Sperner's Theorem and a Problem of Erdős, Katona and Kleitman

SHAGNIK DAS[1], WENYING GAN[1] and BENNY SUDAKOV[2][†]

[1]Department of Mathematics, UCLA, Los Angeles, CA 90095, USA
(e-mail: shagnik@ucla.edu, wgan@math.ucla.edu)
[2]Department of Mathematics, ETH, 8092 Zürich, Switzerland
(e-mail: benjamin.sudakov@math.ethz.ch)

A central result in extremal set theory is the celebrated theorem of Sperner from 1928, which gives the size of the largest family of subsets of $[n]$ not containing a 2-chain, $F_1 \subset F_2$. Erdős extended this theorem to determine the largest family without a k-chain, $F_1 \subset F_2 \subset \cdots \subset F_k$. Erdős and Katona, followed by Kleitman, asked how many chains must appear in families with sizes larger than the corresponding extremal bounds.

In 1966, Kleitman resolved this question for 2-chains, showing that the number of such chains is minimized by taking sets as close to the middle level as possible. Moreover, he conjectured the extremal families were the same for k-chains, for all k. In this paper, making the first progress on this problem, we verify Kleitman's conjecture for the families whose size is at most the size of the $k + 1$ middle levels. We also characterize all extremal configurations.

2010 *Mathematics subject classification*: Primary 05D05

1. Introduction

Sperner's theorem is a central result in extremal set theory, giving the size of the largest family of sets not containing a 2-chain, $F_1 \subset F_2$. Erdős later extended this theorem to determine the largest family without a k-chain, $F_1 \subset F_2 \subset \cdots \subset F_k$. A natural question is to ask how many k-chains must appear in a family larger than this extremal bound.

More precisely, we consider the following problem, first posed by Erdős and Katona and then extended by Kleitman some fifty years ago. Given a family $\mathcal{F}$ of s subsets of $[n]$, how many k-chains must $\mathcal{F}$ contain? We denote this minimum by $c_k(n, s)$, and determine it for a wide range of values of s. This provides a quantitative strengthening of the Erdős result on the size of k-chain-free families.

We shall now discuss the background of Sperner's theorem and this problem further, before presenting our new results.

1.1. Background

Extremal set theory is one of the most rapidly developing areas in combinatorics, having applications to other branches of mathematics and computer science including discrete geometry, functional analysis, number theory and complexity. The typical extremal problem has the following form: How large can a structure be without containing some forbidden configuration? A classical example, considered by many to be the starting point of extremal set theory, is a theorem of Sperner [13]. An *antichain* is a family of subsets of $[n]$ that does not contain sets $F_1 \subset F_2$. Sperner's theorem states that the largest antichain has $\binom{n}{\lfloor n/2 \rfloor}$ sets, a bound that is easily seen to be tight by considering the family of sets of size $\lfloor n/2 \rfloor$. This celebrated result enjoys numerous applications and has many extensions, many of which are discussed in Engel's book [4]. One particular extension, due to Erdős [5], shows that the size of the largest set family without a *k-chain*, that is, k-sets $F_1 \subset F_2 \subset \cdots \subset F_k$, is the sum of the $k-1$ largest binomial coefficients,

$$M_{k-1} = \sum_{i=\lceil (n-k+2)/2 \rceil}^{\lceil (n+k-2)/2 \rceil} \binom{n}{i}.$$

When $k = 2$, we recover Sperner's theorem.

Our problem is what we refer to as an Erdős–Rademacher-type extension of Erdős's theorem, a name we now explain. Arguably the most well-known result in extremal combinatorics is a theorem of Mantel [11] from 1907, which states that an n-vertex triangle-free graph can have at most $\lfloor n^2/4 \rfloor$ edges. In an unpublished result, Rademacher strengthened this theorem by showing that any graph with $\lfloor n^2/4 \rfloor + 1$ edges must contain at least $\lfloor n/2 \rfloor$ triangles. Erdős [6] then extended this to graphs with a linear number of extra edges, and in [7] studied the problem for larger cliques. More generally, for any extremal problem, the corresponding Erdős–Rademacher problem asks how many copies of the forbidden configuration must appear in a structure larger than the extremal bound.

In the context of Sperner's theorem, this problem was first considered by Erdős and Katona, who conjectured that a family with $\binom{n}{\lfloor n/2 \rfloor} + t$ sets must contain at least $t \lceil (n+1)/2 \rceil$ 2-chains. Kleitman [10] confirmed the conjecture, and, in a far-reaching generalization, showed that the minimum number of 2-chains in a family of any fixed size is obtained by choosing sets of size as close to $n/2$ as possible. He then conjectured (see [8, 10]) that the same families minimize the number of k-chains, a problem that has remained open for nearly fifty years.

Conjecture 1.1. *The number of k-chains in a family is minimized by choosing sets of sizes as close to n/2 as possible.*

1.2. Our results

In this paper we study these Erdős–Rademacher-type extensions of the theorems of Sperner and Erdős. We began by considering the case of 2-chains, and determined the

minimum number of 2-chains in a family of any number of sets. Later, we discovered Kleitman had earlier obtained the same result. However, through slightly more careful calculations, and by introducing an additional argument, we are able to characterize all extremal families, as given below.

Theorem 1.2. *Let $\mathcal{F}$ be a family of subsets of $[n]$, with $|\mathcal{F}| = s \geqslant \binom{n}{\lfloor n/2 \rfloor}$. Let $r \in \frac{1}{2}\mathbb{N}$ be the unique half-integer such that*

$$\sum_{i=n/2-r+1}^{n/2+r-1} \binom{n}{i} < s \leqslant \sum_{i=n/2-r}^{n/2+r} \binom{n}{i}.$$

Then $\mathcal{F}$ minimizes the number of 2-chains if and only if the following conditions are satisfied.

(1) *For every $F \in \mathcal{F}$, $n/2 - r \leqslant |F| \leqslant n/2 + r$.*
(2) *For any $A \subset [n]$ with $n/2 - r + 1 \leqslant |A| \leqslant n/2 + r - 1$, we have $A \in \mathcal{F}$.*
(3) *If*

$$s \leqslant \sum_{i=n/2-r}^{n/2+r-1} \binom{n}{i},$$

then $\{F \in \mathcal{F} : |F| = n/2 \pm r\}$ forms an antichain.
(4) *If*

$$s \geqslant \sum_{i=n/2-r}^{n/2+r-1} \binom{n}{i},$$

then $\{F \notin \mathcal{F} : |F| = n/2 \pm r\}$ forms an antichain.

Our main results verify Conjecture 1.1 for families of certain sizes. To begin with, recall that Erdős showed that the largest family without k-chains consists of the $k-1$ middle levels of the hypercube, whose size we denote by M_{k-1}. If we were to add one set to this family, the best we could do would be to add it to the kth level, in which case we would create

$$\binom{\lfloor (n+k)/2 \rfloor}{k-1}(k-1)!$$

k-chains. Indeed, we show that every additional set must contribute at least this many new k-chains, and the above construction shows that this is tight when our extremal family is contained within the k middle levels.

Theorem 1.3. *If $\mathcal{F}$ is a set family over $[n]$ of size $s = M_{k-1} + t$, then $\mathcal{F}$ contains at least*

$$t\binom{\lfloor (n+k)/2 \rfloor}{k-1}(k-1)!$$

k-chains.

We are then able to extend our argument to work for larger set families, obtaining a result that is tight when the extremal family is contained within the $k + 1$ middle levels.

Theorem 1.4. *Provided $n \geqslant 15$ and $k \leqslant n - 6$, if $\mathcal{F}$ is a set family over $[n]$ of size $s = M_k + t$, then the number of k-chains in $\mathcal{F}$ is at least*

$$\binom{n}{\lceil (n-k)/2 \rceil} \binom{\lfloor (n+k)/2 \rfloor}{k-1}(k-1)!$$
$$+ t \left(\binom{\lceil (n+k)/2 \rceil}{k-1} + \binom{\lceil (n+k)/2 \rceil}{k} \binom{k}{2} \right)(k-1)!.$$

In both cases, we actually obtain stronger results (see Theorems 3.1 and 4.2 respectively), providing stability versions of the above theorems, showing that if a family has close to the minimum number of k-chains, it must be close in structure to the extremal example. These stability results are of interest even in the case $k = 2$, as one does not obtain any stability from the Kleitman proof for 2-chains. We then use the stability results to show that when the above bounds are tight, the extremal families are exactly as in Theorem 1.2.

1.3. Outline and notation

The remainder of this paper is organized as follows. Section 2 contains a proof of Theorem 1.2. In Section 3 we prove Theorem 1.3, and then in Section 4 we prove Theorem 1.4. In the final section we present some concluding remarks and open problems. The Appendix contains the proof of a technical proposition needed for Theorem 1.2.

We let $[n]$ denote the set of the first n integers. For a ground set X and an integer i, we denote the family of i-subsets of X by

$$\binom{X}{i} = \{Y \subseteq X : |Y| = i\}.$$

We let

$$M_k = \sum_{i=\lceil n-k+1/2 \rceil}^{\lceil n+k-1/2 \rceil} \binom{n}{i}$$

be the size of the k middle, and thus largest, levels. Given a family $\mathcal{F}$ of subsets of $[n]$, we let

$$\mathcal{F}_i = \mathcal{F} \cap \binom{[n]}{i}$$

denote those sets in $\mathcal{F}$ of size i. The ℓ-shadow of a family is given by

$$\partial^{\ell} \mathcal{F} = \{G : \exists F \in \mathcal{F}, G \subset F, |G| = |F| - \ell\}.$$

For a subset $F \subset [n]$, we define $m(F) = \max\{|F|, n - |F|\}$.

Given a set family $\mathcal{F}$, $c_k(\mathcal{F})$ denotes the number of k-chains in $\mathcal{F}$. For any $n \in \mathbb{N}$ and $0 \leqslant s \leqslant 2^n$, we let $c_k(n, s)$ denote the minimum of $c_k(\mathcal{F})$ over all families $\mathcal{F}$ of s subsets of $[n]$. When $k = 2$, if we have two families $\mathcal{F}$ and $\mathcal{G}$, then we let $c_2(\mathcal{F}, \mathcal{G})$ denote the number of 2-chains with one set from $\mathcal{F}$ and one set from $\mathcal{G}$.

2. Counting 2-chains

In this section we will prove Theorem 1.2, characterizing those families that minimize the number of 2-chains. We essentially show that it is optimal to take sets of sizes as close to $n/2$ as possible. The theorem then prescribes how the boundary sets can be distributed.

Sperner's theorem shows that the largest antichain is given by one of the middle levels, that is, either all sets of size $\lfloor n/2 \rfloor$ or all sets of size $\lceil n/2 \rceil$. Obviously, an antichain minimizes the number of 2-chains, as it has none. This theorem is then a natural extension of Sperner's theorem, as it shows that to construct a family of any size that minimizes the number of 2-chains, one should start by taking sets of size $\lfloor n/2 \rfloor$, then sets of size $\lfloor n/2 \rfloor + 1$, then $\lfloor n/2 \rfloor - 1$, and so on until one has a family of the desired size. As we shall show, these families are optimal, and so we may denote the number of 2-chains in the first s such sets by $c_2(n, s)$.

The idea behind the proof is as follows. If our family $\mathcal{F}$ contains a set F that is too far away from the middle (*i.e.*, $||F| - n/2| > r$), then we will show that we can shift F closer to the middle and decrease the number of 2-chains. Once we have our family contained in the $2r + 1$ middle layers, a simple counting argument will give the characterization of extremal families. As the shifting process is essentially the same as in Kleitman's proof in [10], we relegate the proof of the following proposition to the Appendix.

Proposition 2.1. *Let $\mathcal{F}$ be a family of $s > \binom{n}{\lfloor n/2 \rfloor}$ subsets of $[n]$ minimizing the number of 2-chains. If $A \in \mathcal{F}$ is of maximal cardinality, with $|A| = n/2 + m$, then for any $B \subset A$, $|B| \geqslant n/2 - m + 1$, we have $B \in \mathcal{F}$.*

Assuming this proposition, we shall proceed to prove Theorem 1.2.

Proof of Theorem 1.2. We prove the theorem by induction on s.

For the base case, we take $s = \binom{n}{\lfloor n/2 \rfloor}$. By Sperner's theorem, it follows that any family $\mathcal{F}$ of this size that minimizes the number of 2-chains must be an antichain. It is well known that the only antichains of this size are the family of all sets of size $\lfloor n/2 \rfloor$, or the family of sets of size $\lceil n/2 \rceil$. It is easy to see that these families are the only ones satisfying properties (1)–(4), with $r = 0$ or $1/2$ depending on whether n is even or odd respectively.

For the induction step, assume $s > \binom{n}{\lfloor n/2 \rfloor}$, and let $\mathcal{F}$ be an optimal family of size s. Suppose property (1) were not satisfied. Since $\mathcal{F}$ and $\mathcal{F}' = \{[n] \setminus F : F \in \mathcal{F}\}$ have the same number of 2-chains, we may assume there is a largest set $F \in \mathcal{F}$ with $|F| = n/2 + t$ for some $t > r$. By Proposition 2.1, it follows that for every $G \subset F$, $|G| \geqslant n/2 - t + 1$, we have $G \in \mathcal{F}$. Hence F is in at least

$$\sum_{i=1}^{2t-1} \binom{n/2 + t}{i}$$

2-chains in $\mathcal{F}$. Since $\mathcal{F} \setminus \{F\}$ is a family of $s - 1$ sets, there are at least $c_2(n, s - 1)$ 2-chains in $\mathcal{F}$ not involving F. Thus the number of 2-chains in $\mathcal{F}$ is at least

$$c_2(n, s-1) + \sum_{i=1}^{2t-1} \binom{n/2 + t}{i} > c_2(n, s-1) + \sum_{i=1}^{2r} \binom{n/2 + r}{i} \geqslant c_2(n, s),$$

and so $\mathcal{F}$ cannot be optimal, giving a contradiction. Hence if $\mathcal{F}$ is optimal, each $F \in \mathcal{F}$ has $n/2 - r \leqslant |F| \leqslant n/2 + r$, and so property (1) is established.

Now consider the case

$$\sum_{i=n/2-r+1}^{n/2+r-1} \binom{n}{i} < s \leqslant \sum_{i=n/2-r}^{n/2+r-1} \binom{n}{i}.$$

Since

$$s > \sum_{i=n/2-r+1}^{n/2+r-1} \binom{n}{i},$$

and in light of property (1), it follows that there exists some $F \in \mathcal{F}$ with $|F| = n/2 \pm r$; by symmetry, we may assume $|F| = n/2 + r$. By Proposition 2.1, F must be contained in

$$\sum_{i=1}^{2r-1} \binom{n/2+r}{i} = c_2(n, s) - c_2(n, s - 1)$$

2-chains with sets in $\mathcal{F}$ of sizes between $n/2 - r + 1$ and $n/2 + r - 1$. If F is contained in any 2-chains with sets of size $n/2 - r$, then by induction it follows that $\mathcal{F}$ has more than $c_2(n, s)$ 2-chains, contradicting the optimality of $\mathcal{F}$. Thus F is incomparable to the other sets in $\mathcal{F}$ of sizes $n/2 \pm r$. Removing F, we find that $\mathcal{F} \setminus \{F\}$ must also be optimal, and thus properties (2) and (3) follow.

Finally, suppose

$$\sum_{i=n/2-r}^{n/2+r-1} \binom{n}{i} \leqslant s \leqslant \sum_{i=n/2-r}^{n/2+r} \binom{n}{i}.$$

By property (1), we know all sets in $\mathcal{F}$ have sizes between $n/2 - r$ and $n/2 + r$. Let

$$\mathcal{H} = \bigcup_{i=n/2-r}^{n/2+r} \binom{[n]}{i}$$

be the family of all subsets of $[n]$ of sizes between $n/2 - r$ and $n/2 + r$, and let $\mathcal{G} = \mathcal{H} \setminus \mathcal{F}$ be those sets not in $\mathcal{F}$. We have $c_2(\mathcal{F}) = c_2(\mathcal{H}) - c_2(\mathcal{G}, \mathcal{H}) + c_2(\mathcal{G})$. Here $c_2(\mathcal{H})$ depends only on r, and hence on s, and is independent of the structure of $\mathcal{F}$. We have

$$c_2(\mathcal{G}, \mathcal{H}) = \sum_{G \in \mathcal{G}} c_2(\{G\}, \mathcal{H}) = \sum_{G \in \mathcal{G}} \left(\sum_{i=n/2-r}^{|G|-1} \binom{|G|}{i} + \sum_{i=|G|+1}^{n/2+r} \binom{n-|G|}{i-|G|} \right).$$

The parenthetical term is maximized when $|G| = n/2 \pm r$, and so $c_2(\mathcal{G}, \mathcal{H})$ is maximized when for every $G \in \mathcal{G}$ we have $|G| = n/2 \pm r$. Finally, $c_2(\mathcal{G})$ is minimized when $\mathcal{G}$ is an antichain, in which case $c_2(\mathcal{G}) = 0$. Both of these conditions are satisfied by the construction outlined at the beginning of this section, and hence must also be true of any other extremal family. Thus to minimize $c_2(\mathcal{F})$, $\mathcal{F}$ must contain all sets of sizes between

$n/2 - r + 1$ and $n/2 + r - 1$, and

$$\mathcal{G} = \binom{[n]}{n/2 + r} \cup \binom{[n]}{n/2 - r} \setminus \mathcal{F}$$

should be an antichain, establishing properties (2) and (4). This completes the induction step, and with it the proof of Theorem 1.2. $\square$

3. Counting k-chains

We now seek a similar result for k-chains, and thus to make some progress on Conjecture 1.1. In this section we verify the conjecture when the number of sets is at most that in the k middle levels, and in the next section we shall extend the result to the $k + 1$ middle levels.

If we take all M_{k-1} sets in the $k - 1$ middle levels, that is of sizes between $\lfloor (n-k)/2 \rfloor + 1$ and $\lfloor (n+k)/2 \rfloor - 1$, and then add t sets of size $\lfloor (n+k)/2 \rfloor$, we would create precisely

$$t \binom{\lfloor (n+k)/2 \rfloor}{k - 1}(k - 1)!$$

k-chains. Hence Theorem 1.3 is tight when

$$0 \leqslant t \leqslant \binom{n}{\lfloor (n+k)/2 \rfloor}.$$

We shall in fact prove the following stronger theorem, which provides a stability result. In the following notation, we let $r \in \{(k-1)/2, k/2\}$ be such that the sets in the k middle levels have sizes between $n/2 - r$ and $n/2 + r$, and for a set F, we define $m(F) = \max\{|F|, n - |F|\}$.

Theorem 3.1. *Let $\mathcal{F}$ be a family of subsets of $[n]$. Then the number of k-chains in $\mathcal{F}$ is bounded by*

$$c_k(\mathcal{F}) \geqslant \left(\sum_{i=n/2-r}^{n/2+r} \frac{|\mathcal{F}_i|}{\binom{n}{i}} - (k - 1) \right) \binom{n}{n/2 + r} \binom{n/2 + r}{k - 1}(k - 1)!$$

$$+ \sum_{\substack{F \in \mathcal{F}: \\ ||F| - n/2| \geqslant r+1}} \binom{m(F)}{k - 1}(k - 1)!.$$

Proof. We prove the theorem by induction on $|\mathcal{F}|$. If $|\mathcal{F}| = 0$, then there is nothing to show, as the desired lower bound is negative.

For the induction step, we begin by noting that for every set $F \in \mathcal{F}$ with $||F| - n/2| \geqslant r$, F can be in at most $\binom{m(F)}{k-1}(k - 1)!$ k-chains. If not, then we could remove F, and applying the inductive hypothesis to $\mathcal{F} \setminus \{F\}$, we would have the desired inequality.

We now use an LYM-type inequality, counting the number of k-chains in our family by considering permutations. We say that a permutation $\sigma \in S_n$ *contains* a set $F \subset [n]$, denoted $F \in \sigma$, if $\{\sigma(1), \sigma(2), \ldots, \sigma(|F|)\} = F$; that is, F is an initial segment of σ. Note that if σ contains k sets $F_1, F_2, \ldots, F_k$, then those k sets must form a k-chain. For any

set $F \subset [n]$, we let $S_n[F] = \{\sigma \in S_n : F \in \sigma\}$. Since every permutation containing m sets contributes $m - \binom{m}{k} \leqslant k - 1$ to the right-hand side of the sum below, it follows that

$$(k-1)n! \geqslant \sum_{F \in \mathcal{F}} |S_n[F]| - \sum_{F_1 \subset F_2 \subset \cdots \subset F_k \in \mathcal{F}} |\cap_{i=1}^k S_n[F_i]|. \tag{3.1}$$

As the second sum is over all k-chains in our family $\mathcal{F}$, this inequality will allow us to bound the number of k-chains. Note that for any $F \in \mathcal{F}$, we have $|S_n[F]| = |F|!|[n] \setminus F|!$, and for a k-chain $F_1 \subset F_2 \subset \cdots \subset F_k$,

$$|\cap_{i=1}^k S_n[F_i]| = |F_1|! \prod_{i=1}^{k-1} |F_{i+1} \setminus F_i|! |[n] \setminus F_k|!$$

gives the number of permutations containing the k-chain.

We shall associate every k-chain in $\mathcal{F}$ with either its minimum or maximum set, depending on which is further away from the middle level. For $F \in \mathcal{F}$ with $|F| < n/2$, let $\mathcal{C}(F) = \{F \subset F_2 \subset \cdots \subset F_k : |F| + |F_k| < n\}$, and if $F \in \mathcal{F}$ with $|F| \geqslant n/2$, let

$$\mathcal{C}(F) = \{F_1 \subset \cdots \subset F_{k-1} \subset F : |F_1| + |F| \geqslant n\},$$

and, for convenience, define $C(F) = |\mathcal{C}(F)|$. Note that we have partitioned the set of k-chains $\{F_1 \subset F_2 \subset \cdots \subset F_k : F_i \in \mathcal{F}\}$ into the disjoint sets $\{\mathcal{C}(F)\}_{F \in \mathcal{F}}$. We can thus rewrite inequality (3.1) as follows:

$$(k-1)n! \geqslant \sum_{F \in \mathcal{F}} \left[|S_n[F]| - \sum_{F_1 \subset F_2 \subset \cdots \subset F_k \in \mathcal{C}(F)} |\cap_{i=1}^k S_n[F_i]| \right]. \tag{3.2}$$

To bound $|\cap_{i=1}^k S_n[F_i]|$ appropriately, we require that $\emptyset$ and $[n]$ not be members of our family. This is given by the following lemma.

Lemma 3.2. *If $|\mathcal{F}| \leqslant M_{n-1}$, and $\mathcal{F}$ minimizes the number of k-chains, then $\emptyset \notin \mathcal{F}$ and $[n] \notin \mathcal{F}$.*

Proof. Suppose we had $[n] \in \mathcal{F}$. Since $|F| \leqslant 2^n - 2$, there must be some $\emptyset \neq F \notin \mathcal{F}$. We decrease the number of k-chains in $\mathcal{F}$ by replacing $[n]$ with F, since any new k-chain involving F was a k-chain with $[n]$ before.

Similarly, if $\emptyset \in \mathcal{F}$, we can replace it with any set $[n] \neq F \notin \mathcal{F}$.   □

Note that $M_{n-1} = 2^n - 2$. Thus, if $|\mathcal{F}| > M_{n-1}$, then either we have all subsets of $[n]$, or our family is missing just one set, in which case (as we explained above) it is best to remove either $\emptyset$ or $[n]$. In either case, the bound in Theorem 3.1 remains true.

We now assume $1 \leqslant |F| \leqslant n - 1$ for all $F \in \mathcal{F}$. If we fix $F_1, |F_1| < n/2$, then we maximize

$$|\cap_{i=1}^k S_n[F_i]| = |F_1|! \prod_{i=1}^{k-1} |F_{i+1} \setminus F_i|! |[n] \setminus F_k|!$$

by taking $|F_{i+1} \setminus F_i| = 1$ for $1 \leqslant i \leqslant k - 1$, and so

$$|\cap_{i=1}^k S_n[F_i]| \leqslant |F_1|!(n - |F_1| - (k-1))!.$$

The same holds true if we instead fix $F_k, |F_k| \geqslant n/2$, and thus

$$|\cap_{i=1}^{k} S_n[F_i]| \leqslant (|F_k| - (k-1))!(n - |F_k|)!.$$

We can unify both bounds in the form

$$\frac{|F|!(n - |F|)!}{\binom{m(F)}{k-1}(k-1)!}.$$

Moreover, by definition we must have $\mathcal{C}(F) = \emptyset$ for any F with $||F| - n/2| \leqslant r - 1$. Hence we split our sum based on how $||F| - n/2|$ compares to r. Dividing through by $n!$, inequality (3.2) leads to

$$k - 1 \geqslant \sum_{F \in \mathcal{F}} \frac{1}{\binom{n}{|F|}} \left(1 - \frac{C(F)}{\binom{m(F)}{k-1}(k-1)!}\right) = \Sigma_1 + \Sigma_2 + \Sigma_3,$$

where

$$\Sigma_1 = \sum_{\substack{F \in \mathcal{F} \\ ||F|-n/2| \leqslant r-1}} \frac{1}{\binom{n}{|F|}} = \sum_{i=n/2-r+1}^{n/2+r-1} \frac{|\mathcal{F}_i|}{\binom{n}{i}},$$

$$\Sigma_2 = \sum_{\substack{F \in \mathcal{F} \\ ||F|-n/2|=r}} \frac{1}{\binom{n}{n/2+r}} \left(1 - \frac{C(F)}{\binom{n/2+r}{k-1}(k-1)!}\right)$$

$$= \frac{|\mathcal{F}_{n/2-r}| + |\mathcal{F}_{n/2+r}|}{\binom{n}{n/2+r}} - \frac{\sum_{\substack{F \in \mathcal{F} \\ ||F|-n/2|=r}} C(F)}{\binom{n}{n/2+r}\binom{n/2+r}{k-1}(k-1)!},$$

and

$$\Sigma_3 = \sum_{\substack{F \in \mathcal{F} \\ ||F|-n/2| \geqslant r+1}} \frac{1}{\binom{n}{|F|}} \left(1 - \frac{C(F)}{\binom{m(F)}{k-1}}\right).$$

Note that

$$\sum_{||F|-n/2|=r} C(F) = c_k(\mathcal{F}) - \sum_{||F|-n/2| \geqslant r+1} C(F),$$

and so, substituting in Σ_2, we obtain

$$\frac{c_k(\mathcal{F})}{\binom{n}{n/2+r}\binom{n/2+r}{k-1}(k-1)!} + k - 1 \geqslant \sum_{i=n/2-r}^{n/2+r} \frac{|\mathcal{F}_i|}{\binom{n}{i}}$$

$$+ \sum_{\substack{F \in \mathcal{F}: \\ ||F|-n/2| \geqslant r+1}} \left(\frac{1}{\binom{n}{|F|}} - \frac{C(F)}{(k-1)!}\left(\frac{1}{\binom{n}{|F|}\binom{m(F)}{k-1}} - \frac{1}{\binom{n}{n/2+r}\binom{n/2+r}{k-1}}\right)\right).$$

Now, since

$$\binom{n}{a}\binom{a}{k-1} = \binom{n}{k-1}\binom{n-k+1}{a-k+1},$$

it follows that when $m(F) \geqslant n/2 + r$,

$$\binom{n}{|F|}\binom{m(F)}{k-1} \leqslant \binom{n}{n/2+r}\binom{n/2+r}{k-1},$$

as

$$\frac{1}{2}(n-k+1) \leqslant \frac{n}{2}+r-k+1 \leqslant m(F)-k+1.$$

Hence the summand is minimized when $C(F)$ is as large as possible, which, by our inductive hypothesis, is

$$\binom{m(F)}{k-1}(k-1)!.$$

Substituting this into the inequality above gives

$$\frac{c_k(\mathcal{F})}{\binom{n}{n/2+r}\binom{n/2+r}{k-1}(k-1)!} + k - 1 \geqslant \sum_{i=n/2-r}^{n/2+r} \frac{|\mathcal{F}_i|}{\binom{n}{i}} + \sum_{\substack{F \in \mathcal{F}: \\ ||F|-n/2|\geqslant r+1}} \frac{\binom{m(F)}{k-1}}{\binom{n}{n/2+r}\binom{n/2+r}{k-1}}.$$

Rearranging gives the desired bound

$$c_k(\mathcal{F}) \geqslant \left(\sum_{i=n/2-r}^{n/2+r} \frac{|\mathcal{F}_i|}{\binom{n}{i}} - (k-1) \right) \binom{n}{n/2+r}\binom{n/2+r}{k-1}(k-1)!$$

$$+ \sum_{\substack{F \in \mathcal{F}: \\ ||F|-n/2|\geqslant r+1}} \binom{m(F)}{k-1}(k-1)!. \qquad \square$$

Given Theorem 3.1, it is easy to deduce Theorem 1.3. In fact, we are able to characterize all extremal families.

Proof of Theorem 1.3. Suppose we have a family of sets $\mathcal{F}$, with $|\mathcal{F}| = M_{k-1} + t$. Note that the contribution each set $F \in \mathcal{F}$ makes to the right-hand side above is

$$\frac{\binom{n}{n/2+r}}{\binom{n}{|F|}}\binom{n/2+r}{k-1}(k-1)!$$

if $n/2 - r \leqslant |F| \leqslant n/2 + r$, and $\binom{m(F)}{k-1}(k-1)!$ otherwise. This contribution increases with $||F| - n/2|$, and so to minimize the right-hand side we need all sets to satisfy $||F| - n/2| \leqslant r$. Moreover, since the binomial coefficients $\binom{n}{i}$ are minimized over $n/2 - r \leqslant i \leqslant n/2 + r$ when $i = n/2 \pm r$, any extremal family must contain all sets of sizes between $n/2 - r + 1$ and $n/2 + r - 1$, with the remaining sets having size $n/2 \pm r$. It is easy to see that such a collection of sets gives

$$c_k(\mathcal{F}) \geqslant t\binom{n/2+r}{k-1}(k-1)!$$

above, as required.

To classify the extremal families, note that we already know we must have all sets of sizes between $n/2 - r + 1$ and $n/2 + r - 1$. If $r = k/2$, this consists of the middle $k - 1$ levels, giving M_{k-1} sets. Hence we have t sets of size $n/2 \pm r$. To obtain equality in (3.1), we must have every chain passing through either $k - 1$ or k sets of $\mathcal{F}$. As all the sets in the $k - 1$ middle levels are in $\mathcal{F}$, the chain can contain at most one set from $\mathcal{F}$ of size $n/2 \pm r$. From this, we deduce that $\{F \in \mathcal{F} : |F| = n/2 \pm r\}$ must be an antichain.

If $r = (k - 1)/2$, we know the middle $k - 2$ levels are full, and we have $\binom{n}{n/2+r} + t$ sets of size $n/2 \pm r$. In order to obtain equality in (3.1), we must therefore have every chain pass through at least one set in $\mathcal{F}$ of size $n/2 \pm r$. Hence $\{G \notin \mathcal{F} : |G| = n/2 \pm r\}$ must form an antichain.

In particular, we note that the extremal families are exactly the same as for Theorem 1.2. $\qquad \square$

We remark that Theorem 3.1 is a stability result for Theorem 1.3, as our bound on $c_k(\mathcal{F})$ increases if we are missing sets with $||F| - n/2| \leqslant r - 1$, or have sets with $||F| - n/2| \geqslant r + 1$. Moreover, the stability estimates we obtain can also be tight. For instance, if we replace ℓ sets of size $n/2 \pm r$ with sets of size $n/2 + r + 1$, Theorem 3.1, together with some simple computations, shows that we should gain at least $\ell \binom{n/2+r}{k-2}(k-1)!$ extra k-chains. Moreover, it is easy to check that we gain precisely that many k-chains in the case when our family includes all of the $(n/2 + r)$-sets and none of the $(n/2 - r)$-sets in the shadow of the $(n/2 + r + 1)$-sets which were added.

Similarly, if we replace ℓ sets of size $n/2 + r - 1$ with sets of size $n/2 + r$, the theorem shows that we must gain at least

$$\frac{2r - 1}{n/2 + r} \ell \binom{n/2 + r}{k - 1}(k - 1)!$$

extra k-chains. This is tight again, if our family includes all the sets of size $n/2 + r$ containing any of the replaced sets. As we remarked earlier, these stability results are new even in the case $k = 2$.

4. Larger families

While Theorem 1.3 provides a tight bound on the number of k-chains appearing in families contained within the k middle levels, it underestimates the number of k-chains appearing in larger families. This is because in our calculations we assumed every k-chain had steps $(F_i \setminus F_{i-1})$ of size 1, as this maximizes the number of permutations containing the k-chain. However, when we are working with the $k + 1$ middle levels, we also have k-chains with a larger step of size 2, and so we shall have to make our argument more robust in order to handle these chains.

However, there is one additional difficulty. Recall that the number of permutations containing a k-chain $F_1 \subset F_2 \subset \cdots \subset F_k$ is given by

$$|\cap_{i=1}^{k} S_n[F_i]| = |F_1|! \prod_{i=1}^{k-1} |F_{i+1} \setminus F_i|!(n - |F_k|)!.$$

If we fix F_k, say, then we would hope that for k-chains involving a larger step, the largest this can be is to have one step of size 2 and have all the other steps have size 1, as this is precisely the type of k-chain that appears in our extremal families. Such k-chains are in $2(|F_k| - k)!(n - |F_k|)!$ permutations.

Unfortunately, a chain with $k - 2$ steps of size 1 and one step of size $|F_k| - k + 1$ (so that $|F_1| = 1$) is contained in $(|F_k| - k + 1)!(n - |F_k|)!$ permutations, which is larger than the bound we require. However, recall that we assign k-chains to either F_1 or F_k, depending on which is further from the middle. Thus, unless $|F_k| = n - 1$, we would assign the above k-chain to F_1, and so we would be fixing F_1 and not F_k. Hence the one case we need to avoid is having a k-chain starting with a set of size 1 and ending with a set of size $n - 1$.

We shall later provide a separate argument to show that there cannot be any sets of size $n - 1$ in an extremal family, thus bypassing this problem. In the meantime, for the purposes of our stability results, we shall assume there are no sets of size $n - 1$. We now require two arguments: one to bound the number of k-chains with larger steps, and one to bound the total number of k-chains.

We introduce the following notation for the remainder of this section. Given n and k, we let $a = \lceil (n + k)/2 \rceil$, so that the k middle levels are those sets of sizes between $a - k$ and $a - 1$, and the $(k + 1)$st middle level has sets of size a. For a set family $\mathcal{F}$, we let $\mathcal{C}(\mathcal{F})$ be the set of k-chains in $\mathcal{F}$. We partition these into two subsets: $\mathcal{C}_1(\mathcal{F})$ are those k-chains $F_1 \subset F_2 \subset \cdots \subset F_k$ with $|F_{i+1} \setminus F_i| = 1$ for all $1 \leqslant i \leqslant k - 1$, and $\mathcal{C}_2(\mathcal{F}) = \mathcal{C}(\mathcal{F}) \setminus \mathcal{C}_1(\mathcal{F})$ those k-chains with a larger step. We let $C(\mathcal{F})$, $C_1(\mathcal{F})$ and $C_2(\mathcal{F})$ denote the number of k-chains in these subsets respectively. As in Theorem 1.3, we will again identify a k-chain with one of its endpoints F_1 or F_k, depending which is further from the middle level, giving the partition $\{\mathcal{C}(F)\}_{F \in \mathcal{F}}$ of $\mathcal{C}(\mathcal{F})$. These sets will again be partitioned into $\mathcal{C}_1(F)$ and $\mathcal{C}_2(F)$, depending on whether or not the k-chains have a step of size at least 2. Finally, $C(F)$, $C_1(F)$ and $C_2(F)$ represent the sizes of the corresponding sets of k-chains.

4.1. Counting k-chains with large steps

We begin by showing that large families must contain a number of k-chains with a step of size at least 2. The following proposition also provides some stability, which we shall require to show that an extremal family cannot contain any sets of size $n - 1$.

Proposition 4.1. *Let $\mathcal{F}$ be a set family of size $|\mathcal{F}| = M_k + t_1$, with $1 \leqslant |F| \leqslant n - 2$ for all $F \in \mathcal{F}$, and with at least t_2 sets missing from the middle $k - 1$ levels. Then*

$$C_2(\mathcal{F}) \geqslant \left(t_1 + \left(\frac{k-1}{a} \right) t_2 \right) \binom{a}{k} \binom{k}{2} (k-1)!.$$

Proof. We prove the statement by induction on $t_1 + t_2 \geqslant 0$, noting that we must have $t_2 \geqslant 0$. The base case of $t_1 + t_2 = 0$ is trivial, as in this case the right-hand side is non-positive.

The proof will now run along very similar lines to that of Theorem 1.3, and we shall just make a few changes to count only those chains with a large step. To begin with, when we are counting sets and k-chains in permutations, we only want to consider those k-chains

with a large step. To ensure this, we shall not count k-chains that appear consecutively in some permutation. That is, if $\sigma \in S_n$ contains the sets $F_1 \subset F_2 \subset \cdots \subset F_s$ for some $s \geqslant k+1$, we will count $F_1 \subset F_2 \subset \cdots \subset F_{k-1} \subset F_{k+1}$, but not $F_1 \subset F_2 \subset \cdots \subset F_k$. Thus every k-chain we consider is bound to have some step of size at least 2. If $s \geqslant k$, then the number of such chains is $\binom{s}{k} - (s - (k-1))$, and since $(k-1)s - (\binom{s}{k} - (s - (k-1))) \leqslant k^2 - k$, it follows that

$$(k^2 - k)n! \geqslant (k-1) \sum_{F \in \mathcal{F}} |S_n[F]| - \sum_{F_1 \subset \cdots \subset F_k \in \mathcal{C}_2(\mathcal{F})} |\cap_{i=1}^{k} S_n[F_i]|$$

$$\geqslant \sum_{F \in \mathcal{F}} \left((k-1)|S_n[F]| - \sum_{F_1 \subset \cdots \subset F_k \in \mathcal{C}_2(F)} |\cap_{i=1}^{k} S_n[F_i]| \right).$$

As before, we now seek to maximize the terms $|\cap_{i=1}^{k} S_n[F_i]|$. Provided we have $1 \leqslant |F| \leqslant n-2$ for all sets $F \in \mathcal{F}$, if we fix one of the endpoints of the chain, the number of permutations it is contained in is maximized when we have one step of size 2, and all the other steps of size 1. Thus we can bound $|\cap_{i=1}^{k} S_n[F_i]|$ by $2|F_1|!(n - |F_1| - k)!$ or $2(|F_k| - k)!(n - |F_k|)!$. Dividing through by $n!$, we have that

$$k^2 - k \geqslant \sum_{F \in \mathcal{F}} \frac{1}{\binom{n}{|F|}} \left(k - 1 - \frac{C_2(F)}{\binom{m(F)}{k} \frac{k!}{2}} \right).$$

Now, by definition, we must have $C_2(F) = 0$ for all sets F in the k middle levels; that is, with $a - k \leqslant |F| \leqslant a - 1$. Let $\hat{\mathcal{F}} = \{F \in \mathcal{F} : |F| \leqslant a - k - 1 \text{ or } |F| \geqslant a\}$ be those sets outside the middle k levels. Thus

$$k^2 - k \geqslant \sum_{i=a-k}^{a-1} \frac{(k-1)|\mathcal{F}_i|}{\binom{n}{i}} + \sum_{F \in \hat{\mathcal{F}}} \frac{1}{\binom{n}{|F|}} \left(k - 1 - \frac{C_2(F)}{\binom{m(F)}{k} \frac{k!}{2}} \right).$$

We may assume that for every set $F \in \hat{\mathcal{F}}$,

$$C_2(F) \leqslant \binom{a}{k}\binom{k}{2}(k-1)! = (k-1)\binom{a}{k}\frac{k!}{2},$$

since otherwise we may remove F from $\mathcal{F}$ and we are then done by induction. Hence, since $m(F) \geqslant a$ for all $F \in \hat{\mathcal{F}}$, the parenthetical term in the second sum is always non-negative, and so the right-hand side is minimized by replacing $\binom{n}{|F|}$ by $\binom{n}{a}$, giving

$$k^2 - k \geqslant \sum_{i=a-k}^{a-1} \frac{(k-1)|\mathcal{F}_i|}{\binom{n}{i}} + \sum_{F \in \hat{\mathcal{F}}} \frac{1}{\binom{n}{a}} \left(k - 1 - \frac{C_2(F)}{\binom{m(F)}{k} \frac{k!}{2}} \right)$$

$$\geqslant \sum_{i=a-k}^{a-1} \frac{(k-1)|\mathcal{F}_i|}{\binom{n}{i}} + \sum_{F \in \hat{\mathcal{F}}} \frac{1}{\binom{n}{a}} \left(k - 1 - \frac{C_2(F)}{\binom{a}{k} \frac{k!}{2}} \right)$$

$$= \sum_{i=a-k}^{a-1} \frac{(k-1)|\mathcal{F}_i|}{\binom{n}{i}} + \frac{(k-1)|\hat{\mathcal{F}}|}{\binom{n}{a}} - \frac{C_2(\mathcal{F})}{\binom{n}{a}\binom{a}{k} \frac{k!}{2}},$$

and so

$$\frac{C_2(\mathcal{F})}{\binom{n}{a}\binom{a}{k}\frac{k!}{2}} \geqslant \sum_{i=a-k}^{a-1} \frac{(k-1)|\mathcal{F}_i|}{\binom{n}{i}} - (k^2 - k) + \frac{(k-1)|\hat{\mathcal{F}}|}{\binom{n}{a}}.$$

If the middle k levels were full, then the first sum would equal $k^2 - k$. Since we must have at least t_2 sets missing from the middle $k-1$ levels, the right-hand side is minimized when there are exactly t_2 sets missing, all of size $a-1$. In this case, $|\hat{\mathcal{F}}| = t_1 + t_2$, giving

$$\frac{C_2(\mathcal{F})}{\binom{n}{a}\binom{a}{k}\frac{k!}{2}} \geqslant \frac{(k-1)t_1}{\binom{n}{a}} + (k-1)t_2\left(\frac{1}{\binom{n}{a}} - \frac{1}{\binom{n}{a-1}}\right) = \left(t_1 + \left(\frac{2a-n-1}{a}\right)t_2\right)\frac{k-1}{\binom{n}{a}}.$$

As $a = \lceil (n+k)/2 \rceil$, we have $2a - n - 1 \geqslant k - 1$, and so multiplying through by

$$\binom{n}{a}\binom{a}{k}\frac{k!}{2}$$

gives the desired bound. $\qquad \square$

4.2. Counting all k-chains

As Proposition 4.1 offers us some control over the number of chains with large steps, we can now proceed to bound the total number of k-chains in $\mathcal{F}$. Again, our result provides some stability, as we shall need to forbid sets of size $n-1$.

Theorem 4.2. *Let $\mathcal{F}$ be a set family of size $|\mathcal{F}| = M_k + t_1$, with $1 \leqslant |F| \leqslant n - 2$ for all $F \in \mathcal{F}$, and with at least t_2 sets missing from the middle $k - 1$ levels. Then*

$$C(\mathcal{F}) \geqslant \binom{n}{a-k}\binom{n-a+k}{k-1}(k-1)!$$
$$+ \left(t_1 + \left(\frac{k-1}{a}\right)t_2\right)\left(\binom{a}{k-1} + \binom{a}{k}\binom{k}{2}\right)(k-1)!.$$

Proof. We prove the theorem by induction on $t_1 + t_2 \geqslant 0$, and again must have $t_2 \geqslant 0$. The base case of $t_1 + t_2 = 0$ follows from Theorem 1.3, as when $|\mathcal{F}| = M_k = M_{k-1} + \binom{n}{a-k}$, the right-hand side above is less than the lower bound for $t = \binom{n}{a-k}$ in Theorem 1.3.

We may now assume that any set F not in the middle $k-1$ levels is contained in at most

$$\left(\binom{a}{k-1} + \binom{a}{k}\binom{k}{2}\right)(k-1)!$$

k-chains. If not, then we may remove F from $\mathcal{F}$, thus decreasing t_1 by 1. Applying the inductive hypothesis to $\mathcal{F} \setminus \{F\}$ and adding the k-chains involving F then gives the requisite number of k-chains.

We once again seek to bound the number of k-chains in our family by counting sets and k-chains in permutations, except this time we shall consider all k-chains appearing in the permutations. A permutation with s sets gives rise to $\binom{s}{k}$ k-chains, and since for all s

we have $ks - \binom{s}{k} \leqslant k^2 - 1$, it follows that

$$(k^2 - 1)n! \geqslant k \sum_{F \in \mathcal{F}} |S_n[F]| - \sum_{F_1 \subset \cdots \subset F_k \in \mathcal{C}(\mathcal{F})} |\cap_{i=1}^{k} S_n[F_i]| \tag{4.1}$$

$$= \sum_{F \in \mathcal{F}} \left(k|S_n[F]| - \sum_{F_1 \subset \cdots \subset F_k \in \mathcal{C}(F)} |\cap_{i=1}^{k} S_n[F_i]| \right).$$

To maximize $|\cap_{i=1}^{k} S_n[F_i]|$, since we have no sets of size $n - 1$, we should again take all the gaps to be as small as possible. Those chains in $\mathcal{C}_1(\mathcal{F})$ all have steps of size 1, while those in $\mathcal{C}_2(\mathcal{F})$ should have one step of size 2, and the rest of size 1. Dividing by $n!$ gives

$$k^2 - 1 \geqslant \sum_{F \in \mathcal{F}} \frac{1}{\binom{n}{|F|}} \left(k - \frac{C_1(F)}{\binom{m(F)}{k-1}(k-1)!} - \frac{C_2(F)}{\binom{m(F)}{k} \frac{k!}{2}} \right).$$

By definition, if $a - k + 1 \leqslant |F| \leqslant a - 1$, we must have $C_1(F) = C_2(F) = 0$, and if $|F| = a - k$, then $C_2(F) = 0$. Thus we have three types of k-chains to consider: those in $\mathcal{C}_1(F)$ for $|F| = a - k$, those in $\mathcal{C}_1(F)$ for $F \in \hat{\mathcal{F}} = \{F \in \mathcal{F} : |F| \leqslant a - k - 1 \text{ or } |F| \geqslant a\}$, and those in $\mathcal{C}_2(F)$ for $F \in \hat{\mathcal{F}}$. Splitting our sums thus, we obtain

$$k^2 - 1 \geqslant \sum_{i=a-k}^{a-1} \frac{k|\mathcal{F}_i|}{\binom{n}{i}} - \frac{\sum_{|F|=a-k} C_1(F)}{\binom{n}{a-k}\binom{n-a+k}{k-1}(k-1)!} + \sum_{F \in \hat{\mathcal{F}}} \frac{1}{\binom{n}{|F|}} \left(1 - \frac{C_1(F)}{\binom{m(F)}{k-1}(k-1)!} \right)$$

$$+ \sum_{F \in \hat{\mathcal{F}}} \frac{1}{\binom{n}{|F|}} \left(k - 1 - \frac{C_2(F)}{\binom{m(F)}{k} \frac{k!}{2}} \right).$$

Since

$$\sum_{|F|=a-k} C_1(F) = C(\mathcal{F}) - \sum_{F \in \hat{\mathcal{F}}} C_1(F) - \sum_{F \in \hat{\mathcal{F}}} C_2(F),$$

we can substitute this expression into the second sum, and redistribute, to obtain

$$\frac{C(\mathcal{F})}{\binom{n}{a-k}\binom{n-a+k}{k-1}(k-1)!} \geqslant \sum_{i=a-k}^{a-1} \frac{k|\mathcal{F}_i|}{\binom{n}{i}} - (k^2 - 1) + \Sigma_1 + \Sigma_2,$$

where

$$\Sigma_1 = \sum_{F \in \hat{\mathcal{F}}} \frac{1}{\binom{n}{|F|}} - \frac{C_1(F)}{\binom{n}{|F|}\binom{m(F)}{k-1}(k-1)!} + \frac{C_1(F)}{\binom{n}{a-k}\binom{n-a+k}{k-1}(k-1)!}, \quad \text{and}$$

$$\Sigma_2 = \sum_{F \in \hat{\mathcal{F}}} \frac{1}{\binom{n}{|F|}} \left(k - 1 - \frac{C_2(F)}{\binom{m(F)}{k} \frac{k!}{2}} \right) + \frac{C_2(F)}{\binom{n}{a-k}\binom{n-a+k}{k-1}(k-1)!}.$$

The following lemmas, whose proofs we defer to the end of this subsection, allow us to bound these sums.

Lemma 4.3. *For every $F \in \hat{\mathcal{F}}$, we have*

$$\frac{1}{\binom{n}{|F|}} - \frac{C_1(F)}{\binom{n}{|F|}\binom{m(F)}{k-1}(k-1)!} + \frac{C_1(F)}{\binom{n}{a-k}\binom{n-a+k}{k-1}(k-1)!} \geqslant \frac{\binom{a}{k-1}}{\binom{n}{a-k}\binom{n-a+k}{k-1}}.$$

Lemma 4.4. *For every $F \in \hat{\mathcal{F}}$, we have*

$$\frac{1}{\binom{n}{|F|}}\left(k - 1 - \frac{C_2(F)}{\binom{m(F)}{k}\frac{k!}{2}}\right) \geqslant \frac{1}{\binom{n}{a}}\left(k - 1 - \frac{C_2(F)}{\binom{a}{k}\frac{k!}{2}}\right).$$

We now replace our summands with these lower bounds, obtaining

$$\frac{C(\mathcal{F})}{\binom{n}{a-k}\binom{n-a+k}{k-1}(k-1)!} \geqslant \sum_{i=a-k}^{a-1} \frac{k|\mathcal{F}_i|}{\binom{n}{i}} - (k^2 - 1) + \sum_{F \in \hat{\mathcal{F}}} \frac{\binom{a}{k-1}}{\binom{n}{a-k}\binom{n-a+k}{k-1}}$$

$$+ \sum_{F \in \hat{\mathcal{F}}}\left(\frac{k-1}{\binom{n}{a}} + C_2(F)\left(\frac{1}{\binom{n}{a-k}\binom{n-a+k}{k-1}(k-1)!} - \frac{1}{\binom{n}{a}\binom{a}{k}\frac{k!}{2}}\right)\right)$$

$$= \sum_{i=a-k}^{a-1} \frac{k|\mathcal{F}_i|}{\binom{n}{i}} - (k^2 - 1) + |\hat{\mathcal{F}}|\left(\frac{\binom{a}{k-1}}{\binom{n}{a-k}\binom{n-a+k}{k-1}} + \frac{k-1}{\binom{n}{a}}\right)$$

$$+ \frac{C_2(\mathcal{F})}{\binom{n}{a-k}\binom{n-a+k}{k-1}(k-1)!}\left(1 - \frac{2}{n-a+1}\right).$$

We can use Proposition 4.1 to lower-bound $C_2(\mathcal{F})$. Moreover, as $a \geqslant n - a + k$, it follows that each set in $\hat{\mathcal{F}}$, whose size is not a, has greater weight than any set in the middle k levels. Thus, the right-hand side is minimized when we fill the middle $k - 1$ levels as much as possible. If we were to have the full k middle levels, the first sum would be equal to k^2. However, as we must have at least t_2 sets missing from the middle $k - 1$ levels, it is best to have exactly t_2 sets of size $a - 1$ missing, resulting in

$$\frac{C(\mathcal{F})}{\binom{n}{a-k}\binom{n-a+k}{k-1}(k-1)!} \geqslant 1 - \frac{kt_2}{\binom{n}{a-1}} + (t_1 + t_2)\left(\frac{\binom{a}{k-1}}{\binom{n}{a-k}\binom{n-a+k}{k-1}} + \frac{k-1}{\binom{n}{a}}\right)$$

$$+ \frac{\left(t_1 + \binom{k-1}{a}t_2\right)\binom{a}{k}\binom{k}{2}(k-1)!}{\binom{n}{a-k}\binom{n-a+k}{k-1}(k-1)!}\left(1 - \frac{2}{n-a+1}\right)$$

$$= 1 + A_1 t_1 + A_2 t_2,$$

where, after simplifying the binomial expressions, we find

$$A_1 = \frac{\binom{a}{k-1}}{\binom{n}{a-k}\binom{n-a+k}{k-1}} + \frac{\binom{a}{k}\binom{k}{2}}{\binom{n}{a-k}\binom{n-a+k}{k-1}} + \frac{k-1}{\binom{n}{a}} - \frac{2\binom{a}{k}\binom{k}{2}}{(n-a+1)\binom{n}{a-k}\binom{n-a+k}{k-1}}$$

$$= \frac{\binom{a}{k-1} + \binom{a}{k}\binom{k}{2}}{\binom{n}{a-k}\binom{n-a+k}{k-1}}, \quad \text{and}$$

$$A_2 = \frac{\binom{a}{k-1}}{\binom{n}{a-k}\binom{n-a+k}{k-1}} + \frac{(k-1)\binom{a}{k}\binom{k}{2}}{a\binom{n}{a-k}\binom{n-a+k}{k-1}} + \frac{k-1}{\binom{n}{a}} - \frac{k}{\binom{n}{a-1}} - \frac{2(k-1)\binom{a}{k}\binom{k}{2}}{a(n-a+1)\binom{n}{a-k}\binom{n-a+k}{k-1}}$$

$$= \frac{k-1}{a}\left(\frac{\binom{a}{k-1} + \binom{a}{k}\binom{k}{2}}{\binom{n}{a-k}\binom{n-a+k}{k-1}} + \frac{2a - n - k}{\binom{n}{a}}\right) \geqslant \frac{k-1}{a}\frac{\binom{a}{k-1} + \binom{a}{k}\binom{k}{2}}{\binom{n}{a-k}\binom{n-a+k}{k-1}}.$$

Multiplying through by

$$\binom{n}{a-k}\binom{n-a+k}{k-1}(k-1)!$$

gives the desired bound. ∎

To complete the proof, we now prove the two lemmas.

Proof of Lemma 4.3. Note that

$$\binom{n}{|F|}\binom{m(F)}{k-1} = \binom{n}{k-1}\binom{n-k+1}{m(F)-k+1},$$

and, by the same token,

$$\binom{n}{a-k}\binom{n-a+k}{k-1} = \binom{n}{k-1}\binom{n-k+1}{a-k}.$$

Since $a - k = \lfloor (n-k+1)/2 \rfloor$, and, as $F \in \hat{\mathcal{F}}$, we have $m(F) \geqslant a$, it follows that

$$\binom{n}{|F|}\binom{m(F)}{k-1} \leqslant \binom{n}{a-k}\binom{n-a+k}{k-1}.$$

Thus the left-hand side of the inequality is minimized when we choose $C_1(F)$ as large as possible. By definition,

$$C_1(F) \leqslant \binom{m(F)}{k-1}(k-1)!.$$

Making this substitution gives

$$\frac{1}{\binom{n}{|F|}} - \frac{C_1(F)}{\binom{n}{|F|}\binom{m(F)}{k-1}(k-1)!} + \frac{C_1(F)}{\binom{n}{a-k}\binom{n-a+k}{k-1}(k-1)!} \geqslant \frac{\binom{m(F)}{k-1}}{\binom{n}{a-k}\binom{n-a+k}{k-1}} \geqslant \frac{\binom{a}{k-1}}{\binom{n}{a-k}\binom{n-a+k}{k-1}}.$$

∎

Proof of Lemma 4.4. If $m(F) = a$, then we have equality, so we may assume $m(F) \geqslant a+1$. By induction, we can assume that no set is in more than

$$\left(\binom{a}{k-1} + \binom{a}{k}\binom{k}{2}\right)(k-1)!$$

k-chains, giving a bound on $C_2(F)$. Thus

$$C_2(F) \leqslant \left(\binom{a}{k-1} + \binom{a}{k}\binom{k}{2}\right)(k-1)! \leqslant \binom{a+1}{k}\binom{k}{2}(k-1)! \leqslant (k-1)\binom{m(F)}{k}\frac{k!}{2},$$

and so the factor

$$\left(k - 1 - \frac{C_2(F)}{\binom{m(F)}{k}\frac{k!}{2}}\right)$$

is non-negative. As $\binom{n}{a} \geqslant \binom{n}{|F|}$, we thus have

$$\frac{1}{\binom{n}{|F|}}\left(k - 1 - \frac{C_2(F)}{\binom{m(F)}{k}\frac{k!}{2}}\right) \geqslant \frac{1}{\binom{n}{a}}\left(k - 1 - \frac{C_2(F)}{\binom{m(F)}{k}\frac{k!}{2}}\right) \geqslant \frac{1}{\binom{n}{a}}\left(k - 1 - \frac{C_2(F)}{\binom{a}{k}\frac{k!}{2}}\right).$$

$\square$

4.3. Forbidding large sets

Given the previous theorem, all that remains is to show that an extremal family cannot contain sets of size $n - 1$. The idea behind this is as follows. A set of size $n - 1$ has a very large shadow in the $k - 1$ middle levels. In order for this set to not contain too many k-chains, we must therefore be missing a lot of sets in the $k - 1$ middle levels. By Theorem 4.2, it then follows that the remainder of the family must contain many more k-chains than it ought to. The relevant calculations are given below.

Proposition 4.5. *Suppose $n \geqslant 15$ and $k \leqslant n - 6$, and let $\mathcal{F}$ be a set family with $|F| = M_k + t$. If we do not have $1 \leqslant |F| \leqslant n - 2$ for all $F \in \mathcal{F}$, then*

$$c_k(\mathcal{F}) > \binom{n}{a-k}\binom{n-a+k}{k-1}(k-1)! + t\left(\binom{a}{k-1} + \binom{a}{k}\binom{k}{2}\right)(k-1)!.$$

Proof. The same proof as in Lemma 3.2 shows that we may assume we do not have $\emptyset$ or $[n]$ in $\mathcal{F}$. Hence it suffices to show there are no sets of size $n - 1$ in our family.

Suppose towards contradiction we had some set $F \in \mathcal{F}$ with $|F| = n - 1$. We may, by induction, assume that F is in at most

$$\left(\binom{a}{k-1} + \binom{a}{k}\binom{k}{2}\right)(k-1)!$$

k-chains.

Since F contains $\binom{n-1}{a-1}$ sets of size $a - 1$, there are

$$\binom{n-1}{a-1}\binom{a-1}{k-2}(k-2)!$$

possible k-chains that F might be in which consist of $k - 1$ sets from the $k - 1$ middle levels followed by F. Hence we must be missing a lot of sets from the $k - 1$ middle levels to prevent F from being in too many k-chains. The sets of size $a - 1$ are contained in the most such k-chains, so if we are missing t_2 sets, we must have

$$\left(\binom{n-1}{a-1} - t_2\right)\binom{a-1}{k-2}(k-2)! \leqslant \left(\binom{a}{k-1} + \binom{a}{k}\binom{k}{2}\right)(k-1)!.$$

Solving for t_2 gives

$$t_2 \geqslant \binom{n-1}{a-1} - a\left(1 + \frac{(a-k+1)(k-1)}{2}\right).$$

We now remove from $\mathcal{F}$ all sets of size $n - 1$, thus losing at most n sets, and apply Theorem 4.2 with the above value of t_2. In the theorem, the number of k-chains is

governed by the expression

$$\left(t_1 + \left(\frac{k-1}{a} \right) t_2 \right).$$

We are decreasing t_1 by at most n, but increasing t_2 by at least

$$\binom{n-1}{a-1} - a \left(1 + \frac{(a-k+1)(k-1)}{2} \right),$$

resulting in a net gain in the previous expression of at least

$$\frac{k-1}{a} \left(\binom{n-1}{a-1} - a \left(1 + \frac{(a-k+1)(k-1)}{2} \right) \right) - n.$$

Since $n - k \geqslant 6$, we have $n - a \geqslant 3$. If $n - a$ is some constant, then the first term is at least cubic in n, while the term we are subtracting is quadratic, since in this case $a - k + 1$ will also be constant. One the other hand, if $n - a$ is large, the first term will be at least a large power of n, while the term we subtract is at most cubic in n. Given $n \geqslant 15$, some simple but tedious calculations show that in either case, having a set of size $n - 1$ increases the number of k-chains our family must contain. $\qquad \square$

Note that the condition $k \leqslant n - 6$ is near-optimal, since if $k = n - 3$ then $M_k = 2^n - 2n - 2$, and so by volume considerations alone there must be extremal families with sets of size at least $n - 1$.

Theorem 1.4 now follows easily.

Proof of Theorem 1.4. Since $n \geqslant 15$ and $k \leqslant n - 6$, Proposition 4.5 shows that $1 \leqslant |F| \leqslant n - 2$ for all $F \in \mathcal{F}$. We may then apply Theorem 4.2 with $t_1 = t$ and $t_2 = 0$ to obtain the bound

$$C(\mathcal{F}) \geqslant \binom{n}{a-k} \binom{n-a+k}{k-1} (k-1)! + t \left(\binom{a}{k-1} + \binom{a}{k} \binom{k}{2} \right) (k-1)!.$$

Recalling that $a = \lceil (n+k)/2 \rceil$, this is precisely the desired lower bound.

We can again deduce a characterization of the extremal families. Note that we must have $t_2 = 0$ for the above bound to hold, and so the $k - 1$ middle levels must be full. In order to have equality in Lemma 4.4, we also needed $m(F) = a$ for all $F \in \hat{\mathcal{F}}$. If $a = (n+k)/2$, then the remaining $\binom{n}{a} + t$ sets must have size $n/2 \pm k/2$. To obtain equality in (4.1), every chain must pass through either k or $k + 1$ sets of $\mathcal{F}$, and so we must have $\{ G \notin \mathcal{F} : |G| = n/2 \pm k/2 \}$, forming an antichain.

If, on the other hand, $a = (n + k + 1)/2$, then sets of size $a = n/2 + (k+1)/2$ carry greater weight than sets of size $a - k = n/2 - (k-1)/2$. We can then redefine t_2 above to be the number of sets missing in the k middle levels and obtain the same result. Hence it follows that we must have all sets in the k middle levels, with the remaining sets in $\hat{\mathcal{F}}$ of size $n/2 \pm (k+1)/2$. In order to maintain equality in (4.1), every chain must pass through at most one set $F \in \mathcal{F}$ with $|F| = n/2 \pm (k+1)/2$, and so $\{ F \in \mathcal{F} : |F| = n/2 \pm (k+1)/2 \}$ must be an antichain.

Thus, once again, the extremal families are exactly the same as those that minimize the number of 2-chains, as given by Theorem 1.2. □

5. Concluding remarks and open problems

In this paper, we have partially answered Kleitman's conjecture by showing that the families that minimize the number of 2-chains also minimize the number of k-chains when they occupy up to the $k + 1$ middle levels. While we strongly believe the conjecture is true in general, we suspect new ideas are needed to deal with larger families. As the number of levels grows with respect to k, the number of different types of chains – in terms of the sizes of the steps between sets – grows rapidly, and these would all need to be controlled to obtain a precise result. In this direction, though, the same methods we have used above can be applied to show the following: if we have integers $\alpha_1, \alpha_2, \ldots, \alpha_{k-1}$ with $\sum_i \alpha_i = \ell - 1$, then, provided $|\mathcal{F}| \leqslant M_\ell$ and the largest set in our family has size at most $n - \max_i \alpha_i$, the number of k-chains $F_1 \subset F_2 \subset \cdots \subset F_k$ with $|F_{i+1} \setminus F_i| \geqslant \alpha_i$ is minimized by taking sets in the middle ℓ levels.

Considering the case of 2-chains, our paper has focused on showing that a family with more than $\binom{n}{\lfloor n/2 \rfloor}$ sets must contain many 2-chains. A closely related problem is to determine whether such a family must have any sets contained in many 2-chains. This type of question has been studied before in other settings. For example, when one is considering the number of triangles in a graph, Erdős showed in [6] that any graph with $\lfloor n^2/4 \rfloor + 1$ edges must contain an edge in at least $n/6 + o(n)$ triangles. It is well known and easy to see that the hypercube, a graph whose vertices are subsets of $[n]$, with two vertices adjacent if they are comparable and differ in exactly one element, has independence number 2^{n-1}. Chung, Füredi, Graham and Seymour [1] proved any induced subgraph on $2^{n-1} + 1$ vertices contains a vertex of degree at least $(1/2 + o(1)) \log_2 n$. It is an open problem to determine whether or not this bound is tight (the corresponding upper bound is $O(\sqrt{n})$), and the answer to this question has ramifications in theoretical computer science.

In the context of Sperner's theorem the above problem has a negative answer, which may be surprising given the previous two examples. For convenience, let us assume $n = 2m + 1$ is odd, and consider the following set family. Let

$$\mathcal{F} = \{F : 1 \notin F, |F| = m\} \cup \{F : 1 \in F, |F| = m + 1\}.$$

This family contains

$$2\binom{2m}{m} = \left(1 + \frac{1}{n}\right)\binom{n}{\lfloor n/2 \rfloor}$$

sets, and so we are indeed beyond the Sperner bound. However, it is easy to see that the only pairs of comparable sets are of the form $\{F, \{1\} \cup F\}$ for every $F \in \mathcal{F}$ with $|F| = m$. Hence each set of the family is in only one pair of comparable sets. In fact, for this family we have $c_2(\mathcal{F}) = c_2(n, |\mathcal{F}|)$, so it is possible to have an extremal family with the comparable pairs distributed as evenly as possible. Theorem 1.2 shows that any family with $2\binom{2m}{m} + 1$ sets must contain a set in at least two 2-chains. It is an open problem as

to whether this is also the largest family without a set that contains two other sets (and hence is the maximum set in two 2-chains). This configuration is known as a 2-fork, and the upper bound, which can be obtained using similar arguments as in Theorem 1.3, is

$$\left(1 + \frac{2}{n}\right)\binom{n}{\lfloor n/2 \rfloor},$$

as shown by Katona and Tarján [9].

We find most exciting the prospect of studying Erdős–Rademacher-type problems in other settings. Within the context of Sperner's theorem, a paper of Qian, Engel and Xu [12] studied an extension for multiset families, where the same set may be chosen multiple times. In [2], we derive an Erdős–Rademacher-type strengthening of the Erdős–Ko–Rado theorem. However, as one can investigate similar extensions for any extremal result, there is truly no end to the number of directions in which this project can be continued. We hope that further work of this nature will lead to many interesting results and a greater understanding of classical theorems in extremal combinatorics.

Note added in proof. During the preparation of this manuscript, it came to our attention that Dove, Griggs, Kang and Sereni [3] have independently obtained Theorem 1.3.

Appendix: The shifting proposition

In this Appendix we prove Proposition 2.1, which enables us to perform the shifting necessary for Theorem 1.2. As mentioned in Section 2, this is essentially the same shifting argument used in the original proof of Kleitman in [10]. We provide the proof here, as the details of the calculations are not included in Kleitman's paper.

Proof of Proposition 2.1.   Suppose not. Note that we must have $m \geqslant 1$, otherwise there is nothing to prove. Let $\ell \leqslant 2m - 1$ be the minimal integer such that there exists a largest set of size $n/2 + m$ with a subset of size $n/2 + m - \ell$ that is not in the family. Let

$$\mathcal{A} = \left\{ A \in \mathcal{F} : |A| = \frac{n}{2} + m, \partial^{\ell} A \notin \mathcal{F} \right\},$$

and let $\mathcal{B} = \partial^{\ell} \mathcal{A} \setminus \mathcal{F}$. We can construct an auxiliary bipartite inclusion graph on $\mathcal{A} \cup \mathcal{B}$, with an edge (A, B) if and only if $B \subset A$.

Consider first the case where we have a matching $M : \mathcal{A} \to \mathcal{B}$, so that for every set $A \in \mathcal{A}$ there exists a set $M(A) \subset A$, $M(A) \notin \mathcal{F}$. We shift the family from $\mathcal{F}$ to $\tilde{\mathcal{F}}$ by replacing each set $A \in \mathcal{A}$ by $M(A) \in \mathcal{B}$, and claim that this reduces the number of 2-chains. Note that if $B = M(A)$ is a newly-introduced set, and $C \in \mathcal{F}$ is a set with $C \subset B$, then we must have had $C \subset A$ as well. Thus the only 2-chains that we need to consider are those between the levels $n/2 + m$ and $n/2 + m - \ell$; we call these *intermediate chains*.

Suppose $\ell > 1$. By the minimality of our choice of ℓ, we must have $\partial^i \mathcal{A} \subset \mathcal{F}$ for every $1 \leqslant i \leqslant \ell - 1$. Thus the number of intermediate chains in $\mathcal{F}$ that we lose is at least

$$|\mathcal{A}| \sum_{i=1}^{\ell-1} \binom{n/2 + m}{i}.$$

On the other hand, all sets of size $n/2 + m$ in $\tilde{\mathcal{F}}$ are the sets from $\mathcal{F}$ with i-shadow completely in $\mathcal{F}$ for all $i \leqslant \ell$. These sets cannot be involved in any 2-chains with sets in $\mathcal{B}$, and therefore we only gain intermediate chains between the levels $n/2 + m - 1$ and $n/2 + m - \ell$. The number of such chains that we gain is at most

$$|\mathcal{A}| \sum_{i=1}^{\ell-1} \binom{n/2 - m + \ell}{i}.$$

Since $\ell \leqslant 2m - 1$, it follows that

$$\binom{n/2 - m + \ell}{i} < \binom{n/2 + m}{i} \quad \text{for every } 1 \leqslant i \leqslant \ell - 1,$$

and hence the number of 2-chains decreases.

Thus we may assume $\ell = 1$. If we had $A \in \mathcal{A}$ and $B \in \mathcal{F}$ with $B \subset A$, $|B| = n/2 + m - 1$, then upon shifting to $\tilde{\mathcal{F}}$, we lose the 2-chain $B \subset A$ and gain no pairs. Hence we may assume $\partial \mathcal{A} \cap \mathcal{F} = \emptyset$, so $\mathcal{B} = \partial \mathcal{A}$. We now claim that the sets $A \in \mathcal{A}$ cannot be involved in any 2-chains $C \subset A$ in $\mathcal{F}$. Suppose to the contrary that we had such a 2-chain. Let $x \in C$ be an arbitrary element of C, and shift A to $A \setminus \{x\}$ (recall that $A \setminus \{x\} \notin \mathcal{F}$). Shift the remaining sets in $\mathcal{A}$ by an arbitrary matching from $\mathcal{A}' = \mathcal{A} \setminus \{A\}$ to $\mathcal{B}' = \partial \mathcal{A} \setminus \{A \setminus \{x\}\}$; we can do this by Hall's theorem, since every set in $\mathcal{A}'$ has at least $n/2 + m - 1$ neighbours in $\mathcal{B}'$, while each set in $\mathcal{B}'$ has at most $n/2 - m + 1$ neighbours. In this shifted set we have lost the 2-chain $C \subset A$, and hence $\tilde{\mathcal{F}}$ has fewer 2-chains.

Hence we may assume that there are no 2-chains in $\mathcal{F}$ involving sets in $\mathcal{A}$. Thus in the shifted family $\tilde{\mathcal{F}}$, sets in $\partial \mathcal{A}$ will also not be in any 2-chains. Now, since $|\mathcal{F}| > \binom{n}{\lfloor n/2 \rfloor}$, it follows from Sperner's theorem that there is some 2-chain $C \subset D$ in $\mathcal{F}$. In $\tilde{\mathcal{F}}$, we may also shift D to some set in $\partial \mathcal{A}$, since $m \geqslant 1$ implies $|\partial \mathcal{A}| > |\mathcal{A}|$. As no set in $\partial \mathcal{A}$ is involved in any 2-chain, this reduces the number of 2-chains, which contradicts the minimality of $\mathcal{F}$.

Therefore we conclude that there cannot be a matching from $\mathcal{A}$ to $\mathcal{B}$ in the auxiliary bipartite inclusion graph, and so we will not shift all sets in $\mathcal{A}$. Instead, we use the following lemma, to be proved shortly, to find a collection of sets to shift.

Lemma A.1. *Let G be a bipartite graph on $U \cup V$ with minimum degree $\delta_U \geqslant 1$ in U and maximum degree Δ_V in V. Suppose there is no matching from U to V. Then there exist non-empty subsets $U_1 \subset U$ and $V_1 \subset V$ with a perfect matching $M : U_1 \to V_1$ and $e(U_1, V) + e(U \setminus U_1, V_1) \leqslant |U_1| \Delta_V$.*

Our auxiliary graph satisfies the conditions of the lemma, with

$$U = \mathcal{A}, V = \mathcal{B}, \Delta_V = \binom{n/2 - m + \ell}{\ell} < \binom{n/2 + m}{\ell},$$

and so we can find a collection of sets $\mathcal{A}_1 \subset \mathcal{A}$ and a matching $M : \mathcal{A}_1 \to \mathcal{B}_1 \subset \mathcal{B}$ as given by the lemma. Consider the shifted family $\tilde{\mathcal{F}}$, where we replace the sets in $\mathcal{A}_1$ by the corresponding sets in $\mathcal{B}_1$. As before, since for every $A \in \mathcal{A}_1$ we have $M(A) \subset A$, we need only consider the intermediate chains.

Again, by the minimality of ℓ, we know that $\mathcal{A}_1$ has full shadow in $\mathcal{F}$ up until the ℓth shadow, and so the same calculation as before implies that we remove more chains than we gain, and thus have fewer intermediate chains in $\tilde{\mathcal{F}}$. Hence it suffices to consider only the new chains formed between levels $n/2 + m$ and $n/2 + m - \ell$.

The number of new chains between these levels we gain is exactly $e(\mathcal{A} \setminus \mathcal{A}_1, \mathcal{B}_1)$. On the other hand, we lose all chains between $\mathcal{A}_1$ and $\partial^\ell \mathcal{A}_1 \cap \mathcal{F} = \partial^\ell \mathcal{A}_1 \setminus \mathcal{B}$. Thus the number of chains we are losing is

$$|\mathcal{A}_1| \binom{n/2 + m}{\ell} - e(\mathcal{A}_1, \mathcal{B}).$$

By the lemma, we have

$$|\mathcal{A}_1| \binom{n/2 + m}{\ell} - e(\mathcal{A}_1, \mathcal{B}) > |\mathcal{A}_1| \binom{n/2 - m + \ell}{\ell} - e(\mathcal{A}_1, \mathcal{B}) \geqslant e(\mathcal{A} \setminus \mathcal{A}_1, \mathcal{B}_1),$$

and hence $\tilde{\mathcal{F}}$ has fewer 2-chains than $\mathcal{F}$, contradicting the optimality of $\mathcal{F}$.

Thus if $\mathcal{F}$ minimizes the number of 2-chains, and A is the largest set in $\mathcal{F}$ with $|A| = n/2 + m$, then whenever $B \subset A$ with $|B| \geqslant n/2 - m + 1$, we must have $B \in \mathcal{F}$ as well. $\qquad\square$

It remains to furnish a proof of Lemma A.1, which we now provide.

Proof of Lemma A.1. As there is no matching from U to V, by Hall's theorem there exists a minimal subset $U_0 \subseteq U$ with $|N(U_0)| < |U_0|$. Since $\delta_U \geqslant 1$, we must have $|U_0| \geqslant 2$. Let $u \in U_0$ be an arbitrary element, and take $U_1 = U_0 \setminus \{u\}$. By the minimality of U_0, it follows that $|N(U_1)| \geqslant |U_1|$, and so we must have $|N(U_1)| = |U_1|$. Set $V_1 = N(U_1)$. Again by the minimality of U_0, for any subset $X \subset U_1$, $|N(X)| \geqslant |X|$, and so by Hall's theorem there exists a perfect matching $M : U_1 \to V_1$.

Now

$$\begin{aligned} e(U_1, V) + e(U \setminus U_1, V_1) &= e(U_1, V) + e(U, V_1) - e(U_1, V_1) \\ &= e(U, V_1) \leqslant |V_1| \Delta_V \\ &= |U_1| \Delta_V, \end{aligned}$$

where the second equality follows from the fact that $N(U_1) = V_1$, and thus we have the desired inequality. $\qquad\square$

References

[1] Chung, F. R. K., Füredi, Z., Graham, R. L. and Seymour, P. (1988) On induced subgraphs of the cube. *J. Combin. Theory Ser. A* **49** 180–187.
[2] Das, S., Gan, W. and Sudakov, B. (2014) The minimum number of disjoint pairs in set systems and related problems. Combinatorica, to appear.
[3] Dove, A. P., Griggs, J. R., Kang, R. J. and Sereni, J. S. (2014) Supersaturation in the Boolean lattice. *Integers* **14A** 1–7.
[4] Engel, K. (1997) *Sperner Theory*, Cambridge University Press.
[5] Erdős, P. (1945) On a lemma of Littlewood and Offord. *Bull. Amer. Math. Soc.* **51** 898–902.

[6] Erdős, P. (1962) On a theorem of Rademacher–Turán. *Illinois J. Math.* **6** 122–127.

[7] Erdős, P. (1962) On the number of complete subgraphs contained in certain graphs. *Magy. Tud. Acad. Mat. Kut. Int. Közl.* **7** 459–474.

[8] Erdős, P. and Kleitman, D. (1974) Extremal problems among subsets of a set. *Discrete Math.* **8** 281–294.

[9] Katona, G. O. H. and Tarján, T. G. (1983) Extremal problems with excluded subgraphs in the *n*-cube. In *Graph Theory* (M. Borowiecki *et al.*, eds), Vol. 1018 of *Lecture Notes in Mathematics*, Springer, pp. 84–93.

[10] Kleitman, D. (1966) A conjecture of Erdős–Katona on commensurable pairs among subsets of an *n*-set. In *Theory of Graphs: Proc. Colloq. Tihany*, pp. 215–218.

[11] Mantel, W. (1907) Problem 28. *Wiskundige Opgaven* **10** 60–61.

[12] Qian, J., Engel, K. and Xu, W. (2009) A generalization of Sperner's theorem and an application to graph orientations. *Discrete Appl. Math.* **157** 2170–2176.

[13] Sperner, E. (1928) Ein Satz über Untermengen einer endlichen Menge. *Math. Z.* **27** 544–548.