Supersaturation and stability for forbidden subposet problems

Balázs Patkós*

June 10, 2014

Abstract

We address a supersaturation problem in the context of forbidden subposets. A family $F$ of sets is said to contain the poset $P$ if there is an injection $i : P \rightarrow F$ such that $p \leq_P q$ implies $i(p) \subset i(q)$. The poset on four elements $a, b, c, d$ with $a, b \leq c, d$ is called butterfly. The maximum size of a family $F \subseteq 2^{[n]}$ that does not contain a butterfly is $\Sigma(n, 2) = \binom{n}{\lfloor n/2 \rfloor} + \binom{n}{\lfloor n/2 \rfloor+1}$ as proved by De Bonis, Katona, and Swanepoel. We prove that if $F \subseteq 2^{[n]}$ contains $\Sigma(n, 2) + E$ sets, then it has to contain at least $(1 - o(1))E(\lfloor n/2 \rfloor + 1)\binom{\lfloor n/2 \rfloor}{2}$ copies of the butterfly provided $E \leq 2^{n^{1-\varepsilon}}$ for some positive $\varepsilon$. We show by a construction that this is asymptotically tight and for small values of $E$ we show that the minimum number of butterflies contained in $F$ is exactly $E(\lfloor n/2 \rfloor + 1)\binom{\lfloor n/2 \rfloor}{2}$.

1 Introduction

Many problems in extremal combinatorics deal with determining the maximum size $M$ that certain combinatorial structures can have provided they satisfy some prescribed property. Many properties are defined via forbidden substructures. Therefore for any $M' > M$ it is natural to ask what is the minimum number of forbidden substructures that appear in a structure of size $M'$. Such problems are usually called supersaturation or Erdős-Rademacher type problems. The first such result is a strengthening of Mantel’s theorem [20] that states that a triangle-free graph on $n$ vertices may contain at most $\lfloor \frac{n^2}{4} \rfloor$ edges. Rademacher proved that any graph on $n$ vertices with $\lfloor \frac{n^2}{4} \rfloor + 1$ edges contain at least $\lfloor n/2 \rfloor$ triangles and later Erdős [10] improved this to $t \cdot \lfloor n/2 \rfloor$ triangles for graphs with $\lfloor \frac{n^2}{4} \rfloor + t$ edges provided $t \leq cn$

*MTA–ELTE Geometric and Algebraic Combinatorics Research Group, H–1117 Budapest, Pázmányp. sétány 1/C, Hungary. Email: patkosb@cs.elte.hu. Research supported by the János Bolyai Research Scholarship of the Hungarian Academy of Sciences.
where \( c \) is a sufficiently small positive constant. The general \( r \)-clique case was settled (after the work of many researchers) by Reiher [21]. A supersaturation phenomenon that holds for both graphs and hypergraphs was discovered by Erdős and Simonovits [12].

The first appearance of supersaturation problems in extremal set theory was the minimization problem of the number of disjoint pairs of sets in a family \( F \subseteq 2^{[n]} \) of size \( m \). As observed by Erdős, Ko and Rado [11] a maximum intersecting family (a family of sets such that all pairwise intersections are non-empty) has size \( 2^{n-1} \), thus the problem of determining the minimum number of disjoint pairs is interesting if \( m > 2^{n-1} \). Ahlswede [1] and Frankl [13] independently proved that an optimal family must contain \( \left( \begin{array}{c} n \\ \geq \lceil l \rceil \end{array} \right) \) and must be contained in \( \left( \begin{array}{c} n \\ \geq \lceil l \rceil + 1 \end{array} \right) \) for the integer \( l \) defined by \( \sum_{i=l+1}^{n} \left( \begin{array}{c} n \\ i \end{array} \right) \leq m \leq \sum_{i=l}^{n} \left( \begin{array}{c} n \\ i \end{array} \right) \). As any subset \( F \) of \([n]\) of size \( l \) is disjoint from the same number of subsets of size at least \( l+1 \), this reduced the problem to the uniform case. By the celebrated theorem of Erdős, Ko and Rado [11], a maximum intersecting family \( F \subseteq \left( \begin{array}{c} n \\ l \end{array} \right) \) provided \( 2l \leq n \) holds, thus the supersaturation problem is to minimize the number of disjoint pairs for uniform families of size \( m \) where \( m > \left( \begin{array}{c} n \\ l-1 \end{array} \right) \). This was first addressed by Bollobás and Leader [3] and a major improvement was achieved recently by Das, Gan, and Sudakov [5]. (The graph case \( l = 2 \) was first solved by Ahlswede and Katona [2] in a different context.)

The first theorem in extremal set theory is that of Sperner [22]. It states that a family \( F \subseteq 2^{[n]} \) for which there does not exist a pair \( F, G \in F \) with \( F \subseteq G \) can have size at most \( \left( \begin{array}{c} n \\ \lfloor n/2 \rfloor \end{array} \right) \). These families are called Sperner families. A generalization is due to Erdős [9]: a \( k \)-Sperner family is one that does not contain a chain \( F_1 \subseteq F_2 \subseteq \cdots \subseteq F_{k+1} \) of length \( k+1 \). He proved that if \( F \subseteq 2^{[n]} \) is \( k \)-Sperner, then \(|F| \leq \Sigma(n, k) := \sum_{i=1}^{k} \left( \begin{array}{c} n-k+i \\ i \end{array} \right) \) holds and also that the only \( k \)-Sperner families with this size are \( \cup_{i=1}^{k} \left( \begin{array}{c} n-k+i \\ i \end{array} \right) \) and \( \cup_{i=0}^{k-1} \left( \begin{array}{c} n-k+i \\ i \end{array} \right) \) (these two families coincide if \( n+k \) is odd). We denote the set of these families by \( \Sigma^*(n, k) \). The corresponding supersaturation problem (i.e. minimizing the number of \((k+1)\)-chains for families \( F \subseteq 2^{[n]} \) with \(|F| = m > \Sigma(n, k)\)) was addressed by Kleitman [17] for the case \( k = 2 \). For large values of \( k \), the problem was recently settled by Das, Gan, and Sudakov [5] and independently by Dove, Griggs, Kang, and Sereni [8].

Sperner’s and Erdős’s theorem can be formulated in the following more general context. Let \( P \) be a finite poset. We say that a family \( F \) of sets contains \( P \) if there exists an injection \( i : P \to F \) such that \( p \leq_{P} q \) implies \( i(p) \subset i(q) \). A family \( F \) is \( P \)-free if it does not contain \( P \). The size of a largest \( P \)-free family \( F \subseteq 2^{[n]} \) is denoted by \( La(n, P) \). Note that Erdős’s theorem can be restated as \( La(n, P_{k+1}) = \Sigma(n, k) \), where \( P_{k+1} \) denotes the total ordering or path on \( k+1 \) elements. There are not too many posets \( P \) for which the exact value of \( La(n, P) \) is known. One such example is the butterfly poset \( B \) on four elements \( a, b, c, d \) with \( a \leq_{B} c, a \leq_{B} d, b \leq_{B} c, b \leq_{B} d \).

**Theorem 1.1** (De Bonis, Katona, Swanepoel [7]). If \( F \subseteq 2^{[n]} \) is a butterfly free family, then \(|F| \leq \Sigma(n, 2) \) holds and equality holds if and only if \( F \in \Sigma^*(n, 2) \).
In this paper we will address the problem of minimizing the number of butterflies contained in families $F \subseteq 2^{[n]}$ of fixed size $m > \Sigma(n, 2)$. It will be more convenient to count different image sets of injections of $B$ to $F$ as copies of $B$ instead of counting the number of injections. Also, it does not make too much difference as there are exactly four injections (the number of automorphisms of $B$) to any possible image set.

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

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

**Proposition 1.2.** (a) If $S \subset F$ for some $S \in \Sigma^*(n, 2)$, then $F$ contains at least $(|F| - \Sigma(n, 2))f(n)$ copies of butterflies.

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

It is known that it is possible to construct a family $E$ with the above property as long as the number of sets in $E$ is not more than $\frac{c}{n}\binom{n}{k+1}$ for some sufficiently small constant $c$ (for details, see Section 4). The main result of this paper states that this is best possible for all families of size $\Sigma(n, 2) + E$ if $E$ is very small and asymptotically best possible if $E$ is not that small.

**Theorem 1.3.** Let $F \subseteq 2^{[n]}$ be a family of sets with $|F| = \Sigma(n, 2) + E$.

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

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

One of our major tool in proving Theorem 1.3 will be the following stability version of Theorem 1.1.

**Theorem 1.4.** Let $F \subseteq 2^{[n]}$ be a butterfly-free family such that $|F \setminus F^*| \geq m$ for every $F^* \in \Sigma^*(n, 2)$. Then the inequality $|F| \leq \Sigma(n, 2) - \frac{m}{4}$ holds provided $m$ satisfies $\log m = o(n)$. 


Note that the condition of both Theorem 1.3 and Theorem 1.4 will be used in the form $m, E \leq (\varepsilon n / 2)$ for any fixed positive $\varepsilon$ provided $n$ is large enough.

The rest of the paper is organized as follows. In Section 2, we prove Theorem 1.4 then in Section 3 we prove Theorem 1.3. Section 4 contains some concluding remarks.

2 Stability of maximum butterfly free families

In this section we prove Theorem 1.4, a stability version of Theorem 1.1. As butterfly free families of maximum size possess the 2-Sperner property, the proof of Theorem 1.4 consists of two steps: we first establish a stability result (Lemma 2.6) on 2-Sperner families and then we consider how the number of 3-chains in a butterfly-free family $F$ effect the size of $F$. Note that if $n$ is odd, then Lemma 2.6 would easily follow from the characterization of 1-Sperner families as in that case the family in $\Sigma^*(n, 2)$ consists of two Sperner families of equal maximum size. However, if $n$ is even we need a little work to obtain the same result. We start with stating the celebrated LYM-inequality.

**Theorem 2.1** (LYM-inequality for 2-Sperner families). If $F \subseteq 2^{[n]}$ is a 2-Sperner family, then the inequality

$$\sum_{F \in F} \frac{1}{|F|} \leq 2$$

holds.

**Corollary 2.2.** Let $F \subseteq 2^{[n]}$ be a 2-Sperner family such that one of the following holds:

(a) $n$ is odd and the number of sets $|\{G \notin F : |G| = \lfloor n/2 \rfloor \text{ or } |G| = \lceil n/2 \rceil \}|$ is at least $m$,

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

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

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

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

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

$$\left(\frac{n}{n/2 - 1}\right)^{-1} / \left(\frac{n}{n/2}\right)^{-1} = \frac{|n/2| + 1}{n/2} \geq 1 + \frac{1.9}{n}.$$ 

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

$$\left(\frac{n}{n/2 - 2}\right)^{-1} / \left(\frac{n}{n/2 - 1}\right)^{-1} = \frac{n/2 + 2}{n/2 - 1} \geq 1 + \frac{1.9}{n}.$$ 

\qed
We continue with introducing the notions of shadow and shade. If $\mathcal{F}$ is a family of sets, then its $k$-shadow is $\Delta_k(\mathcal{F}) = \{ G : |G| = k, \exists F \in \mathcal{F} \ G \subset F \}$. To define the $k$-shade of the family we have to assume the existence of an underlying set, say $\mathcal{F} \subseteq 2^{[n]}$. If so, then the $k$-shade is defined as $\nabla_k(\mathcal{F}) = \{ G \in \binom{[n]}{k} : \exists F \in \mathcal{F} \ F \subset G \}$. The well known theorem of Kruskal [18] and Katona [15] states which family $\mathcal{F}$ of $k$-sets minimizes the size of $\Delta_{k-1}(\mathcal{F})$ among all families of $m$ sets. For calculations the following version happens to be more useful than the precise result.

**Theorem 2.3** (Lovász, [19]). Let $\mathcal{G}$ be a family of $k$-sets and let $x$ be the real number such that $\binom{x}{k} = \frac{x(x-1)\ldots(x-k+1)}{k!} = |\mathcal{G}|$ holds. Then the family of shadows satisfies $|\Delta_{k-1}(\mathcal{G})| \geq \binom{x}{k-1}$.

We will apply Theorem 2.3 in a slightly more general setting. If $\mathcal{F} \subset \binom{[n]}{l}$ with $l > n/2$, then a simple double counting argument and Hall’s theorem show that there exists a matching from $\mathcal{F}$ to $\Delta_{l-1}(\mathcal{F})$ such that if $F \in \mathcal{F}$ and $G \in \Delta_{l-1}(\mathcal{F})$ are matched, then $G \subset F$. Using this observation and Theorem 2.3 one obtains the following theorem. Part (ii) of the statement follows from the fact that $G \subset F \subset [n]$ holds if and only if $|\mathcal{G}|$ holds.

**Lemma 2.4.** (i) Let $\mathcal{G} \subset \binom{[n]}{l}$ be a Sperner family with $|\mathcal{G}|$ holds. Then the family of shadows satisfies $|\Delta_{k-1}(\mathcal{G})| \geq \binom{x}{k-1}$.

(ii) Let $\mathcal{G} \subset \binom{[n]}{l}$ be a Sperner family with $|\mathcal{G}|$ holds. Then the family of shadows satisfies $|\nabla_{k+1}(\mathcal{G})| \geq \binom{x}{n-k}$.

Let us define the following functions of $l$ and $m$. Let $x = x(l, m)$ defined by the equation $\binom{x}{k} = m$ and write $g(l, m) = \binom{x}{l-1} - \binom{x}{l}$. According to Lemma 2.4 if $\mathcal{G} \subset \binom{[n]}{l}$ is a Sperner family with $|\mathcal{G}| = m$ and $l \geq |\mathcal{G}|$, then $|\Delta_{l-1}(\mathcal{G})| - |\mathcal{G}| \geq g(l, m)$ holds. In the following proposition we gather some properties of $x(l, m)$ and $g(l, m)$.

**Proposition 2.5.** (i) If $m \leq \binom{2l}{l}$, then $x(l, m) \leq x(l + 1, m)$ holds.

(ii) If $x(l, m) \leq 2l - 1$, then $g(l, m) \geq 0$ holds.

(iii) If $m_1 + m_2 = m$ and $x(l, m) \leq 2l - 1$, then $g(l, m_1) + g(l, m_2) \geq g(l, m)$ holds.

(iv) If $m \leq \binom{2l-1}{l}$, then $g(l, m) \leq g(l + 1, m)$ holds.

(v) For every $\varepsilon > 0$ there exists $l_0$ such that if $l \geq l_0$, then $g(l, m)$ is increasing in the interval $0 \leq m \leq \binom{(2l+\varepsilon)}{l}$.

(vi) If $x(l, m) \leq 4l/3$, then $g(l, m) \geq 2m$ holds.

**Proof.** (i) Clearly the polynomial $\binom{x}{l+1}$ is monotone increasing in $x$ if $x \geq l+1$ holds. Observe that $\frac{x(l,m)}{l(l+1)} = \frac{x(l,m)}{l(l+1)} < m$ as $x(l, m) \leq 2l$ by the assumption $m \leq \binom{2l}{l}$. Therefore $x(l, m) \leq x(l + 1, m)$ holds.

To obtain (ii), (iii) and (vi) write $g(l, m)$ in the following form

$$g(l, m) = \frac{x}{l(l-1)} - \frac{x}{l} = \left( \frac{l}{x-l+1} - 1 \right) \left( \frac{x}{l} \right) = \frac{2l-x-1}{x-l+1} m.$$
(ii) is straightforward and as for fixed \( l \) we know that \( x(l, m) \) is an increasing function of \( m \), therefore the fraction \( \frac{2l-x+1}{x(l+1)} \) is decreasing in \( m \).

To obtain (iv), as \( g(l, m) - g(l + 1, m) = \left( \frac{x(l,m)}{l-1} \right) - \left( \frac{x(l+1,m)}{l} \right) \) we need to compare \( \left( \frac{x(l,m)}{l-1} \right) \) and \( \left( \frac{x(l+1,m)}{l} \right) \).

\[
\frac{\left( x(l,m) \right)}{(x(l+1,m) + m)} = \frac{m}{(x(l+1,m) + m)} = \frac{\left( x(l,m) \right)}{(x(l+1,m) + m)} = \frac{\left( x(l+1,m) \right)}{x(l,m) - l} \cdot \frac{x(l, m) - l - 1}{l + 1} < 1,
\]

where we used (i) and the assumption \( m \leq \left( \frac{2l-1}{l} \right) \) and thus \( x(l, m) \leq 2l - 1 \).

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

\[
g(l, m) = \left( \frac{x}{l-1} \right) - \left( \frac{x}{l} \right) = \frac{x(x-1)\ldots(x-l+2)(2l-x-1)}{l!}.
\]

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

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

If \( x \leq (2 - \varepsilon)l \), then for any \( i \geq (1 - \varepsilon/2)l \) the \( j \)th product in the sum is at least \( \varepsilon \) fraction of the product to subtract. Thus if \( l \) is large enough the derivative is positive and thus the function is increasing. As \( x \) is a monotone increasing function of \( m \), the claim holds.

After all these preliminary results we are ready to state and prove the stability result on 2-Sperner families.

**Lemma 2.6.** For every \( \varepsilon > 0 \), there exists an \( n_0 \), such that the following holds: if \( n \geq n_0 \), \( m \leq \left( \frac{1 - \varepsilon}{n/2} \right) \) and \( \mathcal{F} \subseteq [n] \) is a 2-Sperner family with the property that for any \( \mathcal{F}^* \in \Sigma^*(n, 2) \) we have \( |\mathcal{F} \setminus \mathcal{F}^*| \geq m \), then the following upper bound holds on the size of \( \mathcal{F} \):

\[
|\mathcal{F}| \leq \Sigma(n, 2) - g([n/2] + 1, m).
\]

**Proof.** Let \( \varepsilon > 0 \) be fixed and let \( \mathcal{F} \subseteq [n] \) be a 2-Sperner family such that for any \( \mathcal{F}^* \in \Sigma^*(n, 2) \) we have \( |\mathcal{F} \setminus \mathcal{F}^*| \geq m \). Write \( m' = \min_{\mathcal{F}^* \in \Sigma^*(n, 2)} \{|\mathcal{F} \setminus \mathcal{F}^*|\} \). We can assume that \( m' \leq \left( \frac{1}{2} + o(1) \right) \left( \frac{n}{n/2} \right) \) as otherwise \( |\left( \left[ \frac{n}{2} \right] \right) \cup \left( \left[ \frac{n}{2} \right] \right) \setminus \mathcal{F} \| \geq \delta \left( \left[ \frac{n}{2} \right] \right) \) would hold for some positive \( \delta \) and we would be done by Corollary 2.2 part a) or c) depending on the parity of \( n \).

**CASE I** \( m' \geq \left( \frac{1 - \varepsilon/2}{n/2} \right) \).

If \( n \) is odd, then by Corollary 2.2 b), we have \( |\mathcal{F}| \leq \Sigma(n, 2) - \frac{19m}{n} \) and we are done.
If $n$ is even, then by symmetry we can suppose that $m' = |\mathcal{F} \setminus (\binom{[n]}{n/2-1} \cup \binom{[n]}{n/2})|$. Let $\mathcal{F} = \mathcal{F}_1 \cup \mathcal{F}_2$ with $\mathcal{F}_1 = \{F \in \mathcal{F} : \exists F' \in \mathcal{F}, F' \subset F\}$ and $\mathcal{F}_2 = \mathcal{F} \setminus \mathcal{F}_1$. Let us write $\mathcal{F}_{1,+} = \{F \in \mathcal{F}_1 : |F| > n/2\}$, $\mathcal{F}_{2,+} = \{F \in \mathcal{F}_2 : |F| > n/2\}$, $\mathcal{F}_- = \{F \in \mathcal{F} : |F| < n/2 - 1\}$, $\mathcal{G}_{n/2} = \{G \notin \mathcal{F}, |G| = n/2\}$, $\mathcal{G}_{n/2-1} = \{G \notin \mathcal{F}, |G| = n/2 - 1\}$.

Observe the following bounds:

- $|\mathcal{G}_{n/2}| \leq \left(\frac{1-3\varepsilon}{n/2}\right)^n$ as otherwise by Corollary 2.2 d), we are done.
- $|\mathcal{F}_{1,+}| \leq \left(\frac{1-3\varepsilon}{n/2}\right)^n$ as $\Delta_{n/2}(\mathcal{F}_{1,+}) \subseteq \mathcal{G}_{n/2}$ and $|\mathcal{F}_{1,+}| \leq |\Delta_{n/2}(\mathcal{F}_{1,+})|$ hold.
- $|\mathcal{F}_-| \leq \left(\frac{1-3\varepsilon}{n/2}\right)^n$ as otherwise by Corollary 2.2 c), we are done.

By definition all sets in $\Delta_{n/2}(\mathcal{F}_{2,+}) \setminus \mathcal{G}_{n/2}$ must belong to $\mathcal{F}_1$. No set below an arbitrary set of $\mathcal{F}_1$ belongs to $\mathcal{F}$, therefore all sets of $\Delta_{n/2-1}(\mathcal{F}_{2,+})$ belong to $\mathcal{G}_{n/2-1}$ except those whose complete shade belongs to $\mathcal{G}_{n/2}$. By double counting pairs $(G, G')$ with $G' \subset G$, $|G'| = n/2 - 1$, $G \in \mathcal{G}_{n/2}$ and $\nabla_{n/2}(G') \subseteq \mathcal{G}_{n/2}$ we obtain that the number of such exceptional sets is $(1 + o(1))|\mathcal{G}_{n/2}| \leq \left(\frac{1-2\varepsilon}{n/2}\right)^n$. Let $\mathcal{E}$ denote the family of these exceptional sets.

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

$$|\mathcal{F}| \leq \Sigma(n, 2) - |\Delta_{n/2-1}(\mathcal{F}_{2,+})| + |\mathcal{E}| + |\mathcal{F}_{2,+}| + |\mathcal{F}_{1,+}| + |\mathcal{F}_-|$$

$$\leq \Sigma(n, 2) - |\Delta_{n/2-1}(\mathcal{F}_{2,+})| + m'' + 3 \left(\frac{1-2\varepsilon/3}{n/2}\right)^n$$

$$\leq \Sigma(n, 2) - \left(\binom{x''}{n/2-1} - \binom{x''}{n/2+1}\right) + 3 \left(\frac{1-2\varepsilon/3}{n/2}\right)^n$$

$$\leq \Sigma(n, 2) - \frac{1}{n^2} m'' + 3 \left(\frac{1-2\varepsilon/3}{n/2}\right)^n.$$ 

Here the third inequality follows by Lemma 2.4 and and the last one follows from $x'' \leq 2n - 1 + o(1)$ since $m'' \leq m' \leq \left(\frac{1}{2} + o(1)\right) \binom{n}{n/2}$. We are done as $m'' \geq \left(\frac{1-0.6\varepsilon}{n/2}\right)^n$ holds.

**Case II** $m' < \left(\frac{1-\varepsilon/2}{n/2}\right)^n$

Again we may assume that $|\left(\binom{[n]}{n/2-1} \cup \binom{[n]}{n/2}\right) \setminus \mathcal{F}| = m'$. Let $\mathcal{F} = \mathcal{F}_1 \cup \mathcal{F}_2$ with $\mathcal{F}_1 = \{F \in \mathcal{F} : \exists F' \in \mathcal{F}, F' \subset F\}$ and $\mathcal{F}_2 = \mathcal{F} \setminus \mathcal{F}_1$. Let us write $\mathcal{F}_{1,-} = \{F \in \mathcal{F}_1 : |F| < n/2 - 1\}, \mathcal{F}_{1,+} = \{F \in \mathcal{F}_1 : |F| > n/2 - 1\},$
\[ \mathcal{F}_{2,-} = \{ F \in \mathcal{F} : |F| < n/2 \}, \mathcal{F}_{2,+} = \{ F \in \mathcal{F} : |F| > n/2 \}. \]

To bound the size of \( \mathcal{F}_1 \) note that \( \mathcal{F}_1 \) is disjoint both from \( \Delta_{n/2-1}(\mathcal{F}_{1,+}) \) and \( \nabla_{n/2-1}(\mathcal{F}_{1,-}) \). Similarly, \( \mathcal{F}_2 \) is disjoint both from \( \Delta_{n/2}(\mathcal{F}_{2,+}) \) and \( \nabla_{n/2}(\mathcal{F}_{2,-}) \). By Lemma 2.4 we obtain

\[
|\mathcal{F}| = |\mathcal{F}_1| + |\mathcal{F}_2| \leq \binom{n}{n/2} - g(n/2 + 2, |\mathcal{F}_{1,-}|) - g(n/2, |\mathcal{F}_{1,+}|) + \\
\quad + \binom{n}{n/2} - g(n/2 + 1, |\mathcal{F}_{2,-}|) - g(n/2 + 1, |\mathcal{F}_{2,+}|) \\
\quad \leq \Sigma(n, 2) - g(n/2 + 1, |\mathcal{F}_{1,-}|) - |\mathcal{F}_{2,+}| - |\mathcal{F}_{2,-}| + |\mathcal{F}_{1,+}| - g(n/2, |\mathcal{F}_{1,+}|),
\]

where we used Proposition 2.6 (iii) and (iv). Let us partition \( \mathcal{F}_{1,+} \) into \( \mathcal{F}_{1,+} \cup \mathcal{F}_{1,+} \cup \Delta_{n/2}(\mathcal{F}_{1,+}) \) are disjoint and thus \( |\mathcal{F}_{1,+}| + |\mathcal{F}_{2,+}| \). Also \( s = |\mathcal{F}_{1,+} \cup \Delta_{n/2}(\mathcal{F}_{1,+})| \leq \binom{n}{n/2} \) and thus \( g(s) \geq 0 \) holds. Therefore we obtain \( |\mathcal{F}_1| \leq \binom{n}{n/2} - g(n/2 + 1, |\mathcal{F}_{1,+}|) - g(n/2 + 2, |\mathcal{F}_{1,-}|) \). By Proposition 2.6 (iii), this strengthens the above arrayed inequality to

\[
|\mathcal{F}| \leq \Sigma(n, 2) - g(n/2 + 1, |\mathcal{F}_{1,+}|) + |\mathcal{F}_{1,-}| + |\mathcal{F}_{2,-}| - |\mathcal{F}_{2,+}|.
\]

Note that \( m \leq |\mathcal{F}_{1,+} \cup \mathcal{F}_{1,-} \cup \mathcal{F}_{2,-} \cup \mathcal{F}_{2,+} \) as \( \mathcal{F}_{1,+} \cup \mathcal{F}_{1,-} \cup \mathcal{F}_{2,-} \cup \mathcal{F}_{2,+} \). Therefore we are done by Proposition 2.6 (v).

Having proved Lemma 2.6 we can now turn our attention to butterfly free families containing chains of length 3. Our main tool to bound their size is the following LYM-type inequality.

**Lemma 2.7.** Let \( \mathcal{F} \subset 2^n \setminus \{0, [n]\} \) be a butterfly-free family and let \( \mathcal{M} \) be defined as \( \{ M \in \mathcal{F} : \exists F, F' \in \mathcal{F} \text{ such that } F \subset M \subset F' \} \).

\[
\sum_{F \in \mathcal{F} \setminus \mathcal{M}} \frac{1}{\binom{n}{|F|}} + \sum_{M \in \mathcal{M}} \left( 1 + \frac{(|M| - 1)(n - |M| - 1) - 1}{|M|(n - |M|)} \right) \frac{1}{\binom{n}{|M|}} \leq 2
\]

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

\[
\sum_{F \in \mathcal{F}} |F|!(n - |F|)! \leq 2n! + \sum_{M \in \mathcal{M}} (|M| - 1)! (n - |M| - 1)! - \sum_{M \in \mathcal{M}} ((|M| - 1)(|M| - 1)!)((n - |M| - 1)(n - |M| - 1)!).
\]

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

Corollary 2.8. Let \(\mathcal{F} \subseteq 2^{[n]}\) be a butterfly-free family with \(\emptyset, [n] \notin \mathcal{F}\) and let us write \(\mathcal{M} = \{M \in \mathcal{F} : \exists F, F' \in \mathcal{F} \ F \subset M \subset F'\}\). Then \(|\mathcal{F}| \leq \Sigma(n, 2) - |\mathcal{M}|/3\).

Proof. As \(\emptyset, [n] \notin \mathcal{F}\), for any \(M \in \mathcal{M}\) we have \(2 \leq |M| \leq n - 2\) and thus \(\frac{(|M|-1)(n-|M|-1-1)!}{|M|!(n-|M|)!} \geq 1/3\) if \(n \geq 8\).

Now we are ready to prove Theorem 1.4, the statement of which we recall here below.

Theorem 1.4. Let \(\mathcal{F} \subseteq 2^{[n]}\) be a butterfly-free family such that \(|\mathcal{F} \setminus \mathcal{F}^*| \geq m\) for every \(\mathcal{F}^* \in \Sigma^*(n, 2)\). Then the inequality \(|\mathcal{F}| \leq \Sigma(n, 2) - \frac{m}{6}\) holds provided \(m = o\left(\frac{n^2 + \log n}{n^2}\right)\).

Proof. Let \(\mathcal{F} \subseteq 2^{[n]}\) be a butterfly-free family satisfying the conditions of the theorem. If \(\emptyset \in \mathcal{F}\) or \([n] \in \mathcal{F}\), then \(\mathcal{F} \setminus \{\emptyset\}\) or \(\mathcal{F} \setminus \{[n]\}\) does not contain the poset \(V\) or \(\wedge\), where \(V\) is the poset with three elements one smaller than the other two and \(\wedge\) is the poset with three elements one larger than the other two. In either case by a theorem of Katona and Tarján [16], we have \(|\mathcal{F}| \leq (1 + O(\frac{1}{n}))\binom{n}{n/2}\). Thus we may assume \(\emptyset, [n] \notin \mathcal{F}\). If \(\mathcal{M} = \{M \in \mathcal{F} : \exists F, F' \in \mathcal{F} \ F \subset M \subset F'\}\) contains at least \(m/2\) sets, then we are done by Corollary 2.8. If \(|\mathcal{M}| \leq m/2\), then \(\mathcal{F} \setminus \mathcal{M}\) is 2-Sperner and \(|(\mathcal{F} \setminus \mathcal{M}) \setminus \mathcal{F}^*| \geq m/2\) for every \(\mathcal{F}^* \in \Sigma^*(n, 2)\) and thus by Lemma 2.6 we obtain \(|\mathcal{F} \setminus \mathcal{M}| \leq \Sigma(n, 2) - g(\lceil n/2 \rceil + 1, m/2) \leq \Sigma(n, 2) - m\) as we can use Proposition 2.5 (vi). Therefore \(|\mathcal{F}| \leq \Sigma(n, 2) - m/2\) holds.

3 Proof of the main result

In this section we prove Theorem 1.3. Our main tool is the stability result Theorem 1.4 proved in the previous section. This tells us that if the structure of a butterfly free family is very different from that of the extremal family, then it contains much fewer sets. Since any new set yields an additional copy of the butterfly poset, a family with few butterflies must contain an almost extremal butterfly free family. To deal with families \(\mathcal{F}\) containing almost extremal butterfly free families \(\mathcal{G}\), we have to prove that most sets in \(\mathcal{F} \setminus \mathcal{G}\) behave very similarly to the extra sets in the conjectured extremal families. We formalize this handwaving statement in the following theorem.
Theorem 3.1. For any $\varepsilon > 0$ there exists an $n_0$ such that for any $n \geq n_0$ the following holds provided $m$ satisfies $\log m = o(n)$ and $n/2 - \sqrt{n} \leq k \leq n/2 + \sqrt{n}$: let $F \subseteq \binom{[n]}{k}$ with $|F| = \binom{n}{k} - m$. Then the number of sets in $\binom{[n]}{\geq k+1}$ that contain less than $(1-\varepsilon)k$ sets from $F$ is $o(m)$.

Before we start the proof of Theorem 3.1 let us introduce some notation and an isoperimetric problem due to Kleitman and West (according to Harper [14]). Given a graph $G$ and a positive integer $m \leq |V(G)/2|$, the isoperimetric problem asks for the minimum number of edges $e(X, V(G) \setminus X)$ that go between an $m$-element subset $X$ of $V(G)$ and its complement. For regular graphs, this problem is equivalent to finding the maximum number of edges $e(X)$ in a subgraph of $G$ induced by an $m$-subset $X$ of the vertices. Indeed, in a $d$-regular graph we have $d|X| = 2e(X) + e(X, V(G) \setminus X)$.

Kleitman and West asked for the solution of the isoperimetric problem in the Hamming graph $H(n, k)$ whose vertex set is $\binom{[n]}{k}$ and two $k$-subsets are connected if their intersection has size $k - 1$. Harper [14] introduced and solved a continuous version of this problem. Here we summarize some of his findings. The shift operation $\tau_{i,j}$ is a widely used tool in extremal set theory. It is defined by

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

(1)

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

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

The characteristic vector of a subset $F$ of $[n]$ is a $0-1$ vector $x_F$ of length $n$ with $x_F(i) = 1$ if $i \in F$ and $x_F(i) = 0$ if $i \notin F$. 0-1 vectors with exactly $k$ one entries are clearly in one-to-one correspondence with $\binom{[n]}{k}$. But also, one can consider non-negative integer vectors of length $k$ for any set $F \in \binom{[n]}{k}$ such that $y_F(j) = i_j - j$ where $i_j$ is the index of the $j$th one entry of $x_F$. For any set $F \in \binom{[n]}{k}$ the entries of $y_F$ are non-decreasing as $i_j - j$ is the number of zero coordinates of $x_F$ before the $j$th 1-coordinate. Also, $0 \leq y_F(1) \leq y_F(2) \leq \cdots \leq y_F(k) \leq n-k$ hold.

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

As a short detour, let us mention that Harper considered and solved the following continuous version of the above isoperimetric problem: let \( L_k = \{ z : 0 \leq z(1) \leq z(2) \leq \cdots \leq z(k) \leq 1 \} \) be partially ordered coordinatewise. For fixed \( s \) with \( 0 \leq s \leq \frac{1}{2} \), maximize \( \int_D r(z)dz \) over all downsets of \( L_k \) with volume \( s \). We will use only the following simple observation to prove Theorem 3.1.

**Proposition 3.2.** Let \( \mathcal{F} \subseteq \binom{[n]}{k} \) be a family of size \( m < \binom{\delta n}{\delta n/2} \), then in \( H(n, k) \) we have \( e(\mathcal{F}) \leq \delta mn^2 \).

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

Now we are ready to prove Theorem 3.1.

**Proof of Theorem 3.1.** Let \( \overline{\mathcal{F}} = \binom{[n]}{k} \setminus \mathcal{F} \) and thus \( |\overline{\mathcal{F}}| = m \). We want to bound the number of sets of which the shadow is contained in \( \overline{\mathcal{F}} \) with the exception of at most \( (1 - \varepsilon)k \) sets. Let \( \mathcal{G} \subseteq 2^{[n]} \) denote the family of such sets and write \( \mathcal{G}_l = \{ G \in \mathcal{G} : |G| = l \} \). To bound \( |\mathcal{G}_{k+1}| \) we double count the pairs \( F_1, F_2 \) of sets in \( \overline{\mathcal{F}} \) with \( |F_1 \cap F_2| = k - 1 \). As \( \overline{\mathcal{F}} \) has size \( m \), by applying Proposition 3.2 with a sequence \( \delta_n \to 0 \), we obtain the number of such pairs is \( o(mn^2) \). On the other hand for every such pair there exists at most one \( G \in \mathcal{G}_{k+1} \) with \( F_1, F_2 \subseteq G \) (namely, \( F_1 \cup F_2 \)). Thus the number of such pairs is at least \( |\mathcal{G}_{k+1}| \binom{\delta m}{\delta n/2} \). Therefore, we obtain \( |\mathcal{G}_{k+1}| \binom{\delta m}{\delta n/2} = o(mn^2) \). Rearranging and the assumption on \( k \) yields that \( |\mathcal{G}_{k+1}| = o(m) \).

To bound \( |\mathcal{G}_l| \) for values of \( l \) larger than \( k+1 \), observe that \( \Delta_{k+1}(\mathcal{G}_l) \subseteq \mathcal{G}_{k+1} \) holds for all \( l > k+1 \). Let \( x \) denote the real number for which \( |\mathcal{G}_{k+1}| = \binom{x}{k+1} \) holds. By Theorem 2.3, we obtain that \( |\mathcal{G}_l| \leq \binom{x}{l} \) holds. By the assumption on \( m \) and \( k \), we see that \( x = k+1 + o(k) \) and thus by Proposition 2.5 (vi) we have \( \binom{x}{l} \leq \frac{1}{2} \binom{x}{l+1} \). This gives

\[
|\mathcal{G}| = \sum_{l = k+1}^{n} |\mathcal{G}_l| \leq \sum_{l = k+1}^{n} \binom{x}{l} \leq 2 |\mathcal{G}_{k+1}| = o(m).
\]
Now we are ready to prove our main result that we recall here below. Note that by Proposition 1.2 we only have to deal with families \( F \subseteq 2^n \) that do not contain any \( S \in \Sigma(n, 2) \). When bounding the number of butterflies in \( F \) we will only count copies \( F, F_1, F_2, F_3 \in F \) where \( F \in \cap(\geq [n/2] - 1), F_1 \in ([n/2]), F_2, F_3 \in (\geq [n/2] - 1) \) with \( F_2, F_3 \subset F_1 \subset F \) or \( F \in \cap([n/2] - 2), F_1 \in ([n/2] - 1), F_2, F_3 \in (\geq [n/2]) \) with \( F_2, F_3 \supset F_1 \supset F \).

**Theorem 1.3.** Let \( F \subseteq 2^n \) be a family of sets with \( |F| = \Sigma(n, 2) + E \).

(a) If \( E = E(n) \) satisfies \( \log E = o(n) \), then the number of butterflies contained by \( F \) is at least \( E \cdot f(n) \).

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

**Proof of Theorem 1.3 part (a).** Let \( F \subseteq 2^n \) be a family containing \( \Sigma(n, 2) + E \) sets and let \( F' \) be a maximum size butterfly free subfamily of \( F \). Let \( m \) be defined by \( \min_{S \in \Sigma(n, 2)} |S \setminus F'| \). If \( m \geq 6f(n)E \) holds, then by Theorem 1.4 we have \( |F'| \leq \Sigma(n, 2) - E \cdot f(n) \) and thus \( |F \setminus F'| \geq E \cdot f(n) + 1 \). As \( F' \) is a maximum butterfly free subfamily of \( F \), every set \( F \in F \setminus F' \) forms a butterfly with 3 other sets from \( F' \). Thus the number of butterflies in \( F \) is at least \( |F \setminus F'| \). This finishes the proof if \( m \geq 6f(n)E \) holds.

Suppose next that \( m \leq 6f(n)E \) holds. Note that if \( f(n) \leq n^3 \), \( m \) satisfies \( o\left(\frac{\varepsilon n}{\varepsilon n/2}\right) \) for any positive \( \varepsilon \). Without loss of generality we can assume that \(|\left(\left([n/2] - 1\right) \cup \left([n/2]\right)\right) \setminus F| = m \) and thus \(|F \setminus \left(\left([n/2] - 1\right) \cup \left([n/2]\right)\right)\) = \( m + E \) holds. Let us write \( k = \left\lceil n/2 \right\rceil - 1 \) and fix an \( \varepsilon > 0 \) and pick \( \varepsilon' > 0 \) with the property that \((1 - \varepsilon')^4/2 \geq 1 - \varepsilon \). Applying Theorem 3.1 to \( F \cap \left(\left([n/2] - 1\right) \cup \left([n/2]\right)\right) \) we obtain that the family \( \mathcal{F}_{o,k+1} = \{F \in F \cap \left(\left([n/2] - 1\right) \cup \left([n/2]\right)\right) : |\Delta_k(F) \setminus F| \leq (1 - \varepsilon')k\} \) has size \( o(m) \). Let us apply Theorem 3.1 again, this time to \( \mathcal{F}_{g,k+1} = (F \cap \left(\left([n/2] - 1\right) \cup \left([n/2]\right)\right)) \setminus \mathcal{F}_{b,k+1} \). We obtain that the family \( \mathcal{F}_{o,k+2} = \{F \in \left(\left([n/2] - 1\right) \cup \left([n/2]\right)\right) : |\Delta_{k+1}(F) \setminus \mathcal{F}_{b,k+1}| \leq (1 - \varepsilon')k\} \) has size \( o(m) \). With an identical argument applied to \( \mathcal{F} = \{[n] : F \in F\} \), one can show that the families \( \mathcal{F}_{b,k} = \{F \in F \cap \left(\left([n/2] - 1\right) \cup \left([n/2]\right)\right) : |\nabla_{k+1}(F) \setminus F| \leq (1 - \varepsilon')k\} \) and \( \mathcal{F}_{b,k-1} = \{F \in F \cap \left(\left([n/2] - 1\right) \cup \left([n/2]\right)\right) : |\nabla_{k}(F) \setminus (F \setminus \mathcal{F}_{b,k})| \leq (1 - \varepsilon')k\} \) both have size \( o(m) \).

Let us pick a set \( F \in \mathcal{F}_{g} = \mathcal{F} \setminus \left(\left([n/2] \cup \left([n/2]\right)\right) \cup \mathcal{F}_{b,k-1} \cup \mathcal{F}_{b,k+2}\right) \) and note that the number of such sets is \( m + E - o(m) \geq E \). The following claim finishes the proof of Theorem 1.3 part a).

**Claim 3.3.** For every \( F \in \mathcal{F}_{g} \) there exist at least \( (1 - \varepsilon)f(n) \) copies of the butterfly poset that contains only \( F \) from \( \mathcal{F} \setminus \left(\left([n/2] \cup \left([n/2]\right)\right)\right) \).

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

To obtain part b) of Theorem 1.3 we need better bounds on the number of “bad sets”. We start with the following folklore proposition.

**Proposition 3.4.** Let \(U_1, \ldots, U_l\) be sets of size \(u\) such that \(|U_i \cap U_j| \leq 1\) holds for any \(1 \leq i < j \leq l\). Then we have \(|\bigcup_{i=1}^l U_i| \geq l \cdot \frac{2u - l}{2}\) provided \(l \leq u\) holds.

**Proof.** By the condition on the intersection sizes we have \(|U_i \cup \bigcup_{j=1}^{i-1} U_j| \geq u - i + 1\) and thus \(|\bigcup_{i=1}^l U_i| \geq \sum_{i=1}^l (u - i + 1)\). \(\square\)

**Corollary 3.5.** Let \(F \subseteq \binom{[n]}{k}\) with \(|F| = \binom{[n]}{k} - m\). Then the number of sets \(G\) in \(\binom{[n]}{k+1}\) that contain less than \(k + 1 - 2\sqrt{m}\) sets from \(F\) is at most \(\sqrt{m}\) provided \(m \leq k^3\) and \(n/2 - \sqrt{n} \leq k \leq n/2 + \sqrt{n}\). The number of such sets from \(\binom{[n]}{k+1}\) is at most \(2\sqrt{m}\).

**Proof.** For any set \(G \in \binom{[n]}{k+1}\) with \(|\Delta_k(G) \cap F| < k + 1 - 2\sqrt{m}\) one can consider a family \(\mathcal{H}_G\) of \(2\sqrt{m}\) sets from \(\Delta_k(G) \setminus F\). Clearly, \(|\mathcal{H}_G \cap \mathcal{H}_G'| \leq |\Delta_k(G) \cap \Delta_k(G')| \leq 1\). The sets \(\mathcal{H}_G\) satisfy the condition of Proposition 3.4. Thus if the number of such \(G\)'s is more than \(\sqrt{m}\), then \(|F| > 2\sqrt{m} \cdot \frac{2\sqrt{m} - \sqrt{m}}{2} = m\) which is a contradiction.

The proof of the second statement that deals with sets of larger size is as in the proof of Theorem 3.1. \(\square\)

**Proof of Theorem 1.3 part (b).** Let \(\mathcal{F} \subseteq 2^{[n]}\) be a family of sets with \(|\mathcal{F}| = \Sigma(n, 2) + E\) where \(E = E_n \leq \frac{n}{100}\). Let \(m\) be defined by \(\min_{S \in 2^{[n]}(n, 2)} |S \setminus \mathcal{F}|\). We will write \(k + 1 = \lceil n/2 \rceil\) and assume that \(m = (\binom{n}{k} \cup \binom{n}{k+1}) \setminus \mathcal{F}\). We will consider four cases with respect to \(m\).

**CASE I** \(m \geq 6f(n)E\)

Just as in the proof of Theorem 1.3 part (a), we consider a maximal butterfly free subfamily \(\mathcal{F}' \subseteq \mathcal{F}\) with \((\binom{n}{k} \cup \binom{n}{k+1}) \cap \mathcal{F} \subseteq \mathcal{F}'\). By Corollary 2.5, \(|\mathcal{F}'| < \Sigma(n, 2) - f(n)E\) and thus \(\mathcal{F}\) contains at least \(|\mathcal{F}| - |\mathcal{F}'| > Ef(n)\) copies of the butterfly poset.

**CASE II** \(\frac{n}{10} \leq m < 6f(n)E\)

We again repeat the argument of part (a). By applying Theorem 3.1 twice with \(\varepsilon = 1/4\), we obtain that for \(E + m - o(m) \geq (11 - o(1))E\) sets \(F \in \mathcal{F} \setminus ((\binom{n}{k} \cup \binom{n}{k+1}))\) the number of copies of the butterfly poset that contains only \(F\) from \(\mathcal{F} \setminus ((\binom{n}{k} \cup \binom{n}{k+1}))\) is at least \((\frac{27}{16} - o(1))f(n)\) and thus the number of butterflies in \(\mathcal{F}\) is much larger than \(f(n)E\).

**CASE III** \(50 \leq m < \frac{n}{10}\)
We try to imitate the proof of the second case of part (a). Applying Corollary 3.5 to \( F \cap \binom{n}{k} \) we obtain that the family \( F_{b,k+1} = \{ F \in F \cap \binom{n}{k+1} : |\Delta_k(F) \setminus F| \leq k+1-2\sqrt{m} \} \) has size at most \( \sqrt{m} \). Let us apply Corollary 3.5 again, this time to \( F_{g,k+1} = (F \cap \binom{n}{k+1}) \setminus F_{b,k+1} \). We obtain that the family \( F_{b,k+2} = \{ F \in \binom{n}{k+2} : |\Delta_{k+1}(F) \cap F_{g,k+1}| \leq k+2-2\sqrt{m} \} \) has size at most \( 2m+\sqrt{m} \leq 3m \). With an identical argument applied to \( F = \{ [n] \setminus F : F \in F \} \), one can show that the families \( F_{b,k} = \{ F \in F \cap \binom{n}{k} : |\nabla_{k+1}(F) \setminus F| \leq n-k-2\sqrt{m} \} \) and \( F_{b,k-1} = \{ F \in F \cap \binom{n}{k-1} : |\nabla_{k}(F) \cap (F \setminus F_{b,k})| \leq n-k+1-2\sqrt{m} \} \) both have size at most \( 3\sqrt{m} \).

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

\[
(E + m - 6\sqrt{m})(k + 2 - 2\sqrt{m}) \left( \frac{k + 1 - 2\sqrt{m}}{2} \right) \geq (E + \sqrt{m})(k + 2 - 2\sqrt{m}) \left( \frac{k + 1 - 2\sqrt{m}}{2} \right) \geq E f(n) + \sqrt{m} \cdot \frac{k^3}{4} - E \sqrt{m}(k+2)^2/2 \geq E f(n),
\]

where we used \( m - 6\sqrt{m} \geq \sqrt{m} \) as \( m \geq 50 \), \( k - 2\sqrt{m} = (1 - o(1))k \) as \( m \leq n/10 \) and also \( E \leq n/100 \).

**Case IV** \( 0 < m < 50 \)

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

The case when \( m \) equals 0, was dealt with in the Introduction by Proposition 1.2. □

## 4 Concluding remarks

As mentioned in the introduction, if we add sets \( G_1, G_2, \ldots, G_E \in \binom{n}{[n/2]+1} \) to the family \( F = (\binom{n}{[n/2]} \cup \binom{n}{[n/2]}) \) with the property that \( |G_i \cap G_j| \leq k-1 \) holds for all pairs \( 1 \leq i < j \leq E \), then the number of butterflies in \( F \cup \{ G_1, G_2, \ldots, G_E \} = Ef(n) \). The size of a largest
family of sets with this property is denoted by $K(n, \lceil n/2 \rceil + 1)$. We propose the following conjecture.

**Conjecture 4.1.** Let $E = E(n) \leq K(n, \lceil n/2 \rceil + 1)$. If $n$ is large enough, then the minimum number of butterflies a family $\mathcal{F} \subset 2^{[n]}$ of size $\Sigma(n, 2) + E$ must contain is $Ef(n)$.

If $k$ tends to infinity, determining the asymptotics of $K(n, k) = \max_{G \subset \binom{[n]}{k}} \{|G| : |G_1 \cap G_2| \leq k - 2 \ \forall G_1, G_2 \in G \}$ is one of the most important open problems in coding theory. When $k$ is roughly $n/2$, then by a trivial volume argument one has that $K(n, k) \geq (4 - o(1)) \frac{(n^2)}{n^2}$ and by a probabilistic argument one can obtain $K(n, k) = O\left(\frac{n}{\log n}^2\right)$. The bounds in almost all the steps towards Theorem 1.3 can be improved with a little work, but we do not see how to get anywhere near $K(n, \lceil n/2 \rceil + 1)$ even for the asymptotic statement.

An important step in the proof of Theorem 1.4 is Corollary 2.8 that bounds the size of a butterfly free family as a function of the number of missing middle sets. The bound of Corollary 2.8 is most probably very far from being sharp. We propose the following conjecture.

**Conjecture 4.2.** Let $F \subseteq 2^{[n]}$ be a butterfly-free family with $\emptyset, [n] \notin F$ and let us write $\mathcal{M} = \{M \in F : \exists F, F' \in F \ F \subset M \subset F'\}$. Then $|F| \leq \Sigma(n, 2) - \Omega(n|\mathcal{M}|)$ if $|\mathcal{M}|$ is not too large.

The exact value of $La(n, P)$ is known for very few posets $P$. There is a general conjecture that $La(n, P)/\binom{n}{\lceil n/2 \rceil}$ tends to an integer $e(P)$, where the value of $e(P)$ is defined to be the largest integer such that a family from $\Sigma^*(n, e(P))$ is $P$-free. One of the nicest results in the area was proved by Bukh [1] and states that this conjecture is true for posets $T$ of which the Hasse diagram is a tree (the Hasse diagram of a poset $P$ is the directed graph of which the vertices are elements of $P$ and $p$ and $q$ are joined by an edge directed towards $q$ if $p$ precedes $q$, i.e. $p <_P q$ and there does not exist $z \in P$ with $p <_P z <_P q$). As the exact value of $La(n, T)$ is not known even for tree posets $T$, it is not easy to formulate a conjecture on the minimum number of copies of $T$ in a family of size $La(n, T) + E$, but maybe the following structural statement can have a not very complicated proof.

**Conjecture 4.3.** Let $T$ be a poset of which the Hasse diagram is a tree. Then there exists a constant $c = c(T)$ with $0 < c \leq 1$ such that if $E \leq c \binom{n}{\lceil n/2 \rceil}$, then there exists a family $\mathcal{F} \subseteq 2^{[n]}$ that minimizes the number of copies of $T$ over families of size $La(n, T) + E$ and contains sets only of $e(T) + 1$ different sizes and the set sizes are consecutive integers.

**References**

[1] R. Ahlswede, Simple hypergraphs with maximal number of adjacent pairs of edges, J. Combinatorial Theory (B) 28 (1980), 164 – 167.
[2] R. Ahlswede, G.O.H. Katona, Graphs with maximal number of adjacent pairs of edges, Acta Mathematica Hungarica, 32.1 (1978), 97–120.

[3] B. Bollobás, I. Leader, Set systems with few disjoint pairs, Combinatorica 23.4 (2003), 559 – 570.

[4] B. Bukh, Set families with a forbidden subposet, Electronic J. of Combinatorics, 16 (2009), R142.

[5] S. Das, W. Gan, B. Sudakov, The minimum number of disjoint pairs in set systems and related problems, arxiv.org:1305.6715 (2013).

[6] S. Das, W. Gan, B. Sudakov, Sperner’s Theorem and a Problem of Erdős-Katona-Kleitman, arxiv.org:1302.5210 (2013).

[7] A. De Bonis, G.O.H. Katona, K. Swanepoel, Largest family without $A \cup B \subseteq C \cap D$, J. Combin. Theory Ser. A 111 (2005) 331 – 336.

[8] A. P. Dove, J. R. Griggs, R. J. Kang, J. S. Sereni, Supersaturation in the Boolean lattice, arxiv.org:1303.4336 (2013).

[9] P. Erdős, On a lemma of Littlewood and Offord, Bulletin of the American Mathematical Society 51 (1945), 898 – 902.

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

[11] P. Erdős, C. Ko, R. Rado, Intersection theorems for systems of finite sets, The Quarterly Journal of Mathematics 12.1 (1961), 313 – 320.

[12] P. Erdős, M. Simonovits, Supersaturated graphs and hypergraphs, Combinatorica 3 (1983), 181–192.

[13] P. Frankl, On the minimum number of disjoint pairs in a family of finite sets, J. Combinatorial Theory (A), 22 (1977), 249–251.

[14] L.H. Harper, On a problem of Kleitman and West, Discrete Mathematics 93 (1991) 169–182.

[15] G. Katona, A theorem of finite sets, Theory of Graphs, Proc. Coll. held at Tihany, Hungary, September, 1966 (Akadémiai Kiadó, Budapest, 1968) 187-207.

[16] G.O.H. Katona, T.G. Tarján, Extremal problems with excluded subgraphs in the n-cube, Graph Theory, Springer Berlin Heidelberg (1983), 84 – 93.

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

[18] J.B. Kruskal, The number of simplices in a complex, 1963 Mathematical optimization techniques pp. 251-278 Univ. of California Press, Berkeley, Calif.

[19] L. Lovász, Combinatorial problems and exercises. North-Holland Publishing Co., Amsterdam-New York, 1979. 551 pp.

[20] W. Mantel, Wiskundige Opgaven 10 (1906), 60 - 61.

[21] C. Reiher, The clique density theorem, arxiv.org:1212.2454 (2012).

[22] E. Sperner, Ein Satz über Untermengen einer endlichen Menge, Mathematische Zeitschrift 27 (1928), 544 – 548.