(P) \cup \mathrm{Var}(Q)$ and $\mathrm{Fr}(P \wedge Q) = \mathrm{Fr}(P \vee Q) = \mathrm{Fr}(P) \cup \mathrm{Fr}(Q)$

(iv) If P is a formula and $X_i \in \mathrm{Fr}(P)$, then $\exists X_i P$ and $\forall X_i P$ are formulas with the same variables as P, and $\mathrm{Fr}(\exists X_i P) = \mathrm{Fr}(\forall X_i P) = \mathrm{Fr}(P) \setminus \{X_i\}$

(v) If P is a formula and $x_i \in \mathrm{Fr}(P)$, then $\exists x_i P$ and $\forall x_i P$ are formulas with the same variables as P, and $\mathrm{Fr}(\exists x_i P) = \mathrm{Fr}(\forall x_i P) = \mathrm{Fr}(P) \setminus \{x_i\}$

Finally, a *sentence* in MSOL is a formula P satisfying $\mathrm{Fr}(P) = \emptyset$.

Now we need to understand what it means for a matroid M to satisfy a formula P. There is nothing particularly surprising to this definition; indeed it roughly means that P is true when we interpret the notation of MSOL in the most obvious way with respect to matroid theory and drawing variables from $E(M)$. However we ought to offer a more formal definition (and note that even this is somewhat abbreviated: see [20] for more).

An *interpretation* of P is a set E and a pair $(\phi_{\mathcal{E}}, \phi_{\mathcal{S}})$ of maps such that

$\phi_{\mathcal{E}} : \mathrm{Fr}(P) \cap \mathcal{E} \to E$ and $\phi_{\mathcal{S}} : \mathrm{Fr}(P) \cap \mathcal{S} \to \mathcal{P}(E)$, where $\mathcal{P}(E)$ denotes the power set of E. Let $P_{\phi_{\mathcal{E}}, \phi_{\mathcal{S}}}$ denote the expression created by taking P and replacing every variable in $x \in \mathrm{Fr}(P) \cap \mathcal{E}$ with $\phi_{\mathcal{E}}(x)$ and every variable $X \in \mathrm{Fr}(P) \cap \mathcal{S}$ with $\phi_{\mathcal{E}}(X)$.

When evaluating the expression $P_{\phi_{\mathcal{E}}, \phi_{\mathcal{S}}}$ we may understand all symbols of MSOL to take their usual meaning with respect to matroid theory.

Suppose we have a structure M (which we shall assume is a matroid) equipped with a groundset $E(M)$ and a rank function r on the power set of $E(M)$. We shall say M *satisfies* a formula P if there exists an interpretation $(E(M), \phi_{\mathcal{E}}, \phi_{\mathcal{S}})$ such that the following conditions are met:

(i) If P is an atomic formula, then $P_{\phi_{\mathcal{E}}, \phi_{\mathcal{S}}}$ is true

(ii) If $P = \exists X_i Q$ for $X_i \in \mathcal{S}$, then there is some subset $X'_i \subseteq E(M)$ such that $Q_{\phi_{\mathcal{E}} \cup (X_i, X'_i), \phi_{\mathcal{S}}}$ is true; if $P = \forall X_i Q$ for $X_i \in \mathcal{S}$, then $Q_{\phi_{\mathcal{E}} \cup (X_i, X'_i), \phi_{\mathcal{S}}}$ is true for all $X'_i \subseteq E(M)$

(iii) If $P = \exists x_i Q$ for $x_i \in \mathcal{E}$, then there is some element $x'_i \subseteq E(M)$ such that $Q_{\phi_{\mathcal{E}} \cup (x_i, x'_i), \phi_{\mathcal{S}}}$ is true; if $P = \forall x_i Q$ for $X_i \in \mathcal{S}$, then $Q_{\phi_{\mathcal{E}} \cup (x_i, x'_i), \phi_{\mathcal{S}}}$ is true for all $x'_i \subseteq E(M)$

(iv) If $P = \neg Q$ then $Q_{\phi_{\mathcal{E}}, \phi_{\mathcal{S}}}$ is not true

(v) If $P = Q \wedge R$ then $Q_{\phi_{\mathcal{E}}, \phi_{\mathcal{S}}}$ and $R_{\phi_{\mathcal{E}}, \phi_{\mathcal{S}}}$ are true; if $P = Q \vee R$ then $Q_{\phi_{\mathcal{E}}, \phi_{\mathcal{S}}}$ or $R_{\phi_{\mathcal{E}}, \phi_{\mathcal{S}}}$ is true.

Finally, we can introduce M-logic, defined as a subset of the formulas of MSOL.

Definition 4.5.2 (*M*-logic)**.** A formula in MSOL is in *M-logic* if it takes one of the following forms:

(i) $\exists X_1 \cdots \exists X_l \exists x_1 \cdots \exists x_m P$

(ii) $\exists X_1 \cdots \exists X_l \forall x_1 \cdots \forall x_m P$

(iii) $\forall X_1 \cdots \forall X_l \exists x_1 \cdots \exists x_m P$

(iv) $\forall X_1 \cdots \forall X_l \forall x_1 \cdots \forall x_m P$

where in each of the above L, m may be any non-negative integer, $X_1, \ldots, X_l$ are sets of elements, $x_1, \ldots, x_m$ are elements, and P is a *quantifier-free* formula over those variables (meaning that $\exists$ and $\forall$ are not used in P).

We note that M-logic is reasonably powerful: in particular, it is sufficient to define any minor-closed class with finitely many excluded minors (Corollary 3.3 in [20]). This includes many significant classes of matroids such as sparse paving matroids, and - due to the recent proof of Rota's Conjecture by Geelen, Gerards and Whittle [5] - the class of matroids representable over any given finite field.

Theorem 4.5.3. *The class of transversal matroids is not axiomatisable in M-logic*

Proof. To prove this we must show that there is no k-variable sentence that axiomatises transversal matroids. We shall construct two rank-r sparse paving matroids on n elements such that one is transversal and the other is not; we then demonstrate that no k-variable sentence in M-logic can distinguish between these.

Let $r > 2^{2^{2k}}$ and $n > 2r^2$ and let M, M' be rank-r sparse paving matroids on

groundset $[n]$, defined by their non-bases as follows:

$$\mathcal{L}(M) = \{1,\dots,r\}, \{r+1,\dots,2r\}, \dots, \{r^2 - r + 1, \dots, r^2\}$$

$$\mathcal{L}(M') = \mathcal{L}(M) \cup \{\{r^2 + 1, \dots, r^2 + r\}\}$$

We note that M' is non-transversal (since $|\mathcal{L}(M)| = r + 1$ and so $\beta(\emptyset) = -1$), and the tightening of M by any non-basis C, chosen from subsets of $\{r^2 + 1, \dots, 2r^2\}$ is isomorphic to M'. We also claim that M is transversal, and note that the relaxation of any non-basis in M' gives a matroid isomorphic to M.

(The transversality of M is easily established: there are r disjoint non-bases which we may label $L_1, \dots, LC_r$. A bipartite presentation of M is thus given by the graph with parts $E(M)$ and $\{y_1, \dots, y_r\}$ where for $x \in E(M)$ we have $x \sim y_i$ if and only if $x \notin L_i$.)

There are 4 cases to consider, corresponding to the 4 structures of sentences in M-logic. In each case we shall show that we can find matroids isomorphic to M and M' such that the truth of the sentence does not change, thereby demonstrating that M and M' cannot be distinguished between by a k-variable sentence in M-logic.

In proving that no case can axiomatise transversal matroids, we shall make use of *sigma algebras*.

Definition 4.5.4 (Sigma algebra)**.** Let $\mathcal{S} = \{S_1, \dots, S_n\}$ be a collection of sets. The *sigma algebra* $\Sigma(\mathcal{S})$ of $\mathcal{S}$ is the closure of $\mathcal{S}$ under operations of complement, intersection and union.

For a finite collection $\mathcal{S}$ we have

$$|\Sigma(\mathcal{S})| \leq 2^{2^{|\mathcal{S}|}}$$

To see this, let E be the set of elements contained in the union of all sets in $\mathcal{S}$. Partition the elements of E by the collection of sets which contain them. There are at most $2^{\{|\mathcal{S}|}$ nonempty parts in this partition. And members of the sigma algebra of $\mathcal{S}$ correspond precisely to subsets of this partition, giving at most $2^{2^{|\mathcal{S}|}}$ members.

Definition 4.5.5 (Hamming distance). Given a set E and subsets $S_1, S_2 \subseteq E$, the *Hamming distance* $d_H(S_1, S_2)$ between S_1 and S_2 is the size of the symmetric difference:

$$d_H(S_1, S_2) = |S_1 \Delta S_2|$$

Lemma 4.5.6. *Fix some integers k, m. Then for sufficiently large n, given any collection of k subsets $S_1, \ldots, S_k \subseteq [n]$, we can find some subset $X \subseteq [n]$ such that for any $1 \leq i \leq k$, we have $d_H(S_i, X) \geq m$.*

Proof. Draw X uniformly at random from subsets of n. There are 2^n possible choices of X; however if the symmetric difference $X \Delta S_i$ has size j, there are only $\binom{n}{j}$ choices for it. So as $n \to \infty$ we have (for any $0 \leq i \leq k$)

$$\Pr(d_H(X, S_i) < m) \leq \frac{1}{2^n} \sum_{j=1}^{m} \binom{n}{j} \to 0$$

But

$$\Pr(\exists i : d_H(X, S_i) < m) \leq k.\Pr(d_H(X, S_1) < m) \to 0$$

$\square$

Now we return to the four cases of the main proof:

Case (i)

Suppose we have axiomatised transversal matroids with a k variable sentence of form

$$\exists X_1 \cdots \exists X_l \exists x_1 \cdots \exists x_m P$$

with $l + m = k$ and P quantifier-free. Since M is transversal, we are able to find some $X_1, \ldots, X_l$ and $x_1, \ldots, x_m$ satisfying our quanitifier-free formula P. However our sentence can only make statements about the rank, and independence of sets in the sigma algebra of $\{X_1, \ldots, X_l, \{x_1\}, \ldots, \{x_m\}\}$. This sigma algebra has at most 2^{2^k} members. This means that provided n is sufficiently large we can certainly find some r-element subset $X \subseteq E(M)$ that is not in the sigma algebra - in fact we can further specify that $X \subseteq E(M) \setminus \bigcup_{L \in \mathcal{L}(M)} L$. If we now "tighten" M by defining such an X to be a non-basis, we get a matroid isomorphic to M', i.e. that is non-transversal. But since the rank and independence of every set other than X is unchanged, M' must also satisfies our formula. And so a formula of this kind cannot distinguish between M and M'.

Case (ii)

Suppose we have a k variable sentence of form

$$\exists X_1 \cdots \exists X_l \forall x_1 \cdots \forall x_m P$$

with $l + m = k$ and P quanitfier-free, that axiomatises transversal matroids.

Again assume we have satisfied the existence quantifiers, that is we have found $X_1, \ldots, X_l \subseteq E(M)$ that satisfy our formula. In this case we have to find an r-set X that is not only not in the sigma algebra of $\{X_1, \ldots, X_l\}$, but has Hamming distance at least $m + 1$ from any set in the sigma algebra - since P can only describe sets in the sigma algebra, perturbed by at most m elements $x_1, \ldots, x_m$ (see Proposition 5.2 in [20]). By Lemma 4.5.6 we can choose such an X, provided n is sufficiently large. Now we tighten M by defining X to be a non-basis. The matroid we arrive at is isomorphic to M' - without loss of generality suppose it is M'. But now our formula will still be satisfied for M', since the rank and independence has only changed in the set X, which cannot be described by P. Hence again we are unable to distinguish between M and M'.

Cases (iii) and (iv) are very similar to Cases (i) and (ii). In Case (iii) we start with M' and assume the converse of our formula holds - this reverses the quantifiers and from this point the argument is similar to Case (ii), except that we first relax by a non-basis, and in the second step tighten by a non-basis. Similar in Case (iv) the converse has the same pattern of quantifiers are Case (i), and again the proof only differs in that we first relax and then tighten. □

4.5.1 Axiomatisability of gammoids

We also note that the construction given by Mayhew, Newman and Whittle [20] shows gammoids to be non-axiomatisable. Essentially our work will

amount to showing that the representable matroids given by that construction are also gammoids, and so the argument will also show gammoids to be non-representable.

Theorem 4.5.7. *The class of gammoids is not axiomatisable in M-logic.*

The argument used by Mayhew, Newman and Whittle relies on describing three closely related matroids. The first is the *Kinser matroid*, $\text{Kin}(r)$, which is representable over infinite fields, when $r \geq 4$. The second, formed by relaxing a circuit-hyperplane of $\text{Kin}(r)$, is not representable over any field, and the third, formed by the further relaxation of a second hyperplane, is again representable over an infinite field. They then show that any sentence in M-logic will, for sufficiently large r, be unable to describe the first and third matroids without also describing the second, thereby proving that representability is not axiomatisable in M-logic. Clearly we can prove an equivalent result for the class of gammoids if we can show that the second matroid does not fall in this class, but the first and third do.

Proof. Clearly the second matroid described above, being not representable over any field, will also not be a gammoid. We thus only need to show that the first and third matroids are gammoids (which we shall in fact do by showing them to be minors of transversal matroids). Let's now describe those matroids and show they are both gammoids.

To define the *Kinser* matroid $\text{Kin}(r)$ we first define a rank-$r + 1$ matroid, M_{r+1}. In fact M_{r+1} is defined to be a transversal matroid. We define the elements of M_{r+1} to be partitioned into r pairwise disjoint sets $H_1, \ldots, H_r$, such that $H_1, \cdots H_r$ each have cardinality $r - 2$, and H_r has cardinality 2 (we shall say $H_r = \{e, f\}$). Now define the set system $\mathcal{A}_0, \mathcal{A}_1, \ldots, \mathcal{A}_r$ by

- $\mathcal{A}_0 = H_1 \cup \cdots \cup H_r = E(M_{r+1})$

- $\mathcal{A}_r = H_r$

- for $i \in \{1, \ldots, r-1\}$, $\mathcal{A}_i = E(M_{r+1}) - (H_r \cup H_{i-1} \cup H_i)$ where the subscript is calculated modulo $r - 1$.

We define $\mathrm{Kin}(r)$ to be the *truncation* of M_{r+1} - that is to say, $E(\mathrm{Kin}(r)) = E(M_{r+1})$ and $X \subseteq E(\mathrm{Kin}(r))$ is independent in $\mathrm{Kin}(r)$ if and only if $|X| \leq r$, and X is independent in M_{r+1}.

A truncation may simply be viewed as the addition of a free element followed by contracting by that element. The addition of a free element to a transversal matroid is clearly still transversal, and the contraction of any transversal matroid is a gammoid, therefore $\mathrm{Kin}(r)$ - as the truncation of our transversal matroid M_{k+1} - is a gammoid.

The third matroid described by Mayhew, Newman and Whittle is the matroid M formed from $\mathrm{Kin}(r)$ by relaxing the circuit-hyperplanes $(H_s \cup H_r)$ and $(H_{r-1} \cup H_r)$, where $s \in \{1, \ldots, r-2\}$. In order to show this is representable over an infinite field, they demonstrate that it can be constructed from M_{r+1} as follows: construct M' on the groundset $E(M_{r+1}) \setminus \{e, f\}) \cup \{p, q\}$ where p, q are distinct elements not in $E(M_{r+1})$. Let M' be the transversal matroid defined by sets $\mathcal{A}_0, \mathcal{A}'_1, \ldots, \mathcal{A}'_{r-1}$ where for $i \in \{1, \ldots, s\}$ we have $\mathcal{A}'_i = \mathcal{A}_i \cup p$, and for $i \in \{s+1, \ldots, r-1\}$ we have $\mathcal{A}'_i = \mathcal{A}_i \cup q$.

From M' we can (as described in the original paper) construct M by freely adding elements e, f to the line spanned by p and q, and then deleting p and q. This can actually be achieved by the following steps: firstly form M'' from M' by series extensions on the elements p and q, giving series pairs (p, p') and (q, q'). This is now a transversal matroid with p' and q' sitting at vertices of a

simplex presentation of M''. Now if we add e, f freely to the line between p' and q' this clearly forms a presentation of a new transversal matroid. Finally we can reach M by contracting down by p', q' and then deleting p and q. This makes M a minor of a transversal matroid, and therefore a gammoid. $\square$

Bibliography

[1] J. E. Blackburn, H. H. Crapo, and D. A. Higgs. A catalogue of combinatorial geometries. *Mathematics of Computation*, 27(121):155–s95, 1973.

[2] J. E. Bonin. An introduction to transversal matroids. 2010.

[3] H. H. Crapo and G.-C. Rota. On the foundations of combinatorial theory ii. combinatorial geometries. *Studies in Applied Mathematics.*, 49(2):109–133, 1970.

[4] J. Edmonds and D. Fulkerson. Transversals and matroid partition. *J. Res. Nat. Bur. Standards Sect. B*, (69B):147–153, 1965.

[5] J. Geelen, B. Gerards, and G. Whittle. Solving Rota's Conjecture. 61:736, 08 2014.

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

[7] P. Hall. On Representatives of Subsets. *Journal of the London Mathematical Society*, s1-10(1):26–30, 1935.

[8] A. W. Ingleton. *Transversal Matroids and Related Structures.*, pages 117–131. Springer Netherlands, 1977.

[9] J. L. W. V. Jensen. Sur les fonctions convexes et les inégalités entre les

valeurs moyennes. *Acta Math.*, 30:175–193, 1906.

[10] P. Keevash. The existence of designs., Jan. 2014.

[11] D. E. Knuth. The asymptotic number of geometries. *J. Comb. Theory, Ser. A*, 16(3):398–400, 1974.

[12] M. Las Vergnas. Sur les systèmes de représentants distincts d'une famille d'ensembles. *C. R. Acad. Sci. Paris Sér. A-B*, (270):A501–A503, 1970.

[13] L. Lowrance, J. Oxley, C. Semple, and D. Welsh. On properties of almost all matroids. *Advances in Applied Mathematics*, 50(1):115–124, 2013. Special issue in honor of Geoff Whittle on his 60th birthday.

[14] J. Mason. *A characterization of transversal independence spaces.*, volume 211 of *Springer.* Springer.

[15] J. H. Mason. On a class of matroids arising from paths in graphs. *Proceedings of the London Mathematical Society*, s3-25(1):55–74, 1972.

[16] D. Mayhew. The antichain of excluded minors for the class of gammoids is maximal., Feb. 2016.

[17] D. Mayhew, M. Newman, D. Welsh, and G. Whittle. On the asymptotic proportion of connected matroids. *Eur. J. Comb.*, 32(6):882–890, 2011.

[18] D. Mayhew, M. Newman, and G. Whittle. On excluded minors for real-representability. *J. Comb. Theory, Ser. B*, 99(4):685–689, 2009.

[19] D. Mayhew and G. F. Royle. Matroids with nine elements. *J. Comb. Theory, Ser. B*, 98(2):415–431, 2008.

[20] D. Mayhew, G. Whittle, and M. Newman. Is the missing axiom of matroid theory lost forever? *The Quarterly Journal of Mathematics*, 65(4):1397–1415, 2014.

[21] J. G. Oxley. *Matroid theory.* Oxford University Press, 1992.

[22] R. Pendavingh and J. van der Pol. Counting matroids by entropy. *Matroid Union weblog*, Apr. 2016.

[23] R. Pendavingh and J. van der Pol. On the number of bases of almost all matroids. *ArXiv e-prints*, Feb. 2016.

[24] R. Pendavingh and J. G. van der Pol. Counting matroids in minor-closed classes. *J. Comb. Theory, Ser. B*, 111:126–147, 2015.

[25] R. Pendavingh and J. G. van der Pol. Enumerating matroids of fixed rank. *Electr. J. Comb.*, 24(1):P1.8, 2017.

[26] M. J. Piff and D. J. A. Welsh. On the number of combinatorial geometries. *Bulletin of the London Mathematical Society*, 3(1):55–56, 1971.