

Set systems with no singleton intersection

Peter Keevash ^{*} Dhruv Mubayi [†] Richard M. Wilson [‡]

Abstract

Let \mathcal{F} be a k -uniform set system defined on a ground set of size n with no singleton intersection, i.e. no pair $A, B \in \mathcal{F}$ has $|A \cap B| = 1$. Frankl showed that $|\mathcal{F}| \leq \binom{n-2}{k-2}$ for $k \geq 4$ and n sufficiently large, confirming a conjecture of Erdős and Sós. We determine the maximum size of \mathcal{F} for $k = 4$ and all n , and also establish a stability result for general k , showing that any \mathcal{F} with size asymptotic to that of the best construction must be structurally similar to it.

1 Introduction

Say that a set system \mathcal{F} is L -intersecting if for every $A, B \in \mathcal{F}$ we have $|A \cap B| \in L$. Ray-Chaudhuri and Wilson [18] and Frankl and Wilson [10] obtained tight bounds for L -intersecting set systems. They showed that if L is a set of s non-negative integers and \mathcal{F} is an L -intersecting system on $[n] = \{1, \dots, n\}$ then $|\mathcal{F}|$ is at most $\sum_{i=0}^s \binom{n}{i}$, and at most $\binom{n}{s}$ if \mathcal{F} is k -uniform for some k , i.e. $|A| = k$ for each $A \in \mathcal{F}$. Equality can hold in these bounds when $L = \{0, \dots, s-1\}$. It is natural to ask for the best possible bound for each specific set L , and in general it is an open problem to even determine the order of magnitude. A detailed compilation of results on the uniform version of this problem can be found in [9].

We will consider the problem of finding the largest k -uniform family with no singleton intersection, i.e. L -intersecting with $L = \{0, 2, 3, \dots, k\}$. One construction of such a family is to take all k -subsets of $[n]$ that contain two specified points; this gives a family of $\binom{n-2}{k-2}$ sets with no singleton intersection, which also happens to have no empty intersection. Erdős and Sós [4] conjectured that this is the maximum number for $k \geq 4$ and sufficiently large n , and this was proved by Frankl [6]. (Note that when $n = 3$ the maximum number is n , which can be achieved when n is divisible by 4 by taking $n/4$ vertex disjoint copies of $K_4^{(3)}$, i.e. the complete triple system on 4 points.)

For a more complete understanding of the problem, one might hope to find the maximum number for all n , and to describe the structure of the maximum systems. Our first theorem achieves this when

^{*}Department of Mathematics, Caltech, Pasadena, CA 91125, USA. Email: keevash@caltech.edu. Research supported in part by NSF grant DMS-0555755.

[†]Department of Mathematics, Statistics and Computer Science, University of Illinois, Chicago, IL 60607. Email: mubayi@math.uic.edu. Research supported in part by NSF grant DMS-0400812, and by an Alfred P. Sloan fellowship.

[‡]Department of Mathematics, Caltech, Pasadena, CA 91125, USA. Email: rmw@caltech.edu.

$k = 4$, and our approach gives some additional structural information for general k . Our basic idea is to consider a maximum matching and estimate the rest of the family based on the intersections of its sets with this matching. The same technique has recently been successful for various other extremal problems, such as in [13] and [16].

Before stating our first theorem, we should mention the fundamental intersection theorem of Erdős, Ko and Rado [5]. Say that a set system \mathcal{F} is t -intersecting if for every $A, B \in \mathcal{F}$ we have $|A \cap B| \geq t$. They showed that, if \mathcal{F} is k -uniform and t -intersecting on $[n]$ with n sufficiently large, then $|\mathcal{F}| \leq \binom{n-t}{k-t}$. (The case $t = 2$ is pertinent to our current discussion.) Confirming a conjecture of Erdős, Wilson [19] showed that this bound in fact holds for $n \geq (t+1)(k-t+1)$ (which is the best possible strengthening), and furthermore that the unique maximum system consists of all k -sets containing some fixed t -set. To describe the complete solution for all n we need to define the t -intersecting systems $\mathcal{F}_i^{k,t}(n) = \{A \subset [n] : |A| = k, |A \cap [t+2i]| \geq t+i\}$, for $0 \leq i \leq k-t$. The complete intersection theorem, conjectured by Frankl, and proved by Ahlswede and Khachatrian [1], is that a maximum size k -uniform t -intersecting family on $[n]$ is isomorphic to $\mathcal{F}_i^{k,t}(n)$, for some i which can easily be computed given n . Note that $\mathcal{F}_0^{k,t}(n)$ is the system of all k -sets containing some fixed t -set. These constructions also appear in our analysis for 4-uniform systems with no singleton intersection.

Theorem 1.1 *Suppose \mathcal{F} is a 4-uniform set system on $[n]$ with no pair $A, B \in \mathcal{F}$ satisfying $|A \cap B| = 1$. Then*

$$|\mathcal{F}| \leq \begin{cases} \binom{n}{4} & n = 4, 5, 6 \\ 15 & n = 7 \\ 17 & n = 8 \\ \binom{n-2}{2} & n \geq 9 \end{cases}$$

Furthermore, the only cases of equality are $K_n^{(4)}$ for $n = 4, 5$, $\mathcal{F}_2^{4,2}(n) = K_6^{(4)}$ for $n = 6, 7$, $\mathcal{F}_1^{4,2}(8)$ for $n = 8$ and $\mathcal{F}_0^{4,2}(n)$ for $n \geq 9$.

Many extremal problems have a property known as stability, meaning that not only do they have a unique maximising construction, but also any family with size asymptotic to that of the best construction must be structurally similar to it. Stability theorems can be useful tools for establishing exact results (e.g. [15]), and for enumerating discrete structures (e.g. [3]). They are also interesting in their own right, as they provide information about the problem that is structural, rather than just numerical, and they often motivate new proof techniques where the original ones do not suffice.

A strong stability version of the Erdős-Ko-Rado theorem was obtained by Frankl [7], extending an earlier result of Hilton and Milner [11]. A similar result with different assumptions on the parameters was also obtained by Anstee and Keevash [2]. A simple consequence of Frankl's theorem (which is also easy to prove directly) is that, for any k there is $c(k)$ such that, if \mathcal{F} is k -uniform and t -intersecting on $[n]$ with $|\mathcal{F}| > c(k)n^{k-t-1}$ and n sufficiently large, then there is a set of t points that is contained in every set of \mathcal{F} .

These stability theorems are stronger than the usual stability paradigm in two senses: firstly the supposed lower bound on $|\mathcal{F}|$ is of a lower order of magnitude than the maximum possible (rather than

asymptotic to it), and secondly the conclusion is that \mathcal{F} is contained in the best construction (rather than structurally similar to it). An example of a stability theorem for set systems that is not strong was given by Mubayi [17]. Also, a strong stability theorem cannot hold for our problem of having no singleton intersection. To see this, note that if \mathcal{A} and \mathcal{B} are families on disjoint sets X and Y with no singleton intersection then $\mathcal{A} \cup \mathcal{B}$ is a family on $A \cup Y$ with no singleton intersection. If $X \cup Y = [n]$ and $|Y| = o(n)$ we can take $|\mathcal{A}| \sim \binom{n-2}{k-2}$ but there need not be two points that belong to all of the sets. Our next result is a (normal) stability theorem for systems having no singleton intersection.

Theorem 1.2 *For any $\epsilon > 0$ there is $\delta > 0$ such that if \mathcal{F} is a k -uniform family on $[n]$ with no singleton intersection and $|\mathcal{F}| \geq (1 - \delta) \binom{n-2}{k-2}$, then there are two points x, y so that all but at most ϵn^{k-2} sets of \mathcal{F} contain both x and y .*

A result that is useful in the proof of Theorem 1.2, and is of independent interest, is the following bound, which is slightly suboptimal, but has the advantage of being valid for all n .

Theorem 1.3 *Let \mathcal{F} be a k -uniform family on $[n]$ with no singleton intersection, where $k \geq 3$. Then $|\mathcal{F}| \leq \binom{n}{k-2}$.*

The rest of this paper is organised as follows. We start, in the next section, by quickly deducing Theorem 1.3 from a result of Frankl and Wilson [10]. Then we prove Theorem 1.1 in section 3. Some lemmas used in the proof of Theorem 1.2 are given in section 4, and the proof itself in section 5.

Notation: We write $[n] = \{1, \dots, n\}$. Typically \mathcal{F} is a k -uniform set system (or family, or hypergraph) with ground set $[n]$. Given $A \subset [n]$, the link of \mathcal{F} from A is $\mathcal{F}(A) = \{F \setminus A : A \subset F \in \mathcal{F}\}$. The complete r -uniform hypergraph on s vertices is denoted $K_s^{(r)}$. For $0 \leq i \leq k - t$ we define $\mathcal{F}_i^{k,t}(n) = \{A \subset [n] : |A| = k, |A \cap [t + 2i]| \geq t + i\}$.

2 A bound for all n

In this section we prove Theorem 1.3. It is a simple consequence of the following theorem of Frankl and Wilson, implicit in [10]. For the convenience of the reader we briefly reproduce their proof.

Theorem 2.1 *Suppose p is prime, $k \in \mathbb{N}$, $L \subset \{0, \dots, k-1\}$, and $f(x)$ is an integer-valued polynomial of degree $d \leq k$ such that $f(\ell) \equiv 0 \pmod{p}$ for $\ell \in L$ and $f(k) \not\equiv 0 \pmod{p}$. If \mathcal{F} is a k -uniform L -intersecting set system on $[n]$ then $|\mathcal{F}| \leq \binom{n}{d}$.*

Proof. Let $W_{i,j}$ be the matrix with rows indexed by the i -subsets of $[n]$ and columns by the j -subsets of $[n]$, where given $|A| = i$ and $|B| = j$, the entry $W_{i,j}(A, B)$ is 1 if $A \subset B$ and 0 if $A \not\subset B$. Let V be the row space of $W_{d,k}$. The identity $W_{i,d}W_{d,k} = \binom{k-i}{d-i}W_{i,k}$ implies that V contains the row space of $W_{i,k}$ for all $i \leq d$. Since f is integer-valued there are integers a_0, \dots, a_d such that $f(x) = \sum_{i=0}^d a_i \binom{x}{i}$,

where $\binom{x}{i}$ is the polynomial $\frac{1}{i!}x(x-1)\cdots(x-i+1)$. Consider the matrix $M = \sum_{i=0}^d a_i W_{i,k}^T W_{i,k}$. The row space of M is contained in V , so $\text{rank } M \leq \dim V \leq \binom{n}{d}$. On the other hand, given k -sets A, B we have $M(A, B) = \sum_{i=0}^d a_i \binom{|A \cap B|}{i} = f(|A \cap B|)$. Let M_0 be the submatrix of M consisting of elements $M(A, B)$ with $A, B \in \mathcal{F}$. By our assumptions $M(A, B) \equiv 0 \pmod{p}$ for $A \neq B$ and $M(A, A) \not\equiv 0 \pmod{p}$, so M_0 is non-singular. Therefore $|\mathcal{F}| = \text{rank } M_0 \leq \text{rank } M \leq \binom{n}{d}$. \square

Proof of Theorem 1.3. Let p be a prime that divides $k-1$ and $f(x) = \binom{x-2}{k-2}$, a polynomial of degree $k-2$. Then $f(i) = 0$ for $2 \leq i \leq k-1$, $f(0) = (-1)^{k-2}(k-1) \equiv 0 \pmod{p}$ and $f(k) = 1$. By Theorem 2.1, if \mathcal{F} is a k -uniform family on $[n]$ with no singleton intersection, then $|\mathcal{F}| \leq \binom{n}{k-2}$. \square

3 Solution for 4-uniform families

Throughout we suppose that \mathcal{F} is a 4-uniform set system on $[n]$ with no singleton intersection, i.e. there is no pair $A, B \in \mathcal{F}$ with $|A \cap B| = 1$. In this section we will prove Theorem 1.1, which describes such families \mathcal{F} of maximum size. We start by discussing the small values of n . Trivially $K_n^{(4)}$ is the maximum family for $n = 4, 5, 6$. Also, when $n = 7$ then \mathcal{F} cannot contain two disjoint sets, so is 2-intersecting, and the complete intersection theorem shows that the maximum family is $\mathcal{F}_2^{4,2}(7) = K_6^{(4)}$. Next suppose that $n = 8$. If \mathcal{F} does not contain two disjoint sets then as before it is 2-intersecting, so contains at most 17 sets, with equality only for $\mathcal{F}_1^{4,2}(8)$. In fact, this is the maximum family, as shown by the case $t = 2$ of the next lemma.

Lemma 3.1 *Suppose that \mathcal{F} is a 4-uniform family on $[n]$ with no singleton intersection, and contains a perfect matching A_1, \dots, A_t , with $t \geq 2$. Then $|\mathcal{F}| \leq 3\binom{2t}{2} - 2t$.*

Proof. We argue by induction on t . First we do the base case, where $t = 2$ and it is required to show that $|\mathcal{F}| \leq 14$. Note that every set in \mathcal{F} other than A_1 or A_2 has two points in each of A_1 and A_2 . Given a pair uv in A_1 let $\mathcal{F}(uv)$ be its link in A_2 , i.e. the set of pairs xy in A_2 for which $uvxy$ is in \mathcal{F} , and write $d(uv) = |\mathcal{F}(uv)|$. Since \mathcal{F} has no singleton intersection the links have the following properties:

(i) If uv and wx are disjoint pairs in A_1 and a, b, c are distinct points of A_2 then we do not have $ab \in \mathcal{F}(uv)$ and $ac \in \mathcal{F}(wx)$.

(ii) If ab and cd are disjoint pairs in A_2 and u, v, w are distinct points of A_1 then we do not have $ab \in \mathcal{F}(uv)$ and $cd \in \mathcal{F}(uw)$.

We consider cases according to the maximum value of $d(uv)$. The above properties imply that if there is a pair uv in A_1 with $d(uv) = 6$ then $d(u'v') = 0$ for all other pairs $u'v'$ in A_1 , and if there is a pair uv in A_1 with $d(uv) = 5$ then $d(u'v') \leq 1$ for all other pairs $u'v'$ in A_1 . In either case we have $|\mathcal{F}| = 2 + \sum_{u,v \in A_1} d(uv) < 14$. Otherwise, if $d(uv) \leq 4$ for all pairs uv in A_1 , we claim that for any two opposite pairs uv, wx in A_1 we have $d(uv) + d(wx) \leq 4$. To see this, we can suppose that say $d(uv) \geq 3$. But now, if $ab \in \mathcal{F}(wx)$, by property (i) $\mathcal{F}(uv)$ can only contain ab or $A_2 \setminus ab$,

contradicting the assumption that $d(uv) \geq 3$. Therefore $d(wx) = 0$, so $d(uv) + d(wx) \leq 4$. Since K_4 can be decomposed into 3 matchings, $|\mathcal{F}| = 2 + \sum_{u,v \in A_1} d(uv) \leq 2 + 3 \cdot 4 = 14$, as required.

Now suppose $t \geq 3$. By the case $t = 2$, for every $1 \leq i \leq t-1$ there are at most 12 sets with 2 points in each of A_i and A_t . Thus there are at most $12(t-1) + 1$ sets incident to A_t . By induction hypothesis there are at most $3\binom{2(t-1)}{2} - 2(t-1)$ sets within $\cup_{i=1}^{t-1} A_i$, so in total we have at most $3\binom{2t}{2} - 2t$. \square

The heart of the proof of Theorem 1.1 is contained in the following theorem, which in the case when \mathcal{F} is not intersecting gives a stronger bound on its size. We define

$$b_2(n) = 13 + \max \left\{ 7(n-8), \binom{n-6}{2} \right\}, \quad \text{and}$$

$$b_t(n) = 3\binom{2t}{2} - 2t - 1 + \max \left\{ 3t(n-4t), \binom{n-4t+2}{2} \right\} \quad \text{for } t \geq 3.$$

Theorem 3.2 *Suppose that \mathcal{F} is a 4-uniform family on $[n]$ with no singleton intersection. Let A_1, \dots, A_t be a maximum matching in \mathcal{F} , and suppose $t \geq 2$. Then $|\mathcal{F}| \leq b_t(n)$.*

Proof. Let $A = \cup_{i=1}^t A_i$ and $B = [n] \setminus A$. By maximality of t there are no sets of \mathcal{F} contained in B . The sets contained within A may be estimated by Lemma 3.1: there are at most $3\binom{2t}{2} - 2t$ of them. The remaining sets intersect both A and B , and since there are no singleton intersections they have two possible types: 2 points in some A_i and 2 in B or 3 points in some A_i and 1 in B .

Say that a pair xy in B has colour i if there is a pair ab in A_i such that $abxy$ is a set of \mathcal{F} . Note that a pair may have more than one colour, or be uncoloured. Let M be the set of all pairs xy in B which are coloured but do not intersect any other coloured pair. Thus M is a perfect matching on some set $D \subset B$. Now if a pair xy has more than one colour there can be no set of \mathcal{F} that intersects it in one point: this would create a singleton intersection. In this case all sets in \mathcal{F} meeting xy consist of xy together with a pair in some A_i , so there are at most $6t$ such sets. On the other hand, if xy has a unique colour i then all sets meeting it are contained in $A_i \cup \{x, y\}$, so there are at most $\binom{6}{4} - 1 = 14$ such sets. Thus the number of sets of \mathcal{F} meeting a coloured pair xy is at most $\max\{6t, 14\}$. Setting $d = |D| = 2|M|$, this gives at most $\max\{3td, 7d\}$ sets of \mathcal{F} meeting D .

All other coloured pairs are contained in $B \setminus D$. Let G_i be those of colour i and C_i be those vertices contained in some pair of G_i . Note that C_i can be empty. The crucial observation of the proof is that C_1, \dots, C_t are disjoint (and so the same is true of G_1, \dots, G_t). To see this, suppose to the contrary that $x \in C_i \cap C_j$. Then $xy \in G_i$ and $xz \in G_j$ for some y, z . If $y \neq z$ then we would have a singleton intersection in \mathcal{F} . On the other hand, if $y = z$, we note that since $xy \notin M$ there is another coloured pair P that intersects it. A colour of P is different from at least one of i and j , so again we have a singleton intersection. Thus C_1, \dots, C_t are disjoint.

Let $C = \cup_{i=1}^t C_i$ and $E = B \setminus (C \cup D)$. Any set in \mathcal{F} meeting E has 1 point in E and 3 points in some A_i , which must be uniquely specified to avoid a singleton intersection. Thus there are at most $4e$ such sets, where $e = |E|$. All other sets in \mathcal{F} meet C , so are contained in $A_i \cup C_i$ for some i .

Next we note that the sets in \mathcal{F} within $A_i \cup C_i$ form a 2-intersecting family; for there are no singleton intersections, and if $A_i \cup C_i$ contained two disjoint sets we could enlarge the matching A_1, \dots, A_t . Let $c_i = |C_i|$, so that $|A_i \cup C_i| = c_i + 4$. Note that c_i is either 0 or ≥ 3 , as $c_i = 2$ would correspond to a coloured pair that does not intersect any other coloured pair, but by definition these pairs belong to D , not C . By the complete intersection theorem, the number of sets within $A_i \cup C_i$ is at most $f(c_i)$, defined to be $\binom{c_i+2}{2}$ for $c_i \geq 5$, 17 for $c_i = 4$, 15 for $c_i = 3$, 1 for $c_i = 0$. Now given $|C| = c = \sum_{i=1}^t c_i$, we claim that $\sum_{i=1}^t f(c_i) \leq f(c) + t - 1$, with equality holding when $c_i = c$ for some i , and $c_j = 0$ otherwise. This follows from a variational argument, using the inequalities $f(a+1) + f(b-1) \geq f(a) + f(b)$ for $a \geq b \geq 4$ and $f(a+3) + f(0) \geq f(a) + f(3)$ for $a \geq 3$, which are easy to verify. Excluding the sets A_1, \dots, A_t , we conclude that the number of sets in \mathcal{F} meeting C is at most $f(c) - 1$.

Putting everything together we have $|\mathcal{F}| \leq 3\binom{2t}{2} - 2t + \max\{3td, 7d\} + 4e + f(c) - 1$, where $n = 4t + c + d + e$. For $t \geq 3$ we can write $|\mathcal{F}| \leq 3\binom{2t}{2} - 2t + 3t(n - 4t - c) + f(c) - 1$. This is a quadratic in c with positive coefficient of c^2 for $5 \leq c \leq n - 4t$, so in this range its maximum occurs at $c = 5$ or $c = n - 4t$. Furthermore, it is easy to see that the value at $c = 0$ is larger than at $c = 3, 4, 5$ (and $c = 2$ is impossible as no c_i equals 1 or 2). Therefore the overall maximum occurs at $c = 0$ or $c = n - 4t$, which gives the stated bound. The bound for $t = 2$ follows in the same way, replacing $3td$ by $7d$ in the upper bound for \mathcal{F} . \square

Proof of Theorem 1.1. Suppose that \mathcal{F} is a 4-uniform family on $[n]$ with no singleton intersection. If \mathcal{F} is intersecting then it is 2-intersecting, so we are done by the complete intersection theorem. Otherwise, suppose the maximum matching has size $t \geq 2$, so $n \geq 8$. We have an upper bound on $|\mathcal{F}|$ given in Theorem 3.2, and we claim that this is always less than $\binom{n-2}{2}$.

First we consider the case $t = 2$. When $n = 8$ we have $\binom{n-2}{2} = 15$ and $b_2(n) = 14$; when $n = 9$ we have $\binom{n-2}{2} = 21$ and $b_2(n) = 20$. For $n \geq 10$ we note that $\binom{n-2}{2} - \binom{n-1-2}{2} = n - 3$ and $b_2(n) - b_2(n-1) \leq \max\{7, n-6\} \leq n-3$, where we use the inequality $\max\{a, b\} - \max\{a', b'\} \leq \max\{a - a', b - b'\}$. Therefore $b_2(n) < \binom{n-2}{2}$ for all $n \geq 8$.

For general t , when $n = 4t$ we have $\binom{n-2}{2} - b_t(4t) = (2t-3)(t-1) > 0$. Also, for $n > 4t$ we have $\binom{n-2}{2} - \binom{n-1-2}{2} = n-3$ and $b_t(n) - b_t(n-1) \leq \max\{3t, n-4t+1\} \leq n-3$, so $b_t(n) < \binom{n-2}{2}$ for all $n \geq 4t$. \square

4 Three lemmas

Here we prove some lemmas that will be used in the next section. Our first lemma concerns a multicoloured version of our problem, in the sense of [12].

Lemma 4.1 *Suppose $\mathcal{F}_1, \dots, \mathcal{F}_c$ are k -uniform families on $[n]$ so that there is no $X \in \mathcal{F}_i, Y \in \mathcal{F}_j$ with $|X \cap Y| = 1$ for any $i \neq j$. Then $\sum |\mathcal{F}_i| \leq c \binom{n}{k-2} + \binom{n}{k}$.*

Proof. Let \mathcal{A} be the family of sets that occur in more than one \mathcal{F}_i and \mathcal{B} the family of sets that occur in exactly one \mathcal{F}_i . Then \mathcal{A} has no singleton intersection, so $|\mathcal{A}| \leq \binom{n}{k-2}$ by Theorem 1.3. Therefore

$$\sum |\mathcal{F}_i| \leq c|\mathcal{A}| + |\mathcal{B}| \leq c\binom{n}{k-2} + \binom{n}{k}. \quad \square$$

Remark. By analogy with [14] one might expect that the bound can be improved to $\max\{c\binom{n}{k-2}, \binom{n}{k}\}$, but we do not need such a bound here.

Next we have a lemma on matchings. The argument is similar to one given by Frankl [8] (Proposition 11.6). Here also, it should be possible to replace the summation with a maximum.

Lemma 4.2 *Suppose X and Y are disjoint sets with $|X| = x$, $|Y| = y$ and \mathcal{F} is a set system on $X \cup Y$ such that $|F \cap X| = s$, $|F \cap Y| = t$ for every $F \in \mathcal{F}$. If \mathcal{F} contains no matching of size m then $|\mathcal{F}| < m \left(\binom{x-1}{s-1} \binom{y}{t} + \binom{y-1}{t-1} \binom{x}{s} \right)$.*

Proof. We argue by induction on s, t, x, y . First we note that in the case $x \leq ms$ the number of possible intersections of a set F with X is $\binom{x}{s} \leq m \binom{x-1}{s-1}$, so trivially $|\mathcal{F}| \leq \binom{x}{s} \binom{y}{t} < m \left(\binom{x-1}{s-1} \binom{y}{t} + \binom{y-1}{t-1} \binom{x}{s} \right)$. Similarly we are done when $y \leq mt$. To complete the base of the induction, note that in the case $s = t = 1$ the system \mathcal{F} is a bipartite graph with no matching of size m , and it is easy to see (e.g. by König's theorem) that $|\mathcal{F}| < m \max\{x, y\} \leq m(x + y)$.

For the general case, we use the compression method of Erdős, Ko and Rado [5]. Define arbitrary linear orders $<_X$ on X and $<_Y$ on Y . Given $a, b \in X$ $a <_X b$ we define the ab -shift S_{ab} by $S_{ab}(\mathcal{F}) = \{S_{ab}(F) : F \in \mathcal{F}\}$, where $S_{ab}(F)$ is equal to $F' = F \setminus \{b\} \cup \{a\}$ if $F' \notin \mathcal{F}$, but equal to F if $F' \in \mathcal{F}$. The same definition applies for $a, b \in Y$ with $a <_Y b$. Clearly $|S_{ab}(\mathcal{F})| = |\mathcal{F}|$. A well-known easy property of the shift is that the maximum matching in $S_{ab}(\mathcal{F})$ is no larger than that in \mathcal{F} . Iterating these shifts will eventually produce a family which is invariant with respect to S_{ab} , for any $a, b \in X$ or $a, b \in Y$. We can assume that \mathcal{F} has this property.

Suppose, without loss of generality, that $s > 1$. Let a be the maximal element of X . Consider the systems $\mathcal{F}_0 = \{F : a \notin F \in \mathcal{F}\}$ and $\mathcal{F}_1 = \{F \setminus \{a\} : a \in F \in \mathcal{F}\}$ defined on $X \setminus \{a\} \cup Y$. Since \mathcal{F}_0 does not have a matching of size m we have $|\mathcal{F}_0| \leq m \left(\binom{x-2}{s-1} \binom{y}{t} + \binom{y-1}{t-1} \binom{x-1}{s} \right)$ by induction. Also \mathcal{F}_1 contains no matching of size m . For suppose F_1, \dots, F_m are disjoint sets in \mathcal{F}_1 . Each has $s-1$ points in X , so we can find distinct points a_1, \dots, a_m in $X \setminus \cup_{i=1}^m F_i$. Since \mathcal{F} is invariant with respect to ab -shifts with $a, b \in X$ it contains the sets $F_i \cup \{a_i\}$. However, these form a matching, so indeed \mathcal{F}_1 contains no matching of size m . Therefore $|\mathcal{F}_1| \leq m \left(\binom{x-2}{s-2} \binom{y}{t} + \binom{y-1}{t-1} \binom{x-1}{s-1} \right)$ by induction.

$$\text{We conclude that } |\mathcal{F}| = |\mathcal{F}_0| + |\mathcal{F}_1| \leq m \left(\binom{x-1}{s-1} \binom{y}{t} + \binom{y-1}{t-1} \binom{x}{s} \right). \quad \square$$

Finally, we give a simple optimisation lemma concerning sums of binomial coefficients.

Lemma 4.3 *Consider a function $f(z) = \sum_{j=1}^m c_j \binom{z+s_j}{t_j}$ where $c_j \geq 0$ and s_j, t_j are non-negative integers with $s_j \geq t_j - 1$ for all j . For any positive integers x_1, \dots, x_n , writing $x = \sum_{i=1}^n x_i$, we have $\sum_{i=1}^n f(x_i) \leq f(x) + (n-1)f(0)$.*

Proof. Note that $\binom{x_i+1+s_j}{t_j} + \binom{x_{i'}-1+s_j}{t_j} - \left(\binom{x_i+s_j}{t_j} + \binom{x_{i'}+s_j}{t_j} \right) = \binom{x_i+s_j}{t_j-1} - \binom{x_{i'}-1+s_j}{t_j-1} \geq 0$ if $x_i \geq x_{i'} - 1$. So starting from any sequence x_1, \dots, x_n we can move to the sequence $x, 0, \dots, 0$ without decreasing

the function $\sum_{i=1}^n f(x_i)$, and the final value gives the stated upper bound. \square

5 A stability result

In this section we prove Theorem 1.2, which states: for any $\epsilon > 0$ there is $\delta > 0$ such that if \mathcal{F} is a k -uniform family on $[n]$ with no singleton intersection and $|\mathcal{F}| \geq (1 - \delta) \binom{n-2}{k-2}$, then there are two points x, y so that all but at most ϵn^{k-2} sets of \mathcal{F} contain both x and y .

Proof of Theorem 1.2. Suppose that \mathcal{F} is a k -uniform family on $[n]$ with no singleton intersection, and $|\mathcal{F}| \geq (1 - \delta) \binom{n-2}{k-2}$. We can suppose in all estimates that δ is sufficiently small, and n is sufficiently large (by making δ small). Let A_1, \dots, A_t be a matching in \mathcal{F} with t as large as possible. If $t = 1$ then \mathcal{F} is intersecting, so 2-intersecting. As we mentioned in the introduction, a result of Frankl implies that there is a constant $c(k)$ such that if \mathcal{F} is 2-intersecting and $|\mathcal{F}| > c(k)n^{k-3}$ then there are two points x, y contained in every set of \mathcal{F} . Since $|\mathcal{F}| \geq (1 - \delta) \binom{n-2}{k-2} > c(k)n^{k-3}$ for large n , we are done in the case $t = 1$. Now suppose $t \geq 2$. Let $A = \cup_{i=1}^t A_i$, $B = [n] \setminus A$. Note that all sets in \mathcal{F} meet A , and if they meet any A_i they meet it in at least 2 points.

Let $\mathcal{F}' \subset \mathcal{F}$ be the family of sets meeting exactly one A_i , i.e.

$$\mathcal{F}' = \{F \in \mathcal{F} : \exists 1 \leq i(F) \leq t, F \cap A_{i(F)} \neq \emptyset, F \cap A_j = \emptyset \ \forall j \neq i(F)\}.$$

Let $\mathcal{G} = \{F \cap B : F \in \mathcal{F}'\}$. Say that $G \in \mathcal{G}$ has colour i if $G = F \cap B$ for some F that meets A_i . (A set can have more than one colour.) For $b \in B$ a ‘flower’ on b is a system $\{G_1, \dots, G_{k-2}\} \subset \mathcal{G}$, so that $G_i \cap G_j = \{b\}$ for every $i \neq j$. The key observation is that if there is a flower on b then there is a unique i so that all sets in \mathcal{G} containing b have colour i and no other colour. To see this first note that all the sets in the flower must have the same colour (say i), and no other, to avoid a singleton intersection. Now consider any $G \in \mathcal{G}$ that contains b . Then $|G| \leq k - 2$ so there are at most $k - 3$ sets in the flower that intersect G in a point other than b . Therefore we can find $1 \leq j \leq k - 2$ so that $G_j \cap G = b$, and so to avoid a singleton intersection G_j and G cannot have two different colours, i.e. both have only colour i .

Let X_i be the set of all b for which there is a flower of colour i on b . It follows from the above observation that X_1, \dots, X_t are pairwise disjoint. We also note for future reference that there are no two disjoint sets of \mathcal{F} contained in $A_i \cup X_i$ for any i ; otherwise we could use them instead of A_i to find a larger matching in \mathcal{F} . Since there are no singleton intersections the sets of \mathcal{F} contained in $A_i \cup X_i$ form a 2-intersecting family. Write $X = \cup_{i=1}^t X_i$, $x = |X|$, $x_i = |X_i|$, $Y = B \setminus X$, $y = |Y|$.

Estimate of $|\mathcal{F}'|$

(1) First we count sets corresponding to those elements of \mathcal{G} contained within X , so X_i for some i . By Theorem 1.3 $A_i \cup X_i$ contains at most $\binom{x_i+k}{k-2}$ sets. (In fact, we have noted that these sets form a 2-intersecting family, so we could even obtain a stronger bound from the complete intersection theorem mentioned in the introduction, but this expression will be more convenient.)

(2) Next we count sets corresponding to $\mathcal{J} = \{G : G \in \mathcal{G}, G \subset Y\}$. Let $\mathcal{J}^s = \{G : G \in \mathcal{J}, |G| = s\}$,

and partition $\mathcal{J}^s = \mathcal{J}_1^s \cup \mathcal{J}_2^s$, where \mathcal{J}_1^s contains those G with exactly one colour and \mathcal{J}_2^s those with more than one colour. Now \mathcal{J}_2^s has no singleton intersection, or there would be corresponding sets in \mathcal{F}' with singleton intersection, so $|\mathcal{J}_2^s| \leq \binom{y}{s-2}$ by Theorem 1.3. It follows that at most $t \binom{k}{s} \binom{y}{s-2}$ sets in \mathcal{F}' correspond to sets of \mathcal{J}_2^s .

Also, for each $a \in Y$, $s \geq 2$ the link $\mathcal{J}_1^s(a)$ is a $(s-1)$ -uniform system on Y with no matching of size $k-2$. This is immediate from the definition of X , as if $\mathcal{J}_1^s(a)$ has a matching of size $k-2$ then there is a flower on a , i.e. $a \in X$. By Lemma 4.2 $|\mathcal{J}_1^s(a)| \leq (k-2) \binom{y-1}{s-2}$. Therefore the number of sets in \mathcal{F}' corresponding to sets of \mathcal{J}_1^s is at most $\binom{k}{s} \sum_a |\mathcal{J}_1^s(a)| < \binom{k}{s} y (k-2) \binom{y-1}{s-2}$. (In fact we are even overestimating by an extra factor of s corresponding to the different choices of a in a set of \mathcal{J}_1^s .) For $s=1$ we clearly have $|\mathcal{J}_1^1| \leq y$, corresponding to at most ky sets of \mathcal{F}' .

In total, the number of sets in \mathcal{F}' corresponding to elements of \mathcal{J} is at most

$$\begin{aligned} ky + \sum_{s=2}^{k-2} \left(t \binom{k}{s} \binom{y}{s-2} + \binom{k}{s} y (k-2) \binom{y-1}{s-2} \right) &< ky + (t + ky) \binom{y+k}{k-4} \sum_{s=2}^{k-2} \binom{k}{s} \\ &< 2^k (ky + t) \binom{y+k}{k-4}. \end{aligned}$$

(3) Finally, consider those sets corresponding to \mathcal{K} , defined as those $G \in \mathcal{G}$ that meet both X and Y . Such a G is contained in $X_i \cup Y$ for some i , and has colour i but no other colour. For $2 \leq s \leq k-2$, let $\mathcal{K}_i^s = \{G : G \in \mathcal{K}, |G| = s, G \subset X_i \cup Y\}$. As in estimate (2), for each $a \in Y$, $s \geq 2$ the link $\mathcal{K}_i^s(a)$ is a $(s-1)$ -uniform system with no matching of size $k-2$. Considering each possible intersection size with X_i and Y separately, we apply Lemma 4.2 to get $|\mathcal{K}_i^s(a)| \leq \sum_{\alpha=1}^{s-2} (k-2) \left(\binom{x_i-1}{\alpha-1} \binom{y-1}{s-\alpha-1} + \binom{y-2}{s-\alpha-2} \binom{x_i}{\alpha} \right)$. Applying Lemma 4.3, we can bound the number of sets in \mathcal{F}' corresponding to elements of \mathcal{K} by

$$\begin{aligned} \sum_{i=1}^t \sum_{s=2}^{k-2} \sum_{a \in Y} \binom{k}{s} |\mathcal{K}_i^s(a)| &= \sum_{s=2}^{k-2} \binom{k}{s} \sum_{a \in Y} \sum_{i=1}^t |\mathcal{K}_i^s(a)| \\ &\leq \sum_{s=2}^{k-2} \binom{k}{s} y \sum_{i=1}^t \sum_{\alpha=1}^{s-2} (k-2) \left(\binom{x_i}{\alpha-1} \binom{y}{s-\alpha-1} + \binom{y}{s-\alpha-2} \binom{x_i}{\alpha} \right) \\ &\leq \sum_{s=2}^{k-2} \binom{k}{s} y (k-2) \sum_{\alpha=1}^{s-2} \left(\binom{x}{\alpha-1} \binom{y}{s-\alpha-1} + \binom{y}{s-\alpha-2} \binom{x}{\alpha} \right) \\ &\leq \sum_{s=2}^{k-2} \binom{k}{s} y (k-2) \cdot 2 \binom{x+y}{s-2} \\ &< 2^{k+1} ky \binom{x+y+k}{k-4}. \end{aligned}$$

Adding the estimates (1), (2) and (3) we have

$$|\mathcal{F}'| \leq \sum_{i=1}^t \binom{x_i+k}{k-2} + 2^k (ky + t) \binom{y+k}{k-4} + 2^{k+1} ky \binom{x+y+k}{k-4}.$$

Estimate of $|\mathcal{F} \setminus \mathcal{F}'|$

Suppose $2 \leq \alpha \leq t$ and $\beta = (\beta_1, \dots, \beta_\alpha)$ with $\beta_j \geq 2$ and $\beta^* = \sum_{j=1}^\alpha \beta_j \leq k$ are given. Let \mathcal{H}_β be the collection of all sets $H \subset \cup_i A_i$ such that the list $|H \cap A_i|$, $1 \leq i \leq t$ consists of β and $t - \alpha$ zeroes, in some order. Then $|\mathcal{H}_\beta| \leq \alpha! \binom{t}{\alpha} \prod_{j=1}^\alpha \binom{k}{\beta_j} < 2^{k^2} t^\alpha$, where we crudely estimate that each product term $\binom{k}{\beta_j}$ is at most 2^k and there are at most k terms (as $\beta^* \leq k$).

We can obtain a matching of size t in \mathcal{H}_β as follows. For $1 \leq i \leq t$ let A_i^1, \dots, A_i^α be disjoint subsets of A_i with $|A_i^j| = \beta_j$ for $1 \leq j \leq \alpha$. Let $M_\gamma = \cup_{j=1}^\alpha A_{\gamma+j}^j$ for $1 \leq \gamma \leq t$ where $A_{\gamma+j}$ is to be interpreted as $A_{\gamma+j-t}$ for $\gamma + j > t$. Then $\mathcal{M} = \{M_1, \dots, M_t\}$ is a matching in \mathcal{H}_β . Let $\mathcal{G}_\gamma = \{F \cap B : F \cap \cup_i A_i = M_\gamma\}$. Then \mathcal{G}_γ for $1 \leq \gamma \leq t$ are $(k - \beta^*)$ -uniform systems satisfying the hypothesis of Lemma 4.1, so have total size

$$\sum_{\gamma=1}^t |\mathcal{G}_\gamma| \leq t \binom{n - kt}{k - \beta^* - 2} + \binom{n - kt}{k - \beta^*}.$$

Now we average this estimate over all possible isomorphic choices of the matching \mathcal{M} in \mathcal{H}_β . Let m be the number of such matchings and m' be the number of such matchings that contain some fixed set $M \in \mathcal{H}_\beta$ (this is independent of M). By counting pairs (M, \mathcal{M}) where \mathcal{M} is a maximum matching containing a set M we see that $mt = |\mathcal{H}_\beta| m'$. Writing

$$\mathcal{F}_\beta = \{F \in \mathcal{F} : F \cap \cup_i A_i \in \mathcal{H}_\beta\},$$

we have (recalling that $\mathcal{F}_\beta(M)$ denotes the link of \mathcal{F}_β from M)

$$\begin{aligned} |\mathcal{F}_\beta| &= \sum_{M \in \mathcal{H}_\beta} |\mathcal{F}_\beta(M)| = \sum_{M \in \mathcal{H}_\beta} \frac{1}{m'} \sum_{\mathcal{M} \ni M} |\mathcal{F}_\beta(M)| \\ &= \frac{1}{m'} \sum_{\mathcal{M}} \sum_{M \in \mathcal{M}} |\mathcal{F}_\beta(M)| \leq \frac{m}{m'} \left(t \binom{n - kt}{k - \beta^* - 2} + \binom{n - kt}{k - \beta^*} \right) \\ &= |\mathcal{H}_\beta| \left(\binom{n - kt}{k - \beta^* - 2} + t^{-1} \binom{n - kt}{k - \beta^*} \right). \end{aligned}$$

Since $\beta^* = \sum_j \beta_j$ satisfies $2\alpha \leq \beta^* \leq k$ and $\alpha \geq 2$ we have

$$\begin{aligned} |\mathcal{F} \setminus \mathcal{F}'| &\leq \sum_{\alpha, \beta} |\mathcal{F}_\beta| \leq \sum_{\alpha, \beta} 2^{k^2} t^\alpha \left(\binom{n - kt}{k - \beta^* - 2} + t^{-1} \binom{n - kt}{k - \beta^*} \right) \\ &\leq 2^{k^2} \max_{\alpha} t^\alpha (n^{k-2\alpha-2} + t^{-1} n^{k-2\alpha}) \cdot \sum_{\alpha, \beta} 1 \\ &< 2^{3k^2} n^{k-3}, \end{aligned}$$

where we crudely estimate that there are at most $k^{k+1} < 2^{2k^2}$ ways to choose the numbers $\alpha, \beta_1, \dots, \beta_\alpha$.

Adding the estimates for $|\mathcal{F}'|$ and $|\mathcal{F} \setminus \mathcal{F}'|$ we obtain

$$|\mathcal{F}| \leq \sum_{i=1}^t \binom{x_i + k}{k - 2} + \delta \binom{n}{k - 2},$$

for n sufficiently large. By the hypothesis of the theorem this gives $\sum_{i=1}^t \binom{x_i+k}{k-2} \geq (1-2\delta)\binom{n-2}{k-2}$. Suppose, without loss of generality, that $x_1 \geq x_i$ for all i .

Now some routine calculations imply that $x_1 > (1-8\delta)n$. For the convenience of the reader we will give the details here, but the casual reader may skip to the last paragraph of the proof. Write $1/(r+1) < x_1/n \leq 1/r$ for some natural number r . It follows easily from Lemma 4.3 and induction that $\sum_{i=1}^t \binom{x_i+k}{k-2} \leq r\binom{n/r+k}{k-2} + 1_{t>r}(t-r)\binom{k}{k-2}$. This is less than $(1-2\delta)\binom{n-2}{k-2}$ if $r \geq 2$ (since $k \geq 4$) so we have $r = 1$. Now Lemma 4.3 gives

$$\sum_{i=1}^t \binom{x_i+k}{k-2} \leq \binom{x_1+k}{k-2} + \binom{x-x_1+k}{k-2} + (t-2)\binom{k}{k-2}.$$

From the identity $\binom{a+b}{c} = \sum_i \binom{a}{i} \binom{b}{c-i}$ we have

$$\binom{x_1+k}{k-2} + \binom{x-x_1+k}{k-2} \leq \binom{x+2k}{k-2} - (x-x_1+k)\binom{x_1+k}{k-3},$$

so

$$(1-3\delta)\binom{n-2}{k-2} \leq \binom{x+2k}{k-2} - (x-x_1+k)\binom{x_1+k}{k-3}.$$

In particular $(1-3\delta)\binom{n-2}{k-2} \leq \binom{x+2k}{k-2}$, so $x > (1-4\delta)n$. Also, since $\binom{x+2k}{k-2} < (1+\delta)\binom{n-2}{k-2}$, we must have $(x-x_1+k)\binom{x_1+k}{k-3} < 4\delta\binom{n-2}{k-2}$. Now $f(q) = (x-q+k)\binom{q+k}{k-3}$ is a concave function of q ; to see this, note that

$$\frac{f(q)^2}{f(q-1)f(q+1)} = \frac{(q+4)(q+k)(x-q+k)^2}{(q+3)(q+k+1)((x-q+k)^2-1)} > \frac{(q+4)(q+k)}{(q+3)(q+k+1)} > 1.$$

If $x_1 \leq (1-8\delta)n$, since $x_1 \geq n/2$, it follows that $f(x_1) \geq \min\{f(n/2), f((1-8\delta)n)\} > 4\delta\binom{n-2}{k-2}$, contradiction. Therefore $x_1 > (1-8\delta)n$, as claimed.

The number of sets of \mathcal{F} not contained in $A_1 \cup X_1$ is at most $\sum_{i=2}^t \binom{x_i+k}{k-2} + \delta\binom{n}{k-2} < \binom{8\delta n+k}{k-2} + (t-2)\binom{k}{k-2} + \delta\binom{n}{k-2} < \epsilon n^{k-2}$, for small δ . Also, the sets of \mathcal{F} contained in $A_1 \cup X_1$ form a 2-intersecting family, and as in the first paragraph of the proof, it follows that there are two points x, y such that every set in $A_1 \cup X_1$ contains both x and y . This completes the proof. \square

Remarks.

1. The proof shows not only that there are at most ϵn^{k-2} sets that do not contain both x and y but also that all such sets intersect a set $\overline{A_1 \cup X_1}$ of size at most say ϵn .

2. A more careful analysis of the argument gives a new proof of Frankl's result, and some numerical experiments indicate that the smallest n for which the proof works is considerably smaller than his value, perhaps $n = k^5$ will do, compared with $k^{\Theta(k)}$. We will not attempt to present these calculations here, as the main goal should be to prove the result for all n .

References

- [1] R. Ahlswede and L.H. Khachatrian, The complete intersection theorem for systems of finite sets, *European J. Combin.* **18** (1997), 125–136.
- [2] R. P. Anstee and P. Keevash, Pairwise intersections and forbidden configurations, submitted.
- [3] J. Balogh, B. Bollobás and M. Simonovits, The number of graphs without forbidden subgraphs, *J. Combin. Theory Ser. B* **91** (2004), 1–24.
- [4] P. Erdős, Problems and results in graph theory and combinatorial analysis, Proceedings of the Fifth British Combinatorial Conference (Univ. Aberdeen, Aberdeen, 1975), pp. 169–192, *Congressus Numerantium*, No. XV, Utilitas Math., Winnipeg, Man., 1976.
- [5] P. Erdős, C. Ko and R. Rado, Intersection theorems for systems of finite sets, *Quart. J. Math. Oxford Ser.* **12** (1961), 313–320.
- [6] P. Frankl, On families of finite sets no two of which intersect in a singleton, *Bull. Austral. Math. Soc.* **17** (1977), 125–134.
- [7] P. Frankl, On intersecting families of finite sets, *J. Combin. Theory Ser. A* **24** (1978), 146–161.
- [8] P. Frankl, **Extremal set systems**, Handbook of combinatorics, 1293–1329, Elsevier, Amsterdam, 1995.
- [9] P. Frankl, K. Ota, and N. Tokushige, Exponents of uniform L -systems, *J. Combin. Theory Ser. A* **75** (1996), 23–43.
- [10] P. Frankl and R. M. Wilson, Intersection theorems with geometric consequences, *Combinatorica* **1** (1981), 357–368.
- [11] A.J.W. Hilton and E.C. Milner, Some intersection theorems for systems of finite sets. *Quart. J. Math. Oxford Ser.* **18** (1967), 369–384.
- [12] P. Keevash, M. Saks, B. Sudakov and J. Verstraete, Multicolour Turan problems, *Advances in Applied Mathematics* **33** (2004), 238–262.
- [13] P. Keevash and B. Sudakov, Local density in graphs with forbidden subgraphs, *Combin. Probab. Comput.* **12** (2003), 139–153.
- [14] P. Keevash and B. Sudakov, Set systems with restricted cross-intersections and the minimum rank of inclusion matrices, *SIAM J. Discrete Math.* **18** (2005), 713–727.
- [15] P. Keevash and B. Sudakov, The Turán number of the Fano plane, *Combinatorica* **25** (2005), 561–574.
- [16] D. Mubayi, Erdős-Ko-Rado for three sets, *J. Combin. Theory Ser. A*, to appear.

- [17] D. Mubayi, Structure and stability of triangle-free set systems, *Trans. Amer. Math. Soc.*, to appear.
- [18] D. K. Ray-Chaudhuri and R. M. Wilson, On t -designs, *Osaka J. Math* **12** (1975), 735–744.
- [19] R. M. Wilson, The exact bound in the Erdős-Ko-Rado theorem, *Combinatorica* **4** (1984), 247–257.