Problems in Extremal Set Theory

Balázs Patkós

Doctoral dissertation submitted to the Hungarian Academy of Sciences

Budapest, 2021

Contents

1	Intr	oduction	1
2	For	rbidden subposet problems	
	2.1	Complete multipartite posets	14
		2.1.1 Preliminaries	14
		2.1.2 Proofs of Theorems 7, 8, 9, 10	18
	2.2	Stability and supersaturation for the butterfly poset	24
	2.3	Generalized forbidden subposet problems	35
3	Res	ults concerning traces of sets	40
	3.1	Families avoiding chains as traces	45
	3.2	VC-Saturating families	53
		3.2.1 Random construction	55
		3.2.2 Explicit constructions	56
4	q-an	alog results	63
	4.1	Shadows of subspaces	67
	4.2	Intersecting families of subspaces and colorings of $q\operatorname{-Kneser}$ graphs $\ . \ .$	78

1 Introduction

Extremal combinatorics is that area of mathematics which is concerned with finding the largest, smallest, or otherwise optimal structures with a given property. When the structures in question are graphs or hypergraphs, we talk about extremal (hyper)graph theory - a topic initiated by Pál Turán in his famous paper [109] on the maximum number of edges in K_r -free graphs. [109] contained many problems some of which are still unsolved eighty years later. If the structures addressed are collections, *families* of subsets of a finite ground set, we dive into the area of extremal finite set theory. The two milestone results of this topic are the theorems of Sperner [107] and that of Erdős, Ko, and Rado [36] that determine the maximum number of sets in an antichain of $2^{[n]}$ and an intersecting family in $\binom{[n]}{r}$, respectively. With Dániel Gerbner, we devoted a complete book [53] to gather the most important developments of the last almost-acentury in the area of extremal finite set theory.

This dissertation concentrates on three subtopics and is based on papers [10, 22, 50, 95, 96, 97, 98, 100, 99, 39]. Here below, we describe the main findings and we present the results in more details and include the context and preceeding results in the last three sections.

In the paper [34] appearing in a conference proceedings, Pál Erdős and Daniel Kleitman summarized the operations that are natural to consider for problems on (extremal) set systems (four years later, the paper was published in a peer reviewed journal [35], and on the request of Erdős¹, it also appeared in the book containing selected writings of Erdős published for his sixtieth birthday): (A) intersection, (B) union, (C) disjointness, (D) complement, (E) containment, (F) rank (size). All later three sections of this dissertation address problems defined via these natural operations and their vector space analogs.

In Section 2, we consider so-called forbidden subposet problems - an area that deals with questions formulated with operation (E) from the above list. Erdős [32] generalized Sperner's theorem about antichains to set families not containing k nested sets (a k-chain). Then in the 80s, Katona and Tarján [71] introduced a framework to deal with forbidden *containment* patterns. A poset P has a copy in a set family \mathcal{F} if we can find sets from \mathcal{F} corresponding to elements of P such that if $p \leq_P p'$, then the set corresponding to p is contained in the set corresponding to p'. If these sets are in containment if and only if p < p', then we talk about strong copies, otherwise the copy is weak. A family \mathcal{F} is weak *P*-free if \mathcal{F} does not contain either weak or strong copies of P, while \mathcal{F} is strong P-free if it does not contain strong copies of \mathcal{F} . The extremal number of the largest possible size of a weak or strong P-free family in $2^{[n]}$ is denoted by La(n, P) and $La^*(n, P)$, respectively. Observe that Sperner's theorem corresponds to the case $P = P_2$, and Erdős's generalization is the case $P = P_k$, where P_k is the path poset on k elements, i.e. the totally ordered set on k elements. For any poset Pone can consider the most number e(P) and $e^*(P)$ of middle levels of $2^{[n]}$ that one can have without creating a weak or a strong copy of P. It is widely believed that this trivial construction is asymptotically optimal for any poset, i.e. La(n, P) = (e(P) + e(P))

¹a piece of information from Miklós Simonovits

 $o(1))\binom{n}{\lfloor n/2 \rfloor}$ and $La^*(n, P) = (e^*(P) + o(1))\binom{n}{\lfloor n/2 \rfloor}$ hold. This conjecture was folklore among researchers of the area but appeared explicitly only in [17, 60]. In the past more than 35 years, the conjecture has been verified for many classes of posets. Probably the most important result is due to Bukh [17] in the weak case and to Boehnlein and Jiang [11] in the strong case settling the conjecture for all tree posets. Our first contributions to the area are Theorem 7, Theorem 8, Theorem 9, and Theorem 10 that prove that both the weak and the strong versions of the conjecture are true for some infinite classes of complete multipartite posets.

Whenever an extremal problem has a unique extremal structure, the next step is to decide whether the problem admits stability. In general, this means that an object with almost extremal size must be very close to the extremal one in structure. We obtained such a stability result, Theorem 12 for butterfly-free set families. We apply this stability theorem to prove a supersaturation result, Theorem 13 that determines the minimum number of copies of the butterfly poset in a set family over all families of size La(n, B) + E if $E = 2^{o(n)}$.

At the end of this section, we address the so-called generalized forbidden subposet problems. In line with research initiated by Alon and Schickelman [4] on generalized Turán problems of graphs, with Balázs Keszegh and Dániel Gerbner [49], we denoted for two posets P and Q the maximum number of weak copies of Q in a P-free family $\mathcal{F} \subseteq 2^{[n]}$ by La(n, Q, P) and obtained the first results on this quantity for some "ad hoc" pairs of posets. Already these results showed that there is a wider variety of (asymptotically) extremal families for the genrealized question than for the original one: non-consecutive levels can maximize La(n, Q, P), and in some cases unions of full levels are not even asymptotically optimal.

In [50], with Dániel Gerbner, Abhishek Methuku, Dániel Nagy, and Máté Vizer, we started investigating the generalized forbidden subposet problem more systematically. Just as cliques are the simplest subgraphs to count, totally ordered subposets (chains) are the simplest subposets to address. Our investigations resulted in observing a phenomenon somewhat similar to original forbidden subgraph problem: the asymptotics of most number of edges that *n*-vertex *F*-free graph can have is determined by the celebrated Erős-Stone-Simonovits theorem [38] if the chromatic number of *F* is at least three, while there are lots of open problems for the so-called degenerate case of *F* being bipartite (a survey on the topic is [48]). When we count copies of P_k , then the height h(P) of the forbidden subposet *P* takes the role of the chromatic number. Theorem 16 determines the order of magnitude of $La(n, P_k, P)$ if the height of *P* is strictly larger than *k* and gives some upper and lower bounds if the height is at most *k*.

Section 3 is devoted to results on traces of set families. The trace of a family \mathcal{F} on the set S is the collection of all *intersections* $F \cap S$ where F ranges over all sets in \mathcal{F} , so this concept uses operation (A) from the list of Erdős and Kleitman. The starting point of this branch of extremal finite set theory is the so-called Sauer Lemma (proved independently by Sauer [103], Shelah [104] and Vapnik and Chervonenkis [110]) that states that a family $\mathcal{F} \subseteq 2^{[n]}$ that does not shatter any (k + 1)-subet of [n] has size at most $\sum_{i=0}^{k} {n \choose i}$.

The uniform version of this problem was first studied by Frankl and Pach [45] who

conjectured that the maximum size of a k-uniform family in $\binom{[n]}{k}$ that des not shatter a set of size k is $\binom{n-1}{k-1}$. This was disproved by Ahlswede and Khatchatrian [2] and later Mubayi and Zhao [90] found an exponential number of counterexamples. In [96], we showed how to strengthen the condition of not shettering a set of size k so that the bound conjectured by Frankl and Pach become valid.

In the original non-uniform Sauer-lemma, the extremal families are not yet characterized. In [95], we showed that the same strengthening of the condition as in the uniform case reduces the number of extremal families to 2: two families that are complements of each other. The new condition is as follows: there should not exist a k-set S such that the trace of the family on S contains a chain of length k+1, i.e a maximal chain. This led us to introduce the notion of l-trace k-Sperner families: families for which there does not exist any l-set S such that the trace of the family on S contains a chain longer than k. This notion relates the topic of set traces to the area of Sperner theorems. If l is a constant, then, as proved in [95], the bound of the Sauer-lemma holds. As we proved in the papers [97, 98], if l is close to the size of the ground set of the family, then results of the flavor of Erdős's generalization of Sperner's theorem can be obtained. The proof of Theorem 46 uses two tools that show how topics of extremal combinatorics are interconnected: Bukh's already mentioned theorem on tree-poset free families is combined with a lemma (Lemma 57) that gives a bound on k-uniform set families avoiding tight paths of fixed length l, where k is a function of the size of the underlying set, as results in the literature are mostly concerned with the case when both k and l are fixed and the underlying set is large enough.

It is quite common to consider the saturation counterpart of extremal problems. While an extremal problem ask for the maximum size of a combinatorial structure that possesses some prescribed property, saturation results address the problem of determining the smallest possible size of a combinatorial structure that has the property but is unextendable in any way so that the prescirbed property is maintained. The property appeearing in Sauer's Lemma is not shattering a set of size k + 1. If k equals 0, it is trivial that any family consisting of a single set possesses this property, while any pair of distinct sets shatter a singleton. Dudley [28] showed that any family $\mathcal{F} \subseteq 2^{[n]}$ that is maximal with respect to the property of not shattering any 2-sets, has size n+1. With Nóra Frankl, Sergey Kiselev, and Andrey Kupavksii we showed [39] that for any $k \geq 2$ there exist families $\mathcal{F} \subseteq 2^{[n]}$ of size bounded independently of n such that \mathcal{F} does not shatter any (k + 1)-set, but any extension of \mathcal{F} does.

In Section 4, we consider q-analogs of some well-known problems of extremal finite set theory. In cases of q-analogs, one considers a problem for set families and replaces sets by vector spaces over a finite field \mathbf{F}_q , subsets by subspaces, size by dimension and so the list of Erdős and Kleitman is changed to (B') jointly generated subpace (D') orthogonal subspace (F') dimension. In many cases, the obtained new problem can be solved very similarly to how the proof of the original problem works - these are the uninteresting cases. On the other hand, sometimes techniques for the set family problem are not strong enough to settle the q-analog and new and/or different methods are needed. Even more interestingly, it can happen that the extremal structure to the q-analog is a little or completely different from the ones in the original subset case. This phenomenon might make problems easier (the t-intersection problem for subspaces

was solved by Frankl and Wilson [46] more than ten years earlier than the subset tintersection problem was settled finally by Ahlswede and Khatchatrian [1]), but more often q-analogs are harder to tackle. The *shadow theorem* of Kruskal [77] and Katona [69] is a very much used statement in extremal combinatorics. Several proofs of the theorem are known almost all of which rely on the so-called shifting operation. Unfortunately, the definition of this operation cannot be extended to subspaces. Keevash [72] obtained a new and surprisingly simple proof of an approximate version of the shadow theorem due to Lovász [78]. The first q-analog result, Theorem 76 of the dissertation modifies the proof of Keevash to obtain an approximate shadow theorem for subspaces. As a consequence, we prove Theorem 78 on r-wise intersecting families of subspaces a q-analog of a result of Frankl [40].

The Hilton-Milner theorem [65] states what is the size of the second largest maximal intersecting family in $\binom{[n]}{r}$ provided $n \ge 2r+1$ holds. It can be read as a prototype stability theorem: if an intersecting family \mathcal{F} in $\binom{[n]}{r}$ has size larger than $\binom{n-1}{r-1} - \binom{n-r-1}{r-1} + 1$, then \mathcal{F} must be a subfamily of the unique largest intersecting family determined by the Erdős-Ko-Rado theorem. The bound comes from the family consisting of all rsets containing a fixed element x and meeting a fixed r-set R together with the set R itself. When considering the q-analog of this problem, one would think that the corresponding family of r-dimensional subspaces consists of all r-subspaces containing a fixed 1-dimensional subspace E and having non-trivial intersection with a fixed r-dimensional subspace R with $E \leq R$. But this is not maximal, one can add all rsubspaces of $\langle R, E \rangle$. In a seven authored paper, we were able to prove Theorem 80 that states that this modified family is indeed the second largest intersecting family in $\begin{bmatrix} V \\ r \end{bmatrix}$ where V is an n-dimensional vector space over \mathbf{F}_q . Intersecting families in $\binom{[n]}{k}$ are in one-to-one correspondence with independent sets in the Kneser graph $Kn_{n:k}$, the vertex set of which is $\binom{[n]}{k}$ and vertices corresponding to F and G are joined with an edge if and only if $F \cap G$ are disjoint. The famous result of Lovász states [79] that the natural coloring (color F with its maximal element unless this element is smaller than 2k) is optimal.

As a corollary of Theorem 80, we could obtain Theorem 81 that determines the chromatic number of q-Kneser graphs (vertex set is the set of all k-subspaces of an n-dimensional vector space over \mathbf{F}_q and U, U' are joined by an edge if and only if dim $(U \cap U') = 0$). A slight alteration of the subset coloring is optimal: fixing an (n - k + 1)-dimensional subspace F, one colors U by any 1-dimensional subspace of $F \cap U$.

Notation, definitions, terminology

Sets: for any set X we denote by 2^X , $\binom{X}{k}$, $\binom{X}{\leq k}$ and $\binom{X}{\geq k}$ the power set of X, the family of all k-subsets pf X, the family of all subsets of \overline{X} of size at least k and the family of all subsets of X of size at most k, respectively. We write [n] to denote the set of the first n positive integers. For two sets $A \subseteq B$, we denote by [A, B] the *interval* $\{F : A \subseteq F \subseteq B\}$. If the underlying set [n] is given, then \overline{A} denotes the complement $[n] \setminus A$ of A, and for a family $\mathcal{F} \subseteq 2^{[n]}$, we write $\overline{\mathcal{F}}$ to denote $\{\overline{F} : F \in \mathcal{F}\}$.

Partially ordered sets: we will always use the short term *poset* and with a little abuse of notation we will identify the poset with its set of elements and, if necessary,

define its partial ordering when introducing the poset. So we will write P-free instead of (P, \leq) -free, etc.

Vector spaces: The finite field of q elements is denoted by \mathbf{F}_q . The set of all k-dimensional subspaces of vector space V is denoted by $\begin{bmatrix} V \\ k \end{bmatrix}$ jelöli. For the number of all k-dimensional subspaces of an n-dimensional vector space over \mathbf{F}_q , we use Gaussian q-binomial coefficient

$$\begin{bmatrix} n \\ k \end{bmatrix}_q := \prod_{0 \le i < k} \frac{q^{n-i} - 1}{q^{k-i} - 1}.$$

The index q is often omitted if its value is clear from context. Both binomial and Gaussian coefficients can be extended to real numbers: $\binom{x}{k} = \prod_{0 \le i < k} \frac{x-i}{k-i}$ illetve $\begin{bmatrix} x \\ k \end{bmatrix}_q := \prod_{0 \le i < k} \frac{q^{x-i}-1}{q^{x-i}-1}$.

2 Forbidden subposet problems

The earliest result of extremal finite set theory is the following theorem of Sperner.

Theorem 1 (Sperner [107]). If a family $\mathcal{F} \subseteq 2^{[n]}$ does not contain two sets F, F' with $F \subsetneq F'$, then $|\mathcal{F}| \leq {n \choose \lfloor n/2 \rfloor}$ holds. Moreover, the only families satisfying the property and achieving the maximum size are ${[n] \choose \lfloor n/2 \rfloor}$ and ${[n] \choose \lfloor n/2 \rfloor}$.

Families satisfying the condition of Theorem 1 are called *antichains* or Sperner families. Theorem 1 was generalized by Erdős who proved the following result: a chain \mathcal{C} of length k (a k-chain for short) is a k-tuple of nested sets $C_1 \subsetneq C_2 \subsetneq \cdots \subsetneq C_k$. A family \mathcal{F} of sets is said to be k-Sperner if all chains in \mathcal{F} have length at most k. We define $\Sigma(n,k)$ to be the sum of the k largest binomial coefficients of order n, i.e. $\Sigma(n,k) = \sum_{i=1}^{k} {\binom{n}{\lfloor \frac{n-k}{2} \rfloor + i}}$. Let $\Sigma^*(n,k)$ be the collection of families consisting of the corresponding full levels, i.e. if n + k is odd, then $\Sigma^*(n,k)$ contains one family $\bigcup_{i=1}^{k} {\binom{[n]}{\lfloor \frac{n-k}{2} \rfloor + i}}$, while if n + k is even, then $\Sigma^*(n,k)$ contains two families of the same size $\bigcup_{i=0}^{k-1} {\binom{[n]}{\frac{n-k}{2} + i}}$ and $\bigcup_{i=1}^{k} {\binom{[n]}{\frac{n-k}{2} + i}}$.

Theorem 2 (Erdős, [32]). If $\mathcal{F} \subseteq 2^{[n]}$ is a k-Sperner family, then $|\mathcal{F}| \leq \Sigma(n,k)$ holds. Moreover, if $|\mathcal{F}| = \Sigma(n,k)$, then $\mathcal{F} \in \Sigma^*(n,k)$.

Theorem 1 and Theorem 2 have many generalizations and applications. A not very recent monograph on Sperner theory is [29] and our book [53] devotes several chapters to this topic. As mentioned in the Introduction, Erdős and Kleitman in their 1974 paper [35] philosophized that all interesting extremal set theory problems can be formulated using the following six basic concepts: intersection, union, disjointness, somplement, containment, and size. The forbidden structures of Theorem 1 and Theorem 2 can be considered as containment patterns. Katona and Tarján intorduced [71] a framework in which such forbidden containment patterns can be dealt with in general. The containment patterns are described by posets. Any family $\mathcal{F} \subseteq 2^{[n]}$ of sets is a poset under the containment relation, so any property described in the language of partially ordered sets can be translated to a property of set families. We say that a subposet Q' of Q is a (weak) copy of P, if there exists a bijection $f: P \to Q'$, such that for any $p, p' \in P$, the relation $p <_P p'$ implies $f(p) <_Q f(p')$. We say that Q' is an strong copy of Q, if, in addition to this, $f(p) <_Q f(p')$ also implies $p <_P p'$. If a poset Q does not contain a copy of P, then it is (weak) P-free. Katona and Tarján [71] introduced the problem of determining

$$La(n, P) = \max\{|\mathcal{F}| : \mathcal{F} \subseteq 2^{[n]} \text{ is } P\text{-free}\}.$$

More generally, for a family \mathcal{P} of posets, one can consider $La(n, \mathcal{P}) = \max\{|\mathcal{F}| : \mathcal{F} \subseteq 2^{[n]} \text{ is } P\text{-free for all } P \in \mathcal{P}\}$. In this language, Theorem 2 can be stated as

$$La(n, P_{k+1}) = \sum_{i=1}^{k} \binom{n}{\lfloor \frac{n-k}{2} \rfloor + i} = \Sigma(n, k),$$

where P_k is the chain on k elements. As every poset P is a weak subposet of $P_{|P|}$, applying Theorem 2 yields $La(n, P) \leq (|P| - 1) \binom{n}{\lfloor n/2 \rfloor}$. On the other hand, for any

poset P containing a comparable pair, we have $La(n, P) \ge \binom{n}{\lfloor n/2 \rfloor}$ as shown by the family $\binom{[n]}{\lfloor n/2 \rfloor}$. Therefore, it is natural to introduce $\pi(P) = \lim_{n \to \infty} \frac{La(n, P)}{\binom{n}{\lfloor n/2 \rfloor}}$. It is still unknown whether $\pi(P)$ exists for every poset P.

The following conjecture was folklore for long time, before first being published in [60] and [17].

Conjecture 3. For any poset P, let e(P) denote the largest integer k such that for any j and n, the family $\bigcup_{i=1}^{k} {[n] \choose i+i}$ is P-free. Then $\pi(P)$ exists and is equal to e(P).

The conjecture has been verified for many specific posets [24, 25, 58] and several infinite classes of posets [60, 59, 80]. General bounds on La(n, P) involving the size and the height of P were obtained by Burcsi and Nagy [18], Chen and Li [20], and Grósz, Methuku, and Tompkins [61]. The most notable result of the area is due to Boris Bukh. The Hasse diagram H(P) of a poset P is a graph on vertex set P with p and q joined by an edge in H(P) if and only if $p <_P q$ and there does not exist $z \in P$ with $p <_P z <_P q$. A poset P is a *tree poset* if its Hasse diagram is a tree.

Theorem 4 (Bukh [17]). Conjecture 3 holds for any tree poset T. More precisely, $La(n,T) = (e(T) + O_T(\frac{1}{n})) \binom{n}{\lfloor n/2 \rfloor}$ holds for any tree poset T.

As we described above, there exist two versions of a copy of a poset P, the other being the strong copy. All previous definitions can be introduced for strong copies. A family is strong P-free if it does not contain any strong copies of P. The extremal number $La^*(n, P)$ is the maximum size that a strong P-free family $\mathcal{F} \subseteq 2^{[n]}$ can have. The strong analog of Conjecture 3 is the following.

Conjecture 5. Let P be a poset and let $e^*(P)$ denote the largest integer k such that for any j and n the family $\bigcup_{i=1}^k {[n] \choose j+i}$ is strong P-free. Then $\pi^*(P) = \lim_{n \to \infty} \frac{La^*(n,P)}{\binom{n}{\lfloor \frac{n}{2} \rfloor}}$ exists and is equal to $e^*(P)$.

The strong versions of forbidden subposet problems are somewhat less studied compared to their original counterparts. The first such results were obtained by Caroll and Katona [19]. Bukh's result, Theorem 4 was strengthened as follows. Note that for any tree poset T, we have $e(T) = e^*(T) = h(T) - 1$, where h(t) is the height of the poset T. However, the error term of the next theorem is weaker than that of Theorem 4.

Theorem 6 (Boehnlein, Jiang [11]). Conjecture 5 holds for any tree poset T. Equivalently, $La^*(n,T) = (e^*(T) + o(1)) \binom{n}{\lfloor n/2 \rfloor}$ holds for any tree poset T.

Let us mention that even determining the order of magnitude of $La^*(n, P)$ is far from trivial. (Remember, $La(n, P) = \Theta(\binom{n}{\lfloor n/2 \rfloor})$ is a trivial consequence of Theorem 2.) Lu and Milans proved [81] $La^*(n, P) = \Theta_P(\binom{n}{\lfloor n/2 \rfloor})$ for posets P of height 2. Finally, Methuku and Pálvölgyi settled [88] this problem by proving the same statement for arbitrary posets. Méroueh strengthened [84] their result and showed that for any poset P there exists a constant C_P such that the Lubell-mass of any strong P-free family is at most C_P . (For the definition of Lubell-mass and more details see Section 2.1.)

The smallest poset for which Conjecture 3 and Conjecture 5 are yet to be proven is the so-called diamond poset D_2 on four elements a, b, c, d with a < b, c < d. Griggs, Li, and Lu verified [59] Conjecture 3 for an infinite subclass of the generalized diamond posets D_s on s + 2 elements $a, b_1, b_2, \ldots, b_s, c$ with $a < b_1, b_2, \ldots, b_s < c$.

Let us now define the infinite classes of posets for which we verified Conjecture 3 and Conjecture 5. They generalize results of Griggs, Li, and Lu in the weak case and prove their counterparts in the strong case. Let the K_{r_1,r_2,\ldots,r_s} denote the *complete multi-level poset* on $\sum_{i=1}^{s} r_i$ elements $a_1^1, a_2^1, \ldots, a_{r_1}^1, a_1^2, a_2^2, \ldots, a_{r_2}^2, \ldots, a_1^s, a_2^s, \ldots, a_{r_s}^s$ with $a_{\alpha}^i < a_{\beta}^j$ if and only if i < j. The rank $r(a_l^i)$ of the element a_l^i is *i*. Our first result gives not only the asymptotics of $La^*(n, K_{r,t})$, but also the order of magnitude of the second order term of the extremal value.

Theorem 7 (Patkós [99]). For any positive integers $2 \le r, t$ we have $\Sigma(n, 2) + \left(\frac{r+t-2}{n} - O_{r,t}(\frac{1}{n^2})\right) \binom{n}{\lfloor n/2 \rfloor} \le La^*(n, K_{r,t}) \le \left(2 + \frac{2(r+t-2)}{n} + o(\frac{1}{n})\right) \binom{n}{\lfloor n/2 \rfloor}.$

Note that the same upper bound for $La(n, K_{r,t})$ follows from Theorem 4 as $K_{r,t}$ is a (strong) subposet of $K_{r,1,s}$ and $K_{r,1,s}$ is a tree poset. By the same argument, Theorem 6 implies the asymptotics of $La^*(n, K_{r,t})$ but its error term is worse than that of Theorem 7. Let us remark that $La(n, K_{2,2}) = \Sigma(n, 2)$ was shown by De Bonis, Katona, Swanepoel [25]. As they also showed the uniqueness of the extremal family, it was known that the strict inequality $La(n, K_{2,2}) < La^*(n, K_{2,2})$ holds. Theorem 7 tells us the order of magnitude of the gap between these two parameters.

Then we turn our attention to the three level case of $K_{r,s,t}$. To do so we need to introduce the following notation: for positive integers r, t let

$$f(r,t) = \begin{cases} 0 & \text{if } r = t = 1, \\ 1 & \text{if } r = 1, t > 1 & \text{or } r > 1, t = 1, \\ 2 & \text{if } r, t > 2. \end{cases}$$

Also, for any integer $s \ge 2$ let us define $m = m_s = \lceil \log_2(s - f(r, t) + 2) \rceil$ and $m^* = m_s^* = \min\{m : s \le \binom{m}{\lceil m/2 \rceil}\}$ and for any real number z, let z^+ denote $\max\{0, z\}$. Note that m_s^* is the minimum integer m such that $2^{[m]}$ contains an antichain of size s and thus an interval [A, B] contains an antichain of size s if and only if $|B \setminus A| \ge m_s^*$. Another equivalent formulation is to say that an interval [A, B] contains a strong copy of $K_{1,s,1}$ if and only if $|B \setminus A| \ge m_s^*$. Similarly, an interval [A, B] contains a weak copy of $K_{1,s,1}$ if and only if $|B \setminus A| \ge m_s^*$. Similarly, an interval [A, B] contains to denote 0 by f(1, 1), but we will see later how the function f comes into picture.

Our next theorem deals with the weak problem for complete three-level posets $K_{r,s,t}$. The main term of all of our bounds depends on the value of r and t via the function f. For most values of s we can determine $\pi(K_{r,s,t})$, for the rest we obtain an upper bound that is bigger than our lower bound by less than one. For a real z we denote its positive part max $\{0, z\}$ by z^+ .

Theorem 8 (Patkós [99]). Let $s - f(r, t) \ge 2$.

$$\begin{array}{ll} (1) \ If \ s - f(r,t) \ \in \ \left[2^{m_s-1} - 1, 2^{m_s} - \binom{m_s}{\lceil \frac{m_s}{2} \rceil}\right) - 1\right], \ then \ \pi(K_{r,s,t}) \ = \ e(K_{r,s,t}) \ = \ e(K_{r,s,t}) \ = \ m_s + f(r,t) \ holds. \ In \ particular, \ we \ have \\ \Sigma(n,m_s+f(r,t)) + \left(\frac{(r-2)^+ + (t-2)^+}{n} - O_{r,t}(\frac{1}{n^2})\right) \binom{n}{\lceil \frac{n}{2} \rceil} \le La(n,K_{r,s,t}) \le (m_s+f(r,t) + \frac{2(r+t-2)}{n} + o(\frac{1}{n}))\binom{n}{\lceil \frac{n}{2} \rceil}. \\ (2) \ If \ s - f(r,t) \ \in \left[2^{m_s} - \binom{m_s}{\lceil \frac{m_s}{2} \rceil}\right), 2^{m_s} - 2\right], \ then \\ \Sigma(n,m_s+f(r,t)) + \left(\frac{(r-2)^+ + (t-2)^+}{n} - O_{r,t}(\frac{1}{n^2})\right)\binom{n}{\lceil \frac{n}{2} \rceil} \le La(n,K_{r,s,t}) \le (m_s+f(r,t) + 1 - \frac{2^{m_s-s+f(r,t)-1}}{\binom{m_s}{\lceil \frac{m_s}{2} \rceil}}\binom{n}{n} \ holds. \end{array}$$

Note that the special case r = t = 1 of Theorem 8 was already obtained by Griggs, Li and Lu [59]. Let us state a result that covers the case s = 2, f(r,t) > 0.

Theorem 9 (Patkós [99]). For any pair of integers r, t with f(r, t) > 0 we have $\Sigma(n, 3) + (\frac{(r-2)^+ + (t-2)^+}{n} - O_{r,t}(\frac{1}{n^2})) {n \choose \lceil \frac{n}{2} \rceil} \le La(n, K_{r,2,t}) \le (3 + \frac{2(r+t-2)}{n} + o(\frac{1}{n})) {n \choose \lceil \frac{n}{2} \rceil}.$ In particular, $\pi(K_{r,2,t}) = 3$ holds.

We turn our attention to the general case of $K_{r,s_1,s_2,...,s_j,t}$. As there are more technical details in calculating $e(K_{r,s_1,s_2,...,s_j,t})$ than in calculating $e^*(K_{r,s_1,s_2,...,s_j,t})$ we will only consider the strong problem in its full generality.

Theorem 10 (Patkós [99]). (i) For any positive integers $1 \le r, t$ we have $\Sigma(n, 4 + f(r,t)) + (\frac{r+t-2}{n} - O_{r,t}(\frac{1}{n^2})) \binom{n}{\lfloor n/2 \rfloor} \le La^*(n, K_{r,4,t}) \le (4 + f(r,t) + \frac{2(r+t-2)}{n} + o(\frac{1}{n})) \binom{n}{\lfloor n/2 \rfloor}$. In particular, $\pi^*(K_{r,4,t}) = 4 + f(r,t)$ holds.

(ii) For any constant c with 1/2 < c < 1 there exists an integer s_c such that if $s \ge s_c$ and $s \le c\binom{m_s^*}{\lceil m_s^*/2 \rceil}$, then we have $\Sigma(n, m_s^* + f(r, t)) + (\frac{r+t-2}{n} - O_{r,t}(\frac{1}{n^2}))\binom{n}{\lceil n/2 \rceil} \le La^*(n, K_{r,s,t}) \le (m_s^* + f(r, t) + \frac{2(r+t-2)}{n} + o(\frac{1}{n}))\binom{n}{\lfloor n/2 \rfloor}$. In particular, $\pi^*(K_{r,s,t}) = m_s^* + f(r, t)$ holds.

(iii) There exists an integer s_0 such that for any r, s, t with $s \ge s_0$ we have $\Sigma(n, m_s^* + f(r,t)) + (\frac{r+t-2}{n} - O_{r,t}(\frac{1}{n^2})) {n \choose \lfloor n/2 \rfloor} \le La^*(n, K_{r,s,t}) \le (m_s^* + 1 + f(r,t) + \frac{2(r+t-2)}{n} + o(\frac{1}{n})) {n \choose \lfloor n/2 \rfloor}.$

(iv) For any constant c with 1/2 < c < 1 there exists an integer s_c such that if all s_i 's satisfy that either $s_i = 4$ or $s_i \ge s_c$ and $s_i \le c \binom{m_{s_i}}{\lceil m_{s_i}^*/2 \rceil}$, then we have $La^*(n, K_{r,s_1,s_2,...,s_j,t}) = (e^*(K_{r,s_1,s_2,...,s_j,t}) + O_{r,t}(\frac{1}{n})) \binom{n}{\lfloor n/2 \rfloor}.$

Let us elaborate on the technical condition for the s_i s required by Theorem 10. For the sequence $a_j := \binom{j}{\lfloor j/2 \rfloor}$, we have $\lim_{j\to\infty} a_j/a_{j+1} = 1/2$. So the closer c is to 1, the larger fraction of the numbers in the interval $[a_j, a_{j+1}]$ satisfy the condition of Theorem 8. Part (ii) and (iv) of the theorem states that if we pick large enough numbers, then integers of an arbitrary large fraction of these intervals can be picked to play the role of s_i s.

Once an extremal result is obtained, one might get interested in describing the structures achieving the extremum. Does there exist a unique extremal object? If yes or if

only a small number of extremal objects exist, then one might ask whether almost extremal objects have very similar structure to the extremal ones. If yes, the phenomenon is called *stability*. The first stability theorem concerning set families is due to Hilton and Milner [65] even though they did not use this terminology. About the same time in extremal graph theory, Erdős [33] and Simonovits [105] were the first to consider stability problems sistematically. When the class of combinatorial objects we consider are defined via some forbidden substructures, then another line of research that arises after determining the extremal number is the so-called supersaturation problem. Let us illustrate this with the first such result in the history of extremal graph theory. The cornerstone result of Mantel [83] states that the maximum number of edges in an *n*-vertex triangle free graph is $\lfloor \frac{n^2}{4} \rfloor$. One can restate this theorem as follows: in any *n*-vertex graph with $\lfloor \frac{n^2}{4} \rfloor + E$ edges, there exist at least E triangles. Rademacher in an unpublished manuscript proved that for E = 1 something much stronger holds: in any *n*-vertex graph with $\lfloor \frac{n^2}{4} \rfloor + 1$ edges, there exist at least $\lfloor n/2 \rfloor$ triangles. Erdős [30] extended this result proving that *n*-vertex graphs with $\lfloor \frac{n^2}{4} \rfloor + E$ edges contain at least E[n/2] triangles if $E < \varepsilon n$ for some fixed positive ε . Since then, supersaturation problems are studied in all brunches of extremal combinatorics. Very often, the proof of a supersaturation result uses a stability theorem. The simple strategy is as follows: let ex be the extremal size and let \mathcal{O} be an object of size ex + E. Take a maximal subobject $\mathcal M$ of $\mathcal O$ that does not contain the forbidden structure. One has to consider two cases: using the stability result \mathcal{M} is either (i) close to the extremal objects in structure and then using this extra information one has a hope to establish the desired lower bound on the number of forbidden structures in \mathcal{O} , or (ii) not close to the extremal structre. In the latter case, the stability result implies that the size of \mathcal{M} is much less than exand the trivial bound $ex + E - |\mathcal{M}|$ on the number of forbidden substructures suffices to prove the supersaturation result. The contributions of the thesis will demonstrate how this strategy works in one particular case of the forbidden subposet problem. Before elaborating on the details, let us mention that not too many supersaturation results are known in the area. Answering a question of Erdős and Katona, Kleitman determined [75] those set families that minimize the number of pairs in containment over all families $\mathcal{F} \subset 2^{[n]}$ of given size. This is the supersaturation version of Sperner's result, Theorem 1. He immediately conjectured a solution for the supersaturation problem corresponding to Erdős's generalization, Theorem 2. The problem was unchallenged for a couple of decades, and then suddenly several researchers made progress [23, 93, 8] on the problem. Finally, Samotij proved [102] Kleitman's conjecture in full generality. A very recent joint paper of ours with Gerbner, Nagy, and Vizer [51] contains further supersaturation results in the area.

In section 2.2, we will consider stability and supersaturation of the forbidden subposet problem with the forbidden containment pattern being the butterfly poset Bnamed after the look of its Hasse-diagram \bowtie . The poset B has four elements a, b, c, dwith a, b < c, d, i.e. with the notation of complete multipartite posets $B = K_{2,2}$. The value of La(n, B) was determined by DeBonis, Katona, and Swanepoel.

Theorem 11 (DeBonis, Katona, Swanepoel [25]). For any $n \ge 3$, we have $La(n, B) = \Sigma(n, 2)$. Moreover, if $\mathcal{F} \subseteq 2^{[n]}$ is a B-free family with $|\mathcal{F}| = \Sigma(n, 2)$, then $\mathcal{F} \in \Sigma^*(n, 2)$.

We will prove a stability version of Theorem 11. Note that a 4-chain is a special weak copy of B, therefore every B-free family is 3-Sperner. Theorem 11 states that any maximum size B-free family is actually 2-Sperner. First we will prove a stability theorem that states that the more "middle sets" (sets that are middle elements in a 3-chain in \mathcal{F}) the B-free \mathcal{F} contains, the smaller size it must have. This will follow from a rather standard argument. A more involved proof will yield the following stronger result that sates that if a B-free family \mathcal{F} differs "much" from the extremal family of Theorem 11, then its size is "much" smaller than $\Sigma(n, 2)$.

Theorem 12 (Patkós [100]). Let *m* be a non-negative integer with $m \leq {\binom{2n}{3}-1}{\lceil n/2 \rceil}$ and let $\mathcal{F} \subseteq 2^{[n]}$ be a butterfly-free family such that $|\mathcal{F} \setminus \mathcal{F}^*| \geq m$ for every $\mathcal{F}^* \in \Sigma^*(n, 2)$. Then the inequality $|\mathcal{F}| \leq \Sigma(n, 2) - \frac{m}{4}$ holds if *n* is large enough.

Theorem 12 will be one of our main tools to deduce the following supersaturation result. We will also show that the bounds of the next result are sharp.

Theorem 13 (Patkós [100]). Set $f(n) = (\lceil n/2 \rceil + 1) \binom{\lceil n/2 \rceil}{2}$. Let $\mathcal{F} \subseteq 2^{[n]}$ be a family of sets with $|\mathcal{F}| = \Sigma(n, 2) + E$.

(a) If E = E(n) satisfies $\log E = o(n)$, then the number of weak copies of B contained by \mathcal{F} is at least $(1 - o(1))E \cdot f(n)$.

(b) Furthermore, if $E \leq \frac{n}{100}$, then the number of weak copies of B contained by \mathcal{F} is at least $E \cdot f(n)$.

Although, in this thesis, we do not present results in this topic, let us mention that beyond stability and supersaturation it is common to address counting problems. Clearly, if a family \mathcal{F} of sets does not contain any copies of some forbidden poset P, then neither does any of its subfamilies. Therefore for any n and P there exist at least $2^{La(n,P)}$ P-free families in $2^{[n]}$. One aims to prove an upper bound $2^{(1+o(1))La(n,P)}$ on this number. Note that if $P = P_2$, then this is Dedekind's problem [26] of determining the number of antichains in $2^{[n]}$, which was settled by Korshunov [76].

There existed some theorems in extremal graph theory about the maximum number ex(n, H, F) of *H*-subgraphs in *n*-vertex *F*-free graphs. Notable results were the cases $H = C_3$, $F = C_5$ [13] and vice versa $H = C_5$ $F = C_3$ [13, 62, 64]. Recently, Alon and Shikelmann [4] started addressing the general parameter ex(n, H, F) and many researchers have found interest in proving bounds on this number.

With Dániel Gerbner and Balázs Keszegh [49], we initiated the investigation of the forbidden subposet analog of the topic: counting the maximum number of copies of a poset in a family $\mathcal{F} \subseteq 2^{[n]}$ that is \mathcal{P} -free. More formally, we introduced the following quantity: Let $\mathcal{F} \subseteq 2^{[n]}$ and P be a poset, then let $c(P, \mathcal{F})$ denote the number of weak copies of P in \mathcal{F} .

Definition 14. For families of posets \mathcal{P} and \mathcal{Q} let

$$La(n, \mathcal{P}, \mathcal{Q}) := \max \left\{ \sum_{Q \in \mathcal{Q}} c(Q, \mathcal{F}) : \mathcal{F} \subseteq 2^{[n]}, \ \mathcal{F} \ is \ \mathcal{P}\text{-}free \right\}.$$

If either $\mathcal{Q} = \{Q\}$ or $\mathcal{P} = \{P\}$, then we simply write $La(n, P, \mathcal{Q})$, $La(n, \mathcal{P}, Q)$, $La(n, \mathcal{P}, Q)$. Note that $La(n, \mathcal{P}) = La(n, \mathcal{P}, P_1)$.

There are not many results in the literature where other posets are counted. Katona [70] determined the maximum number of 2-chains (copies of P_2) in a 2-Sperner (P_3 -free) family $\mathcal{F} \subseteq 2^{[n]}$ by showing $La(n, P_3, P_2) = \binom{n}{i_1}\binom{i_1}{i_2}$ where i_1, i_2 are chosen such that $n - i_1, i_1 - i_2$ and i_2 differ by at most one, so i_1 is roughly 2n/3 and i_2 is roughly n/3. This was reproved in [95] and generalized by Gerbner and Patkós in [52], where we proved the following result. To state the theorem and for later purposes we will use the multinomial coefficient that counts the number of k-chains $F_1 \subsetneq F_2 \subsetneq \cdots \subsetneq F_k$ in $2^{[n]}$ with $|F_i| = l_i: \binom{n}{l_1, l_2, \dots, l_k} = \prod_{i=1}^k \binom{l_{k-i+2}}{l_{k-i+1}}$, where $i_{k+1} = n$.

Theorem 15 (Gerbner, Patkós [52]). For any pair $k > l \ge 1$ of integers we have

$$La(n, P_k, P_l) = \max_{0 \le i_1 < i_2 < \dots < i_{k-1} \le n} c\left(P_l, \bigcup_{j=1}^{k-1} {[n] \choose i_j}\right) = \max_{0 \le i_1 < i_2 < \dots < i_{k-1} \le n} {n \choose l_{k-1}, \dots, l_1}$$

Moreover, if k = l + 1, then the above maximum is attained when the integers $i_1, i_2 - i_1, \ldots, i_{k-1} - i_{k-2}, n - i_{k-1}$ differ by at most one.

Already these results show that as opposed to Conjecture 3, in this generalized setting the extremal structures not always consist of consecutive levels. In our paper [49], among other specific results, we showed examples of P and Q for which the asymptotics of La(n, P, Q) is not attained by any family that is the union of some levels.

With Gerbner, Methuku, Nagy, and Vizer we studied the maximum number of k-chains in P-free families and observed a phenomenon that somewhat resembles to the statement of the Erdős-Stone-Simonovits theorem in extremal graph theory: if the chromatic number of the forbidden graph H is at least 3, then there exists dense H-free graphs, while if H is bipartite, then all H-free graphs have subquadratic number of edges. Thus the distinguishing parameter for graphs is the chromatic number of the forbidden subgraph. Less surprisingly, in our case the important parameter is the height h(P) of the forbidden subposet. Note that by Theorem 15, the quantity $La(n, P_{k+1}, P_k)$ grows exponentially in n and the base is increasing in k, so despite the polynomial factor in the upper bound of (ii) of the next Theorem, the parts (i) and (ii) cover essentially different behaviors of $La(n, P, P_k)$.

Theorem 16 (Gerbner, Methuku, Nagy, Patkós, Vizer [50]). Let l be the height of P. (i) If l > k, then

$$La(n, P, P_k) = \Theta(La(n, P_{k+1}, P_k)).$$

Moreover,

 $La(n, P_{k+1}, P_k) \le La(n, P, P_k) \le La(n, P_{|P|}, P_k) \le \binom{|P| - 1}{k} La(n, P_{k+1}, P_k).$ (ii) If $l \le k$, then

$$La(n, P, P_k) = O(n^{2k-1/2}La(n, P_l, P_{l-1})),$$

and there exists a poset P of height l such that $La(n, P, P_k) = \Theta(La(n, P_l, P_{l-1}))$ holds.

We remark that in a forthcoming manuscript with József Balogh, Ryan R. Martin, and Dániel Nagy, we obtained sharp results for the case k = 2 and showed that for any poset P of height 2, one has $La(n, P, P_2) = O(n\binom{n}{\lfloor n/2 \rfloor})$ and this order of magnitude cannot be improved as $La(n, B, P_2) = \lceil n/2 \rceil \binom{n}{\lfloor n/2 \rfloor}$.

Let us briefly discuss the impact of the theorems of this section. The half sentence "the conjecture was proved for some infinite classes of complete multipartite posets by Patkós [99]" does not hurt, so Theorems 7, 8, 9, 10 are often cited. On the other hand, different subtopics in the area of forbidden subposet problems attract very different amount of interest. Saturation problems, an area not included in this dissertation, seem to be very popular recently, while supersaturation phenomena were addressed only in the case of chains - until this problem was completely resolved by Samotij [102]. Therefore, most citations of [100] come from PhD theses of graduate students who try to cover the whole area. Theorem 16 and the papers [49, 50] on the generalized forbidden subposet problem are quite recent. We included this result to the dissertation as it points to some interesting open problems and to some analogy with degenerate Turán problems for graphs.

Complete multipartite posets 2.1

2.1.1**Preliminaries**

In this subsection, we gather all lemmas needed for the proofs of Theorems 7, 8, 9, 10. We start with a technical proposition

Proposition 17. (i) If $s_i \ge 2$ holds for all $1 \le i \le j$, then we have $e^*(K_{r,s_1,s_2,\ldots,s_j,t}) =$

 $\begin{array}{l} f(r,t) + \sum_{i=1}^{j} m_{s_{i}}^{*}.\\ (ii) \ Let \ us \ write \ w = |\{i : s_{i-1} = s_{i} = 1\}|, \ where \ r = s_{0} \ and \ t = s_{j+1}. \ Then \\ (ii) \ Let \ us \ write \ w = |\{i : s_{i-1} = s_{i} = 1\}|, \ where \ r = s_{0} \ and \ t = s_{j+1}. \end{array}$ $e^*(K_{r,s_1,s_2,\ldots,s_j,t}) = w + e^*(K_{r,\sigma_1,\sigma_2,\ldots,\sigma_{j'},t})$, where $\sigma_1, \sigma_2, \ldots, \sigma_{j'}$ is the sequence obtained from s_1, s_2, \ldots, s_j by removing all its ones.

Proof. To see (i), let \mathcal{F} consist of $f(r,t) + \sum_{i=1}^{j} m_{s_i}^*$ consecutive levels of $2^{[n]}$ and suppose we find a strong copy of $K_{r,s_1,s_2,\ldots,s_j,t}$. If F_1,\ldots,F_r and F'_1,\ldots,F'_t play the role of the bottom r and the top t sets, then $|\bigcap_{i=1}^{t} F'_i| - |\bigcup_{k=1}^{r} F_j| < \sum_{l=1}^{j} m_{s_i}^*$ holds. If $F_1^i, \ldots, F_{s_i}^i$ play the role of the sets of the *i*th middle level of $K_{r,s_1,s_2,\ldots,s_j,t}$, then $|\cup_{j=1}^{s_i} F_j^i| \ge |\cup_{j=1}^{s_{i-1}} F_j^{i-1}| + s_j$ must hold. Thus one would need $\sum_{i=1}^j m_{s_i}^*$ more levels for the j middle levels of $K_{r,s_1,s_2,\ldots,s_j,t}$. It is easy to see that $f(r,t) + \sum_{i=1}^j m_{s_i}^* + 1$ consecutive levels do contain a strong copy of $K_{r,s_1,s_2,\ldots,s_i,t}$.

To see (ii), assume \mathcal{G} is a copy of a strong $K_{r,s_1,s_2,\ldots,s_j,t}$ in $2^{[n]}$. Let i, i + p be two indices such that $s_i, s_{i+p+1} \geq 2$ and $s_{i+h} = 1$ for all $1 \leq h \leq p$. Let $G_1^i, \ldots, G_{s_i}^i$ and $G_1^{i+p+1}, \ldots, G_{s_{i+p+1}}^{i+p+1}$ denote the sets in \mathcal{G} corresponding to the *i*th and (i+p+1)st level of $K_{r,s_1,s_2,\ldots,s_j,t}$. Then for $I = \bigcup_{l=1}^{s_i} G_l^i$ and $J = \bigcap_{l=1}^{s_{i+p+1}} G_l^{i+p+1}$ we must have $I \subseteq J$ and $|J| - |I| \ge p - 1$ as \mathcal{G} contains a chain of length p in [I, J]. For $G \in \mathcal{G}$ let us write r(G) for the rank of the element corresponding to G. Then $\mathcal{G}' = \{G \in \mathcal{G} : r(G) \leq G\}$ $i \} \cup \{G \setminus (J \setminus I) : G \in \mathcal{G}, r(G) \ge i + p + 1\}$ is a strong copy of $K_{r,s_1,s_2,\ldots,s_i,s_{i+p+1},\ldots,s_j,t}$ such that the size of the largest set in \mathcal{G}' is (p-1) less than that the size of the largest set in \mathcal{G} . Continuing this process we obtain a copy of $K_{r,\sigma_1,\sigma_2,\ldots,\sigma_{s'},t}$ where the size of the largest set is w less than the size of the largest set in \mathcal{G} . This shows $e^*(K_{r,\sigma_1,\sigma_2,\ldots,\sigma_{j'},t}) \leq e^*(K_{r,s_1,s_2,\ldots,s_j,t}) - w$. To see the other inequality, one has to reverse the above procedure. We leave the details to the reader.

The proofs of Theorems 7, 8, 9, 10 will use the chain-partition method of Griggs, Li, and Lu [59]. The often used LYM-inequality [82, 85, 111] is derived by counting pairs all (F, C) with $F \in \mathcal{F}$ and \mathcal{C} being a maximal chain. When proving the LYMinequality, one gets an upper bound on the number of such pairs by observing that every maximal chain can contain at most a bounded number of sets in \mathcal{F} . The idea of the chain-partition method is that even if a certain bound does not hold for some particular chains, then the same bound might be true on average in every part of some partition of \mathbf{C}_n , the set of maximal chains in [n]. Griggs, Li, and Lu introduced min-, max-, min-max-partitions. We will (define and) use these and some alterations as well. We start with some notation. For a family $\mathcal{F} \subseteq 2^{[n]}$ of sets and $A \subseteq [n]$ we define $s_{\mathcal{F}}^{-}(A)$ to be the maximum size of an antichain in $\mathcal{F} \cap 2^A$ and $s^+_{\mathcal{F}}(A)$ to be the maximum size of an antichain in $\{F \in \mathcal{F} : A \subseteq F\}$. For a set $A \subseteq [n]$ and a family \mathcal{F} of sets let $\mathbf{C}_{A,k,-}$ denote the set of those maximal chains \mathcal{C} from \emptyset to A for which for every $C \in \mathcal{C} \setminus \{A\}$

we have $s_{\mathcal{F}}^-(C) < k$ and let $\mathbf{C}_{A,k,+}$ denote the set of those maximal chains \mathcal{C} from A to [n] for which for every $C \in \mathcal{C} \setminus \{A\}$ we have $s_{\mathcal{F}}^+(C) < k$.

The min-max-*partition* of \mathbf{C}_n with respect to a family $\mathcal{F} \subseteq 2^{[n]}$ is $\{\mathbf{C}_{A,B} : A \subseteq B \subseteq [n]\}$ where $\mathbf{C}_{A,B}$ consists of those maximal chains \mathcal{C} in \mathbf{C}_n for which the smallest set of $\mathcal{C} \cap \mathcal{F}$ is A and the largest set of $\mathcal{C} \cap \mathcal{F}$ is B. To obtain a real partition of \mathbf{C}_n one has to add $\mathbf{C}_{\emptyset} = \{\mathcal{C} \in \mathbf{C}_n : \mathcal{C} \cap \mathcal{F} = \emptyset\}.$

For $r \geq 2$ let us now define the min_r-partition of \mathbf{C}_n with respect to \mathcal{F} . For a set A with $s_{\mathcal{F}}^-(A) \geq r$ we set $\mathbf{C}_{A,r} = \{\mathcal{C} \in \mathbf{C}_n : A \in \mathcal{C}, \forall C \subset A, C \in \mathcal{C} : s_{\mathcal{F}}^-(C) < r\}$. Note that every $\mathcal{C} \in \mathbf{C}_n$ belongs to exactly one set $\mathbf{C}_{A,r}$ provided \mathcal{F} contains an antichain of size r as then $s_{\mathcal{F}}^-([n]) \geq r$ and [n] is contained in all maximal chains $\mathcal{C} \in \mathbf{C}_n$.

Now we define the $\min_r - \max_t$ partition of \mathbf{C}_n . Before introducing the formal definition, we describe the idea of the partition. For the sake of simplicity assume that both r and t are at least 2. For every chain $\mathcal{C} \in \mathbf{C}_n$ we want to introduce two markers $A, B \in \mathcal{C}$ with the property that A is the smallest set in \mathcal{C} below which there exists an antichain of size r in \mathcal{F} (i.e., $s_{\mathcal{F}}^-(A) \geq r$) and B is the largest set in \mathcal{C} above which there exists an antichain of size t in \mathcal{F} (i.e., $s_{\mathcal{F}}^+(B) \geq t$). If \mathcal{F} is $K_{r,s,t}$ -free, we know that $[A, B] \cap \mathcal{F}$ contains less than s sets, while if \mathcal{F} is strong $K_{r,s,t}$ -free, then [A, B] does not contain an antichain of size s. The problem with the above reasoning is that $B \subsetneq A$ might hold, thus we will have to distinguish two cases.

Let us start with introducing $S = \{S \in 2^{[n]} : s_{\mathcal{F}}(S) \geq r\}$, the family of those sets that can play the role of A in the above argument. We partition S into two subfamilies: $S^- = \{S \in S : s_{\mathcal{F}}^+(S) < t\}$ and $S^+ = S \setminus S^-$. Clearly, if $A \in S^-$ is the smallest set in the chain $\mathcal{C} \in \mathbf{C}_n$ with $s_{\mathcal{F}}^-(A) \geq r$, then for the largest set B in \mathcal{C} with $s_{\mathcal{F}}^+(B) \geq t$ we will have $B \subsetneq A$.

For any set $S \in S^-$ let \mathbf{C}_S denote the set of those maximal chains \mathcal{C} in \mathbf{C}_n in which

- if r = 1, then S is the smallest set in $\mathcal{F} \cap \mathcal{C}$,
- if $r \ge 2$, then S is the smallest set in C with $s_{\mathcal{F}}(S) \ge r$.

For any set $A \in S^+$ and B with $A \subseteq B$ let $\mathbf{C}_{A,B} = \mathbf{C}_{A,r,B,t}$ denote the set of those maximal chains \mathcal{C} in \mathbf{C}_n in which

- if r = 1, then A is the smallest set in $\mathcal{F} \cap \mathcal{C}$,
- if $r \ge 2$, then A is the smallest set in \mathcal{C} with $s_{\mathcal{F}}(A) \ge r$,
- if t = 1, then B is the largest set in $\mathcal{F} \cap \mathcal{C}$,
- if $t \ge 2$, then B is the largest set in C with $s_{\mathcal{F}}^+(B) \ge t$.

The min_r – max_t partition of \mathbf{C}_n is $\{\mathbf{C}_S : S \in \mathcal{S}^-\} \cup \{\mathbf{C}_{A,B} : A \in \mathcal{S}^+, A \subseteq B\}$. Consider a maximal chain $\mathcal{C} \in \mathbf{C}_n$. If $r \geq 2$ and the size z of the largest antichain in \mathcal{F} satisfies $z = s_{\mathcal{F}}^-([n]) \geq \max\{r, t\}$, then there is a smallest set H of \mathcal{C} with $s_{\mathcal{F}}^-(H) \geq r$. If $H \in \mathcal{S}^-$, then \mathcal{C} belongs to \mathbf{C}_H . If not, then $H \in \mathcal{S}^+$ and thus for the largest set H' of \mathcal{C} with $s_{\mathcal{F}}^+ \geq t$ we have $H \subseteq H'$ and therefore $\mathcal{C} \in \mathbf{C}_{H,H'}$ holds. We obtained that the min_r – max_t partition of \mathbf{C}_n is indeed a partition if $r \geq 2$. If r = 1, then we need to add the set $\mathbf{C}_{\emptyset} = \{\mathcal{C} \in \mathbf{C}_n : \mathcal{C} \cap \mathcal{F} = \emptyset\}$.

After introducing the necessary definitions, we start to prove our preliminary lemmas that will serve as building blocks of our proofs in the next subsection.

Lemma 18. Let $\mathcal{F} \subseteq 2^{[n]}$ be a family such that all $F \in \mathcal{F}$ have size in $[n/2 - n^{2/3}, n/2 + n^{2/3}]$.

(i) Let $A \subset [n]$ with $s_{\mathcal{F}}(A) < k$. Then the number of pairs (F, \mathcal{C}) where \mathcal{C} is a maximal chain from \emptyset to A and $F \in \mathcal{F} \cap (\mathcal{C} \setminus \{A\})$ is $\frac{2(k-1)}{n}|A|! + o(\frac{1}{n}|A|!)$. (ii) Let $A \subset [n]$ with $s_{\mathcal{F}}^+(A) < k$. Then the number of pairs (F, \mathcal{C}) where \mathcal{C} is a

(ii) Let $A \subset [n]$ with $s_{\mathcal{F}}^+(A) < k$. Then the number of pairs (F, \mathcal{C}) where \mathcal{C} is a maximal chain from A to [n] and $F \in \mathcal{F} \cap (\mathcal{C} \setminus \{A\})$ is $\frac{2(k-1)}{n}|A|! + o(\frac{1}{n}|A|!)$.

Proof. The property possessed by A and \mathcal{F} ensures that $\mathcal{F}_A := \{F \in \mathcal{F} : F \subset A\}$ contains at most k - 1 sets of each possible size. Thus the number of pairs (F, \mathcal{C}) in question is at most

$$\sum_{i=n/2-n^{2/3}}^{\min\{n/2+n^{2/3},|A|-1\}} (k-1)i!(|A|-i)! \le \frac{k-1}{|A|}|A|! + \frac{2(k-1)}{|A|(|A|-1)}|A|! + \frac{12(k-1)n^{2/3}}{|A|(|A|-1)(|A|-2)}|A|! \le \frac{2(k-1)}{n}|A|! + O_k(\frac{1}{n^{4/3}}|A|!)$$

if $|A| \ge (1/2 - n^{1/3})n$. If $|A| \le (1/2 - n^{1/3})n$, then \mathcal{F} does not contain any subset F of A. This completes the proof of (i) and (ii) follows by applying (i) to the set \overline{A} and the family $\overline{\mathcal{F}}$.

Remark. Note that $n^{2/3}$ could be replaced by any function f(n) satisfying $4 \log n \sqrt{n} \le f(n) = o(n)$. In the proof of Lemma 18 we used f(n) = o(n) and at the beginning of the proofs of upper bounds in the next subsection, we will need a calculation involving Chernoff's inequality where the assumption $4 \log n \sqrt{n} \le f(n)$ will be used.

Corollary 19. Let $\mathcal{F} \subseteq 2^{[n]}$ be a family such that all $F \in \mathcal{F}$ have size in $[n/2 - n^{2/3}, n/2 + n^{2/3}]$.

(i) Let $A \subset [n]$ with $s_{\mathcal{F}}^-(A) \geq k$. Then the number of pairs (F, \mathcal{C}) where $\mathcal{C} \in \mathbf{C}_{A,k,-}$ and $F \in \mathcal{F} \cap (\mathcal{C} \setminus \{A\})$ is $(1 + \frac{2(k-1)}{n})|\mathbf{C}_{A,k,-}| + o(\frac{1}{n}|\mathbf{C}_{A,k,-}|)$. (ii) Let $A \subset [n]$ with $s_{\mathcal{F}}^+(A) \geq k$. Then the number of pairs (F, \mathcal{C}) where $\mathcal{C} \in \mathbf{C}_{A,k,+}$

(ii) Let $A \subset [n]$ with $s_{\mathcal{F}}^+(A) \geq k$. Then the number of pairs (F, \mathcal{C}) where $\mathcal{C} \in \mathbf{C}_{A,k,+}$ and $F \in \mathcal{F} \cap (\mathcal{C} \setminus \{A\})$ is $(1 + \frac{2(k-1)}{n})|\mathbf{C}_{A,k,+}| + o(\frac{1}{n}|\mathbf{C}_{A,k,+}|).$

Proof. First we prove (i). Let $A_1, \ldots, A_j, A_{j+1}, \ldots, A_{|A|}$ denote the subsets of A of size |A|-1 such that $s_{\mathcal{F}}(A_i) < k$ if and only if $1 \leq i \leq j$. (If $s_{\mathcal{F}}(A) \geq k$ for all i, then $\mathbf{C}_{A,k,-}$ is empty and there is nothing to prove.) Note that if $S_1 \subset S_2$, then $s_{\mathcal{F}}(S_2) < k$ implies $s_{\mathcal{F}}(S_1) < k$. Therefore $\mathbf{C}_{A,k,-} = \bigcup_{i=1}^{j} \mathbf{C}_{A_i,A}$, where $\mathbf{C}_{A_i,A}$ denotes the set of those maximal chains from \emptyset to A that contain A_i . Indeed, $\mathbf{C}_{A_i,A} \subset \mathbf{C}_{A,k,-}$ for $1 \leq i \leq j$ as by the above A is the first set in a chain $\mathcal{C} \in \mathbf{C}_{A_i,A}$ with $s_{\mathcal{F}}(A)$ at least k, while for all $i \geq j+1$ we have $s_{\mathcal{F}}(A_j) \geq k$ and thus $\mathbf{C}_{A_i,A} \cap \mathbf{C}_{A,k,-} = \emptyset$.

Let us fix *i* with $1 \leq i \leq j$ and consider pairs (F, \mathcal{C}) with $F \in \mathcal{F} \cap \mathcal{C}$ and $\mathcal{C} \in \mathbb{C}_{A_i, A}$. As $s_{\mathcal{F}}^-(A_i) < k$, we can apply Lemma 18 (i) to \mathcal{F} and A_i , and obtain that the number of such pairs with $F \subsetneq A_i$ is at most $\frac{2(k-1)}{n}|A_i|! + o(\frac{1}{n}|A_i|!)$. Even if all A_i 's belong to

 \mathcal{F} , then every chain $\mathcal{C} \in \mathbf{C}_{A,k,-}$ can contain one more set from \mathcal{F} , namely one of the A_i 's. This completes the proof of (i) and (ii) follows by applying (i) to the set \overline{A} and the family $\overline{\mathcal{F}}$.

Lemma 20. (i) Let $\mathcal{G} \subseteq 2^{[k]}$ be a family of sets such that any antichain $\mathcal{A} \subset \mathcal{G}$ has size at most 3. Then the number of pairs (G, \mathcal{C}) with $G \in \mathcal{G} \cap \mathcal{C}$ and $\mathcal{C} \in \mathbf{C}_k$ is at most 4k!.

(ii) For any constant c with 1/2 < c < 1 there exists an integer s_c such that if $s \ge s_c$ and $s \le c \binom{m_s^*}{\lfloor m_s^*/2 \rfloor}$, then the following holds: if $\mathcal{G} \subseteq 2^{[k]}$ is a family of sets such that any antichain $\mathcal{A} \subset \mathcal{G}$ has size less than s, then the number of pairs (G, \mathcal{C}) with $G \in \mathcal{G} \cap \mathcal{C}$ and $\mathcal{C} \in \mathbf{C}_k$ is at most $m_s^* k!$.

(iii) There exists an integer s_0 such that if $s \ge s_0$ and $\mathcal{G} \subseteq 2^{[k]}$ is a family of sets such that any antichain $\mathcal{A} \subset \mathcal{G}$ has size at most s, then the number of pairs (G, \mathcal{C}) with $G \in \mathcal{G} \cap \mathcal{C}$ and $\mathcal{C} \in \mathbf{C}_k$ is at most $(m_s^* + 1)k!$.

Proof. First we prove (i). We may assume that $\emptyset, [k] \in \mathcal{G}$ holds as adding them will not result in violating the condition of the lemma and the number of pairs to be counted can only increase. These two sets are in k! maximal chains each, thus giving 2k! pairs. All other sets belong to $|G|!(k - |G|)! = \frac{k!}{\binom{k}{|G|}}$ chains in \mathbf{C}_k . Sets of same size form an antichain, therefore for every $1 \leq i \leq k - 1$ there exist at most 3 sets of size i in \mathcal{G} and thus the total number of pairs (G, \mathcal{C}) is at most

$$S(k) = 2k! + 3k! \sum_{i=1}^{k-1} \frac{1}{\binom{k}{i}}.$$

For k = 2, 3, 4, 5 the sum S(k) equals 3.5k!, 4k!, 4k!, 3.8k!, respectively. Furthermore, it is an easy exercise to show that $\frac{S(k)}{k!}$ is monotone decreasing for $k \ge 5$ and therefore $\frac{S(k)}{k!} \le 4$ holds for all positive integers k. This completes the proof of (i).

Now we prove (ii). Clearly, as long as $k < m_s^*$ we can have $\mathcal{G} = 2^{[k]}$ and then the number of pairs is $(k+1)k! \leq m_s^*k!$. When $k \geq m_s^*$ we again use the observation that for any $0 \leq j \leq k$ we have $|\{G \in \mathcal{G} \cap {[k] \choose j}| < s$ and thus the number of pairs (G, \mathcal{C}) is at most $S(k) = \sum_{j=0}^{k} \min\{s-1, {k \choose j}\}j!(n-j)!$. We need to show that $R(k) := \frac{S(k)}{k!} = \sum_{j=0}^{k} \min\{\frac{s-1}{{k \choose j}}, 1\} \leq m_s^*$ holds for all $k \geq m_s^*$. Consider the case $k = m_s^*$. If s is large enough (and thus m_s^* and k), then ${m_s^*/2 \choose m_s^*/2} = (1+o(1)){m_s^*/2 \choose m_s^*/2}$ holds provided $|j| \leq \sqrt{m_s^*}/\log m_s^*$. Therefore, by the assumption $s \leq c {m_s^* \choose m_s^*/2}$ we have at least $2\sqrt{m_s^*}/\log m_s^*$ summands in $R(m_s^*)$ that are not more than $\frac{1+c}{2}$, a constant smaller than 1. Thus, if m_s^* is large enough, their subsum

$$\sum_{i=\lceil m_{s}^{*}/2\rceil-\sqrt{m_{s}^{*}}/\log m_{s}^{*}}^{\lceil m_{s}^{*}/2\rceil+\sqrt{m_{s}^{*}}/\log m_{s}^{*}}\frac{s-1}{\binom{m_{s}^{*}}{j}}$$

is less than $2\sqrt{m_s^*}/\log m_s^* - 1$ and since all other summands are not more than 1, we obtain $R(m_s^*) < m_s^*$.

To finish the proof of (ii), we prove that if $k \ge m_s^*$ holds, then we have $R(k+1) \le R(k)$. First note that if $r_{k,j}$ denotes the *j*th summand in R(k), then we have $r_{k,j} \ge r_{k+1,j}$ and $r_{k,k-j} \ge r_{k+1,k+1-j}$. Thus it is enough to show

$$\sum_{i=-1}^{1} r_{k,\lceil k/2\rceil+i} \ge \sum_{i=-1}^{2} r_{k+1,\lceil k/2\rceil+i}.$$

By the definition of m_s^* , we know that $r_{k,\lceil k/2\rceil} < 1$. Since $\binom{k}{\lceil k/2\rceil} = (1/2 + o(1))\binom{k+1}{\lceil k/2\rceil}$ we have that the left hand side is $(3 + o(1))r_{k,\lceil k/2\rceil}$ while the right hand side is $(4 + o(1))r_{k,\lceil k/2\rceil}/2 = (2 + o(1))r_{k,\lceil k/2\rceil}$. This finishes the proof of (ii).

Finally, we prove (iii). Clearly, as long as $k \leq m_s^*$ for any family $\mathcal{G} \subseteq 2^{[k]}$ the number of pairs is $(k+1)k! \leq (m_s^*+1)k!$. We need to show that $R(k) \leq m_s^*+1$ holds for all $k > m_s^*$. As in (ii) the proof of $R(k+1) \leq R(k)$ for $k \geq m_s^*$ did not require the assumption on s and c, we obtain that $R(k) \leq m_s^*+1$ holds for all k. \Box

Our last auxiliary lemma was proved by Griggs, Li and Lu [59].

Lemma 21 (Griggs, Li, Lu, during the proof of Theorem 2.5 in [59]). Let $s \ge 2$, and define $m'_s := \lceil \log_2(s+2) \rceil$.

(1) If $s \in [2^{m'_s-1}-1, 2^{m'_s}-\binom{m'_s}{\lceil \frac{m'_s}{2} \rceil})-1]$, then if $\mathcal{G} \subseteq 2^{[k]}$ is a $K_{1,s,1}$ -free family of sets, then the number of pairs (G, \mathcal{C}) with $G \in \mathcal{G} \cap \mathcal{C}$ and $\mathcal{C} \in \mathbf{C}_k$ is at most $m'_s k!$.

sets, then the number of pairs $(G, \mathcal{C})^{*}$ with $G \in \mathcal{G} \cap \mathcal{C}$ and $\mathcal{C} \in \mathbf{C}_k$ is at most $m'_s k!$. (2) If $s \in [2^{m'_s} - \binom{m'_s}{\lceil \frac{m'_s}{2} \rceil}, 2^{m'_s} - 2]$, then if $\mathcal{G} \subseteq 2^{[k]}$ is a $K_{1,s,1}$ -free family of sets, then the number of pairs (G, \mathcal{C}) with $G \in \mathcal{G} \cap \mathcal{C}$ and $\mathcal{C} \in \mathbf{C}_k$ is at most $(m'_s + 1 - \frac{2^{m'_s - s - 1}}{\binom{m'_s}{\lceil \frac{m'_s}{2} \rceil}})k!$.

2.1.2 Proofs of Theorems 7, 8, 9, 10

Let us start with constructions to see the lower bounds. We partition $\binom{[n]}{k}$ into n classes: $\mathcal{F}_{n,k,i} = \{F \in \binom{[n]}{k} : \sum_{j \in F} j \equiv i \pmod{n}\}$ and denote the union of the r largest classes by $\binom{[n]}{k}_{r,mod}$. Clearly, $|\binom{[n]}{k}_{r,mod}| \geq \frac{r}{n} \binom{n}{k}$. Furthermore, it has the property that for any distinct r + 1 sets $F_1, F_2, \ldots, F_{r+1} \in \binom{[n]}{k}_{r,mod}$ we have $|\bigcap_{i=1}^{r+1} F_i| \leq k-2$ and $|\bigcup_{i=1}^{r+1} F_i| \geq k+2$.

- For Theorem 7 consider the family $\mathcal{F} := {\binom{[n]}{\lceil n/2 \rceil 2}}_{r-1,mod} \cup {\binom{[n]}{\lceil n/2 \rceil 1}} \cup {\binom{[n]}{\lceil n/2 \rceil}} \cup {\binom{[n]}{\lceil n/2 \rceil}} \cup {\binom{[n]}{\lceil n/2 \rceil 1}} \cup {\binom{[n]}{\lceil n/2 \rceil}} \cup {\binom{[n]}{\rceil}} \bigcirc {\binom{[n]}{\rceil}} \cup {\binom{[n]}{\rceil}} \bigcirc {\binom{[n]}{\rceil}} \square {\binom{[n]}{\rceil}} \square {\binom{[n]}{\rceil}} \bigcirc {\binom{[n]}{\rceil}} \square {\binom{[n]}{\rceil}} \square {\binom{[n]}{\rceil}} \bigcirc {\binom{[n]}{\rceil}} \bigcirc {\binom{[n]}{\rceil}} (\binom{[n]}{\rceil} \bigcirc {\binom{[n]}{\rceil}} (\binom{[n]}{\rceil} \bigcirc {\binom{[n]}{\rceil}} (\binom{[n]}{\rceil}) (\binom{[n]}{\rceil} (\binom{[n]}{\rceil}) (\binom{[n]}{\rceil}) (\binom{[n]}{\rceil}) (\binom{[n]}{\rceil} (\binom{[n]}{\rceil}) (\binom{[n]}{\rceil}$
- For Theorem 8 let k be the index of the level below the $m_s + f(r, t)$ middle levels, i.e., $k = \lceil \frac{n-m_s - f(r,t)}{2} \rceil - 1$. Write $l = k + m_s + f(r,t) + 1$ and let us consider the family

$$\mathcal{F} := \binom{[n]}{k}_{(r-2)^+,mod} \cup \bigcup_{i=1}^{m_s+f(r,t)} \binom{[n]}{k+i} \cup \binom{[n]}{l}_{(t-2)^+,mod}.$$

We claim that \mathcal{F} is $K_{r,s,t}$ -free. Assume not and let $A_1, A_2, \ldots, A_r, B_1, B_2, \ldots, B_s$, $C_1, C_2, \ldots, C_t \in \mathcal{F}$ form a copy of $K_{r,s,t}$. If $r \geq 2$, then $|\cup_{i=1}^r A_i| \geq k+2$ and if r = 1, then $|A_1| \geq k+1$ (note that if r = 1, 2, then $(r-2)^+ = 0$ and thus the smallest set size in \mathcal{F} is k+1). Similarly, if $t \geq 2$, then $|\cap_{j=1}^t C_j| \leq l-2$ and if t = 1, then $|C_1| \leq l-1$. In any case, $|\cap_{t=1}^t C_j| - |\cup_{i=1}^r A_i| \leq m_s - 1$ and thus there is no place for B_1, B_2, \ldots, B_s - a contradiction.

- The construction showing the lower bound of Theorem 9 is a special case of the one for Theorem 8.
- For Theorem 10 (i), (ii) and (iii), let k be the index of the level below the $m_s^* + f(r,t)$ middle levels, i.e., $k = \lceil \frac{n-m_s^*-f(r,t)}{2} \rceil 1$. Write $l = k + m_s^* + f(r,t) + 1$ and let us consider the family

$$\mathcal{F} := \binom{[n]}{k}_{r-1,mod} \cup \bigcup_{i=1}^{m_s^* + f(r,t)} \binom{[n]}{k+i} \cup \binom{[n]}{l}_{t-1,mod}$$

One can see that for any antichains $A_1, A_2, \ldots, A_r \in \mathcal{F}$ and $C_1, C_2, \ldots, C_t \in \mathcal{F}$ we have $|\bigcap_{i=1}^t C_i| - |\bigcup_{j=1}^r A_j| \le m_s^* - 1$ and thus there is no room for an antichain of size s in between. Note that when s = 4, then $m_s^* = 4$ as $\binom{4}{2} = 6 \ge 4$, but $\binom{3}{2} = 3 < 4$.

Let us now start proving the upper bounds of our results. First of all, from here on every family $\mathcal{F} \subseteq 2^{[n]}$ contains sets only of size from the interval $[n/2 - n^{2/3}, n/2 + n^{2/3}]$. This leaves all our proofs valid as by Chernoff's inequality $|\{F \subseteq [n] : ||F| - n/2| \ge n^{2/3}\}| \le 2e^{-2n^{1/3}} = o(\frac{1}{n^2} {n \choose \lfloor n/2 \rfloor})$.

As we mentioned earlier, for all proofs we will use the chain partition method. This works in the following way: for a family $\mathcal{F} \subseteq 2^{[n]}$ suppose we can partition \mathbf{C}_n into $\mathbf{C}_{n,1}, \mathbf{C}_{n,2}, \ldots, \mathbf{C}_{n,l}$ such that for all $1 \leq i \leq l$ the number of pairs (F, \mathcal{C}) with $F \in \mathcal{F} \cap \mathcal{C}$ and $\mathcal{C} \in \mathbf{C}_{n,i}$ is at most $b|\mathbf{C}_{n,i}|$. Then clearly the number of pairs (F, \mathcal{C}) with $F \in \mathcal{F} \cap \mathcal{C}$ and $\mathcal{C} \in \mathbf{C}_n$ is at most $b|\mathbf{C}_n|$. Since the number of such pairs is exactly $\sum_{F \in \mathcal{F}} |F|!(n - |F|)!$, we obtain the LYM-type inequality

$$\sum_{F \in \mathcal{F}} \frac{1}{\binom{n}{|F|}} \le b,$$

and thus $|\mathcal{F}| \leq b \binom{n}{\lceil n/2 \rceil}$ holds. Therefore, in the proofs below we will end our reasoning whenever we reach a bound on the appropriate partition as mentioned above.

Proof of the upper bound in Theorem 7. Let \mathcal{F} be a strong $K_{r,t}$ -free family. We can assume that \mathcal{F} contains an antichain of size at least r as otherwise \mathcal{F} could contain at most r-1 sets of the same size and thus we would obtain $|\mathcal{F}| \leq (r-1)(n+1)$.

Let us recall the min_r-partition of \mathbf{C}_n with respect to \mathcal{F} . For a set A with $s_{\mathcal{F}}(A) \geq r$ we have $\mathbf{C}_{A,r} = \{ \mathcal{C} \in \mathbf{C}_n : A \in \mathcal{C}, \forall C \subset A, C \in \mathcal{C} : s_{\mathcal{F}}(C) < r \}.$

We claim that the number of pairs (F, \mathcal{C}) with $F \in \mathcal{F} \cap \mathcal{C}$ and $\mathcal{C} \in \mathbf{C}_{A,r}$ is at most $(2 + \frac{2(r+s-2)}{n} + o(\frac{1}{n}))|\mathbf{C}_{A,r}|$. We distinguish three types of pairs:

- 1. if $A \in \mathcal{F}$, then there are exactly $|\mathbf{C}_{A,r}|$ pairs with F = A,
- 2. any chain in $\mathbf{C}_{A,r,-}$ can be extended to (n-|A|)! chains in $\mathbf{C}_{A,r}$, thus by Corolary 19 (i) there are $(1 + \frac{2(r-1)}{n} + o(\frac{1}{n}))|\mathbf{C}_{A,r}|$ pairs with $F \subsetneq A$,
- 3. finally, any maximal chain from A to [n] can be extended to $|\mathbf{C}_{A,r,-}|$ chains in $\mathbf{C}_{A,r}$, thus Lemma 18 (ii) implies that there are $(\frac{2(s-1)}{n} + o(\frac{1}{n}))|\mathbf{C}_{A,r}|$ pairs with $A \subsetneq F$

This gives us a total of at most $\left(2 + \frac{2(r+s-2)}{n} + o(\frac{1}{n})\right) |\mathbf{C}_{A,r}|$ pairs, which completes the proof.

Now we turn our attention to complete three level posets.

Proof of the upper bound in Theorem 8. Let \mathcal{F} be a $K_{r,s,t}$ -free family. We can assume that \mathcal{F} contains an antichain of size at least $z := \max\{r, t\}$ as otherwise \mathcal{F} could contain at most z-1 sets of the same size and thus we would obtain $|\mathcal{F}| \leq (z-1)(n+1)$.

We consider the $\min_{t} - \max_{t}$ partition of \mathbf{C}_{n} and we claim that the number of pairs (F, \mathcal{C}) with $F \in \mathcal{F} \cap \mathcal{C}$ and $\mathcal{C} \in \mathbf{C}_S, \mathcal{C} \in \mathbf{C}_{A,B}$ is at most $b|\mathbf{C}_S|, b|\mathbf{C}_{A,B}|$ respectively, where b is the bound stated in Theorem 8.

First consider the "degenerate" case of \mathbf{C}_S with $S \in \mathcal{S}^-$. A chain $\mathcal{C} \in \mathbf{C}_S$ goes from \emptyset until one of the subsets S_1, S_2, \ldots, S_k of S with size |S| - 1 for which $s_{\mathcal{F}}(S_i) < r$. Then \mathcal{C} must go through S, and finally \mathcal{C} must contain a maximal chain from S to [n]. Thus $|\mathbf{C}_S| = k(|S| - 1)!(n - |S|)!$. We distinguish two types of pairs to count.

- 1. If $r \ge 2$, then applying Corollary 19 (i) we obtain that there are at most (1 + 1) $\frac{2(r-1)}{n} + o(\frac{1}{n})|\mathbf{C}_S|$ pairs (F, \mathcal{C}) with $F \subsetneq S$. Together with $\{(S, \mathcal{C}) : \mathcal{C} \in \mathbf{C}_S\}$ we have $(2 + \frac{2(r-1)}{n} + o(\frac{1}{n}))|\mathbf{C}_S|$ pairs. If r = 1, then by definition the number of pairs (F, \mathcal{C}) with $F \subseteq S$ is at most $|\mathbf{C}_S|$ as for all such pairs we must have F = S.
- 2. Applying Lemma 18 (ii) we obtain that there are at most $(\frac{2(t-1)}{n} + o(\frac{1}{n}))|\mathbf{C}_S|$ pairs (F, \mathcal{C}) with $S \subsetneq F$.

This gives a total of at most $(2 + \frac{2(r+t-2)}{n} + o(\frac{1}{n}))|\mathbf{C}_S|$ pairs. We now consider the "more natural" $A \in \mathcal{S}^+$, $A \subseteq B$ case. As there are sets in the interval [A, B], this time we distinguish three types of pairs:

- 1. If r = 1, then there is no pair (F, \mathcal{C}) with $F \subsetneq A$. If $r \ge 2$, then applying Corollary 19 (i) we obtain that there are at most $(1 + \frac{2(r-1)}{n} + o(\frac{1}{n}))|\mathbf{C}_{A,B}|$ pairs (F, \mathcal{C}) with $F \subseteq A$.
- 2. If t = 1, then there is no pair (F, \mathcal{C}) with $B \subsetneq F$. If $t \ge 2$, then applying Corollary 19 (ii) we obtain that there are at most $(1 + \frac{2(t-1)}{n} + o(\frac{1}{n}))|\mathbf{C}_{A,B}|$ pairs (F, \mathcal{C}) with $B \subsetneq F$.
- 3. If \mathcal{F} is a $K_{r,s,t}$ -free family, then $\{F \in \mathcal{F} : A \subseteq F \subseteq B\}$ is a $K_{1,s-f(r,t),1}$ -free family. Indeed, if f(r,t) = 2, then $|\{F \in \mathcal{F} : A \subseteq F \subseteq B\}| \leq s$ as these sets together with the sets of the antichain of size r below A and the sets of the antichain

of size t above B would form a copy of $K_{r,s,t}$ in \mathcal{F} . If f(r,t) = 1, say r = 1, then by the definition of the $\min_1 - \max_t$ partition, we have $A \in \mathcal{F}$ and thus $|\{F \in \mathcal{F} : A \subsetneq F \subseteq B\}| \leq s$, in particular together with A they are $K_{1,s-1,1}$ -free. If f(r,t) = 0, then the $K_{1,s-f(r,t),1}$ -free property is the same as the $K_{1,s,1}$ -free property which is possessed by $\{F \in \mathcal{F} : A \subseteq F \subseteq B\}$ as it is a subfamily of \mathcal{F} . By Lemma 21, in case (1) of Theorem 8 the number of pairs (F, \mathcal{C}) with $A \subseteq F \subseteq$ *B* is at most $m_s|\mathbf{C}_{A,B}|$, while in case (2) of Theorem 8 the number of pairs (F, \mathcal{C}) with $A \subseteq F \subseteq B$ is at most $(m_s + 1 - \frac{2^{m_s} - s + f(r,t) - 1}{\binom{m_s}{\lceil m_s/2 \rceil}})|\mathbf{C}_{A,B}|$.

Adding up the number of three types of pairs we obtain that the total number of pairs is not more than $(m_s + f(r,t) + \frac{2(r+t-2)}{n} + o(\frac{1}{n}))|\mathbf{C}_{A,B}|$ and $(m_s + 1 + f(r,t) - \frac{2(r+t-2)}{n} + o(\frac{1}{n}))|\mathbf{C}_{A,B}|$ $\frac{2^{m_s}-s+f(r,t)-1}{\binom{m_s}{\lceil m_s/2\rceil}} + \frac{2(r+t-2)}{n} + o(\frac{1}{n}))|\mathbf{C}_{A,B}| \text{ in the two respective cases of Theorem 8.}$

We continue with the proof of Theorem 9.

Proof of the upper bound of Theorem 9. Let \mathcal{F} be a $K_{r,2,t}$ -free family and let us write $r^{++} = \max\{r, 2\}, t^{++} = \max\{t, 2\}.$ We consider the $\min_{r^{++}} - \max_{t^{++}} - partition of C_n$. Just as in the proof of Theorem 8, we obtain that if $S \in S^-$ than the number of pairs (F, \mathcal{C}) with $F \in \mathcal{F} \cap \mathcal{C}$ and $\mathcal{C} \in \mathbf{C}_S$ is at most $(2 + O(\frac{1}{n}))|\mathbf{C}_S|$. Note that if $A \subseteq B$, then $|\mathcal{F} \cap \{G \in 2^{[n]} : A \subseteq G \subseteq B\}| \leq 1$ as by definition of the $\min_{r+1} - \max_{t+1}$ -partition two such sets would make \mathcal{F} contain a copy of $K_{r,2,t}$.

- Applying Corollary 19 (i) we obtain that there are at most $(1 + \frac{2(r^{++}-1)}{n} + O_r(\frac{1}{n^2}))|\mathbf{C}_{A,B}|$ pairs (F, \mathcal{C}) with $F \subsetneq A$.
- Applying Corollary 19 (ii) we obtain that there are at most $\left(1 + \frac{2(t^{++}-1)}{n} + \frac{1}{n}\right)$ $O_t(\frac{1}{n^2})|\mathbf{C}_{A,B}|$ pairs (F,\mathcal{C}) with $B \subsetneq F$.
- By the observation above, the number of pairs (F, \mathcal{C}) with $A \subseteq F \subseteq B$ is at most $|\mathbf{C}_{A,B}|.$

Proof of the upper bound of Theorem 10. Throughout the proof we will assume that all s'_i s are at least 2. This will be needed for the fact that all $m^*_{s_i}$'s are larger than 1.

First we prove (i), (ii), and (iii). Let \mathcal{F} be a strong $K_{r,s,t}$ -free family. We can assume that \mathcal{F} contains an antichain of size at least $z := \max\{r, t\}$ as otherwise \mathcal{F} could contain at most z-1 sets of the same size and thus we would obtain $|\mathcal{F}| \leq (z-1)(n+1)$. We again consider the min_r – max_t partition of \mathbf{C}_n and count the number of pairs (F, \mathcal{C}) with $F \in \mathcal{F} \cap \mathcal{C}$ and $\mathcal{C} \in \mathbf{C}_n$.

The degenerate case is identical to what we had in the proof of Theorem 8, thus we only consider the case when $A \in S^+$, $A \subseteq B$. The three types of pairs:

1. If r = 1, then there is no pair (F, \mathcal{C}) with $F \subsetneq A$. If $r \ge 2$, then applying Corollary 19 (i) we obtain that there are at most $(1 + \frac{2(r-1)}{n} + o(\frac{1}{n}))|\mathbf{C}_{A,B}|$ pairs (F, \mathcal{C}) with $F \subsetneq A.$

- 2. If t = 1, then there is no pair (F, \mathcal{C}) with $B \subsetneq F$. If $t \ge 2$, then applying Corollary 19 (ii) we obtain that there are at most $(1 + \frac{2(t-1)}{n} + o(\frac{1}{n}))|\mathbf{C}_{A,B}|$ pairs (F, \mathcal{C}) with $B \subsetneq F$.
- 3. Note that $\{F \in \mathcal{F} : A \subseteq F \subseteq B\}$ cannot contain an antichain of size s as otherwise \mathcal{F} would contain a strong copy of $K_{r,s,t}$.
 - (a) If \mathcal{F} is a strong $K_{r,4,t}$ -free family, then by Lemma 20 (i) the number of pairs (F, \mathcal{C}) with $A \subseteq F \subseteq B$ is at most $4|\mathbf{C}_{A,B}|$.
 - (b) If \mathcal{F} is a strong $K_{r,s,t}$ -free family with $s \leq c \binom{m_s^*}{\lceil m_s^*/2 \rceil}$ and s large enough, then by Lemma 20 (ii) the number of pairs (F, \mathcal{C}) with $A \subseteq F \subseteq B$ is at most $m_s^* |\mathbf{C}_{A,B}|$.
 - (c) If \mathcal{F} is a strong $K_{r,s,t}$ -free family with s large enough, then by Lemma 20 (iii) the number of pairs (F, \mathcal{C}) with $A \subseteq F \subseteq B$ is at most $(m_s^* + 1)|\mathbf{C}_{A,B}|$.

Altogether these bounds yield that the total number of pairs is at most

- 1. $(4 + f(r,t) + \frac{2(r+t-2)}{n} + o(\frac{1}{n}))|\mathbf{C}_n|$ if \mathcal{F} is strong $K_{r,4,t}$ -free,
- 2. $(m_s^* + f(r,t) + \frac{2(r+t-2)}{n} + o(\frac{1}{n}))|\mathbf{C}_n|$ if \mathcal{F} is strong $K_{r,s,t}$ -free, $s \leq c \binom{m_s^*}{\lceil m_s^*/2 \rceil}$ and s large enough,
- 3. $(m_s^*+1+f(r,t)+\frac{2(r+t-2)}{n}+o(\frac{1}{n}))|\mathbf{C}_n|$ if \mathcal{F} is strong $K_{r,s,t}$ -free and s large enough.

Now we prove (iv). Let \mathcal{F} be a strong $K_{r,s_1,s_2,\ldots,s_i,t}$ -free family. We can assume that \mathcal{F} contains an antichain of size at least $z := \max\{r, t\}$ as otherwise \mathcal{F} could contain at most z-1 sets of the same size and thus we would obtain $|\mathcal{F}| \leq (z-1)(n+1)$. Before proceeding with the formal proof, let us briefly summarize the ideas of the partition of \mathbf{C}_n that we are going to use. Just as in the case of the min_r - max_t partition we try to assign markers A_0, A_1, \ldots, A_j to every chain $\mathcal{C} \in \mathbf{C}_n$ with the following properties: (a) A_0 is the smallest set in \mathcal{C} with $s_{\mathcal{F}}(A_0) \ge r$ and (b) for every $1 \le i \le j A_i$ is the smallest set in C above A_{i-1} such that $[A_{i-1}, A_i]$ contains an antichain of size s_i . This definition enables us to build the *i*th middle level of $K_{r,s_1,\ldots,s_j,t}$ between A_{i-1} and A_i for all i with $1 \leq i \leq j$ and thus we obtain that $s^+_{\mathcal{F}}(A_j) < t$ must hold. If we were able to define all those markers, then we could apply our lemmas from the previous subsection to bound the number of pairs (F, \mathcal{C}) with $F \in \mathcal{F} \cap \mathcal{C}, \mathcal{C} \in \mathbf{C}_n$ in the different intervals $[A_i, A_{i+1}]$. Unfortunately, it might happen that not all markers can be defined. However we will index the parts of the partition of \mathbf{C}_n by chains of length at most j+1. Instead of giving formal definitions of the $\mathbf{C}_{A_0,\ldots,A_i}$'s and then verifying that they indeed form a partition of \mathbf{C}_n , we consider an arbitrary maximal chain $\mathcal{C} \in \mathbf{C}_n$ and describe the procedure how to define its markers.

- If r = 1, then A_0 is the smallest set in $\mathcal{F} \cap \mathcal{C}$,
- if $r \ge 2$, then A_0 is the smallest set in \mathcal{C} with $s_{\mathcal{F}}(A_0) \ge r$.

Note that by the assumption $s_{\mathcal{F}}([n]) \geq \max\{r, t\}$ the marker A_0 is defined for all chains $\mathcal{C} \in \mathbf{C}_n$. Let us now assume that A_{i-1} has been defined for some $1 \leq i \leq j$. If $s_{\mathcal{F}}^+(A_{i-1}) < s_i$, then our procedure is finished and \mathcal{C} belongs to $\mathbf{C}_{A_0,A_1,\ldots,A_{i-1}}$. If $s_{\mathcal{F}}^+(A_{i-1}) \geq s_i$ holds, then

• A_i is the smallest set in C such that $[A_{i-1}, A_i]$ contains an antichain of size s_i .

Note that if the procedure does not stop at A_{i-1} , then A_i exists as $[n] \in \mathcal{C}$ and $s_i \leq s_{\mathcal{F}}^+(A_{i-1})$.

Observe that a chain C in C_{A_0,\ldots,A_i} contains all A_i 's and for every $0 \le k \le j$ it goes through one of the $(|A_k| - 1)$ -subsets $A_1^k, \ldots A_{l_k}^k$ of A_k for which $[A_{k-1}, A_l^k]$ does not contain an antichain of size s_k where $A_{-1} = \emptyset$ and $s_{-1} = r$.

We now count the pairs (F, \mathcal{C}) with $F \in \mathcal{F} \cap \mathcal{C}$ and $\mathcal{C} \in \mathbf{C}_{A_0, \dots, A_i}$.

- Pairs with $F \subsetneq A_0$. If r = 1, then there is no such pair by definition of A_0 , otherwise we can apply Corollary 19 to all $(|A_0|-1)$ -subsets A of A_0 with $s_{\mathcal{F}}(A) < r$ to obtain that the number of such pairs is at most $(1 + \frac{2(r-1)}{n} + o(\frac{1}{n}))|\mathbf{C}_{A,\dots,A_i}|$.
- Pairs with $A_0 \subseteq F \subsetneq A_i$. For any $1 \leq k \leq i$ one can apply Lemma 20 to A_{k-1} and all $A_1^k, \ldots, A_{l_k}^k$ to obtain that the number of pairs with $F \in [A_{k-1}, A_l^k]$ for some $1 \leq l \leq l_k$ is at most $m_{s_k}^* | \mathbf{C}_{A_0, \ldots, A_i} |$.
- Pairs with $F \supseteq A_i$.
 - If i < j, then by definition of how we declared our process finished, we obtain $s_{\mathcal{F}}^+(A_i) < s_{i+1}$. Thus we can apply Lemma 18 (ii) to obtain that the number of such pairs is at most $(1 + \frac{2(s_{i+1}-1)}{n} + o(\frac{1}{n}))|\mathbf{C}_{A,\dots,A_i}|$.
 - If i = j and t = 1, then by definition of A_j there is no such pair.
 - If i = j and t > 1, then as \mathcal{F} is strong $K_{r,s_1,\ldots,s_j,t}$ -free, we obtain that $s_{\mathcal{F}}^+(A_j) < t$. Thus we can apply Lemma 18 (ii) to obtain that the number of such pairs is at most $(1 + \frac{2(t-1)}{n} + o(\frac{1}{n}))|\mathbf{C}_{A,\ldots,A_j}|$.

Adding up these bounds we obtain that if i = j, then the total number of pairs is at most $(f(r,t) + \sum_{k=1}^{j} m_{s_k}^* + O(\frac{1}{n}))|\mathbf{C}_{A_0,\dots,A_j}|$. If i < j holds the upper bound we obtain is $(f(r,t) + 1 + \sum_{k=1}^{i} m_{s_k}^* + O(\frac{1}{n}))|\mathbf{C}_{A_0,\dots,A_j}|$, where the extra 1 is needed only for the case t = 1. But since $s_j > 1$ holds, we have $(f(r,t) + 1 + \sum_{k=1}^{i} m_{s_k}^* + O(\frac{1}{n}))|\mathbf{C}_{A_0,\dots,A_j}| \leq (f(r,t) + \sum_{k=1}^{j} m_{s_k}^* + O(\frac{1}{n}))|\mathbf{C}_{A_0,\dots,A_j}|$.

2.2Stability and supersaturation for the butterfly poset

In this subsection, we will address the problem of minimizing the number of butterflies contained in families $\mathcal{F} \subseteq 2^{[n]}$ of fixed size $m > \Sigma(n, 2)$. It will be more convenient to count different image sets of injections of B to \mathcal{F} as *copies* of B instead of counting the number of injections. Also, the problems are equivalent as there are exactly four injections (the number of automorphisms of B) to any possible image set.

Note that whenever we add a set G to a family $\mathcal{F} \in \Sigma^*(n,2)$, the number of newly constructed butterflies in $\mathcal{F} \cup \{G\}$ will be minimized if G is "closest to the middle". If n = 2k and $\mathcal{F} = {\binom{[n]}{k-1}} \cup {\binom{[n]}{k}}$, then G should be picked from ${\binom{[n]}{k+1}}$. In this case, if F_1, F_2, F_3, G is a newly created butterfly, then $F_1, F_2 \subset F_3 \subset G$ must hold. If one adds a set G to a family \mathcal{F} from $\Sigma^*(n,2)$ with |G| > k+1, then the number of butterflies with $F_1, F_2 \subset F_3 \subset G, F_i \in \mathcal{F}$ is already larger than in the previous case. Thus, independently of parity, the minimum number of butterflies appearing when adding one new set to a family in $\Sigma^*(n,2)$ is

$$f(n) = (\lceil n/2 \rceil + 1) \binom{\lceil n/2 \rceil}{2}.$$

Therefore, if adding E new sets to a family $\mathcal{F} \in \Sigma^*(n,2)$ we will have at least $E \cdot f(n)$ butterflies. Note that if $G_1, G_2 \in {[n] \choose k+1}$ are such that $|G_1 \cap G_2| \leq k-1$, then there are no butterflies in $\mathcal{F} \cup \{G_1, G_2\}$ that contain both G_1 and G_2 . Thus it is possible to have only $E \cdot f(n)$ copies of butterfly as long as we can pick sets from $\binom{[n]}{k+1}$ with this property. We summarize our findings in the following proposition.

Proposition 22. (a) If $\mathcal{S} \subset \mathcal{F}$ for some $\mathcal{S} \in \Sigma^*(n, 2)$, then \mathcal{F} contains at least

 $(|\mathcal{F}| - \Sigma(n, 2))f(n) \text{ copies of butterflies.}$ $(b) \text{ If } \mathcal{F} = {[n] \choose \lceil n/2 \rceil - 1} \cup {[n] \choose \lceil n/2 \rceil} \cup \mathcal{E} \text{ where } \mathcal{E} \subset {[n] \choose \lceil n/2 \rceil + 1} \text{ such that } |E_1 \cap E_2| < \lceil n/2 \rceil$ $\text{ holds for all } E_1, E_2 \in \mathcal{E}, \text{ then } \mathcal{F} \text{ contains exactly } |\mathcal{E}| \cdot f(n) \text{ copies of butterflies.}$

It is known that it is possible to construct a family \mathcal{E} with the above property as long as the number of sets in \mathcal{E} is not more than $\frac{1}{n} \binom{n}{k+1}$: the families $\mathcal{E}_j = \{E \in \binom{[n]}{k+1}: \sum_{i \in E} i \equiv j \pmod{n}\}$ all possess this property, therefore the largest among them must be of size at least $\frac{1}{n} \binom{n}{k+1}$. The main result of this section, Theorem 13 states that this is best possible for all families of size $\Sigma(n,2) + E$, if E is very small and asymptotically best possible, if E is not that small.

We start with stating the celebrated LYM-inequality [12, 82, 85, 111]. This was originally stated for Sperner families, but using the fact that any k-Sperner family can be decomposed into k antichains, the statement generalizes easily to k-Sperner families. As we will not need the result in its full generality, we state it in the case k = 2.

Theorem 23 (LYM-inequality for 2-Sperner families). If $\mathcal{F} \subseteq 2^{[n]}$ is a 2-Sperner family, then the inequality

$$\sum_{F \in \mathcal{F}} \frac{1}{\binom{n}{|F|}} \le 2$$

holds.

The sum $\sum_{F \in \mathcal{F}} \frac{1}{\binom{n}{|F|}}$ is called the *Lubell-mass* of \mathcal{F} .

Corollary 24. Let $\mathcal{F} \subseteq 2^{[n]}$ be a 2-Sperner family such that one of the following holds: (a) n is odd and the number of sets $|\{G \notin \mathcal{F} : |G| = \lceil n/2 \rceil \text{ or } |G| = \lfloor n/2 \rfloor\}|$ is at least m,

(b) n is odd and the number of sets $|\{F \in \mathcal{F} : |F| \neq \lceil n/2 \rceil, \lfloor n/2 \rfloor\}|$ is at least m,

(c) n is even and the number of sets $|\{G \notin \mathcal{F} : |G| = n/2\}|$ is at least m,

(d) n is even and the number of sets $|\{F \in \mathcal{F} : |F| \neq n/2 - 1, n/2, n/2 + 1\}|$ is at least m.

Then we have the inequality $|\mathcal{F}| \leq \Sigma(n,2) - \frac{1.9m}{n}$.

Proof. Parts (a),(b),(c) immediately follow from the LYM-inequality and from the fact that the ratio of the two smallest possible terms in the Lubell-mass is

$$\binom{n}{\lfloor n/2 \rfloor - 1}^{-1} / \binom{n}{\lfloor n/2 \rfloor}^{-1} = \frac{\lfloor n/2 \rfloor + 1}{\lfloor n/2 \rfloor} \ge 1 + \frac{1.9}{n}.$$

To obtain (d), observe that the ratio of the second and third smallest possible terms in the Lubell-mass is

$$\binom{n}{n/2-2}^{-1} / \binom{n}{n/2-1}^{-1} = \frac{n/2+2}{n/2-1} \ge 1 + \frac{1.9}{n}.$$

We continue with introducing the notions of *shadow* and *shade*. If \mathcal{F} is a family of sets, then its k-shadow is $\Delta_k(\mathcal{F}) = \{G : |G| = k, \exists F \in \mathcal{F} \ G \subset F\}$. To define the k-shade of the family we have to assume the existence of an underlying set, say $\mathcal{F} \subseteq 2^{[n]}$. If so, then the k-shade is defined as $\nabla_k(\mathcal{F}) = \{G \in \binom{[n]}{k} : \exists F \in \mathcal{F} \ F \subset G\}$. The well-known theorem of Kruskal [77] and Katona [68] states which family \mathcal{F} of ksets minimizes the size of $\Delta_{k-1}(\mathcal{F})$ among all families of m sets. For calculations the following version happens to be more useful than the precise result.

Theorem 25 (Lovász, [78]). Let \mathcal{G} be a family of k-sets and let $x \ge k$ be the real number such that $\binom{x}{k} = |\mathcal{G}|$ holds. Then the family of shadows satisfies $|\Delta_{k-1}(\mathcal{G})| \ge \binom{x}{k-1}$

We will apply Theorem 25 in a slightly more general setting. If $\mathcal{F} \subset {\binom{[n]}{l}}$ with l > n/2, then a simple double counting argument and Hall's theorem show that there exists a matching from \mathcal{F} to $\Delta_{l-1}(\mathcal{F})$ such that if $F \in \mathcal{F}$ and $G \in \Delta_{l-1}(\mathcal{F})$ are matched, then $G \subset F$. Using this observation and Theorem 25, one obtains the following lemma. Part (ii) of the statement follows from the fact that $G \subset F \subset [n]$ holds if and only if $[n] \setminus F \subset [n] \setminus G$.

Lemma 26. (a) Let $\mathcal{G} \subseteq {\binom{[n]}{\geq k}}$ be a Sperner family with $\lfloor n/2 \rfloor \leq k$ and let $x \geq k$ be the real number such that $\binom{x}{k} = |\mathcal{G}|$ holds. Then the family of shadows satisfies $|\Delta_{k-1}(\mathcal{G})| \geq \binom{x}{k-1}$.

(b) Let $\mathcal{G} \subseteq \begin{pmatrix} [n] \\ \leq k \end{pmatrix}$ be a Sperner family with $\lceil n/2 \rceil \geq k$ and let $x \geq k$ be the real number such that $\binom{x}{n-k} = |\mathcal{G}|$ holds. Then the family of shades satisfies $|\nabla_{k+1}(\mathcal{G})| \geq \binom{x}{n-k-1}$.

Let us define the following functions of l and m. Let x = x(l, m) be defined by the equation $\binom{x}{l} = m$ and write $g(l, m) = \binom{x}{l-1} - \binom{x}{l}$. According to Lemma 26, if $\mathcal{G} \subseteq \binom{[n]}{\geq l}$ is a Sperner family with $|\mathcal{G}| = m$ and $l \geq \lfloor n/2 \rfloor$, then $|\Delta_{l-1}(\mathcal{G})| - |\mathcal{G}| \geq g(l, m)$ holds. This will be crucial in the proof of Lemma 28, the stability result on 2-Sperner families. When comparing the size of a 2-Sperner family \mathcal{F} to $\Sigma(n, 2)$, we will split \mathcal{F} into an upper and lower Sperner family \mathcal{F}_u and \mathcal{F}_l . Then $\mathcal{F}'_u = \{F \in \mathcal{F}_u : |F| > n/2\}$ will be replaced by $\Delta_{n/2}(\mathcal{F}'_u)$ and $\mathcal{F}'_l = \{F \in \mathcal{F}_l : |F| > n/2 - 1\}$ will be replaced by $\Delta_{n/2-1}(\mathcal{F}'_l)$. As the resulting family is 2-Sperner, we will have $|\mathcal{F}| \leq \Sigma(n, 2) - g(n/2, |\mathcal{F}'_u|) - g(n/2 - 1, |\mathcal{F}'_l|)$. To be able to do calculations with expressions involving the g function, we gather some properties of x(l, m) and g(l, m) in the following proposition.

Proposition 27. (a) If $m \le \binom{2l}{l}$, then $x(l,m) \le x(l+1,m) \le x(l,m) + 1$ holds. (b) If $x(l,m) \le 2l - 1$, then $g(l,m) \ge 0$ holds.

(c) If $m_1 + m_2 = m$ and $x(l,m) \le 2l - 1$, then $g(l,m_1) + g(l,m_2) \ge g(l,m)$ holds. (d) If $m \le \binom{2l-1}{l}$, then $g(l,m) \le g(l+1,m)$ holds. (e) For every $\varepsilon > 0$ there exists l_0 such that if $l \ge l_0$, then g(l,m) is increasing in

(e) For every $\varepsilon > 0$ there exists l_0 such that if $l \ge l_0$, then g(l,m) is increasing in the interval $0 \le m \le \binom{(2-\varepsilon)l}{l}$.

(f) If $x(l,m) \leq \overline{4l/3} - 1$, then $2m \leq g(l,m) \leq 2lm$ holds.

Proof. (a) Clearly the polynomial $\binom{x}{l+1}$ is monotone increasing in x if $x \ge l+1$ holds. Observe that $\binom{x(l,m)}{l+1} = \frac{x(l,m)-l}{l+1} \binom{x(l,m)}{l} < m$ as $x(l,m) \le 2l$ by the assumption $m \le \binom{2l}{l}$. Therefore, $x(l,m) \le x(l+1,m)$ holds. Similarly, $\binom{x(l,m)+1}{l+1} = \frac{x(l,m)+1}{l+1} \binom{x(l,m)}{l} > m$ and therefore $x(l+1,m) \le x(l,m) + 1$ holds.

To obtain (b), (c), and (f) write g(l,m) in the following form

$$g(l,m) = \binom{x}{l-1} - \binom{x}{l} = \left(\frac{l}{x-l+1} - 1\right)\binom{x}{l} = \frac{2l-x-1}{x-l+1}m.$$

(b) and (f) are straightforward and to obtain (c) note that as for fixed l we know that x(l,m) is an increasing function of m, the fraction $\frac{2l-x-1}{x-l+1}$ is decreasing in m.

To obtain (d), as $g(l,m) - g(l+1,m) = {\binom{x(l,m)}{l-1}} - {\binom{x(l+1,m)}{l}}$ we need to compare ${\binom{x(l,m)}{l-1}}$ and ${\binom{x(l+1,m)}{l}}$.

$$\frac{\binom{x(l,m)}{l-1}}{\binom{x(l+1,m)}{l}} = \frac{\binom{x(l,m)}{l-1}}{m} \frac{m}{\binom{x(l+1,m)}{l}} = \frac{\binom{x(l,m)}{l-1}}{\binom{x(l,m)}{l}} \frac{\binom{x(l+1,m)}{l+1}}{\binom{x(l+1,m)}{l}} = \frac{l}{x(l,m)-l+1} \cdot \frac{x(l+1,m)-l}{l+1} < 1.$$

where we used $x(l+1,m) \le x(l,m) + 1$ of (a).

To obtain (e) consider g(l,m) in the following form

$$g(l,m) = \binom{x}{l-1} - \binom{x}{l} = \frac{x(x-1)\dots(x-l+2)(2l-x-1)}{l!}$$

As a function of x it is a polynomial with no multiple roots, therefore between l-2 and 2l-1 it is a concave function with one maximum. Its derivative is

$$\frac{1}{l!} \left((2l - x - 1) \sum_{i=0}^{l-2} \prod_{j=0, j \neq i}^{l-2} (x - j) - \prod_{i=0}^{l-2} (x - i) \right)$$

If $x \leq (2 - \varepsilon)l$, then any product in the sum is at least an ε fraction of the product to subtract. Thus if l is large enough the derivative is positive and thus the function is increasing. As x is a monotone increasing function of m, the claim holds.

After all these preliminary results we are ready to state and prove the stability result on 2-Sperner families. Let us remind the reader that we would like to obtain a lemma that states that if a 2-Sperner family \mathcal{F} is very different from the largest one(s) (i.e. that/those in $\Sigma^*(n, 2)$), then it should be much smaller than the extremal size. The parameter with which we measure this difference is the number of sets in \mathcal{F} that do not belong to the closest extremal family, i.e. min $|\mathcal{F} \setminus \mathcal{F}^*|$ where the minimum is taken over the families in $\Sigma^*(n, 2)$.

Lemma 28. For every $\varepsilon > 0$, there exists an n_0 such that the following holds: if $n \ge n_0, m \le \binom{(1-\varepsilon)n}{\lfloor n/2 \rfloor}$ and $\mathcal{F} \subseteq 2^{[n]}$ is a 2-Sperner family with the property that for any $\mathcal{F}^* \in \Sigma^*(n,2)$ we have $|\mathcal{F} \setminus \mathcal{F}^*| \ge m$, then the following upper bound holds on the size of \mathcal{F} :

$$|\mathcal{F}| \le \Sigma(n,2) - g(\lceil n/2 \rceil + 1,m).$$

Proof. Let $\varepsilon > 0$ be fixed and let $\mathcal{F} \subseteq 2^{[n]}$ be a 2-Sperner family such that for any $\mathcal{F}^* \in \Sigma^*(n, 2)$ we have $|\mathcal{F} \setminus \mathcal{F}^*| \ge m$. Note that by Proportion 27 (f) we have $g(\lceil n/2 \rceil + 1, m) \le 2nm \le \binom{(1-0.95\varepsilon)n}{n/2}$ if *n* is large enough. Write $m' = \min_{\mathcal{F}^* \in \Sigma^*(n,2)} \{|\mathcal{F} \setminus \mathcal{F}^*|\}$. We can assume that $m' \le (\frac{1}{2} + o(1))\binom{n}{\lceil n/2 \rceil + 1}$ as otherwise $|(\binom{[n]}{\lfloor n/2 \rfloor}) \cup \binom{[n]}{\lceil n/2 \rceil}) \setminus \mathcal{F}| \ge \delta\binom{n}{\lfloor n/2 \rfloor}$ would hold for some positive δ and we would be done by Corollary 24 part (a) or (c) depending on the parity of *n*.

CASE I. $m' \ge \binom{(1-\varepsilon/2)n}{n/2}$.

If n is odd, then by Corollary 24 (b), we have $|\mathcal{F}| \leq \Sigma(n,2) - \frac{1.9m'}{n}$ and we are done as $\frac{1.9m'}{n} \geq \frac{1.9}{n} \binom{(1-\varepsilon/2)n}{n/2} \geq \binom{(1-3\varepsilon/4)n}{n/2}$ for n large enough.

If n is even, then by symmetry we can suppose that $m' = |\mathcal{F} \setminus (\binom{[n]}{n/2-1} \cup \binom{[n]}{n/2})|$. Let $\mathcal{F} = \mathcal{F}_1 \cup \mathcal{F}_2$ with $\mathcal{F}_1 = \{F \in \mathcal{F} : \exists F' \in \mathcal{F}, F' \subset F\}$ and $\mathcal{F}_2 = \mathcal{F} \setminus \mathcal{F}_1$. Let us write

$$\mathcal{F}_{1,+} = \{F \in \mathcal{F}_1 : |F| > n/2\}, \mathcal{F}_{2,+} = \{F \in \mathcal{F}_2 : |F| > n/2\}, \mathcal{F}_{-} = \{F \in \mathcal{F} : |F| < n/2 - 1\}$$

$$\mathcal{G}_{n/2} = \{ G \notin \mathcal{F}, |G| = n/2 \}, \quad \mathcal{G}_{n/2-1} = \{ G \notin \mathcal{F}, |G| = n/2 - 1 \}.$$

Observe the following bounds:

- $|\mathcal{G}_{n/2}| \leq {\binom{(1-3\varepsilon/4)n}{n/2}}$ as otherwise by Corollary 24 (d), we are done.
- $|\mathcal{F}_{1,+}| \leq {\binom{(1-3\varepsilon/4)n}{n/2}}$ as $\Delta_{n/2}(\mathcal{F}_{1,+}) \subseteq \mathcal{G}_{n/2}$ and $|\mathcal{F}_{1,+}| \leq |\Delta_{n/2}(\mathcal{F}_{1,+})|$ hold.
- $|\mathcal{F}_{-}| \leq {\binom{(1-3\varepsilon/4)n}{n/2}}$ as otherwise by Corollary 24 (c), we are done.
- By definition all sets in $\Delta_{n/2}(\mathcal{F}_{2,+}) \setminus \mathcal{G}_{n/2}$ must belong to \mathcal{F}_1 . No set below an arbitrary set of \mathcal{F}_1 belongs to \mathcal{F} , therefore all sets of $\Delta_{n/2-1}(\mathcal{F}_{2,+})$ belong to $\mathcal{G}_{n/2-1}$ except those whose complete shade belongs to $\mathcal{G}_{n/2}$. By double counting

pairs (G, G') with $G' \subset G$, |G'| = n/2 - 1, $G \in \mathcal{G}_{n/2}$ and $\nabla_{n/2}(G') \subseteq \mathcal{G}_{n/2}$ we obtain that the number of such exceptional sets is $(1 + o(1))|\mathcal{G}_{n/2}| \leq \binom{(1-2\varepsilon/3)n}{n/2}$. Let \mathcal{E} denote the family of these exceptional sets.

Let $m'' = |\mathcal{F}_{2,+}|$. By the above, we have $m' \ge m'' = m' - |\mathcal{F}_{-}| - |\mathcal{F}_{1,+}| \ge \binom{(1-0.6\varepsilon)n}{n/2}$. Also, writing $m'' = \binom{x''}{n/2+1}$ we have

$$\begin{split} |\mathcal{F}| &\leq \Sigma(n,2) - |\Delta_{n/2-1}(\mathcal{F}_{2,+})| + |\mathcal{E}| + |\mathcal{F}_{2,+}| + |\mathcal{F}_{1,+}| + |\mathcal{F}_{-}| \\ &\leq \Sigma(n,2) - |\Delta_{n/2-1}(\mathcal{F}_{2,+})| + m'' + 3\binom{(1-2\varepsilon/3)n}{n/2} \\ &\leq \Sigma(n,2) - \binom{x''}{n/2-1} - \binom{x''}{n/2+1} + 3\binom{(1-2\varepsilon/3)n}{n/2} \\ &\leq \Sigma(n,2) - \frac{1}{n^2}m'' + 3\binom{(1-2\varepsilon/3)n}{n/2}. \end{split}$$

Here the third inequality follows by Lemma 26 and the last one follows from $x'' \leq 2n - 1 + o(1)$ since $m'' \leq m' \leq (\frac{1}{2} + o(1)) \binom{n}{n/2}$. We are done as $m'' \geq \binom{(1-0.6\varepsilon)n}{n/2}$ holds.

CASE II. $m' < \binom{(1-\varepsilon/2)n}{n/2}$

Again we may assume that $|(\binom{[n]}{n/2-1} \cup \binom{[n]}{n/2}) \setminus \mathcal{F}| = m'$. Let $\mathcal{F} = \mathcal{F}_1 \cup \mathcal{F}_2$ with $\mathcal{F}_1 = \{F \in \mathcal{F} : \not \supseteq F' \in \mathcal{F}, F' \subset F\}$ and $\mathcal{F}_2 = \mathcal{F} \setminus \mathcal{F}_1$. Let us write

$$\mathcal{F}_{1,-} = \{ F \in \mathcal{F}_1 : |F| < n/2 - 1 \}, \quad \mathcal{F}_{1,+} = \{ F \in \mathcal{F}_1 : |F| > n/2 - 1 \},$$
$$\mathcal{F}_{2,-} = \{ F \in \mathcal{F}_1 : |F| < n/2 \}, \quad \mathcal{F}_{2,+} = \{ F \in \mathcal{F}_1 : |F| > n/2 \}.$$

To bound the size of \mathcal{F}_1 note that \mathcal{F}_1 is disjoint both from $\Delta_{n/2-1}(\mathcal{F}_{1,+})$ and $\nabla_{n/2-1}(\mathcal{F}_{1,-})$. Similarly, \mathcal{F}_2 is disjoint both from $\Delta_{n/2}(\mathcal{F}_{2,+})$ and $\nabla_{n/2}(\mathcal{F}_{2,-})$. By Lemma 26 we obtain

$$\begin{aligned} |\mathcal{F}| &= |\mathcal{F}_1| + |\mathcal{F}_2| \le \binom{n}{n/2 - 1} - g(n/2 + 2, |\mathcal{F}_{1,-}|) - g(n/2, |\mathcal{F}_{1,+}|) + \\ &+ \binom{n}{n/2} - g(n/2 + 1, |\mathcal{F}_{2,-}|) - g(n/2 + 1, |\mathcal{F}_{2,+}|) \\ &\le \Sigma(n, 2) - g(n/2 + 1, |\mathcal{F}_{2,-}| + |\mathcal{F}_{2,+}| + |\mathcal{F}_{1,-}|) - g(n/2, |\mathcal{F}_{1,+}|), \end{aligned}$$

where we used Proposition 27 (c) and (d). Let us partition $\mathcal{F}_{1,+}$ into $\mathcal{F}_{1,+,n/2} \cup \mathcal{F}_{1,+,+}$ with $\mathcal{F}_{1,+,n/2} = \{F \in \mathcal{F}_{1,+} : |F| = n/2\}$. As $\mathcal{F}_{1,+}$ is Sperner, $\mathcal{F}_{1,+,n/2}$ and $\Delta_{n/2}(\mathcal{F}_{1,+,+})$ are disjoint and thus $|\mathcal{F}_{1,+,n/2} \cup \Delta_{n/2}(\mathcal{F}_{1,+,+})| \geq |\mathcal{F}_{1,+}| + g(n/2+1, |\mathcal{F}_{1,+,+}|)$. Also $s = |\mathcal{F}_{1,+,n/2} \cup \Delta_{n/2}(\mathcal{F}_{1,+,+})| \leq {n-1 \choose n/2}$ and thus $g(s) \geq 0$ holds. Therefore, we obtain $|\mathcal{F}_1| \leq {n \choose n/2-1} - g(n/2+1, |\mathcal{F}_{1,+,+}|) - g(n/2+2, |\mathcal{F}_{1,-}|)$. By Proposition 27 (c), this strengthens the above arrayed inequality to

$$|\mathcal{F}| \le \Sigma(n,2) - g(n/2+1, |\mathcal{F}_{1,+,+}| + |\mathcal{F}_{1,-} \cup \mathcal{F}_{2,-}| + |\mathcal{F}_{2,+}|).$$

Note that $m \leq |\mathcal{F}_{1,+,+} \cup \mathcal{F}_{1,-} \cup \mathcal{F}_{2,-} \cup \mathcal{F}_{2,+}|$ as $\mathcal{F}_{1,+,n/2} \subseteq \binom{n}{n/2}$. Therefore, we are done by Proposition 27 (e).

Having proved Lemma 28 we can now turn our attention to butterfly-free families containing chains of length 3. Our main tool to bound their size is the following LYM-type inequality.

Lemma 29. Let $\mathcal{F} \subset 2^{[n]} \setminus \{\emptyset, [n]\}$ be a butterfly-free family and let \mathcal{M} be defined as $\{M \in \mathcal{F} : \exists F, F' \in \mathcal{F} \mid F \subsetneq M \subsetneq F'\}.$

$$\sum_{F \in \mathcal{F}} \frac{1}{\binom{n}{|F|}} + \sum_{M \in \mathcal{M}} \left(1 - \frac{n}{|M|(n-|M|)} \right) \frac{1}{\binom{n}{|M|}} \le 2.$$

Proof. We count the pairs (F, \mathcal{C}) where \mathcal{C} is a maximal chain in [n] and $F \in \mathcal{F} \cap \mathcal{C}$ holds. For fixed F there are |F|!(n - |F|)! maximal chains containing F. For any maximal chain \mathcal{C} , we have $|\mathcal{F} \cap \mathcal{C}| \leq 3$ as a 4-chain is a butterfly. If $|\mathcal{C} \cap \mathcal{F}| = 3$, then \mathcal{C} contains exactly one member $M \in \mathcal{M}$ as otherwise \mathcal{F} would contain a 4-chain. Note that for any $M \in \mathcal{M}$ there exist unique sets $F_{1,M}, F_{2,M} \in \mathcal{F}$ with $F_{1,M} \subset M \subset F_{2,M}$. Indeed, sets with these containment properties exist by definition of \mathcal{M} and M cannot contain two sets $F', F'' \in \mathcal{F}$ as $F_{2,M}, M, F', F''$ would constitute a butterfly. Similarly if $M \subset F^*, F^{**} \in \mathcal{F}$ holds, then $F_{1,M}, M, F^*, F^{**}$ would constitute a butterfly. Therefore, all maximal chains \mathcal{C} that contain M with $|\mathcal{F} \cap \mathcal{C}| = 3$ must contain $F_{1,M}$ and $F_{2,M}$ and thus their number is at most (|M| - 1)!(n - |M| - 1)!. (Here we used that $\emptyset, [n] \notin \mathcal{F}$.) Moreover, for any maximal chain \mathcal{C} with $M \in \mathcal{C}, F_{1,M}, F_{2,M} \notin \mathcal{C}$, we have $|\mathcal{C} \cap \mathcal{F}| = 1$ and the number of such chains is at least (|M|! - (|M| - 1)!)((n - |M|)! - (n - |M| - 1)!). We obtained the following inequality

$$\begin{split} \sum_{F \in \mathcal{F}} |F|!(n-|F|)! &\leq 2n! + \sum_{M \in \mathcal{M}} (|M|-1)!(n-|M|-1)! - \\ &- \sum_{M \in \mathcal{M}} ((|M|-1)(|M|-1)!)((n-|M|-1)(n-|M|-1)!). \end{split}$$

Rearranging and dividing by n! we obtain the claim of the lemma.

Corollary 30. Let $\mathcal{F} \subseteq 2^{[n]}$ be a butterfly-free family with \emptyset , $[n] \notin \mathcal{F}$ and let us write $\mathcal{M} = \{M \in \mathcal{F} : \exists F, F' \in \mathcal{F} \ F \subset M \subset F'\}$. If n is large enough, then $|\mathcal{F}| \leq \Sigma(n, 2) - 9|\mathcal{M}|/20$.

Proof. As \emptyset , $[n] \notin \mathcal{F}$, for any $M \in \mathcal{M}$ we have $2 \leq |M| \leq n-2$ and thus $1 - \frac{n}{|M|(n-|M|)} \geq 9/20$ for every $M \in \mathcal{M}$ if n is large enough. Therefore, the two summands in Lemma 29 corresponding to a set $M \in \mathcal{M}$ is at least 29/20 as much as the summand corresponding to a set $F \in \mathcal{F} \setminus \mathcal{M}$ with |F| = |M|. The number of possible summands in Lemma 29 is $\Sigma(n, 2)$ if $\mathcal{M} = \emptyset$ and each pair of sets in \mathcal{M} leaves place for one less summand.

Now we have all auxiliary results in hand to prove the stability result on B-free families.

Proof of Theorem 12. Let $\mathcal{F} \subseteq 2^{[n]}$ be a butterfly-free family satisfying the conditions of the theorem. If $\emptyset \in \mathcal{F}$ or $[n] \in \mathcal{F}$, then $\mathcal{F} \setminus \{\emptyset\}$ or $\mathcal{F} \setminus \{[n]\}$ does not contain the poset \vee or \wedge , where \vee is the poset with three elements one smaller than the other

two and \wedge is the poset with three elements one larger than the other two. In either case by a theorem of Katona and Tarján [71], we have $|\mathcal{F}| \leq (1 + O(\frac{1}{n})) \binom{n}{n/2}$. Thus we may assume \emptyset , $[n] \notin \mathcal{F}$. If $\mathcal{M} = \{M \in \mathcal{F} : \exists F, F' \in \mathcal{F} \ F \subset M \subset F'\}$ contains at least 10m/19 sets, then we are done by Corollary 30. If $|\mathcal{M}| \leq 10m/19$, then $\mathcal{F} \setminus \mathcal{M}$ is 2-Sperner and $|(\mathcal{F} \setminus \mathcal{M}) \setminus \mathcal{F}^*| \geq 9m/19$ for every $\mathcal{F}^* \in \Sigma^*(n, 2)$ and thus by Lemma 28 we obtain $|\mathcal{F} \setminus \mathcal{M}| \leq \Sigma(n, 2) - g(\lceil n/2 \rceil + 1, 9m/19) \leq \Sigma(n, 2) - 18m/19$ as we can use Proposition 27 (f) by the assumption on m. Therefore, $|\mathcal{F}| \leq \Sigma(n, 2) - 8m/19$ holds.

We can start working towards the proof of the supersaturation result. Since any new set added to a maximal butterfly-free family yields an additional copy of the butterfly poset, a family with few butterflies must contain an almost extremal butterfly-free family. To deal with families \mathcal{F} containing almost extremal butterfly-free families \mathcal{G} , we have to prove that most sets in $\mathcal{F} \setminus \mathcal{G}$ behave very similarly to the extra sets in the conjectured extremal families. We formalize this handwaving statement in the following theorem.

Theorem 31. For any $\varepsilon > 0$ there exists an n_0 such that for any $n \ge n_0$ the following holds provided m satisfies $\log m = o(n)$ and $n/2 - \sqrt{n} \le k \le n/2 + \sqrt{n}$: let $\mathcal{F} \subset {[n] \choose k}$ with $|\mathcal{F}| = {n \choose k} - m$. Then the number of sets in ${[n] \choose \ge k+1}$ that contain fewer than $(1 - \varepsilon)k$ sets from \mathcal{F} is o(m).

Before we start the proof of Theorem 31, let us introduce some notation and an *isoperimetric problem* due to Kleitman and West (according to Harper [63]). Given a graph G and a positive integer $m \leq |V(G)/2|$, the isoperimetric problem asks for the minimum number of edges $e(X, V(G) \setminus X)$ that go between an *m*-element subset X of V(G) and its complement. For regular graphs, this problem is equivalent to finding the the maximum number of edges e(X) in an subgraph of G induced by an *m*-subset X of the vertices. Indeed, in a d-regular graph we have $d|X| = 2e(X) + e(X, V(G) \setminus X)$.

Kleitman and West asked [73, 74] for the solution of the isoperimetric problem in the Hamming graph H(n, k) whose vertex set is $\binom{[n]}{k}$ and two k-subsets are connected if their intersection has size k-1. Harper [63] introduced and solved a continuous version of this problem. Here we summarize some of his findings. The shift operation $\tau_{i,j}$ is a widely used tool in extremal set theory. It is defined by

$$\tau_{i,j}(F) = \begin{cases} F \setminus \{j\} \cup \{i\} & \text{if } j \in F, i \notin F \text{ and } F \setminus \{j\} \cup \{i\} \notin \mathcal{F} \\ F & \text{otherwise.} \end{cases}$$

And the shift of a family is defined as $\tau_{i,j}(\mathcal{F}) = \{\tau_{i,j}(F) : F \in \mathcal{F}\}.$

Harper proved that in the Hamming graph we have $e(\mathcal{F}) \leq e(\tau_{i,j}(\mathcal{F}))$ for any family $\mathcal{F} \subseteq {\binom{[n]}{k}}$ and $i, j \in [n]$. Therefore, it is enough to consider the isoperimetric problem for *left shifted* families, i.e. families for which $\mathcal{F} = \tau_{i,j}(\mathcal{F})$ holds for all pairs i < j.

The characteristic vector of a subset F of [n] is a 0-1 vector x_F of length n with $x_F(i) = 1$ if $i \in F$ and $x_F(i) = 0$ if $i \notin F$. 0-1 vectors of length n with exactly k one entries are clearly in one-to-one correspondence with $\binom{[n]}{k}$. But also, one can consider non-negative integer vectors of length k for any set $F \in \binom{[n]}{k}$ such that $y_F(j) = i_j - j$

where i_j is the index of the *j*th one entry of x_F . For any set $F \in {\binom{[n]}{k}}$ the entries of y_F are non-decreasing, as $i_j - j$ is the number of zero coordinates of x_F before the *j*th 1-coordinate. Also, $0 \le y_F(1) \le y_F(2) \le \cdots \le y_F(k) \le n-k$ hold.

Such vectors form the poset $L_{k,n-k}$ under coordinatewise ordering, i.e. $L_{a,b} = \{x \in [0,b]^a : x(1) \leq x(2) \leq \cdots \leq x(a)\}$ and $x \leq_{L_{a,b}} y$ if and only if $x(i) \leq y(i)$ for all $1 \leq i \leq a$. It was shown by Harper that a family $\mathcal{F} \subset {[n] \choose k}$ is left shifted if and only if the set $\{y_F : F \in \mathcal{F}\}$ is a *downset* in $L_{k,n-k}$ (a set D is a downset in a poset P if $d' \leq_P d \in D$ implies $d' \in D$). If F, F' are endpoints of an edge in H(n,k), then for some $i \neq j$ we have $x_F(i) = 0, x_{F'}(i) = 1, x_F(j) = 1, x_{F'}(j) = 0$ and $x_F(l) = x_{F'}(l)$ for all $l \in [n], l \neq i, j$. If i < j, then this means that F' could be obtained from F by using $\tau_{i,j}$ and therefore $y_{F'} \leq_{L_{k,n-k}} y_F$ holds. Moreover, the number of edges for which F is the "upper endpoint" is $r(y_F) = \sum_{i=1}^k y_F(i)$. If \mathcal{F} is left shifted and $F \in \mathcal{F}$, then all lower endpoints of such edges belong to \mathcal{F} , thus the number of edges spanned by \mathcal{F} in H(n,k) is $\sum_{F \in \mathcal{F}} r(y_F)$. Therefore, the isoperimetric problem in H(n,k) is equivalent to maximizing $\sum_{y \in Y} r(y)$ over all downsets $Y \subset L_{k,n-k}$ of a fixed size.

We will use only the following simple observation to prove Theorem 31.

Proposition 32. Let $\mathcal{F} \subseteq {\binom{[n]}{k}}$ be a family of size $m < {\binom{\delta n}{\delta n/2}}$, with $0 \le \delta < 1$ and $k > \delta n/2$. Then in H(n,k) we have $e(\mathcal{F}) \le \delta mn^2$.

Proof. Suppose not and let \mathcal{F} be a left shifted counterexample and thus we have $\sum_{F \in \mathcal{F}} r(y_F) \geq \delta m n^2$. Therefore, there must be an $F \in \mathcal{F}$ with $r(y_F) \geq \delta n^2$. Note that for such a vector, we have $y_F(k - \delta n/2) \geq \delta n/2$ as otherwise $r(y_F) \leq r(y^*) \leq \delta n^2$ would hold where $y^*(i) = \delta n/2$ if $i \leq k - \delta n/2$ and $y^*(i) = n - k$ if $i > k - \delta n/2$. As \mathcal{F} is left shifted, the set $Y_{\mathcal{F}} = \{y_F : F \in \mathcal{F}\}$ is a downset in $L_{k,n-k}$. Any vector $y \in L_{k,n-k}$ with $y_i = 0$ for $i \leq k - \delta n/2$ and $y(i) \leq \delta n/2$ for $i > k - \delta n/2$ satisfies $y \leq_{L_{k,n-k}} y_F$. Therefore, all those vectors belong to $Y_{\mathcal{F}}$. The number of such vectors is $\binom{\delta n}{\delta n/2}$.

Proof of Theorem 31. Let $\overline{\mathcal{F}} = {\binom{[n]}{k}} \setminus \mathcal{F}$ and thus $|\overline{\mathcal{F}}| = m$. We want to bound the number of sets of which the shadow is contained in $\overline{\mathcal{F}}$ with the exception of at most $(1 - \varepsilon)k$ sets. Let $\mathcal{G} \subset 2^{[n]}$ denote the family of such sets and write $\mathcal{G}_l = \{G \in \mathcal{G} : |G| = l\}$. To bound $|\mathcal{G}_{k+1}|$ we double count the pairs F_1, F_2 of sets in $\overline{\mathcal{F}}$ with $|F_1 \cap F_2| = k - 1$. As $\overline{\mathcal{F}}$ has size m, by applying Proposition 32 with a sequence $\delta_n \to 0$, we obtain the number of such pairs is $o(mn^2)$. On the other hand for every such pair there exists at most one $G \in \mathcal{G}_{k+1}$ with $F_1, F_2 \subset G$ (namely, $F_1 \cup F_2$). Thus the number of such pairs is at least $|\mathcal{G}_{k+1}| {\varepsilon \choose 2}$. Therefore, we obtain $|\mathcal{G}_{k+1}| {\varepsilon \choose 2} = o(mn^2)$. Rearranging and the assumption on k yields that $|\mathcal{G}_{k+1}| = o(m)$.

To bound $|\mathcal{G}_l|$ for values of l larger than k + 1, observe that $\Delta_{k+1}(\mathcal{G}_l) \subseteq \mathcal{G}_{k+1}$ holds for all l > k + 1. Let x denote the real number for which $|\mathcal{G}_{k+1}| = \binom{x}{k+1}$ holds. By Theorem 25, we obtain that $|\mathcal{G}_l| \leq \binom{x}{l}$ holds. By the assumption on m and k, we see that x = k + 1 + o(k) and thus by Proposition 27 (f) we have $\binom{x}{l+1} \leq \frac{1}{2}\binom{x}{l}$. This gives

$$|\mathcal{G}| = \sum_{l=k+1}^{n} |\mathcal{G}_l| \le \sum_{l=k+1}^{n} \binom{x}{l} \le 2|\mathcal{G}_{k+1}| = o(m).$$

We will apply Theorem 31 only with $k = \lfloor n/2 \rfloor - 1, \lfloor n/2 \rfloor, \lfloor n/2 \rfloor + 1$ and $\lfloor n/2 \rfloor + 2$.

Now we are ready to prove our main result. Note that by Proposition 22 we only have to deal with families $\mathcal{F} \subseteq 2^{[n]}$ that do not contain any $\mathcal{S} \in \Sigma^*(n,2)$. When bounding the number of butterflies in \mathcal{F} we will distinguish two cases depending on $m = \min_{\mathcal{S} \in \Sigma^*(n,2)} |\mathcal{S} \setminus \mathcal{F}'|$. In the harder case, when m is small, we will only count copies $F, F_1, F_2, F_3 \in \mathcal{F}$ where $F \in {[n] \choose {\geq \lceil n/2 \rceil + 1}}$, $F_1 \in {[n] \choose {\lceil n/2 \rceil}}$, $F_2, F_3 \in {[n] \choose {\lceil n/2 \rceil - 1}}$ with $F_2, F_3 \subset$ $F_1 \subset F$ or $F \in {[n] \choose {\leq \lceil n/2 \rceil - 2}}$, $F_1 \in {[n] \choose {\lceil n/2 \rceil - 1}}$, $F_2, F_3 \in {[n] \choose {\lceil n/2 \rceil}}$ with $F_2, F_3 \supset F_1 \supset F$.

Proof of Theorem 13 part (a) Let $\mathcal{F} \subseteq 2^{[n]}$ be a family containing $\Sigma(n,2) + E$ sets and let \mathcal{F}' be a maximum size butterfly-free subfamily of \mathcal{F} . Let m be defined by $\min_{\mathcal{S}\in\Sigma^*(n,2)} |\mathcal{S}\setminus\mathcal{F}'|$. If $m \geq 6f(n)E$ holds, then by Theorem 12 we have $|\mathcal{F}'| \leq \Sigma(n,2) - Ef(n)$ and thus $|\mathcal{F}\setminus\mathcal{F}'| \geq E(f(n)+1)$. As \mathcal{F}' is a maximum butterfly-free subfamily of \mathcal{F} , every set $F \in \mathcal{F} \setminus \mathcal{F}'$ forms a butterfly with 3 other sets from \mathcal{F}' . Thus the number of butterflies in \mathcal{F} is at least $|\mathcal{F}\setminus\mathcal{F}'|$. This finishes the proof if $m \geq 6f(n)E$ holds.

Suppose next that $m \leq 6f(n)E$ holds. Note that as $f(n) \leq n^3$, $m \leq 6n^3E = o(\binom{\varepsilon n}{\varepsilon n/2})$ for any positive ε . Without loss of generality we can assume that $|(\binom{[n]}{\lfloor n/2 \rfloor - 1}) \cup \binom{[n]}{\lfloor n/2 \rfloor})| \leq F| = m$ and thus $|\mathcal{F} \setminus (\binom{[n]}{\lfloor n/2 \rfloor - 1} \cup \binom{[n]}{\lfloor n/2 \rfloor})| = m + E$ hold. Let us write $k = \lceil n/2 \rceil - 1$ and fix an $\varepsilon > 0$ and pick $\varepsilon' > 0$ with the property that $(1 - \varepsilon')^4/2 \geq 1 - \varepsilon$. Applying Theorem 31 to $\mathcal{F} \cap \binom{[n]}{k}$ we obtain that the family $\mathcal{F}_{b,k+1} = \{F \in \mathcal{F} \cap \binom{[n]}{k+1}\}$: $|\Delta_k(F) \cap \mathcal{F}| \leq (1 - \varepsilon')k\}$ has size o(m). Let us apply Theorem 31 again, this time to $\mathcal{F}_{g,k+1} = (\mathcal{F} \cap \binom{[n]}{k+1}) \setminus \mathcal{F}_{b,k+1}$. We obtain that the family $\mathcal{F}_{b,\geq k+2} = \{F \in \binom{[n]}{\geq k+2}\}$: $|\Delta_{k+1}(F) \cap \mathcal{F}_{g,k+1}| \leq (1 - \varepsilon')k\}$ has size o(m). With an identical argument applied to $\overline{\mathcal{F}} = \{[n] \setminus F : F \in \mathcal{F}\}$, one can show that the families $\mathcal{F}_{b,k} = \{F \in \mathcal{F} \cap \binom{[n]}{k} : |\nabla_{k+1}(F) \setminus \mathcal{F}_{b,k}| \le (1 - \varepsilon')k\}$ and $\mathcal{F}_{b,\leq k-1} = \{F \in \mathcal{F} \cap \binom{[n]}{\leq k-1}\} : |\nabla_k(F) \cap (\mathcal{F} \setminus \mathcal{F}_{b,k})| \le (1 - \varepsilon')k\}$ both have size o(m).

Let us pick a set $F \in \mathcal{F}_g = \mathcal{F} \setminus \left(\binom{[n]}{k} \cup \binom{[n]}{k+1} \cup \mathcal{F}_{b, \leq k-1} \cup \mathcal{F}_{b, \geq k+2} \right)$ and note that the number of such sets is $m + E - o(m) \geq E$.

Claim 33. For every $F \in \mathcal{F}_g$ there exist at least $(1 - \varepsilon)f(n)$ copies of the butterfly poset that contain only F from $\mathcal{F} \setminus (\binom{[n]}{k} \cup \binom{[n]}{k+1})$.

Proof of Claim Assume $|F| \ge k+2$. Then as $F \in \mathcal{F}_g$ there are at least $(1-\varepsilon)k$ sets F' in $\Delta_{k+1}(F) \cap \mathcal{F}_{g,k+1}$. For every $F' \in \mathcal{F}_{g,k+1}$ we have $|\Delta_k(F') \cap \mathcal{F}| \ge (1-\varepsilon')k$. Since every four-tuple F, F', F_1, F_2 forms a butterfly where $F_1, F_2 \in \Delta_k(F') \cap \mathcal{F}$ we obtain that the number of butterflies containing only F from $\mathcal{F} \setminus \left({n \choose k} \cup {n \choose k+1} \right)$ is at least $(1-\varepsilon')k {(1-\varepsilon')k \choose 2} \ge (1-\varepsilon')^4 k^3/2 \ge (1-\varepsilon)f(n)$ if n and thus k are large enough. The proof of the case when $|F| \le k-1$ is similar.

The above claim finishes the proof of Theorem 13 part (a).

To obtain part (b) of Theorem 13 we need better bounds on the number of "bad sets". We start with the following folklore proposition.

Proposition 34. Let U_1, \ldots, U_l be sets of size u such that $|U_i \cap U_j| \leq 1$ holds for any $1 \leq i < j \leq l$. Then we have $|\bigcup_{i=1}^l U_i| \geq l \cdot \frac{2u-l}{2}$.

Proof. By the condition on the intersection sizes we have $|U_i \setminus \bigcup_{j=1}^{i-1} U_j| \ge u - i + 1$ and thus $|\bigcup_{i=1}^{l} U_i| \ge \sum_{i=1}^{l} u - i + 1$.

Corollary 35. Let $\mathcal{F} \subset {\binom{[n]}{k}}$ with $|\mathcal{F}| = {\binom{n}{k}} - m$. Then the number of sets G in ${\binom{[n]}{k+1}}$ that contain fewer than $k+1-2\sqrt{m}$ sets from \mathcal{F} is at most \sqrt{m} provided $m \leq k^2$ and $n/2 - \sqrt{n} \leq k \leq n/2 + \sqrt{n}$. The number of such sets from ${\binom{[n]}{\geq k+1}}$ is at most $2\sqrt{m}$.

Proof. For any set $G \in {[n] \choose k+1}$ with $|\Delta_k(G) \cap \mathcal{F}| < k+1-2\sqrt{m}$ one can consider a family \mathcal{H}_G of $2\sqrt{m}$ sets from $\Delta_k(G) \setminus \mathcal{F}$. Clearly, if $G' \in {[n] \choose k+1}$ is another set with $|\Delta_k(G) \cap \mathcal{F}| < k+1-2\sqrt{m}$, then $|\mathcal{H}_G \cap \mathcal{H}_{G'}| \leq |\Delta_k(G) \cap \Delta_k(G')| \leq 1$. The sets \mathcal{H}_G satisfy the condition of Proposition 34. Thus if the number of such G's is more than \sqrt{m} , then $|\overline{\mathcal{F}}| > \sqrt{m} \cdot \frac{4\sqrt{m}-\sqrt{m}}{2} = m$ which is a contradiction.

The proof of the second statement that deals with sets of larger size is as in the proof of Theorem 31. $\hfill\blacksquare$

Proof of Theorem 13 part (b) Let $\mathcal{F} \subseteq 2^{[n]}$ be a family of sets with $|\mathcal{F}| = \Sigma(n,2) + E$ where $E = E_n \leq \frac{n}{100}$. Let m be defined by $\min_{\mathcal{S} \in \Sigma^*(n,2)} |\mathcal{S} \setminus \mathcal{F}|$. We will write $k + 1 = \lceil n/2 \rceil$ and assume that $m = (\binom{[n]}{k} \cup \binom{[n]}{k+1}) \setminus \mathcal{F}$. We will consider four cases with respect to m.

CASE I. $m \ge 6f(n)E$

Just as in the proof of Theorem 13 part a), we consider a maximal butterfly-free subfamily $\mathcal{F}' \subseteq \mathcal{F}$ with $(\binom{[n]}{k} \cup \binom{[n]}{k+1}) \cap \mathcal{F} \subseteq \mathcal{F}'$. By Corollary 30, $|\mathcal{F}'| < \Sigma(n, 2) - f(n)E$ and thus \mathcal{F} contains at least $|\mathcal{F}| - |\mathcal{F}'| > Ef(n)$ copies of the butterfly poset.

CASE II. $\frac{n}{10} \le m < 6f(n)E$

We again repeat the argument of part (a). By applying Theorem 31 twice with $\varepsilon = 1/4$, we obtain that for $E + m - o(m) \ge (11 - o(1))E$ sets $F \in \mathcal{F} \setminus (\binom{[n]}{k} \cup \binom{[n]}{k+1})$ the number of copies of the butterfly poset that contains only F from $\mathcal{F} \setminus (\binom{[n]}{k} \cup \binom{[n]}{k+1})$ is at least $(\frac{27}{64} - o(1))f(n)$ and thus the number of butterflies in \mathcal{F} is much larger than f(n)E.

CASE III. $50 \le m < \frac{n}{10}$

We try to imitate the proof of the second case of part (a). Applying Corollary 35 to $\mathcal{F} \cap {\binom{[n]}{k}}$ we obtain that the family $\mathcal{F}_{b,k+1} = \{F \in \mathcal{F} \cap {\binom{[n]}{k+1}} : |\Delta_k(F) \setminus \mathcal{F}| \leq k+1-2\sqrt{m}\}$ has size at most \sqrt{m} . Let us apply Corollary 35 again, this time to $\mathcal{F}_{g,k+1} = (\mathcal{F} \cap {\binom{[n]}{k+1}}) \setminus \mathcal{F}_{b,k+1}$. We obtain that the family $\mathcal{F}_{b,\geq k+2} = \{F \in {\binom{[n]}{\geq k+2}} : |\Delta_{k+1}(F) \cap \mathcal{F}_{g,k+1}| \leq k+2-2\sqrt{m}\}$ has size ar most $2\sqrt{m} + \sqrt{m} \leq 3\sqrt{m}$. With an identical argument applied to $\overline{\mathcal{F}} = \{[n] \setminus F : F \in \mathcal{F}\}$, one can show that the families $\mathcal{F}_{b,k} = \{F \in \mathcal{F} \cap {\binom{[n]}{k}} : |\nabla_{k+1}(F) \setminus \mathcal{F}| \leq n-k-2\sqrt{m}\}$ and $\mathcal{F}_{b,\leq k-1} = \{F \in \mathcal{F} \cap {\binom{[n]}{\leq k-1}} : |\nabla_k(F) \cap (\mathcal{F} \setminus \mathcal{F}_{b,k})| \leq n-k+1-2\sqrt{m}\}$ both have size at most $3\sqrt{m}$.

Let us pick a set $F \in \mathcal{F}_g = \mathcal{F} \setminus (\binom{[n]}{k} \cup \binom{[n]}{k+1} \cup \mathcal{F}_{b, \leq k-1} \cup \mathcal{F}_{b, \geq k+2})$ and note that the number of such sets is at least $m + E - 6\sqrt{m}$. The number of copies of butterflies in \mathcal{F} with F being the only member from $\mathcal{F} \setminus (\binom{[n]}{k} \cup \binom{[n]}{k+1})$ is at least $(k+2-2\sqrt{m})\binom{k+1-2\sqrt{m}}{2}$, and thus the number of butterflies in \mathcal{F} is at least

$$\begin{split} (E+m-6\sqrt{m})(k+2-2\sqrt{m})\binom{k+1-2\sqrt{m}}{2} &\geq (E+\sqrt{m})(k+2-2\sqrt{m})\binom{k+1-2\sqrt{m}}{2} \\ &\geq Ef(n)+\sqrt{m}\cdot\frac{k^3}{4}-E\frac{\sqrt{m}(k+2)^2}{2} > Ef(n), \end{split}$$

where we used $m - 6\sqrt{m} \ge \sqrt{m}$ as $m \ge 50$, $k - 2\sqrt{m} = (1 - o(1))k$ as $m \le \frac{n}{10}$ and also $E \le \frac{n}{100}$.

CASE IV. 0 < m < 50

In this case, every set in $\binom{[n]}{\geq k+2}$ contains at least k+2-m sets from $\mathcal{F} \cap \binom{[n]}{k+1}$ and every set in $\binom{[n]}{\geq k+1}$ contains at least k+1-m sets from $\mathcal{F} \cap \binom{[n]}{k}$. Similar statements hold for sets in $\binom{[n]}{\leq k}$ and $\binom{[n]}{k-1}$. Therefore, all E+m sets F of $\mathcal{F} \setminus (\binom{[n]}{k} \cup \binom{[n]}{k+1})$ are contained in at least $(k+2-m)\binom{k+1-m}{2}$ butterflies that contain only F from $\mathcal{F} \setminus (\binom{[n]}{k} \cup \binom{[n]}{k+1})$. Thus the number of butterflies in \mathcal{F} is at least $(E+m)(k+2-m)\binom{k+1-m}{2} \ge E(k+2)(k+1)k/2 + m(k+2)(k+1)k/2 - 7Em(k+2)^2/4$. This is strictly larger than Ef(n) = E(k+2)(k+1)k as $E \le n/10 \le k/4$.

The case when m equals 0, was dealt with by Proposition 22.
2.3 Generalized forbidden subposet problems

In this subsection, we prove bounds on the maximum number of k-chains in P-free families to obtain Theorem 16. In our proofs, we will use the class of complete multilevel posets $K_{a_1,a_2,...,a_s}$. Observe that any poset P of height l is a subposet of the l-level poset $K_{|P|-l+1,|P|-l+1,...,|P|-l+1}$. First, we prove the special case on the number of 2-chains.

Theorem 36. (i) For any poset P of height at least 3, we have

$$La(n, P, P_2) = \Theta(La(n, P_3, P_2)).$$

Moreover,

$$La(n, P_3, P_2) \le La(n, P, P_2) \le La(n, P_{|P|}, P_2) \le \left(\left\lfloor \frac{(|P| - 1)^2}{4} \right\rfloor + o(1) \right) \cdot La(n, P_3, P_2) \le La(n, P_3, P_3) \le La(n,$$

(ii) For any connected poset P of height 2 with at least 3 elements, we have

$$\Omega\left(\binom{n}{\lfloor n/2 \rfloor}\right) = La(n, P, P_2) = O(n2^n).$$

Proof of Theorem 36. To prove (i), observe first that any *P*-free family is $P_{|P|}$ -free, and if the height of *P* is at least 3, then any P_3 -free family is *P*-free. This shows the first two inequalities. To prove the last inequality first observe that by Theorem 15, it is enough to consider families \mathcal{F} consisting of |P| - 1 full levels and determine the value

$$\max_{0 \le i_1 < i_2 < \dots < i_{|P|-1} \le n} \sum_{1 \le l < j \le |P|-1} \binom{n}{i_j} \binom{i_j}{i_l}.$$

As claimed by Theorem 15, $\binom{n}{i_j}\binom{i_j}{i_l}$ is maximized when $i_l, i_j - i_l$, and $n - i_j$ differ by at most 1. Furthermore, if $i_j \notin ((2/3 - \varepsilon)n, (2/3 + \varepsilon)n)$ or $i_l \notin ((1/3 - \varepsilon)n, (1/3 + \varepsilon)n)$, then $\binom{n}{i_j}\binom{i_j}{i_l} = o(\binom{n}{2n/3}\binom{2n/3}{n/3}) = o(La(n, P_3, P_2))$. So, if we consider the graph Gwith vertex set $\{i_1, i_2, \ldots, i_{|P|-1}\}$ where $i_s < i_t$ are joined by an edge if and only if $\binom{n}{i_t}\binom{i_t}{i_s} = \Theta(La(n, P_3, P_2))$, then G is triangle-free. Therefore the number of edges in G is at most $\lfloor \frac{(|P|-1)^2}{4} \rfloor$. This finishes the proof of part (i).

The lower bound of (ii) is given by the family $\mathcal{F}_{\wedge,\vee}$ constructed by Katona and Tarján [71]:

$$\mathcal{F}_{\wedge,\vee} = \binom{[n-1]}{\lfloor \frac{n-1}{2} \rfloor} \cup \left\{ F \cup \{n\} : F \in \binom{[n-1]}{\lfloor \frac{n-1}{2} \rfloor} \right\}$$

Indeed, all the connected components of the comparability graph of $\mathcal{F}_{\wedge,\vee}$ have size two, so $\mathcal{F}_{\wedge,\vee}$ is $\{\wedge_2,\vee_2\}$ -free and $c(P_2,\mathcal{F}_{\wedge,\vee}) = \binom{n-1}{\lfloor \frac{n-1}{2} \rfloor} = \Omega\left(\binom{n}{\lfloor n/2 \rfloor}\right)$. To prove the upper bound of (ii) observe that for any poset P of height 2, if a family

To prove the upper bound of (ii) observe that for any poset P of height 2, if a family $\mathcal{F} \subseteq 2^{[n]}$ is P-free, then in particular it is $K_{|P|-1,|P|-1}$ -free, so we obtain

$$La(n, P, P_2) \le La(n, K_{|P|-1, |P|-1}, P_2).$$

Therefore to finish the proof it is enough to show $La(n, K_{s,s}, P_2) \leq O_s(n2^n)$ for any given integer s. Let $\mathcal{G} \subseteq 2^{[n]}$ be a $K_{s,s}$ -free family and for any pair $G, G' \in \mathcal{G}$ with $G \subset G'$ let us define M = M(G, G') to be a set with $G \subseteq M \subseteq G'$ which is maximal with respect to the property that there exist at least s sets $G_1, G_2, \ldots, G_s \in \mathcal{G}$ with $M \subsetneq G_i \ i = 1, 2, \ldots, s$.

First let us consider those pairs $G \subset G'$ for which M cannot be defined. This means that G is contained in at most s - 1 other sets of \mathcal{G} and thus the number of such pairs is at most $(s - 1)|\mathcal{G}| = O_s(2^n)$.

Next, for a fixed $M \in 2^{[n]}$, let us consider the pairs $G \subset G'$ with M = M(G, G') = G'. Note that since M is contained in s sets of \mathcal{G} (and \mathcal{G} is $K_{s,s}$ -free), M can contain at most s-1 sets from \mathcal{G} . In particular, the number of pairs $G \subset G'$ with M(G, G') = G' is at most $(s-1)|\mathcal{G}| = O_s(2^n)$.

Finally, let us consider the pairs $G \subset G'$ with $M = M(G, G') \subsetneq G'$. Each such G' contains a set $M' = M \cup \{x\}$ with $x \notin M$. The number of such M' is $|G'| - |M| \le n$, and for a given M', the number of G' containing M' is at most s (namely, M' and s - 1 other sets from \mathcal{G}) as otherwise M' would be fit to play the role of M(G, G'). So the number of G' containing M is at most sn. (Moreover, M can contain at most s - 1 sets from \mathcal{G} , so there are at most (s - 1) choices for G.) Therefore the number of pairs $G \subset G'$ with M(G, G') = M is at most $(s - 1) \cdot sn$. Summing over all sets M and adding the other types of pairs in containment we obtain

$$c(P_2, \mathcal{G}) \le O_s(2^n) + s(s-1)n2^n = O_s(n2^n).$$

which finishes the proof of the upper bound of (ii).

Let us repeat that with Dániel Nagy in a forthcoming paper, we improved the upper bound $O(n2^n)$ of Theorem 36 to $O(n\binom{n}{\lfloor n/2 \rfloor})$ which is best possible as we also showed $La(n, B, P_2) = \lceil n/2 \rceil \binom{n}{\lceil n/2 \rceil}$. Before turning to the proof of the general result Theorem 16, let us determine the order of magnitude of $La(n, T, P_2)$ for any tree poset T of height 2.

Proposition 37 (Gerbner, Methuku, Nagy, Patkós, Vizer [50]). For any tree poset T of height 2 with $|T| \ge 3$, we have

$$La(n, T, P_2) = \Theta\left(\binom{n}{\lfloor n/2 \rfloor}\right).$$

Proof. The lower bound follows from the lower bound of Theorem 36 (ii).

Now we prove the upper bound. Note that a *T*-free family \mathcal{F} does not contain a chain of length |T|. Therefore, we can partition \mathcal{F} into antichains $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_m$, such that $m \leq |T| - 1$ where \mathcal{A}_i is the family of minimal elements in $\mathcal{F} \setminus (\bigcup_{k=1}^{i-1} \mathcal{A}_k)$ for every *i*. Sperner's theorem implies $|\mathcal{A}_i| \leq {n \choose \lfloor n/2 \rfloor}$ for every *i*.

For i < j, let $n_{i,j}$ be the number of containments $A \subset B$ such that $A \in \mathcal{A}_i, B \in \mathcal{A}_j$. Notice that it is impossible that $A \subset B$ for $A \in \mathcal{A}_i$ and $B \in \mathcal{A}_j$ when i > j, so the number of P_2 's in \mathcal{F} is $\sum_{i < j} n_{i,j}$.

We claim that for any $1 \leq i < j \leq m$, we have $n_{i,j} \leq |T|(|\mathcal{A}_i| + |\mathcal{A}_j|)$. Indeed, suppose otherwise, and consider the comparability graph $G = G_{\mathcal{A}_i \cup \mathcal{A}_j}$. Then in G, there are at least |T||V(G)| edges, so the average degree in G is at least 2|T|. It is easy to find a subgraph G' of G with minimum degree at least |T|, and one can then embed T greedily into G', giving an embedding of T into \mathcal{F} , a contradiction.

Therefore, the number of P_2 's in \mathcal{F} is

$$\sum_{1 \le i < j \le m} n_{i,j} \le \sum_{1 \le i < j \le m} |T|(|\mathcal{A}_i| + |\mathcal{A}_j|) < |T|^3 \binom{n}{\lfloor n/2 \rfloor}.$$

For all posets P of height 2, we have e(P) = 1 or e(P) = 2. Clearly, if e(P) = 2, then $La(n, P, P_2) = \Theta(n\binom{n}{\lfloor n/2 \rfloor})$ as the middle two levels have that many copies of P_2 and our recent result with Balogh, Martin, and Nagy provides the upper bound. It is tempting to pose the following conjecture.

Conjecture 38.

If P is a finite poset of height 2 with e(P) = 1, then there exists a constant c_P such that $La(n, P, P_2) \leq c_P\binom{n}{\lfloor n/2 \rfloor}$ holds.

After this short detour, we finish this section with the proof of the theorem on the number of k-chains in P-free families.

Proof of Theorem 16. The proof of (i) is similar to the proof of (i) in Theorem 36. Observe first that any *P*-free family is $P_{|P|}$ -free and any P_l -free family is *P*-free. This shows the first two inequalities. To prove the last inequality, consider the canonical partition of a *P*-free family \mathcal{F} into at most |P| antichains. We can choose k of them $\binom{|P|}{k}$ ways, and in each of the resulting k-Sperner families there are at most $La(n, P_{k+1}, P_k)$ k-chains. Note that we counted every k-chain in \mathcal{F} once.

To prove the bound in (ii), let K and K' be the complete l-level and (l-1)-level posets with parts of size s = |P| - 1. Observe that if a family $\mathcal{F} \subseteq 2^{[n]}$ is P-free, then in particular it is K-free, so we obtain $La(n, P, P_k) \leq La(n, K, P_k)$. We use induction on k. The base case k = 2 is given by Theorem 36, and we note the proof is similar to the proof of Theorem 36. Let us also mention that the statement is trivial for l = 1, hence we can assume $l \geq 2$.

Let $\mathcal{G} \subseteq 2^{[n]}$ be a K-free family and consider a k-chain \mathcal{C} consisting of the sets $G_1 \subset \cdots \subset G_k$ in \mathcal{G} . Let $M = M(\mathcal{C})$ be a set with $G_1 \subset \cdots \subset G_{k-1} \subseteq M \subseteq G_k$ which is maximal with respect to the property that there exist at least s sets $H_1, H_2, \ldots, H_s \in \mathcal{G}$ with $M \subsetneq H_i$ $i = 1, 2, \ldots, s$. Let $\mathcal{M} = \{M(\mathcal{C}) \mid \mathcal{C} \text{ is a k-chain in } \mathcal{G}\}.$

We will upper bound the number of k chains $\mathcal{C} = \{G_1 \subset G_2 \subset \cdots \subset G_k\}$ (with $G_i \in \mathcal{G}$) in each of the following 3 cases separately: $M(\mathcal{C})$ is not defined at all, $M(\mathcal{C}) = G_k$ and finally $M(\mathcal{C}) \subsetneq G_k$.

First let us consider those k-chains for which M cannot be defined. This means that G_{k-1} is contained in at most s-1 other sets of \mathcal{G} and thus the number of such k-chains is at most (s-1) times the number of (k-1)-chains. If $l \leq k-1$, then by induction,

the number of (k-1)-chains is at most $O(n^{2k-2-1/2}La(n, P_l, P_{l-1}))$, otherwise l = k (recall that we assumed $l \leq k$) and then (i) shows that the number of (k-1)-chains is $O(La(n, P_l, P_{l-1}))$.

Next, consider a fixed set $M \in \mathcal{M}$. Note that \mathcal{G} is K-free and M is contained in s sets of \mathcal{G} , therefore M cannot contain K' in \mathcal{G} . In particular, the number of chains of length k-1 contained in M is at most $O(n^{2k-2-1/2}La(|M|, P_{l-1}, P_{l-2}))$ by induction. In particular, the number of k-chains $\mathcal{C} = \{G_1 \subset G_2 \subset \cdots \subset G_k\}$ for which $M(\mathcal{C}) = G_k$ is $s \cdot O(n^{2k-2-1/2}La(|M|, P_{l-1}, P_{l-2}))$.

Finally, let us now count the chains \mathcal{C} with $M = M(\mathcal{C}) \subsetneq G_k$. Given such a chain $G_1 \subset G_2 \subset \cdots \subset G_k$, we know that G_k contains a set $M' = M \cup \{x\}$ with $x \notin M$, as M is its proper subset. The number of such sets M' is $n - |M| \leq n$ and for a given M' the number of sets in \mathcal{G} containing M' is at most s (namely, M' and s - 1 other sets from \mathcal{G}), as otherwise M' would be fit to play the role of $M(\mathcal{C})$. It means, given the bottom k - 1 sets in a chain that are contained in M, there are at most sn ways to pick G_k . Thus the number of k-chains \mathcal{C} with $M = M(\mathcal{C})$ is at most sn times the number of chains of length k - 1 contained in M, which is at most

$$O(n^{2k-2-1/2})La(|M|, P_{l-1}, P_{l-2}),$$

by induction.

The total number of k-chains for which $M(\mathcal{C})$ could be defined is then

$$\sum_{M \in \mathcal{M}} O(sn \cdot n^{2k-2-1/2} La(|M|, P_{l-1}, P_{l-2})) = O(n^{2k-1-1/2}) \sum_{i=0}^{n} \sum_{M \in \binom{n}{i} \cap \mathcal{M}} La(i, P_{l-1}, P_{l-2})$$
(1)

Claim 39. For any $i \leq n$ we have

$$\sum_{M \in \binom{n}{i} \cap \mathcal{M}} La(i, P_{l-1}, P_{l-2}) \le La(n, P_l, P_{l-1}).$$

Proof. By Theorem 15, there are integers i_1, \ldots, i_{l-2} such that $La(i, P_{l-1}, P_{l-2})$ is the number of (l-2)-chains in the family consisting of all sets of sizes i_1, \ldots, i_{l-2} in $2^{[i]}$. Therefore, $\sum_{M \in \binom{n}{i}} La(i, P_{l-1}, P_{l-2})$ is equal to the number of (l-1)-chains consisting of sets of size i_1, \ldots, i_{l-2}, i . As these l-1 levels do not contain P_l , the number of (l-1)-chains is at most $La(n, P_l, P_{l-1})$ by definition.

Using the above claim and (1), we obtain that the number of k-chains for which $M(\mathcal{C})$ could be defined is $\sum_{i=0}^{n} O(n^{2k-1-1/2})La(n, P_l, P_{l-1}) = O(n^{2k-1/2})La(n, P_l, P_{l-1})$. We already obtained that the number of k-chains for which $M(\mathcal{C})$ could not be defined is $O(n^{2k-2-1/2})La(n, P_l, P_{l-1})$, which finishes the proof of the upper bound in (ii).

Now we prove the remaining part of (ii). Let $K = K_{s,s,\ldots,s}$ be the complete *l*-level poset with s > k-l+1. Let $i_1, i_2, \ldots, i_{l-1}$ be integers such that $n-(k-l+1)-i_1, i_1-i_2, \ldots, i_{l-2}-i_{l-1}, i_{l-1}$ differ by at most 1. By Theorem 15, we know that the family $\mathcal{G}' = \bigcup_{j=1}^{l-1} \binom{[n-(k-l+1)]}{i_j}$ realizes $La(n-(k-l+1), P_l, P_{l-1})$. Since \mathcal{G}' is (l-1)-Sperner,

if we add sets to \mathcal{G}' such that no two $G, G' \in \binom{[n-(k-l+1)]}{i_{l-1}}$ are contained in the same newly added sets, then the resulting family will be *K*-free. If we add the sets

$$\left\{ [n - (k - l + 1) + 1, n - (k - l + 1) + j] \cup G : j \in [k - l + 1], G \in \binom{[n - (k - l + 1)]}{i_{l-1}} \right\}$$

and denote the resulting family by \mathcal{G} , then we have

$$c(\mathcal{G}, P_k) = c(\mathcal{G}', P_{l-1}) = La(n - (k - l + 1), P_l, P_{l-1}) = \Omega_{k,l}(La(n, P_l, P_{l-1})).$$

This finishes the proof.

3 Results concerning traces of sets

Usually, set family properties capture how sets within a family \mathcal{F} relate to each other: what is the size of their pairwise or *r*-wise intersections, what kind of configurations involving containment relations can be found in \mathcal{F} , etc. Sometimes, one is interested in how the sets of the family behave when we restrict them to a subset of the underlying set. These restrictions are called traces. Formally, the *trace* of a set F on another set X is $F \cap X$, and is denoted by $F|_X$. The trace of a family \mathcal{F} of sets on X is $\mathcal{F}|_X =$ $\{F|_X : F \in \mathcal{F}\}$. No matter how many sets F have the same trace, it is counted only once in $\mathcal{F}|_X$. Let us remark that the meta-statement by Erdős and Kleitman [35] that properties of interest for set families are defined via six concepts (intersection, union, disjointness, complement, containment, size) includes problems concerning traces.

We say that a family shatters a set X (the terminology traces is also used very often), if all subsets of X appear as trace, i.e. $\mathcal{F}|_X = 2^X$. The collection of sets that are shattered by \mathcal{F} is denoted by $sh(\mathcal{F})$. The Vapnik-Chervonenkis dimension (or VC-dimension for short) of \mathcal{F} is the size of the largest set in $sh(\mathcal{F})$, and is denoted by $\dim_{VC}(\mathcal{F})$.

The fundamental result concerning traces of families was proven in the early 1970s independently by Sauer [103], Shelah [104], and Vapnik and Chervonenkis [110]. It is very often referred to as the Sauer Lemma.

Theorem 40 (Sauer [103], Shelah [104], Vapnik, Chervonenkis [110]). If $|\mathcal{F}| > \sum_{i=0}^{k-1} {n \choose i}$, then \mathcal{F} traces a subset X of [n] with |X| = k (and this is sharp as ${\binom{[n]}{\leq k-1}}$ shows).

Theorem 40 has several proofs. The one using the so-called down-shifting technique (used independently by Alon [3] and Frankl [42]) gives a little bit more. By definition, sh(F) is downward closed, i.e. $G \subset F \in \mathcal{F}$ implies $G \in \mathcal{F}$. Also, for any downward closed family \mathcal{F} , one has $sh(\mathcal{F}) = \mathcal{F}$. As observed by Pajor [94], the down-shifting argument yields the following theorem, a somewhat stronger form of Theorem 40.

Theorem 41 (Pajor [94]). For any set system $\mathcal{F} \subseteq 2^{[n]}$ we have

$$|sh(\mathcal{F})| \ge |\mathcal{F}|.$$

Let us shortly summarize results on extremal families in Theorem 40 and Theorem 41. The former direction seems much harder and attracted less researchers. Frankl [42] and Dudley [28] characterized families $\mathcal{F} \subseteq 2^{[n]}$ of size n+1 with $\dim_{VC}(\mathcal{F}) = 1$. There are quite many known families of size $\sum_{i=1}^{d} \binom{n}{i}$ with VC-dimension d (see e.g. [47]), but more results are available concerning families satisfying $|sh(\mathcal{F})| = |\mathcal{F}|$. A statement very similar to Theorem 41 was proved by Bollobás, Leader and Radcliffe [14]. We say that a family \mathcal{F} strongly shatters the set X, if there exists a set S disjoint from Xsuch that $S + 2^X := \{Y \cup S : Y \in 2^X\} \subseteq \mathcal{F}$ holds. The set S is a support of X, and the family of supports is denoted by $\mathbf{S}(X)$. We write $ssh(\mathcal{F}) = \{X : \mathbf{S}(X) \neq \emptyset\}$. By definition, we have $ssh(\mathcal{F}) \subseteq sh(\mathcal{F})$. Observe that for any downward closed \mathcal{D} we have $ssh(\mathcal{D}) = \mathcal{D}$.

Theorem 42 (Bollobás, Leader, Radcliffe [14]). For any set system $\mathcal{F} \subseteq 2^{[n]}$ we have

 $|ssh(\mathcal{F})| \le |\mathcal{F}|.$

There exist several papers (e.g. [86, 87]) dealing with describing the families for which the inequalities of Theorem 41 and Theorem 42 are satisfied with equality. It is the following result of Bollobás and Radcliffe that tells us that equality holds in both cases for the same class of families.

Theorem 43 (Bollobás, Radcliffe [15]). For any set system $\mathcal{F} \subseteq 2^{[n]}$ the following two properties are equivalent

 $(a) |\mathcal{F}| = |sh(\mathcal{F})|,$ (b) $|\mathcal{F}| = |ssh(\mathcal{F})|$.

Apart from describing extremal families, Theorem 40 leads in several directions. We now focus on the direction that seems somewhat similar to forbidden subposet problems, the topic of the previous section. We will require that traces of families of interest should avoid copies of a poset P on all sets of a fixed size. Theorem 40 and 41 can be interpreted in this context as well: the forbidden poset should be B_k , the Boolean poset of all sets in $2^{[k]}$ ordered by inclusion. In this dissertation, we will only consider chains as forbidden structures, but with Dániel Gerbner and Máté Vizer [54]. we introduced this topic in full generality and obtained further results.

Definition: A family $\mathcal{F} \subset 2^X$ of sets is said to be *l*-trace k-Sperner if for any subset Y of X of size l, the trace of \mathcal{F} on Y does not contain any chain of length k+1.

The *l*-trace *k*-Sperner property can be formalized through forbidden traces, too. One has to exclude $\binom{l+1}{k+1}$ families as trace (all possibilities how we can choose k+1levels out of the l+1 that $2^{[l]}$ possesses). We will be interested in the maximum size f(n,k,l) that an *l*-trace k-Sperner family $\mathcal{F} \subseteq 2^{[n]}$ can have. Part (a) of the following result is a trivial consequence of Theorem 40, the main statement, part (b), asserts that the more restrictive condition of forbidding a maximal chain instead of the complete Boolean lattice, ensures that there will be only two extremal families.

Theorem 44 (Patkós [95]). (a) $f(n,k,k) = \sum_{i=0}^{k-1} {n \choose i}$. (b) If $\mathcal{F} \subseteq 2^{[n]}$ is k-trace k-Sperner with $|\mathcal{F}| = \sum_{i=0}^{k-1} {n \choose i}$, then either $\mathcal{F} = {[n] \choose \leq k-1}$ or $\mathcal{F} = {[n] \choose \geq n-k+1}.$

The statement of Theorem 44 remains valid (for for families with large enough underlying sets) even if we enlarge the size of those subsets on which we consider the trace of the set family.

Theorem 45 (Patkós [95]). For every pair of integers k and l $(1 \le k \le l)$ there exists N(k,l) such that if $n \ge N(k,l)$, then $f(n,k,l) = \sum_{i=0}^{k-1} \binom{n}{i}$. Furthermore, if $2 \le k \le l$, then the only optimal l-trace k-Sperner families are $\binom{[n]}{\le k-1}$ and $\binom{[n]}{\ge n-k+1}$.

On the other hand, if l, the size of those sets on which we consider the trace of set families, is close to the size of the ground set, then the problem becomes more similar to the original k-Sperner problem and thus our results will be similar to Theorem 2 of Erdős. Indeed, suppose, when considering traces, we only want to omit a fixed number l' of elements of the underlying set, and assume that l' < k, where k + 1 is the length of the chain that we want to avoid in traces on subsets of size n - l'. Then consider a family $\mathcal{F} = \binom{n}{m} \cup \binom{n}{m+1} \cup \cdots \cup \binom{n}{m+k-l'-1}$ of k - l' consecutive levels. Clearly, for any $X \subset [n]$ with |X| = n - l' and $F \in \mathcal{F}$, we have $m - l' \leq |F|_X| \leq m + k - l' - 1$. So there are only k possible set sizes in $\mathcal{F}|_X$ and thus \mathcal{F} is (n - l')-trace k-Sperner. To obtain the largest such family, the levels should be taken "from the middle", and our next theorem states these families are asymptotically optimal.

Theorem 46 (Patkós [98]). Let k and l' be positive integers with l' < k. Then if $\mathcal{F} \subseteq 2^{[n]}$ is an (n-l')-trace k-Sperner family, then $|\mathcal{F}| \leq (k-l'+o(1))\binom{n}{\lfloor n/2 \rfloor}$. In particular, $f(n, n-l', k) = (k-l'+o(1))\binom{n}{\lfloor n/2 \rfloor}$.

Let us remark that in [98], we determined the exact maximum possible size of an (n-1)-trace k-Sperner family for all $k \ge 2$ by showing $f(n, n-1, k) = \Sigma(n, k-1)$ if n is large enough.

The maximum size of uniform families of bounded VC-dimension was first studied by Frankl and Pach [45]. Clearly, the VC-dimension of a k-uniform family can be at most k. Frankl and Pach proved that if a family $\mathcal{F} \subseteq {\binom{[n]}{k}}$ has VC-dimension strictly smaller than k, then we must have $|\mathcal{F}| \leq {\binom{n}{k-1}}$. They also observed that any k-uniform intersecting family \mathcal{G} has VC-dimension at most k-1 as if a k-set G belongs to \mathcal{G} , then $\emptyset \notin \mathcal{G}|_G$, otherwise $G \notin \mathcal{G}|_G$. By the Erdős-Ko-Rado theorem, the star (the family of all k-sets containing a fixed element of the ground set) is the unique largest intersecting family if 2k < n, Frankl and Pach conjectured that the star is the largest kuniform family with VC-dimension at most k-1. This was disproved by Ahlswede and Khatchatrian [2], who, for $k \geq 3$, constructed a k-uniform family of size $\binom{n-1}{k-1} + \binom{n-4}{k-3}$ of VC-dimension k-1. A decade later, Mubayi and Zhao obtained [90] exponentially many pairwise non-isomorphic families each achieving the size of the family of Ahlswede and Khatchatrian. If the size of these constructions turns out to be extremal, then this could explain why determining the exact extremal value seems hard for this problem. Mubayi and Zhao also proved the first improvement on the $\binom{n}{k-1}$ upper bound of Frankl and Pach.

Just as in the non-uniform case, one may strengthen the condition that \mathcal{F} does not shatter any k-subsets to $\mathcal{F}|_K$ does not contain maximal chains for every k-set K. Observe that k-uniform intersecting families \mathcal{F} will satisfy this stronger condition for the same reason as before: for any k-set K, either the empty set or K itself will be missing from $\mathcal{F}|_K$, and these sets belong to all maximal chains in K. The next theorem shows not only that the conjecture of Frankl and Pach becomes true with this strengthened condition, but any family satisfying the trace condition with at least half the extremal size, must be a subfamily of a star.

Theorem 47 (Patkós [96]). For every integer $2 \le k$ and real 1/2 < c < 1 there exists an $N_0(k,c)$ such that for any $n \ge N_0(k)$ if $\mathcal{F} \subseteq {\binom{[n]}{k}}$ has size larger than $c{\binom{n-1}{k-1}}$ and

there is no subset X of [n] with |X| = k such that $\mathcal{F}|_X$ contains a maximal chain, then there exists $x \in [n]$ such that $x \in F$ for all $F \in \mathcal{F}$.

Most extremal problems ask for the size of the largest combinatorial structure satisfying a prescribed property. Some of these have a natural saturation counterpart problem that asks for the smallest possible size. Whenever the extremal problem is defined via a forbidden substructure, for example if the property in interest is not containing a certain subgraph as in Turán proglems, a containment pattern as in forbidden subposet problems or certain traces, then the empty graph or the empty set system will clearly satisfy the required property. Also, whenever a graph or set family has the property, so does any of its subgraphs or subfamilies. Therefore, the meaningful saturation question is to determine the minimum possible size of a maximal (unextendable) structure. So for traces the saturation problem corresponding to Theorem 40 is to determine $sat_{VC}(n,d)$ the minimum size of a family $\mathcal{F} \subseteq 2^{[n]}$ such that the VC-dimension of \mathcal{F} is d, but for any $G \in 2^{[n]} \setminus \mathcal{F}$ the family $\mathcal{F} \cup \{G\}$ has VC-dimension d+1. Clearly, any family consisting of a single set has VC-dimension 1 and any family containing at least two sets has VC-dimension at least 2, and thus we have $sat_{VC}(n,0) = 1$ for any positive integer n. Dudley [28] showed that any maximal family $\mathcal{F} \subseteq 2^{[n]}$ with VC-dimension 1 has size n + 1 and thus the number $sat_{VC}(n, 1) = n + 1$ matches the one we obtain from Theorem 40. After some thoughts, it is not hard to come up with maximal families $\mathcal{F} \subset 2^{[n]}$ of VC-dimension $d \geq 2$ of size strictly smaller than $\sum_{j=0}^{d-1} \binom{n}{j}$, but somewhat surprisingly, the situation is "completely" the opposite for $d \geq 2$. As shown by the next theorem, there exist maximal families $\mathcal{F} \subseteq 2^{[n]}$ of VC-dimension d such that their size is bounded independently of n.

Theorem 48 (Frankl, Kiselev, Kupavskii, Patkós [39]). For any $d \ge 3$, $sat_{VC}(n, d-1) \le 4^d$ holds for any $n \ge 2d$. Moreover, if d is odd or if $d \ge 14$, then we can replace 4^d with $\frac{1}{2}\binom{2d}{d}$.

Let us finish the introduction of this section by reviewing the impact of the results and how they fit into recent research concerning traces of set systems. The area of forbidden configuration problems, initiated by Richard Anstee (see the survey [5]) and with numerous contributions from Attila Sali and some elegant and important results from Zoltán Füredi, deals with uniform set families that avoid certain trace patterns. Results are formulated in the language of 0-1 matrices, but all statements can be converted to results about traces by considering the matrix with the characteristic vectors of sets. Theorem 47 fits into this setting and is cited several times by papers of the area. The construction of Theorem 47 can be generalized to larger uniformity: for any $k \leq m$, the family $\mathcal{F} \subseteq {m \choose m}$ consisting all *m*-sets containing a fixed (m - k + 1)-set M, has the property that for any k-set X, we have $\mathcal{C}_k \not\subseteq \mathcal{F}_X$, i.e. \mathcal{F} is k-trace k-Sperner. We conjectured if n is large enough, then this family is extremal. This cojecture was veerified by Tan [108].

When l is close to the size of the underlying set, the problem of finding the largest l-trace k-Sperner family becomes more similar to problems on antichains. The paper [98] containing Theorem 46 is mostly not cited for its main result Theorem 46, but rather for one of its lemmas (Lemma 57 in this dissertation) about uniform set families avoiding

so-called tight paths. There are several notions of paths and cycles for hypergraphs, tight paths are one of them. Lemma 57 and Corollary 58 give the current best bound on the size of k-uniform families not containing tight paths of length l in the case when l < 0.8k and the ground set is large enough.

3.1 Families avoiding chains as traces

We start by proving theorems on traces of non-uniform families on sets of fixed size.

Proof of Theorem 44. The statement about f(n, k, k) is straightforward from Theorem 40.

To prove (b), let us consider a k-trace k-Sperner set system $\mathcal{F} \subseteq 2^{[n]}$ with $|\mathcal{F}| = \sum_{i=0}^{k-1} {n \choose i}$. By Theorem 41 we have $|sh(\mathcal{F})| \ge \sum_{i=0}^{k-1} {n \choose i}$. But if \mathcal{F} shatters a k-subset of [n], then it is not k-trace k-Sperner, so $sh(\mathcal{F}) = {[n] \choose {\le k-1}}$ and in particular, $|\mathcal{F}| = |sh(\mathcal{F})|$. From Theorem 42, it follows that $|\mathcal{F}| = |ssh(\mathcal{F})|$, and thus by $ssh(\mathcal{F}) \subseteq tr(\mathcal{F})$, we have that $ssh(\mathcal{F}) = sh(\mathcal{F}) = {[n] \choose {\le k-1}}$.

Now let us consider a set $F \in \mathcal{F}$ with minimum size. If |F| > n - k + 1, then $|\mathcal{F}| < f(n, k, k)$ - a contradiction. Therefore $|F| \leq n - k + 1$, so there exists $X \subseteq [n] \setminus F$ with |X| = k - 1. By the paragraph above, we have $X \in ssh(\mathcal{F})$. Let us take an arbitrary $S(X) \in \mathbf{S}(X)$. We claim that there is no element $s \in S(X) \setminus F$. Indeed, if there is, then let us consider $\mathcal{F}|_{X \cup \{s\}}$. Since $F \in \mathcal{F}$ and $s \notin F$, we have $\emptyset \in \mathcal{F}|_{X \cup \{s\}}$. Since $X \in ssh(\mathcal{F})$ and $s \in S(X)$, there is a chain in $\mathcal{F}|_{X \cup \{s\}}$ of length k with set sizes 1, 2, ..., k, which together with the empty set form a chain of length k + 1 - a contradiction. Thus $S(X) \subseteq F$, but since by the definition of support, $S(X) \cup \emptyset = S(X) \in \mathcal{F}$ and F is of minimum size, we must have S(X) = F and thus $F + 2^X \subseteq \mathcal{F}$. As X was chosen arbitrarily, we obtain that for any Y with $Y \cap F = \emptyset$ and |Y| = k - 1, we have $F + 2^Y \subseteq \mathcal{F}$.

We claim that for any such Y, the set $F \cup Y$ is maximal in \mathcal{F} . Indeed, if not, then $F \cup Y \cup A \in \mathcal{F}$ for some non empty A. Therefore, for some $a \in A$, the trace $\mathcal{F}|_{Y \cup \{a\}}$ contains a chain of length k + 1 (the trace of $F \cup Y \cup A$ is $Y \cup \{a\}$ and from the trace of $F + 2^Y$ we can pick the other k sets) - a contradiction.

We claim that for any $Y' \subseteq F \cup Y$ with |Y'| = k - 1 we have $F \cup Y \setminus Y' + 2^{Y'} \subseteq \mathcal{F}$ (and $F \cup Y \setminus Y'$ is minimal in \mathcal{F}). To see this observe that $S(Y') = F \cup Y \setminus Y'$. Indeed, if there was an element $s \in (F \cup Y \setminus Y' \setminus S(Y'))$, then we would have a chain of length k+1 in $\mathcal{F}|_{Y'\cup s}$. Thus $S(Y') \supseteq F \cup Y \setminus Y'$ and $S(Y') \supseteq F \cup Y \setminus Y')$ would contradict the maximality of $F \cup Y \cup Y'$ as $S(Y') \cup Y' \in \mathcal{F}$ by definition. The minimality of $F \cup Y \setminus Y'$ follows just as the maximality of $F \cup Y$.

We obtained that for any Y, Y' with |Y| = |Y'| = k - 1 we have $F \cup Y \setminus Y' + 2^{Y'} \subseteq \mathcal{F}$ and $F \cup Y \setminus Y'$ is minimal in \mathcal{F} , so we could have started with $F \cup Y \setminus Y'$ in place of F. Thus we get, that for any $Y_1, Y'_1, Y_2, Y'_2, ..., Y_m, Y'_m$ the set $((((F \cup Y_1 \setminus Y'_1) \cup Y_2 \setminus Y'_2)...) \cup Y_m \setminus Y'_m)$ is minimal in \mathcal{F} and $((((F \cup Y_1 \setminus Y'_1) \cup Y_2 \setminus Y'_2)...) \cup Y_m \setminus Y'_m) + 2^{Y'_m} \subseteq \mathcal{F}$. That is for any $G \subseteq [n]$ with $|F| \leq |G| \leq |F| + k - 1$, we have $G \in \mathcal{F}$. But because of (a), it is possible if and only if $\mathcal{F} = {[n] \choose \leq k-1}$ or $\mathcal{F} = {[n] \choose \geq n-k+1}$.

Proof of Theorem 45. If k = 1, then N(1, l) = 2l - 1 is a good choice. Indeed, let $n \ge 2l - 1$ and assume that $|\mathcal{F}| \ge 2$.

Case 1: \mathcal{F} contains two members A, B with $A \subset B$. Then picking any *l*-subset L of [n] which contains an element from $B \setminus A$ and considering the *L*-trace would yield a contradiction.

Case 2: \mathcal{F} is a Sperner family. Then (since $n \geq 2l - 1$) for any $A, B \in \mathcal{F}$ we have

that either $\overline{B} \cup (A \cap B)$ or $\overline{A} \cup (B \cap A)$ is of size at least l, so we can find an l-subset, where the traces of the sets are in inclusion. (In fact, 2l - 1 is sharp as shown by any pair $A, \overline{A} \subset [2l-2], |A| = l - 1$.)

Since in the case k = 1 there is no uniqueness for the extremal family, we still have to establish the base case k = 2, but as this case and the inductive step is very similar, we describe them simultaneously. Suppose that for some fixed k and l, we have already proved the statement of the theorem for every k', l' with $k' \leq k, l' \leq l$ and with at least one of k' and l' strictly smaller than k or l. Let M denote the maximum of N(k', l'), where k', l' are as above and put $N = M + k + \sum_{i=0}^{l-1} {M+k \choose i}$. We will prove that the statement about f(n, k, l) is true if $n \geq N$, and the statement about the optimal families holds provided $n \geq N + 1$.

Before we proceed to the actual proof, we need to introduce some notation. For any family $\mathcal{F} \subseteq 2^X$ and $x \in X$, we put $\mathcal{F}_x^0 := \{F \in \mathcal{F} : x \notin F, F \cup \{x\} \in \mathcal{F}\}, \mathcal{F}_x^1 := \{F \in \mathcal{F} : x \in F, F \setminus \{x\} \in \mathcal{F}\}$ and $\mathcal{F}_{\overline{x}} := \mathcal{F} \setminus (\mathcal{F}_x^0 \cup \mathcal{F}_x^1)$. Trivially $|\mathcal{F}_x^0| = |\mathcal{F}_x^1|,$ $|\mathcal{F}| = |\mathcal{F}_x^0| + |\mathcal{F}_x^1| + |\mathcal{F}_{\overline{x}}|$ and $|\mathcal{F}|_{X \setminus \{x\}}| = |\mathcal{F}_x^0| + |\mathcal{F}_{\overline{x}}|$.

Lemma 49. If \mathcal{F} is *l*-trace *k*-Sperner on the underlying set X, then for any $x \in X$, \mathcal{F}_x^0 is (l-1)-trace (k-1)-Sperner on $X \setminus \{x\}$.

Proof of Lemma: Suppose not. Then there exist an l-1-set $L' \subseteq X \setminus \{x\}$ and $F_1, F_2, ..., F_k \in \mathcal{F}^0_x$ such that $F_1|_{L'} \subset F_2|_{L'} \subset ... \subset F_k|_{L'}$. But then, putting $F_{k+1} = F_k \cup \{x\} \in \mathcal{F}$ and $L = L' \cup \{x\}$, we would have $F_1|_L \subset F_2|_L \subset ... \subset F_k|_L \subset F_{k+1}|_L$ - a contradiction.

Suppose there exists an *l*-trace *k*-Sperner family $\mathcal{F} \subseteq 2^{[n]}$ $(n \geq N)$ with $|\mathcal{F}| = \sum_{i=0}^{k-1} {n \choose i} + C$ (where *C* is positive). We claim that there is a subset $X \subseteq [n]$ with $|X| \geq M + k$ such that for any element $x \in X$ we have $|\mathcal{F}|_X| \geq \sum_{i=0}^{k-1} {|X| \choose i} + C$ and $|(\mathcal{F}|_X)_x^0| = \sum_{i=0}^{k-2} {|X|-1 \choose i}.$

 $\begin{aligned} |\langle \mathcal{F}|_X \rangle_X | &= \mathcal{L}_{i=0} (i) \\ \text{We know that for any } x \in X \subseteq [n] \text{ with } |X| \geq M+1 \text{ we have } |(\mathcal{F}|_X)_x^0| \leq \\ \sum_{i=0}^{k-2} \binom{|X|-1}{i}, \text{ because of Lemma 49 and the inductive hypothesis on } f(n, k-1, l-1). \end{aligned}$ Therefore if $|(\mathcal{F}|_X)_x^0| \neq \sum_{i=0}^{k-2} \binom{|X|-1}{i}$ then we must have $|(\mathcal{F}|_X)_x^0| < \sum_{i=0}^{k-2} \binom{|X|-1}{i}.$ Thus if $|\mathcal{F}|_X| \geq \sum_{i=0}^{k-1} \binom{|X|}{i}$, then we have $|\mathcal{F}|_{X\setminus\{x\}}| \geq \sum_{i=0}^{k-1} \binom{|X|-1}{i} + 1$. We obtain that if X = [n] is not a good choice for our claim, then there is an $x_1 \in [n]$ which shows this fact and $|\mathcal{F}|_{[n]\setminus\{x_1\}}| \geq \sum_{i=0}^{k-1} \binom{n-1}{i} + C+1$. If $X = [n]\setminus\{x_1\}$ is not good either, then some $x_2 \in [n] \setminus \{x_1\}$ shows this and we have that $|\mathcal{F}|_{[n]\setminus\{x_1,x_2\}}| \geq \sum_{i=0}^{k-1} \binom{n-2}{i} + C+2$. Continuing in this way we get that if there is no good set, then there is a subset $Y \subset [n]$ with |Y| = M+k such that we have $|\mathcal{F}|_Y| > n - (M+k) \geq N - (M+k) \geq \sum_{i=0}^{l-1} \binom{M+k}{i}$. But then, by Theorem 41, $\mathcal{F}|_Y$ (and so \mathcal{F} as well) shatters a set of size l contradicting the l-trace k-Sperner property.

So we established that for some $X \subseteq [n]$ with $|X| \ge M + k$ and any of its elements $x \in X$ we have $|\mathcal{F}|_X| \ge \sum_{i=0}^{k-1} {|X| \choose i} + C$ and $|(\mathcal{F}|_X)_x^0| = \sum_{i=0}^{k-2} {|X|-1 \choose i}$. If $(\mathcal{F}|_X)_x^0 = {X \setminus \{x\} \choose \leq k-2}$ or $(\mathcal{F}|_X)_x^0 = {X \setminus \{x\} \choose \geq |X|-k+1}$, then $\mathcal{F}|_X$ contains ${X \choose \leq k-1}$ or ${X \choose \geq |X|-k+1}$ and at least one additional set which contradicts the *l*-trace *k*-Sperner property. Why is it true that $(\mathcal{F}|_X)_x^0 = {X \setminus \{x\} \choose \leq k-2}$ or $(\mathcal{F}|_X)_x^0 = {X \setminus \{x\} \choose \leq k-2}$ or $(\mathcal{F}|_X)_x^0 = {X \setminus \{x\} \choose \leq k-2}$ or $(\mathcal{F}|_X)_x^0 = {X \setminus \{x\} \choose \geq |X|-k+1}$? If $k \ge 3$, this is simply the inductive

hypothesis for the uniqueness of the extremal systems. If k = 2 we need to work a bit more.

In this case, what we have already proved is that for the above set X and for any $x \in X$ we have $(\mathcal{F}|_X)_x^0 \neq \emptyset$, i.e. the singleton $\{x\}$ is strongly shattered by $\mathcal{F}|_X$. Since $\mathcal{F}|_X$ is *l*-trace 2-Sperner, we need the following lemma.

Lemma 50. If for some l with $2l \leq n$, the family $\mathcal{F} \subseteq 2^{[n]}$ is l-trace 2-Sperner and \mathcal{F} strongly shatters all singletons, then $\mathcal{F} = {[n] \choose \leq 1}$ or $\mathcal{F} = {[n] \choose \geq n-1}$.

Proof: If there is a singleton $x \in [n]$ with $\emptyset \in \mathbf{S}(x)$, then $\emptyset, \{x\} \in \mathcal{F}$ and $x \in F \in \mathcal{F}$ implies $F = \{x\}$ because of the *l*-trace 2-Sperner property. Therefore $\mathcal{F}|_{[n]\setminus\{x\}}$ is (l-1)-trace 2-Sperner. Since $2(l-1) \leq n-1 = |[n] \setminus \{x\}|$, we obtain by induction that $\mathcal{F}|_{[n]\setminus\{x\}} = \binom{[n]\setminus\{x\}}{\leq 1}$ or $\mathcal{F}|_{[n]\setminus\{x\}} = \binom{[n]\setminus\{x\}}{\geq n-2}$, but the latter is impossible as $\emptyset \in \mathcal{F}|_{[n]\setminus\{x\}}$. Thus we have $\mathcal{F}|_{[n]\setminus\{x\}} = \binom{[n]\setminus\{x\}}{\leq 1}$ and $\mathcal{F} = \binom{[n]}{\leq 1}$. Likewise, if there is a singleton such that one of the supports with respect to \mathcal{F} is

Likewise, if there is a singleton such that one of the supports with respect to \mathcal{F} is the complement set, then $\mathcal{F} = {[n] \choose \geq n-k+1}$. So we may assume that for any singleton x we have $\emptyset, [n] \setminus \{x\} \notin \mathbf{S}(x)$. Let us pick x such that (one of) its support $S(x) \in \mathbf{S}(x)$ is of minimum size.

Claim 51. For any singleton $x' \subset [n] \setminus S(x)$ we have $\{S(x)\} = \mathbf{S}(x')$.

Proof of Claim: Let us consider an arbitrary $S(x') \in \mathbf{S}(x')$, we will show that $S(x') \subseteq S(x)$, so by the minimality of S(x) we will have S(x') = S(x). Suppose there is an element $s \notin S(x)$ belonging to S(x'). Let us put $L = \{x'\} \cup \{s\} \cup L'$, where $s \notin L' \subseteq ([n] \setminus S(x)) \cup S(x')$, with |L'| = l - 2 (the existence of such a set follows from the assumption $2l \leq n$ and the minimality of S(x)). But then $\mathcal{F}|_L$ would contain a chain of length 3 as shown by $(S(x') \cup \{x'\})|_L$, $S(x')|_L$ and $S(x)|_L$.

Claim 52. For every $y \in S(x)$ and $S(y) \in \mathbf{S}(y)$, we have |S(x)| = |S(y)| and $|S(x) \cup S(y)| = |S(x)| + 1$.

Proof of Claim: If S(y) contained two elements $x_1, x_2 \notin S(x)$, then putting $L = \{x_1, x_2\} \cup L'$, where $x_1, x_2 \notin L' \subseteq ([n] \setminus S(x_1)) \cup S(y)$ with |L'| = l - 2 (the existence of such L' follows from the assumption $2l \leq n$ and the minimality of $S(x) = S(x_1)$, which holds by the previous claim), $\mathcal{F}|_L$ would contain the 3-chain: $S(x_1)|_L \subset S(x) \cup \{x_1\}|_L \subset S(y)|_L$.

Because of Claim 52, Claim 51 could be applied to y and an arbitrary $x' \notin S(x) \cup S(y)$ (there is such x' as $|S(x) \cup S(y)| = |S(x)| + 1$), giving S(x) = S(x') = S(y) - a contradiction as $y \in S(x), y \notin S(y)$.

We obtained that the support of any singleton is either the empty set or the complement of the singleton, so the proof of the lemma is complete by the paragraph preceeding the claims. $\hfill\square$

We still have to show, that if $n \ge N+1$, then the only optimal families are $\binom{[n]}{\le k-1}$ and $\binom{[n]}{>n-k+1}$. Let $\mathcal{F} \subseteq 2^{[n]}$ be an *l*-trace *k*-Sperner family with $n \ge N+1$. If for any

 $\begin{aligned} x \in [n], &\text{we had } |\mathcal{F}_x^0| < \sum_{i=0}^{k-2} \binom{n-1}{i}, \text{ then } |\mathcal{F}|_{[n] \setminus \{x\}}| > \sum_{i=0}^{k-1} \binom{n-1}{i} \text{ would hold, but this cannot happen, since } n-1 \geq N \text{ and we have already proved that for any } n' \geq N \text{ we have } f(n',k,l) = \sum_{i=0}^{k-1} \binom{n'}{i}. \text{ So if } k=2 \text{ we can apply Lemma 50 to obtain that } \mathcal{F} \text{ is either } \binom{[n]}{\leq k-1} \text{ or } \binom{[n]}{\geq n-k+1}, \text{ while if } k>2 \text{ by the induction hypothesis for any } x \in [n] \text{ we have that } \mathcal{F}_x^0 \text{ is } \binom{[n] \setminus \{x\}}{\leq k-2} \text{ or } \binom{[n] \setminus \{x\}}{\geq n-k+2}. \end{aligned}$

Next, we consider traces of uniform families on subsets of the underlying set of fixed size. We will use the celebrated Hilton-Milner theorem on non-trivial intersecting families, i.e. for which there does not exist any element that belong to all sets of the family. In this section, we will only use the upper bound the theorem gives, and will only elaborate on the extremal families in the next section, so we only state this weakened version here.

Theorem 53. (Hilton, Milner [65]) Let $\mathcal{F} \subset {X \choose k}$ be an intersecting family with $k \ge 3$, $n \ge 2k+1$ such that no element belongs to sets of \mathcal{F} . Then $|\mathcal{F}| \le {n-1 \choose k-1} - {n-k-1 \choose k-1} + 1$.

We start showing Theorem 47 by first proving a lemma stating that if we want to have an "almost" maximal chain $C_k^- = \{[1], [2], ..., [k]\}$ as trace, then much smaller families suffice.

Lemma 54. For every integer $2 \le k$ and real 1/2 < c' < 1 there exists an $N'_0(k,c)$ such that for any $n \ge N'_0(k)$ if $\mathcal{F} \subseteq {[n] \choose k}$ has size larger than $c' {n-1 \choose k-1}$ then there exists a set $X \subset [n]$ with |X| = k such that $\mathcal{C}_k^- \subseteq \mathcal{F}|_X$.

Proof. We proceed by induction on k. For k = 2, if there exists an intersecting pair of 2-sets $F_1, F_2 \in \mathcal{F}$, then $\emptyset \neq F_1|_{F_2} \subset F_2$ is a C_2^- . Therefore \mathcal{F} is a pairwise disjoint family and thus $|\mathcal{F}| \leq n/2 < c'(n-1)$ for any 1/2 < c' if n is large enough.

Now suppose the lemma is proved for k-1 and any real between 1/2 and 1. For a real c' fix an $M > N'(k-1, \frac{c'+1/2}{2})$ such that the following inequalities hold for all $n \ge M$

$$\frac{c'-1/2}{2}\binom{n-2}{k-2} > \binom{n-2}{k-2} - \binom{n-k-2}{k-2},\tag{2}$$

$$c'\left(\binom{n-2}{k-2} + \binom{n-3}{k-2}\right) > \binom{n-2}{k-2}.$$
(3)

The existence of such N' for (2) follows from the fact that if we consider the two sides of (3) as polynomials of n, then the degree of the left hand side is one more than the degree of the right hand side and for (3) from c' > 1/2 and from $\lim_{n \to \infty} {\binom{n-2}{k-2}} / {\binom{n-3}{k-2}} = 1$.

Let $N'(k,c') = M + 1 + 2\binom{M+1}{k-1}$, $n \ge N'(k,c')$ and $\mathcal{F} \subseteq \binom{[n]}{k}$ a family with $|\mathcal{F}| \ge c'\binom{n-1}{k-1}$. Let $x_1 \in [n]$ be an element with maximum degree which is at least the average degree $c'\binom{n-1}{k-1}\frac{k}{n} \ge c'\binom{n-2}{k-2}$ and consider $\mathcal{F}_{\overline{x}_1}$. By the inductive hypothesis there exists a (k-1)-subset $X \subset [n] \setminus \{x_1\}$ such that $\mathcal{F}_{\overline{x}_1}|_X$ contains \mathcal{C}_{k-1}^- . Just by removing these sets one after the other and repeatedly using the inductive hypothesis we get that $\mathcal{G} = \{X \in \mathcal{F}_{\overline{x}_1} : \mathcal{C}_{k-1}^- \subseteq \mathcal{F}_{\overline{x}_1}|_X\}$ has size at least $(c' - \frac{c'+1/2}{2})\binom{n-2}{k-2} = \frac{c'-1/2}{2}\binom{n-2}{k-2}$. If two

sets $X_1, X_2 \in \mathcal{G}$ are disjoint, then writing $F_1 = X_1 \cup \{x_1\}, F_2 = X_2 \cup \{x_1\}$ both $\mathcal{F}|_{F_1}$ and $\mathcal{F}|_{F_2}$ contain \mathcal{C}_k^- as $F_1|_{F_2} = F_1 \cap F_2 = F_2|_{F_1} = \{x\}$. Thus we may assume that \mathcal{G} is intersecting and thus by Theorem 53 and (2) there exists a $x_2 \in [n] \setminus \{x_1\}$ such that $x_2 \in X$ for all $X \in \mathcal{G}$.

If there is a set $F' \in \mathcal{F}_{x_1}$ with $x_2 \notin F$, then we claim that there is a set $X \in \mathcal{G}$ such that $\mathcal{F}' \cap X = \emptyset$. Indeed, the number of k-1-sets containing x_2 and meeting F is $\binom{n-2}{k-2} - \binom{n-k-2}{k-2}$, thus again by (1) there is a set $X \in \mathcal{G}$ as claimed. Therefore writing $F = X \cup \{x\}$ we have $F'|_F = \{x_1\}$ and thus $\mathcal{C}_k^- \subseteq \mathcal{F}|_F$.

Therefore we may assume that all $F' \in \mathcal{F}_{x_1}$ contain x_2 and thus as x_1 is of maximum degree x_1 and x_2 are contained in the same sets of \mathcal{F} . The number of sets in \mathcal{F} containing both x_1 and x_2 is at most $\binom{n-2}{k-2}$, thus removing these sets from \mathcal{F} there remains a family \mathcal{F}_1 of subsets of $[n] \setminus \{x_1, x_2\}$ of size at least $c'\binom{n-1}{k-1} - \binom{n-2}{k-2}$ which is by (3) greater or equal to $c'\binom{n-3}{k-1} + 1$.

Repeating the above argument l times, we either find a set X such that $\mathcal{C}_k^- \subseteq \mathcal{F}_{l-1}|_X \subseteq \mathcal{F}|_X$ or subfamily $\mathcal{F}_l \subseteq \binom{[n] \setminus \{x_1, x_2, z_3, x_4, \dots, x_{2l-1}, x_{2l}\}}{k} \cap \mathcal{F}_{l-1}$ with size at least $c'\binom{n-2l-1}{k-1} + l$. Thus we either find a set X such that $\mathcal{C}_k^- \subseteq \mathcal{F}_l|_X \subseteq \mathcal{F}|_X$ for some $l \leq \frac{n-M}{2}$ or as $n \geq M + 1 + 2\binom{M+1}{k-1}$ we obtain a subfamily of \mathcal{F} on M or M + 1 elements (depending on the parity of n) with size at least $\binom{M+1}{k-1}$, and thus by the result of Frankl and Pach [45], we even find a $2^{[k]}$ as trace which proves the lemma.

Proof of Theorem 47. For some k and c as in the statement of the theorem, let us fix an integer N(k,c) larger than $N'(k,\frac{c+1/2}{2})$ of Lemma 54 such that for any $n \ge N(k,c)$ the following inequality holds

$$\frac{c-1/2}{2}\binom{n-1}{k-1} > \binom{n-1}{k-1} - \binom{n-k-1}{k-1} + 1.$$
(4)

Let $\mathcal{F} \subset {\binom{[n]}{k}}$ be a family with size at least $c\binom{n-1}{k-1}$. We claim that the size of the set $\mathcal{H} = \{X \subset [n] : |X| = k, \mathcal{C}_k^- \subseteq \mathcal{F}_X\}$ is at least $\frac{c-1/2}{2} \binom{n-1}{k-1}$. Indeed, applying Lemma 54 to \mathcal{F} we obtain 1 set in \mathcal{H} , then removing this set from \mathcal{F} and applying Lemma 54 again we get another set and so on until the remaining family contains less than $\frac{c+1/2}{2} \binom{n-1}{k-1}$ sets. If there is a pair of disjoint sets $X_1, X_2 \in \mathcal{H}$, then the trace $X_1 \cap X_2 = \emptyset$ extends this to a \mathcal{C}_k , thus we may assume that those sets form an intersecting family, therefore by Theorem 53 and (4) there must exist an element $x \in [n]$ such that $x \in X$ for all $X \in \mathcal{H}$. Any set $F \in \mathcal{F} \setminus \mathcal{H}$ must meet all sets in \mathcal{H} as otherwise $F|_X = F \cap X = \emptyset$ would complete $\mathcal{C}_k^- \subseteq \mathcal{F}|_X$ to \mathcal{C}_k . But this can happen only if F contains x as otherwise the number of k-sets containing x and meeting F would be $\binom{n-1}{k-1} - \binom{n-k-1}{k-1}$ which is by (4) smaller than $\frac{c-1/2}{2} \binom{n-1}{k-1} \leq |\mathcal{H}|$. Thus all sets in \mathcal{F} contain x which finishes the proof of the theorem.

Finally, we consider traces of non-uniform families on subsets that contain almost all elements of the ground set. The proof of Theorem 46 will be given in two parts each of which will have a somewhat surprising relation to different combinatorial problems.

To obtain the first step, we will apply Bukh's result, Theorem 4 on the size of families that avoid a forbidden tree poset. For the second step, we will need and prove a lemma on families avoiding *tight paths*, a problem very much considered in the literature.

To be able to use Theorem 4, we need to define the following directed graph: $T_{h,c}$ is a tree with height h such that all arcs are directed towards the root and each vertex, with the exception of the leaves, has exactly c children. Let $P_{h,c}$ denote the poset with $T_{h,c}$ as its Hasse diagram. As the height of $P_{h,c}$ is h, the following two statements and Bukh's result Theorem 4 will immediately yield Theorem 46.

Theorem 55 (Patkós [98]).

Let k, l' be positive integers with l' < k. Then the following inequality holds:

$$f(n,k,n-l') \le f(n,l',n-l') + La(n,P_{k-l'+1,2^{l'}}).$$

Theorem 56 (Patkós [98]).

For any positive integer l', the size of an (n-l')-trace l'-Sperner family $\mathcal{F} \subseteq 2^{[n]}$ is $O_{l'}(n^{-1/3}\binom{n}{\lfloor n/2 \rfloor})$.

Proof of Theorem 55. Let $\mathcal{F} \subseteq 2^{[n]}$ be a set family of size $f(n, l', n - l') + La(n, P_{k-l'+1,2^{l'}}) + 1$. We will find an l'-subset $L \subset [n]$ and a chain of length k + 1 in $\mathcal{F}|_{[n] \setminus L}$. By the size of \mathcal{F} , there exists a copy of $P_{k-l'+1,2^{l'}}$ in \mathcal{F} . We remove the set corresponding to the root of $T_{k-l'+1,2^{l'}}$ and repeat this procedure until there exists no more copy of $P_{k-l'+1,2^{l'}}$ in the remaining family. As $|\mathcal{F}| = f(n, l', n-l') + La(n, P_{k-l'+1,2^{l'}}) + 1$, we must have removed at least f(n, l', n - l') + 1 sets. Thus, there exists an l'-subset $L \subseteq [n]$ and l' + 1 removed sets $F_{k-l'+1}, F_{k-l'+2}, \dots, F_k, F_{k+1}$ such that

$$F_{k-l'+1}|_{[n]\setminus L} \subsetneq F_{k-l'+2}|_{[n]\setminus L} \subsetneq \dots \subsetneq F_{k}|_{[n]\setminus L} \subsetneq F_{k+1}|_{[n]\setminus L}$$

holds.

As $F_{k-l'+1}$ is a removed set, there exists a copy of $P_{k-l'+1,2^{l'}}$ such that $F_{k-l'+1}$ corresponds to its largest element. Therefore there are lots of chains of length k-l' in \mathcal{F} such that all of their elements are subsets of $F_{k-l'+1}$. Clearly, if $G \subseteq G'$, then $G|_{[n]\setminus L} \subseteq G'|_{[n]\setminus L}$, but we also require the sets of the chain not to coincide when considering their traces on [n] - L. Thus, we need a chain $F_1 \subsetneq F_2 \subsetneq \ldots \subsetneq F_{k-l'} \subsetneq F_{k-l'+1}$ such that $F_{i+1} \setminus F_i$ is not contained in L for all $i = 1, \ldots, k-l'$. Suppose we have already picked F_j from the *j*th level of the copy of $P_{k-l'+1,2^{l'}}$ for all $j = i+1, \ldots, k-l'+1$. Then F_{i+1} has $2^{l'}$ children in $P_{k-l'+1,2^{l'}}$. As for any F of these sets, we have $F_{i+1} \setminus F \neq \emptyset$, and L has $2^{l'} - 1$ non-empty subsets, at least one such F will satisfy $F|_{[n]\setminus L} \subseteq F_{i+1}|_{[n]\setminus L}$. Letting this F be F_i we continue to define all F_j 's and we get a chain of length k+1 in $\mathcal{F}|_{[n]\setminus L}$. This shows that \mathcal{F} cannot be (n-l')-trace k-Sperner.

Proof of Theorem 56. Let $\mathcal{F} \subseteq 2^{[n]}$ be an (n-l')-trace l'-Sperner family and let $\mathcal{F}_i = \{F \in \mathcal{F} : |F| = i\}$ for all i = 0, 1, ..., n. Note that if $\mathcal{H} \subseteq {[n] \choose i}$ is (n-l')-trace l'-Sperner, then \mathcal{H} does not contain sets $H_1, H_2, ..., H_{l'+1}$ such that for some $x_1, x_2, ..., x_{i+l'} \in [n]$ we have $H_j = \{x_j, x_{j+1}, ..., x_{j+i-1}\}$ for all j = 1, 2, ..., l' + 1 (sets satisfying these

conditions are often said to form a *tight path of length* l' + 1). Indeed, if such sets exist, then the traces of the H_j 's form a chain of length l' + 1 on the set $[n] \setminus \{x_1, x_2, ..., x_{l'}\}$ provided $i \ge l'$.

The results in the literature concerning uniform families not containing tight paths of given length concentrate on families with fixed uniformity and obtain bounds on their size from this view point. As "almost all sets in $2^{[n]}$ " have size around n/2, those results are not strong enough for our purposes, thus we prove the following lemma. The Turán number $ex_i(n, P_{l+1}^{tight})$ of the tight path of length l + 1 is the maximum number of sets in an *i*-uniform family \mathcal{F} that does not contain any tight paths of length l + 1.

Lemma 57 (Patkós [98]).

For any pair i, l of positive integers with i > l we have

$$ex_i(n, P_{l+1}^{tight}) \le ex_i(n, P_l^{tight}) + \frac{l}{i-l+1} \binom{n}{i-1}.$$

Proof. Let \mathcal{H} be a hypergraph on n vertices with $|\mathcal{H}| = ex_i(n, P_l^{tight}) + \frac{l}{i-l+1} \binom{n}{i-1} + 1$. By definition of the Turán number, \mathcal{H} must contain a tight path of length l. Removing the last edge of this path we can still find another tight path of length l. In this way, we find $\frac{l}{i-l+1} \binom{n}{i-1} + 1$ different edges in \mathcal{H} such that each of them is the last set in a certain tight path of length l.

Let \mathcal{H}_1 denote the subhypergraph of these edges and consider an edge $H \in E(\mathcal{H}_1)$. Let H' denote the first edge of (one of) the tight path(s) to which H belongs, i.e. if the vertices of the tight path are $x_1, x_2, ..., x_{i+l-1}$ and $H = \{x_l, x_{l+1}, ..., x_{i+l-1}\}$, then $H' = \{x_1, x_2, ..., x_i\}$. Let the modified shadow of H with respect to H' be $\{H \setminus \{x_j\} :$ $l \leq j \leq i\}$. The size of the modified shadow determined by any tight path is i - l + 1. Therefore, there exists an (i - 1)-set G that belongs to the modified shadows of at least l + 1 edges $H^1, H^2, ..., H^{l+1}$ from $E(\mathcal{H}_1)$.

Let $P_1, P_2, ..., P_l = H^1$ be a tight path of length l on the vertices $\{y_1, y_2, ..., y_{i+l-1}\}$ with $\{y_j, y_{j+1}, ..., y_{j+i-1}\} = P_j \in E(\mathcal{H})$ for all j = 1, 2, ..., l and let $G = H^1 \setminus \{y_t\}$ for some $l \leq t \leq i$. As the H^j 's are all different containing G and have size i at least one of them, say H^2 , is of the form $G \cup \{z\}$ such that $z \notin \{y_1, y_2, ..., y_{l-1}, y_t\}$. But then the sets $P_1, P_2, ..., P_l = H^1, H^2$ form a tight path of length l + 1 on the vertices $\{y_1, y_2, ..., y_{l-1}, y_t, y_l, y_{l+1}, ..., y_{k+l-1}, z\}$.

Corollary 58. For any pair *i*, *l* of integers with $l \leq i$ we have

$$ex_i(n, P_l^{tight}) \le \sum_{j=2}^l \frac{j-1}{i-j+2} \binom{n}{i-1}.$$

Proof. For a k-set S its shadow $\Delta(S)$ is simply $\binom{S}{k-1}$. Observe that for l = 2, the statement follows from the fact that if a k-uniform hypergraph \mathcal{H} on n vertices is P_2^{tight} -free, then the collection $\{\Delta(H) : H \in E(\mathcal{H})\}$ of shadows is pairwise disjoint, therefore $k \cdot |\mathcal{F}| \leq \binom{n}{k-1}$ must hold. The general statement then follows by induction on l using Lemma 57.

It is well known that $|\{X \subseteq [n] : ||X| - n/2| \ge n^{2/3}\}| = o(\frac{1}{n} \binom{n}{\lfloor n/2 \rfloor})$. Therefore applying Corollary 58 with l = l', we obtain

$$|\mathcal{F}| = o\left(\frac{1}{n}\binom{n}{\lfloor n/2 \rfloor}\right) + \sum_{i=n/2-n^{2/3}}^{n/2+n^{2/3}} |\mathcal{F}_i| = 2n^{2/3}O_{l'}\left(\frac{1}{n}\binom{n}{\lfloor n/2 \rfloor}\right) = O_{l'}\left(n^{-1/3}\binom{n}{\lfloor n/2 \rfloor}\right).$$

3.2 VC-Saturating families

In this subsection, we prove Theorem 48. Observe that in order to obtain an upper bound on the saturation number, one needs constructions. We will define random and explicit constructions in the next subsections, but before we gather the ideas that will be needed later and also present a toy example for the case d = 3.

The following proposition gives us an idea on how constant-sized saturated families should look like. In order to formulate it, we need some definitions. For a family $\mathcal{F} \subset 2^{[n]}$ and $x, y \in [n]$, we say that x and y are *duplicates*, if, for any $F \in \mathcal{F}$, $x \in F$ if and only if $y \in F$. Let $D(x) \subset [n]$ be the class of all duplicates of x with x included. Define the reduced family $\mathcal{R}(\mathcal{F})$ to be the projection of \mathcal{F} on W, where $W \subset [n]$ is obtained by keeping exactly one element out of each class of duplicates. Note that $\mathcal{R}(\mathcal{F})$ is defined up to relabeling of the ground set, $|\mathcal{R}(\mathcal{F})| = |\mathcal{F}|$ and, informally, $\mathcal{R}(\mathcal{F})$ captures the structure of \mathcal{F} . In the next proposition Δ denotes the symmetric difference of sets.

Proposition 59. Let $d \geq 2$ and consider a d-saturated family $\mathcal{F} \subset 2^{[n]}$.

(i) Assume that $\mathcal{F} = \mathcal{R}(\mathcal{F})$. If $x_1, \ldots, x_m \in [n]$ are such that, for any $F \in \mathcal{F}$ and x_i , the set $F \triangle \{x_i\}$ is not contained in \mathcal{F} , then a family \mathcal{F}' (on a larger ground set) that is obtained from \mathcal{F} by duplicating some of x_1, \ldots, x_m is d-saturated.

(ii) If there exists $x \in [n]$ such that $|D(x)| \ge 2$, then for any such x the family $\mathcal{R}(\mathcal{F})$ must satisfy the property from (i) w.r.t. (a duplicate of) x. That is, for any $F \in \mathcal{F}$, the set $F \triangle D(x)$ is not contained in \mathcal{F} .

We note that the condition of (ii) definitely holds for some x if $n > 2^{|\mathcal{F}|}$. This proposition implies that a constant-sized *d*-saturated family for any sufficiently large n is reducible to a saturated family as in (i).

Proof. (i) For simplicity, assume that \mathcal{F}' is obtained from \mathcal{F} by duplicating x_1 several times, and let $D(x_1)$ be the class of duplicates of x_1 . Assume that \mathcal{F}' is not saturated, that is, there is a set $X \notin \mathcal{F}'$ such that $\mathcal{F}' \cup \{X\}$ has VC-dimension d. Recall that $\mathcal{F}'|_{[n]} = \mathcal{F}$, and \mathcal{F} is saturated. Thus, $X|_{[n]} = F|_{[n]}$ for some $F \in \mathcal{F}'$. In other words, $\emptyset \neq F\Delta X \subset D(x_1) \setminus \{x_1\}$. Take any $y \in F\Delta X$, define $Y := [n] \setminus \{x_1\} \cup \{y\}$ and consider $\mathcal{F}_0 := \mathcal{F}'|_Y$. Then \mathcal{F}_0 is isomorphic to \mathcal{F} . By the choice of $y, F|_Y\Delta X|_Y = \{y\}$, and thus, by the definition of x_1 (and y being the duplicate of x_1 for \mathcal{F}'), only at most one of $F|_Y$ and $X|_Y$ can be contained in \mathcal{F}_0 . Therefore, $X|_Y \notin \mathcal{F}_0$, and thus $VC(\mathcal{F}_0 \cup \{X|_Y\}) > VC(\mathcal{F}_0)$, a contradiction.

(ii) The proof of this part is largely the proof of (i) in reverse. Assume that this is not the case. Take $F, F\Delta D(x) \in \mathcal{F}$, put $Y := (F \setminus D(x)) \cup \{x\}$ and consider the family $\mathcal{F}_1 := \mathcal{F} \cup \{Y\}$. Clearly, $|\mathcal{F}_1| = |\mathcal{F}| + 1$. Next, we show that $VC(\mathcal{F}_1) = VC(\mathcal{F}) = d$, contradicting the saturation property of \mathcal{F} . Indeed, assume that some (d+1)-element set S is shattered by \mathcal{F}_1 . Then, clearly, $S \cap D(x) \neq \emptyset$. Moreover, if $|S \cap D(x)| = 1$ then $Y|_S \in \{F|_S, F\Delta D(x)|_S\}$, and thus such S should have been shattered by \mathcal{F} . Therefore, $|S \cap D(x)| \ge 2$. However, by definition, there is at most 1 set in \mathcal{F}_1 that does not either contain or is disjoint with $S \cap D(x)$, while, in order to shatter S, one needs at least 2^d such sets. This contradiction shows that $VC(\mathcal{F}_1) = d$, and thus \mathcal{F} was not saturated in the first place.

One of the challenges in proving Theorem 48 is to find the right class of families to search for constructions in. Proposition 59 suggests to search for (reduced) saturated families such that the Hamming distance between any two sets in the family is at least 2. One natural way to achieve this is to consider *uniform families*, i.e., families in which all sets have the same size.

It turns out that we can find (d-1)-saturated families among intersecting families in $\binom{[2d]}{d}$. The following proposition gives us a sufficient condition for such a family to be (d-1)-saturated. We say that \mathcal{F} almost shatters X if $\mathcal{F}|_X = 2^X \setminus \{\emptyset\}$ or $\mathcal{F}|_X = 2^X \setminus \{X\}$.

Proposition 60. If a family $\mathcal{F} \subset {\binom{[2d]}{d}}$ almost shatters any $A \in {\binom{[2d]}{d}}$, then \mathcal{F} is (d-1)-saturated.

Proof. Since \mathcal{F} is almost shattered, adding a *d*-set to \mathcal{F} will result in shattering that set. We thus need to show that adding a set *B* of size other than *d* also results in some *d*-set being shattered. The argument is symmetric for sets of size smaller/larger than *d*, and we present the case |B| < d only. Consider a family $\mathcal{F}' := \mathcal{F} \cup \{B\}, |B| < d$. Take a set $X \subset \overline{B}, |X| = d - |B|$. Then $X \in \mathcal{F}|_{B \cup X}$, and so by the assumption on \mathcal{F} there is a *d*-set $A \in \mathcal{F}$ such that $A \cap (B \cup X) = X$. Therefore, $B \cap A = \emptyset$ and thus *A* is shattered by \mathcal{F}' .

Proposition 59 implies that any uniform *d*-saturated family on a ground set of size n can be transformed into a *d*-saturated family of the same size on any larger ground set. Proposition 60 tells us that it is sufficient to find a family $\mathcal{F} \subset {\binom{[2d]}{d}}$ that almost shatters any *d*-subset of [2d]. The latter property implies that, for any *d*-set S, exactly one of S, \bar{S} must be contained in \mathcal{F} . In other words, $\mathcal{F} \subset {\binom{[2d]}{d}}$ must be an intersecting family of size $\frac{1}{2} {\binom{2d}{d}}$.

In Section 3.2.1, we show that for $d \ge 16$, if we pick one set from each such complementary pair independently and uniformly at random, then with positive probability, we obtain a family that almost shatters every *d*-set.

In Section 3.2.2, we give explicit constructions of saturated families for any $d \ge 4$ that are based on intersecting families as above and have an additive combinatorics flavour. For odd d, we also obtain a certain classification result.

Before going on to constructions for general d, let us give a concrete example of a saturated family for d = 3, which proves Theorem 48 for that case, as well as gives an idea of what type of intersecting families we are going to use for explicit constructions.

Let $\mathcal{F} \subset {\binom{[6]}{3}}$ be the family of all 3-tuples in which the sum of the elements belongs to $H = \{1, 3, 4\} \mod 6$. Note that $\sum_{i=0}^{5} i = 3 \pmod{6}$ and that $H \cap (3 - H) = \emptyset$, where here and in what follows the operations are mod 6. This implies that \mathcal{F} contains exactly 1 set out of each complementary pair of 3-sets and that, in particular, \mathcal{F} is intersecting.

Claim 61. Every $A \in {\binom{[6]}{3}}$ is almost shattered by \mathcal{F} .

Proof. To prove the claim, it is sufficient to show that, for any $S' \subset S \in {\binom{[6]}{3}}, |S'| \in \{1,2\}$, there exists a set $F \in \mathcal{F}$ such that $F \cap S = S'$. Assume that the sum of the elements from S is x (mod 6) and the sum of elements from S' is y (mod 6).

If |S'| = 2 then we need to find $z \in \overline{S}$ such that $y + z \in H$. If there is no such z then $\{y + z : z \in \overline{S}\} = \{0, 2, 5\}$ and so $\{y + z : z \in S'\} \subset \{1, 3, 4\}$. But then, assuming $S' = \{z_1, z_2\}$, we have that the sum of the two elements $(y+z_1)+(y+z_2) = 3(z_1+z_2) = 0 \pmod{3}$, but on the other hand, it must be one of the numbers in $\{1 + 3, 1 + 4, 3 + 4\} \pmod{6}$, and none of those numbers is divisible by 3. This contradiction implies that there must be z with the desired property.

The case |S'| = 1 is similar. Put $\bar{S} = \{z_1, z_2, z_3\}$. Assuming that there is no pair $z_i, z_j \in \bar{S}, i \neq j$, such that $y + z_i + z_j \in \{1, 3, 4\}$, we get that $y + \{z_1 + z_2, z_1 + z_3, z_2 + z_3\} = \{0, 2, 5\}$, which, passing to the complements and using that $\sum_{i=0}^{5} i = 3 \pmod{6}$, means that $y' + y'' + \{z_1, z_2, z_3\} = \{1, 3, 4\}$, where $\{y, y', y''\} = S$. But then $y' + y'' + \{y, y', y''\} = \{0, 2, 5\}$, and, in particular, $3(y' + y'') \in \{0 + 2, 2 + 5, 0 + 5\}$, which is a contradiction. This concludes the proof of the claim.

Equipped with this claim, we apply Propositions 60, concluding that \mathcal{F} is saturated. We then apply Proposition 59 (i) and duplicate arbitrary elements sufficiently many times to get a saturated family of VC-dimension 2 for $n \geq 6$.

3.2.1 Random construction

Consider a random family $\mathcal{F} \subset {\binom{[2d]}{d}}$, obtained in the following way: for each pair A, \overline{A} of complementary *d*-element sets, we include one of them in \mathcal{F} independently and uniformly at random. Let H_A be an event that $A \in \mathcal{F}$.

For any d-set A and set $X \subseteq A$ let $Q_{A,X}$ stand for the event that $X \notin \mathcal{F}|_A$. This event happens if and only if for each pair of complementary d-sets B, \overline{B} such that $A \cap B = X$, we added \overline{B} to \mathcal{A} . In particular,

$$\mathsf{P}[Q_{A,X}] = 2^{-\binom{d}{|X|}}.$$

Theorem 62. If $d \ge 14$, we have $\mathsf{P}\left[\bigcap_{A \in \binom{[2d]}{d}, \emptyset \neq X \subset A} \bar{Q}_{A,X}\right] > 0$, *i.e.*, with positive probability \mathcal{F} almost shatters every $A \in \binom{[2d]}{d}$.

Equipped with this theorem, we can conclude the proof of Theorem 48 as in the case of d = 3.

We will use the Lovász Local Lemma to show the validity of Theorem 62.

Lemma 63 (Lovász Local Lemma [37, 106]). Let B_1, \ldots, B_m be events in an arbitrary probability space. For each i, let $S_i \subset [m]$ be such that B_i is independent of the sigmaalgebra generated by the events $\{B_j : j \notin S_i \cup \{i\}\}$. Assume that there are real numbers x_1, \ldots, x_m such that $0 \leq x_i < 1$ and

$$\mathsf{P}[B_i] \le x_i \prod_{j \in S_i} (1 - x_j).$$

Then with positive probability no event B_i holds.

Whether or not H_B holds only depends on those events $Q_{A,X}$ for which $A \cap B = X$. Thus, an event H_B depends on $\binom{d}{k}^2$ events $Q_{A,X}$ with |X| = k. Therefore, an event $Q_{A,X}$ depends on

$$d_{|X|,l} := \binom{d}{|X|} \binom{d}{l}^2$$

events $Q_{B,Y}$ with |Y| = l.

To apply Theorem 63, we need to choose the coefficients x_i . We put

$$x_{A,X} := p_{|X|} := 2^{-\max\{\binom{d-1}{|X|}, \binom{d-1}{d-|X|}\}}.$$

The cases $|X| \leq d/2$ and $|X| \geq d/2$ are symmetric, and thus, in what follows, we assume that $|X| \leq d/2$. Then the maximum in the expression above is attained on the first binomial coefficient and $\mathsf{P}[Q_{A,X}]/p_{|X|} = 2^{-\binom{d}{|X|} + \binom{d-1}{|X|}} = 2^{-\binom{d-1}{|X|-1}}$. We need to show that for each $1 \leq k \leq d/2$ and |X| = k we have

$$\mathsf{P}[Q_{A,X}] \le p_{|X|} \prod_{l=1}^{d-1} (1-p_l)^{d_{k,l}} \qquad \Leftrightarrow \qquad \prod_{l=1}^{d-1} (1-p_l)^{d_{k,l}} \ge 2^{-\binom{d-1}{k-1}}.$$

Recall that $d_{k,l} = {\binom{d}{k}} {\binom{d}{\ell}}^2$ and that ${\binom{d-1}{k-1}}/{\binom{d}{k}} = \frac{k}{d}$ and is minimized for k = 1. Thus, to verify the last displayed inequality, it is sufficient to show that

$$\prod_{l=1}^{d-1} (1-p_l)^{\binom{d}{l}^2} \ge 2^{-\frac{1}{d}}.$$
(5)

For $d/2 > l \ge 2$ and $d \ge 10$

$$\frac{\binom{d}{l}^2 2^{-\binom{d-1}{l}}}{\binom{d}{l-1}^2 2^{-\binom{d-1}{l-1}}} = \frac{(d-l)^2}{l^2} 2^{-\frac{d-2l}{d-l}\binom{d-1}{l}} \le (d-1)^2 2^{-\frac{(d-1)(d-4)}{2}} < \frac{1}{10}$$

and so we have

$$\prod_{l=1}^{d-1} (1-p_l) {\binom{d}{l}}^2 \ge \prod_{l=1}^{d/2} (1-p_l)^{2\binom{d}{l}}^2 \ge 1-2\sum_{l=1}^{d/2} {\binom{d}{l}}^2 2^{-\binom{d-1}{\ell}} \ge 1-3d^2 2^{1-d}.$$

The last expression is at least $1 - \frac{2}{d}$ for any d such that $2^d \ge 12d^3$. The latter holds for $d \ge 16$. On the other hand, for $d \ge 10$ we have $2^{-1/d} < 1 - \frac{1}{2d}$, and thus (5) holds for $d \ge 16$.

3.2.2 Explicit constructions

For odd $d \ge 7$ we find explicit constructions of intersecting families $\mathcal{F} \in {\binom{[2d]}{d}}$ which almost shatter any $A \in {\binom{[2d]}{d}}$. We then conclude the proof as in the case d = 3.

For even $d \ge 6$ the explicit constructions we found are slightly different. They consist of a maximal intersecting family in $\binom{[2d]}{d}$ and a few other sets, which form a saturated (and not necessarily almost-shattering) family.

Thus these constructions are not necessarily uniform, moreover they may contain two sets whose Hamming distance is one. Therefore, in order to use Proposition 59 (i) and extend the construction to larger n, we cannot simply duplicate an arbitrary element. However, we will make sure to have a distinguished element, for which the condition of Proposition 59 (i) holds.

Finally, we give two examples of saturated families for d = 4, 5. Those together with the example for d = 3 in the introduction cover all values of $d \ge 3$, as stated in Theorem 48.

The case of odd d

Fix an integer d and consider a set $X \subset [2d]$ of size d. Define

$$\mathcal{F}(X) := \left\{ F \in \binom{[2d]}{d} : \sum_{i \in F} i \in X \pmod{2d} \right\}$$

Theorem 64. Let d = 2k + 1. Then $\mathcal{F}(X)$ almost shatters every $S \in {\binom{[2d]}{d}}$ if and only if the following three conditions hold:

- 1. |X| = d and $X \cap (d X) = \emptyset \pmod{2d};$
- 2. X contains both odd and even elements;
- 3. for every $u \in X$, $\sum_{w \in X \setminus \{u\}} w \neq 0 \pmod{d}$.

It is not difficult to find residue classes that satisfy the three conditions from the theorem. We use the notation $[a, b] := \{a, a + 1, \dots, b\}$. Then one example is

$$X := [1, k] \cup [2k + 1, 3k + 1]$$

for odd k. Indeed, both 1. and 2. are straightforward to check. To see 3., we note that all elements of X are $0, 1, \ldots, k \pmod{2k+1}$, while

$$\sum_{x \in X} x = 2 \cdot \binom{k+1}{2} = k(k+1) = k+1 + \frac{k-1}{2}(2k+2) = \frac{3k+1}{2} \pmod{2k+1}.$$

Thus, condition 3. is satisfied.

Another example for odd k is

$$X := \{1, 3, \dots, 2k - 1, 2k + 1, 2k + 2, 2k + 4, \dots, 4k\}.$$

The previous examples give constructions in the case d = 4r + 3 for some positive integer r. In case $d = 4r + 1 \ge 9$, we can take

$$X := \{0\} \cup A \cup (d+A) \pmod{2d}, \text{ where } A = \{1, \dots, 2r-t-1, 2r-t, 2r+1, 2r+2, \dots, 2r+t\}$$

for some appropriately chosen $t \ge 1$. E.g., for k = 4 we can take t = 1, getting the set $\{1, 2, 3, 5\} \pmod{9}$. It is not difficult to check the first two conditions. As for the third condition, note that $2(\sum_{i=1}^{2r-t} i + \sum_{i=2r+1}^{2r+t} i) = r + t^2 \pmod{4r+1}$. By taking t such

that $2r + t < r + t^2 < 4r + 1$, which is always possible for $r \ge 2$, we make sure that the third condition is satisfied.

Proof of Theorem 64. Take any X as in the theorem. In what follows, we treat X as a set of residues modulo 2d, and the inclusions/equality between X and other sets should be interpreted as those for sets of residues modulo 2d. Most sums are also taken modulo 2d, which should be clear from the context. The proof of the theorem consists of the following lemmas.

Lemma 65. For any $A \in {\binom{[2d]}{d}}$, $\mathcal{F}(X)|_A$ contains exactly 1 out of \emptyset , A if and only if the first condition from Theorem 64 holds.

Proof of Lemma Note that $\sum_{i=1}^{2d} i = d \pmod{2d}$. Thus, the first condition in the theorem is equivalent to saying that for any $B \in \binom{[2d]}{d}$, $B \in \mathcal{F}(X)$ if and only if $\overline{B} \notin \mathcal{F}(X)$.

We will need some lower bounds on a special instance of the generalized Erdős– Heilbronn problem (originally [31], for a recent survey see Chapter IV A.3 in [7]). In a group G, the restricted s-sumset $\sum {A \choose s}$ for some $A \subseteq G$ and integer $s \ge 2$ is the set of all different sums of s distinct elements from A (in the number theory literature, the notation $s \land A$ is used). We will be interested in the case $G = \mathbb{Z}_{2d}$ and |A| = d. We will use a special case of the following result from [55]. A 2-coset is a coset of a subgroup with all non-zero elements having order 2, and an *almost 2-coset* is a 2-coset with possibly one element removed.

Theorem 66 (Girard, Griffiths, Hamidoune [55]). Let A be a subset of the abelian group G and let $2 \le s \le |A| - 2$. Then $|\sum {A \choose s}| \ge |A|$ unless $s \in \{2, |A| - 2\}$ and A is a 2-coset. Furthermore $|\sum {A \choose s}| > |A|$ unless A is a coset of a subgroup of G or $s \in \{2, |A| - 2\}$ and

(i) A is an almost 2-coset, or

(ii) |A| = 4 and A is the union of two cosets of a subgroup of order 2.

Corollary 67. Consider $A \subset \mathbb{Z}_{2d}$, $|A| = d \ge 5$, such that A contains both odd and even elements. Then for each $2 \le s \le d-2$, we have $|\sum {\binom{A}{s}}| > d$.

Lemma 68. The family $\mathcal{F}(X)|_A$ contains all the sets of size $s, 2 \leq s \leq d-2$ for any $A \in \binom{[2d]}{d}$ if and only if condition 2 from Theorem 64 holds.

Proof of Lemma. If X contains only even elements then it is not difficult to see that the $\mathcal{F}(X)|_A$, where A is the set of all odd elements, misses all projections of odd size. Similarly, if X contains only odd elements then $\mathcal{F}(X)|_A$, where A is the set of all even elements, misses all projections of even size. Thus, condition 2 from the theorem is necessary.

Conversely, take any such A and a particular subset $A' \subset A$ of size s. We need to show that it is possible to complement it with d-s elements in \overline{A} so that the sum of all elements in the resulting set belongs to X modulo 2d. Applying Corollary 67, we have $\left(\sum_{a \in A'} a + \sum {A \choose d-s}\right) \cap X \neq \emptyset \pmod{2d}$ by the pigeon-hole principle for any \overline{A} containing both odd and even elements. In case when \overline{A} is the set of all even or all odd

dc 1942 21

elements, $\sum {A \choose d-s}$ contains either all even or all odd elements. In any case, condition 2 from the theorem implies that the aforementioned intersection is non-empty as well. \Box

Lemma 69. The family $\mathcal{F}(X)|_A$ contains all the sets of size 1 and d-1 for any $A \in {\binom{[2d]}{d}}$ if and only if condition 3 from Theorem 64 holds.

Proof of Lemma. If $u \in X$ is such that $\sum_{w \in X \setminus \{u\}} w = 0 \pmod{2d}$ then $\mathcal{F}(X)|_X$ does not contain $X \setminus \{u\}$. Indeed, we need an element y from \overline{X} to complement $X \setminus \{u\}$ to obtain a set with sum in X. But then $y \in X \cap \overline{X}$. If $u \in X$ is such that $\sum_{w \in X \setminus \{u\}} w = d \pmod{2d}, \text{ then consider the sets } A = d + X \text{ and } A' = A \setminus \{u + d\}.$ We have $\sum_{a \in A'} a = (d - 1)d + \sum_{x \in X \setminus \{u\}} x = d \pmod{2d}, \text{ because } (d - 1)d = 0 \pmod{2d} \text{ (here we use the fact that } d \text{ is odd}). We claim that there is no } F \in \mathcal{F}(X)$ such that $F \cap A = A'$. Indeed, if there is $b \in \overline{A}$ such that $\sum_{y \in A' \cup \{b\}} y \in X$ then $b \in X - \sum_{y \in A'} y = X - d = A$, a contradiction. Thus, condition 3 from the theorem is necessary.

Conversely, fix some A. Let us first show that, for any $A' \subset A$ of size d-1, there is $b \in B := \overline{A}$ such that $A' \cup \{b\} \in \mathcal{F}(X)$.

Assume that this is not the case. Then the set $B + \sum_{a \in A'} a \subset Y := \overline{X} \pmod{2d}$, and, given that |B| = |Y| = d, we have

$$B + \sum_{a \in A'} a = Y.$$

Let us put $z := \sum_{a \in A'} a \pmod{2d}$. Then, since $B \sqcup A = X \sqcup Y$, we have A + z = X. Put $X' := \{a' + z : a' \in A'\}$. We get that $z = \sum_{a \in A'} a = \sum_{x \in X'} x - z|X'|$. Rewriting this, we have $\sum_{x \in X'} x = dz$, and thus $\sum_{x \in X'} x = 0 \pmod{d}$. This contradicts condition 3 from the theorem.

The case when $A' = \{a\}$ is very similar. We show that for any a there is $b \in B$ such that $\{a\} \cup B \setminus \{b\} \in \mathcal{F}(X)$.

Assume that this is not the case. Then

$$\left(a + \sum_{b \in B} b\right) - B = Y.$$

Let us put $z := a + \sum_{b \in B} b$. Then, since $B \sqcup A = X \sqcup Y$, we have z - A = X. Put x' := z - a and note that $x' \in X$. We get that $z = a + \sum_{b \in B} b = z - x' + dz - \sum_{y \in Y} y$. Recall that $\sum_{y \in Y} y + \sum_{x \in X} x = d$. Using the last two equations, we again get $\sum_{x \in X \setminus \{x'\}} x = 0 \pmod{d}$. This contradicts condition 3 from the theorem. \Box

This concludes the proof of Theorem 64.

The case of even d

The first part of the argument in this case follows essentially the same steps as the argument in the case of odd d. Fix an even integer d and consider a set $X \subset [2d]$ of size d - 1. Define

$$\mathcal{F}_1(X) = \left\{ F \in \binom{[2d]}{d} : 2d \in F \text{ and } \sum_{i \in F} i \in X \cup \left\{ \frac{d}{2} \right\} \pmod{2d} \right\}.$$

$$\mathcal{F}_2(X) = \Big\{ F \in \binom{[2d]}{d} : 2d \notin F \text{ and } \sum_{i \in F} i \in X \cup \Big\{ \frac{3d}{2} \Big\} \pmod{2d} \Big\}.$$

Proposition 70. Let d = 2k. For $X \subset [2d] \setminus \{\frac{d}{2}, \frac{3d}{2}\}$ the family $\mathcal{F} := \mathcal{F}_1(X) \cup \mathcal{F}_2(X)$ almost shatters every $S \in \binom{[2d]}{d}$ with $2d \in S$ if the following three conditions hold.

- 1. |X| = d 1 and $X \cap (d X) = \emptyset \pmod{2d}$;
- 2. X contains both odd and even elements.
- 3. For $X_1 = X \cup \{\frac{d}{2}\}$ and for every $u \in X_1$ we have $\sum_{w \in X_1 \setminus \{u\}} w \neq 0 \pmod{d}$. Equivalently, for $X_1 = X \cup \{\frac{3d}{2}\}$ and for every $u \in X_1$ we have $\sum_{w \in X_1 \setminus \{u\}} w \neq 0 \pmod{d}$.

The following three claims imply the proposition.

Claim 71. For every $S \in {\binom{[2d]}{d}}$ exactly one of S, \overline{S} belongs to \mathcal{F} .

Proof. If $\sum_{i \in S} i \in X$ it follows from the first condition as in Lemma 65. If $\sum_{i \in S} i = \frac{d}{2} \pmod{2d}$ or $\sum_{i \in S} i = \frac{3d}{2} \pmod{2d}$, then $\sum_{i \in S} i = \sum_{i \in \overline{S}} i \pmod{2d}$, thus $S \in \mathcal{F}$ and $\overline{S} \notin \mathcal{F}$.

Claim 72. For any $S \in {\binom{[2d]}{d}}$ with $2d \in S$, every subset of size $2 \leq s \leq d-2$ of S appears in $\mathcal{F}|_S$.

Proof. Let $S' \subset S$ be a subset of size s. We consider two cases.

Case 1: $2d \in S'$. In this case we have to show that it is possible to complement it with d-s elements in \overline{S} such that the sum of all elements in the resulting set belongs to $X \cup \{\frac{d}{2}\}$. If \overline{S} is the set of all even or all odd elements, $\sum {\binom{\overline{S}}{d-s}}$ contains all even or all odd elements. Otherwise, Corollary 67 implies $|\sum {\binom{\overline{S}}{d-s}}| > d$, and we are done since $|X \cup \{\frac{d}{2}\}| = d$.

Case 2: $2d \notin S'$. In this case we have to show that it is possible to complement it with d-s elements in \overline{S} such that the sum of all elements in the resulting set belongs to $X \cup \{\frac{3d}{2}\}$. This can be done in the same way as we handled Case 1.

Claim 73. For any $S \in {\binom{[2d]}{d}}$ with $2d \in S$ we have ${\binom{S}{1}} \cup {\binom{S}{d-1}} \subseteq \mathcal{F}|_S$.

Proof. There are 4 types of sets to consider.

Type 1: $S' \in {S \choose d-1}$ with $2d \in S'$. To prove that S' belongs to $\mathcal{F}|_S$, we have to show that there is a $b \in \overline{S}$ such that $S' \cup \{b\} \in \mathcal{F}_1(X)$.

Assume that this is not the case. Let $X_1 = X \cup \{ \frac{d}{2} \}$ and $\overline{S} + \sum_{s \in S'} s \subset Y := \overline{X}_1 \pmod{2d}$. Given that $|\overline{S}| = |Y| = d$, we have

$$\overline{S} + \sum_{s \in S'} s = Y.$$

Let us put $z := \sum_{s \in S'} s \pmod{2d}$. Then, since $\overline{S} \sqcup S = X_1 \sqcup Y$, we have $S + z = X_1$. Put $X' := \{s' + z : s' \in S'\}$. We get that $z = \sum_{s \in S'} s = \sum_{x \in X'} x - z|X'|$. Rewriting this,

we have $\sum_{x \in X'} x = dz$, and thus $\sum_{x \in X'} x = 0 \pmod{d}$. This contradicts condition 3 from the proposition.

Type 2: $S' \in {S \choose d-1}$ with $2d \notin S'$. To prove that S' belongs to $\mathcal{F}|_S$, we have to show that there is a $b \in \overline{S}$ such that $S' \cup \{b\} \in \mathcal{F}_2(X)$. We can proceed in the same way as in the case of Type 1, with letting $X_1 = X \cup \{\frac{3d}{2}\}$.

Type 3: $S' = \{a\}$ with $a \neq 2d$. To prove that S' belongs to $\mathcal{F}|_S$, we have to show that there is a $b \in \overline{S}$ such that $\overline{S} \setminus \{b\} \cup \{a\} \in \mathcal{F}_2(X)$.

Assume that this is not the case and let $X_1 = X \cup \{\frac{3d}{2}\}$. Then

$$\left(a + \sum_{b \in \overline{S}} b\right) - \overline{S} = \overline{X}_1.$$

Let us put $z := a + \sum_{b \in \overline{S}} b$. Then, since $\overline{S} \sqcup S = X_1 \sqcup \overline{X}_1$, we have $z - S = X_1$. Put x' := z - a and note that $x' \in X_1$. We get that $z = a + \sum_{b \in \overline{S}} b = z - x' + dz - \sum_{x \in \overline{X}_1} x$. Recall that $\sum_{x \in X_1} x + \sum_{x \in \overline{X}_1} x = d \pmod{2d}$. Using the last two equations, we again get $\sum_{x \in X_1 \setminus \{x'\}} x = 0 \pmod{d}$. This contradicts condition 3 from the proposition.

Type 4: $S' = \{a\}$ with a = 2d. To prove that S' belongs to $\mathcal{F}|_S$, we have to show that there is a $b \in \overline{S}$ such that $\overline{S} \setminus \{b\} \cup \{a\} \in \mathcal{F}_1(X)$. We can proceed in the same way as in the case of Type 3, with letting $X_1 = X \cup \{\frac{d}{2}\}$.

Proposition 74. If X is as in Proposition 70 and $A \cup \mathcal{F}_1(X) \cup \mathcal{F}_2(X)$ has VC-dimension (d-1) for some $A \subset [2d]$ then $2d \in S$ if |A| < d and $2d \notin A$ if |A| > d.

Proof. We only prove the first half of the statement, the second can be done similarly. Assume |A| < d and $2d \notin A$. Since any A' of size d that contains $A \cup \{2d\}$ is almost shattered by Proposition 70, there is an $S \in \mathcal{F}_1(X)$ such that $S \cap A' = \{2d\}$. This S is disjoint with A. By Proposition 70, S is almost shattered, and since $S \in \mathcal{F}_1(X)$, A cannot be added without increasing the dimension, as it gives the missing empty projection of S.

Now we are ready to find the constructions even d = 2k with $d \ge 6$. It is not hard to check that

$$X = \left\{2d, \frac{d}{2} + 1\right\} \cup \left[2, \frac{d}{2} - 1\right] \cup \left[d + 1, \frac{3d}{2} - 1\right]$$

satisfies the conditions in Proposition 70.

Take $\mathcal{F}_1(X) \cup \mathcal{F}_2(X)$, which is a maximal intersecting family in $\binom{[2d]}{d}$, and add some other sets to it, until we obtain a saturated family \mathcal{F} . Then it follows from Proposition 74 that for any $F \in \mathcal{F}$ the set $F\Delta\{2d\}$ is not contained in \mathcal{F} . Thus we can duplicate $\{2d\}$, and obtain a construction on any ground set [n] for $n \geq 2d$.

The cases d = 4, 5

To complete the proof of Theorem 48, we need to handle the cases d = 4, 5. We provide two constructions which we have found using computer search.

Let σ be the cyclic permutation $(1 \ 2 \dots 2d - 1)$ and let $\mathcal{F} \subset {\binom{[2d]}{d}}$ be a family. Then we put $\mathcal{P}(\mathcal{F}) := \{\sigma^i F, F \in \mathcal{F}, i = 1, \dots, 2d - 1\}$. For d = 4 we take

$$\mathcal{F}_4 = \big\{ \{1, 2, 3, 4\}, \{1, 2, 3, 5\}, \{1, 2, 4, 5\}, \{1, 2, 4, 8\}, \{1, 3, 5, 8\} \big\},\$$

and for d = 5 we take

$$\mathcal{F}_5 = \{\{1, 2, 3, 4, 5\}, \{1, 2, 3, 4, 6\}, \{1, 2, 3, 4, 8\}, \{1, 2, 3, 5, 6\}, \{1, 2, 3, 5, 7\}, \\ \{1, 2, 3, 5, 8\}, \{1, 2, 3, 6, 8\}, \{1, 2, 3, 6, 10\}, \{1, 2, 4, 5, 8\}, \{1, 2, 4, 5, 10\}, \\ \{1, 2, 5, 6, 10\}, \{1, 2, 5, 7, 10\}, \{1, 2, 5, 8, 10\}, \{1, 3, 5, 7, 10\}\}.$$

We then check using computer that $\mathcal{P}(\mathcal{F}_4), \mathcal{P}(\mathcal{F}_5)$ are saturated.

Note that $\mathcal{P}(\mathcal{F}_4)$, $\mathcal{P}(\mathcal{F}_5)$ are intersecting, and if adding a set $A \in 2^{[2d]}$ to the family increases the VC-dimension, then adding $\sigma^i A$ increases it as well. We also note that $\mathcal{P}(\mathcal{F}_4)$, $\mathcal{P}(\mathcal{F}_5)$ do not have the almost-shattering property.

4 *q*-analog results

In this last section of the dissertation, we describe and prove so-called q-analogs of results in extremal set theory. Roughly speaking, to introduce a q-analog problem, one needs to replace the underlying set [n] of a set system problem by an n-dimensional vector space over the finite field of order q and set size by dimension. Sometimes the obtained q-analog can be tackled very similarly to the original problem: these are not very interesting cases. But often the methods used to solve the set system problem are not applicable in the subspace setting.

We repeat some of the definitions and classical results from Section 2. Let X be an *n*-element set and, for $0 \le k \le n$, let $\binom{X}{k}$ denote the family of all subsets of X of cardinality k. For a family $\mathcal{F} \subset \binom{X}{k}$, we define the shadow of \mathcal{F} , denoted $\partial \mathcal{F}$, to consist of those (k-1)-subsets of X contained in at least one member of \mathcal{F} ,

$$\partial \mathcal{F} := \left\{ E \in \begin{pmatrix} X \\ k-1 \end{pmatrix} : E \subset F \in \mathcal{F} \right\}.$$

Kruskal [77] and Katona [69] determined the minimum size of the shadow of \mathcal{F} as a function of k and the size of \mathcal{F} . Recall that the binomial coefficient

$$\binom{y}{k} := \frac{y(y-1)\cdots(y-k+1)}{k!}$$

can be defined for all $y \in \mathbb{R}$ and $k \in \mathbb{Z}^+$. Lovász [78, Ex 13.31(b)] proved the following weaker but more convenient version of the Kruskal-Katona theorem.

Theorem 75 (Lovász). Let $\mathcal{F} \subset {X \choose k}$ and let $y \ge k$ be the real number defined by $|\mathcal{F}| = {y \choose k}$. Then $|\partial \mathcal{F}| \ge {y \choose k-1}$. If equality holds, then $y \in \mathbb{Z}^+$ and $\mathcal{F} = {Y \choose k}$, where Y is a y-subset of X.

The main contribution of Section 4.1 of the dissertation is the following analog of Theorem 75. The proof adapts the approach of Keevash used in a simple proof [72] of Theorem 75.

Theorem 76 (Chowdhury, Patkós [22]). Let $\mathcal{F} \subset \begin{bmatrix} V \\ k \end{bmatrix}$ and let $y \ge k$ be the real number defined by $|\mathcal{F}| = \begin{bmatrix} y \\ k \end{bmatrix}$. Then

$$|\partial \mathcal{F}| \ge \begin{bmatrix} y\\ k-1 \end{bmatrix}.$$

If equality holds, then $y \in \mathbb{Z}^+$ and $\mathcal{F} = \begin{bmatrix} Y \\ k \end{bmatrix}$, where Y is a y-dimensional subspace of V.

Not much was known about shadows in vector spaces before the publication of our paper [22]. In [9], a partial analog of the Kruskal-Katona theorem is given when V is a vector space over the field \mathbb{F}_2 . The only other result on shadows in vector spaces, which was known, appears in [43].

The Kruskal-Katona shadow theorem and its approximate but more computable version Theorem 75 have many applications. As another contribution of the dissertation, we will present an application of Theorem 76 to obtain the q-analog of the following result of Frankl. A family \mathcal{F} of sets is said to be r-wise intersecting if for any $F_1, F_2, \ldots, F_r \in \mathcal{F}$ the intersection $F_1 \cap F_2 \cap \ldots, \cap F_r$ is non-empty.

Theorem 77 (Frankl, [40]). Suppose that $\mathcal{F} \subset {\binom{X}{k}}$ is r-wise intersecting and $(r-1)n \ge rk$. Then $|\mathcal{F}| \le {\binom{n-1}{k-1}}$. Moreover, excepting the case r = 2 and n = 2k, equality holds if and only if $\mathcal{F} = \left\{ F \in {\binom{X}{k}} : x \in F \right\}$ for some $x \in X$.

We will use Theorem 76 to extend Theorem 77 to vector spaces. A family $\mathcal{F} \subset \begin{bmatrix} V \\ k \end{bmatrix}$ is called *r*-wise intersecting if for all $F_1, \ldots, F_r \in \mathcal{F}$ we have $\bigcap_{i=1}^r F_i \neq \{0\}$.

Theorem 78 (Chowdhury, Patkós [22]). Suppose $\mathcal{F} \subset \begin{bmatrix} V \\ k \end{bmatrix}$ is r-wise intersecting and $(r-1)n \geq rk$. Then

$$|\mathcal{F}| \le {n-1 \choose k-1}.$$

Moreover, equality holds if and only if $\mathcal{F} = \left\{F \in \begin{bmatrix} V \\ k \end{bmatrix} : v \subset F\right\}$ for some onedimensional subspace $v \subset V$, unless r = 2 and n = 2k.

The case r = 2 of Theorem 78 is the Erdős-Ko-Rado theorem for vector spaces, which has been extensively studied. Hsieh [66] first proved the Erdős-Ko-Rado theorem for vector spaces, but not for all relevant n and his proof involves many lengthy calculations. Later, Frankl and Wilson [46] proved the Erdős-Ko-Rado theorem for vector spaces, essentially by computing the eigenvalues of the so-called q-Kneser graph; the q-Kneser graph has the k-dimensional subspaces of V as its vertices, where two subspaces α, β are adjacent if $\alpha \cap \beta = \{0\}$. While Frankl and Wilson's method is less computational than Hsieh's, finding the eigenvalues of the q-Kneser graph still requires some calculations.

It is unclear where the characterization of equality in the case n = 2k of the Erdős-Ko-Rado theorem for vector spaces first appeared in the literature. Godsil and Newman [56, 92] gave a characterization of equality in this case using techniques similar to those of Frankl and Wilson [46].

Greene and Kleitman [57] gave a very elegant proof to the Erdős-Ko-Rado theorem for vector spaces when k|n. Deza and Frankl [27] sketched an inductive proof of the Erdős-Ko-Rado theorem for vector spaces using Greene and Kleitman's proof for the base case n = 2k and a generalization of the shifting technique.

In Section 4.2, we present the q-analog of the Hilton-Milner theorem and, as a consequence, we determine the chromatic number of q-Kneser graphs. The Hilton-Milner theorem is a strong stability version of the Erdős-Ko-Rado theorem on uniform intersecting families. It states how large can an intersecting family $\mathcal{F} \subseteq {\binom{[n]}{k}}$ be if it is not a subfamily of a trivially intersecting family. Or equivalently if the covering number $\tau(\mathcal{F})$, the smallest possible size of a set S (a cover of \mathcal{F}) that meets every set $F \in \mathcal{F}$, is at least two. Let us remark that there exist sharp upper bounds on the size of intersecting set families with covering number at least 3 or at least 4 [41, 44].

Theorem 79. (Hilton, Milner [65]) Let $\mathcal{F} \subset {X \choose k}$ be an intersecting family with $k \geq 3$, $n \geq 2k + 1$ and $\tau(\mathcal{F}) \geq 2$. Then $|\mathcal{F}| \leq {n-1 \choose k-1} - {n-k-1 \choose k-1} + 1$. The families achieving that size are

(i) for any k-subset F and $x \in X \setminus F$ the family

$$\mathcal{F}_{HM} = \{F\} \cup \{G \in \binom{X}{k} : x \in G, \ F \cap G \neq \emptyset\},\$$

(ii) if k = 3, then for any 3-subset S the family

$$\mathcal{F}_3 = \{F \in \binom{X}{3} : |F \cap S| \ge 2\}.$$

We will prove a q-analog of Theorem 79 for intersecting families of subspaces with $\tau(\mathcal{F}) \geq 2$. Trivially intersecting families of subspaces are called point-pencils. The point of a point-pencil is the one-dimensional subspace contained in all k-subspaces (the pencils) of the point-pencil.

Let us first remark that for a fixed 1-subspace $E \leq V$ and a k-subspace U with $E \leq U$ the family $\mathcal{F}_{E,U} = \{U\} \cup \{W \in \begin{bmatrix} V \\ k \end{bmatrix} : E \leq W, \dim(W \cap U) \geq 1\}$ is not maximal as we can add all subspaces in $\begin{bmatrix} E \lor U \\ k \end{bmatrix}$. We will say that \mathcal{F} is an *HM-type family* if

$$\mathcal{F} = \left\{ W \in \begin{bmatrix} V \\ k \end{bmatrix} : E \leqslant W, \ \dim(W \cap U) \ge 1 \right\} \cup \begin{bmatrix} E \lor U \\ k \end{bmatrix}$$

for some fixed $E \in {V \brack 1}$ and $U \in {V \brack k}$ with $E \not\leq U$. Note that the size of an HM-type family is ${n-1 \brack k-1} - q^{k(k-1)} {n-k-1 \brack k-1} + q^k$.

Theorem 80 (Blokhuis, Brouwer, Chowdhury, Frankl, Mussche, Patkós, Szőnyi [10]). Let V be an n-dimensional vector space over $\mathbf{F}(q)$, where $q \ge 3$ and $n \ge 2k + 1$ or q = 2 and $n \ge 2k + 2$, $k \ge 3$. Then for any intersecting family $\mathcal{F} \subseteq \begin{bmatrix} V \\ k \end{bmatrix}$ with $\tau(\mathcal{F}) \ge 2$ we have

$$|\mathcal{F}| \le {\binom{n-1}{k-1}} - q^{k(k-1)} {\binom{n-k-1}{k-1}} + q^k.$$

When equality holds, either \mathcal{F} is an HM-type family, or k = 3 and $\mathcal{F} = \mathcal{F}_3 = \{F \in \begin{bmatrix} V \\ k \end{bmatrix} : \dim(S \cap F) \ge 2\}$ for some $S \in \begin{bmatrix} V \\ 3 \end{bmatrix}$.

Furthermore, if $k \ge 4$, then there exists an $\varepsilon > 0$ (independent of n, q, k) such that if $|\mathcal{F}| \ge (1 - \varepsilon) \left({n-1 \brack k-1} - q^{k(k-1)} {n-k-1 \brack k-1} + q^k \right)$, then \mathcal{F} is a subfamily of an HM-type family.

After proving the above theorem, we apply this result to determine the chromatic number of q-Kneser graphs. The vertex set of the q-Kneser graph $qK_{n:k}$ is $\begin{bmatrix} V \\ k \end{bmatrix}$, where V is an n-dimensional vector space over \mathbf{F}_q . Two vertices of $qK_{n:k}$ are adjacent if and only if the corresponding k-subspaces are disjoint. We obtain the following theorem.

Theorem 81 (Blokhuis, Brouwer, Chowdhury, Frankl, Mussche, Patkós, Szőnyi [10]). If $k \ge 3$, $n \ge 2k+1$ and $q \ge 3$ or $n \ge 2k+2$, and q = 2, then for the chromatic number of the q-Kneser graph we have $\chi(qK_{n:k}) = {n-k+1 \choose 1}$. Moreover, each color class of a minimum coloring is a point-pencil and the points determining a color are the points of an (n-k+1)-dimensional subspace.

Let us comment on the past ten years of the results presented in this section. The paper containing Theorem 76 and Theorem 78 is the most often cited paper of the author. This is mainly due to the fact that any paper that proves q-analogs of some extremal set theory results, mentions some other examples, and since the Kruskal-Katona theorem is one of the classic results in extremal set theory, our q-analog is often cited. However, our result (so far) has far fewer applications than the original shadow theorem. One possible reason to this (apart from the obvious fact that q-analogs are less studied, than set families) could be that for subpsace family sizes between xbrackk and $\begin{bmatrix} x+1\\ k \end{bmatrix}$, our result gives bounds less sharp than Theorem 75, so additional results might be needed as in [101].

The paper containing the proof of Theorem 80 and Theorem 81 has the most number of essential citations as it is very much applicable in obtaining structural results in q-Kneser graphs and similar objects. Let us finally mention that Ihringer determined [67] the chromatic number $\chi(qK_{n:k})$ in most of the cases left open in Theorem 81: $q \ge 5$, n = 2k.

Notation. In the next two subsections, for subpaces A and B we denote by $A \vee B$ the subpace generated by $A \cup B$ and if B is a 1-space generated by the vector v, then we just write $A \vee v$. For a subpace $A \leq V$ we write A^{\perp} to denote the subspace of all vectors $v \in V$ that are perpendicular to all vectors of A.

4.1 Shadows of subspaces

The most likely reason for which no q-analog of the Kruskal-Katona shadow theorem had been obtained for so long is that all proofs of the original statement and its relaxations relied heavily on shifting techniques and the shifting operation cannot be introduced in the vector space setting. In 2008, Peter Keevash [72] gave a short new inductive proof of Theorem 75 that did not make use of the shifting operation. In this subsection, we adapt his argument to prove Theorem 76.

We first collect some definitions and facts that will be used in the proof of Theorem 76. If $\mathcal{F} \subset \begin{bmatrix} V \\ k \end{bmatrix}$, then

$$K_{k+1}^{k}(\mathcal{F}) := \left\{ T \in \begin{bmatrix} V\\ k+1 \end{bmatrix} : \begin{bmatrix} T\\ k \end{bmatrix} \subset \mathcal{F} \right\}$$

is the family of (k+1)-dimensional subspaces in V all of whose k-dimensional subspaces lie in \mathcal{F} . If $v \in \begin{bmatrix} V \\ 1 \end{bmatrix}$, then

$$K_{k+1}^{k}(v) := \{ T \in K_{k+1}^{k}(\mathcal{F}) : v \subset T \}$$

is the family of (k + 1)-dimensional subspaces in $K_{k+1}^k(\mathcal{F})$ that contain v. For $v \in \begin{bmatrix} V \\ 1 \end{bmatrix}$, define the degree of v, which is denoted by d(v), to be the number of elements of \mathcal{F} that contain v. If $v \in \begin{bmatrix} V \\ 1 \end{bmatrix}$ and $U \subset V$ is an (n-1)-dimensional subspace not containing v then

$$L_U(v) := \left\{ A \in \begin{bmatrix} U\\ k-1 \end{bmatrix} : A \lor v \in \mathcal{F} \right\}$$

is the family of (k-1)-dimensional spaces in U whose linear span with v is an element of \mathcal{F} . Observe that $d(v) = |L_U(v)|$.

Finally, we collect some notation and facts regarding the Gaussian binomial coefficients. When k = 1, we will write the Gaussian binomial coefficient $\begin{bmatrix} a \\ 1 \end{bmatrix}$ as [a]. For $a \in \mathbb{Z}^+$, we define $[a]! = \prod_{j=1}^{a} [j]$. A familiar relation involving binomial coefficients is Pascal's identity. We note two similar relations involving Gaussian binomial coefficients.

Lemma 82. If $a \in \mathbb{R}$ and $k \in \mathbb{Z}^+$, then

$$\begin{bmatrix} a \\ k \end{bmatrix} = q^{a-k} \begin{bmatrix} a-1 \\ k-1 \end{bmatrix} + \begin{bmatrix} a-1 \\ k \end{bmatrix} = \begin{bmatrix} a-1 \\ k-1 \end{bmatrix} + q^k \begin{bmatrix} a-1 \\ k \end{bmatrix}.$$

Instead of proving Theorem 76, we will show the following result.

Theorem 83. Let $\mathcal{F} \subset \begin{bmatrix} V \\ k \end{bmatrix}$ and let $y \ge k$ be the real number defined by $|\mathcal{F}| = \begin{bmatrix} y \\ k \end{bmatrix}$. Then

$$|K_{k+1}^k(\mathcal{F})| \le \begin{bmatrix} y\\k+1 \end{bmatrix}.$$

Equality holds if and only if $y \in \mathbb{Z}^+$ and $\mathcal{F} = \begin{bmatrix} Y \\ k \end{bmatrix}$ for some y-dimensional subspace $Y \subset V$.

We observe that Theorem 83 implies Theorem 76. Indeed, let \mathcal{F} be as in Theorem 76, and let $x \ge k-1$ be the real number defined by $|\partial \mathcal{F}| = \begin{bmatrix} x \\ k-1 \end{bmatrix}$. By Theorem 83, we have

$$\begin{bmatrix} y \\ k \end{bmatrix} = |\mathcal{F}| \le |K_k^{k-1}(\partial \mathcal{F})| \le \begin{bmatrix} x \\ k \end{bmatrix}$$

because $\mathcal{F} \subset K_k^{k-1}(\partial \mathcal{F})$. Hence $x \geq y$ so $|\partial \mathcal{F}| = \begin{bmatrix} x \\ k-1 \end{bmatrix} \geq \begin{bmatrix} y \\ k-1 \end{bmatrix}$. If $|\partial \mathcal{F}| = \begin{bmatrix} y \\ k-1 \end{bmatrix}$ then x = y. Hence, $|K_k^{k-1}(\partial \mathcal{F})| = \begin{bmatrix} y \\ k \end{bmatrix}$ and $\mathcal{F} = K_k^{k-1}(\partial \mathcal{F})$. By Theorem 83, this implies $y \in \mathbb{Z}^+$ and $\partial \mathcal{F} = \begin{bmatrix} Y \\ k-1 \end{bmatrix}$ for some y-dimensional subspace $Y \subset V$. Clearly, $\begin{bmatrix} Y \\ k \end{bmatrix} = K_k^{k-1}(\partial \mathcal{F}) = \mathcal{F}$.

Proof of Theorem 83: We argue by induction on k. The base case k = 1 is easy: Suppose $\mathcal{F} \subset \begin{bmatrix} V \\ 1 \end{bmatrix}$ and $|\mathcal{F}| = [y]$. Since there are q + 1 one-dimensional spaces in a two-dimensional space, $|K_2^1(v)| \leq (1/q)([y] - 1)$ if $v \in \mathcal{F}$ and $|K_2^1(v)| = 0$ otherwise. Now

$$(q+1)|K_2^1(\mathcal{F})| = \sum_{v \in {V \brack 1}} |K_2^1(v)| \le \frac{|y|(|y|-1)}{q},\tag{6}$$

which implies that $|K_2^1(\mathcal{F})| \leq {y \choose 2}$.

Suppose $T \in K_{k+1}^k(v)$. Observe that the q^k k-dimensional subspaces in T that do not contain v are elements of \mathcal{F} that do not contain v. Moreover, if $U \subset V$ is an (n-1)-dimensional subspace that does not contain v, then $T \cap U$ is a k-dimensional subspace in $K_k^{k-1}(L_U(v))$. The first condition implies that

$$q^{k}|K_{k+1}^{k}(v)| = \left| \left\{ S \in \begin{bmatrix} V\\ k \end{bmatrix} : v \not\subset S \subset T \in K_{k+1}^{k}(v) \right\} \right| \le |\mathcal{F}| - d(v),$$

and hence that $|K_{k+1}^k(v)| \leq (1/q^k)(|\mathcal{F}| - d(v))$. The second condition implies that $|K_{k+1}^k(v)| \leq |K_k^{k-1}(L_U(v))|$ because if T_1, T_2 are distinct elements of $K_{k+1}^k(v)$ then $T_1 \cap U$ and $T_2 \cap U$ are distinct elements of $K_k^{k-1}(L_U(v))$.

We claim that $|K_{k+1}^k(v)| \leq ([y-k]/[k])d(v)$ for all $v \in {V \brack 1}$, which is trivial if d(v) = 0. Furthermore, if $d(v) \neq 0$, then equality is possible only when $d(v) = {y-1 \brack k-1}$. To see this, suppose first that $d(v) \geq {y-1 \brack k-1}$. Then by the first condition and Lemma 82, it suffices to observe that $(1/q^k)({y \brack k} - d(v)) \leq ([y-k]/[k])d(v)$. On the other hand, if $d(v) \leq {y-1 \brack k-1}$, then define the real number $y_v \geq k$ by $d(v) = {y_v-1 \brack k-1}$. Since $d(v) = |L_U(v)|$, the second condition and the induction hypothesis imply that

$$|K_{k+1}^k(v)| \le |K_k^{k-1}(L_U(v))| \le \begin{bmatrix} y_v - 1 \\ k \end{bmatrix} = \frac{[y_v - k]}{[k]} d(v) \le \frac{[y - k]}{[k]} d(v).$$

The equality conditions are clear so the claim holds in either case. Now

$$\begin{split} [k+1]|K_{k+1}^{k}(\mathcal{F})| &= \sum_{v \in {V \brack 1}} |K_{k+1}^{k}(v)| \le \frac{|y-k|}{[k]} \sum_{v \in {V \brack 1}} d(v) = \frac{|y-k]}{[k]} [k]|\mathcal{F}| \\ &= [y-k] {y \brack k} = [k+1] {y \brack k+1}. \end{split}$$

Therefore, $|K_{k+1}^k(\mathcal{F})| \leq {y \choose k+1}$, and equality holds only when all one-dimensional subspaces v with non-zero degree satisfy $d(v) = {y-1 \choose k-1}$.

We now characterize the case of equality. Again the proof proceeds by induction on k. The base case k = 1 is easy: Suppose $\mathcal{F} \subset \begin{bmatrix} V \\ 1 \end{bmatrix}$, $|\mathcal{F}| = [y]$, and $|K_2^1(\mathcal{F})| = \begin{bmatrix} y \\ 2 \end{bmatrix}$. Then (6) implies that $|K_2^1(v)| = (1/q)([y] - 1)$ for all $v \in \mathcal{F}$. Hence, if v, w are distinct elements of \mathcal{F} , then every one-dimensional space in the two-dimensional space spanned by v and w lies in \mathcal{F} . It is easy to see by induction that if A is a subspace of dimension $1 \leq d < \lceil y \rceil$ such that $\begin{bmatrix} A \\ 1 \end{bmatrix} \subset \mathcal{F}$, then there exists a subspace B of dimension d+1 that contains A and for which $\begin{bmatrix} B \\ 1 \end{bmatrix} \subset \mathcal{F}$. In particular, this proves that $y \in \mathbb{Z}^+$ and $\mathcal{F} = \begin{bmatrix} Y \\ 1 \end{bmatrix}$ for some y-dimensional subspace Y.

Now suppose $\mathcal{F} \subset \begin{bmatrix} V \\ k \end{bmatrix}$, $|\mathcal{F}| = \begin{bmatrix} y \\ k \end{bmatrix}$, and $|K_{k+1}^k(\mathcal{F})| = \begin{bmatrix} y \\ k+1 \end{bmatrix}$. Choose $v \in \begin{bmatrix} V \\ 1 \end{bmatrix}$ for which $d(v) \neq 0$. Since $|K_{k+1}^k(\mathcal{F})| = \begin{bmatrix} y \\ k+1 \end{bmatrix}$, we have $d(v) = \begin{bmatrix} y^{-1} \\ k^{-1} \end{bmatrix}$ and $|K_{k+1}^k(v)| = \begin{bmatrix} y^{-1} \\ k \end{bmatrix}$. Let U be an (n-1)-dimensional subspace not containing v. We have $|L_U(v)| = d(v) = \begin{bmatrix} y^{-1} \\ k^{-1} \end{bmatrix}$ so

$$\begin{bmatrix} y-1\\ k \end{bmatrix} = |K_{k+1}^k(v)| \le |K_k^{k-1}(L_U(v))| \le \begin{bmatrix} y-1\\ k \end{bmatrix},$$

which implies that $|K_k^{k-1}(L_U(v))| = {y-1 \brack k}$. By the induction hypothesis, $L_U(v) = {W \brack k-1}$ for some (y-1)-dimensional space W, which implies $y \in \mathbb{Z}^+$. Moreover, for every k-dimensional subspace A in $K_k^{k-1}(L_U(v)) = {W \brack k}$, we have $A \lor v \in K_{k+1}^k(v)$. Hence all k-dimensional subspaces in $Y := W \lor v$ lie in \mathcal{F} . Since $|\mathcal{F}| = {y \brack k}$ and dim(Y) = y, we must have $\mathcal{F} = {Y \brack k}$.

As an application of Theorem 76, we obtain Theorem 78, the q-analog of Theorem 77. We will first prove the bound in Theorem 78 and characterize equality when (r - 1)n > rk. We finish this section by the characterization of equality for the case (r - 1)n = rk.

The proof proceeds by induction on $(r-1)n - rk \in \mathbb{N}$. For the base case (r-1)n - rk = 0, we generalize Greene and Kleitman's argument in [57]. A family S of t-dimensional subspaces of V is called a t-spread if every one-dimensional subspace of V is contained in exactly one t-dimensional subspace in S. If the elements in S that lie in a subspace U form a t-spread of U then we say that S induces a t-spread on U. A t-spread S is called geometric if S induces a t-spread on each 2t-dimensional subspace generated by a pair of elements in S. It is well-known [6] that V possesses a geometric t-spread if and only if t|n. In the base case (r-1)n - rk = 0, we have n = r(n-k) so V has a geometric (n-k)-spread. The following facts concerning geometric t-spreads are easy to establish.

Lemma 84. If S is a geometric t-spread of V, then S induces a geometric t-spread on any subspace of V that is generated by elements of S.

Lemma 85. If S is a geometric t-spread of V, then for any isomorphism $\pi \in GL(V)$, the family $\pi(S) := {\pi(S) : S \in S}$ is also a geometric t-spread of V.

Suppose $r, n, k \in \mathbb{Z}^+$ satisfy (r-1)n - rk = 0 and let $\mathcal{F} \subset \begin{bmatrix} V \\ k \end{bmatrix}$ be an *r*-wise intersecting family. Endow V with the usual inner product, and consider the family

$$\mathcal{F}^{\perp} := \{ F^{\perp} : F \in \mathcal{F} \} \subset \begin{bmatrix} V \\ n-k \end{bmatrix}.$$

Let \mathcal{B} be a geometric (n-k)-spread of V. We want to determine the maximum number of elements of \mathcal{B} that lie in \mathcal{F}^{\perp} . Since \mathcal{F} is r-wise intersecting, we have that \mathcal{F}^{\perp} is r-wise co-intersecting; that is, any r elements of \mathcal{F}^{\perp} are contained in a common (n-1)-dimensional space. If r = 2 and n = 2k, the family \mathcal{F}^{\perp} is both intersecting and co-intersecting; hence only one element of the spread \mathcal{B} can lie in \mathcal{F}^{\perp} in this case. Lemma 86 determines the maximum number of elements of \mathcal{B} that lie in \mathcal{F}^{\perp} whenever $r, n, k \in \mathbb{Z}^+$ satisfy (r-1)n - rk = 0.

Lemma 86. Let $r, n, k \in \mathbb{Z}^+$ satisfy (r-1)n - rk = 0. Let \mathcal{B} be a geometric (n-k)-spread of V. If $\mathcal{B}' \subset \mathcal{B}$ is a r-wise co-intersecting subfamily, then

$$|\mathcal{B}'| \le \frac{q^{(r-1)(n-k)} - 1}{q^{n-k} - 1}$$

If equality holds, \mathcal{B}' is a (n-k)-spread of a (r-1)(n-k)-dimensional space.

Proof. Let B_1, \ldots, B_m be a maximum subfamily of \mathcal{B}' such that dim $(\bigvee_{i=1}^m B_i) = m(n-k)$. Hence, if $B \in \mathcal{B}'$ then $B \cap \bigvee_{i=1}^m B_i \neq \{0\}$. Since \mathcal{B} is geometric, \mathcal{B} induces a spread on $\bigvee_{i=1}^m B_i$ by Lemma 84. As $B \cap \bigvee_{i=1}^m B_i \neq \{0\}$ for every B in \mathcal{B}' , all elements in \mathcal{B}' lie in $\bigvee_{i=1}^m B_i$. Since \mathcal{B}' is r-wise co-intersecting, we must have $m \leq r-1$. Therefore,

$$|\mathcal{B}'| \le \frac{q^{(r-1)(n-k)} - 1}{q^{n-k} - 1}$$

which is the number of elements in a (n-k)-spread of a (r-1)(n-k)-dimensional space. Also, if equality holds, \mathcal{B}' is a (n-k)-spread of a (r-1)(n-k)-dimensional space.

Now we prove the base case of Theorem 78; the case r = 2 of Lemma 87 is a result of Greene and Kleitman [57].

Lemma 87. Suppose $r, n, k \in \mathbb{Z}^+$ satisfy (r-1)n - rk = 0. If $\mathcal{F} \subset \begin{bmatrix} V \\ k \end{bmatrix}$ is r-wise intersecting, then $|\mathcal{F}| \leq \begin{bmatrix} n-1 \\ k-1 \end{bmatrix}$.

Proof. Let \mathcal{B} be a geometric (n-k)-spread of V and let $\pi \in GL(V)$ be an isomorphism. By Lemma 85, the spread $\pi(\mathcal{B})$ is also geometric. Consider the family $\mathcal{F}^{\perp} \subset \begin{bmatrix} V \\ n-k \end{bmatrix}$. Since \mathcal{F} is *r*-wise intersecting, \mathcal{F}^{\perp} is *r*-wise co-intersecting. By Lemma 86,

$$|\mathcal{F}^{\perp} \cap \pi(\mathcal{B})| \le \frac{q^{(r-1)(n-k)} - 1}{q^{n-k} - 1} = \frac{q^k - 1}{q^{n-k} - 1}$$
(7)

because $\mathcal{F}^{\perp} \cap \pi(\mathcal{B})$ is a *r*-wise co-intersecting subfamily of $\pi(\mathcal{B})$ and because we have k = (r-1)(n-k) when r, n, k satisfy (r-1)n - rk = 0.
As $|GL(V)| = q^{n(n-1)/2}(q-1)^n [n]!$, we have

$$\sum_{\pi \in GL(V)} |\mathcal{F}^{\perp} \cap \pi(\mathcal{B})| \le \frac{q^{\kappa} - 1}{q^{n-k} - 1} \cdot q^{n(n-1)/2} (q-1)^n [n]!.$$

Now, given $F^{\perp} \in \mathcal{F}^{\perp}$ and $B \in \mathcal{B}$ there are $q^{n(n-1)/2}(q-1)^n[n-k]![k]!$ isomorphisms $\pi \in GL(V)$ such that $\pi(B) = F^{\perp}$. Consequently,

$$\begin{aligned} & \left(\frac{q^n - 1}{q^{n-k} - 1}\right) |\mathcal{F}^{\perp}| \; q^{n(n-1)/2} (q-1)^n [n-k]! [k]! \\ = & |\mathcal{B}| \; |\mathcal{F}^{\perp}| \; |\{\pi \in GL(V) : \pi(B) = F^{\perp}\}| \\ = & \sum_{\pi \in GL(V)} |\mathcal{F}^{\perp} \cap \pi(\mathcal{B})| \\ \leq & \frac{q^k - 1}{q^{n-k} - 1} \; \cdot \; q^{n(n-1)/2} (q-1)^n [n]!. \end{aligned}$$

Since $|\mathcal{F}| = |\mathcal{F}^{\perp}|$, we have

$$\begin{aligned} |\mathcal{F}| &\leq \left(\frac{q^{n(n-1)/2}(q-1)^n[n]!}{q^{n(n-1)/2}(q-1)^n[n-k]![k]!}\right) \left(\frac{q^{n-k}-1}{q^n-1}\right) \left(\frac{q^k-1}{q^{n-k}-1}\right) \\ &= \left[\binom{n-1}{k-1}\right]. \end{aligned}$$

With Lemma 87 in hand, we are ready to prove Theorem 78.

Proof of Theorem 78. The proof proceeds by induction on $(r-1)n - rk \in \mathbb{N}$. The base case (r-1)n - rk = 0 was proved in Lemma 87. Suppose Theorem 78 holds when r, n, ksatisfy (r-1)n - rk = p for $p \ge 0$. We will prove Theorem 78 holds when r, n, k satisfy (r-1)n - rk = p + 1. Let $\mathcal{F} \subset \begin{bmatrix} V \\ k \end{bmatrix}$ be a maximum size r-wise intersecting family. Now the family $\mathcal{P} := \{P \in \begin{bmatrix} V \\ k \end{bmatrix} : v \subset P\}$, where $v \subset V$ is some one-dimensional subspace, is r-wise intersecting so $|\mathcal{F}| \ge |\mathcal{P}| = \begin{bmatrix} n-1 \\ k-1 \end{bmatrix}$. Let W be an (n + 1)-dimensional space over \mathbf{F}_q that contains V. Define the family

$$\mathcal{A} := \left\{ A \in \begin{bmatrix} W \\ k+1 \end{bmatrix} : \exists \ F \in \mathcal{F} \text{ with } F \subset A
ight\}$$

to be the family of all (k + 1)-dimensional spaces in W that contain some $F \in \mathcal{F}$. We will partition \mathcal{A} into the following subfamilies:

$$\mathcal{A}_1 := \{ A \in \mathcal{A} : A \not\subset V \}, \quad \mathcal{A}_2 := \mathcal{A} \setminus \mathcal{A}_1.$$

First let us compute the size of \mathcal{A}_1 . Observe that if $A \in \begin{bmatrix} W \\ k+1 \end{bmatrix}$ and A does not lie in V, then A intersects V in exactly a k-dimensional space. Therefore, A cannot

contain two distinct k-dimensional spaces in \mathcal{F} . Any $F \in \mathcal{F}$ can be extended to a (k+1)-dimensional space in \mathcal{A}_1 in q^{n-k} ways. Therefore, $|\mathcal{A}_1| = q^{n-k}|\mathcal{F}| \ge q^{n-k} {n-1 \brack k-1}$.

Now we will compute the size of \mathcal{A}_2 . Observe that, by duality, we have $F \subset A \in \mathcal{A}_2$ for some $F \in \mathcal{F}$ if and only if $F^{\perp} \supset A^{\perp} \in \begin{bmatrix} V \\ n-k-1 \end{bmatrix}$. Therefore, $|\mathcal{A}_2| = |\partial \mathcal{F}^{\perp}|$. Since

$$|\mathcal{F}^{\perp}| = |\mathcal{F}| \ge {\binom{n-1}{k-1}} = {\binom{n-1}{n-k}},\tag{8}$$

by applying Theorem 76 we obtain

$$|\mathcal{A}_2| = \left|\partial \mathcal{F}^{\perp}\right| \ge {n-1 \choose n-k-1} = {n-1 \choose k}.$$
(9)

As $\mathcal{A} = \mathcal{A}_1 \dot{\cup} \mathcal{A}_2$, we have by Lemma 82 that

$$|\mathcal{A}| = |\mathcal{A}_1| + |\mathcal{A}_2| \ge q^{n-k} {n-1 \brack k-1} + {n-1 \brack k} = {n \brack k}.$$
 (10)

Now \mathcal{F} is *r*-wise intersecting so \mathcal{A} is an *r*-wise intersecting family of (k+1)-dimensional spaces in W. Observe that r, n+1, k+1 satisfy

$$(r-1)(n+1) - r(k+1) = (r-1)n - rk - 1 = (p+1) - 1 = p.$$

By the induction hypothesis $|\mathcal{A}| \leq {n \brack k}$, which implies equality everywhere in (8), (9), and (10). As a result, $q^{n-k}|\mathcal{F}| = |\mathcal{A}_1| = q^{n-k} {n-1 \brack k-1}$, which implies $|\mathcal{F}| = {n-1 \brack k-1}$. Moreover, $|\mathcal{F}^{\perp}| = {n-1 \brack n-k}$ and $|\partial \mathcal{F}^{\perp}| = |\mathcal{A}_2| = {n-k \brack n-k-1}$. Therefore \mathcal{F}^{\perp} satisfies equality in Theorem 76, which implies that $\mathcal{F}^{\perp} = {Y \brack n-k}$ for some (n-1)-dimensional subspace $Y \subset V$. By duality, $\mathcal{F} = \{F \in {V \brack k} : v \subset F\}$ for some one-dimensional subspace $v \subset V$.

All what is left to do is characterizing equality in the base case of Theorem 78 when (r-1)n - rk = 0. Godsil and Newman [56, 92] recently characterized equality in the Erdős-Ko-Rado theorem for vector spaces using the methods of [46]. Recall that the Erdős-Ko-Rado theorem for vector spaces is the case r = 2 of Theorem 78. In particular, they obtained the following result.

Theorem 88 (Godsil and Newman). If n = 2k and $\mathcal{F} \subset \begin{bmatrix} V \\ k \end{bmatrix}$ is a maximum size intersecting family, then $\mathcal{F} = \{F \in \begin{bmatrix} V \\ k \end{bmatrix} : v \subset F\}$ for some one-dimensional subspace $v \subset V$ or $\mathcal{F} = \begin{bmatrix} U \\ k \end{bmatrix}$ for some (2k-1)-dimensional subspace $U \subset V$.

We use their result to characterize equality in Theorem 78 when (r-1)n - rk = 0and $r \ge 3$. The proof proceeds by induction on r; the base case r = 2 and n = 2kis Theorem 88. Let $\mathcal{F} \subset \begin{bmatrix} V \\ k \end{bmatrix}$ be a maximum size r-wise intersecting family. It will be more natural to state results in terms of $\mathcal{F}^{\perp} \subset \begin{bmatrix} V \\ n-k \end{bmatrix}$ so we make the following simple observation.

Lemma 89. We have $\mathcal{F} \subset \begin{bmatrix} V \\ k \end{bmatrix}$ is a maximum size r-wise intersecting family if and only if $\mathcal{F}^{\perp} \subset \begin{bmatrix} V \\ n-k \end{bmatrix}$ is a maximum size r-wise co-intersecting family, i.e. for any $G_1, G_2, \ldots, G_r \in \mathcal{F}^{\perp}$ we have $G_1 \lor G_2 \lor \cdots \lor G_r \neq V$. Lemma 92 will allow us to use induction. We first state two simple corollaries of Lemma 87 that will be used in the proof of Lemma 92. Recall that V is r(n - k)dimensional since r, n, k satisfy (r - 1)n - rk = 0.

Corollary 90. Suppose r, n, k satisfy (r-1)n - rk = 0. Let $\mathcal{F} \subset \begin{bmatrix} V \\ k \end{bmatrix}$ be r-wise intersecting. If there is a geometric (n-k)-spread \mathcal{B} of V such that equality holds in (7) for all $\pi \in GL(V)$, then \mathcal{F} has maximum size.

Corollary 91. Suppose r, n, k satisfy (r-1)n - rk = 0. If $\mathcal{F} \subset \begin{bmatrix} V \\ k \end{bmatrix}$ is a maximum size r-wise intersecting family, then equality holds in (7) for every geometric (n-k)-spread \mathcal{B} of V and for every $\pi \in GL(V)$.

Lemma 92. Let $\mathcal{F} \subset \begin{bmatrix} V \\ k \end{bmatrix}$ be a maximum size r-wise intersecting family. Fix F^{\perp} in \mathcal{F}^{\perp} and let $U \subset V$ be an (r-1)(n-k)-dimensional space that intersects F^{\perp} trivially; that is $F^{\perp} \cap U = \{0\}$. Then

$$\mathcal{F}^{\perp}|_U := \{ E \in \mathcal{F}^{\perp} : E \subset U \}$$

is a maximum size (r-1)-wise co-intersecting family in $\begin{bmatrix} U\\ n-k \end{bmatrix}$.

Proof. Let S be a geometric (n - k)-spread of V. Choose S_1, \ldots, S_r in S such that $\bigvee_{i=1}^r S_i = V$. Since $F^{\perp} \cap U = \{0\}$, there exists an isomorphism $\rho \in GL(V)$ such that $\rho(S_1) = F^{\perp}$ and $\rho(\bigvee_{i=2}^r S_i) = U$. The (n-k)-spread $\mathcal{B} := \rho(S)$ is geometric by Lemma 85, and $F^{\perp} \in \mathcal{B}$; moreover $U = \bigvee_{i=2}^r \rho(S_i)$ so \mathcal{B} induces a geometric (n-k)-spread \mathcal{B}' on U by Lemma 84.

Observe that $\mathcal{F}^{\perp}|_U$ is (r-1)-wise co-intersecting since $F^{\perp} \cap U = \{0\}$. To prove that $\mathcal{F}^{\perp}|_U \subset \begin{bmatrix} U\\ n-k \end{bmatrix}$ is a maximum size (r-1)-wise co-intersecting family, we will apply Lemma 89 and Corollary 90. That is, we will show that if $\alpha \in GL(U)$ then equality holds in (7):

$$\left|\mathcal{F}^{\perp}|_{U} \cap \alpha(\mathcal{B}')\right| = \frac{q^{(r-2)(n-k)} - 1}{q^{n-k} - 1}.$$

Let $\pi \in GL(V)$ be an isomorphism such that $\pi(F^{\perp}) = F^{\perp}$, $\pi(U) = U$, and $\pi|_U = \alpha$. Since \mathcal{F}^{\perp} is a maximum size *r*-wise co-intersecting family, $\mathcal{F}^{\perp} \cap \pi(\mathcal{B})$ is a (n-k)-spread of a (r-1)(n-k)-dimensional space W_{π} by Lemma 86 and Corollary 91. Consider the subspace $W_{\pi} \cap U$ and observe that $\dim(W_{\pi} \cap U) = (r-2)(n-k)$ since F^{\perp} is contained in W_{π} and intersects U trivially.

The spread $\pi(\mathcal{B})$ induces the spread $\mathcal{F}^{\perp} \cap \pi(\mathcal{B})$ on W_{π} and induces the spread $\alpha(\mathcal{B}')$ on U. Consider the elements of $\alpha(\mathcal{B}')$ that intersect $W_{\pi} \cap U$ non-trivially; as these elements are in $\pi(\mathcal{B})$ and intersect W_{π} , they must lie in W_{π} and hence in $W_{\pi} \cap U$. Hence, the elements of $\alpha(\mathcal{B}')$ that intersect $W_{\pi} \cap U$ non-trivially form a spread of $W_{\pi} \cap U$. Moreover, these elements lie in $\mathcal{F}^{\perp} \cap \pi(\mathcal{B})$ so

$$\mathcal{F}^{\perp}|_U \cap \alpha(\mathcal{B}') = (\mathcal{F}^{\perp} \cap \pi(\mathcal{B})) \cap \alpha(\mathcal{B}')$$

is the spread $\pi(\mathcal{B})$ induces on $W_{\pi} \cap U$. Since $W_{\pi} \cap U$ is (r-2)(n-k)-dimensional, $|\mathcal{F}^{\perp}|_U \cap \alpha(\mathcal{B}')|$ satisfies (7) with equality. By Lemma 89 and Corollary 90, $\mathcal{F}^{\perp}|_U$ is a maximum size (r-1)-wise co-intersecting family in $\begin{bmatrix} U\\ n-k \end{bmatrix}$.

Characterizing Equality in Theorem 78 when (r-1)n - rk = 0 and $r \ge 3$: The proof proceeds by induction on r; the base case r = 2 and n = 2k is Theorem 88.

Let $r \geq 3$ and suppose the statement is proved for any $2 \leq r' < r$. Let $\mathcal{F} \subset \begin{bmatrix} V \\ k \end{bmatrix}$ be a maximum size *r*-wise intersecting family and observe that $\mathcal{F}^{\perp} \subset \begin{bmatrix} V \\ n-k \end{bmatrix}$ is a maximum size *r*-wise co-intersecting family. We will show that $\mathcal{F}^{\perp} = \begin{bmatrix} H \\ n-k \end{bmatrix}$ where *H* is a (n-1)dimensional space of *V*. By duality, this implies that $\mathcal{F} = \{F \in \begin{bmatrix} V \\ k \end{bmatrix} : v \subset F\}$ for some one-dimensional subspace $v \subset V$, which is the desired conclusion.

Fix some $F^{\perp} \in \mathcal{F}^{\perp}$. By Lemma 92, if U is a (r-1)(n-k)-dimensional subspace that intersects F^{\perp} trivially, then $\mathcal{F}^{\perp}|_{U}$ is a maximum size (r-1)-wise co-intersecting family in $\begin{bmatrix} U\\n-k \end{bmatrix}$. When r = 3, then dim U = 2(n-k) and $\mathcal{F}^{\perp}|_{U}$ is a maximum size intersecting and co-intersecting family in $\begin{bmatrix} U\\n-k \end{bmatrix}$; hence by Theorem 88

- 1. $\mathcal{F}^{\perp}|_{U} = \{E \in \begin{bmatrix} U \\ n-k \end{bmatrix} : u \subset E\}$ for some one-dimensional subspace $u \subset U$ or
- 2. $\mathcal{F}^{\perp}|_{U} = \begin{bmatrix} U'\\ n-k \end{bmatrix}$ for some (2(n-k)-1)-dimensional subspace $U' \subset U$.

If r > 3 then, by the induction hypothesis and duality, $\mathcal{F}^{\perp}|_{U} = \begin{bmatrix} U'\\ n-k \end{bmatrix}$, where $U' \subset U$ is some ((r-1)(n-k)-1)-dimensional subspace.

Our first task is to eliminate the possibility that $\mathcal{F}^{\perp}|_{U} = \{E \in \begin{bmatrix} U \\ n-k \end{bmatrix} : u \subset E\}$ for some one-dimensional subspace $u \subset U$ in the case r = 3. We now show that if $\mathcal{F}^{\perp}|_{U} = \{E \in \begin{bmatrix} U \\ n-k \end{bmatrix} : u \subset E\}$ for some one-dimensional subspace $u \subset U$, then every element of \mathcal{F}^{\perp} must intersect $F^{\perp} \lor u$ non-trivially.

Claim 93. If $\mathcal{F}^{\perp}|_{U} = \{E \in \begin{bmatrix} U \\ n-k \end{bmatrix} : u \subset E\}$ for some one-dimensional subspace $u \subset U$, then for all $G \in \mathcal{F}^{\perp}$ we have $G \cap (F^{\perp} \lor u) \neq \{0\}$.

Proof. Suppose, for a contradiction, that there exists $G \in \mathcal{F}^{\perp}$ such that G intersects $F^{\perp} \lor u$ trivially. We have $\dim((F^{\perp} \lor G) \cap U) = n - k$ because F^{\perp} intersects both G and U trivially. Since u does not lie in $F^{\perp} \lor G$ and $\mathcal{F}^{\perp}|_U = \{E \in \begin{bmatrix} U \\ n-k \end{bmatrix} : u \subset E\}$, we can find $E' \in \mathcal{F}^{\perp}|_U$ that intersects $F^{\perp} \lor G$ trivially. Hence $F^{\perp} \lor G \lor E' = V$, which contradicts the fact that \mathcal{F}^{\perp} is 3-wise co-intersecting.

We now show that if $\mathcal{F}^{\perp}|_{U} = \{E \in \begin{bmatrix} U\\ n-k \end{bmatrix} : u \subset E\}$ for some one-dimensional subspace $u \subset U$, then any (n-k)-dimensional space that meets F^{\perp} trivially but meets $F^{\perp} \lor u$ non-trivially must lie in \mathcal{F}^{\perp} .

Claim 94. Suppose $\mathcal{F}^{\perp}|_U = \{E \in \begin{bmatrix} U \\ n-k \end{bmatrix} : u \subset E\}$ for some one-dimensional subspace $u \subset U$. If $G \in \begin{bmatrix} V \\ n-k \end{bmatrix}$, $G \cap F^{\perp} = \{0\}$, and $G \cap (F^{\perp} \lor u) \neq \{0\}$, then $G \in \mathcal{F}^{\perp}$.

Proof. There exists a geometric (n - k)-spread \mathcal{B} of V that contains both G and F^{\perp} because $G \cap F^{\perp} = \{0\}$. Since \mathcal{B} is a spread, all subspaces in $(\mathcal{F}^{\perp} \cap \mathcal{B}) \setminus \{F^{\perp}\}$ meet $F^{\perp} \lor u$ in a one-dimensional subspace that does not lie in F^{\perp} by Claim 93. Lemma 86 and Corollary 91 imply that $\mathcal{F}^{\perp} \cap \mathcal{B}$ is a spread of a 2(n - k)-dimensional space so $|(\mathcal{F}^{\perp} \cap \mathcal{B}) \setminus \{F^{\perp}\}| = q^{n-k}$. There are q^{n-k} one-dimensional subspaces in $F^{\perp} \lor u$ that do not lie in F^{\perp} . Hence, each one-dimensional subspace in $(F^{\perp} \lor u) \setminus F^{\perp}$ meets a unique

subspace in $(\mathcal{F}^{\perp} \cap \mathcal{B}) \setminus \{F^{\perp}\}$. Since G meets $F^{\perp} \vee u$ in a one-dimensional subspace that does not lie in F^{\perp} and $G \in \mathcal{B}$, we must have $G \in \mathcal{F}^{\perp} \cap \mathcal{B} \subset \mathcal{F}^{\perp}$.

We now eliminate the possibility that $\mathcal{F}^{\perp}|_{U} = \{E \in \begin{bmatrix} U\\ n-k \end{bmatrix} : u \subset E\}$ for some onedimensional subspace $u \subset U$. We will construct three (n-k)-dimensional subspaces that together span V, and intersect $F^{\perp} \vee u$ in a one-dimensional subspace not lying in F^{\perp} . By Claim 94, these three spaces lie in \mathcal{F}^{\perp} , which contradicts \mathcal{F}^{\perp} being 3-wise co-intersecting. To build these three subspaces, we first choose three one-dimensional subspaces v_1^1, v_2^1, v_3^1 in $(F^{\perp} \vee u) \setminus F^{\perp}$ such that $v_3^1 \not\subset v_1^1 \vee v_2^1$. These one-dimensional subspaces exist because $\dim(F^{\perp} \vee u) = (n-k) + 1 \geq 3$ so, after picking v_1^1 and v_2^1 , any one-dimensional subspace of $F^{\perp} \vee u$ not in $F^{\perp} \cup (v_1^1 \vee v_2^1)$ will do. Since the number of one-dimensional subspaces in $(F^{\perp} \vee u) \setminus (F^{\perp} \cup (v_1^1 \vee v_2^1))$ is $q^{n-k} - q > 0$, we can indeed choose v_3^1 .

We construct a family of one-dimensional subspaces

$$\{v_i^j : i \in \{1, 2, 3\}, j \in \{1, \dots, n-k\}\}$$

such that, for each $i \in \{1, 2, 3\}$, the subspace $V_i = \bigvee_{j=1}^{n-k} v_i^j$ intersects $F^{\perp} \lor u$ in the one-dimensional subspace $v_i^1 \not\subset F^{\perp}$ and $\bigvee_{i=1}^3 V_i = V$. The subspaces V_1, V_2, V_3 are the desired three (n-k)-dimensional subspaces. We pick the one-dimensional subspaces one after the other; we have to show that at each step there is a possible one-dimensional subspace to pick. When picking the last one-dimensional subspace v_3^{n-k} we must choose a one-dimensional subspace from V that is not in $V_1 \lor V_2 \lor \bigvee_{j=1}^{n-k-1} v_j^j$ nor in $F^{\perp} \lor \bigvee_{j=1}^{n-k-1} v_3^j$. By inclusion-exclusion, there are $q^{3(n-k)-1} - q^{2(n-k)-2} > 0$ one-dimensional subspaces in V that do not lie in $V_1 \lor V_2 \lor \bigvee_{j=1}^{n-k-1} v_3^j$ nor in $F^{\perp} \lor \bigvee_{j=1}^{n-k-1} v_3^j$; thus it is indeed possible to construct the desired three (n-k)-dimensional subspaces. Therefore, we have eliminated the possibility that $\mathcal{F}^{\perp}|_U = \{E \in \begin{bmatrix} U \\ n-k \end{bmatrix} : u \subset E\}$ for some one-dimensional subspace $u \subset U$ in the case r = 3.

We may now assume that $r \geq 3$ and that if U is a (r-1)(n-k)-dimensional space that intersects F^{\perp} trivially then $\mathcal{F}^{\perp}|_{U} = \begin{bmatrix} U'\\n-k \end{bmatrix}$ for some ((r-1)(n-k)-1)-dimensional subspace $U' \subset U$. Our ultimate goal is to prove that $\mathcal{F}^{\perp} = \begin{bmatrix} F^{\perp} \lor U'\\n-k \end{bmatrix}$. Naturally, we first show that if U_1, U_2 are two (r-1)(n-k)-dimensional subspaces that intersect F^{\perp} trivially, then $F^{\perp} \lor U'_1 = F^{\perp} \lor U'_2$.

Claim 95. Let U_1, U_2 be two (r-1)(n-k)-dimensional subspaces of V that intersect F^{\perp} trivially. Let U'_1, U'_2 be the ((r-1)(n-k)-1)-dimensional subspaces of U_1 and U_2 such that $\mathcal{F}^{\perp}|_{U_1} = \begin{bmatrix} U'_1 \\ n-k \end{bmatrix}$ and $\mathcal{F}^{\perp}|_{U_2} = \begin{bmatrix} U'_2 \\ n-k \end{bmatrix}$. Then $F^{\perp} \vee U'_1 = F^{\perp} \vee U'_2$.

Proof. Suppose, for a contradiction, that $F^{\perp} \vee U'_1 \neq F^{\perp} \vee U'_2$. Choose subspaces W_1, \ldots, W_{r-2} in $\begin{bmatrix} U'_1 \\ n-k \end{bmatrix}$ such that W_1 is not contained in $F^{\perp} \vee U'_2$ and dim $\left(\bigvee_{i=1}^{r-2} W_i\right) = (r-2)(n-k)$.

The subspace $F^{\perp} \vee \bigvee_{i=1}^{r-2} W_i$ is (r-1)(n-k)-dimensional because U_1 intersects F^{\perp} trivially. The subspace U'_2 is ((r-1)(n-k)-1)-dimensional and intersects F^{\perp} trivially so

$$(r-2)(n-k) - 1 \le \dim\left(U_2' \cap \left(F^{\perp} \lor \bigvee_{i=1}^{r-2} W_i\right)\right) \le (r-2)(n-k).$$

Suppose that dim $\left(U_2' \cap \left(F^{\perp} \vee \bigvee_{i=1}^{r-2} W_i\right)\right) = (r-2)(n-k)$ for a contradiction. By definition of W_1 , we can choose a one-dimensional subspace $w \subset W_1$ that does not lie in $F^{\perp} \vee U_2'$. The subspace $F^{\perp} \vee w$ is (n-k+1)-dimensional. The subspace $F^{\perp} \vee \bigvee_{i=1}^{r-2} W_i$ is (r-1)(n-k)-dimensional and contains $F^{\perp} \vee w$. If dim $\left(U_2' \cap \left(F^{\perp} \vee \bigvee_{i=1}^{r-2} W_i\right)\right) = (r-2)(n-k)$, then $F^{\perp} \vee W_2'$ by construction. Therefore, dim $\left(U_2' \cap \left(F^{\perp} \vee \bigvee_{i=1}^{r-2} W_i\right)\right) = (r-2)(n-k) - 1$.

Since U'_2 is ((r-1)(n-k)-1)-dimensional, this implies that there exists a subspace Z in $\begin{bmatrix} U'_2 \\ n-k \end{bmatrix}$ that intersects $F^{\perp} \vee \bigvee_{i=1}^{r-2} W_i$ trivially. Now $F^{\perp}, W_1, \ldots, W_{r-2}, Z$ lie in \mathcal{F}^{\perp} since $\mathcal{F}^{\perp}|_{U_1} = \begin{bmatrix} U'_1 \\ n-k \end{bmatrix}$ and $\mathcal{F}^{\perp}|_{U_2} = \begin{bmatrix} U'_2 \\ n-k \end{bmatrix}$. By construction, $F^{\perp} \vee \bigvee_{i=1}^{r-2} W_i \vee Z = V$, which contradicts \mathcal{F}^{\perp} being *r*-wise co-intersecting. This proves $F^{\perp} \vee U'_1 = F^{\perp} \vee U'_2$.

Now we show that any (n-k)-dimensional subspace in $F^{\perp} \vee U'$ that intersects F^{\perp} trivially must lie in \mathcal{F}^{\perp} .

Claim 96. If $G \in \begin{bmatrix} F^{\perp} \lor U' \\ n-k \end{bmatrix}$ and $G \cap F^{\perp} = \{0\}$, then $G \in \mathcal{F}^{\perp}$.

Proof. Since $G \cap F^{\perp} = \{0\}$, there exists a (r-1)(n-k)-dimensional subspace U(G) that contains G and intersects F^{\perp} trivially. Let U(G)' be the ((r-1)(n-k)-1)-dimensional subspace of U(G) such that $\mathcal{F}^{\perp}|_{U(G)} = \begin{bmatrix} U(G)'\\ n-k \end{bmatrix}$. By Claim 95,

$$G \subset (F^{\perp} \lor U') \cap U(G) = (F^{\perp} \lor U(G)') \cap U(G) = U(G)'.$$

$$[U(G)'] \subset F^{\perp}$$

Hence $G \in \begin{bmatrix} U(G)' \\ n-k \end{bmatrix} \subset \mathcal{F}^{\perp}$.

Now we are ready to prove $\mathcal{F}^{\perp} = \begin{bmatrix} F^{\perp} \lor U' \\ n-k \end{bmatrix}$. Suppose, for a contradiction, that there exists a subspace $H \in \mathcal{F}^{\perp}$ that is not in $\begin{bmatrix} F^{\perp} \lor U' \\ n-k \end{bmatrix}$. We will construct r-1 subspaces in $\begin{bmatrix} F^{\perp} \lor U' \\ n-k \end{bmatrix}$ that each intersect F^{\perp} trivially and that together with H span V. By Claim 96, these r-1 subspaces lie in \mathcal{F}^{\perp} which contradicts \mathcal{F}^{\perp} being r-wise co-intersecting.

To build these r-1 subspaces, we construct a family of one-dimensional subspaces

$$\{v_i^j: i \in \{1, \dots, r-1\}, j \in \{1, \dots, n-k\}\}$$

such that for each $i \in \{1, \ldots, r-1\}$, the subspace $G_i = \bigvee_{j=1}^{n-k} v_i^j$ lies in $F^{\perp} \lor U'$, intersects F^{\perp} trivially, and $\bigvee_{i=1}^{r-1} G_i \lor H = V$. The subspaces G_1, \ldots, G_{r-1} are the desired r-1 subspaces. We pick the one-dimensional subspaces one after the other; we have to show that at each step there is a possible one-dimensional subspace to pick. When picking the last one-dimensional subspace v_{r-1}^{n-k} we must pick a one-dimensional subspace from $F^{\perp} \lor U'$ that is not in $H \lor \bigvee_{i=1}^{r-2} G_i \lor \bigvee_{j=1}^{n-k-1} v_{r-1}^j$ nor in $F^{\perp} \lor \bigvee_{j=1}^{n-k-1} v_{r-1}^j$. Since H is not contained in $F^{\perp} \lor U'$, we have

$$\dim\left(\left(F^{\perp} \vee U'\right) \cap \left(H \vee \bigvee_{i=1}^{r-2} G_i \vee \bigvee_{j=1}^{n-k-1} v_{r-1}^j\right)\right) = r(n-k) - 2.$$

Hence, there are at least

$$q^{r(n-k)-2} - (q^{2(n-k)-2} + q^{2(n-k)-3} + \dots + 1) > 0$$

one-dimensional subspaces of $F^{\perp} \vee U'$ that do not lie in $H \vee \bigvee_{i=1}^{r-2} G_i \vee \bigvee_{j=1}^{n-k-1} v_{r-1}^j$ nor in $F^{\perp} \vee \bigvee_{j=1}^{n-k-1} v_{r-1}^j$; thus it is indeed possible to construct the desired r-1 subspaces. This proves that $\mathcal{F}^{\perp} \subseteq {F^{\perp} \vee U' \choose n-k}$, and since $|\mathcal{F}^{\perp}| = {n-1 \choose k-1}$ we have $\mathcal{F}^{\perp} = {F^{\perp} \vee U' \choose n-k}$. The subspace $F^{\perp} \vee U'$ is (n-1)-dimensional; by duality, $\mathcal{F} = \{F \in {V \choose k} : v \subset F\}$ for some one-dimensional subspace $v \subset V$, which is the desired conclusion.

4.2 Intersecting families of subspaces and colorings of *q*-Kneser graphs

Intersection theorems on vector spaces have a long history. Let us start this subsection by stating the result of Frankl and Wilson that we will use several times in our proof of Theorem 80. It determines the largest size of t-intersecting families of subpaces. It was published a decade earlier than the corresponding set family result by Ahlswede and Khatchatrian [1].

Theorem 97. (Frankl, Wilson [46]) Let V be a vector space over $\mathbf{F}(q)$ of dimension n. For any t-intersecting family $\mathcal{F} \subseteq \begin{bmatrix} V \\ k \end{bmatrix}$ we have

$$|\mathcal{F}| \leq {n-t \brack k-t}$$
 if $2k \leq n$,

and

$$|\mathcal{F}| \leq {\binom{2k-t}{k}} \quad if \ 2k-t \leq n \leq 2k.$$

These bounds are best possible.

We start the proof of Theorem 80 by stating an easy technical lemma for q-binomial coefficients that will simplify our computations.

Lemma 98. Let $a \ge 0$ and $n \ge k \ge a + 1$ and $q \ge 2$. Then

$$\begin{bmatrix} k\\1\end{bmatrix} \begin{bmatrix} n-a-1\\k-a-1\end{bmatrix} < \frac{1}{(q-1)q^{n-2k}} \begin{bmatrix} n-a\\k-a\end{bmatrix}.$$

Proof. The inequality to be proved simplifies to

$$(q^{k-a}-1)(q^k-1)q^{n-2k} < q^{n-a}-1.$$

Lemma 99. Let $E \in {V \brack 1}$. If $E \leq L \leq V$, where L is an l-subspace, then the number of k-subspaces of V containing E and intersecting L is at least ${l \brack l} {n-2 \brack k-2} - q {l \brack l} {n-3 \brack k-3}$ (with equality for l = 2), and at most ${l \brack l} {n-2 \brack k-2}$.

Proof. The k-spaces containing E and intersecting L in a 1-dimensional space are counted exactly once in the first term. Those subspaces that intersect L in a 2-dimensional space are counted $\begin{bmatrix} 2\\1 \end{bmatrix} = q + 1$ times in the first term and -q times in the second term, thus once overall. If a subspace intersects L in a subspace of dimension $i \geq 3$, then it is counted $\begin{bmatrix} i\\1 \end{bmatrix}$ times in the first term and $-q \begin{bmatrix} i\\2 \end{bmatrix}$ times in the second term, thus a negative number of times overall.

Our next lemma gives bounds on the size of a HM-type family that are easier to work with than the precise formula mentioned in the introduction of Section 4.

Lemma 100. Let $n \ge 2k + 1$, $k \ge 3$ and $q \ge 2$. If \mathcal{F} is a HM-type family, then $(1 - \frac{1}{q^3 - q}) {k \brack 1} {n-2 \brack k-2} < {k \brack 1} {n-2 \brack k-2} - q {k \brack 2} {n-3 \brack k-3} \le |\mathcal{F}| \le {k \brack 1} {n-2 \brack k-2}.$

Proof. The first inequality follows immediately from Lemma 98 by noting that $q \begin{bmatrix} k \\ 2 \end{bmatrix} = {k \brack 1} {k \brack 1} - 1)/{(q+1)}$ and $n \ge 2k + 1$. Let \mathcal{F} be the HM-type family defined by the 1-space E and the k-space U. Then \mathcal{F} contains all k-subspaces of V containing E and

intersecting U, so the second inequality follows from Lemma 100. For the last inequality, Lemma 100 almost suffices, but we also have to count the k-subspaces of $\begin{bmatrix} E \lor U \\ k \end{bmatrix}$ that do not contain E. Each (k-1)-subspace W of U is contained in q+1 such subspaces, one of which is $E \lor W$. On the other hand, $E \lor W$ was counted at least q+1 times since $k \ge 3$. This proves the last inequality. \Box

For any $A \leq V$ and $\mathcal{F} \subseteq \begin{bmatrix} V \\ k \end{bmatrix}$, we write $\mathcal{F}_A = \{F \in \mathcal{F} : A \leq F\}$.

Lemma 101. If a subspace S does not intersect each element of \mathcal{F} , then there is a subspace T > S with dim $T = \dim S + 1$ and $|\mathcal{F}_T| \ge |\mathcal{F}_S| / {k \choose 1}$.

Proof. There is an $F \in \mathcal{F}$ such that $S \cap F = 0$. Average over all $T = S \vee E$ where E is a 1-subspace of F.

Lemma 102. If an s-dimensional subspace S does not intersect each element of \mathcal{F} , then $|\mathcal{F}_S| \leq {k \choose 1} {n-s-1 \choose k-s-1}$.

Proof. There is a
$$(s+1)$$
-space T with $\binom{n-s-1}{k-s-1} \ge |\mathcal{F}_T| \ge |\mathcal{F}_S|/\binom{k}{1}$.

Corollary 103. Let $\mathcal{F} \subseteq \begin{bmatrix} V \\ k \end{bmatrix}$ be an intersecting family with $\tau(\mathcal{F}) \geq s$. Then for any *i*-space $L \leq V$ with $i \leq s$ we have $|\mathcal{F}_L| \leq \begin{bmatrix} k \\ 1 \end{bmatrix}^{s-i} \begin{bmatrix} n-s \\ k-s \end{bmatrix}$.

Proof. If i = s, then clearly $|\mathcal{F}_L| \leq {n-s \choose k-s}$. If i < s, then there exists an $F \in \mathcal{F}$ such that $F \cap L = 0$; now apply Lemma 101 s - i times.

Before proving the q-analog of the theorem of Hilton-Milner we describe the essential part of maximal intersecting families with $\tau(\mathcal{F}) = 2$. Let us define \mathcal{T} to be the family of 2-spaces of V that intersect all subspaces in \mathcal{F} .

Proposition 104. Let \mathcal{F} be a maximal intersecting family with $\tau(\mathcal{F}) = 2$. Then \mathcal{F} contains all k-spaces containing an element of \mathcal{T} and we have one of the following three possibilities:

- (i) $|\mathcal{T}| = 1$ and $\binom{n-2}{k-2} < |\mathcal{F}| < \binom{n-2}{k-2} + (q+1)\left(\binom{k}{1} 1\right)\binom{k}{1}\binom{n-3}{k-3};$
- $\begin{array}{l} (ii) \ |\mathcal{T}| > 1, \ \tau(\mathcal{T}) = 1, \ and \ there \ is \ an \ (l+1) \text{-space} \ W \ (with \ 2 \le l \le k) \ and \ a \ 1\text{-space} \\ E \leqslant W \ so \ that \ \mathcal{T} = \{M : E \leqslant M \leqslant W, \ \dim M = 2\}. \ In \ this \ case \\ \begin{bmatrix} l \\ 1 \end{bmatrix} \begin{bmatrix} n-2 \\ k-2 \end{bmatrix} q \begin{bmatrix} l \\ 2 \end{bmatrix} \begin{bmatrix} n-3 \\ k-3 \end{bmatrix} \le |\mathcal{F}| \le \begin{bmatrix} l \\ 1 \end{bmatrix} \begin{bmatrix} n-2 \\ k-2 \end{bmatrix} + \begin{bmatrix} k \\ 1 \end{bmatrix} (\begin{bmatrix} k \\ 1 \end{bmatrix} \begin{bmatrix} l \\ 1 \end{bmatrix}) \begin{bmatrix} n-3 \\ k-3 \end{bmatrix} + q^l \begin{bmatrix} n-l \\ k-l \end{bmatrix}. \\ For \ l = 2 \ the \ upper \ bound \ here \ can \ be \ strengthened \ to \\ |\mathcal{F}| \le (q+1) \begin{bmatrix} n-2 \\ k-2 \end{bmatrix} q \begin{bmatrix} n-3 \\ k-3 \end{bmatrix} + \begin{bmatrix} k \\ 1 \end{bmatrix} (\begin{bmatrix} k \\ 1 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \end{bmatrix}) \begin{bmatrix} n-3 \\ k-3 \end{bmatrix} + q^2 \begin{bmatrix} k \\ 1 \end{bmatrix} \begin{bmatrix} n-3 \\ k-3 \end{bmatrix}; \end{array}$
- (iii) $\mathcal{T} = \begin{bmatrix} A \\ 2 \end{bmatrix}$ for some 3-subspace A and $\mathcal{F} = \{U \in \begin{bmatrix} V \\ k \end{bmatrix} : \dim(U \cap A) \ge 2\}$ and $|\mathcal{F}| = (q^2 + q + 1)(\begin{bmatrix} n-2 \\ k-2 \end{bmatrix} \begin{bmatrix} n-3 \\ k-3 \end{bmatrix}) + \begin{bmatrix} n-3 \\ k-3 \end{bmatrix}.$

In case (ii) there is a 1-space E and an l-space L such that \mathcal{F} contains the set $\mathcal{F}_{E,L}$ of all k-spaces containing E and intersecting L. The last two terms of the upper bound for $|\mathcal{F}|$ in (ii) give an upper bound on $|\mathcal{F} \setminus \mathcal{F}_{E,L}|$.

Proof. Let \mathcal{F} be a maximal intersecting family with $\tau(\mathcal{F}) = 2$. Since \mathcal{F} is maximal, it contains all k-spaces containing a $T \in \mathcal{T}$. Since $n \geq 2k$ and $k \geq 2$ two disjoint elements of \mathcal{T} would be contained in disjoint elements of \mathcal{F} , which is impossible. So \mathcal{T} is intersecting.

The following observation is immediate: if $A, B \in \mathcal{T}$ and $A \cap B < C < A \lor B$, then $C \in \mathcal{T}$. As an intersecting family of 2-spaces is either a family of 2-spaces containing some fixed 1-space E or a set of 2-subspaces of a 3-space, we get the following:

(*): \mathcal{T} is either a family of all 2-subspaces in a given (l+1)-space containing some fixed 1-space E (and $k \ge l \ge 1$), or \mathcal{T} is the set of all 2-subspaces of a 3-space.

(i) : If $|\mathcal{T}| = 1$, then let S denote the only 2-space in \mathcal{T} and let $E \leq S$ be any 1-space. Since $\tau(\mathcal{F}) > 1$ there exists an $F \in \mathcal{F}$ with $E \leq F$, for which we must have $\dim(F \cap S) = 1$. Since S is the only subspace of \mathcal{T} , for any 1-subspace E' of F different from $F \cap S$, $\mathcal{F}_{E \vee E'} \leq {k \brack 1} {n-3 \brack k-3}$ by Lemma 102, hence the number of subspaces containing E but not containing S is at most $({k \brack 1} - 1){k \brack k-3}$. This gives the upper bound.

(ii): Assume that $\tau(\mathcal{T}) = 1$ and $|\mathcal{T}| > 1$. By (*), \mathcal{T} is the set of 2-spaces in an (l+1)-space W (with $l \geq 2$) containing some fixed 1-space E. Every $F \in \mathcal{F} \setminus \mathcal{F}_E$ intersects W in a hyperplane. Let L be a hyperplane in W not on E. Then \mathcal{F} contains all k-spaces on E that intersect L. Hence the lower bound and the first term in the upper bound come from Lemma 99. The second term comes from counting the k-spaces of \mathcal{F} that contain E and intersect a given $F \in \mathcal{F}$ (not containing E) in a point of $F \setminus W$. Here Lemma 102 is used. If $l \geq 3$, then there are q^l hyperplanes in W not containing E and there are $\begin{bmatrix} n-l \\ k-l \end{bmatrix}$ k-spaces through such a hyperplane. For l = 2 there are q^2 hyperplanes in W and they cannot be in \mathcal{T} . Using Lemma 102 gives the bound.

(*iii*) is immediate.

Corollary 105. Let $\mathcal{F} \subseteq {V \brack k}$ be a maximal intersecting family with $\tau(\mathcal{F}) = 2$. Suppose $q \geq 3$ and $n \geq 2k+1$, or q = 2 and $n \geq 2k+2$. If \mathcal{F} is at least as large as an HM-type family and k > 3, then \mathcal{F} is an HM-type family. If k = 3, then \mathcal{F} is an HM-type family or an \mathcal{F}_3 -type family.

There exists an $\varepsilon > 0$ (independent of n, k, q) such that if $k \ge 4$ and $|\mathcal{F}|$ is at least $(1 - \varepsilon)$ times the size of an HM-type family, then \mathcal{F} is an HM-type family.

Proof. Apply Proposition 104. Note that the HM-type families are precisely those from case (ii) with l = k. Let n = 2k + r where $r \ge 1$. We have $|\mathcal{F}| / {n-2 \choose k-2} < 1 + \frac{q+1}{(q-1)q^r} {k \choose 1}$ in case (i) of Proposition 104 by Lemma 98. We have $|\mathcal{F}| / {n-2 \choose k-2} < (\frac{1}{q} + \frac{1}{(q-1)q^r}) {k \choose 1} + \frac{q^2}{(q-1)q^r}$ in case (ii) when l < k. In both cases, for $q \ge 3$ and $k \ge 3$, or $q = 2, k \ge 4$, and $r \ge 2$, this is less than $(1 - \varepsilon)$ times the lower bound on the size of an HM-type family given in Lemma 100. Using the stronger estimate in Lemma 100, we find the same conclusion for q = 2, k = 3, and $r \ge 2$.

In case (iii), $|\mathcal{F}_3| = \begin{bmatrix} 3\\2 \end{bmatrix} \begin{bmatrix} n-2\\k-2 \end{bmatrix} - \frac{q^3-q}{q-1} \begin{bmatrix} n-3\\k-3 \end{bmatrix}$. For $k \ge 4$, this is much smaller than the size of the HM-type families. For k = 3, the two families have the same size.

From now on we can suppose that $\mathcal{F} \subseteq \begin{bmatrix} V \\ k \end{bmatrix}$ is an intersecting family and $\tau(\mathcal{F}) = l > 2$. We shall derive a contradiction from $|\mathcal{F}| \ge f(n, k, q)$, and even from $|\mathcal{F}| > (1 - \varepsilon)f(n, k, q)$ for some $\varepsilon > 0$ (independent of n, k, q), where f(n, k, q) is the size $\begin{bmatrix} n-1 \\ k-1 \end{bmatrix} - q^{k(k-1)} \begin{bmatrix} n-k-1 \\ k-1 \end{bmatrix} + q^k$ of an HM-type family.

First consider the case l = k. Then $|\mathcal{F}| \leq {k \choose 1}^k$ by Corollary 2.6. On the other hand,

$$|\mathcal{F}| > \left(1 - \frac{1}{q^3 - q}\right) {k \brack 1} {n-2 \brack k-2} > \left(1 - \frac{1}{q^3 - q}\right) {k \brack 1}^{k-1} ((q-1)q^{n-2k})^{k-2}$$

by Lemma 99 and Lemma 98. If either $q \ge 3$, $n \ge 2k+1$ or q = 2, $n \ge 2k+2$, then either k+3, (n,k,q) = (9,4,3), or (n,k,q) = (10,4,2). If (n,k,q) = (9,4,3), then f(n,k,q) = 3837721, and $40^4 = 2560000$, which gives a contradiction. If (n,k,q) = (10,4,2), then f(n,k,q) = 153171, and 154 = 50625, which again gives a contradiction. Hence k = 3. Now $|\mathcal{F}| > (1 - \frac{1}{q^3 - q}) {k \choose 1} {n-2 \choose k-2}$ gives a contradiction for $n \ge 8$, so n = 7. Therefore, if we assume that $n \ge 2k + 1$ and either $q \ge 3$, $(n,k) \ne (7,3)$ or q = 2, $n \ge 2k + 2$ then we are not in the case l = k.

It remains to settle the case n = 7, k = l = 3, and $q \ge 3$. By Lemma 101, we can choose a 1-space E such that $|\mathcal{F}_E| > |\mathcal{F}|/{3 \brack 1}$ and a 2-space S on E such that $|\mathcal{F}_S| \ge |\mathcal{F}_E|/{3 \brack 1}$. Then $|\mathcal{F}_S| > q + 1$ since $|\mathcal{F}| > {2 \brack 1} {3 \brack 1}^2$. Pick $F' \in \mathcal{F}$ disjoint from S and define $H := S \lor F'$. All $F \in \mathcal{F}_S$ are contained in the 5-space H. Since $|\mathcal{F}| > {5 \brack 3}$, there is an $F_0 \in \mathcal{F}$ not contained in H. If $F_0 \cap S = 0$, then each $F \in \mathcal{F}_S$ is contained in $S \lor (H \cap F_0)$; this implies $|\mathcal{F}_S| \le q + 1$, which is impossible. Thus, all elements of \mathcal{F} disjoint from S are in H.

Now F_0 must meet F' and S, so F_0 meets H in a 2-space S_0 . Since $|\mathcal{F}_S| > q + 1$, we can find two elements F_1, F_2 of \mathcal{F}_S with the property that S_0 is not contained in the 4-space $F_1 \vee F_2$. Since any $F \in \mathcal{F}$ disjoint from S is contained in H and meets F_0 , it must meet S_0 and also F_1 and F_2 . Hence the number of such F's is at most q^5 . Altogether $|\mathcal{F}| \leq q^5 + {2 \brack 1} {3 \brack 1}^2$; the first term comes from counting $F \in \mathcal{F}$ disjoint from S and the second term comes from counting $F \in \mathcal{F}$ on a given one-dimensional subspace E < S. This contradicts $|\mathcal{F}| > (1 - \frac{1}{q^3 - q}) {3 \brack 1} {5 \brack 1}$. Finally, we can assume l < k. Suppose, for the moment, that there are two l-

Finally, we can assume l < k. Suppose, for the moment, that there are two l-subspaces in V that non-trivially intersect all $F \in \mathcal{F}$, and that these two l-spaces meet in an m-space, where $0 \le m \le l-1$. By Corollary 103, for each 1-subspace P we have $|\mathcal{F}_P| \le {k \brack l^{l-1} {n-l \brack k-l}}$, and for each 2-subspace L we have $|\mathcal{F}_L| \le {k \brack l^{l-2} {n-l \brack k-l}}$. Consequently,

$$|\mathcal{F}| \leq {m \choose 1} {k \choose l}^{l-1} {n-l \choose k-l} + \left({l \choose 1} - {m \choose 1} \right)^2 {k \choose 1}^{l-2} {n-l \choose k-l}.$$
 (11)

The upper bound (11) is a quadratic in $x = \begin{bmatrix} m \\ 1 \end{bmatrix}$ and is largest at one of the extreme values x = 0 and $x = \begin{bmatrix} l-1 \\ 1 \end{bmatrix}$. The maximum is taken at x = 0 only when $\begin{bmatrix} l \\ 1 \end{bmatrix} - \frac{1}{2} \begin{bmatrix} k \\ 1 \end{bmatrix} > \frac{1}{2} \begin{bmatrix} l-1 \\ 1 \end{bmatrix}$; that is, when k = l. Since we assume that l < k, the upper bound in (11) is largest for m = l - 1. We find

$$|\mathcal{F}| \leq {\binom{l-1}{1}} {\binom{k}{1}}^{l-1} {\binom{n-l}{k-l}} + \left({\binom{l}{1}} - {\binom{l-1}{1}}\right)^2 {\binom{k}{1}}^{l-2} {\binom{n-l}{k-l}}.$$

On the other hand,

$$|\mathcal{F}| > \left(1 - \frac{1}{q^3 - q}\right) {k \brack 1} {n-2 \brack k-2} > \left(1 - \frac{1}{q^3 - q}\right) {k \brack 1}^{l-1} {n-l \brack k-l} ((q-1)q^{n-2k})^{l-2}.$$

Comparing these, and using k > l, $n \ge 2k + 1$, and $n \ge 2k + 2$ if q = 2, we find either (n, k, l, q) = (9, 4, 3, 3) or q = 2, n = 2k + 2, l = 3, and $k \le 5$. If (n, k, l, q) =(9, 4, 3, 3) then f(n, k, q) = 3837721, while the upper bound is 3508960, which is a contradiction. If (n, k, l, q) = (12, 5, 3, 2) then f(n, k, q) = 183628563, while the upper bound is 146766865, which is a contradiction. If (n, k, l, q) = (10, 4, 3, 2) then f(n, k, q) =153171, while the upper bound is 116205, which is a contradiction. Hence, under our assumption that there are two distinct *l*-spaces that meet all $F \in \mathcal{F}$, the case 2 < l < kcannot occur. We now assume that there is a unique *l*-space *T* that meets all $F \in \mathcal{F}$. We can pick a 1-space E < T such that $|\mathcal{F}_E| > |\mathcal{F}|/{\binom{l}{1}}$. Now there is some $F' \in \mathcal{F}$ not on *E*, so *E* is in $\binom{k}{1}$ lines such that each $F \in \mathcal{F}_E$ contains at least one of these lines. Suppose *L* is one of these lines and *L* does not lie in *T*; we can enlarge *L* to an *l*-space that still does not meet all elements of \mathcal{F} , so $|\mathcal{F}_L| \leq {\binom{k}{1}}^{l-1} {\binom{n-l-1}{k-l-1}}$ by Lemma 101 and Lemma 102. If *L* does lie on *T*, we have $|\mathcal{F}_L| \leq {\binom{k}{1}}^{l-2} {\binom{n-l}{k-l}}$ by Corollary 103. Hence,

$$|\mathcal{F}| \leq {l \choose 1} |\mathcal{F}_E| \leq {l \choose 1} \left({l-1 \choose 1} {k \choose 1}^{l-2} {n-l \choose k-l} + \left({k \choose 1} - {l-1 \choose 1} \right) {k \choose 1}^{l-1} {n-l-1 \choose k-l-1} \right).$$

On the other hand, we have $|\mathcal{F}| > (1 - \frac{1}{q^3 - q})((q - 1)q^{n-2k})^{l-2} {k \brack 1}^{l-1} {n-l \brack k-l}$. Under our standard assumptions $n \ge 2k+1$ and $n \ge 2k+2$ if q = 2, this implies q = 2, n = 2k+2, l = 3, which gives a contradiction. We showed: If $q \ge 3$ and $n \ge 2k+1$ or if q = 2 and $n \ge 2k+2$, then an intersecting family $\mathcal{F} \subseteq {V \brack k}$ with $|\mathcal{F}| > f(n, k, q)$ must satisfy $\tau(\mathcal{F}) \ge 2$. Together with Corollary 105, this proves Theorem 80.

Finally, we turn to the proof of Theorem 81, that is, we show that $\chi(qK_{n:k}) = {\binom{n-k+1}{1}}$. The case k = 2 was proven in [21] and the general case for $q > q_k$ in [91]. We will need the following result of Bose and Burton [16].

Theorem 106 (Bose, Burton [16]). If V is an n-dimensional vector space over $\mathbf{F}(q)$ and \mathcal{E} is a family of 1-subspaces of V such that any k-subspace of V contains at least one element of \mathcal{E} , then $|\mathcal{E}| \geq {n-k+1 \choose 1}$. Furthermore, equality holds if and only if $\mathcal{E} = {H \choose 1}$ for some (n-k+1)-subspace H of V.

For the proof of Theorem 81, we will also need the natural extension of the Bose-Burton result.

Proposition 107 (Metsch [89]). If V is an n-dimensional vector space over \mathbf{F}_q and \mathcal{E} is a family of $\binom{n-k+1}{1} - \varepsilon$ 1-subspaces of V, then the number of k-subspaces of V that are disjoint from all $E \in \mathcal{E}$ is at least $\varepsilon q^{(k-1)(n-k+1)}$.

Proof of Theorem 81. Suppose that we have a coloring with at most $\binom{n-k+1}{1}$ colors. Let G (the good colors) be the set of colors that are point-pencils and let B (the bad colors) be the remaining set of colors. Then $|G| + |B| \leq \binom{n-k+1}{1}$. Suppose $|B| = \varepsilon > 0$. By Proposition 107, the number of k-spaces with a color in B is at least $\varepsilon q^{(k-1)(n-k)}$,

so that the average size of a bad color class is at least $q^{(k-1)(n-k)}$. This must be smaller than the size of a HM-type family. Thus, by Lemma 100,

$$q^{(k-1)(n-k)} < \begin{bmatrix} k\\1 \end{bmatrix} \begin{bmatrix} n-2\\k-2 \end{bmatrix}.$$

For $k \ge 3$ and $q \ge 3$, $n \ge 2k+1$ or q+2, $n \ge 2k+2$ this is a contradiction. The weaker form of Proposition 107, as stated in [89], suffices unless q = 2, n = 2k+2.)

If |B| = 0, then all color classes are point-pencils, and we are done by Theorem 106.

References

- [1] R. Ahlswede and L. H. Khachatrian. The complete intersection theorem for systems of finite sets. *European Journal of Combinatorics*, 18(2):125–136, 1997.
- [2] R. Ahlswede and L. H. Khachatrian. Counterexample to the Frankl-Pach conjecture for uniform, dense families. *Combinatorica*, 17(2):299–301, 1997.
- [3] N. Alon. On the density of sets of vectors. Discrete Mathematics, 46(2):199-202, 1983.
- [4] N. Alon and C. Shikhelman. Many T copies in H-free graphs. Journal of Combinatorial Theory, Series B, 121:146–172, 2016.
- [5] R. Anstee. A survey of forbidden configuration results. *Electronic Journal of Combinatorics*, 1000:DS20–Jan, 2013.
- [6] R. Baer. Partitionen abelscher gruppen. Archiv der Mathematik, 14(1):73–83, 1963.
- [7] B. Bajnok. Additive combinatorics: A menu of research problems. CRC Press, 2018.
- [8] J. Balogh and A. Z. Wagner. Kleitman's conjecture about families of given size minimizing the number of k-chains. Advances in Mathematics, 330:229–252, 2018.
- [9] S. Bezrukov and A. Blokhuis. A Kruskal–Katona type theorem for the linear lattice. *European Journal of Combinatorics*, 20(2):123–130, 1999.
- [10] A. Blokhuis, A. E. Brouwer, A. Chowdhury, P. Frankl, T. Mussche, B. Patkós, and T. Szönyi. A Hilton-Milner theorem for vector spaces. *The Electronic Journal* of Combinatorics [electronic only], 17(1):R71, 2010.
- [11] E. Boehnlein and T. Jiang. Set families with a forbidden induced subposet. Combinatorics, Probability and Computing, 21(04):496–511, 2012.
- [12] B. Bollobás. On generalized graphs. Acta Mathematica Hungarica, 16(3-4):447– 452, 1965.
- [13] B. Bollobás and E. Győri. Pentagons vs. triangles. Discrete Mathematics, 308(19):4332–4336, 2008.
- [14] B. Bollobás, I. Leader, and A. Radcliffe. Reverse Kleitman inequalities. Proceedings of the London Mathematical Society, 3(1):153–168, 1989.
- [15] B. Bollobás and A. J. Radcliffe. Defect Sauer results. Journal of Combinatorial Theory, Series A, 72(2):189–208, 1995.
- [16] R. Bose and R. Burton. A characterization of flat spaces in a finite geometry and the uniqueness of the Hamming and the MacDonald codes. *Journal of Combina*torial Theory, 1(1):96–104, 1966.

- [17] B. Bukh. Set families with a forbidden subposet. Electronic Journal of Combinatorics, 16(1):R142, 2009.
- [18] P. Burcsi and D. T. Nagy. The method of double chains for largest families with excluded subposets. *Electronic Journal of Graph Theory and Applications*, 1(1), 2013.
- [19] T. Carroll and G. O. H. Katona. Bounds on maximal families of sets not containing three sets with $A \cap B \subset C$, $A \not\subset B$. Order, 25(3):229–236, 2008.
- [20] H.-B. Chen and W.-T. Li. A note on the largest size of families of sets with a forbidden poset. Order, 31(1):137–142, 2014.
- [21] A. Chowdhury, C. Godsil, and G. Royle. Colouring lines in projective space. Journal of Combinatorial Theory, Series A, 113(1):39–52, 2006.
- [22] A. Chowdhury and B. Patkós. Shadows and intersections in vector spaces. Journal of Combinatorial Theory, Series A, 117(8):1095–1106, 2010.
- [23] S. Das, W. Gan, and B. Sudakov. Sperner's theorem and a problem of Erdős, Katona and Kleitman. *Combinatorics, Probability and Computing*, 24(04):585– 608, 2015.
- [24] A. De Bonis and G. O. H. Katona. Largest families without an r-fork. Order, 24(3):181–191, 2007.
- [25] A. De Bonis, G. O. H. Katona, and K. J. Swanepoel. Largest family without $A \cup B \subseteq C \cap D$. Journal of Combinatorial Theory, Series A, 111(2):331–336, 2005.
- [26] R. Dedekind. Über Zerlegungen von Zahlen Durch Ihre Grössten Gemeinsamen Theiler, chapter 1, pages 1–40. Vieweg+Teubner Verlag, Wiesbaden, 1897.
- [27] M. Deza and P. Frankl. Erdős–Ko–Rado theorem—22 years later. SIAM Journal on Algebraic Discrete Methods, 4(4):419–431, 1983.
- [28] R. M. Dudley. The structure of some Vapnik-Chervonenkis classes. In Proceedings of the Berkeley Conference in Honor of Jerzy Neyman, volume 2, pages 495–507, 1985.
- [29] K. Engel. Sperner Theory, volume 65. Cambridge University Press, 1997.
- [30] P. Erdős. On a theorem of Rademacher-Turán. Illinois J. Math, 6(122-127):1–3, 1962.
- [31] P. Erdős and H. Heilbronn. On the addition of residue classes mod p. Acta Arithmetica, 9:149–159, 1964.
- [32] P. Erdős. On a lemma of Littlewood and Offord. Bulletin of the American Mathematical Society, 51(12):898–902, 1945.

- [33] P. Erdős. On some new inequalities concerning extremal properties of graphs. In Theory of Graphs (Proceedings of Colloquium, Tihany, 1966), pages 77–81, 1968.
- [34] P. Erdős and D. J. Kleitman. Extremal problems among subsets of a set. In Proc. Second Chapel Hill Conf. on Combinatorial Mathematics and its Applications, pages 146–170. Univ. North Carolina, Chapel Hill, N.C., 1970.
- [35] P. Erdős and D. J. Kleitman. Extremal problems among subsets of a set. Discrete Mathematics, 8(3):281–294, 1974.
- [36] P. Erdős, C. Ko, and R. Rado. Intersection theorems for systems of finite sets. The Quarterly Journal of Mathematics, 12(1):313–320, 1961.
- [37] P. Erdős and L. Lovász. Problems and results on 3-chromatic hypergraphs and some related questions. *Infinite and Finite Sets*, 10(2):609–627, 1975.
- [38] P. Erdős and M. Simonovits. A limit theorem in graph theory. Studia Sci. Math. Hung, 1:51–57, 1966.
- [39] N. Frankl, S. Kiselev, A. Kupavskii, and B. Patkós. VC-saturated set systems. arXiv preprint arXiv:2005.12545, 2020.
- [40] P. Frankl. On Sperner families satisfying an additional condition. Journal of Combinatorial Theory, Series A, 20(1):1–11, 1976.
- [41] P. Frankl. On intersecting families of finite sets. Bulletin of the Australian Mathematical Society, 21(03):363–372, 1980.
- [42] P. Frankl. On the trace of finite sets. Journal of Combinatorial Theory, Series A, 34(1):41–45, 1983.
- [43] P. Frankl and R. L. Graham. Intersection theorems for vector spaces. European Journal of Combinatorics, 6(2):183–187, 1985.
- [44] P. Frankl, K. Ota, and N. Tokushige. Uniform intersecting families with covering number four. Journal of Combinatorial Theory, Series A, 71(1):127–145, 1995.
- [45] P. Frankl and J. Pach. On disjointly representable sets. Combinatorica, 4(1):39– 45, 1984.
- [46] P. Frankl and R. M. Wilson. The Erdős-Ko-Rado theorem for vector spaces. Journal of Combinatorial Theory, Series A, 43(2):228–236, 1986.
- [47] Z. Füredi and F. Quinn. Traces of finite sets. Ars Combin, 18:195–200, 1984.
- [48] Z. Füredi and M. Simonovits. The history of degenerate (bipartite) extremal graph problems. In *Erdős Centennial*, pages 169–264. Springer, 2013.
- [49] D. Gerbner, B. Keszegh, and B. Patkós. Generalized forbidden subposet problems. Order, 37:389–410, 2020.

- [50] D. Gerbner, A. Methuku, D. T. Nagy, B. Patkós, and M. Vizer. On the number of containments in *P*-free families. *Graphs and Combinatorics*, 35(6):1519–1540, 2019.
- [51] D. Gerbner, D. Nagy, B. Patkós, and M. Vizer. Supersaturation, counting, and randomness in forbidden subposet problems. the Electronic Journal of Combinatorics, 28(1), 2021.
- [52] D. Gerbner and B. Patkós. *l*-chain profile vectors. SIAM Journal on Discrete Mathematics, 22(1):185–193, 2008.
- [53] D. Gerbner and B. Patkós. Extremal finite set theory. CRC Press, 2018.
- [54] D. Gerbner, B. Patkós, and M. Vizer. Forbidden subposet problems for traces of set families. *The Electronic Journal of Combinatorics*, pages P3–49, 2018.
- [55] B. Girard, S. Griffiths, and Y. O. Hamidoune. k-sums in abelian groups. Combinatorics, Probability and Computing, 21(4):582–596, 2012.
- [56] C. Godsil and M. Newman. Independent sets in association schemes. Combinatorica, 4(26):431–443, 2006.
- [57] C. Greene and D. Kleitman. Proof techniques in the ordered sets. Studies in Combinatorics, pages 22–79, 1978.
- [58] J. R. Griggs and G. O. H. Katona. No four subsets forming an N. Journal of Combinatorial Theory, Series A, 115(4):677–685, 2008.
- [59] J. R. Griggs, W.-T. Li, and L. Lu. Diamond-free families. Journal of Combinatorial Theory, Series A, 119(2):310–322, 2012.
- [60] J. R. Griggs and L. Lu. On families of subsets with a forbidden subposet. Combinatorics, Probability and Computing, 18(05):731–748, 2009.
- [61] D. Grósz, A. Methuku, and C. Tompkins. An improvement of the general bound on the largest family of subsets avoiding a subposet. *Order*, pages 1–13, 2016.
- [62] A. Grzesik. On the maximum number of five-cycles in a triangle-free graph. Journal of Combinatorial Theory, Series B, 102(5):1061–1066, 2012.
- [63] L. H. Harper. On a problem of Kleitman and West. Discrete Mathematics, 93(2):169–182, 1991.
- [64] H. Hatami, J. Hladky, D. Král', S. Norine, and A. Razborov. On the number of pentagons in triangle-free graphs. *Journal of Combinatorial Theory, Series A*, 120(3):722–732, 2013.
- [65] A. J. Hilton and E. C. Milner. Some intersection theorems for systems of finite sets. The Quarterly Journal of Mathematics, 18(1):369–384, 1967.

- [66] W. Hsieh. Intersection theorems for systems of finite vector spaces. Discrete Mathematics, 12(1):1–16, 1975.
- [67] F. Ihringer. The chromatic number of the q-kneser graph for large q. the Electronic Journal of Combinatorics, pages P1–50, 2019.
- [68] G. O. H. Katona. On a conjecture of Erdős and a stronger form of Sperner's theorem. Studia Scientarum Mathematicarum Hungarica, 1:59–63, 1966.
- [69] G. O. H. Katona. A theorem of finite sets. In Theory of Graphs, Proceedings of Colloquium held at Tihany, Hungary, pages 187–207, Budapest, 1968. Akadémia Kiadó.
- [70] G. O. H. Katona. Two applications of Sperner type theorems (for search theory and truth functions). *Period. Math. Hungar*, 3:19–26, 1973.
- [71] G. O. H. Katona and T. G. Tarján. Extremal problems with excluded subgraphs in the *n*-cube. In *Graph Theory*, pages 84–93. Springer, 1983.
- [72] P. Keevash. Shadows and intersections: stability and new proofs. Advances in Mathematics, 218(5):1685–1703, 2008.
- [73] D. Kleitman. Extremal hypergraph problems. Surveys in combinatorics, pages 44–65, 1979.
- [74] D. Kleitman. Extremal problems on hypergraphs. In Extremal Problems for Finite Sets, volume 3, pages 355–374, 1991.
- [75] D. J. Kleitman. A conjecture of Erdős-Katona on commensurable pairs among subsets of an n-set. In Theory of Graphs, Proceedings of Colloquium held at Tihany, Hungary, pages 187–207, 1966.
- [76] A. D. Korshunov. The number of monotone Boolean functions (in Russian). Problemy Kibernetiki, 38:5–108, 1981.
- [77] J. B. Kruskal. The number of simplices in a complex. Mathematical Optimization Techniques, pages 251–268, 1963.
- [78] L. Lovász. Combinatorial Problems and Exercises. Akadémiai Kiadó North Holland, 1979.
- [79] L. Lovász. Kneser's conjecture, chromatic number, and homotopy. Journal of Combinatorial Theory, Series A, 25(3):319–324, 1978.
- [80] L. Lu. On crown-free families of subsets. Journal of Combinatorial Theory, Series A, 126:216–231, 2014.
- [81] L. Lu and K. G. Milans. Set families with forbidden subposets. Journal of Combinatorial Theory, Series A, 136:126–142, 2015.

- [82] D. Lubell. A short proof of Sperner's lemma. Journal of Combinatorial Theory, 1(2):299, 1966.
- [83] W. Mantel. Problem 28. Wiskundige Opgaven, 10(60-61):320, 1907.
- [84] A. Méroueh. A LYM inequality for induced posets. Journal of Combinatorial Theory, Series A, 155:398–417, 2018.
- [85] L. D. Meshalkin. Generalization of Sperner's theorem on the number of subsets of a finite set. Theory of Probability & Its Applications, 8(2):203–204, 1963.
- [86] T. Mészáros and L. Rónyai. Shattering-extremal set systems of small vcdimension. International Scholarly Research Notices, 2013, 2013.
- [87] T. Mészáros and L. Rónyai. Shattering-extremal set systems of VC dimension at most 2. Electronic Journal of Combinatorics, 21(4), 2014.
- [88] A. Methuku and D. Pálvölgyi. Forbidden hypermatrices imply general bounds on induced forbidden subposet problems. *Combinatorics, Probability and Computing*, 26(4):593–602, 2017.
- [89] K. Metsch. How many s-subspaces must miss a point set in PG(d, q). Journal of Geometry, 86(1):154–164, 2007.
- [90] D. Mubayi and Y. Zhao. On the VC-dimension of uniform hypergraphs. Journal of Algebraic Combinatorics, 25(1):101–110, 2007.
- [91] T. Mussche. *Extremal combinatorics in generalized Kneser graphs*. PhD thesis, Technical University Eindhoven, 2009.
- [92] M. W. Newman. Independent sets and eigenspaces. PhD thesis, University of Waterloo, 2004.
- [93] J. A. Noel, A. Scott, and B. Sudakov. Supersaturation in posets and applications involving the container method. *Journal of Combinatorial Theory, Series A*, 154:247–284, 2018.
- [94] A. Pajor. Sous-espaces l_1^n des espaces de Banach. Hermann, 1985.
- [95] B. Patkós. l-trace k-Sperner families of sets. Journal of Combinatorial Theory, Series A, 116(5):1047–1055, 2009.
- [96] B. Patkós. Traces of uniform families of sets. *Electronic Journal of Combinatorics*, 16:N8, 2009.
- [97] B. Patkós. A note on traces of set families. Moscow Journal of Combinatorics and Number Theory, 2(1):47–55, 2012.
- [98] B. Patkós. Families that remain k-Sperner even after omitting an element of their ground set. The Electronic Journal of Combinatorics, 20(1):P32, 2013.

- [99] B. Patkós. Induced and non-induced forbidden subposet problems. *Electronic Journal of Combinatorics*, 22(1):P1–30, 2015.
- [100] B. Patkós. Supersaturation and stability for forbidden subposet problems. Journal of Combinatorial Theory, Series A, 136:220–237, 2015.
- [101] B. Rossman. Thresholds in the lattice of subspaces of \mathbb{F}_q^n . In Latin American Symposium on Theoretical Informatics, pages 504–515. Springer, 2021.
- [102] W. Samotij. Subsets of posets minimising the number of chains. Transactions of the American Mathematical Society, 371(10):7259–7274, 2019.
- [103] N. Sauer. On the density of families of sets. Journal of Combinatorial Theory, Series A, 13(1):145–147, 1972.
- [104] S. Shelah. A combinatorial problem; stability and order for models and theories in infinitary languages. *Pacific Journal of Mathematics*, 41(1):247–261, 1972.
- [105] M. Simonovits. A method for solving extremal problems in graph theory, stability problems. In *Theory of Graphs (Proceedings of Colloquium, Tihany, 1966)*, pages 279–319, 1968.
- [106] J. Spencer. Asymptotic lower bounds for ramsey functions. Discrete Mathematics, 20:69–76, 1977.
- [107] E. Sperner. Ein Satz über Untermengen einer endlichen Menge. Math Z, 27(1):585–592, 1928.
- [108] T. S. Tan. Traces without maximal chains. *Electronic Journal of Combinatorics*, 17(1):N16, 2010.
- [109] P. Turán. On an extremal problem in graph theory (in Hungarian). Matematikai és Fizikai Lapok, 48:0436–452, 1941.
- [110] V. N. Vapnik and A. Y. Chervonenkis. On uniform convergence of the frequencies of events to their probabilities. *Teoriya Veroyatnostei i ee Primeneniya*, 16(2):264–279, 1971.
- [111] K. Yamamoto. Logarithmic order of free distributive lattice. Journal of the Mathematical Society of Japan, 6(3-4):343–353, 1954.