# Set families with a forbidden subposet

Boris Bukh *

Department of Pure Mathematics and Mathematical Statistics
Cambridge, CB3 0WA, UK

*and*

Churchill College
Cambridge, CB3 0DS, UK

B.Bukh@dpmms.cam.ac.uk

Submitted: Dec 14, 2008; Accepted: Nov 4, 2009; Published: Nov 30, 2009
Mathematics Subject Classification: 06A07, 05D05

**Abstract**

We asymptotically determine the size of the largest family $\mathcal{F}$ of subsets of $\{1,\ldots,n\}$ not containing a given poset $P$ if the Hasse diagram of $P$ is a tree. This is a qualitative generalization of several known results including Sperner's theorem.

## Introduction

We say that a poset $P$ is a *subposet* of a poset $P'$ if there is an injective map $f\colon P \to P'$ such that $a \leqslant_P b$ implies $f(a) \leqslant_{P'} f(b)$. A poset $P$ is an *induced subposet* of $P'$ if there is an injective map $f\colon P \to P'$ for which $a \leqslant_P b$ if and only if $f(a) \leqslant_{P'} f(b)$. For instance, $\Lambda$ is a subposet of $\vdots$, but not an induced subposet. For a poset $P$ a *Hasse diagram*, denoted by $H(P)$, is a graph whose vertices are elements of $P$, and $xy$ is an edge if $x < y$ and for no other element $z$ of $P$ we have $x < z < y$.

Let $[n] = \{1,\ldots,n\}$, and denote by $2^{[n]}$ the collection of all subsets of $[n]$. One can think of a family $\mathcal{F}$ of subsets of $[n]$ as a poset under inclusion. In this way $\mathcal{F}$ becomes an induced subposet of the Boolean lattice. In this paper we examine the size of the largest family $\mathcal{F} \subset 2^{[n]}$ subject to the condition that $\mathcal{F}$ does not contain a fixed finite subposet $P$. We do not require $P$ to be an induced subposet. A set family $\mathcal{F}$ not containing a subposet $P$ will be called a $P$-free family. We denote by $\operatorname{ex}(P,n)$ the size of the largest $P$-free family $\mathcal{F} \subset 2^{[n]}$. For example, the classical Sperner's theorem [Spe28] asserts that $\operatorname{ex}(:,n) = \binom{n}{\lfloor n/2 \rfloor}$.

*This work was done while the author studied at Princeton University. Its intellectual and financial support are gratefully acknowledged.

Erdős [Erd45] extended Sperner's result, and proved that if $C_l$ denotes the chain of length $l$, then $\mathrm{ex}(C_l, n)$ is equal to the sum of $l-1$ largest binomial coefficients of order $n$. Katona and Tarján[KT83] proved that $\mathrm{ex}(\wedge, n) = \binom{n}{\lfloor n/2 \rfloor}\bigl(1 + O(1/n)\bigr)$. A common generalization of results of Erdős and Katona-Tarján was established by Thanh[Tha98] who showed that if $P_{k,l}$ is any fixed poset with vertex set $\{1, \dots, k\} \cup \{1', \dots, l'\}$ in which the relations are $1 < 2 < \cdots < k$ and $1' < 1$, $2' < 1$, $\dots l' < 1$, then $\mathrm{ex}(P_{k,l}) = k\binom{n}{\lfloor n/2 \rfloor}\bigl(1 + O(1/n)\bigr)$ (the error term was subsequently improved by de Bonis and Katona[DBK07]). For example, $\mathrm{ex}(\curlywedge, n) = 2\binom{n}{\lfloor n/2 \rfloor}\bigl(1 + O(1/n)\bigr)$. In [DBKS05] it is shown that $\mathrm{ex}(\bowtie, n) = \binom{n}{\lfloor n/2 \rfloor} + \binom{n}{\lfloor n/2 \rfloor + 1}$. Griggs and Katona [GK08] proved that $\mathrm{ex}(\mathcal{N}, n) = \binom{n}{\lfloor n/2 \rfloor}\bigl(1 + O(1/n)\bigr)$. Recently, Griggs and Lu[GL] proved that $\mathrm{ex}(L_{4k}, n) = \binom{n}{\lfloor n/2 \rfloor}\bigl(1 + O(1/n)\bigr)$, where $L_{4k}$ is the loop of length $4k$ on two adjacent levels of the Boolean lattice.

Let $\mathrm{Mon}(\mathbb{Z})$ be the set of all functions $f \colon \mathbb{Z} \to \{0, 1\}$ such that $f(n) = 1$ and $f(-n) = 0$ for all sufficiently large $n$. The elements of $\mathrm{Mon}(\mathbb{Z})$ will be called eventually monotone functions. For $f, g \in \mathrm{Mon}(\mathbb{Z})$ write $f \leqslant g$ if $f(n) \leqslant g(n)$ for all $n$. Then $(\mathrm{Mon}(\mathbb{Z}), <)$ is a distributive lattice. Note that $\sum_n f(n) - g(n)$ is well-defined for every $f, g \in \mathrm{Mon}(\mathbb{Z})$. We define a *level* of $\mathrm{Mon}(\mathbb{Z})$ to be a maximal family $L \subset \mathrm{Mon}(\mathbb{Z})$ satisfying $\sum_n f(n) - g(n) = 0$ for all $f, g \in L$. Note that a level of $\mathrm{Mon}(\mathbb{Z})$ is an antichain. The poset $\mathrm{Mon}(\mathbb{Z})$ can be thought of as the induced subposet of $2^{\mathbb{Z}}$ spanned by the set $\{X \subset \mathbb{Z} : |X \triangle Y| < \infty\}$, for some fixed $Y \subset \mathbb{Z}$ which is neither finite nor cofinite.

The simplest explanation for all the results above is the following conjecture.

**Conjecture.** *For a finite poset $P$ let $l(P)$ be the maximum number of levels in $\mathrm{Mon}(\mathbb{Z})$ so that their union does not contain $P$ as a subposet, then*

$$\mathrm{ex}(P, n) = l(P)\binom{n}{\lfloor n/2 \rfloor}\bigl(1 + O(1/n)\bigr).$$

Intuitively the conjecture asserts that the largest $P$-free family is essentially the union of the maximum number of middle levels of the Boolean lattice $2^{[n]}$ not containing $P$. If true, the conjecture would be an analogue of Erdős-Stone-Simonovits theorem from the extremal graph theory which asserts that the largest graph not containing a given graph $G$ is essentially the largest complete partite graph not containing $G$

In this paper, we establish the conjecture whenever $H(P)$ is a tree, generalizing several of the results mentioned above. Unlike the papers above we are not concerned with establishing the best possible bounds inside the $O(1/n)$ term, which allows us to give a rather short proof. The rest of the paper is occupied by the proof of the following theorem.

**Theorem 1.** *If $P$ is a finite poset and $H(P)$ is a tree, then*

$$\mathrm{ex}(P, n) = (h(P) - 1)\binom{n}{\lfloor n/2 \rfloor}\bigl(1 + O(1/n)\bigr)$$

*where $h(P)$ is the height of $P$, i.e., the length of the longest chain in $P$. Moreover, $l(P) = h(P) - 1$ for such a $P$.*

For the case $h(P) = 2$, the theorem was also independently proved by Griggs and Lu[GL] by a very different argument.

## Proof idea

Before embarking on the proof of Theorem 1, we first non-rigorously sketch a simple proof for the special case $h(P) = 2$. The proof unfortunately does not generalize to $h(P) \geqslant 3$, but it will motivate the otherwise hard-to-follow technical details of the more general proof. We shall need a strengthening of Sperner's lemma, to the effect that if $|\mathcal{F}| > \binom{n}{\lfloor n/2 \rfloor}$, then not only there are pairs of comparable sets, but a plentitude of such pairs. The following statement is easy and can be proved similarly to Lemma 4 below (see [Kle68] for a sharper result for small $\epsilon$).

**Lemma 2.** *If $|\mathcal{F}| \geqslant (1+\epsilon)\binom{n}{\lfloor n/2 \rfloor}$, then there are at least $(1/10)n\epsilon|\mathcal{F}|$ pairs of sets $F_1 \subset F_2$ contained in $\mathcal{F}$.*

*Proof of the case $h(P) = 2$ of Theorem 1 assuming Lemma 2.* Suppose $\mathcal{F} \subset 2^{[n]}$ is a family of $|\mathcal{F}| \geqslant (1 + 20|P|/n)\binom{n}{\lfloor n/2 \rfloor}$ sets, where $P$ is a poset. We will show that $\mathcal{F}$ contains a copy of $P$. Let $G$ be the graph with vertex set $\mathcal{F}$ where a pair of distinct sets $F_1, F_2 \in \mathcal{F}$ connected by an edge if $F_1 \subset F_2$ or $F_2 \subset F_1$. By Lemma 2 the average degree of $G$ is at least $4|P|$. Since every graph of of average degree $d$ contains a non-empty subgraph of minimum degree at least $d/2$, the graph $G$ contains a subgraph $G'$ of minimum degree at least $2|P|$.

Then we shall embed elements of $P$ one-by-one into $V(G')$, in such a way that at each step embedding is injective, and preserves order. More precisely, we assume that $P$ is not a subposet of $V(G')$, and use this assumption to construct a sequence of embeddings $\pi_i\colon P_i \to V(G')$, where

a) $P_i$ is an induced subposet of $P$, and $H(P_i)$ is a tree,

b) $P_i \setminus P_{i-1}$ consists of a single element,

c) $v \leqslant_{P_i} u$ implies $\pi_i(v) \leqslant \pi_i(u)$ for all $u, v \in P_i$.

First, we embed some element of $P$ arbitrarily into $V(G')$, and let $P_1$ to consist of that single element. Then, for each $i \geqslant 2$ let $v \in P \setminus P_{i-1}$ be a not-yet-embedded element of $P$, which is comparable to some $u \in P_{i-1}$, and let $P_i = P_{i-1} \cup \{v\}$. Since $h(P) = 2$, and $H(P)$ is a tree, $u$ is the only element of $P_{i-1}$ comparable to $v$. We shall define embedding $\pi_i$ that agrees with $\pi_{i-1}$ on $P_{i-1} \setminus \{u\}$.

Without loss of generality we can assume that $u < v$. Set $\tau_1 = \pi_{i-1}$, and write $u_1 = \tau_1(u)$. Since $u_1 \in V(G')$, and degree of every vertex in $V(G')$ is at least $2|P| \geqslant |P|+1$ there is at least one neighbor $u_2$ of $u_1$ in $G'$ which is not in the image of $\tau_1$. If $u_1 < u_2$, then we let $\pi_i$ to be an extension of $\tau_1$ sending $v$ to $u_2$. Suppose $u_1 > u_2$. Let $\tau_2$ be the embedding obtained from $\tau_1$ by mapping $u$ to $u_2$ instead of $u_1$. The embedding $\tau_2$ satisfies the same properties (a)–(c) that $\pi_{i-1}$ does. As $2|P| \geqslant |P| + 2$ we can again find a neighbor $u_3$ of $u_2$ which is neither $u_1$ nor in the image of $\pi_2$, and we can assume that $u_2 > u_3$. Repeating this process $|P|$ times yields a chain $u_1 > u_2 > \ldots > u_{|P|}$ of elements of $V(G')$. However $P$ is a subposet (but not necessarily an induced subposet) of any linear extension of itself, and thus embeds into $\{u_1, \ldots, u_{|P|}\} \subset \mathcal{F}$. □

The proof above is clearly similar to the proof that every tree on $d$ vertices embeds into a graph of average degree $2d$. To extend the proof above to the case $h(P) = 3$, for example, it is tempting to replace the graph $G$ by a 3-uniform hypergraph of triples of sets in a chain. The problem with this approach is lack of any good replacement for the concept of minimum degree. The solution is therefore to eliminate minimum degree from the proof entirely. To see how it is done, we present an alternative way of embedding trees in graphs of large average degree. It is far more wasteful, but avoids minimum degrees.

**Proposition 3.** *For every tree $T$ there is a $d_0(T)$ such that $T$ embeds into every finite graph of average degree at least $d_0(T)$.*

*Proof.* The proof is by induction on $T$, with $|T| = 1$ being the trivial base case. Assume $|T| \geqslant 2$. Let $v$ be any leaf of $T$, and $u$ be its unique neighbor. Set $T' = T \setminus v$. We will show that we can take $d_0(T) = d_0(T') + 2|T|$. Suppose $G$ is of average degree at least $d_0(T') + 2|T|$. Let $B = \{v \in G : \deg_G(v) \leqslant |T|\}$. Then the graph $G' = G \setminus B$ is only $|B||T|$ edges smaller than $G$, and thus has has average degree at least $(d_0(T') + 2|T|) - 2|B||T|/|V| \geqslant d_0(T')$. By the induction hypothesis there is an embedding $\pi\colon T' \to G'$. As $\pi(u) \in G'$, we infer $\deg_G(\pi(u)) > |T|$, implying that there is at least one neighbor of $\pi(u)$ that is not in $\pi(T')$. Use this neighbor to extend the embedding $\pi$ of $T'$ to an embedding of $T' \cup \{v\} = T$. □

For technical reasons, it turns out to be easier to work with marked chains rather than hypergraphs; intuitively this change corresponds to hypergraphs with weighted edges. In the other respects, the proof of Theorem 1 is a straightforward generalization of the two arguments above.

# Proof of Theorem 1

By an *interval* in a poset $P$ we mean a set of the form $[x, y] = \{z \in P : x \leqslant z \leqslant y\}$. A *maximal chain* in a poset $P$ is a chain, which is not contained in any other chain. In particular, a maximal chain in $2^{[n]}$ is a chain of sets $\emptyset = S_0 \subset S_1 \subset \cdots \subset S_n = [n]$ with $|S_i| = i$. A *$k$-marked chain* with markers $F_1, \ldots, F_k$ is a $k+1$-tuple $(M, F_1, \ldots, F_k)$ where $M$ is a maximal chain in $2^{[n]}$, $F_1 \supset \cdots \supset F_k$ and $F_1, \ldots, F_k$ belong to $M$.

**Lemma 4.** *If $\mathcal{F} \subset 2^{[n]}$ is of size*

$$|\mathcal{F}| \geqslant (k - 1 + \epsilon)\binom{n}{\lfloor n/2 \rfloor},$$

*then there are at least $(\epsilon/k)n!$ $k$-marked chains whose markers belong to $\mathcal{F}$.*

*Proof.* Let $C_i$ be the number of maximal chains that contain exactly $i$ sets from $\mathcal{F}$. Counting the number of pairs $(F, M)$ where $F \in \mathcal{F}$ is an element of a maximal chain $M$ in two different ways, we obtain

$$\sum_i iC_i = \sum_{F\in\mathcal{F}} \frac{n!}{\binom{n}{|F|}} \geqslant |\mathcal{F}| \frac{n!}{\binom{n}{\lfloor n/2\rfloor}} \geqslant (k-1+\epsilon)n!.$$

From this and $\sum C_i = n!$, we infer that

$$\sum_{i\geqslant k} iC_i \geqslant \sum_i iC_i - (k-1)\sum_{i\leqslant k-1} C_i \geqslant \epsilon n!.$$

The number of $k$-marked chains with markers in $\mathcal{F}$ is

$$\sum_{i\geqslant k}\binom{i}{k}C_i = \sum_{i\geqslant k}\binom{i-1}{k-1}\frac{i}{k}C_i \geqslant \frac{1}{k}\sum_{i\geqslant k} iC_i \geqslant \frac{\epsilon}{k}n!. \qquad \square$$

Call a poset $P$ of height $k$ *saturated* if every maximal chain is of length $k$. For us the maximal chains in $P$ play the role analogous to the edges of a tree $T$ in Proposition 3. However, in general the edges might have different sizes, which is analogous to dealing with non-uniform hypergraphs. The saturated posets are the analogues of uniform hypergraphs.

The next two lemmas establish a couple of intuitively obvious, but annoyingly hard to rigorously prove facts about saturated posets whose Hasse diagram is a tree. The first lemma will be used to reduce the problem of embedding an arbitrary $P$ to the problem of embedding saturated $P$. The second lemma will allow us to do induction on $|P|$.

**Lemma 5.** *If $P$ is a finite poset with $H(P)$ being a tree, then $P$ is an induced subposet of some saturated finite poset $\tilde{P}$ with $H(\tilde{P})$ being a tree, and $h(P) = h(\tilde{P})$.*

*Proof.* For the purpose of this proof let $s(P)$ be the number of maximal chains in $P$. Since every element is contained in some maximal chain, $|P| \leqslant s(P)h(P)$, implying that for fixed $s$ and $k$ there only finitely many posets $P$ with $s(P) = s$ and $h(P) = k$. Assume there is a counterexample to the lemma. Let $P$ be a counterexample with the largest number of elements for given $s(P)$ and $h(P)$.

Since $P$ is not saturated there is a maximal chain $v_1 < \cdots < v_t$ in $P$ of length $t < k$. For each $i = 0, \ldots, t-1$ we can define a new poset $P_i$ which is obtained from $P$ by adding a new element $v$ and two new relations $v_i < v$ and $v < v_{i+1}$ (in the case $i = 0$ we add only one new relation). Clearly, $s(P_i) = s(P)$ and $P$ is an induced subposet of $P_i$. We will show by induction on $i$ that for each $i = 0, \ldots, t-1$ either $h(P_j) = h(P)$ for some $j \leqslant i$, or there is a chain in $P$ of length $k - i$ whose smallest element is $v_{i+1}$.

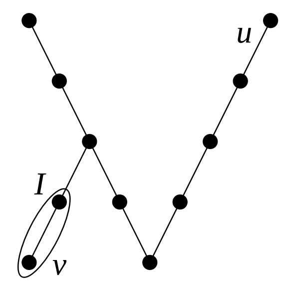
If $h(P_0) > k$, then there is a chain $C$ in $P_0$ of length $k+1$. Since $C$ is not a chain of $P$, it contains $v$. Thus $v_1 \in C$, and $C \setminus \{v\}$ contains $v_1$ and has length $k$. Now suppose $i \geqslant 1$ and we have established the inductive claim for all smaller values of $i$. If $h(P_j) > h(P)$ for all $j = 0, \ldots, i-1$, then by the induction hypothesis there is a chain $C \subset P$ of length

$k-i+1$ whose smallest element is $v_i$. Therefore, all chains in $P$ whose largest element is $v_i$ have lengths not exceeding $i$. Since the length of the chain $v_1 < \cdots < v_i$ is $i$, the assumption $h(P_i) > k$ implies that there is a chain in $P$ of length at least $k-i$ whose smallest element is $v_{i+1}$. This establishes the inductive claim.

Hence, if $h(P_i) > h(P)$ for all $i$, there is a chain of length $k-t+1 \geqslant 2$ whose smallest element is $v_t$. This contradicts the maximality of $v_1 < \cdots < v_t$, implying that $h(P_i) = h(P)$ for some $i$. However, $P$ was assumed to be the largest counterexample with given $s(P)$ and $h(P)$. Therefore, there are no counterexamples to the lemma. $\square$

**Lemma 6.** *Suppose $P$ is a saturated finite poset of height $h(P) = k \geqslant 2$, and $H(P)$ is a tree. Furthermore, assume that $P$ is not a chain. Then there is a $v \in P$, which is a leaf in $H(P)$, and an interval $I$ of length $|I| \leqslant k-1$ containing $v$ such that $H(P \setminus I)$ is a tree, and the poset $P \setminus I$ is a saturated poset of height $h(P)$.*

*Proof.* A sequence $v_1, \ldots, v_l \in P$ is said to be a *poset path* of length $l-1$ if $v_i$ and $v_{i+1}$ are comparable for all $i$. A *poset distance* between $v$ and $u$, denoted $\operatorname{pdist}(v,u)$, is the shortest length of a poset path connecting them.




Let $v$ and $u$ be a pair of leaves maximizing $\operatorname{pdist}(v,u)$, and let $v = v_1, v_2, v_3, \ldots, v_l = u$ be the shortest poset path between them. Observe that $\operatorname{pdist}(v,u) \geqslant 2$, for $P$ is not a chain. Without loss of generality we can assume that $v < v_2$, for we can consider the opposite poset otherwise. Let $I_0$ be the longest interval containing $v$, all of whose elements have degree at most two in $H(P)$. If $|I_0| < h(P)$, set $I = I_0$. If $|I_0| = h(P)$, set $I = I_0 \setminus \{\max I_0\}$, where $\max I_0$ is the largest element of $I_0$.

Let $C$ be an arbitrary maximal chain in $P \setminus I$. We will show that $C$ has length $h(P)$. Suppose $C$ contains an element $w$ which is incomparable with $I$. Let then $M$ be a maximal chain in $P$ containing $C$. Since $w$ is comparable to every element of $M$, the chain $M$ is disjoint from $I$, implying $C = M$ and $h(C) = h(M) = h(P)$. So we can suppose that all elements of $C$ are comparable to $I$.

If $|I_0| = h(P)$, then the only element in $P \setminus I$ which is comparable to $I$ is $\max I_0$. Since $P$ is not a chain, there is a chain of length two containing $\max I_0$, which is disjoint from $I$, contradicting the maximality of $C$. Hence, we can assume that $|I_0| < h(P)$. Let $w$ be any element of $P \setminus I$ comparable to an element $z$ of $I$. If $w < z$, then $\min([w,z] \cap I)$ has degree at least 3 in $H(P)$. If $z < w$ and $\max I \not< w$, then $\max([z,w] \cap I)$ has degree at least 3 in $H(P)$. In either case, we reach a contradiction with the choice of $I$. Hence $w$ exceeds all elements of $I$. Taking $w = \min C$, it follows that $\max I < \min C$. By the maximality of $C$, there is no $z \in P \setminus I$ satisfying $\max I < z < \min C$. Upon taking $w = v_2$, it also follows that $\max I < v_2$, implying $\min C \leqslant v_2$. If $v_2 = \min C$, then $C \cup \{v_3\}$ is also a chain, contradicting the maximality of $C$. Let $y = \max C$. The only path from $u$ to $y$ in $H(P)$ goes through $v_2$ and $\min C$. Therefore every poset path from $u$ to $y$ has to go through $v_2$ and $\min C$, implying $\operatorname{pdist}(y,u) > \operatorname{pdist}(v,u)$. Though the element $y$ needs not to be a leaf itself, if $z$ is any leaf of $H(P)$ such that the path from $u$ to $z$ goes through $y$, then $\operatorname{pdist}(z,u) \geqslant \operatorname{pdist}(y,u) > \operatorname{pdist}(v,u)$, contradicting the choice of $v$ and $u$. $\square$

The core of the proof of Theorem 1 is contained in the following lemma. In the lemma the family of $k$-marked chains $\mathcal{L}$ plays analogous role to the graph $G$ in the Proposition 3, with $\mathcal{F}$ being analogous to the vertex set of $G$. The condition that $\mathcal{F}$ does not contain a chain of length $K$ comes from the fact that every poset embeds into every sufficiently long chain.

**Lemma 7.** *Let $P$ be a saturated finite poset of height $h(P) = k \geqslant 2$, whose Hasse diagram is a tree. Suppose $\mathcal{F} \subset 2^{[n]}$ is a set family, such that no chain contains more than $K$ sets from $\mathcal{F}$, and all sets in $\mathcal{F}$ are of size between $n/4$ and $3n/4$. Moreover, suppose $\mathcal{L}$ is a family of $k$-marked chains with markers in $\mathcal{F}$ of size*

$$|\mathcal{L}| \geqslant \frac{\binom{|P|+1}{2} 4^{K+1}}{n} n!.$$

*Then there is an embedding of $P$ into $\mathcal{F}$ in which every maximal chain of $P$ is mapped to the set of markers of some $k$-marked chain in $\mathcal{L}$.*

*Proof.* The proof is by induction on $|P|$. If $P$ is the chain of length $k$, then finding the required embedding is easy: marked elements on any $L \in \mathcal{L}$ form the desired chain. Now suppose we want to embed $P$, and have already established the lemma for all smaller saturated posets. Use the preceding lemma to obtain a leaf $v$ and an interval $I \ni v$ such that $P \setminus I$ is a still a saturated poset of height $k$. By passing to the opposite poset to $P$ and replacing $\mathcal{F}$ by $\bar{\mathcal{F}} = \{[n] \setminus F : F \in \mathcal{F}\}$ if necessary, we can assume that $v$ is smaller than any element that is comparable with $v$. Let $C$ be a maximal chain containing $I$. Let $s = |C \setminus I| = k - |I|$. Note that $s \geqslant 1$.

Call a chain $F_1 \supset \cdots \supset F_s$ of length $s$ a *bottleneck* if there is a set $\mathcal{S} \subset \mathcal{F}$ with than $|P|$ elements such that for every $k$-marked chain in $\mathcal{L}$ of the form $(M, F_1, \ldots, F_s, F_{s+1}, \ldots, F_k)$ we have $\mathcal{S} \cap \{F_{s+1}, \ldots, F_k\} \neq \emptyset$. Such an $\mathcal{S}$ is said to be a *witness* to the fact that $F_1 \supset \cdots \supset F_s$ is a bottleneck. Note that without loss of generality, a witness contains only proper subsets of $F_s$. For each bottleneck $F_1 \supset \cdots \supset F_s$, let $\mathcal{S}(F_1, \ldots, F_s)$ be a fixed witness containing only proper subsets of $F_s$. Call a $k$-marked chain $(M, F_1, \ldots, F_k) \in \mathcal{L}$ *bad* if for some $s$ the chain $F_1 \supset \cdots \supset F_s$ is a bottleneck.

Consider any $s$-element set $R = \{r_1, \ldots, r_s\}$ of integers with $1 \leqslant r_1 < \cdots < r_s \leqslant K$. If $M$ is a maximal chain in $2^{[n]}$ containing at least $r_s$ elements from $\mathcal{F}$, let $F_1 \supset F_2 \supset \cdots \supset F_{r_s}$ be the $r_s$ largest of these elements. The subchain of $F_1 \supset F_2 \supset \cdots \supset F_{r_s}$ indexed by $R$ is $F_{r_1} \supset F_{r_2} \supset \cdots \supset F_{r_s}$, and we denote it by $C_R(M)$. If $C_R(M)$ is a bottleneck, and $\mathcal{L}$ contains a $k$-marked chain of the form $(M, \ldots)$, whose $s$ largest markers are $C_R(M)$, then we say that $M$ is *$R$-bad*. Intuitively, the $R$-bad chains correspond to the edges adjacent to the vertices of low degree in Proposition 3.

Pick a maximal chain $M$ in $2^{[n]}$ uniformly at random. Let $B_R$ be the event that $M$ is $R$-bad. We will estimate $\Pr[B_R]$ for each fixed $R$ individually.

One way to pick a random maximal chain $M$ of $2^{[n]}$ is to start with $[n]$ and remove elements one by one, each step choosing an element uniformly at random among the remaining elements. Thus one can generate chain $M$ in two stages. In the first stage, we remove elements from $[n]$ at random until either we encounter $r_s$ sets from $\mathcal{F}$, or until we run out of elements to remove. Denote by $T$ the chain obtained at the end of the first stage (it is not a maximal chain, unless we ran out of elements). In the second stage, we resume removing elements at random from $\min T$, until no elements are left. If $T$ is not maximal, then $C_R(M)$ is independent of what happens in the second stage, and $C_R(T)$ is defined in the obvious way.

If $T$ is a maximal chain, or $C_R(T)$ is not a bottleneck, then $B_R$ does not hold. Otherwise, let $\mathcal{S} = \mathcal{S}(C_R(T))$ be the witness that $C_R(T)$ is a bottleneck. Recall that $\mathcal{S} \subset \mathcal{F}$, $|\mathcal{S}| < |P|$ and $\mathcal{S}$ meets every $k$-chain in $\mathcal{L}$ whose top $s$ markers are $C_R(T)$. Let

$$\mathcal{T}_R = \{\text{chain } T_0 \text{ in } 2^{[n]} : |T_0 \cap \mathcal{F}| = r_s, C_R(T_0) \text{ is a bottleneck}\}.$$

The probability that $M$ meets $S$ is thus

$$\begin{aligned}
\Pr[M \cap \mathcal{S} \neq \emptyset \,|\, C_R(T) \text{ is a bottleneck}] &\leqslant \max_{T_0 \in \mathcal{T}_R} \Pr[M \cap \mathcal{S} \neq \emptyset | T = T_0] \\
&\leqslant \max_{T_0 \in \mathcal{T}_R} |\mathcal{S}(C_R(T_0))| \max_{\substack{F \in \mathcal{F} \\ F < \min T_0}} \Pr[F \in M | T = T_0] \\
&\leqslant |P| \max_{T_0 \in \mathcal{T}_R} \max_{\substack{F \in \mathcal{F} \\ F < \min T_0}} \Pr[F \in M | T = T_0] \\
&\leqslant |P| \max_{T \in \mathcal{T}_R} \max_{\substack{F \in \mathcal{F} \\ F < \min T_0}} \frac{1}{|F| + 1} \\
&\leqslant |P| \max_{F \in \mathcal{F}} \frac{1}{|F| + 1}, \\
&\leqslant 4|P|/n
\end{aligned}$$

where the fourth inequality follows because at the step before obtaining $F$, we have $|F| + 1$ choices as to which element to remove, with at most one choice yielding $F$. If $M \cap \mathcal{S} = \emptyset$, then there is no $k$-marked chain of the form $(M, \dots)$ in $\mathcal{L}$, and $B_R$ does not hold. Therefore

$$\begin{aligned}
\Pr[B_R] &= \Pr[C_R(T) \text{ is a bottleneck}] \Pr[B_R | C_R(T) \text{ is a bottleneck}] \\
&\leqslant \Pr[C_R(T) \text{ is a bottleneck}] \Pr[M \cap \mathcal{S} \neq \emptyset \,|\, C_R(T) \text{ is a bottleneck}] \leqslant 4|P|/n.
\end{aligned}$$

Since $R$ is a subset of $[K]$, the number of pairs $(M, R)$ where $M$ is an $R$-bad maximal chain is at most $(4|P|\binom{K}{s}/n)n!$. Since no chain contains more than $K$ elements of $\mathcal{F}$, every bad $k$-marked chain gives rise to one such pair $(M, R)$. Since $R \subset [K]$, every pair $(M, R)$ arises in at most $\binom{K}{s}$ ways, implying that there are no more than $(4|P|\binom{K}{s}^2/n)n! \leqslant (|P|4^{K+1}/n)n!$ bad $k$-marked chains in $\mathcal{L}$.

Let $\mathcal{L}'$ be the set of all good $k$-marked chains in $\mathcal{L}$. There are

$$|\mathcal{L}'| \geqslant |\mathcal{L}| - \frac{|P|4^{K+1}}{n} n! \geqslant \frac{4^{K+1}\left[\binom{|P|+1}{2} - |P|\right]}{n} n! = \frac{\binom{|P|}{2} 4^{K+1}}{n} n!$$

of them. By the induction hypothesis there is an embedding $\pi \colon P \setminus I \to \mathcal{F}$. Recall that $C$ was a maximal chain containing $I$, and look at $C \setminus I$. Since $P \setminus I$ is saturated,

$C \setminus I$ is contained in some chain $C'$ of length $k$ in $P \setminus I$. Therefore $\pi(C')$ is contained in some $L \in \mathcal{L}'$. Since all $k$-marked chains in $\mathcal{L}'$ are good, $\pi(C \setminus I)$ is not a bottleneck. In particular, since $|P \setminus I| < |P|$, we infer that $\pi(P \setminus I)$ is not a witness that $\pi(C \setminus I)$ is a bottleneck. Thus there is a $k$-marked chain $\tilde{L} \in \mathcal{L}$ containing $\pi(C \setminus I)$ as markers, but not containing any other element of $\pi(P \setminus I)$ as a marker. Therefore, we can map $I$ to the bottom $k - s$ markers of $\tilde{L}$, completing the desired embedding. $\square$

With most of the work already done, we are ready to prove our main result.

*Proof of Theorem 1.* If $h(P) = 1$, then $P$ is a single-element poset, and the theorem is trivially true. So, assume that $h(P) \geqslant 2$. Consider the case when $P$ is a saturated poset, and suppose

$$|\mathcal{F}| \geqslant (h(P)-1)\binom{n}{\lfloor n/2 \rfloor}\Big(1 + \frac{h(P)|P|^2 4^{|P|+2}}{n}\Big)$$

and $n$ is sufficiently large. We will show that $\mathcal{F}$ contains $P$. The number of sets $F \in 2^{[n]}$ with fewer than $n/4$ or more than $3n/4$ elements is equal $2^n$ times the probability that for a randomly chosen $F \in 2^{[n]}$ we have $||F| - n/2| > n/4$. Thus by Chernoff's inequality the number of such sets $F \in 2^{[n]}$ is at most $2^n \cdot 2\exp\big(-2(n/4)^2/n\big) = o\big(\binom{n}{\lfloor n/2 \rfloor}/n\big)$. Let $\mathcal{F}' = \{F \in F : |F - n/2| \leqslant n/4\}$. As our $n$ is sufficiently large,

$$|\mathcal{F}'| \geqslant (h(P)-1)\binom{n}{\lfloor n/2 \rfloor}\Big(1 + \frac{h(P)|P|^2 4^{|P|+1}}{n}\Big).$$

Therefore, from Lemma 4 and Lemma 7 applied with $k = h(P)$ and $K = |P|$ it follows that either there is an embedding of $P$ into $\mathcal{F}'$ or $\mathcal{F}'$ contains a chain $C$ of length $|P|$. In the latter case, we can find an embedding $\pi$ of $P$ into $\mathcal{F}'$ anyway by simply letting $\pi \colon P \to C$ be any linear extension of $P$.

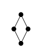
If $H(P)$ is a tree and $P$ is not a saturated poset, then by Lemma 5 it is contained in some saturated poset $P'$ of height $h(P') = h(P)$, such that $H(P')$ is a tree. Therefore $\mathrm{ex}(P, n) = (h(P)-1)\binom{n}{\lfloor n/2 \rfloor}(1 + O(1/n))$ for every poset $P$, for which $H(P)$ is a tree.

It remains to prove that $l(P) = h(P) - 1$. The inequality $l(P) \geqslant h(P) - 1$ is clear, as a union of $h(P) - 1$ levels does not contain a chain of length exceeding $h(P) - 1$, and hence does not contain $P$. Let $L_1, \ldots, L_h$ be $h$ distinct levels of $\mathrm{Mon}(\mathbb{Z})$, and let $L$ be their union. Suppose furthermore that the levels $L_i$ are so ordered that for any functions $f_i \in L_i$, the inequality $\sum_n f_i(n) - f_j(n) > 0$ holds whenever $i > j$ (by the definition of a level, if the inequality holds between a pair functions in levels $L_i$ and $L_j$, then it holds for all pairs).

To complete the proof, we need to exhibit an embedding of $P$ into $L$. By Lemma 5 it suffices to treat the case when $P$ is saturated. We will prove the existence of the embedding by induction on $|P|$. If $P$ is a chain of length $h$, then the embedding is obvious. Suppose $P$ is not a chain, we can find embedding for all smaller saturated $P$ of height $h$. By Lemma 6 there is a leaf $v$ and an interval $I$ of the form $I = [v, u)$ such that $P \setminus I$ is a saturated poset of height $h$. By induction $P \setminus I$ is embeddable into $L$. Fix any such embedding. Since $\pi(u)$ is contained in a chain of length $k$ in $L$, and $P$ is saturated, it follows that $\pi(u) \in L_{|I|+1}$. Let $n_0$ be a large enough that $(\pi(w))(n) = 1$ for all $w \in P \setminus I$ and $n \geqslant n_0$. Complete the embedding by mapping the interval $I$ to the interval of functions $f_1, \ldots, f_{|I|} \in \mathrm{Mon}(\mathbb{Z})$ defined by

$$f_i(n) = \begin{cases} 0, & \text{if } n_0 \leqslant n \leqslant n_0 + i - 1, \\ (\pi(w))(n), & \text{otherwise.} \end{cases} \qquad \square$$

# Concluding remarks

Though it would be interesting to determine exactly or find very good asymptotic estimates for $\mathrm{ex}(P, n)$ in general, a first step is to find the leading term in the asymptotic. In this paper we found the leading term of $\mathrm{ex}(P, n)$ whenever $H(P)$ is a tree. For some posets $P$ whose Hasse diagram is not a tree, one can find a poset $P'$ that contains $P$ and whose Hasse diagram is a tree with $l(P) = l(P')$, to obtain that $\mathrm{ex}(P, n) \sim \mathrm{ex}(P', n)$. For example, ex(, $n$) $\sim$ ex(, $n$), and similarly for other complete two-level posets, thus recovering the results from [DBK07, Section 5]. The simplest two posets that are not subposets of trees with the same value of $l(P)$ are  and , and the asymptotics of the function ex for these posets is not known.

It is conceivable that the conjecture in this paper is even true if its premise that $\mathcal{F}$ does not contain a subposet $P$ is replaced by the weaker premise that $\mathcal{F}$ does not contain $P$ as an induced subposet.

**Acknowledgement.** I thank Máté Matolcsi for reading a preliminary version of this paper, and two referees for useful suggestions.

# References


[DBK07] Annalisa De Bonis and Gyula O. H. Katona. Largest families without an $r$-fork. *Order*, 24(3):181–191, 2007.

[DBKS05] Annalisa De Bonis, Gyula O. H. Katona, and Konrad J. Swanepoel. Largest family without $A \cup B \subseteq C \cap D$. *J. Combin. Theory Ser. A*, 111(2):331–336, 2005. arXiv:math/0407373v1.

[Erd45] P. Erdös. On a lemma of Littlewood and Offord. *Bull. Amer. Math. Soc.*, 51:898–902, 1945.

[GK08] Jerrold R. Griggs and Gyula O. H. Katona. No four subsets forming an N. *J. Combin. Theory Ser. A*, 115(4):677–685, 2008. `http://www.math.sc.edu/~IMI/technical/07papers/0704.pdf`.

[GL] Jerrold R. Griggs and Linyuan Lu. On families of subsets with a forbidden subposet. arXiv:0807.3702v1.

[Kle68] D. Kleitman. A conjecture of Erdős-Katona on commensurable pairs among subsets of an $n$-set. In *Theory of Graphs (Proc. Colloq., Tihany, 1966)*, pages 215–218. Academic Press, New York, 1968.

[KT83] G. O. H. Katona and T. G. Tarján. Extremal problems with excluded subgraphs in the $n$-cube. In *Graph theory (Łagów, 1981)*, volume 1018 of *Lecture Notes in Math.*, pages 84–93. Springer, Berlin, 1983.

[Spe28] Emanuel Sperner. Ein Satz über Untermengen einer endlichen Menge. *Math. Z.*, 27(1):544–548, 1928.

[Tha98] Hai Tran Thanh. An extremal problem with excluded subposet in the Boolean lattice. *Order*, 15(1):51–57, 1998.