2016

Random and deterministic versions of extremal poset problems

Kirsten Ann Hogenson
Iowa State University

Random and deterministic versions of extremal poset problems

by

Kirsten Ann Hogenson

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Mathematics

Program of Study Committee:

Ryan R. Martin, Major Professor

Steve Butler

Paul Sacks

Sung-Yell Song

Ananda Weerasinghe

Iowa State University

Ames, Iowa

2016

DEDICATION

To my parents, Dave and Paula, for your wisdom and encouragement.

And to Todd, for your love and support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor, Dr. Ryan R. Martin, for his guidance, insight, and patience throughout the completion of this work. I would also like to thank all of the fellow graduate students and collaborators I have met during my time at Iowa State for their unwavering friendship and encouragement.

Additionally, I would like to thank Melanie Erickson, Ellen Olson, and all of the other office staff for absolutely everything they do for the department.

ABSTRACT

Let $\mathcal{P}(n)$ denote the set of all subsets of $\{1, \ldots, n\}$ and let $\mathcal{P}(n, p)$ be the set obtained from $\mathcal{P}(n)$ by selecting elements independently at random with probability p. The Boolean lattice is a partially ordered set, or poset, consisting of the elements of $\mathcal{P}(n)$, partially ordered by set inclusion. A basic question in extremal poset theory asks the following: Given a poset P, how big is the largest family of sets in the Boolean lattice which does not contain the structure P as a subposet? The following random analogue of this question is also of interest: Given a poset P, how big is the largest family of sets in $\mathcal{P}(n, p)$ which does not contain the structure P as a subposet? In this thesis, we present new proofs for a collection of deterministic extremal subposet problems. We also discuss a new technique called the Hypergraph Container Method in depth and use it to prove a random version of De Bonis and Katona's $(r + 1)$-fork-free theorem.

CHAPTER 1. INTRODUCTION

A partially ordered set, or *poset*, is a pair $P = (X, \preceq)$ where X is a set and $\preceq$ is a binary relation on X that is reflexive, antisymmetric, and transitive. The relation $\preceq$ is commonly referred to as a *partial ordering*. For elements $a, b \in X$, we say that b *covers* a if $a \preceq b$ and there is no element $c \in X$ such that $a \preceq c \preceq b$. In this thesis, the set X will be assumed to be finite, though in general this restriction is not required.

$P' = (X', \preceq')$ is a *weak subposet* of $P = (X, \preceq)$ if there exists an injection of X' into X which preserves the relationships in P'. More formally, there is an injective function $\varphi : X' \to X$ such that if $a, b \in X'$, $a \preceq' b$ then $\varphi(a) \preceq \varphi(b)$. Note that this definition allows there to be additional relationships in $(\varphi(X'), \preceq)$ which are not in $(X', \preceq')$. If we add the restriction that $a \not\preceq' b$ implies $\varphi(a) \not\preceq \varphi(b)$, then we say that P' is a *strong*, or *induced*, subposet of P. All of the subposet relationships discussed in this thesis should be assumed to be weak unless otherwise specified.

A simple *graph* G is a pair (V, E) where V is a set of *vertices* and E is a set of two element subsets of V called *edges*. We will typically denote the number of vertices as $|V|$ and the number of edges as $|E|$ or $e(G)$. For a vertex $v \in V$, we let $N_G(v)$ denote the *neighborhood* of v in G, which is the set $\{u \in V \mid \{u, v\} \in E\}$. The *degree* of a vertex v is the size of its neighborhood in G and is denoted $\deg_G(v)$. If it is clear to which graph we are referring, we will simplify the notation to $N(v)$ and $\deg(v)$, respectively.

A *directed graph*, or *digraph*, is a graph where the edge set E is replaced with a set A of ordered pairs of vertices called *arcs*. In the case of digraphs, we will be concerned with both the *in-neighborhood* and the *out-neighborhood* of a particular vertex. These neighborhoods are defined as

$$N_G^-(v) := \{u \in V \mid (u,v) \in A\} \quad \text{and} \quad N_G^+(v) := \{u \in V \mid (v,u) \in A\},$$

respectively. Similarly, the *in-degree* and *out-degree* are defined as

$$\deg_G^-(v) := |N_G^-(v)| \quad \text{and} \quad \deg_G^+(v) := |N_G^+(v)|,$$

respectively. As with undirected graphs, we will neglect the subscript when it is clear to which digraph we are referring.

Graphs are an extremely useful way to visually represent relationships in discrete mathematics, and they can be used to diagram posets in a number of ways.

For a given poset $P = (X, \preceq)$, the *comparability digraph* of P is a digraph G with $V = X$ and $(a,b) \in A$ if and only if $a, b \in X$, $a \preceq b$, $a \neq b$. The undirected underlying graph of G is known as the *comparability graph* of P. Though these graphs have many uses, they can be cumbersome to examine visually if $\preceq$ contains many relationships. For this reason, posets are often presented in Hasse diagrams. A *Hasse diagram* is a transitive reduction of a typical comparability graph, where the poset elements are drawn in the plane as vertices, and an edge goes upward from a to b if b covers a. Figure 1.1 shows these three different visualizations for the example poset $P = (\{A, B, C, D\}, \{A \preceq B, A \preceq C, A \preceq D, C \preceq D\})$.

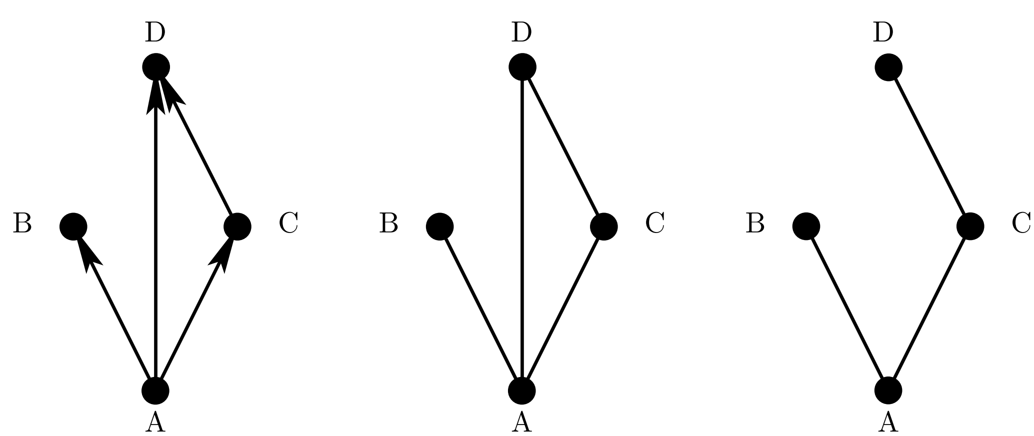

Figure 1.1 From left to right: The comparability digraph, comparability graph, and Hasse diagram for the poset P.

Let $[n] = \{1, 2, \ldots, n\}$, the set of integers 1 through n, and let $\mathcal{P}(n)$ denote the set of all subsets of $[n]$. $\mathcal{P}(n)$ is referred to as the *power set* of $[n]$.

One commonly studied poset structure is the *Boolean lattice*, which is the poset on $\mathcal{P}(n)$, partially ordered by set inclusion. Formally, we may write $\mathcal{B}_n = (\mathcal{P}(n), \subseteq)$. The Boolean lattice can be diagrammed as pictured in Figure 1.2, where it is arranged into layers. The top layer consists of the set $[n]$, the bottom layer consists of the empty set $\emptyset$, and the k^{th} layer consists of all subsets of $[n]$ of size k. The family of all sets in the k^{th} layer is often denoted as $\binom{[n]}{k}$, as it has size $\binom{n}{k}$. Similarly, the family of all sets of size at most k is $\binom{[n]}{\leq k}$, and it has size $\binom{n}{\leq k} = \sum_{0 \leq i \leq k} \binom{n}{i}$. Depending on the parity of n, the Boolean lattice will have either one or two middle layers, at least one of which is $\binom{[n]}{\lfloor n/2 \rfloor}$. In this thesis the size of one middle layer will be denoted as $M = \binom{n}{\lfloor n/2 \rfloor}$, and the total size of k middle layers will be denoted as M_k.

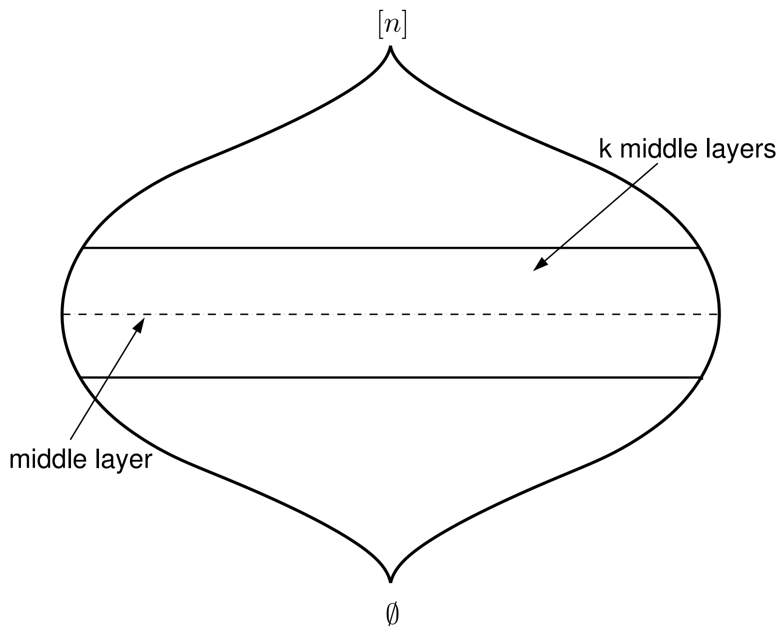

Figure 1.2 The Boolean lattice $\mathcal{B}_n$.

1.1 Forbidden subposet problems

A basic question in extremal poset theory asks the following: Given a poset P, how big is the largest family $\mathcal{F}$ of sets in the Boolean lattice $\mathcal{B}(n)$ such that $(\mathcal{F}, \subseteq)$ does *not* contain

P as a subposet? The maximum size of such a family is denoted $La(n, P)$, where n denotes the dimension of the Boolean lattice, and P denotes the forbidden subposet. This problem has been answered for a number of subposets P, either exactly or asymptotically, but in general is not easy to compute. We will discuss some of the noteworthy known results here.

Many of these results employ "Big Oh" notation, which is defined in the following way: For two sequences a_n and b_n, we say that a_n is "Big Oh" of b_n, or $a_n = O(b_n)$, if as $n \to \infty$, there exist constants C and n_0 such that $|a_n| \leq Cb_n$ for $n \geq n_0$. This means that the *order of magnitude* of the sequence a_n is at most the order of magnitude of b_n. Equivalently, we can use "Big Omega" notation to describe the same relationship between a_n and b_n. Specifically, b_n is "Big Omega" of a_n, or $b_n = \Omega(a_n)$.

The 2-chain poset, denoted $\mathcal{P}_2$, consists of two related elements. A generalization of the 2-chain poset is the k-chain, denoted $\mathcal{P}_k$, which consists of k elements that are pairwise related. Hasse diagrams for the 2-chain and the k-chain can be seen in Figure 1.3. Notice that these Hasse diagrams look like path graphs on 2 and k vertices, respectively. Because of this, the notation used for these chain posets mimics the graph theory notation for path graphs. The *height* of a poset is the minimum number of layers of the Boolean lattice that it must inhabit. Notice that the height of $\mathcal{P}_2$ is 2, because one set cannot be contained in a different set of the same size. By extending this argument, we can see that the height of $\mathcal{P}_k$ is k.

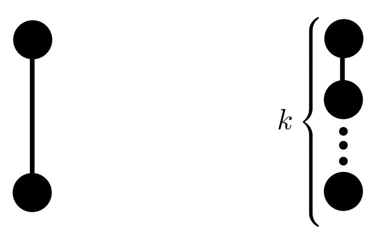

Figure 1.3 From left to right: The 2-chain $\mathcal{P}_2$ and the k-chain $\mathcal{P}_k$.

A family of sets which does not contain $\mathcal{P}_2$ as a subposet is referred to as an *antichain.* One famous result from Sperner [51] states that the size of the largest antichain in $\mathcal{B}_n$ is M, the size of a middle layer. We can see that every layer of the Boolean lattice is an antichain, due to the fact that $\mathcal{P}_2$ is a height 2 poset. Therefore, one example of a maximum-size antichain is a single middle layer of the Boolean lattice. Sperner's result was generalized by Erdős [21], who showed that a k-chain-free family in $\mathcal{B}_n$ is no larger than M_{k-1}, the size of $k-1$ middle layers. To see this, notice that $\mathcal{P}_k$ is a height k poset. Thus, it is impossible to obtain a k-chain as a subposet of a family which exists in $k-1$ layers. The largest $k-1$ layers of the Boolean lattice lie in the middle, meaning one example of a k-chain-free family consists of all $k-1$ middle layers. These results are summarized below in Theorems 1.1.1 and 1.1.2. Note that if we set $k=2$, these results are identical.

Theorem 1.1.1 (Sperner [51])**.** $La(n, \mathcal{P}_2) = M$

Theorem 1.1.2 (Erdős [21])**.** $La(n, \mathcal{P}_k) = M_{k-1}$

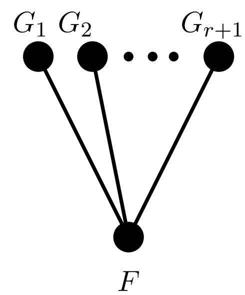

Figure 1.4 The $(r+1)$-fork $\mathcal{V}_{r+1}$.

The $(r+1)$-fork poset, denoted $\mathcal{V}_{r+1}$, is the family of distinct sets $F, G_1, \ldots, G_{r+1}$ such that $F \subset G_i$ for all i. A Hasse diagram for the $(r+1)$-fork can be seen in Figure 1.4. Notice that I have chosen to represent the $(r+1)$-fork with no relationships amongst the G_i sets. In order for the $(r+1)$-fork to be present as a weak subposet, only the pictured relationships are required, and the relationships between the G_i sets may or may not be present. This drawing convention will be used for each of the following posets discussed in this section. Theorem 1.1.3, which was originally proved by De Bonis and Katona in [18], gives an upper bound for the largest $(r+1)$-fork-free family in $\mathcal{B}_n$. It is worth noting that for large n, this upper bound approaches M. This means that, asymptotically, $(r+1)$-fork-free families behave similarly to antichains.

Theorem 1.1.3 (De Bonis and Katona [18])**.** $La(n, \mathcal{V}_{r+1}) \leq \left(1 + \frac{2r}{n} + O\left(\frac{1}{n^2}\right)\right) M.$

The butterfly poset, denoted $\bowtie$, is the family of four distinct sets A, B, C, D such that $(A \cup B) \subset (C \cap D)$. A Hasse diagram for the butterfly can be seen in Figure 1.5. The upper

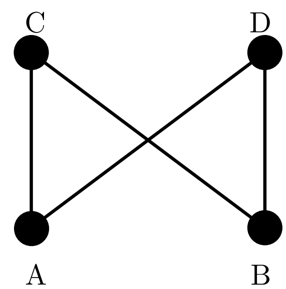

Figure 1.5 The butterfly $\bowtie$.

bound presented in Theorem 1.1.4 was originally proved in [19] by De Bonis, Katona, and Swanepoel. In fact, they used the cyclic permutation method to show that $La(n, \bowtie) = M_2$. That is, the extremal butterfly-free family consists of exactly two middle layers of the Boolean lattice. Griggs and Li used the partition method to reprove this result in [26]. To see that this extremal family makes sense, note that the butterfly structure can never appear in two consecutive layers of the Boolean lattice i.e. $A, B \in \binom{[n]}{k}$ and $C, D \in \binom{[n]}{k+1}$. This is because the symmetric difference of C and D would be at least two 2, forcing $|C \cap D| \leq k$. So there is either a unique set of size k in $C \cap D$ or $|C \cap D| < k$. Therefore, two consecutive middle layers of $\mathcal{B}_n$ form a maximum butterfly-free family.

Theorem 1.1.4 (De Bonis, Katona, and Swanepoel [19])**.** $La(n, \bowtie) \leq 2M$.

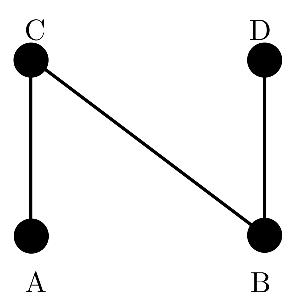

Figure 1.6 The $\mathcal{N}$ poset.

The $\mathcal{N}$ poset is the family of four distinct sets A, B, C, D such that $A \subset C$ and $B \subset (C \cap D)$. A Hasse diagram for $\mathcal{N}$ can be seen in Figure 1.6. Theorem 1.1.5 was originally proved by Griggs and Katona in [25] using a linear programming technique. They found that a maximum $\mathcal{N}$-free family is asymptotically about the same size as a maximum $\mathcal{V}_{r+1}$-free family.

Theorem 1.1.5 (Griggs and Katona [25])**.** $La(n, \mathcal{N}) \le \left(1 + \frac{2}{n} + O\left(\frac{1}{n^2}\right)\right) M.$

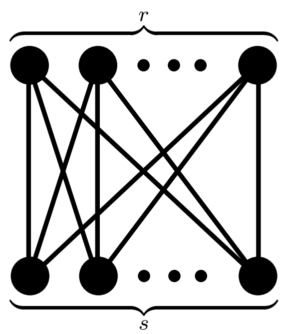

Figure 1.7 The $\mathcal{K}_{r,s}$ poset.

Let $r, s \ge 2$ with at least one of r and s greater than 2. Let $\mathcal{K}_{r,s}$ denote the family of $r+s$ distinct sets $A_1, \ldots, A_s$ and $B_1, \ldots, B_r$ which satisfy $A_i \subset B_j$ for all $1 \le i \le s$, $1 \le j \le r$. A Hasse diagram for $\mathcal{K}_{r,s}$ is pictured in Figure 1.7. Notice that the $\mathcal{K}_{r,s}$ poset's Hasse diagram looks like a complete bipartite graph, and as such the notation for this poset mimics graph theory notation. The forbidden subposet result for $\mathcal{K}_{r,s}$ is presented in Theorem 1.1.6 and was originally proved by De Bonis and Katona [18].

Theorem 1.1.6 (De Bonis and Katona [18])**.** $La(n, \mathcal{K}_{r,s}) \le \left(2 + 2\left(\frac{r+s-3}{n}\right) + O\left(\frac{1}{n^2}\right)\right) M.$

Notice that the $\mathcal{K}_{r,s}$ is one possible generalization of the butterfly poset, which is the excluded case $r = s = 2$. (Another generalization of $\bowtie$ will be discussed in Section 2.1.5.) If we set $r = s = 2$ in Theorem 1.1.6, we get a result that is asymptotically equivalent to the upper bound in Theorem 1.1.4. However, since an exact result is known for the butterfly, we keep that result separate from this weaker asymptotic upper bound.

Let $s, t \ge 1$ and $k \ge 3$. Then $\mathcal{P}_k(s, t)$ denotes the broom – a height k poset with s minimal elements and t maximal elements. Alternately, the broom can be thought of as a path on k elements in which the minimal element is replicated $s-1$ times and the maximal element is replicated $t-1$ times. A Hasse diagram for $\mathcal{P}_k(s, t)$ is pictured in Figure 1.8. Brooms

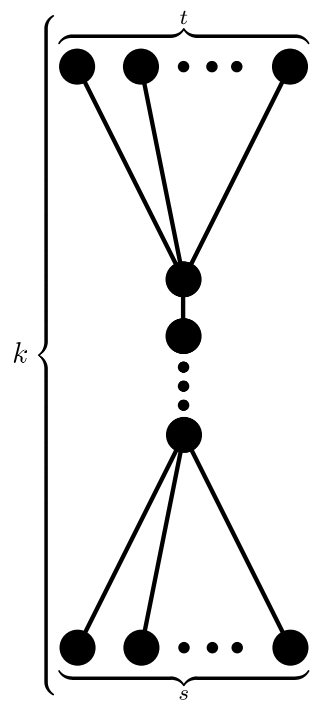

Figure 1.8 The broom $\mathcal{P}_k(s,t)$.

were introduced by Griggs and Lu in [28]. They referred to brooms as batons and used chain counting and probabilistic techniques to prove the result in Theorem 1.1.7.

Theorem 1.1.7 (Griggs and Lu [28]). $La(n, \mathcal{P}_k(s,t)) \leq \left(k - 1 + O\left(\frac{1}{n}\right)\right) M.$

Notice that brooms behave similarly to k-chains. This is a consequence of the fact that a $\mathcal{P}_k(s,t)$ is like a $(k-2)$-chain with a $\mathcal{V}_t$ glued onto the top and a $\mathcal{V}_s$ glued onto the bottom. We saw in Theorem 1.1.3 that forks behave asymptotically like antichains, so it makes sense that a $\mathcal{P}_k(s,t)$ would behave asymptotically like a k-chain. Further, notice that setting $k=3$ yields a broom which contains $\mathcal{K}_{s,t}$ as a subposet. Thus, unsurprisingly, we get the same asymptotic upper bound for $La(n, \mathcal{P}_3(s,t))$ as we do for $La(n, \mathcal{K}_{s,t})$.

This is just a sampling of the forbidden subposet problems that have been solved. There are many additional posets which have been extensively studied, and occasionally reproofs of existing results are published to demonstrate the power of new techniques. In fact, I will present new reproofs of four of these results in Chapter 2.

1.2 Random forbidden subposet problems

Let $\mathcal{P}(n,p)$ be the set obtained from $\mathcal{P}(n)$ by selecting elements independently at random with probability p. We use the notation $b_n \gg a_n$ if $a_n \geq 0$ and $a_n = o(b_n)$. Further, this "Little Oh" notation indicates that as $n \to \infty$, $a_n/b_n \to 0$. We say that an event $\mathcal{E}_n$ happens "with high probability" if the probability of that event occurring approaches 1 as n tends to infinity.

Recently, there has been interest in random versions of known forbidden subposet theorems. In particular, a random version of Sperner's Theorem (Theorem 1.1.1) was proved independently by Balogh, Mycroft, and Treglown [9] and by Collares Neto and Morris [16]. In fact, Collares Neto and Morris proved a random version of Erdős's Theorem (Theorem 1.1.2), which yields Sperner's Theorem for $k = 2$. These results are presented below in Theorems 1.2.1 and 1.2.2.

Theorem 1.2.1 (Balogh, Mycroft, and Treglown [9]). *If $p \gg 1/n$, then with high probability, the largest antichain in $\mathcal{P}(n,p)$ has size at most $(1 + o(1))pM$.*

Notice that this upper bound is approximately pM, which is the expected size of a middle layer in $\mathcal{P}(n,p)$. Further, if we set $p = 1$, this result is identical to the statement of Theorem 1.1.1. The random version of Erdős's Theorem yields a similar result.

Theorem 1.2.2 (Collares Neto and Morris [16]). *If $p \gg 1/n$, then with high probability, the largest $\mathcal{P}_k$-free family in $\mathcal{P}(n,p)$ has size at most $(k - 1 + o(1))pM$.*

In this case, the upper bound is approximately $p(k-1)M \approx pM_{k-1}$, which is the expected size of $k-1$ middle layers of $\mathcal{P}(n,p)$. Again, setting $p = 1$ gives us the deterministic result from Theorem 1.1.2.

Both of these random results use the 'hypergraph container' method, which was independently developed by Balogh, Morris, and Samotij in [8] and by Saxton and Thomason in [49]. Though their techniques are very similar, it is interesting to note that the proof methods for these two random results are in fact different. Balogh, Mycroft, and Treglown [9] used the proof method from [8], meaning they encoded the poset in a simple graph and constructed containers with a small number of vertices. Meanwhile, Collares Neto and Morris [16] used the proof technique from [49], meaning they encoded the poset in a hypergraph and constructed containers with a small number of edges. This method, along with some additional relevant definitions, is discussed further in Chapter 3.

Lower bounds for the random Sperner and Erdős theorems are also known. In fact, they follow from a result of Osthus [44] and are presented below in Theorem 1.2.3 and its corollary, Corollary 1.2.4. Notice that in both cases, we can add approximately one additional layer of $\mathcal{P}(n,p)$ without encountering the forbidden subposet.

Theorem 1.2.3 (Osthus [44])**.** *If $p \ll 1/n$, then with high probability, the largest antichain in $\mathcal{P}(n,p)$ has size at least $(2-o(1))pM$.*

Corollary 1.2.4. *If $p \ll 1/n$, then with high probability, the largest $\mathcal{P}_k$-free family in $\mathcal{P}(n,p)$ has size at least $(k-o(1))pM$.*

A quick proof of Corollary 1.2.4 is presented below.

Proof of Corollary 1.2.4. Using Theorem 1.2.3, note that, with high probability, we may begin with an antichain of size at least $(2-o(1))pM$. To that antichain, we may add up to $k-2$ layers of $\mathcal{P}(n,p)$ without creating a k-chain. Thus, with high probability, we have a k-chain free family of size at least

$$(2-o(1))pM + (k-2)pM = (k-o(1))pM,$$

as desired. □

The bounds for the size of maximum chain-free families in $\mathcal{P}(n,p)$ depend on the value of p. In particular, the tipping point for this radical change occurs at $p = 1/n$. Therefore, we can call $\hat{p}(n) = 1/n$ a *threshold* for this behavior. As noted in *Random Graphs* [32], a threshold is not unique since all functions of the same order will be thresholds as well. However, for convenience, it is still customary to refer to *the* threshold.

One of my main results, presented in Theorem 1.2.5, is a random version of De Bonis and Katona's $\mathcal{V}_{r+1}$-free theorem. It is accompanied by a lower bound theorem, Theorem 1.2.6, which is again a corollary of the Osthus result. Notice that, as in the deterministic case, the $(r+1)$-fork-free families behave similarly to antichains.

Theorem 1.2.5 (The Random $\mathcal{V}_{r+1}$-Free Theorem)**.** *If $p \gg 1/n$, then with high probability, the largest $\mathcal{V}_{r+1}$-free family in $\mathcal{P}(n,p)$ has size at most $(1+o(1))pM$.*

Theorem 1.2.6. *If $p \ll 1/n$, then with high probability, the largest $\mathcal{V}_{r+1}$-free family in $\mathcal{P}(n,p)$ has size at least $(2-o(1))pM$.*

Again, the behavior of these random poset-free families depends on the value of p, and the threshold for p is presented in Corollary 1.2.7. It is interesting to note that this threshold is the same as the threshold for the largest antichain in $\mathcal{P}(n,p)$ having size at most $(1+o(1))pM$ yielded by Balogh, Mycroft, and Treglown [9] and by Collares Neto and Morris [16].

Corollary 1.2.7. *The threshold for the largest $\mathcal{V}_{r+1}$-free family in $\mathcal{P}(n,p)$ having size at most $(1+o(1))pM$ is $\hat{p}(n) = 1/n$.*

Detailed proofs for Theorems 1.2.5 and 1.2.6 and some additional discussion of threshold functions can be found in Chapter 4.

1.3 Organization of thesis

In Chapter 2, we discuss deterministic extremal poset results. First, Section 2.1 revisits four known forbidden subposet problems, and we reprove them using a supersaturation result from Kleitman. Subsection 2.1.1 covers the $(r+1)$-fork, Subsection 2.1.2 covers the $\mathcal{N}$ poset, Subsection 2.1.3 covers the $\mathcal{K}_{r,s}$ poset, and Subsection 2.1.4 covers brooms. We discuss the limitations of this technique in Subsection 2.1.5, where we give an example of one poset family for which it does not work. Next, in Section 2.2, we discuss the notorious diamond-free subposet problem. In particular, we present the best current upper and lower bounds on the size of a maximum diamond-free family in $\mathcal{B}_n$.

In Chapter 3, we discuss the hypergraph container method in more detail. First, Section 3.1 covers the technique's origins and inspirations. Then Section 3.2 describes how to apply this method. In particular, a container-building algorithm is presented and the role of supersaturation results is mentioned. A short exploration of various results which have been proved using this technique is available in Section 3.3. The topics discussed include list colorings

(Subsection 3.3.1), arithmetic progressions (Subsection 3.3.2), sum-free sets (Subsection 3.3.3), intersecting families (Subsection 3.3.4), and union-free families (Subsection 3.3.5).

In Chapter 4, we provide a full proof for Theorem 1.2.5. We begin with a proof of Theorem 1.2.6, the lower bound theorem, in Section 4.1. Then in Section 4.2, we prove our main piece of machinery, the $\mathcal{V}_{r+1}$-Free Container Lemma. This includes a thorough description of how to build the container given a $\mathcal{V}_{r+1}$-free family I, as well as a proof that the container algorithm yields well-defined functions which describe all containers. Once the necessary lemma is established, we present the proof of Theorem 1.2.5 in Section 4.3. Finally, in Section 4.4, we discuss future work and present some open questions.

CHAPTER 2. DETERMINISTIC POSET RESULTS

There are many small posets P for which the question of calculating $La(n, P)$ has been asked. In some cases, techniques like chain counting and the partition method have yielded either exact or asymptotic values for $La(n, P)$. But in others, only bounds for $La(n, P)$ are known. In this chapter, we will investigate both situations. First, we will revisit and reprove optimal upper bounds of $La(n, P)$ for four posets. Then, we will discuss the smallest poset for which an optimal upper bound has not been proved: the notorious diamond $\Diamond$.

2.1 Reproofs using a supersaturation result of Kleitman

As new proof techniques are discovered and developed, it is not unusual to see reproofs of known results. These reproofs can offer new mathematical insights and a deeper understanding of the theorems with which they are concerned. They can also showcase the power of new techniques, or expose new applications of long known facts. In this section, we will see four new proofs of known extremal subposet problems that employ an old theorem from Daniel Kleitman. His result concerned what he referred to as "commensurable pairs" and was presented as Theorem I in [34]. A reworded version of this theorem is presented below in Theorem 2.1.1.

Theorem 2.1.1 (Kleitman [34]). *If a family $\mathcal{F}$ of subsets of an n-set has $\binom{n}{\lfloor n/2 \rfloor} + x$ members, there must be at least $(\lfloor n/2 \rfloor + 1)\, x$ distinct pairs (A, B) of members of $\mathcal{F}$ satisfying $A \subset B$, $A \neq B$.*

If we explore the meaning of this result in the context of posets, it says the following: If a family of sets in the Boolean lattice has $M + x$ members, then it must contain at least $(\lfloor n/2 \rfloor + 1)\, x$ distinct pairs of related sets. This sort of theorem is often referred to as a 'supersaturation' result, because it implies that adding just a few (x many) elements to a maximum antichain yields numerous (at least $(\lfloor n/2\rfloor + 1)\,x$ many) 2-chains.

Recall that for a given family $\mathcal{F} \subseteq \mathcal{B}_n$, the comparability digraph of $\mathcal{F}$ is a digraph G with $V = \mathcal{F}$ and $(S, T) \in A$ if and only if $S, T \in \mathcal{F}$, $S \subset T$, $S \neq T$. The subgraph of G *induced* by a set of vertices $V_1 \subset V$ is the digraph with vertex set V_1 and arc set $A_1 = \{(u, v) \mid u, v \in V_1 \text{ and } (u, v) \in A\}$. This induced subgraph is denoted as $G[V_1]$. It follows immediately from Theorem 2.1.1 that the comparability digraph of a family $\mathcal{F}$ with $|\mathcal{F}| = M + x$ contains at least $(\lfloor n/2\rfloor + 1)\,x$ arcs. This fact will be instrumental in the following four reproofs.

2.1.1 The $(r + 1)$-fork poset

Recall that the $(r + 1)$-fork $\mathcal{V}_{r+1}$ is the family of distinct sets $F, G_1, \ldots, G_{r+1}$ such that $F \subset G_i$ for all i. Theorem 1.1.3 states De Bonis and Katona's result that the upper bound for the size of a maximum $(r + 1)$-fork-free family in $\mathcal{B}_n$ approaches M as n approaches infinity. They proved their result using chain-counting and the famous LYMB-inequality [38; 55; 41; 12]. Griggs and Li [26] later reproved it, albeit with a worse error term, using the partition method. Our new proof uses Theorem 2.1.1 and yields the better error term.

New Proof of Theorem 1.1.3. Let $G = (V, A)$ be the comparability digraph of a $\mathcal{V}_{r+1}$-free family $\mathcal{F}$. Since G has no $(r + 1)$-fork, the maximum outdegree of a vertex in G is r. The number of arcs in a digraph can be computed by $|A| = \sum_{v\in V} \deg^+(v) = \sum_{v\in V} \deg^-(v)$, so this implies that $|A| \leq r|V|$.

Set $x = |V| - M$ and note that $|\mathcal{F}| = M + x = |V|$. By Theorem 2.1.1, the number of arcs induced by $\mathcal{F}$ is at least $(\lfloor n/2\rfloor + 1)\,x$. This means $(\lfloor n/2\rfloor + 1)\,(|V| - M) \leq r|V|$, which implies

$$
\begin{aligned}
|V| &\leq \left(\frac{\lfloor n/2\rfloor + 1}{\lfloor n/2\rfloor - r + 1}\right) M \\
&= \left(1 + \frac{2r}{n} + \left(\frac{r}{\lfloor n/2\rfloor - r + 1} - \frac{2r}{n}\right)\right) M \\
&= \left(1 + \frac{2r}{n} + O\left(\frac{1}{n^2}\right)\right) M,
\end{aligned}
$$

as desired. □

Note that if $r = o(n)$, we can replace the $O\left(1/n^2\right)$ term in the final upper bound with $O\left(r^2/n^2\right)$.

2.1.2 The $\mathcal{N}$ poset

Recall that $\mathcal{N}$ is the family of four distinct sets A, B, C, and D which satisfy $A \subset C$, $B \subset (C \cap D)$. Theorem 1.1.5 states Griggs and Katona's result that the upper bound for the size of a maximum $\mathcal{N}$-free family in $\mathcal{B}_n$ approaches M as n tends to infinity. Griggs and Li later used the partition method to reprove it with a weaker error term in [26]. We apply Theorem 2.1.1 to reprove Theorem 1.1.5, and our new proof yields the better error term.

New Proof of Theorem 1.1.5. Let $G = (V, A)$ be the comparability digraph of an $\mathcal{N}$-free family $\mathcal{F}$. Let $\mathcal{S} = \{F \in \mathcal{F} \mid \deg^-(F) = 0\}$, $\mathcal{U} = \{F \in \mathcal{F} \mid \deg^+(F) = 0\}$, and $\mathcal{T} = \{F \in \mathcal{F} \mid \deg^-(F) > 0 \text{ and } \deg^+(F) > 0\}$. Note that $\mathcal{T}$ is an antichain, because an arc in $G[\mathcal{T}]$ would imply a directed P_4 in G which in turn implies an $\mathcal{N}$ in $\mathcal{F}$.

Because G is $\mathcal{N}$-free, notice that for any $F \in \mathcal{T}$, F is the center vertex in a directed P_3 from $\mathcal{S}$ to $\mathcal{U}$. This corresponds to a transitive K_3 in G. Remove all K_3-s from G and call the resulting graph G'. Let $\mathcal{S}' = G' \cap \mathcal{S}$, $\mathcal{T}' = G' \cap \mathcal{T}$, and $\mathcal{U}' = G' \cap \mathcal{U}$. Note that $\mathcal{T}' = \emptyset$, $|\mathcal{T}| = |\mathcal{S} \setminus \mathcal{S}'| = |\mathcal{U} \setminus \mathcal{U}'|$, and the graph G' consists of disjoint stars.

Thus, the total number of arcs in G is bounded above by

$$|\mathcal{U}'| + |\mathcal{S}'| + 3|\mathcal{T}| = |\mathcal{U}| + |\mathcal{S}| + |\mathcal{T}| \leq |V|.$$

Set $x = |V| - M$ and note that $|\mathcal{F}| = M + x = |V|$. By Theorem 2.1.1, the number of arcs induced by $\mathcal{F}$ is at least $(\lfloor n/2 \rfloor + 1)\, x$. This means

$$(\lfloor n/2 \rfloor + 1)\,(|V| - M) \leq |V|,$$

which implies

$$\begin{aligned} |V| &\leq \left(\frac{\lfloor n/2 \rfloor + 1}{\lfloor n/2 \rfloor}\right) M \\ &= \left(1 + \frac{2}{n} + \left(\frac{1}{\lfloor n/2 \rfloor} - \frac{2}{n}\right)\right) M \\ &= \left(1 + \frac{2}{n} + O\left(\frac{1}{n^2}\right)\right) M, \end{aligned}$$

as desired. $\square$

2.1.3 The $\mathcal{K}_{r,s}$ poset

Let $r, s \geq 2$ with at least one of r and s greater than 2. Recall that $\mathcal{K}_{r,s}$ is the family of $r+s$ distinct sets $A_1, \ldots, A_s$ and $B_1, \ldots, B_r$ which satisfy $A_i \subset B_j$ for all $1 \leq i \leq s$, $1 \leq j \leq r$. Theorem 1.1.6 states De Bonis and Katona's result that the upper bound for the maximum size of a $\mathcal{K}_{r,s}$-free family in $\mathcal{B}_n$ approaches $2M$ as n tends toward infinity. We reprove it here using Theorem 2.1.1.

New Proof of Theorem 1.1.6. Let $G = (V, A)$ be the comparability digraph of a $\mathcal{K}_{r,s}$-free family $\mathcal{F}$ for $r, s \geq 2$ and at least one of $r, s \geq 3$. Define $\mathcal{S} = \{F \in \mathcal{F} \mid \deg^+(F) \geq r\}$, $\mathcal{U} = \{F \in \mathcal{F} \mid \deg^-(F) \geq s\}$, and $\mathcal{T} = \{F \in \mathcal{F} \mid \deg^+(F) < r \text{ and } \deg^-(F) < s\}$. Note that in $\mathcal{S}$, $\deg^+(F) \geq r$ implies $\deg^-(F) < s$, otherwise we would get a configuration in G which corresponds to a $\mathcal{K}_{r,s}$ in $\mathcal{F}$. Similarly, in $\mathcal{U}$, $\deg^-(F) \geq s$ implies $\deg^+(F) < r$.

We will transform G into the digraph G' by removing all arcs going out of $\mathcal{U}$, all arcs going into $\mathcal{S}$, and all arcs contained within $\mathcal{T}$. Notice that we have removed at most $(r+s)|V(G)|$ arcs from G. Consider $G'[\mathcal{S} \cup \mathcal{T}]$. Due to the degree conditions, this induced subgraph contains at most $s|\mathcal{S} \cup \mathcal{T}|$ arcs.

Set $x = |\mathcal{S} \cup \mathcal{T}| - M$. By Theorem 2.1.1, the number of arcs induced by $\mathcal{S} \cup \mathcal{T}$ is at least $(\lfloor n/2 \rfloor + 1)\, x$. This means

$$(\lfloor n/2 \rfloor + 1)\,(|\mathcal{S} \cup \mathcal{T}| - M) \leq s|\mathcal{S} \cup \mathcal{T}|,$$

which implies

$$\begin{aligned} |\mathcal{S} \cup \mathcal{T}| &\leq \left(\frac{\lfloor n/2 \rfloor + 1}{\lfloor n/2 \rfloor - s + 1} \right) M \\ &= \left(1 + \frac{s}{\lfloor n/2 \rfloor - s + 1} \right) M. \end{aligned}$$

A similar argument will show

$$|\mathcal{U} \cup \mathcal{T}| \leq \left(1 + \frac{r}{\lfloor n/2 \rfloor - r + 1} \right) M.$$

Thus,

$$
\begin{aligned}
|\mathcal{F}| &\le |\mathcal{S} \cup \mathcal{T}| + |\mathcal{U} \cup \mathcal{T}| \\
&\le \left(2 + 2\left(\frac{r+s-3}{n}\right) + \frac{s}{\lfloor n/2 \rfloor - s + 1} + \frac{r}{\lfloor n/2 \rfloor - r + 1} - 2\left(\frac{r+s-3}{n}\right)\right) M \\
&= \left(2 + 2\left(\frac{r+s-3}{n}\right) + O\left(\frac{1}{n^2}\right)\right) M,
\end{aligned}
$$

as desired. □

It should be noted that De Bonis and Katona's proof follows a similar outline to this reproof. They partition their $\mathcal{K}_{r,s}$-free family into two fork-free families, one with no "up-fork" and the other with no "down-fork". They are able to directly apply their $\mathcal{V}_{r+1}$-free theorem to the family with no "up-fork". For the family with no "down-fork", they argue that the $\mathcal{V}_{r+1}$-free result can again be applied since complementation preserves inclusion. In contrast, we split up our family in a way that allows for overlap in the two fork-free subfamilies, $\mathcal{S} \cup \mathcal{T}$ and $\mathcal{U} \cup \mathcal{T}$. This allows us to obtain the same result with less restriction.

2.1.4 The broom poset

Recall that the broom $\mathcal{P}_k(s,t)$ is a height k poset with s minimal elements and t maximal elements, for $s,t \ge 1$ and $k \ge 3$. Theorem 1.1.7 states Griggs and Lu's result that the upper bound for the size of a maximum broom-free family in $\mathcal{B}_n$ approaches $(k-1)M$ as n tends to infinity. Notice that the broom can be considered a generalization of the chain, fork, butterfly, and $\mathcal{K}_{r,s}$ posets. In particular, a broom is a k-chain when $s = t = 1$, it is a fork when $k = s = 1$, $t > 1$, it is a butterfly when $k = s = t = 2$, and it is a $\mathcal{K}_{r,s}$ when $k = 2$, $r, s \ge 2$, and at least one of r and s is greater than 2. The case of $\mathcal{P}_k(1,r)$ is known as an r-fork with a k-shaft and has been examined by De Bonis and Katona [18].

Here we reprove Theorem 1.1.7 using Theorems 2.1.1 and 1.1.2.

New Proof of Theorem 1.1.7. Let $G = (V, A)$ be the comparability digraph of a $\mathcal{P}_k(s,t)$-free family $\mathcal{F}$. Let $\mathcal{S} = \{F \in \mathcal{F} \mid \deg^-(F) < s\}$, $\mathcal{U} = \{F \in \mathcal{F} \mid \deg^+(F) < t\}$, and $\mathcal{T} = V(G) \setminus (\mathcal{S} \cup \mathcal{U})$. Note that $\mathcal{S}$ and $\mathcal{U}$ need not be disjoint. Note further that the number of arcs incident with $\mathcal{S}$ is at most $(s-1)|\mathcal{S}|$, and the number of arcs incident with $\mathcal{U}$ is at most $(t-1)|\mathcal{U}|$.

Because $\mathcal{F}$ is $\mathcal{P}_k(s,t)$-free, the set $\mathcal{T}$ cannot contain a $(k-2)$-chain. Using Theorem 1.1.2, we can conclude that $|\mathcal{T}| \leq M_{k-3}$, the size of $k-3$ middle layers of $\mathcal{B}_n$. Therefore, $|\mathcal{T}| \leq (k-3)M$.

Set $x = |\mathcal{S}| - M$ and note that $|\mathcal{S}| = M + x$. By Theorem 2.1.1, the number of arcs induced by $\mathcal{S}$ is at least $(\lfloor n/2 \rfloor + 1)\, x$. This means

$$(\lfloor n/2 \rfloor + 1)\,(|\mathcal{S}| - M) \leq (s-1)|\mathcal{S}|,$$

which implies

$$\begin{aligned} |\mathcal{S}| &\leq \left(\frac{\lfloor n/2 \rfloor + 1}{\lfloor n/2 \rfloor - s + 2} \right) M \\ &= \left(1 + O\left(\frac{1}{n} \right) \right) M. \end{aligned}$$

A similar argument will show that $|\mathcal{U}| \leq \left(1 + O\left(\frac{1}{n}\right)\right) M$. Thus,

$$|\mathcal{F}| \leq |\mathcal{S}| + |\mathcal{T}| + |\mathcal{U}| \leq \left(k - 1 + O\left(\frac{1}{n} \right) \right) M,$$

as desired. □

2.1.5 Additional applications

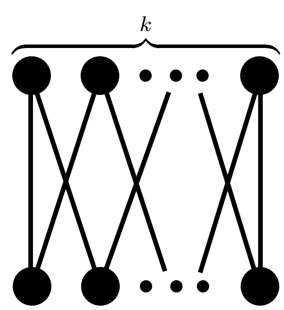

Figure 2.1 The crown $\mathcal{O}_{2k}$.

While this new proof technique works nicely for some height 2 posets it is worth noting that it does not work for all such posets. In particular, it does not work for the generalized crown poset $\mathcal{O}_{2k}$, which is the height 2 poset whose Hasse diagram is an undirected cycle of length $2k$. This poset is another possible generalization of the butterfly $\bowtie$. (The first generalization is the $\mathcal{K}_{r,s}$ poset discussed in Section 1.1.) The Hasse diagram of $\mathcal{O}_{2k}$ can be seen in Figure 2.1.

Despite the difficulties exemplified by the crown, it may be possible to apply this technique to general height 2 posets whose Hasse diagrams are trees. However, it is unclear if this technique could be extended to posets which occupy three or more layers of the Boolean lattice. One exception to this uncertainty is any extension of an acquiescent height 2 poset by a simple k-chain, e.g. the broom poset. In these cases, we can achieve the desired result by applying Theorems 2.1.1 and 1.1.2.

2.2 The diamond poset

In this section, we discuss a density parameter $\pi(P)$, which is closely related to $La(n, P)$. Define $\pi(P)$ to be

$$\pi(P) := \lim_{n\to\infty} \frac{La(n,P)}{\binom{n}{\lfloor n/2 \rfloor}} = \lim_{n\to\infty} La(n,P)M^{-1},$$

when this limit exists. In cases where the limit may not exist, we will use $\pi(P)$ to refer to the partial limit, $\limsup_{n\to\infty} La(n,P)M^{-1}$. After observing that $\pi(P)$ was an integer for all posets where $\pi(P)$ was known, Griggs and Lu [28] proposed the following conjecture:

Conjecture 2.2.1 (Griggs and Lu [28])**.** *For every finite poset P, the limit $\pi(P)$ exists and is an integer.*

The diamond poset, denoted $\Diamond$, is a height 3 poset on four elements, A, B, C, D, such that $A \subset (B \cap C)$ and $B, C \subset D$. Notice that this poset is exactly the Boolean lattice $\mathcal{B}_2$. A Hasse diagram for the diamond can be seen in Figure 2.2.

Some researchers have declared the diamond to be the most challenging poset in extremal poset theory. This is because calculating $La(n, \Diamond)$, and thus $\pi(\Diamond)$, is notoriously difficult. It is not even known if the overall limit $\pi(\Diamond)$ exists, though if it does we trivially have that $2 \leq \pi(\Diamond) \leq 3$. In fact, it has been shown that $\pi(\Diamond) < 3$. This fact, coupled with Conjecture 2.2.1, has led many to guess that $\pi(\Diamond) = 2$. Over the years, there have been a series of improvements on the upper bound for $\pi(\Diamond)$ which are listed in Table 2.1. Note that some of the earlier bounds from Griggs, Li, and Lu were not published on their own, but are discussed in [27] with an improvement. The techniques used to derive these results include simple averaging arguments, the Lubell function method, full chain counting, and flag algebras. Currently, the best known

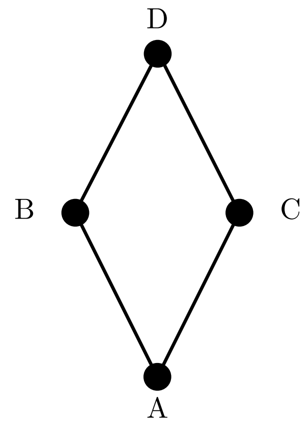

Figure 2.2 The diamond $\Diamond$.

upper bound is from Grósz, Methuku, and Tompkins [29]. Their result was proved using a partition of the maximal chains of a $\Diamond$-free family, coupled with an induction method and is presented in Theorem 2.2.2.

Theorem 2.2.2 (Grósz, Methuku, and Tompkins [29]). $\pi(\Diamond) \leq 2.20711$.

Table 2.1 Upper bounds for $\pi(\Diamond)$.

Bound	**Proved by**	**Year**
2.5	Griggs, Li, and Lu [27]	Pre-2012
2.296	Griggs, Li, and Lu [27]	Pre-2012
2.283	Axenovich, Manske, and Martin [2]	2012
2.273	Griggs, Li, and Lu [27]	2012
2.25	Kramer, Martin, and Young [37]	2013
2.20711	Grósz, Methuku, and Tompkins [29]	2016

Because the standard diamond-free problem is so stubborn, some researchers have considered a special case where the diamond-free family $\mathcal{F}$ consists of subsets of $[n]$ with at most 3 different sizes. In other words, they narrowed the diamond-free problem down to families which live in 3 layers of the Boolean lattice $\mathcal{B}_n$. This is a natural adjustment to make since any diamond-free family can be partitioned into three antichains. A list of known upper bounds for $La(n, \Diamond)$ in three layers can be found in Table 2.2, with the current best known bound being from Balogh, Hu, Lidický, and Liu [5]. Their result was proved using a modified version of the flag algebra technique and is presented in Theorem 2.2.3.

Theorem 2.2.3 (Balogh, Hu, Lidický, and Liu [5])**.** *If $\mathcal{F}$ is a $\Diamond$-free subset of three layers of $\mathcal{B}_n$, then $|F| \leq (2.15121 + o(1))M$.*

Table 2.2 Upper bounds for $La(n, \Diamond)$ in three layers.

Bound	**Proved by**	**Year**
$(2.20711 + o(1))M$	Axenovich, Manske, and Martin [2]	2012
$(2.15471 + o(1))M$	Manske and Shen [40]	2013
$(2.15121 + o(1))M$	Balogh, Hu, Lidický, and Liu [5]	2014

Work is still being done on these problems, but it is made more challenging by the fact that certain known methods cannot be used to obtain better results. For example, Kramer, Martin, and Young [37] note that, without any adjustments, their flag algebra technique will not yield a better upper bound for $\pi(\Diamond)$ than 2.25. This means that any hope of proving either $\pi(\Diamond) = 2$ or $\pi(\Diamond) \neq 2$ will likely depend on the development of new, innovative proof techniques.

CHAPTER 3. THE HYPERGRAPH CONTAINER METHOD

In this chapter, we will discuss the hypergraph container method in more detail. First, however, we need a few additional definitions. A *hypergraph* is a pair $H = (V, E)$ where V is a set of vertices and E is a set of *hyperedges*, which are subsets of V. Note that this is a generalization of a graph which allows edges to consist of more than two vertices. A *k-uniform hypergraph*, or *k-graph*, is a hypergraph for which all hyperedges are k-element subsets of the vertex set. An *independent set* for a hypergraph $H = (V, E)$ is a set $I \subset V$ for which there is no edge $e \in E$ such that $e \subset I$. If a subset of vertices $A \subset V$ contains 'few' edges $e \in E$, we call A *sparse*. The number that constitutes 'few' edges typically depends upon the situation.

Next we will discuss the origins of the hypergraph container method, including its two seminal papers and their treatment of the technique. Then we will explore how to apply the technique to an arbitrary problem. In particular, we will talk about problem reformulations, container building, and supersaturation results. Finally, we will survey some of the various results that have been proved or reproved using this technique.

3.1 History

The hypergraph container method was introduced independently in two papers, namely *Independent sets in hypergraphs* by Balogh, Morris, and Samotij [8] and *Hypergraph containers* by Saxton and Thomason [49]. In both papers, the authors note that a vast number of theorems in combinatorics can be rephrased to concern independent sets in uniform hypergraphs. Their development of this technique follows breakthroughs by Conlon and Gowers [17] and Schacht [50] who came up with two other very different techniques for solving these types of problems.

In [8], Balogh, Morris, and Samotij mention a recent interest in 'sparse random' versions of classical results and assert that the container technique provides a new approach for solving extremal and structural problems of this type. Meanwhile, in [49], Saxton and Thomason emphasize that this technique allows a unified approach to studying combinatorial problems which are concerned, sometimes implicitly, with counting independent sets. Both papers prove the existence of a relatively small collection of well-behaved 'containers' into which all independent sets can be partitioned. In particular, these containers are sparse and of bounded size, meaning they are not much larger than the independent sets they contain. Using containers, the authors of both papers are able to derive numerous interesting results in a straightforward manner. With a bit of extra work, the technique can be applied to a huge variety of problems. This power is a consequence of the method's generality, and a sampling of the problems it has been applied to can be seen in Section 3.3.

3.2 Technique

Proofs using the container technique tend to go through the following steps: First, the initial question must be reformulated to concern counting independent sets in graphs or hypergraphs. Next, it must be shown that all independent sets in a given graph or hypergraph are contained in a small number of relatively sparse vertex subsets called containers. In some cases, a third step is required where the containers are shown to have a particular property and it is deduced that the independent sets also have that property. In the remainder of this section, we will discuss these three steps in more detail.

For the first step, the given problem must be rephrased to concern graphs or hypergraphs and independent sets. How exactly to do this depends heavily on the initial problem. In some cases, the initial question is already concerned with graphs; in others, it may be possible to encode the given information into a graph in some clever way. For example, suppose we are attempting to schedule trains between two stations so that no two trains attempt to use a single track at the same time. To encode this problem in a graph, we could let each vertex correspond to a particular train trip and make two vertices adjacent if the times of the trips overlap. Then any independent set would correspond to a collection of trips which do not overlap and thus they could take place on a single track.

Suppose instead that we are trying to solve a forbidden subposet problem for the poset P. In this case, we may want to consider the comparability graph or digraph for the host lattice. In the case where $P = \mathcal{P}_2$, a $\mathcal{P}_2$-free family is an antichain, and it corresponds exactly to an independent set in the comparability graph. However, for more complicated posets, a P-free family may not be close enough to an independent set to be helpful. When this happens, we can encode P-free families into a $|P|$-graph $H = (V, E)$ in the following way: First, let V consist of all of the poset elements in the host lattice. Next, let e be a hyperedge if the poset elements in e form the subposet P. Then any independent set in H contains no edges, i.e. its vertices form a P-free family. This is a natural extension from comparability graphs to a sort of comparability hypergraph. Indeed, for the case where $P = \mathcal{P}_2$, these two encoding techniques are identical. Though a researcher may want to consider these established encoding techniques for each new problem, it may be necessary to try additional encoding techniques, or even develop new ones, before an optimal technique is discovered.

Once the hypergraph setting has been established, it is time to build the containers. In the case that our initial problem may be encoded in a regular graph $G = (V, E)$ (i.e. a 2-graph), we may employ an algorithm to generate our containers. This algorithm is discussed in [10] where it is credited to Kleitman and Winston [35; 36]. Its outline is presented below as Algorithm 3.2.1. It should be noted that this algorithm is also employed in [9] and that we use a modified version of it to prove Theorem 1.2.5 in Chapter 4.

Algorithm 3.2.1 (The Graph Container Algorithm [10; 35; 36])**.** *Let $n := |V|$ and fix an arbitrary total order $v_1, \ldots, v_n$ of V and some $\Delta > 0$. Let I be an independent set in G. Set $G_0 := G$ and $S := \emptyset$. In Step i of the algorithm we do the following:*

(a) Let u be the vertex of maximum degree in G_{i-1} (ties are broken here by our fixed total ordering);

(b) If $u \notin I$ then define $G_i := G_{i-1} \setminus \{u\}$ and move to Step $i + 1$;

(c) If $u \in I$ and $\deg_{G_{i-1}}(u) \geq \Delta$ then add u to S; define $G_i := G_{i-1} \setminus (\{u\} \cup N_G(u))$ and move to Step $i + 1$;

(d) *If* $u \in I$ *and* $\deg_{G_{i-1}}(u) < \Delta$ *then add* u *to* S*; define* $G_i := G_{i-1} \setminus \{u\}$ *and* $f(S) := V(G_i)$ *and terminate.*

The input for Algorithm 3.2.1 is a graph G and degree parameter Δ, both of which depend on the given problem. What the algorithm does is process the vertices of G one at a time. It begins with high degree vertices and considers vertices in the same order each time since a total ordering is fixed in advance. Vertices that are not in the independent set are ignored, and vertices which are in the independent set are placed into the set S as long as their degree is at least as big as the parameter Δ. When no more vertices have degree at least Δ, the remaining unprocessed vertices are placed into the set $f(S)$. It is clear then that $S \subset I$, $I \subset (S \cup f(S))$, and $S \cap f(S) = \emptyset$.

It is less obvious that this algorithm generates a function f whose domain is the set of subsets of V given by $\binom{V}{\leq |V|/\Delta}$ and whose codomain is the collection of all subsets of V. However, it can be shown that $|S| \leq |V|/\Delta$ and that the function f is well-defined, and we will do so in Section 4.2.

Together, the input and output of the function f form a container for the independent set I, i.e. $I \subset S \cup f(S)$. A sufficiently large choice for Δ will typically ensure that the number of containers yielded by Algorithm 3.2.1 is $2^{o(|I_{max}|)}$ where I_{max} denotes a maximum independent set in G. Further, a sufficiently small choice for Δ will ensure that the containers are sparse. If there does not exist a Δ which ensures both of these container requirements occur, we may apply the algorithm twice. For the first application, we choose a parameter Δ_1 which ensures the correct number of containers is produced; for the second application, we choose a smaller parameter Δ_2 which ensures the containers are sufficiently sparse. To see an example of this double algorithm application, refer to the proof of Lemma 4.2.1 in Section 4.2.

Once the containers are established, it is customary to apply some supersaturation result to ensure that each container is of the proper size. Roughly, a supersaturation result will assert that a set larger than $(1 + o(1))|I_{max}|$ must contain many edges, i.e. it is not sparse. Thus, by virtue of the fact that each container is sparse, it must have size at most $(1 + o(1))|I_{max}|$. This closeness of the sizes of maximum independent sets and containers allows us to deduce that independent sets have a variety of desired properties. For example, in the proof of Theorem 1.2.1

(Theorem 2 in [9]), they argue that the probability that a container in $\mathcal{P}(n,p)$ has size greater than $(1+o(1))pM$ approaches zero as n tends to infinity. (Note that for the random Sperner theorem, $|I_{max}| = pM$.) They then deduce that, with high probability, the largest antichain in $\mathcal{P}(n,p)$ must have size at most $(1+o(1))pM$, which is the desired result. Like the initial problem reformulation, how to perform this final deduction depends heavily on the problem at hand and may require additional clever lemmas.

3.3 Applications

The abstractness of the hypergraph container technique makes it applicable to a wide variety of problems. We have already mentioned that hypergraph containers were used to solve a random version of Sperner's Theorem (See Theorem 1.2.1) and a random version of Erdős's Theorem (See Theorem 1.2.2). In this section, we will discuss a selection of additional results which have also been proved with this technique. For a more in-depth look at these and other hypergraph container results, see [3; 4; 6; 7; 8; 9; 10; 11; 16; 42; 43; 46; 49].

3.3.1 List colorings

Many interesting problems arise in graph theory when we assign colors to graph substructures. Possible substructures that could be labeled with colors include (but are not limited to) vertices, edges, and subgraphs. The most basic type of graph coloring is a *proper vertex coloring* where we assign colors to all vertices such that no two adjacent vertices share the same color. Recall that two vertices u and v are adjacent in a graph or hypergraph if there exists an edge e which contains both u and v. A graph G is said to be k-*colorable* if its vertices can be properly colored with at most k colors, and the *chromatic number* of G, $\chi(G)$, is the minimum value of k such that G is k-colorable.

The concept of vertex coloring can be extended to *list coloring* in the following way: Each vertex v of G is assigned a list of colors $L(v)$, and we attempt to properly color the vertices of G such that the color of v is chosen from the list $L(v)$. G is said to be k-*choosable*, or k-*list-colorable*, if it is possible to properly list-color G whenever each vertex v is assigned a list $L(v)$ of k colors. The *list chromatic number* of G, denoted $\chi_l(G)$, is the smallest k such that G

is k-choosable. This sort of coloring problem was first studied by Erdős, Rubin, and Taylor [22] and by Vizing [53].

In a hypergraph $H = (V, E)$, the degree of a vertex $v \in V$, denoted $\deg(v)$, is the number of edges $e \in E$ such that $v \in e$. A graph or hypergraph is said to be *d-regular* if $\deg(v) = d$ for all $v \in V$. The *average degree* of a graph is $|V|^{-1} \sum_{v \in V} \deg(V)$. Results concerning the list chromatic number of r-graphs which are either d-regular or of average degree d have been proved in [1; 30; 31; 48]. In [49], Saxton and Thomason extend these results to all simple r-graphs with Theorem 3.3.1 (Theorem 2.1 in [49]).

Theorem 3.3.1 (Saxton and Thomason [49])**.** *Let $r \in \mathbb{N}$ be fixed. Let G be a simple r-graph with average degree d. Then as $d \to \infty$,*

$$\chi_l(G) \geq (1 + o(1)) \frac{1}{(r-1)^2} \log_r d$$

holds. Moreover, if G is regular then

$$\chi_l(G) \geq (1 + o(1)) \frac{1}{r-1} \log_r d.$$

To prove this result, they consider an r-graph $G = (V, E)$, a list of t colors $[t]$, and a collection of independent sets $\mathcal{I} = \{I_1, \dots, I_t\}$ which roughly correspond to the sets of vertices which receive colors 1 through t, respectively. A collection of color lists of size ℓ, $\mathcal{L} = \{L(v) \mid v \in V,\ L(v) \subset [t],\ |L(v)| = \ell\}$, is compatible with $\mathcal{I}$ if each vertex v lies within an independent set I_j for which $j \in L(v)$. In this setting, Saxton and Thomason use hypergraph containers to show that as $d \to \infty$, an incompatible collection of lists exists when $\ell < (1 + o(1))\frac{1}{(r-1)^2} \log_r d$ if G has average degree d and when $\ell < (1 + o(1))\frac{1}{r-1} \log_r d$ if G is d-regular. This yields the desired result because the existence of an incompatible collection of lists of size ℓ means that $\chi_l(G) > \ell$.

3.3.2 Arithmetic progressions

A *k-term arithmetic progression*, or k-AP, in $[n]$ is a set of distinct elements of the form $a, a+d, a+2d, \dots, a+(k-1)d$ where $k, d \in \mathbb{N}$. According to a famous theorem of Szemerédi [52], the largest subset $A \subset [n]$ such that A contains no k-AP has $|A| = o(n)$. A sparse random version of this celebrated result, presented in Theorem 3.3.2, has been proved by Conlon and Gowers [17] and by Schacht [50]. It has also been proved using hypergraph containers by Saxton and Thomason [49] and by Balogh, Morris, and Samotij [8].

Theorem 3.3.2 (Sparse random version of Szemerédi's Theorem [8; 17; 49; 50])**.** *Let $k \geq 3$ and $\epsilon > 0$. There exists a constant $C > 0$ such that for $p \geq Cn^{-1/(k-1)}$, if A is a subset of $[n]$ whose elements are chosen independently at random with probability p, then with high probability, any subset of A of size $\epsilon|A|$ contains a k-AP.*

To prove this theorem, Balogh, Morris, and Samotij consider a k-graph $H = (V, E)$ where $V = [n]$ and E is the set of k-APs in $[n]$. An independent set in this hypergraph is a set which contains no k-AP. In this setting, they use hypergraph containers to show that for $\delta > 0$ and $m = \Omega\left(n^{1-1/(k-1)}\right)$, there are at most $\binom{\delta n}{m}$ subsets of $[n]$ of size m with no k-AP. They then use a Chernoff bound to deduce the desired random result.

3.3.3 Sum-free sets

A set of integers A in $[n]$ is *sum-free* if for all $x, y \in A$ (not necessarily distinct), $x + y \notin A$. A is *maximal* if for all $z \in [n] \setminus A$, $A \cup \{z\}$ is not sum-free. A conjecture of Cameron and Erdős [14], which was independently confirmed by Green [23] and Sapozhenko [47], stated that $[n]$ contains $O(n^{n/2})$ sum-free sets. Later, Cameron and Erdős [15] posed a question asking if the number of maximal sum-free subsets is $o(f(n))$, where $f(n)$ denotes the total number of sum-free subsets of $[n]$. This question was answered positively, first by Łuczak and Schoen [39] and later by Wolfovitz [54] who showed that the number of maximal sum-free subsets is at most $2^{3n/8+o(n)}$. Cameron and Erdős [15] also provided a lower bound of $2^{\lfloor n/4 \rfloor}$ for the number of maximal sum-free subsets. In [7], Balogh, Liu, Sharifzadeh, and Treglown use hypergraph containers to show that this lower bound is asymptotically correct. Their result is presented below in Theorem 3.3.3.

Theorem 3.3.3 (Balogh, Liu, Sharifzadeh, and Treglown [7])**.** *There are at most $2^{(1/4+o(1))n}$ maximal sum-free sets in $[n]$.*

To prove this theorem, Balogh, et. al. combine a container result from Green (Proposition 6 in [23]), a removal lemma from Green (Corollary 1.6 in [24]), a characterization of all sum-free sets from Deshouillers, Freiman, Sós, and Temkin [20], and some results about the number of maximal independent sets in graphs which may have loops. Green's container lemma finds containers for all sum-free subsets of $[n]$ which contain at most $o(n^2)$ Schur triples. Though it preceded the development of the hypergraph container technique, this lemma can be obtained as a consequence of the container results in [8; 49]. Green's removal lemma shows that every subset of $[n]$ with $o(n^2)$ Schur triples can be partitioned into a sum-free set B and a set C with $|C| = o(n)$. Deshouillers, Freiman, Sós, and Temkin show that each sum-free subset S of $[n]$ either consists of odd numbers or has size bounded above by one of $2n/5 + 1$ and $\min(S)$.

With these existing results in mind, Balogh, et. al. reformulate the sum-free problem to concern maximal independent sets in what they refer to as link graphs. Before we define a link graph, note that a *Schur triple* is a set of integers $x, y, z \in [n]$ (not necessarily distinct) such that $x + y = z$. Given any subsets $B, S \in [n]$, the *link graph* of S on B, denoted $L_S[B]$, has vertex set B and edges defined as follows:

(a) (x, y) is an edge of $L_S[B]$ if there is some $z \in S$ such that $\{x, y, z\}$ is a Schur triple;

(b) there is a loop at vertex x if $\{x, x, z\}$ is a Schur triple for some $z \in S$ or if $\{x, y, z\}$ is a Schur triple for some $y, z \in S$.

It can be shown that if B and S are sum-free subsets of $[n]$ and $I \subset B$ is a set such that $I \cup S$ is a maximal sum-free subset of $[n]$, then I is a maximal independent set in $L_S[B]$. Choosing the sets B, S, and I in a clever way and applying the container lemma then yields the desired result from Theorem 3.3.3.

3.3.4 Intersecting families

A family $\mathcal{A}$ in $\mathcal{P}(n)$ is *intersecting* if for all $A, B \in \mathcal{A}$, $|A \cap B| \geq 1$. $\mathcal{A}$ is *t-intersecting* if for all $A, B \in \mathcal{A}$, $|A \cap B| \geq t$. Katona's famous intersection theorem [33] states that the largest t-intersecting family in $\mathcal{P}(n)$ has size $\binom{n}{\geq (n+t)/2}$ if $n+t$ is even and size $2\binom{n-1}{\geq (n+t-1)/2}$ otherwise. In particular, the family $\binom{[n]}{\geq (n+t)/2}$ is t-intersecting when $n+t$ is even. In [10], Balogh, Treglown, and Wagner prove a random version of Katona's intersection theorem. Their result is presented below in Theorem 3.3.4.

Theorem 3.3.4 (Balogh, Treglown, and Wagner [10]). *If $p = 2^{-o(\sqrt{n}\log n)}$ and $t = o(\sqrt{n})$ then with high probability the largest t-intersecting family in $\mathcal{P}(n,p)$ has size $(1/2 + o(1))2^n p$.*

To prove the lower bound, Balogh, Treglown, and Wagner use a Chernoff bound to show that the expected number of sets from $\binom{[n]}{\geq (n+t)/2}$ that will be present in $\mathcal{P}(n,p)$ is at least $(1/2 - o(1))2^n p$. They then use a union bound to show that, with high probability, there are sufficiently many sets in $\mathcal{P}(n,p)$ which can be added to $\binom{[n]}{\geq (n+t)/2}$ to form a larger t-intersecting family of size at least $(1/2 + o(1))2^n p$.

For the upper bound, Balogh, Treglown, and Wagner considered a graph G on vertex set $\mathcal{P}(n)$ where $A, B \in \mathcal{P}(n)$ form an edge in G if and only if $A \cap B = \emptyset$. Note that in this graph, an independent set corresponds exactly to an intersecting set because nonadjacent vertices correspond to sets with nonempty intersections. A single application of Algorithm 3.2.1 can then be used to build containers for the intersecting sets, and a Chernoff bound can show that, with high probability, the intersection of any container with $\mathcal{P}(n,p)$ is at most $(1/2 + o(1))2^n p$. Since any t-intersecting family is clearly an intersecting family, we get that the largest t-intersecting family in $\mathcal{P}(n,p)$ has size at most $(1/2 + o(1))2^n p$, as desired.

3.3.5 Union-free families

A family $\mathcal{F}$ in $\mathcal{P}(n)$ is *union-free* if there are no three distinct sets $A, B, C \in \mathcal{F}$ such that $A \cup B = C$. The question asking how many union-free families exist in $\mathcal{P}(n)$ was first addressed by Burosch, Demetrovics, Katona, Kleitman, and Sapozhenko in [13]. They found that this number was between $2^{\binom{n}{n/2}}$ and $2^{2\sqrt{2}\binom{n}{n/2}(1+o(1))}$, and they conjectured that it is actually equal to $2^{\binom{n}{n/2}(1+o(1))}$. Recently, their conjecture was confirmed by Balogh and Wagner [11], who proved the result in Theorem 3.3.5 using the hypergraph container technique.

Theorem 3.3.5 (Balogh and Wagner [11]). *The number of union-free families in $\mathcal{P}(n)$ is $2^{\binom{n}{n/2}(1+o(1))}$.*

In their proof, Balogh and Wagner consider a 3-graph H with vertex set $\mathcal{P}(n)$ where $A, B, C \in \mathcal{P}(n)$ form a hyperedge if and only if $A \cup B = C$. In this context, independent sets and union-free families are equivalent. However, Balogh and Wagner were unable to use any existing container theorems because the co-degrees in this hypergraph could be very large. Indeed, given two sets $A, C \in \mathcal{P}(n)$, there may be many possible sets $B \in \mathcal{P}(n)$ such that $A \cup B = C$. To address this problem, they introduce the notion of an r-rooted hypergraph. Specifically, a 3-graph $H = (V, E)$ is *r-rooted* if there is a function $f : E \to V$ such that for every edge $e \in E$, $f(e) \in e$ and for any pair of vertices $u, v \in V$, there are at most r edges $e \in E$ for which $u, v \in e$ and $f(e) \notin \{u, v\}$. For edge $e \in E$, $f(e)$ is called the *head* of e. Since for any $A, B \in \mathcal{P}(n)$ there is a unique $C \in \mathcal{P}(n)$ such that $A \cup B = C$, the authors choose C to be the head of edge $\{A, B, C\}$, and this rooted (oriented) version of H avoids the co-degree issue.

The next issue that Balogh and Wagner needed to overcome concerned the degrees of link graphs (similar to those discussed in Subsection 3.3.3), which help reduce the problem from 3-graphs to 2-graphs. They managed to address this issue by proving a balanced supersaturation result for the hypergraph H defined above. This result, which we will not discuss here, allowed them to prove an r-rooted container lemma and in turn, Theorem 3.3.5.

Subsections 3.3.1 through 3.3.5 contain just a small sampling of the sort of problems that can be solved with the hypergraph container method. Additional topics to which it can be applied include characterizing graphs without triangles [6] or large cliques [3], calculating thresholds for Maker-Breaker games [43], calculating Folkman numbers [46], examining H-free and induced H-free graphs [49], proving sparse random versions of famous theorems such as Erdős-Stone [8], and counting error-correcting codes [10]. Many of these results are enlightening because of the additional tools that are developed to facilitate their proofs. e.g. balanced supersaturation results, link graphs, and rooted hypergraphs. In fact, some of these tools may be necessary to prove random forbidden subposet results for posets with more structure, like $\bowtie$ or $\mathcal{K}_{r,s}$. We will say more about this potential application in Section 4.4.

CHAPTER 4. THE RANDOM $(r+1)$-FORK-FREE THEOREM

Our main result, Theorem 1.2.5, is the subject of this final chapter, and we will prove it in detail in Section 4.3. First, however, we will prove the lower bound theorem, Theorem 1.2.6, in Section 4.1. After that, we will use the hypergraph container technique described in Chapter 3 to prove our $\mathcal{V}_{r+1}$-free container lemma, Lemma 4.2.1, in Section 4.2. Then once the main result is proved, we will discuss some final conclusions and open questions in Section 4.4.

4.1 Proof of the lower bound theorem

In this section we provide a short proof of Theorem 1.2.6. Like Corollary 1.2.4, this result can be quickly deduced from Theorem 1.2.3.

Proof of Theorem 1.2.6. Suppose $p \ll 1/n$. Then by Theorem 1.2.3, we have that with high probability the largest antichain in $\mathcal{P}(n,p)$ has size at least $(2-o(1))pM$. Every antichain in $\mathcal{P}(n,p)$ is also a $\mathcal{V}_{r+1}$-free family. Thus, with high probability, the largest $\mathcal{V}_{r+1}$-free family in $\mathcal{P}(n,p)$ has size at least $(2-o(1))pM$, as desired. □

4.2 The $\mathcal{V}_{r+1}$-free container lemma

Let $G=(V,A)$ be the comparability digraph of $\mathcal{B}_n$. We will now present the key piece of machinery used to prove Theorem 1.2.5. Lemma 4.2.1, also known as the $\mathcal{V}_{r+1}$-free Container Lemma, is a very technical result which guarantees the existence of a set of containers $\mathcal{C} \subset \mathcal{P}(n)$ with the following three properties:

(1) $|\mathcal{C}| = 2^{o(M)}$;

(2) Every $\mathcal{V}_{r+1}$-free set I in G lies in some $C \in \mathcal{C}$;

(3) $|C| \leq (1+o(1))M$ for each $C \in \mathcal{C}$.

It should be noted that Lemma 4.2.1 is similar to Lemma 6 in [9].

Lemma 4.2.1 ($\mathcal{V}_{r+1}$-free Container Lemma)**.** *Suppose that $t \in \mathbb{N}$, $0 < \epsilon \leq 1/(2t)^{t+1}$ and n is sufficiently large. Then there exist functions $f : \binom{V}{\leq (r+1)2^n n^{-(t+0.9)}} \to \binom{V}{\leq (t+1+\epsilon)M}$ and $g : \binom{V}{\leq (r+1)\frac{(t+2)M}{\epsilon^2 n^t}} \to \binom{V}{\leq (t+\epsilon)M}$ such that, for any $\mathcal{V}_{r+1}$-free set I in G, there are disjoint subsets $S_1, S_2 \subseteq I$ such that $S_1 \cup S_2$ and $g(S_1 \cup S_2)$ are disjoint, $S_2 \subseteq f(S_1)$ and $I \subseteq S_1 \cup S_2 \cup g(S_1 \cup S_2)$.*

This set of containers will be constructed using a modified version of Algorithm 3.2.1 in two phases. Let t and ϵ be as given in the lemma statement and let n be sufficiently large. In the first phase, we process vertices using the parameter $\Delta_1 = n^{t+0.9}$ to ensure that $|\mathcal{C}| = 2^{o(M)}$. All vertices which are added to the container during Phase 1 are put into the set labeled S_1. After Phase 1 terminates, the vertices which have not yet been processed are put into the set labeled $f(S_1)$. Then we run the second phase on the induced subgraph $G[f(S_1)]$ using the parameter $\Delta_2 = \epsilon^2 n^t$ to ensure that each $C \in \mathcal{C}$ satisfies $|C| \leq (1+o(1))M$. All vertices which are added to the container during Phase 2 are put into the set labeled S_2. When Phase 2 terminates, the vertices left in the graph are put into the set labeled $g(S_1 \cup S_2)$ and the container-building process is complete. The result is three pairwise disjoint sets S_1, S_2, and $g(S_1 \cup S_2)$ such that $(S_1 \cup S_2) \subseteq I \subseteq (S_1 \cup S_2 \cup g(S_1 \cup S_2))$. $S_1 \cup S_2$ is often referred to as the *label* of the container $S_1 \cup S_2 \cup g(S_1 \cup S_2)$.

Figure 4.1 diagrams the relationship between containers and $\mathcal{V}_{r+1}$-free families in $\mathcal{B}_n$. A similar figure was presented by W. Samotij in a 2013 talk given at The Hebrew University of Jerusalem. Note that though this picture looks a bit like the Boolean lattice, it actually shows the collection of all families in the Boolean lattice. This means that the maximum element is $\mathcal{P}(n)$ instead of $[n]$, and the minimum element is $\emptyset$.

The modifications we will make to Algorithm 3.2.1 are necessary because the original version of the algorithm assumes we are dealing with an independent set, and for this lemma we are not. Notice that a $\mathcal{V}_{r+1}$ in $\mathcal{B}_n$ corresponds to a vertex with out-degree at least $r+1$ in G. Therefore, a $\mathcal{V}_{r+1}$-free family in $\mathcal{B}_n$ corresponds exactly to a set of vertices $I \subset V$ such that for

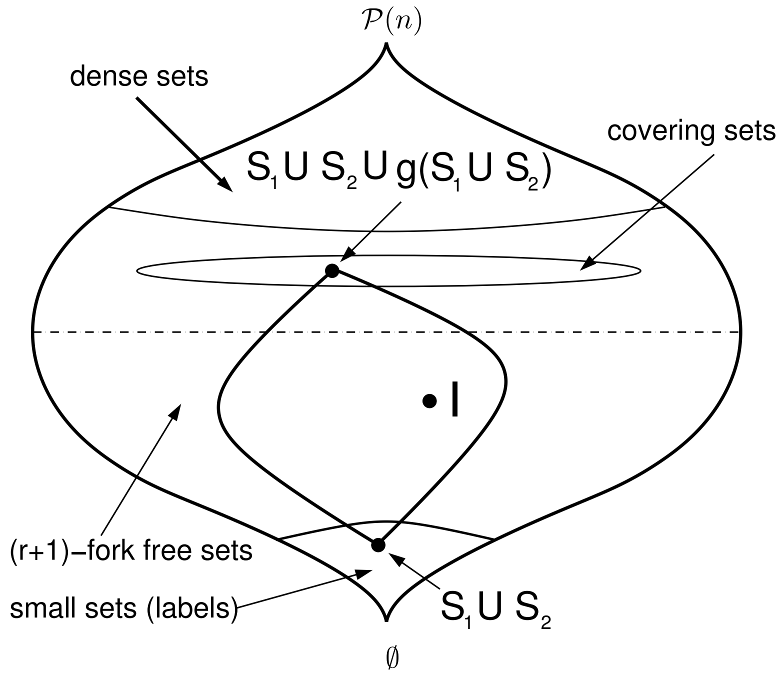

Figure 4.1 A $\mathcal{V}_{r+1}$-free family I and its container $S_1 \cup S_2 \cup g(S_1 \cup S_2)$.

all $v \in I$, $\deg_G^+(v) \leq r$. To account for the few edges that may be present in I, we will make the following changes to Algorithm 3.2.1 cases (c) and (d): In both cases (c) and (d), we will add u and its out-neighbors in I to the set S. In case (d), we will remove u and its out-neighbors in I from the graph G_i before terminating the algorithm. For both of the modifications, we are adding a bounded number of vertices to the set S and removing a bounded number of vertices from the graph G_i. Though this change must be accounted for, it does not prevent a set of containers with the desired properties from being created.

Lemma 4.2.2 was presented as Corollary 5 by Balogh, Mycroft, and Treglown in [9]. It is an immediate corollary of a theorem by Kleitman [34]. We will use it during our proof of Lemma 4.2.1 to verify that the sets $f(S_1)$ and $g(S_1 \cup S_2)$ are of the correct size. Though G represents the comparability graph of $\mathcal{B}_n$ in Lemma 4.2.2, the result also holds for comparability digraphs and is therefore applicable in our proof.

Lemma 4.2.2. *Let $G = (V, E)$ be the comparability graph of $\mathcal{B}_n$, $U \subseteq V$ and suppose that $0 < \epsilon \leq 1/2$ and $t \in \mathbb{N}$. If $|U| \geq (t + \epsilon)M$, then $e(G[U]) > \epsilon n^t |U|/(2t)^{t+1}$.*

We are now ready to prove Lemma 4.2.1.

Proof of Lemma 4.2.1. Fix an arbitrary total ordering on the vertices of G, the comparability digraph of $\mathcal{B}_n$. Let I be a $\mathcal{V}_{r+1}$-free set in G, let $G_0 := G$, and let S_1 and S_2 be initially empty. We will build S_1 and S_2 using an iterative process. As we run through this process, any ties during vertex selection will be broken by the total ordering. Begin at Phase 1 Step 1.

Phase 1: At Step i, let u be a vertex with maximum out-degree in G_{i-1}. If $u \notin I$, let $G_i := G_{i-1} - \{u\}$ and move to Phase 1 Step $i+1$. If $u \in I$ and $\deg^+_{G_{i-1}}(u) \geq n^{t+0.9}$, add u to S_1. Since I is $\mathcal{V}_{r+1}$-free, there are at most r vertices, $v_1, \ldots, v_r$, in $I \cap N^+_{G_{i-1}}(u)$. If any exist, add the v_i to S_1 as well. Let $G_i := G_{i-1} - (\{u\} \cup N^+_G(u))$ and move to Phase 1 Step $i+1$. If $u \in I$ but $\deg^+_{G_{i-1}}(u) < n^{t+0.9}$, add u and any v_i (if they exist) to S_1 and remove both u and the v_i (if they exist) from G_{i-1} to get G_i, but do not remove any other neighbors of u from G_{i-1}. Set $f(S_1) := V(G_i)$, then go to Phase 2 Step $i+1$.

Phase 2: At Step i, let u be a vertex with maximum out-degree in G_{i-1}. If $u \notin I$, let $G_i := G_{i-1} - \{u\}$ and move to Phase 2 Step $i+1$. If $u \in I$ and $\deg^+_{G_{i-1}}(u) \geq \epsilon^2 n^t$, add u to S_2. If any exist, add all $v_i \in I \cap N^+_{G_{i-1}}(u)$ to S_2 as well. Let $G_i := G_{i-1} - (\{u\} \cup N^+_G(u))$ and move to Phase 2 Step $i+1$. If $u \in I$ but $\deg^+_{G_{i-1}}(u) < \epsilon^2 n^t$, add u and any v_i (if they exist) to S_2 and remove both u and the v_i (if any exist) from G_{i-1} to get G_i, but do not remove any other neighbors of u from G_{i-1}. Set $g(S_1 \cup S_2) := V(G_i)$, then terminate the process.

By design, we have that S_1 and S_2 are disjoint subsets of I such that $(S_1 \cup S_2) \cap (g(S_1 \cup S_2)) = \emptyset$, $S_2 \subseteq f(S_1)$, and $I \subseteq S_1 \cup S_2 \cup g(S_1 \cup S_2)$. Note also that $f(S_1)$ represents all vertices left in G_i at the end of Phase 1 and that $g(S_1 \cup S_2)$ represents all vertices left in G_i at the end of Phase 2.

Now, notice that in each step of Phase 1, at most $r+1$ vertices are added to S_1. This means that the number of steps taken during Phase 1 is at least $\frac{1}{r+1}|S_1|$. Further, notice that in Phase 1 Step i, no fewer than $n^{t+0.9}$ out-neighbors of u are deleted from G_{i-1} to form G_i. The only exception happens during the final step of Phase 1 when at most r out-neighbors are deleted from G_{i-1} to form G_i. This means that the total number of steps taken during Phase

1 is at most $1 + \frac{|V|}{n^{t+0.9}+1}$. Putting these two inequalities together gives us

$$\frac{1}{r+1}|S_1| \leq 1 + \frac{|V|}{n^{t+0.9}+1}, \tag{4.1}$$

which gives us the following upper bound on the size of S_1 for n sufficiently large:

$$\begin{aligned} |S_1| &\leq (r+1) + \frac{(r+1)|V|}{n^{t+0.9}+1} \\ &\leq (r+1)2^n(2^{-n} + (n^{t+0.9}+1)^{-1}) \\ &\leq (r+1)2^n n^{-(t+0.9)}. \end{aligned}$$

Further, at the end of Phase 1, every vertex of $G_i = G[f(S_1)]$ has $\deg^+_{G_i}(v) \leq n^{t+0.9}$. So we can upper bound the number of edges in $G[f(S_1)]$ by

$$e(G[f(S_1)]) \leq |f(S_1)|n^{t+0.9} \leq \frac{\epsilon n^{t+1}|f(S_1)|}{2(t+1)^{t+2}}, \tag{4.2}$$

when n is large enough to guarantee that $\frac{(2(t+1))^{t+2}}{n^{0.1}} \leq \epsilon$. Recall that $\epsilon \leq \frac{1}{(2t)^{t+1}}$. Then since $t \in \mathbb{N}$, we have $\epsilon \leq 1/2$ and can apply Lemma 4.2.2 to get $|f(S_1)| < (t+1+\epsilon)M$.

In Phase 2, we begin with a digraph on $|f(S_1)|$ vertices and run through a process similar to Phase 1. So by an argument similar to the justification for the inequality in (4.1), we have

$$\frac{1}{r+1}|S_2| \leq 1 + \frac{|f(S_1)|}{\epsilon^2 n^t + 1},$$

which gives us the following upper bound for the size of S_2:

$$\begin{aligned} |S_2| &\leq (r+1)\left(1 + \frac{(t+1+\epsilon)M}{\epsilon^2 n^t + 1}\right) \\ &\leq (r+1)\left(1 + \frac{(t+1+\epsilon)M}{\epsilon^2 n^t}\right). \end{aligned}$$

Combining the upper bounds for $|S_1|$ and $|S_2|$ allows us to upper bound $|S_1 \cup S_2|$ in the following way:

$$\begin{aligned} |S_1 \cup S_2| &= |S_1| + |S_2| \\ &\leq (r+1)2^n n^{-(t+0.9)} + (r+1)\left(1 + \frac{(t+1+\epsilon)M}{\epsilon^2 n^t}\right) \\ &\leq (r+1)\left(\frac{(t+2)M}{\epsilon^2 n^t}\right). \end{aligned}$$

Further, since every vertex of G_i has $\deg^+_{G_i}(v) \le \epsilon^2 n^t$ at the end of Phase 2, we get a bound for the number of edges in $G[g(S_1 \cup S_2)]$ which is similar to the bound in (4.2):

$$e(G[g(S_1 \cup S_2)]) \le \epsilon^2 n^t |g(S_1 \cup S_2)| \le \frac{\epsilon n^t |g(S_1 \cup S_2)|}{(2t)^{t+1}}.$$

So we can again apply Lemma 4.2.2, and we will get $|g(S_1 \cup S_2)| < (t+\epsilon)M$.

All that remains to be shown is that the functions f and g are well-defined.

To begin, suppose that I and I' are two different $\mathcal{V}_{r+1}$-free sets in G that yield the same set S_1. Note that this immediately implies $S_1 \subset I \cap I'$. We will show that I and I' also yield the same set for $f(S_1)$:

Let G_{i-1} and G'_{i-1} be the graphs associated with I and I', respectively, at the beginning of Step i in Phase 1. Notice that in Phase 1 Step 1, $G_0 = G'_0 = G$.

At Phase 1 Step i, let u be the selected vertex of maximum out-degree. Since we've fixed a total ordering, we will choose the same u regardless of whether we are considering I or I'. There are two cases for u.

Case 1: If u is in $V(G_{i-1}) \setminus (I \cup I')$, then we do not add u to S_1, and we remove u from both G_{i-1} and G'_{i-1} to form G_i and G'_i.

Case 2: If u is in $I \cap I'$, then in both cases u will be added to S_1. If u has any out-neighbors $v_i \in I$, then those same out-neighbors must be contained in I'. This is because the v_i would be added to S_1 when considering I but not when considering I'. During an intermediate step in the phase, u and all of its out-neighbors (including the v_i) would be removed from G'_{i-1} to form G'_i. During the final step in the phase, the v_i would simply not be added to S_1 before Phase 1 terminates. Either way, the v_i would never be added to S_1 when considering I', which contradicts the fact that I and I' yield the same S_1. Thus, if any v_i exist, we have $v_i \in I \cap I'$. If u satisfies the minimum degree condition of Phase 1, then u and all of its out-neighbors are removed from both G_{i-1} and G'_{i-1} to form G_i and G'_i. Alternatively, if u does not satisfy the minimum degree condition, then u and the v_i (if any exist) will be removed from both G_{i-1} and G'_{i-1} to form G_i and G'_i, but no other out-neighbors of u will be removed from either graph.

So regardless of whether we are considering I or I', we will always remove the same vertices from the graph at the end of the step. Thus, at the end of Phase 1, we will have the same graph $G_i = G'_i$, i.e. we will get the same set $f(S_1)$.

A similar argument will show that if I and I' yield the same sets for S_1, $f(S_1)$, and S_2, they will also yield the same set for $g(S_1 \cup S_2)$.

Next, consider two different $\mathcal{V}_{r+1}$-free sets I and I', which yield the sets S_1, S_2 and S'_1, S'_2, respectively, such that $S_1 \cup S_2 = S'_1 \cup S'_2$. We will show that $S_1 = S'_1$ (equivalently $S_2 = S'_2$):

Let G_{i-1} and G'_{i-1} be the graphs associated with I and I', respectively, at the beginning of Step i in Phase 1. Notice that in Phase 1 Step 1, $G_0 = G'_0 = G$ and $S_1 = S'_1 = \emptyset$.

Suppose $S_1 = S'_1$ and $G_{i-1} = G'_{i-1}$ at the beginning of Phase 1 Step i. Because of the fixed total ordering, we will select the same vertex u of maximum outdegree regardless of whether we are considering I or I'. There are three cases for u.

Case 1: If $u \in V(G_{i-1}) \setminus (I \cup I')$, then we do not change S_1 or S'_1, and we remove u from both G_{i-1} and G'_{i-1} to form G_i and G'_i.

Case 2: If $u \in I \setminus I'$, then we add u to S_1 but remove it from both G_{i-1} and G'_{i-1} to form G_i and G'_i. However, this would mean that $u \notin S'_1 \cup S'_2 = S_1 \cup S_2$, which is a contradiction. So $u \notin I \setminus I'$, and by a similar argument, $u \notin I' \setminus I$.

Case 3: If $u \in I \cap I'$, then we add u to both S_1 and S'_1. If u has any out-neighbors v_i in $I \setminus I'$, we will add those v_i to S_1 but not to S'_1. In fact, since $v_i \notin I'$ and $S'_1 \cup S'_2 \subset I'$, we will never add any v_i to $S'_1 \cup S'_2 = S_1 \cup S_2$. This is a contradiction, so any out-neighbor of u in I is contained in $I \cap I'$. By a similar argument, any out-neighbor of u in I' is contained in $I \cap I'$. This means that in this step, we add the same vertices to S_1 and S'_1, and we remove the same vertices from G_{i-1} and G'_{i-1}.

In each case, we have $S_1 = S'_1$ and $G_i = G'_i$ after Phase 1 Step i is complete. A similar argument will hold during Phase 2. Therefore, by induction, we have that $S_1 \cup S_2 = S'_1 \cup S'_2$ implies $S_1 = S'_1$ (equivalently $S_2 = S'_2$).

So the functions f and g are well-defined, and the proof is complete. $\square$

4.3 The random $\mathcal{V}_{r+1}$-free theorem

Now that the $\mathcal{V}_{r+1}$-free container lemma is established, we can finally prove the random $\mathcal{V}_{r+1}$-free theorem. In fact, we will prove the slightly stronger statement in Theorem 4.3.1, which is equivalent to Theorem 1.2.5 when $t = 1$.

Theorem 4.3.1. *For all $\epsilon > 0$ and $r \in \mathbb{N}$, there exists a constant C such that, for $t \in \mathbb{N}$, if $p > \frac{C}{n^t}$ then with high probability, the largest $\mathcal{V}_{r+1}$-free set in $\mathcal{P}(n,p)$ has size at most $(1+\epsilon)pMt$.*

Notice that for this theorem we will be dealing with $\mathcal{P}(n,p)$ rather than $\mathcal{B}_n$. Because of this, we will work with the comparability digraph of $\mathcal{P}(n,p)$, denoted $G_{(n,p)}$, rather than the comparability digraph G of $\mathcal{B}_n$. However, we will still apply Lemma 4.2.1 to G and obtain a set of containers in $\mathcal{P}(n)$. Then, we will explore the implications that come from assuming a large portion of the container is randomly selected to be present in $\mathcal{P}(n,p)$. In particular, we will show that the probability of a large $\mathcal{V}_{r+1}$-free set I, and thus most of a large container C with $I \subset C$, being present in $G_{(n,p)}$ approaches 0 as n tends to infinity. It should be noted that Theorem 4.3.1 is similar to Theorem 2 in [9].

Proof of Theorem 4.3.1. Let $\epsilon > 0$ and $r, t \in \mathbb{N}$ be fixed. Assume that $\epsilon < \frac{1}{(2t)^{t+1}}$. Let $C := (r+1) \cdot 10^{10}\epsilon^{-5}$ and $\epsilon_1 := \epsilon/4$. Let $G_{(n,p)}$ be the graph formed from the comparability digraph G of $\mathcal{B}_n$ by selecting vertices independently at random with probability $p > C/n^t$.

To derive a contradiction, suppose that $G_{(n,p)}$ contains some $\mathcal{V}_{r+1}$-free set I such that $|I| > (1+\epsilon)pMt$. Apply Lemma 4.2.1, the $\mathcal{V}_{r+1}$-free Container Lemma, with ϵ_1 playing the role of ϵ. By this container lemma, we know there are disjoint sets S_1, S_2, and $g(S_1 \cup S_2)$ in $\mathcal{P}(n)$ such that $S_1 \cup S_2 \subseteq I \subseteq S_1 \cup S_2 \cup g(S_1 \cup S_2)$ and

$$|S_1 \cup S_2| \leq \frac{(r+1)(t+2)M}{\epsilon_1^2 n^t} < \frac{(r+1)pMt + 2(r+1)Mp}{\epsilon_1^2 C} \leq (1+\epsilon)pMt < |I|.$$

Since all of I is present in $G_{(n,p)}$, so too are the sets S_1 and S_2, as well as an additional $|I| - |S_1 \cup S_2| \geq (1 + \frac{\epsilon}{2})pMt$ elements of $g(S_1 \cup S_2)$. This is because

$$|I| - |S_1 \cup S_2| \geq (1+\epsilon)pMt - \frac{(r+1)(t+2)M}{\epsilon_1^2 n^t},$$

and

$$\frac{\epsilon}{2}pMt \geq \frac{(r+1)(t+2)M}{\epsilon_1^2 n^t},$$

when $p > C/n^t$ and $C = (r+1)10^{10}\epsilon^{-5}$.

The number of possible sets S_1 is $\binom{2^n}{\leq (r+1)2^n n^{-(t+0.9)}}$, and the probability that a given set S_1 is in $G_{(n,p)}$ is $p^{|S_1|}$. Further, since for a fixed S_1 we have $|f(S_1)| \leq (t+1+\epsilon_1)M \leq (t+2)M$, there are at most $\binom{(t+2)M}{\leq \frac{(r+1)(t+2)M}{\epsilon_1^2 n^t}}$ possibilities for S_2 and the probability that S_2 is in $G_{(n,p)}$ is $p^{|S_2|}$. For a fixed pair of sets S_1 and S_2, $|g(S_1 \cup S_2)| \leq (t+\epsilon_1)M \leq (1+\frac{\epsilon}{4})Mt$. So the expected number of vertices of $g(S_1 \cup S_2)$ which are present in $G_{(n,p)}$ is $\leq (1+\frac{\epsilon}{4})pMt$.

Let $v_1, \ldots, v_{|g(S_1 \cup S_2)|}$ be an enumeration of $g(S_1 \cup S_2)$ and let $X_1, \ldots, X_{|g(S_1 \cup S_2)|}$ be independent 0-1 random variables such that

$$X_i = \begin{cases} 1 & \text{if } v_i \in G_{(n,p)} \\ 0 & \text{if } v_i \notin G_{(n,p)}. \end{cases}.$$

Then $X = \sum X_i$ is a random variable representing the number of elements of $g(S_1 \cup S_2)$ present in $G_{(n,p)}$, and $\mathbb{E}(X) = |g(S_1 \cup S_2)|p$. So we may use a Chernoff bound to find

$$\begin{aligned} Pr\left(X \geq \left(1+\frac{\epsilon}{2}\right)pMt\right) &= Pr\left(X \geq \left(1+\frac{\epsilon/4}{1+\epsilon/4}\right)\left(1+\frac{\epsilon}{4}\right)pMt\right) \\ &\leq \exp\left(-\frac{\left(\frac{\epsilon/4}{1+\epsilon/4}\right)^2}{2+\frac{\epsilon/4}{1+\epsilon/4}}\left(1+\frac{\epsilon}{4}\right)pMt\right) \\ &= \exp\left(-\frac{\epsilon^2 pMt}{32+12\epsilon}\right) \\ &\leq \exp\left(-\frac{\epsilon^2 pMt}{100}\right), \end{aligned}$$

i.e. the probability that at least $(1+\frac{\epsilon}{2})pMt$ vertices from $g(S_1 \cup S_2)$ are present in $G_{(n,p)}$ is at most $\exp\left(-\frac{\epsilon^2 pMt}{100}\right)$.

Next, we will present an upper bound for the probability that $G_{(n,p)}$ has a $\mathcal{V}_{r+1}$-free set I with $|I| > (1+\epsilon)pMt$. Let this probability be represented by Π. We can derive an upper bound by adding up the probabilities that all of S_1 and S_2, and at least $(1+\frac{\epsilon}{2})pMt$ vertices from $g(S_1 \cup S_2)$ are present in $G_{(n,p)}$ over all possible pairs of sets S_1 and S_2. Symbolically, this gives us

$$
\begin{aligned}
\Pi &\leq \sum_{|S_1|}\sum_{|S_2|}\binom{2^n}{|S_1|}p^{|S_1|}\cdot\binom{(t+2)M}{|S_2|}p^{|S_2|}\cdot\exp\left(-\frac{\epsilon^2 pMt}{100}\right)\\
&\leq \left((r+1)2^n n^{-(t+0.9)}+1\right)\left(\frac{(r+1)(t+2)M}{\epsilon_1^2 n^t}+1\right)\cdot\max_{|S_1|}\left\{\binom{2^n}{|S_1|}\right\}\cdot p^{(r+1)2^n n^{-(t+0.9)}}\\
&\quad\cdot\max_{|S_2|}\left\{\binom{(t+2)M}{|S_2|}\right\}\cdot p^{\frac{(r+1)(t+2)M}{\epsilon_1^2 n^t}}\cdot\exp\left(-\frac{\epsilon^2 pMt}{100}\right)\\
&\leq \left((r+1)2^n n^{-(t+0.9)}+1\right)\left(\frac{(r+1)(t+2)M}{\epsilon_1^2 n^t}+1\right)\binom{2^n}{(r+1)2^n n^{-(t+0.9)}}p^{(r+1)2^n n^{-(t+0.9)}}\\
&\quad\cdot\binom{(t+2)M}{\frac{(r+1)(t+2)M}{\epsilon_1^2 n^t}}p^{\frac{(r+1)(t+2)M}{\epsilon_1^2 n^t}}\cdot\exp\left(-\frac{\epsilon^2 pMt}{100}\right)
\end{aligned}
$$

Recall that $t \in \mathbb{N}$, $\epsilon < \frac{1}{(2t)^{t+1}}$, $\epsilon_1 = \epsilon/4$, $C = (r+1)\cdot 10^{10}\epsilon^{-5}$, $p > \frac{C}{n^t}$ and n is sufficiently large. Using this information, we will show that this upper bound for Π is $o(1)$. The product of the first pair of terms can be bounded in the following way:

$$
\begin{aligned}
\left((r+1)2^n n^{-(t+0.9)}+1\right)\left(\frac{(r+1)(t+2)M}{\epsilon_1^2 n^t}+1\right) &\leq \exp\left((r+1)2^n n^{-(t+0.9)}+\frac{(r+1)(t+2)M}{\epsilon_1^2 n^t}\right)\\
&\leq \exp\left(\frac{\epsilon^2 pMt}{400}\right).
\end{aligned}
$$

We can derive the same upper bound for the product of the second pair of terms:

$$
\begin{aligned}
\binom{2^n}{(r+1)2^n n^{-(t+0.9)}}p^{(r+1)2^n n^{-(t+0.9)}} &\leq \left(\frac{2^n ep}{(r+1)2^n n^{-(t+0.9)}}\right)^{(r+1)2^n n^{-(t+0.9)}}\\
&= \exp\big((1+\log p+(t+0.9)\log n-\log(r+1))\\
&\quad\cdot((r+1)2^n n^{-(t+0.9)})\big)\\
&\lesssim \exp\left(\frac{0.9\epsilon^3\log n}{n^{0.4}10^{10}}\epsilon^2 pMt\right)\\
&\leq \exp\left(\frac{\epsilon^2 pMt}{400}\right).
\end{aligned}
$$

And we can again derive the same upper bound for the product of the third pair of terms:

$$\begin{aligned}\binom{(t+2)M}{\frac{(r+1)(t+2)M}{\epsilon_1^2 n^t}} p^{\frac{(r+1)(t+2)M}{\epsilon_1^2 n^t}} &\le \left(\frac{ep\epsilon^2 n^t}{(r+1)4^2}\right)^{\frac{(r+1)(t+2)M}{\epsilon_1^2 n^t}} \\ &= \exp\big((1+\log p + 2\log\epsilon + t\log n - \log((r+1)4^2)) \\ &\quad \cdot \left(\frac{(r+1)(t+2)M}{\epsilon_1^2 n^t}\right)\big) \\ &\lesssim \exp\left(\frac{\epsilon\log(1/\epsilon)}{10^{10}}\epsilon^2 pMt\right) \\ &\le \exp\left(\frac{\epsilon^2 pMt}{400}\right).\end{aligned}$$

Combining these inequalities with the current bound for Π gives us

$$\Pi \le \exp\left(-\frac{\epsilon^2 pMt}{400}\right),$$

which implies that Π is $o(1)$. Thus, with high probability, the largest $\mathcal{V}_{r+1}$-free set in $G_{(n,p)}$ has size at most $(1+\epsilon)pMt$, as desired. □

4.4 Conclusions and future work

Now that the random $\mathcal{V}_{r+1}$-free theorem is proved, there are a number of different research questions that could be considered next. First and foremost, it may be interesting to attempt the same problem using a different graph reformulation. In particular, it could be enlightening to encode $\mathcal{B}_n$ into an $(r+2)$-graph where each hyperedge corresponds to a copy of $\mathcal{V}_{r+1}$. This would alter the proof in that our $\mathcal{V}_{r+1}$-free families would actually be independent sets in hypergraphs, rather than sparse sets in simple graphs. Then instead of applying a container-building algorithm, we would need to prove a useful balanced supersaturation theorem to reach the same conclusion. This sort of result would say that a family with more than $(1+o(1))M$ elements contains many $(r+1)$-forks and those forks can be chosen to be evenly distributed over $\mathcal{B}_n$.

Another natural next step would be to prove random analogs of extremal results for other posets, such as the butterfly $\bowtie$. However, one new obstacle presented by this example is that encoding the problem in a comparability graph or digraph simply will not work. The butterfly is a height 2 poset with at least four relations. This means a $\bowtie$-free family in a comparability graph is not required to even be sparse, let alone an independent set. Encoding the poset problem into a 4-graph where each edge corresponds to a copy of $\bowtie$ may work, but has thus far been resistant to proof. The sticking point is again that we need to prove a balanced supersaturation theorem for this situation. A supersaturation result for the butterfly was recently proved by Patkós [45], but it is unhelpful for us because we need to consider larger families in $\mathcal{P}(n)$ than those it addresses. To gain more insight into this problem, it may be useful to examine the techniques used by Balogh and Wagner [11] in their proof about union-free families and attempt to apply them to the $\bowtie$ 4-graph.

Though it is difficult to find an upper bound for a maximum random $\bowtie$-free family, it is possible to prove a lower bound. This bound is presented below in Proposition 4.4.1.

Proposition 4.4.1. *If $p \ll 1/n^{4/3}$, then with high probability, the largest $\bowtie$-free family in $\mathcal{P}(n,p)$ has size at least $(3-o(1))pM$.*

Before we prove this result, consider three consecutive layers of the Boolean lattice, $L_{k-1} := \binom{[n]}{k-1}$, $L_k := \binom{[n]}{k}$, and $L_{k+1} := \binom{[n]}{k+1}$. If a $\bowtie$ exists within these three layers, it will look like either an "Up Y", a "Down Y", or a true butterfly, as pictured in Figure 4.2.

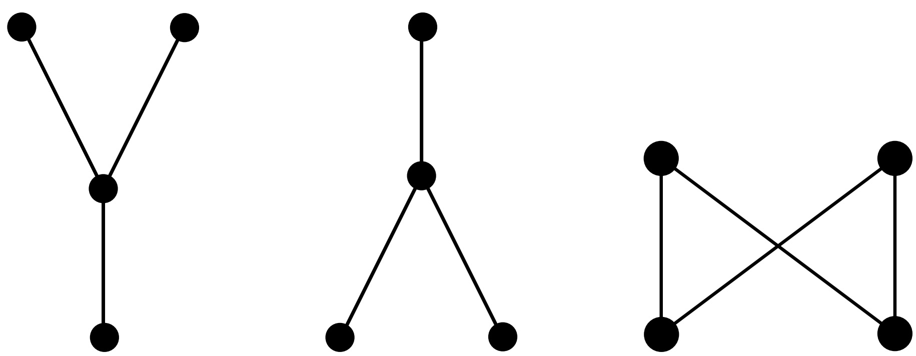

Figure 4.2 From left to right: The Up Y, Down Y, and true butterfly.

Notice that the Up Y and Down Y live in all three of L_{k-1}, L_k, and L_{k+1}, while the true butterfly must live in layers L_{k-1} and L_{k+1}, since it is impossible to have a butterfly in two consecutive layers.

Proof of Proposition 4.4.1. Let $p \ll 1/n^{4/3}$. Then it follows that $p \ll 1/n$. Let $\mathcal{F}$ be the set obtained from $L_{k-1} \cup L_k \cup L_{k+1}$ by selecting elements independently at random with probability p. According to Theorem 1.2.3, we have that with high probability, the sets $\mathcal{F} \cap (L_k \cup L_{k+1})$ and $\mathcal{F} \cap (L_{k-1} \cup L_k)$ will be antichains. We can deduce then that the number of Up Y-s and Down Y-s in $\mathcal{F}$ will approach zero as n tends to infinity.

Let $\beta(n)$ denote the number of true butterflies present in $\mathcal{F} \cap (L_{k-1} \cup L_{k+1})$. We can calculate the expected value of $\beta(n)$, $\mathbb{E}(\beta(n))$, in the following way. First, note that this expectation is the product of the number of butterflies in $L_{k-1} \cup L_{k+1}$ and the probability that the four elements forming each butterfly are present in $\mathcal{F}$. Each butterfly in $L_{k-1} \cup L_{k+1}$ can be found by choosing a set S in L_k, picking two elements in $[n] \setminus S$ to add to S, and picking two elements in S to remove from S. Further, the probability that any four elements are present in $\mathcal{F}$ is p^4. This gives us

$$\begin{aligned}
\mathbb{E}(\beta(n)) &= \binom{n}{k}\binom{n-k}{2}\binom{k}{2}p^4 \\
&\approx \binom{n}{k} \cdot \frac{1}{2}\left(\frac{n}{2}\right)^2 \cdot \frac{1}{2}\left(\frac{n}{2}\right)^2 \cdot p^4 \\
&= pM \cdot \frac{n^4 p^3}{2^6} \quad \text{when } k = \lfloor n/2 \rfloor.
\end{aligned}$$

Since $p \ll 1/n^{4/3}$, we have that $\frac{n^4 p^3}{2^6} \to 0$ as $n \to \infty$. So when $k = \lfloor n/2 \rfloor$, the expected number of true butterflies in $\mathcal{F}$ approaches zero as n tends to infinity. This means that with high probability, $\mathcal{F}$ is a $\bowtie$-free family in $\mathcal{P}(n,p)$ of expected size $3pM$. Therefore $(3 - o(1))pM$ is a lower bound for the size of a maximum $\bowtie$-free family in $\mathcal{P}(n,p)$, as desired.

$\square$

Notice that, as in Theorems 1.2.3, 1.2.4, and 1.2.6, Proposition 4.4.1 implies that we can again add approximately one additional layer of $\mathcal{P}(n,p)$ without encountering the forbidden subposet. Another interesting implication of Proposition 4.4.1 is that the threshold for the random butterfly-free problem may be $1/n^{4/3}$. This possibility leads to a number of questions which we present below.

Question 4.4.2. *Is $\hat{p} = 1/n^{4/3}$ the threshold for the largest $\bowtie$-free family in $\mathcal{P}(n,p)$ having size at most $(2 + o(1))pM$?*

A positive answer to Question 4.4.2 would be noteworthy because it would distinguish the random butterfly-free problem from the random chain-free and fork-free problems. In each of those cases, the threshold has already been determined to be $\hat{p} = 1/n$. Unfortunately, this question may prove difficult to answer directly. Because of this, we pose three additional questions.

Question 4.4.3. *If $p \gg 1/(n \log n)$, can we show that the largest $\bowtie$-free family in $\mathcal{P}(n,p)$ has size at most $(1+o(1))2pM$?*

A positive answer to Question 4.4.3 would not give us a value for the actual threshold for the random butterfly-free problem, but it would imply that $1/n$ is *not* the threshold. However, a positive answer to Question 4.4.4 would do both; it would show that $1/n^{4/3}$ is the threshold and $1/n$ is not.

Question 4.4.4. *If $p \gg 1/n^{4/3}$, can we show that the largest $\bowtie$-free family in $\mathcal{P}(n,p)$ has size at most $(1+o(1))2pM$?*

If Question 4.4.4 is indeed too difficult to answer, we offer a weaker different bound to consider in Question 4.4.5. We conjecture that the answer to this question is yes and note that, like Question 4.4.3, a positive answer here will imply that $1/n$ is not the threshold for the random butterfly-free problem.

Question 4.4.5. *If $p \gg (\log n)/n^{4/3}$, can we show that the largest $\bowtie$-free family in $\mathcal{P}(n,p)$ has size at most $(1+o(1))2pM$?*

Let $\hat{p}(n)$ denote the correct threshold for the random butterfly-free problem and consider the limit

$$\lim_{n\to\infty} \frac{-\log \hat{p}(n)}{\log n}. \tag{4.3}$$

If the answer to Question 4.4.2 is yes, as we believe, then the limit presented in (4.3) should be equal to 4/3. If not, we ask the following question:

Question 4.4.6. *Does* $\liminf\limits_{n\to\infty} \dfrac{-\log \hat{p}(n)}{\log n} = \limsup\limits_{n\to\infty} \dfrac{-\log \hat{p}(n)}{\log n}$*?*

We refer to the value of $\liminf_{n\to\infty} \frac{-\log \hat{p}(n)}{\log n}$ as the "fragility" of the extremal configuration. Larger values for this partial limit should indicate that the extremal structure associated with the given problem is more specific and easily disrupted.

BIBLIOGRAPHY

[1] N. Alon and A. Kostochka, *Hypergraph list coloring and Euclidean Ramsey theory*, Random Structures Algorithms **39** (2011), no. 3, 377–390. MR 2816937

[2] M. Axenovich, J. Manske, and R.R. Martin, *Q_2-free families in the Boolean lattice*, Order **29** (2012), no. 1, 177–191. MR 2948756

[3] J. Balogh, N. Bushaw, M. Collares Neto, H. Liu, R. Morris, and M. Sharifzadeh, *The typical structure of graphs with no large cliques*, ArXiv e-prints:1406.6961 (2014).

[4] J. Balogh, S. Das, M. Delcourt, H. Liu, and M. Sharifzadeh, *Intersecting families of discrete structures are typically trivial*, J. Combin. Theory Ser. A **132** (2015), 224–245. MR 3311345

[5] J. Balogh, P. Hu, B. Lidický, and H. Liu, *Upper bounds on the size of 4- and 6-cycle-free subgraphs of the hypercube*, European J. Combin. **35** (2014), 75–85. MR 3090487

[6] J. Balogh, H. Liu, Š. Petříčková, and M. Sharifzadeh, *The typical structure of maximal triangle-free graphs*, Forum Math. Sigma **3** (2015), e20, 19pp. MR 3482269

[7] J. Balogh, H. Liu, M. Sharifzadeh, and A. Treglown, *The number of maximal sum-free subsets of integers*, Proc. Amer. Math. Soc. **143** (2015), no. 11, 4713–4721. MR 3391030

[8] J. Balogh, R. Morris, and W. Samotij, *Independent sets in hypergraphs*, J. Amer. Math. Soc. **28** (2015), no. 3, 669–709. MR 3327533

[9] J. Balogh, R. Mycroft, and A. Treglown, *A random version of Sperner's theorem*, J. Combin. Theory Ser. A **128** (2014), 104–110. MR 3265918

[10] J. Balogh, A. Treglown, and A. Zsolt Wagner, *Applications of graph containers in the Boolean lattice*, ArXiv e-prints:1602.05870 (2016).

[11] J. Balogh and A. Zsolt Wagner, *On the number of union-free families*, ArXiv e-prints:1601.03659 (2016).

[12] B. Bollobás, *On generalized graphs*, Acta Math. Acad. Sci. Hungar **16** (1965), 447–452. MR 0183653

[13] G. Burosch, J. Demetrovics, G. O. H. Katona, D. J. Kleitman, and A. A. Sapozhenko, *On the number of databases and closure operations*, Theoret. Comput. Sci. **78** (1991), no. 2, (Part A), 377–381. MR 1095988

[14] P. J. Cameron and P. Erdős, *On the number of sets of integers with various properties*, Number theory (Banff, AB, 1988), de Gruyter, Berlin, 1990, pp. 61–79. MR 1106651

[15] ———, *Notes on sum-free and related sets*, Combin. Probab. Comput. **8** (1999), no. 1-2, 95–107, Recent trends in combinatorics (Mátraháza, 1995). MR 1684624

[16] M. Collares Neto and R. Morris, *Maximum-size antichains in random sets*, ArXiv e-prints:1404.5258 (2014).

[17] D. Conlon and W. T. Gowers, *Combinatorial theorems in sparse random sets*, ArXiv e-prints:1011.4310 (2010).

[18] A. De Bonis and G. O. H. Katona, *Largest families without an r-fork*, Order **24** (2007), no. 3, 181–191. MR 2358080 (2008m:05294)

[19] A. De Bonis, G. O. H. Katona, and K. J. Swanepoel, *Largest family without $A \cup B \subseteq C \cap D$*, J. Combin. Theory Ser. A **111** (2005), no. 2, 331–336. MR 2156217

[20] J.-M. Deshouillers, G. A. Freiman, V. Sós, and M. Temkin, *On the structure of sum-free sets. II*, Astérisque (1999), no. 258, xii, 149–161, Structure theory of set addition. MR 1701193

[21] P. Erdős, *On a lemma of Littlewood and Offord*, Bull. Amer. Math. Soc. **51** (1945), 898–902. MR 0014608 (7,309j)

[22] P. Erdős, A. L. Rubin, and H. Taylor, *Choosability in graphs*, Proceedings of the West Coast Conference on Combinatorics, Graph Theory and Computing (Humboldt State Univ., Arcata, Calif., 1979), Congress. Numer., XXVI, Utilitas Math., Winnipeg, Man., 1980, pp. 125–157. MR 593902

[23] B. Green, *The Cameron-Erdős conjecture*, Bull. London Math. Soc. **36** (2004), no. 6, 769–778. MR 2083752

[24] ———, *A Szemerédi-type regularity lemma in abelian groups, with applications*, Geom. Funct. Anal. **15** (2005), no. 2, 340–376. MR 2153903

[25] J. R. Griggs and G. O. H. Katona, *No four subsets forming an N*, J. Combin. Theory Ser. A **115** (2008), no. 4, 677–685. MR 2407920

[26] J. R. Griggs and W.-T. Li, *The partition method for poset-free families*, J. Comb. Optim. **25** (2013), no. 4, 587–596. MR 3044160

[27] J. R. Griggs, W.-T. Li, and L. Lu, *Diamond-free families*, J. Combin. Theory Ser. A **119** (2012), no. 2, 310–322. MR 2860596

[28] J. R. Griggs and L. Lu, *On families of subsets with a forbidden subposet*, Combin. Probab. Comput. **18** (2009), no. 5, 731–748. MR 2534266

[29] D. Grósz, A. Methuku, and C. Tompkins, *An upper bound on the size of diamond-free families of sets*, ArXiv e-prints:1601.06332 (2016).

[30] P. Haxell and M. Pei, *On list coloring Steiner triple systems*, J. Combin. Des. **17** (2009), no. 4, 314–322. MR 2527171

[31] P. Haxell and J. Verstraëte, *List coloring hypergraphs*, Electron. J. Combin. **17** (2010), no. 1, Research Paper 129, 12. MR 2729378

[32] S. Janson, T. Łuczak, and A. Rucinski, *Random graphs*, Wiley-Interscience Series in Discrete Mathematics and Optimization, Wiley-Interscience, New York, 2000. MR 1782847

[33] G. O. H. Katona, *Intersection theorems for systems of finite sets*, Acta Math. Acad. Sci. Hungar **15** (1964), 329–337. MR 0168468

[34] D. Kleitman, *A conjecture of Erdős-Katona on commensurable pairs among subsets of an n-set*, Theory of Graphs (Proc. Colloq., Tihany, 1966), Academic Press, New York, 1968, pp. 215–218. MR 0232685 (38 #1008)

[35] D. Kleitman and K. J. Winston, *The asymptotic number of lattices*, Ann. Discrete Math. **6** (1980), 243–249, Combinatorial mathematics, optimal designs and their applications (Proc. Sympos. Combin. Math. and Optimal Design, Colorado State Univ., Fort Collins, Colo., 1978). MR 593536

[36] ______, *On the number of graphs without* 4*-cycles*, Discrete Math. **41** (1982), no. 2, 167–172. MR 676877

[37] L. Kramer, R.R. Martin, and M. Young, *On diamond-free subposets of the Boolean lattice*, J. Combin. Theory Ser. A **120** (2013), no. 3, 545–560. MR 3007136

[38] D. Lubell, *A short proof of Sperner's lemma*, J. Combinatorial Theory **1** (1966), 299. MR 0194348

[39] T. Łuczak and T. Schoen, *On the number of maximal sum-free sets*, Proc. Amer. Math. Soc. **129** (2001), no. 8, 2205–2207 (electronic). MR 1823901

[40] J. Manske and J. Shen, *Three layer Q_2-free families in the Boolean lattice*, Order **30** (2013), no. 2, 585–592. MR 3063206

[41] L. D. Mešalkin, *A generalization of Sperner's theorem on the number of subsets of a finite set*, Teor. Verojatnost. i Primenen **8** (1963), 219–220. MR 0150049

[42] R. Morris and D. Saxton, *The number of C_2l-free graphs*, ArXiv e-prints:1309.2927 (2013).

[43] R. Nenadov, A. Steger, and M. Stojaković, *On the threshold for the Maker-Breaker H-game*, ArXiv e-prints:1401.4384 (2014).

[44] D. Osthus, *Maximum antichains in random subsets of a finite set*, J. Combin. Theory Ser. A **90** (2000), no. 2, 336–346. MR 1757282 (2001c:06002)

[45] B. Patkós, *Supersaturation and stability for forbidden subposet problems*, J. Combin. Theory Ser. A **136** (2015), 220–237. MR 3383276

[46] V. Rödl, A. Ruciński, and M. Schacht, *An exponential-type upper bound for Folkman numbers*, ArXiv e-prints:1603.00521 (2016).

[47] A. A. Sapozhenko, *The Cameron-Erdős conjecture*, Discrete Math. **308** (2008), no. 19, 4361–4369. MR 2433862

[48] D. Saxton and A. Thomason, *List colourings of regular hypergraphs*, Combin. Probab. Comput. **21** (2012), no. 1-2, 315–322. MR 2900065

[49] ______, *Hypergraph containers*, Invent. Math. **201** (2015), no. 3, 925–992. MR 3385638

[50] M. Schacht, *Extremal results for random discrete structures*, ArXiv e-prints:1603.00894 (2016).

[51] E. Sperner, *Ein Satz über Untermengen einer endlichen Menge*, Math. Z. **27** (1928), no. 1, 544–548. MR 1544925

[52] E. Szemerédi, *On sets of integers containing no k elements in arithmetic progression*, Acta Arith. **27** (1975), 199–245, Collection of articles in memory of Juriĭ Vladimirovič Linnik. MR 0369312

[53] V. G. Vizing, *Coloring the vertices of a graph in prescribed colors*, Diskret. Analiz (1976), no. 29 Metody Diskret. Anal. v Teorii Kodov i Shem, 3–10, 101. MR 0498216

[54] G. Wolfovitz, *Bounds on the number of maximal sum-free sets*, European J. Combin. **30** (2009), no. 7, 1718–1723. MR 2548662

[55] K. Yamamoto, *Logarithmic order of free distributive lattice*, J. Math. Soc. Japan **6** (1954), 343–353. MR 0067086