Combinatorics, Probability and Computing (2012) 21, 496–511.
doi:10.1017/S0963548312000028

Set Families With a Forbidden Induced Subposet

EDWARD BOEHNLEIN and TAO JIANG

Department of Mathematics, Miami University, Oxford, OH 45056, USA
(e-mail: boehnlei@math.sc.edu, jiangt@muohio.edu)

Received 6 June 2011; revised 23 January 2012; first published online 22 March 2012

For each poset H whose Hasse diagram is a tree of height k, we show that the largest size of a family $\mathcal{F}$ of subsets of $[n] = \{1, \ldots, n\}$ not containing H as an induced subposet is asymptotic to $(k-1)\binom{n}{\lfloor n/2 \rfloor}$. This extends a result of Bukh [1], which in turn generalizes several known results including Sperner's theorem.

AMS 2010 *Mathematics subject classification*: Primary 06A07; 05D05

1. Introduction

A poset $G = (S, \leqslant)$ is a set S equipped with a partial ordering $\leqslant$. We say that a poset $G = (S, \leqslant)$ contains another poset $H = (S', \leqslant')$ as a *subposet*, and write $H \subseteq G$ if there exists an injection $f : S' \to S$ such that, for all $u, v \in S'$, if $u \leqslant' v$ then $f(u) \leqslant f(v)$. We say that $G = (S, \leqslant)$ contains $H = (S', \leqslant')$ as an *induced subposet*, and write $H \subseteq^* G$ if there exists an injection $f : S' \to S$ such that, for all $u, v \in S'$, $u \leqslant' v$ if and only if $f(u) \leqslant f(v)$.

Given a positive integer n, let $[n] = \{1, 2, \ldots, n\}$. The *Boolean lattice* $\mathbb{B}_n$ of order n is the poset $(2^{[n]}, \subseteq)$. Throughout this paper we automatically equip any family $\mathcal{F} \subseteq 2^{[n]}$ with the containment relation $\subseteq$ and thus view $\mathcal{F}$ as a subposet of $\mathbb{B}_n$. Given a positive integer n and a poset H, let $\mathrm{La}(n, H)$ denote the largest size of a family $\mathcal{F} \subseteq \mathbb{B}_n$ that does not contain H as a subposet. Let $\mathrm{La}^*(n, H)$ denote the largest size of a family $\mathcal{F} \subseteq \mathbb{B}_n$ that does not contain H as an induced subposet. Note that in some papers, such as in Bukh [1], $\mathrm{ex}(H, n)$ is used instead of $\mathrm{La}(n, H)$, which is perhaps a more natural notation, since this is indeed a Turán function. However, in this paper we will inherit the $\mathrm{La}(n, H)$ notation that is used in most of the earlier papers on the subject. The study of these functions dates back to Sperner's theorem [7], which asserts that the largest size of an antichain in the Boolean lattice of order n equals $\binom{n}{\lfloor n/2 \rfloor}$, with equality attained by taking the middle level of the Boolean lattice. If we use P_2 to denote a chain of two elements, then Sperner's theorem says that $\mathrm{La}(n, P_2) = \mathrm{La}^*(n, P_2) = \binom{n}{\lfloor n/2 \rfloor}$. Erdős [5] extended Sperner's theorem to show that $\mathrm{La}(n, P_k)$, where P_k is the chain of k elements, is the sum of the $k-1$ middle binomial coefficients in n (*i.e.*, the sum of the sizes of the middle $k-1$

levels of $\mathbb{B}_n$). Consequently,

$$\lim_{n\to\infty} \frac{\mathrm{La}(n,P_k)}{\binom{n}{\lfloor n/2\rfloor}} = k-1.$$

A systematic study of $\mathrm{La}(n,H)$ started a few years ago, and a series of results on $\mathrm{La}(n,H)$ were developed. In most of these results H is a poset whose Hasse diagram is a tree or H is a height-2 poset, where the *height* of H is the largest cardinality of a chain in H. We give a brief review of some of these results. Let V_k denote the the height-2 poset that consists of $k+1$ elements $A, B_1, \ldots, B_k$ where, for all $i \in [k]$, $A \leqslant B_i$. We call V_k the *k-fork*. Improving earlier results of Thanh [8], De Bonis and Katona [3] showed that $\mathrm{La}(n,V_k) = \binom{n}{\lfloor n/2\rfloor}(1+\frac{k-1}{n}+\Theta(\frac{1}{n^2}))$. Let B denote the *butterfly* poset on four elements A_1, A_2, B_1, B_2, where, for all $i, j \in [2]$, $A_i \leqslant B_j$. De Bonis, Katona and Swanepoel [4] showed that $\mathrm{La}(n,B) = \binom{n}{\lfloor n/2\rfloor} + \binom{n}{\lfloor n/2\rfloor+1}$. More generally, for $r,s \geqslant 2$ let $K_{r,s}$ denote the two-level poset consisting of elements $A_1, \ldots, A_r, B_1, \ldots, B_s$ where, for all $i \in [r], j \in [s]$, $A_i \leqslant B_j$. De Bonis and Katona [3] showed that $\mathrm{La}(n,K_{r,s}) \sim 2\binom{n}{\lfloor n/2\rfloor}$, as $n \to \infty$. Extending earlier results on tree-like posets, Griggs and Lu [6] showed that if T is any height-2 poset whose Hasse diagram is a tree, then $\mathrm{La}(n,T) \sim \binom{n}{\lfloor n/2\rfloor}$. Independently, Bukh [1] obtained the following more general result.

Theorem 1.1 (Bukh [1]). *If H is a finite poset whose Hasse diagram is a tree of height $k \geqslant 2$, then*

$$\mathrm{La}(n,H) = (k-1)\binom{n}{\lfloor n/2\rfloor}(1+O(1/n)).$$

Note that Bukh's result generalizes (in a loose sense) all prior results on posets whose Hasse diagram is a tree. Furthermore, it also implies De Bonis and Katona's result that $\mathrm{La}(n,K_{r,s}) \leqslant 2\binom{n}{\lfloor n/2\rfloor}(1+O(\frac{1}{n}))$, for the following reason. Consider the three-level poset H that consists of elements $A_1, \ldots, A_r, B, C_1, \ldots, C_t$ where, for all $i \in [r]$, $A_i \leqslant B$, and for all $j \in [t]$, $B \leqslant C_j$. By transitivity, for all $i \in [r], j \in [t]$, $A_i \leqslant C_j$, and so H contains $K_{r,s}$ as a subposet. So, $\mathrm{La}(n,K_{r,s}) \leqslant \mathrm{La}(n,H) \leqslant 2\binom{n}{\lfloor n/2\rfloor}(1+O(1/n))$.

In this paper we are concerned with finding (or avoiding, depending on the perspective) induced subposets in $\mathbb{B}_n$. Generally speaking, induced subposets are harder to force, since we need to enforce non-containment as well as containment among corresponding members. For instance, for a family $\mathcal{F} \subseteq \mathbb{B}_n$ to contain the 2-fork V_2 as an induced subposet, we need to find three members of A, B, C of $\mathcal{F}$ satisfying $A \subseteq B, A \subseteq C$, $B \nsubseteq C$, and $C \nsubseteq B$. By comparison, for $\mathcal{F}$ to contain V_2 just as a subposet, we only need to ensure the existence of $A, B, C \in \mathcal{F}$ satisfying $A \subseteq B, A \subseteq C$.

Since a family $\mathcal{F} \subseteq \mathbb{B}_n$ that does not contain H as a subposet certainly does not contain H as an induced subposet, we always have $\mathrm{La}^*(n,H) \geqslant \mathrm{La}(n,H)$. In general, the determination of $\mathrm{La}^*(n,H)$ seems to be harder than the determination of $\mathrm{La}(n,H)$. The only result on $\mathrm{La}^*(n,H)$ that we are aware of is due to Carroll and Katona [2], who showed that

$$\binom{n}{\lfloor n/2\rfloor}\left(1+\frac{1}{n}+\Omega\left(\frac{1}{n^2}\right)\right) \leqslant \mathrm{La}^*(n,V_2) \leqslant \binom{n}{\lfloor n/2\rfloor}\left(1+\frac{2}{n}+O\left(\frac{1}{n^2}\right)\right).$$

For a lower bound on $\mathrm{La}^*(n,H)$, let $\mathcal{F}$ consist of the middle $k-1$ levels of the Boolean lattice B_n. Clearly $\mathcal{F}$ does not contain H (as an induced subposet) and $|\mathcal{F}| = \binom{n}{\lfloor n/2 \rfloor}(k-1-O(1/n))$. So $\mathrm{La}^*(n,H) \geqslant \binom{n}{\lfloor n/2 \rfloor}(k-1-O(1/n))$. We shall prove the following upper bound.

Theorem 1.2. *Let H be a finite poset whose Hasse diagram is a tree of height $k \geqslant 2$. Then*

$$\mathrm{La}^*(n,H) \leqslant \left(k-1+O\left(\frac{\sqrt{n\ln n}}{n}\right)\right) \cdot \binom{n}{\lfloor n/2 \rfloor}.$$

As an immediate consequence of Theorem 1.2 and the lower bound discussion, we obtain the following extension of Bukh's result (to an induced version).

Corollary 1.3. *Let H be a finite poset whose Hasse diagram is a tree of height $k \geqslant 2$. Then*

$$\mathrm{La}^*(n,H) = (k-1)\binom{n}{\lfloor n/2 \rfloor}(1+o(1)).$$

Note that our error term estimates on $\mathrm{La}^*(n,H)$ are weaker than Bukh's on $\mathrm{La}(n,H)$. It would be interesting to sharpen our bounds on the error term. To prove Theorem 1.2, we first make a quick reduction. As mentioned in [6], using Chernoff's inequality, it is easy to show that the number of sets $F \in 2^{[n]}$ satisfying $||F| - \frac{n}{2}| > 2\sqrt{n \ln n}$ is at most $O(\frac{1}{n^{3/2}}\binom{n}{\lfloor n/2 \rfloor})$. Define

$$\widetilde{\mathbb{B}}_n = \left\{ v \in \mathbb{B}_n : |v| \in \left[\frac{n}{2} - 2\sqrt{n\ln n},\ \frac{n}{2} + 2\sqrt{n \ln n}\right]\right\}.$$

By our discussion above, there are only $O(\frac{1}{n^{3/2}}\binom{n}{\lfloor n/2 \rfloor})$ members of $\mathbb{B}_n$ that lie outside $\widetilde{\mathbb{B}}_n$. So to prove Theorem 1.2 it suffices to prove the following.

Theorem 1.4. *Let $k, h \geqslant 2$ be integers. There exist constants $n_0 = n_0(k,h)$ and $c_{k,h}$ such that the following is true for all $n \geqslant n_0$. Let H be a poset whose Hasse diagram is an h-vertex tree of height k. Let $\mathcal{F} \subseteq \widetilde{\mathbb{B}}_n$ be a family with $|\mathcal{F}| \geqslant (k-1+\frac{c_{k,h}\sqrt{n\ln n}}{n})\binom{n}{\lfloor n/2 \rfloor}$. Then $\mathcal{F}$ contains H as an induced subposet.*

For the rest of the paper, we prove Theorem 1.4.

2. Preliminaries

In this section, we recall some facts from [1] which will be used in our main arguments. Given a poset H, let $D(H)$ denote its Hasse diagram. We call a poset H *k-saturated* if every maximal chain is of length k. Thus, in particular, H has height k.

Lemma 2.1 ([1]). *If H is a finite poset with $D(H)$ being a tree of height k, then H is an induced subposet of some saturated finite poset $\widetilde{H}$ with $D(\widetilde{H})$ being a tree of height k.*

Due to Lemma 2.1, for the rest of the paper we will assume that H is k-saturated. Let H be a poset and $x, y \in H$ where $x \leqslant y$. Define $[x,y] = \{z \in H : x \leqslant z \leqslant y\}$ and call it an *interval*. An interval in H that is a chain is called a *chain interval*. The statement we give below is equivalent to the original one in [1].

Lemma 2.2 ([1]). *Let $k \geqslant 2$. Suppose H is a k-saturated poset that is not a chain and $D(H)$ is a tree. There exists $v \in H$, which is a leaf in $D(H)$, and a chain interval $I = [v, u]$ or $[u, v]$ of length $|I| \leqslant k$ containing v, such that $D(H \setminus I')$ is a tree and the poset $H \setminus I'$ is k-saturated, where $I' = I - \{u\}$.*

Fix a positive integer k. A k-chain in $\mathbb{B}_n$ is just a chain in $\mathbb{B}_n$ with k distinct members. A *full chain* of a Boolean lattice $\mathbb{B}_m$ of order m is just a chain of length $m+1$. So it starts with the top element of the lattice and ends with bottom element of the lattice and contains a member of each cardinality between 0 and m. Let $\mathcal{F} \subseteq \mathbb{B}_n$ be a family. Given a k-chain $Q = (F_1, \ldots, F_k)$, where $F_1 \supset F_2 \supset \cdots \supset F_k$ and, for all $i \in [k], F_i \in \mathcal{F}$, and a full chain M of $\mathbb{B}_n$ that contains Q, we call the pair (M, Q) a *k-marked chain* with markers in $\mathcal{F}$. We call M the *host* of the k-marked chain (M, Q) and say that M *hosts* (M, Q). Throughout our paper, the family $\mathcal{F}$ is fixed. So, if we omit the phrase 'with markers in $\mathcal{F}$', it should be understood that the markers (the F_i) are in $\mathcal{F}$. Note that if M and M' are two distinct full chains of $\mathbb{B}_n$ that contain Q, then (M, Q) and (M', Q) are in fact considered to be two distinct k-marked chains in our definition. The following lemma is a claim contained in the proof of Lemma 4 in [1] (Lemma 2.4 below). We paraphrase it slightly as follows. Recall that $\binom{x}{k}$ is defined to be 0 when $x < k$.

Lemma 2.3. *Let $k \geqslant 2$ and let $\mathcal{F} \subseteq \mathbb{B}_n$. Let $\mathcal{C}(\mathbb{B}_n)$ denote the set of all $n!$ full chains of $\mathbb{B}_n$. For each $M \in \mathcal{C}(\mathbb{B}_n)$, let $x(M)$ denote the number of members of $\mathcal{F}$ contained in M. Let $\mathcal{L}$ denote the family of all the k-marked chains with markers in $\mathcal{F}$. Then*

$$|\mathcal{L}| = \sum_{M \in \mathcal{C}(\mathbb{B}_n)} \binom{x(M)}{k}.$$

Proof. Given any $M \in \mathcal{C}(\mathbb{B}_n)$, M hosts exactly $\binom{x(M)}{k}$ many k-marked chains with markers in $\mathcal{F}$. So there are altogether $\sum_{M \in \mathcal{C}(\mathbb{B}_n)} \binom{x(M)}{k}$ many k-marked chains with markers in $\mathcal{F}$. $\square$

The following lemma is established in [1]. We rephrase the proof slightly differently.

Lemma 2.4 ([1]). *Let ϵ be a positive real. Let $\mathcal{F} \subseteq \mathbb{B}_n$. Let $\mathcal{L}$ denote the family of all the k-marked chains with markers in $\mathcal{F}$. If $|\mathcal{F}| \geqslant (k - 1 + \epsilon)\binom{n}{\lfloor n/2 \rfloor}$. Then*

$$|\mathcal{L}| \geqslant (\epsilon/k)k!.$$

Proof. For each i, let C_i denote the number of full chains M of $\mathbb{B}_n$ with $x(M) = i$. Let X be the random variable that counts the number of members of $\mathcal{F}$ contained in a random full chain M of $\mathbb{B}_n$. For each member $F \in \mathcal{F}$, the probability that M contains F is precisely $\frac{1}{\binom{n}{|F|}}$. Hence

$$\mathbb{E}(X) = \sum_{F \in \mathcal{F}} \frac{1}{\binom{n}{|F|}} \geqslant |\mathcal{F}| \cdot \frac{1}{\binom{n}{\lfloor n/2 \rfloor}} \geqslant k - 1 + \epsilon.$$

On the other hand, by a direct counting argument we have $\mathbb{E}(X) = \sum_i iC_i/n!$. Thus, $\sum_i iC_i \geqslant (k-1+\epsilon)n!$. Clearly, $\sum_{i=1}^{k-1} iC_i \leqslant (k-1)n!$. So, $\sum_{i=k}^{n} iC_i \geqslant \epsilon n!$. For $i \geqslant k$, we have $\binom{i}{k} \geqslant \frac{i}{k}$. By Lemma 2.3, the number of k-marked chains with members in $\mathcal{F}$ equals

$$\sum_i C_i \binom{i}{k} \geqslant \sum_{i=k}^{n} C_i(i/k) = (1/k)\sum_{i=k}^{n} iC_i \geqslant (\epsilon/k)n!. \qquad \square$$

3. Forbidden neighbourhoods

Recall that elements of $\mathbb{B}_n$ are subsets of $[n]$. We refer to elements of $\mathbb{B}_n$ as *vertices* in the lattice. If v is a vertex in $\mathbb{B}_n$, it is also understood to be the subset of $[n]$ that it represents. The *cardinality* or *weight* of v, denoted by $|v|$, is the cardinality of the subsets of $[n]$ that v represents. Even though the partial ordering associated with $\mathbb{B}_n$ is the containment $\subseteq$ relation, we will continue to denote it by $\leqslant$ in most cases. If $u, v \in \mathbb{B}_n$ and $u \leqslant v$, we call u a *descendant of* v and we call v an *ancestor of* u. Given a vertex v in $\mathbb{B}_n$, the *down-set* $D(v)$ of v is defined to be

$$D(v) = \{u \in \mathbb{B}_n : u \leqslant v\}.$$

In other words, $D(v)$ is the set of all descendants of v. Note that if $|v| = m$, then $D(v)$ forms a Boolean lattice $\mathbb{B}_m$ of order m. The *up-set* $U(v)$ of v is defined to be

$$U(v) = \{u \in \mathbb{B}_n : v \leqslant u\}.$$

In other words, $U(v)$ is the set of all ancestors of v. Note that if $|v| = m$, then $U(v)$ forms a Boolean lattice $\mathbb{B}_{n-m}$ of order $n-m$. If S is a set of vertices in $\mathbb{B}_n$, we define

$$D(S) = \bigcup_{v \in S} D(v) \quad \text{and} \quad U(S) = \bigcup_{v \in S} U(v).$$

Given a vertex $v \in \widetilde{\mathbb{B}}_n$ and a set $S \subseteq \widetilde{\mathbb{B}}_n$ with $S \cap U(v) = \emptyset$, define

$$D^*(v,S) = [(D(v) \setminus \{v\}) \cap (U(S) \cup D(S))] \cap \widetilde{\mathbb{B}}_n. \tag{3.1}$$

We call $D^*(v,S)$ the *forbidden neighbourhood* of S *below* v in $\widetilde{\mathbb{B}}_n$. Given a vertex $v \in \widetilde{\mathbb{B}}_n$ and a set $S \subseteq \widetilde{\mathbb{B}}_n$ with $S \cap D(v) = \emptyset$, let

$$U^*(v,S) = [(U(v) \setminus \{v\}) \cap (U(S) \cup D(S))] \cap \widetilde{\mathbb{B}}_n. \tag{3.2}$$

We call $U^*(v,S)$ the *forbidden neighbourhood* of S *above* v in $\widetilde{\mathbb{B}}_n$.

It is crucial to note in the above definitions that sets $D^*(v,S)$ and $U^*(v,S)$ are both contained in $\widetilde{\mathbb{B}}_n$ and so is $\mathcal{F}$. The next two lemmas play an important role in our arguments.

Lemma 3.1. *Let $n \geqslant 2000$. Let $v \in \widetilde{\mathbb{B}}_n, S \subseteq \widetilde{\mathbb{B}}_n$, where $S \cap U(v) = \emptyset$ and $|S| \leqslant n/6$. Let M be a uniformly chosen random full chain of $D(v)$ (among all $|v|!$ full chains of $D(v)$). We have*

$$\mathbb{P}\big(M \cap (D^*(v,S)) \neq \emptyset\big) \leqslant \frac{39|S|\sqrt{n \ln n}}{n}.$$

Proof. It is easy to check that $\frac{n}{2} - 2\sqrt{n \ln n} > \frac{n}{3}$ for all $n \geqslant 2000$. Let $s = |S|$. For any vertex w in $(D(v) \setminus \{v\}) \cap \widetilde{\mathbb{B}}_n$, the probability that M contains w is

$$\frac{1}{\binom{|v|}{|w|}} \leqslant \frac{1}{|v|} \leqslant \frac{1}{n/3} = \frac{3}{n}.$$

Since $|(S \cap (D(v) \setminus \{v\})) \cap \widetilde{\mathbb{B}}_n| \leqslant s$, then

$$\mathbb{P}\big(M \cap [(S \cap (D(v) \setminus \{v\})) \cap \widetilde{\mathbb{B}}_n] \neq \emptyset\big) \leqslant \frac{3s}{n}. \tag{3.3}$$

Let $\ell = |v| - (\frac{n}{2} - 2\sqrt{n \ln n})$. Since $v \in \widetilde{\mathbb{B}}_n$, $\ell \leqslant 4\sqrt{n \ln n}$. To bound the probability that M intersects $U = (U(S) \cap (D(v) \setminus \{v\})) \cap \widetilde{\mathbb{B}}_n = U(S) \cap (D(v) \setminus \{v\})$, we first bound the probability that M is disjoint from U. Note that $U = U(S \cap D(v)) \cap (D(v) \setminus \{v\})$, since only a descendant of v may have an ancestor in $D(v) \setminus \{v\}$. Suppose $S \cap D(v) = \{y_1, \ldots, y_p\}$, where $p \leqslant s$. By our assumptions, for all $i \in [p]$, $y_i \leqslant v$ and $|v| - |y_i| \leqslant \ell$ (since $y_i \in \widetilde{\mathbb{B}}_n$). When we view $v, y_1, \ldots, y_p$ as sets we have $|\bigcap_{i=1}^p y_i| \geqslant |v| - p\ell$. Being a full chain of $D(v)$, we may view M as being obtained by starting with the set v and successively removing an element in it. For M not to enter $U(\{y_1, \ldots, y_p\}) \setminus \{v\}$, it suffices that the first element removed from the set v is in $\bigcap_{i=1}^p y_i$. So the probability that M does not intersect U is at least $|\bigcap_{i=1}^p y_i|/|v| \geqslant 1 - (p\ell/|v|) \geqslant 1 - s\ell/|v|$. Therefore

$$\mathbb{P}(M \cap [(U(S) \cap (D(v) \setminus \{v\})) \cap \widetilde{\mathbb{B}}_n] \neq \emptyset) \leqslant \frac{s\ell}{|v|} \leqslant \frac{4s\sqrt{n \ln n}}{n/3} = \frac{12s\sqrt{n \ln n}}{n}. \tag{3.4}$$

Next, we bound the probability that M intersects $D = (D(S) \cap (D(v)) \setminus \{v\})) \cap \widetilde{\mathbb{B}}_n$. Again, we first bound the probability that M is disjoint from D. Suppose $S = \{z_1, \ldots, z_s\}$. Since $S \cap U(v) = \emptyset$, for all $i \in [s]$, we have $v \nleqslant z_i$. So set v has an element u_i that is not in set z_i. When we form M by successively removing elements of set v, as long as each of the first ℓ steps removes an element outside $\{u_1, \ldots, u_s\}$, M would not enter D. Note that for all $0 < x \leqslant \frac{1}{2}$, $(1-x)^\ell \geqslant e^{-2x\ell} \geqslant 1 - 2x\ell$. Now

$$\begin{aligned}
&\mathbb{P}(M \cap [(D(S) \cap (D(v) \setminus \{v\})) \cap \widetilde{\mathbb{B}}_n] = \emptyset) \\
&\quad \geqslant \frac{(|v|-s)(|v|-s-1)\cdots(|v|-s-\ell+1)}{|v|(|v|-1)\cdots(|v|-\ell+1)} \\
&\quad = \left(1 - \frac{s}{|v|}\right)\left(1 - \frac{s}{|v|-1}\right)\cdots\left(1 - \frac{s}{|v|-\ell+1}\right) \\
&\quad \geqslant \left(1 - \frac{s}{|v|-\ell+1}\right)^\ell \\
&\quad \geqslant \left(1 - \frac{s}{n/3}\right)^\ell \quad \big(\text{since } |v| - \ell + 1 \geqslant \tfrac{n}{2} - 2\sqrt{n \ln n} \geqslant \tfrac{n}{3}\big) \\
&\quad \geqslant 1 - \frac{2s\ell}{n/3} \quad \big(\text{since } s \leqslant \tfrac{n}{6}\big).
\end{aligned}$$

Therefore,

$$\mathbb{P}(M \cap [(D(S) \cap (D(v) \setminus \{v\})) \cap \widetilde{\mathbb{B}}_n] \neq \emptyset) \leqslant \frac{2s\ell}{n/3} = \frac{6s\ell}{n} \leqslant \frac{24s\sqrt{n \ln n}}{n}. \tag{3.5}$$

Combining equations (3.3), (3.4), and (3.5), we get

$$\mathbb{P}(M \cap D^*(v,S) \neq \emptyset) \leqslant \frac{39s\sqrt{n\ln n}}{n}. \qquad \square$$

A similar argument yields the following result.

Lemma 3.2. *Let $n \geqslant 2000$. Let $v \in \widetilde{\mathbb{B}}_n, S \subseteq \widetilde{\mathbb{B}}_n$, where $S \cap D(v) = \emptyset$ and $|S| \leqslant n/6$. Let M be a uniformly chosen random full chain of $U(v)$ (among all $|v|!$ full chains of $U(v)$). We have*

$$\mathbb{P}(M \cap (U^*(v,S)) \neq \emptyset) \leqslant \frac{39|S|\sqrt{n\ln n}}{n}.$$

4. *k*-marked chains and related notions

In this section, as in the rest of the paper, chains are viewed from top to bottom, unless otherwise specified. Let H be a poset whose Hasse diagram is a tree of height k. Let $h = |V(H)|$. Let $n \geqslant \max\{2000, 6h\}$. Let $\mathcal{F} \subseteq \widetilde{\mathbb{B}}_n$. Let $\mathcal{L}$ be a family of k-marked chains with markers in $\mathcal{F}$. For each $v \in \widetilde{\mathbb{B}}_n$ and $d \in [k]$, let

$$\mathcal{L}(v,d) = \{(M,Q) \in \mathcal{L} : \text{ the } d\text{th member of } Q \text{ is } v\}.$$

Let $\mathcal{C}(\mathbb{B}_n)$ denote the set of all $n!$ full chains of $\mathbb{B}_n$. Next, we are going to define the notion of *bad*. This is defined relative to $h = |V(H)|$, which is fixed throughout this section. For each $d \in [k]$, we define a vertex $v \in \widetilde{\mathbb{B}}_n$ to be *d-lower-bad* relative to $\mathcal{L}$ if there exists a set $S \subseteq \widetilde{\mathbb{B}}_n$ with $S \cap U(v) = \emptyset$ and $|S| \leqslant h$ such that

$$\mathcal{L}(v,d) \neq \emptyset \quad \text{and} \quad \forall (M,Q) \in \mathcal{L}(v,d), \quad Q \cap D^*(v,S) \neq \emptyset.$$

We call S a *d-lower-witness* of v relative to $\mathcal{L}$. Similarly, we define a vertex $v \in \widetilde{\mathbb{B}}_n$ to be *d-upper-bad* relative to $\mathcal{L}$ if there exists a set $T \subseteq \widetilde{\mathbb{B}}_n$ with $T \cap D(v) = \emptyset$ and $|T| \leqslant h$ such that

$$\mathcal{L}(v,d) \neq \emptyset \quad \text{and} \quad \forall (M,Q) \in \mathcal{L}(v,d), \quad Q \cap U^*(v,T) \neq \emptyset.$$

We call T a *d-upper-witness* of v relative to $\mathcal{L}$. Let $d \in [k]$. Let $v \in \widetilde{\mathbb{B}}_n$ and $M \in \mathcal{C}(\mathbb{B}_n)$. We say that v is *d-lower-bad* relative to M and $\mathcal{L}$ if v is d-lower-bad relative to $\mathcal{L}$ and there exists at least one Q such that $(M,Q) \in \mathcal{L}(v,d)$. We say that v is *d-upper-bad* relative to M and $\mathcal{L}$ if v is d-upper-bad relative to $\mathcal{L}$ and there exists at least one Q such that $(M,Q) \in \mathcal{L}(v,d)$. A k-marked chain (M,Q) is *good* relative to $\mathcal{L}$ if Q does not contain a vertex v that is either d-lower-bad or d-upper-bad relative to M and $\mathcal{L}$ for any $d \in [k]$. The following proposition follows immediately from the definitions above.

Proposition 4.1. *Let (M,Q) be a member of $\mathcal{L}$ that is good relative to $\mathcal{L}$, and let $v \in Q$. Suppose v is the dth vertex of Q. Then for any set S of at most h vertices of $\mathbb{B}_n$, where $S \cap U(v) = \emptyset$, there exists a member $(M',Q') \in \mathcal{L}(v,d)$ such that M' is disjoint from $D^*(v,S)$. For any set T of at most h vertices, where $T \cap D(v) = \emptyset$, there exists a member $(M'',Q'') \in \mathcal{L}(v,d)$ such that M'' is disjoint from $U^*(v,T)$.*

Proof. Note that $(M, Q) \in \mathcal{L}(v, d)$. By our assumption, v is not d-lower-bad or d-upper-bad relative to $\mathcal{L}$; otherwise v would be either d-lower-bad or d-upper-bad relative to M and $\mathcal{L}$, contradicting (M, Q) being good relative to $\mathcal{L}$. So, there is no d-lower witness of v or d-upper-witness of v of size at most h and the claim follows. $\square$

Now, for each $d \in [k]$ and for each $v \in \widetilde{\mathbb{B}}_n$ that is d-lower-bad relative to $\mathcal{L}$, we fix a corresponding d-lower-witness $S_{v,d}$ of v. For each $d \in [k]$ and each $v \in \widetilde{\mathbb{B}}_n$ that is d-upper-bad relative to $\mathcal{L}$, we fix a corresponding d-upper-witness $T_{v,d}$. A chain $x_1 > y_1 > x_2 > y_2 > \cdots > x_p > y_p$ in $\widetilde{\mathbb{B}}_n$ is called a *d-lower-bad string* (*relative to $\mathcal{L}$*) if, for each $i \in [p]$, x_i is d-lower-bad relative to $\mathcal{L}$ and $y_i \in D^*(x_i, S_{x_i,d})$. Similarly, a chain $x_1 < y_1 < x_2 < y_2 < \cdots < x_p < y_p$ in $\widetilde{\mathbb{B}}_n$ is called a *d-upper-bad string relative to $\mathcal{L}$* if for each $i \in [p]$, x_i is d-upper-bad relative to $\mathcal{L}$ and $y_i \in U^*(x_i, T_{x_i,d})$.

Given a sequence $J = (j_1, j_2, \ldots, j_q)$ of numbers in $[n]$, where either $j_1 < j_2 < \cdots < j_q$ or $j_1 > j_2 > \cdots > j_q$, and a chain C in $\mathbb{B}_n$, let $C[J]$ denote the subchain of C consisting of the j_1th, j_2th, $\ldots$, j_qth members of $\mathcal{F}$ on C (counted from the top). If C contains fewer than j_q members of $\mathcal{F}$, then $C[J]$ is defined to be the empty chain. If J contains only one number j, then we write $C[j]$ for $C[\{j\}]$. In the following two lemmas, we keep our assumptions about n, k, and h described at the beginning of the section.

Lemma 4.2. *Let $d \in [k]$. Let p be a positive integer. Let J be an increasing sequence of $2p$ numbers in $[n]$. Let $v \in \widetilde{\mathbb{B}}_n$. Let M be a uniformly chosen random full chain of $D(v)$. Then*

$$\mathbb{P}(M[J] \textit{ forms a d-lower-bad string}) \leqslant \left(\frac{39h\sqrt{n \ln n}}{n}\right)^p.$$

Proof. Let

$$\gamma = \frac{39h\sqrt{n \ln n}}{n}.$$

We use induction on p. For fixed p, we prove the statement for all J with $2p$ numbers and all $v \in \widetilde{\mathbb{B}}_n$. For the basis step, let $p = 1$. Suppose $J = (j_1, j_2)$, where $j_1 < j_2$. Let $v \in \widetilde{\mathbb{B}}_n$ be given. Let M be a uniformly chosen random full chain of $D(v)$. Recall that if $M[J]$ forms a d-lower-bad string, then its members lie in $\widetilde{\mathbb{B}}_n$. We have

$$\begin{aligned}&\mathbb{P}(M[J] \text{ is a } d\text{-lower-bad string})\\ &\leqslant \sum_{u \in D(v) \cap \widetilde{\mathbb{B}}_n} \mathbb{P}(M[j_1] = u) \cdot \mathbb{P}(M[j_2] \in D^*(u, S_{u,d}) \mid M[j_1] = u).\end{aligned}$$

Fix any $u \in D(v) \cap \widetilde{\mathbb{B}}_n$. The full chains M of $D(v)$ satisfying $M[j_1] = u$ are precisely those concatenations of full chains of $I(v, u)$ (the sublattice consisting of all x satisfying $v \geqslant x \geqslant u$) that contain exactly j_1 members of $\mathcal{F}$ and all full chains of $D(u)$. So, $\mathbb{P}(M[j_2] \in D^*(u, S_{u,d}) \mid M[j_1] = u)$ is the same as the probability that on a uniformly chosen random full chain M' of $D(u)$ the $(j_2 - j_1 + 1)$th member of $\mathcal{F}$ is in $D^*(u, S_{u,d})$. This probability is certainly no more than the probability that M' intersects $D^*(u, S_{u,d})$. Since $u \in \widetilde{\mathbb{B}}_n$, $|D^*(u, S_{u,d})| \leqslant h \leqslant \frac{n}{6}$ and $n \geqslant 2000$, by Lemma 3.1, the latter probability is at most γ. Hence,

$$\begin{aligned}\mathbb{P}(M[J] \text{ is a } d\text{-lower-bad string}) &\leqslant \sum_{u\in D(v)\cap\widetilde{\mathbb{B}}_n} [\mathbb{P}(M[j_1]=u)\cdot\gamma] \\ &= \gamma\cdot\sum_{u\in D(v)\cap\widetilde{\mathbb{B}}_n} \mathbb{P}(M[j_1]=u)\leqslant\gamma,\end{aligned}$$

where the last inequality uses the fact that for different u the events $M[j_1]=u$ are certainly disjoint. This proves the basis step. For the induction step, assume $p\geqslant 2$. Suppose the claim has been proved for all J' and $v\in\widetilde{\mathbb{B}}_n$, where J' is an increasing sequence of $2p-2$ numbers. Given a full chain M of $D(v)$ and a vertex y on M, we let M_y denote the portion of M from y down. Let $J'=(j_3-j_2+1, j_4-j_2+1,\ldots,j_{2p}-j_2+1)$. We have

$$\begin{aligned}&\mathbb{P}(M[J] \text{ is a } d\text{-lower-bad string}) \\ &\quad\leqslant \sum_{u\in D(v)\cap\widetilde{\mathbb{B}}_n}\sum_{y\in D^*(u,S_{u,d})} [\mathbb{P}(M[j_1]=u)\cdot\mathbb{P}(M[j_2]=y\mid M[j_1]=u) \\ &\qquad\cdot\mathbb{P}(M_y[J']) \text{ is a } d\text{-lower-bad string} \mid M[j_1]=u, M[j_2]=y)].\end{aligned}$$

Using reasoning as in the basis step, given $M[j_1]=u, M[j_2]=y$, all full chains of $D(y)$ are equally likely for M_y. So given $M[j_1]=u, M[j_2]=y$, the probability that $M_y[J']$ is a d-lower-bad string is the same as the probability that given a random full chain M' of $D(y)$, $M'[J']$ forms a d-lower-bad string. By the induction hypothesis, this is at most γ^{p-1}. So,

$$\begin{aligned}&\mathbb{P}(M[J] \text{ is a } d\text{-lower-bad string}) \\ &\quad\leqslant \sum_{u\in D(v)\cap\widetilde{\mathbb{B}}_n}\sum_{y\in D^*(u,S_{u,d})} [\mathbb{P}(M[j_1]=u)\cdot\mathbb{P}(M[j_2]=y\mid M[j_1]=u)\cdot\gamma^{p-1}] \\ &\quad= \gamma^{p-1}\cdot\sum_{u\in D(v)\cap\widetilde{\mathbb{B}}_n}\mathbb{P}(M[j_1]=u)\cdot\sum_{y\in D^*(u,S_{u,d})}\mathbb{P}(M[j_2]=y\mid M[j_1]=u) \\ &\quad\leqslant \gamma^{p-1}\cdot\sum_{u\in D(v)\cap\widetilde{\mathbb{B}}_n}\mathbb{P}(M[j_1]=u)\cdot\gamma \quad \text{(see discussion in the basis step)} \\ &\quad= \gamma^{p}\cdot\sum_{u\in D(v)\cap\widetilde{\mathbb{B}}_n}\mathbb{P}(M[j_1]=u)\leqslant\gamma^p.\end{aligned}$$

This completes the induction step and our proof. □

A similar argument yields the following result.

Lemma 4.3. *Let $d\in[k]$. Let p be a positive integer. Let J be a decreasing sequence of $2p$ numbers in $[n]$. Let $v\in\widetilde{\mathbb{B}}_n$. Let M be a uniformly chosen random full chain of $U(v)$. Then*

$$\mathbb{P}(M[J] \textit{ forms a } d\textit{-upper-bad string})\leqslant\left(\frac{39h\sqrt{n\ln n}}{n}\right)^p.$$

5. A nested sequence of dense families of k-marked chains

We show in this section that for families $\mathcal{F} \subseteq \widetilde{\mathbb{B}}_n$ with

$$|\mathcal{F}| \geqslant \left(k - 1 + \Omega\left(\frac{\sqrt{n \ln n}}{n}\right)\right) \binom{n}{\lfloor n/2 \rfloor},$$

we can obtain a sequence of families of k-marked chains with markers in $\mathcal{F}$, $\mathcal{L}_1 \supseteq \mathcal{L}_2 \cdots \supseteq \mathcal{L}_h$, such that for each $i \in [h-1]$ every member of $\mathcal{L}_{i+1}$ is good relative to $\mathcal{L}_i$. Let $\mathcal{C}(\mathbb{B}_n)$ denote the set of full chains of $\mathbb{B}_n$.

Theorem 5.1. *Let $k, h \geqslant 2$ be integers. Let $a_{k,h} = 2^{33k^3h}$ and $c_{k,h} = a_{k,h} \cdot (16k^2h)$. Let n_0 satisfy $n_0 \geqslant \max\{2000, 6h\}$ and*

$$\frac{a_{k,h}\sqrt{n_0 \ln n_0}}{n_0} < \frac{1}{2}.$$

Let $n \geqslant n_0$. Let

$$\epsilon = \frac{c_{k,h}\sqrt{n \ln n}}{n}.$$

Suppose that $\mathcal{F} \subseteq \widetilde{\mathbb{B}}_n$ is a family with $|\mathcal{F}| \geqslant (k-1+\epsilon)\binom{n}{\lfloor n/2 \rfloor}$. For each $M \in \mathcal{C}(\mathbb{B}_n)$, let $Y(M)$ denote the set of members of $\mathcal{F}$ contained in M. There exist functions $X_1, \ldots, X_h$ from $\mathcal{C}(\mathbb{B}_n)$ to $2^{\mathcal{F}}$ satisfying the following.

(1) *For all $M \in \mathcal{C}(\mathbb{B}_n)$, $X_1(M) = Y(M)$.*

(2) *For all $i \in [h-1]$ and $M \in \mathcal{C}(\mathbb{B}_n)$, $X_{i+1}(M) \subseteq X_i(M)$, and if $X_{i+1}(M) \neq \emptyset$ then*

$$\frac{|X_{i+1}(M)|}{|X_i(M)|} \geqslant 1 - \frac{1}{4kh}.$$

(3) *For all $i \in [h]$, the family of k-marked chains $\mathcal{L}_i$ with markers in $\mathcal{F}$, defined by*

$$\mathcal{L}_i = \left\{(M, Q) : M \in \mathcal{C}(\mathbb{B}_n),\ Q \in \binom{X_i(M)}{k}\right\},$$

satisfies

$$|\mathcal{L}_i| \geqslant (\epsilon n!/k)\left(1 - \frac{i}{2h}\right).$$

(4) *For all $i \in [h-1]$, every member of $\mathcal{L}_{i+1}$ is good relative to $\mathcal{L}_i$ (where good and bad are defined with respect to h).*

Proof. If no $\mathcal{F} \subseteq \widetilde{\mathbb{B}}_n$ exists satisfying $|\mathcal{F}| \geqslant (k-1+\epsilon)\binom{n}{\lfloor n/2 \rfloor}$, then the theorem is vacuously true. So we assume that such $\mathcal{F}$ exists. To construct the functions $X_1, \ldots, X_h$, we use induction on i. For the basis step, for each $M \in \mathcal{C}(\mathbb{B}_n)$, we let $X_1(M) = Y(M)$. By Lemma 2.4, we have

$$|\mathcal{L}_1| \geqslant (\epsilon/k)n!. \tag{5.1}$$

So item (3) holds. There is nothing else to prove. For the induction step, let $i \geqslant 1$ and suppose the functions $X_1, \ldots, X_i$ have been defined so that items (1), (2), (3), (4) all hold. We want to define X_{i+1} to satisfy all the requirements.

For each $d \in [k]$ and each $v \in \widetilde{\mathbb{B}}_n$ that is d-lower-bad relative to $\mathcal{L}_i$, we fix a corresponding d-lower-witness $S_{v,d}$. For each $d \in [k]$ and each $v \in \widetilde{\mathbb{B}}_n$ that is d-upper-bad relative to $\mathcal{L}_i$, we fix a corresponding d-upper-witness $T_{v,d}$. To define X_{i+1}, we first classify those M with $X_i(M) \neq \emptyset$ into two types. For each $d \in [k]$, let $B^-_{i,d}(M)$ denote the set of vertices in $X_i(M)$ that are d-lower-bad relative to M and $\mathcal{L}_i$. Let $B^-_i(M) = \bigcup_{d=1}^k B^-_{i,d}(M)$. For each $d \in [k]$, let $B^+_{i,d}(M)$ denote the set of vertices in $X_i(M)$ that are d-upper-bad relative to M and $\mathcal{L}_i$. Let $B^+_i(M) = \bigcup_{d=1}^k B^+_{i,d}(M)$. Let $B_i(M) = B^-_i(M) \cup B^+_i(M)$. Let $x(M) = |X_i(M)|$ and let $b(M) = |B_i(M)|$. Set $C = 4kh$. Let

$$\mathcal{C}_1 = \left\{ M \in \mathcal{C}(\mathbb{B}_n) : X_i(M) \neq \emptyset,\ \frac{b(M)}{x(M)} \leqslant \frac{1}{C} \right\},$$
$$\mathcal{C}_2 = \left\{ M \in \mathcal{C}(\mathbb{B}_n) : X_i(M) \neq \emptyset,\ \frac{b(M)}{x(M)} > \frac{1}{C} \right\}.$$

Now, we define X_{i+1} as follows:

$$\begin{aligned} &\text{If } X_i(M) = \emptyset \text{ or } M \in \mathcal{C}_2, \text{ then let } X_{i+1}(M) = \emptyset. \\ &\text{Otherwise, } M \in \mathcal{C}_1, \text{ and we let } X_{i+1}(M) = X_i(M) \setminus B_i(M). \end{aligned}$$

Clearly, for all $M \in \mathcal{C}(\mathbb{B}_n)$, $X_{i+1}(M) \subseteq X_i(M)$.

Claim 1. We have the following.

(1) For all $M \in \mathcal{C}(\mathbb{B}_n)$, where $X_{i+1}(M) \neq \emptyset$, we have

$$|X_{i+1}(M)| \geqslant \left(1 - \frac{1}{C}\right)|X_i(M)| = \left(1 - \frac{1}{4kh}\right)|X_i(M)|.$$

(2) Each member of $\mathcal{L}_{i+1}$ is good relative to $\mathcal{L}_i$.

(3)

$$\sum_{M \in \mathcal{C}_1} \binom{|X_{i+1}(M)|}{k} \geqslant \left(1 - \frac{k}{C}\right) \sum_{M \in \mathcal{C}_1} \binom{|X_i(M)|}{k} \geqslant \left(1 - \frac{1}{4h}\right) \binom{|X_i(M)|}{k}.$$

Proof of Claim 1. Let $M \in \mathcal{C}(\mathbb{B}_n)$ and suppose $X_{i+1}(M) \neq \emptyset$. Then $M \in \mathcal{C}_1$. By our definition of $\mathcal{C}_1$, we have $|B_i(M)|/|X_i(M)| \leqslant 1/C$. Since $X_{i+1}(M) = X_i(M) \setminus B_i(M)$, item (1) follows immediately. The only members of $\mathcal{L}_{i+1}$ have the form (M, Q), where $M \in \mathcal{C}_1$ and $Q \in \binom{X_{i+1}(M)}{k}$. Fix any such member (M, Q). Since $X_{i+1}(M) = X_i(M) \setminus B_i(M)$, and $Q \in \binom{X_{i+1}(M)}{k}$, Q contains no vertex that is either d-lower-bad or d-upper-bad relative to M and $\mathcal{L}_i$ for any $d \in [k]$. Hence (M, Q) is good relative to $\mathcal{L}_i$. So item (2) holds. As in the definition, let $b(M) = |B_i(M)|$ and $x(M) = |X_i(M)|$. The number of k-subsets of $X_i(M)$ that contain a member of $B_i(M)$ is certainly at most

$$b\binom{x-1}{k-1} = \frac{bk}{x}\binom{x}{k} \leqslant \frac{k}{C}\binom{x}{k}.$$

Therefore, we have

$$\binom{|X_{i+1}(M)|}{k} \geqslant \binom{x}{k} - \frac{k}{C}\binom{x}{k} = \left(1 - \frac{k}{C}\right)\binom{|X_i(M)|}{k} = \left(1 - \frac{1}{4h}\right)\binom{|X_i(M)|}{k}.$$

So item (3) (of Claim 1) holds. This completes the proof of Claim 1. □

Claim 2. We have

$$\sum_{M \in \mathcal{C}_2} \binom{|X_i(M)|}{k} \leqslant \frac{\epsilon}{4kh} \cdot n!.$$

Proof of Claim 2. We further partition $\mathcal{C}_2$ into two subclasses. Let $\mathcal{C}_2^-$ consist of those $M \in \mathcal{C}_2$ with $|B_i^-(M)| \geqslant |B_i(M)|/2 = b(M)/2$ and let $\mathcal{C}_2^+ = \mathcal{C}_2 - \mathcal{C}_2^-$. For each $d \in [k]$, let $\mathcal{C}_{2,d}^-$ consist of those $M \in \mathcal{C}_2^-$ with $|B_{i,d}^-(M)| \geqslant |B_i^-(M)|/k$. Clearly, $\mathcal{C}_2^- = \bigcup_{d=1}^k \mathcal{C}_{2,d}^-$. For each $d \in [k]$, we first bound $\sum_{M \in \mathcal{C}_{2,d}^-(M)} \binom{|X_i(M)|}{k}$.

For each $M \in \mathcal{C}_{2,d}^-$, we define a sequence $R_d^-(M)$, called the *greedy d-lower-bad string generated by* M relative to $\mathcal{L}_i$, as follows. Scan M from top to bottom. Let x_1 be the first vertex in $B_{i,d}^-(M)$ that we encounter. Recall that this means x_1 is d-lower-bad relative to M and $\mathcal{L}$ and we have fixed a d-lower-witness $S_{x_1,d}$ of v (relative to $\mathcal{L}_i$) with $|S_{x_1,d}| \leqslant h$ and there is at least one member (M, Q) of $\mathcal{L}_i(x_1, d)$. Since the members of $\mathcal{L}_i$ on M form $\binom{X_i(M)}{k}$ and $\mathcal{L}_i(x_1, d) \neq \emptyset$, in particular the k consecutive members of $X_i(M)$, with x_1 being the dth one among them, form a Q with $(M, Q) \in \mathcal{L}_i(x_1, d)$. Since x_1 is d-lower-bad relative to $\mathcal{L}_i$, Q must intersect $D^*(x_1, S_{x_1,d})$, which takes place below x_1. Let y_1 be the first member of $X_i(M)$ below x_1 that lies in $D^*(x_1, S_{x_1,d})$. By our discussion above, y_1 is among the $k - d$ members of $X_i(M)$ below x_1. After we encounter y_1, we continue down M. If there are more vertices in $X_i(M)$ that are d-lower-bad relative to M and $\mathcal{L}_i$, then let x_2 denote the next vertex in $X_i(M)$ that is d-lower-bad relative to M and $\mathcal{L}_i$. We then similarly define y_2. We continue like this until we run out of vertices in $X_i(M)$. Following our reasoning for the existence of y_1, whenever an x_i is defined, y_i must exist and is within the $k - d$ members of $X_i(M)$ below x_i. Suppose $R_d^-(M) = (x_1, y_1, x_2, y_2, \ldots, x_p, y_p)$. By our procedure, $p \geqslant \lceil |B_{i,d}^-(M)|/k \rceil$. Let J be the increasing sequence of $2p$ numbers in $[n]$ such that $M[J] = R_d^-(M)$. We denote J by $P_d^-(M)$ and call it the *d-lower-bad profile* of M relative to $\mathcal{L}_i$. Now we organize the terms in $\sum_{M \in \mathcal{C}_{2,d}^-} \binom{X_i(M)}{k}$ by $|P_d^-(M)|$. For convenience, we will view the increasing sequence $P_d^-(M)$ simply as a subset of $[n]$. Let p be any positive integer. Consider $M \in \mathcal{C}_{2,d}^-$ with $|P_d^-(M)| = 2p$. By item (2) of the induction hypothesis,

$$\frac{|X_i(M)|}{|X_1(M)|} \geqslant \left(1 - \frac{1}{4kh}\right)^{i-1} \geqslant \left(1 - \frac{1}{4kh}\right)^{h} \geqslant 1 - \frac{2h}{4kh} \geqslant \frac{1}{2}.$$

So

$$|Y(M)| = |X_1(M)| \leqslant 2|X_i(M)| \leqslant 2|B_i(M)|C \leqslant 4|B_i^-(M)|C \leqslant 4k|B_{i,d}^-(M)|C \leqslant 4k^2pC$$

(recall that $p \geqslant (|B_{i,d}^-(M)|)/k$). Clearly the largest number in $P_d^-(M)$ is no more than $|Y(M)| \leqslant 4k^2pC$. So, $P_d^-(M) \in \binom{[4k^2pC]}{2p}$. Fix any $2p$-subset (increasing sequence) J of $[4k^2pC]$. By our definition of $P_d^-(M)$, if $P_d^-(M) = J$, then certainly $M[J] = R_d^-(M)$ forms a d-lower-bad string relative to $\mathcal{L}_i$ by the definition of $R_d^-(M)$. Thus

$$\begin{aligned}
&|\{M \in \mathcal{C}_{2,d}^- : P_d^-(M) = J\}| \\
&\quad \leqslant |\{M \in \mathcal{C}(\mathbb{B}_n) : M[J] \text{ forms a } d\text{-lower-bad string relative to } \mathcal{L}_i\}| \\
&\quad \leqslant \left(\frac{39h\sqrt{n \ln n}}{n}\right)^p \cdot n! \qquad \text{(by Lemma 4.2).}
\end{aligned}$$

So

$$\begin{aligned}
|\{M \in \mathcal{C}_{2,d}^- : |P_d^-(M)| = 2p\}| &\leqslant \binom{4k^2pC}{2p} \cdot \left(\frac{39h\sqrt{n \ln n}}{n}\right)^p \cdot n! \\
&\leqslant 2^{4k^2pC} \left(\frac{39h\sqrt{n \ln n}}{n}\right)^p \cdot n!.
\end{aligned}$$

Also, for each $M \in \mathcal{C}_{2,d}^-$ with $|P_d^-(M)| = 2p$, we showed earlier that $|Y(M)| \leqslant 4k^2pC$. Hence

$$\binom{|X_i(M)|}{k} \leqslant \binom{|Y(M)|}{k} \leqslant 2^{|Y(M)|} \leqslant 2^{4k^2pC}.$$

So, the contribution to $\sum_{M \in \mathcal{C}_{2,d}^-} \binom{|X_i(M)|}{k}$ from those $M \in \mathcal{C}_{2,d}^-$ with $|P_d^-(M)| = 2p$ is at most

$$2^{4k^2pC} \cdot 2^{4k^2pC} \cdot \left(\frac{39h\sqrt{n \ln n}}{n}\right)^p \cdot n! \leqslant \left(\frac{2^{32k^3h} \cdot 39h\sqrt{n \ln n}}{n}\right)^p \cdot n! < \left(\frac{2^{33k^3h}\sqrt{n \ln n}}{n}\right)^p \cdot n!.$$

Let

$$\beta = \frac{2^{33k^3h}\sqrt{n \ln n}}{n}.$$

By our assumption about n, $\beta < \frac{1}{2}$. Summing over all $p \geqslant 1$, we get

$$\sum_{M \in \mathcal{C}_{2,d}^-} \binom{|X_i(M)|}{k} \leqslant \sum_{p=1}^{\infty} \beta^p \cdot n! \leqslant 2\beta n!.$$

Summing over all $d \in [k]$, we get

$$\sum_{M \in \mathcal{C}_2^-} \binom{|X_i(M)|}{k} \leqslant 2k\beta n!.$$

By a similar argument, we have

$$\sum_{M \in \mathcal{C}_2^+} \binom{|X_i(M)|}{k} \leqslant 2k\beta n!.$$

Recall that

$$\epsilon = \frac{2^{33k^3h}\sqrt{n \ln n}}{n} \cdot (16k^2h) = \beta(16k^2h).$$

We have

$$\sum_{M \in \mathcal{C}_2} \binom{|X_i(M)|}{k} = \sum_{M \in \mathcal{C}_2^-} \binom{|X_i(M)|}{k} + \sum_{M \in \mathcal{C}_2^+} \binom{|X_i(M)|}{k} \leqslant 4k\beta n! = \frac{\epsilon}{4kh} \cdot n!.$$

This completes the proof of Claim 2. □

Claim 3. We have

$$|\mathcal{L}_{i+1}| \geqslant (\epsilon/k)n!\left(1 - \frac{i+1}{2h}\right).$$

Proof of Claim 3. By induction hypothesis,

$$|\mathcal{L}_i| \geqslant (\epsilon/k)n!\left(1 - \frac{i}{2h}\right).$$

By Claim 2,

$$\sum_{M \in \mathcal{C}_2} \binom{|X_i(M)|}{k} \leqslant \frac{\epsilon}{4kh} \cdot n!.$$

So

$$\sum_{M \in \mathcal{C}_1} \binom{|X_i(M)|}{k} \geqslant (\epsilon/k)n!\left(1 - \frac{i}{2h} - \frac{1}{4h}\right).$$

By Claim 1 and our definition of $\mathcal{L}_{i+1}$, we have

$$\begin{aligned}
|\mathcal{L}_{i+1}| = \sum_{M \in \mathcal{C}_1} \binom{|X_{i+1}(M)|}{k} &\geqslant \left(1 - \frac{1}{4h}\right) \sum_{M \in \mathcal{C}_1} \binom{|X_i(M)|}{k} \\
&\geqslant (\epsilon/k)n!\left(1 - \frac{i}{2h} - \frac{1}{4h}\right)\left(1 - \frac{1}{4h}\right) \\
&\geqslant (\epsilon/k)n!\left(1 - \frac{i+1}{2h}\right)
\end{aligned}$$

This completes the proof of Claim 3. □

So item (3) of the theorem holds. This completes the induction step and the proof. □

6. Proof of Theorem 1.4

Now, we are ready to prove Theorem 1.4. We keep all the notation from previous sections. Let k, h, H be given. Let $c_{k,h}$ and n_0 be defined as in Theorem 5.1. Let $\mathcal{F} \subseteq \widetilde{\mathbb{B}}_n$ be a family satisfying

$$|\mathcal{F}| \geqslant \left(k - 1 + \frac{c_{k,h}\sqrt{n \ln n}}{n}\right)\binom{n}{\lfloor n/2 \rfloor}.$$

It is easy to check that all the conditions of Theorem 5.1 are satisfied. Let $\mathcal{L}_1 \supseteq \mathcal{L}_2 \supseteq \cdots \supseteq \mathcal{L}_h$ be the sequence of families of k-marked chains we obtained in Theorem 5.1. We define a sequence of subposets $H_1, H_2, \ldots$ of H as follows. Let $H_1 = H$. Recall that H_1 is k-saturated. Suppose H_1 is not a chain. Then, by Lemma 2.2, H_1 contains a chain interval $I_1 = [v_1, u_1]$ or $[u_1, v_1]$, where v_1 is a leaf in $D(H_1)$ and $H_2 = H_1 \setminus (I - u_1)$ is still k-saturated and $D(H_2)$ is a tree. If H_2 is a chain, then we terminate. Otherwise, H_2 contains a chain interval $I_2 = [v_2, u_2]$ or $[u_2, v_2]$ such that $H_3 = H_2 \setminus (I_2 - u_2)$ is k-saturated. We continue like this until the current subposet, say H_q, is just a k-chain. Clearly $q \leqslant h$. We prove the following proposition, which implies Theorem 1.4. Given a set W of vertices in $\mathbb{B}_n$, we view W as a family of subsets of $[n]$ and define the *sublattice* of $\mathbb{B}_n$ induced by W, denoted by $\mathbb{B}_n[W]$, to be $(W, \subseteq)$. Clearly, $\mathbb{B}_n[W]$ is an induced subposet of $\mathbb{B}_n$.

Proposition 6.1. *There exist subsets $W_1 \supseteq W_2 \supseteq \cdots \supseteq W_q$ of $\mathbb{B}_n$ such that the following hold.*
(1) *For all $i \in [q]$, $\mathbb{B}_n[W_i] = H_i$. (Hence, we will treat W_i as $V(H_i)$).*
(2) *For all $i \in [q]$ and $v \in W_i = V(H_i)$, if v is at level d of H_i (from the top) then $\mathcal{L}_i(v, d) \neq \emptyset$.*

Proof. We use reverse induction on i. For the basis step, let $i = q$. We know that H_q is just a k-chain. By Theorem 5.1, $|\mathcal{L}_q| \geqslant (\epsilon/k)n!(1 - \frac{q}{2h}) > 0$. So there exists $(M, Q) \in \mathcal{L}_q$. We embed H_q using Q. Let $W_q = V(Q)$. Clearly, items (1) and (2) both hold. For the induction step, let $i \leqslant q - 1$. Suppose we have defined $W_{i+1}, \ldots, W_q$ that satisfy all the requirements. Recall that $H_{i+1} = H_i \setminus (I_i - u_i)$, where $I_i = [v_i, u_i]$ or $[u_i, v_i]$ is a chain interval in H_i. Without loss of generality, we may assume $I_i = [v_i, u_i]$, which would put v_i at level k since v_i is a leaf in $D(H_i)$ and each leaf is at level 1 or k. (The case where $I_i = [u_i, v_i]$ can be handled similarly.) Suppose u_i is at level d from the top in H_{i+1}. By item (2) of the induction hypothesis, $\mathcal{L}_{i+1}(u_i, d) \neq \emptyset$. Let $(M, Q) \in \mathcal{L}_{i+1}(u_i, d)$. Then u_i is the dth vertex of Q (from the top). By Theorem 5.1, (M, Q) is good relative to $\mathcal{L}_i$. Let $S = W_{i+1} \setminus U(u_i)$. In other words, S is the set of vertices in H_{i+1} that are not ancestors of u_i. Since $|S| \leqslant h$, by Proposition 4.1, there exists a member $(M', Q') \in \mathcal{L}_i(u_i, d)$ such that M' is disjoint from $D^*(u_i, S)$. We can embed $I_i - u_i$ using the portion Q^* of Q' below u_i. The newly embedded vertices, by design, are not in $D^*(u_i, S)$ and hence are not related to any vertex in S. (They are, however, descendants of u_i and hence are still descendants of the ancestors of u_i in $W_{i+1} = V(H_{i+1})$.) Let $W_i = W_{i+1} \cup V(Q^*)$. Since $\mathbb{B}_n[W_{i+1}] = H_{i+1}$, it follows from our discussion above that $\mathbb{B}_n[W_i] = H_i$. Furthermore, because of the existence of (M', Q'), it is easy to see that the newly embedded vertices (namely those in Q^*) still satisfy item (2) of the theorem. This completes the induction step and the proof. □

7. Concluding remarks

7.1. Comments on the approach

Even though our approach follows that of Bukh, we needed to use several key new ideas. In Bukh's argument, it is crucial to assume that on each full chain the number of members of $\mathcal{F}$ is bounded. Indeed, if some full chain contains h members of $\mathcal{F}$ then $\mathcal{F}$ contains an h-chain, which already contains H as a subposet. However, for the induced version, this is no longer the case. One can have an unbounded number of members of $\mathcal{F}$ on a full chain without forcing an induced H. To overcome this difficulty, we consider two types of full chains. In one type of full chain the number of bad members of $\mathcal{F}$ is negligible compared to the number of members of $\mathcal{F}$. In a second type of full chain, the number of bad members of $\mathcal{F}$ is comparable to the number of members of $\mathcal{F}$. For the second type, the key observation is that the number of k-marked chains on type 2 full chains decreases exponentially fast as the number of bad members of $\mathcal{F}$ that lie on the full chain. This still allows us to limit the total number of bad k-marked chains and build our nested sequence of dense families of k-marked chains, which is then used to embed H

iteratively. Another major departure from Bukh's approach is that we no longer insist on using entire k-marked chains to embed maximal chains of H. Rather, we use k-marked chains to locate good vertices to embed H, while preserving the levels of vertices.

7.2. Induced versus non-induced

We showed that when H is a poset whose Hasse diagram is a tree, $\mathrm{La}(n,H)$ and $\mathrm{La}^*(n,H)$ are asymptotically equal, both asymptotic to $(k-1)\binom{n}{\lfloor n/2\rfloor}$, where k is the height of $D(H)$. For other posets though, $\mathrm{La}^*(n,H)$ can be very different from $\mathrm{La}(n,H)$. For instance, since $\mathrm{La}(n,K_{r,s}) \leqslant (2+o(1))\binom{n}{\lfloor n/2\rfloor}$, for any two-level poset H, we have $\mathrm{La}(n,H) \leqslant (2+o(1))\binom{n}{\lfloor n/2\rfloor}$. However, we now show that for every fixed m, there exists a two-level poset H_m satisfying $\mathrm{La}^*(n,H_m) \geqslant (m-1-o(1))\binom{n}{\lfloor n/2\rfloor}$. Specifically, let H_m be the two-level poset consisting of $x_1,x_2,\ldots,x_m$ at the lower level and $y_1,y_2,\ldots,y_m$ at the upper level. For each $i\in[m]$, let $x_i \leqslant y_j$ for $j=i,i+1,\ldots,m$. Suppose $\mathcal{G}\subseteq\mathbb{B}_n$ is a family that contains H_m as an induced subposet with members $A_1,\ldots,A_m$ playing the roles of $x_1,\ldots,x_m$, respectively and members $B_1,\ldots,B_m$ playing the role of $y_1,\ldots,y_m$, respectively. For each $i\in[m]$, let $S_i=\bigcap_{j=i}^m B_i$. Note that $S_m \supseteq S_{m-1} \supseteq \cdots \supseteq S_1$. Also, by our assumption, for all $i\in[m]$, $A_i\subseteq S_i$ and if $i\geqslant 2$ then also $A_i \nsubseteq S_{i-1}$. In particular, this implies that $S_1,\ldots,S_m$ must be distinct sets. So $|S_m|-|S_1|\geqslant m-1$. It follows that $|B_m|-|A_1|\geqslant m-1$. Now, let $\mathcal{F}\subseteq\mathbb{B}_n$ be a family that consists of the middle $m-1$ levels of $\mathbb{B}_n$. Since the cardinalities of any two members of $\mathcal{F}$ differ by at most $m-2$, $\mathcal{F}$ does not contain H_m as an induced subposet. Since $|\mathcal{F}|=(m-1-o(1))\binom{n}{\lfloor n/2\rfloor}$, we have

$$\mathrm{La}^*(n,H_m) \geqslant (m-1-o(1))\binom{n}{\lfloor n/2\rfloor}.$$

Acknowledgements

The authors thank the anonymous referee for his/her helpful suggestions and Robert Seiver for his helpful discussions.

References

[1] Bukh, B. (2009) Set families with a forbidden subposet. *Electron. J. Combin.* **16** #R142.

[2] Carroll, T. and Katona, G. O. H. (2008) Bounds on maximal families of sets not containing three sets with $A\cup B\subset C, A\not\subset B$. *Order* **25** 229–236.

[3] De Bonis, A. and Katona, G. O. H. (2007) Largest families without an r-fork. *Order* **24** 181–191.

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

[5] Erdős, P. (1945) On a lemma of Littlewood and Offord. *Bull. Amer. Math. Soc.* **51** 898–902.

[6] Griggs, J. and Lu, L. (2009) On families of subsets with a forbidden subposet. *Combin. Probab. Comput.* **18** 731–748.

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

[8] Thanh, H. T. (1998) An extremal problem with excluded subposets in the Boolean lattice. *Order* **15** 51–57.