

ANTI-RAMSEY FORBIDDEN POSET PROBLEMS

BALÁZS PATKÓS

ABSTRACT. A family \mathcal{G} of sets is a weak copy of a poset P if there is a bijection $f : P \rightarrow \mathcal{G}$ such that $p \leq q$ implies $f(p) \subseteq f(q)$. If f satisfies $p \leq q$ if and only if $f(p) \subseteq f(q)$, the \mathcal{G} is a strong copy of P . We study the anti-Ramsey numbers $\text{ar}(n, P)$, $\text{ar}^*(n, P)$, the maximum number of colors used in a coloring of $2^{[n]}$ that does not admit a rainbow weak or strong copy of P , respectively. We establish connections to the well-studied extremal numbers $\text{La}(n, P)$ and $\text{La}^*(n, P)$ and determine asymptotically $\text{ar}^*(n, T)$ for all tree posets T and $\text{ar}^*(n, O_{2k})$ for all crown posets O_{2k} .

1. INTRODUCTION

In extremal combinatorics, Turán type problems ask for determining the maximum possible size of a combinatorial object that does not contain a fixed forbidden object. An object is *rainbow* with respect to a coloring if all its elements receive distinct colors. Rainbow Turán problems (introduced for graphs by Keevash, Mubayi, Sudakov, and Verstraete [15]) ask for determining the maximum possible size of an object that does not admit rainbow forbidden objects with respect to any ‘proper’ coloring. Another variant of rainbow problems, anti-Ramsey problems (introduced for graphs by Erdős, Simonovits, and Sós [7]), ask to maximize the number of colors used when coloring a big object under the condition that no rainbow forbidden small object appears.

We will be interested in the latter type of problems in the area of forbidden subposets. A family \mathcal{G} of sets is a *weak copy* of a poset (P, \leq) , if there exists a bijection $f : P \rightarrow \mathcal{G}$ such that $p \leq q$ implies $f(p) \subseteq f(q)$, and \mathcal{G} is a *strong copy* of (P, \leq) , if there exists a bijection $f : P \rightarrow \mathcal{G}$ such that $p \leq q$ if and only if $f(p) \subseteq f(q)$. With some abuse of notation, we will write P instead of (P, \leq) and say that a family \mathcal{F} is *weak (strong) P -free* if it does not contain weak (strong) copies of P . The *forbidden subposet problem* of determining $\text{La}(n, P)$ ($\text{La}^*(n, P)$), the maximum size of a weak (strong) P -free family $\mathcal{F} \subseteq 2^{[n]} := \{S : S \subseteq \{1, 2, \dots, n\}\}$, was introduced by Katona and Tarján [14] in the early 80s. For results and open problems, see the survey [9] and Chapter 7 of [8].

Rainbow Turán forbidden problems were introduced and studied in [19] and another rainbow variant, Gallai-Ramsey numbers of poset pairs in [13]. In this paper, we start investigating the anti-Ramsey version of the forbidden subposet problem. Let $\text{ar}(n, P)$ ($\text{ar}^*(n, P)$) denote the maximum number of colors used in a coloring of $2^{[n]}$ that does not admit a rainbow weak (strong) copy of P . By definition, we immediately obtain the inequalities

$$(1) \quad \text{ar}(n, P) \leq \text{La}(n, P) \quad \text{and} \quad \text{ar}^*(n, P) \leq \text{La}^*(n, P),$$

as if a coloring of $2^{[n]}$ uses more than $\text{La}(n, P)$ ($\text{La}^*(n, P)$) colors, then the family we get by picking an arbitrary set from each color class contains a weak (strong) copy of P which is rainbow. The first inequality of (1) and an old result of Erdős [6] imply $\text{ar}(n, P) = O\left(\binom{n}{\lfloor n/2 \rfloor}\right)$ and the second inequality of (1) together with a result of Methuku and Pálvölgyi [17] imply $\text{ar}^*(n, P) = O\left(\binom{n}{\lfloor n/2 \rfloor}\right)$

Let us make further quick observations that improve the inequalities of (1) and that give lower bounds on $\text{ar}(n, P)$ and $\text{ar}^*(n, P)$ and still connect the anti-Ramsey number to the original forbidden subposet problem. We need two concepts. A family \mathcal{F} is said to be *convex* if $F \subseteq G \subseteq F'$ and $F, F' \in \mathcal{F}$ imply

$G \in \mathcal{F}$. Let $\text{La}_{\text{con}}(n, P)$ ($\text{La}_{\text{con}}^*(n, P)$) denote the maximum size of a weak (strong) P -free convex family $\mathcal{F} \subseteq 2^{[n]}$.

For any poset P , we write $\mathcal{P}^-(P)$ to denote the set $\{P \setminus \{m\} : m \text{ is maximal or minimal in } P\}$. Also, for a set \mathcal{P} of posets, we say that the family \mathcal{F} is weak (strong) \mathcal{P} -free if it is weak (strong) P -free for all $P \in \mathcal{P}$. $\text{La}(n, \mathcal{P})$ ($\text{La}^*(n, \mathcal{P})$) denotes the maximum size of a weak (strong) \mathcal{P} -free family $\mathcal{F} \subseteq 2^{[n]}$.

Proposition 1.1. *For any poset P , we have*

- (1) $1 + \text{La}_{\text{con}}(n, \mathcal{P}^-(P)) \leq \text{ar}(n, P)$,
- (2) $1 + \text{La}_{\text{con}}^*(n, \mathcal{P}^-(P)) \leq \text{ar}^*(n, P)$.

Proof. Let $\mathcal{F} \subseteq 2^{[n]}$ be a convex weak (strong) $\mathcal{P}^-(P)$ -free family of size $\text{La}_{\text{con}}(n, \mathcal{P}^-(P))$ ($\text{La}_{\text{con}}^*(n, \mathcal{P}^-(P))$). Color each $F \in \mathcal{F}$ with distinct colors and use the same new color for all sets $2^{[n]} \setminus \mathcal{F}$. By the convexity of \mathcal{F} , sets of the new color can only be maximal or minimal elements in a weak (strong) rainbow copy of P , so the remaining sets of the copy should form a weak (strong) copy of a poset in $\mathcal{P}^-(P)$ within \mathcal{F} . \square

As a corollary, we immediately obtain the following. Let C_k denote the chain (totally ordered set) on k elements. Then if P contains a weak copy of $2C_2$ (two unrelated copies of C_2) or a copy of C_3 , then $\text{ar}^*(n, P), \text{ar}(n, P) = \Theta(\binom{n}{\lfloor n/2 \rfloor})$. Indeed, then any $P' \in \mathcal{P}^-(P)$ contains a pair of elements in relation, so both $\text{La}_{\text{con}}^*(n, \mathcal{P}^-(P))$ and $\text{La}_{\text{con}}(n, \mathcal{P}^-(P))$ are at least $\binom{n}{\lfloor n/2 \rfloor}$ as shown by the convex family of the middle layer of $2^{[n]}$.

Proposition 1.2.

- (1) *For any poset P , we have $\text{ar}(n, P) \leq 2 + \text{La}(n, \mathcal{P}^-(P))$.*
- (2) *If a poset P contains a smallest or largest element m , then $\text{ar}^*(n, P) \leq 1 + \text{La}^*(n, P \setminus \{m\})$.*

Proof. For any coloring c of $2^{[n]}$ without a rainbow weak copy of P , pick one set from every color class of c except for the color classes of \emptyset and $[n]$. The family obtained must be $\mathcal{P}^-(P)$ -free, as such a copy is rainbow by definition, and can be extended by \emptyset or by $[n]$ to obtain a rainbow weak copy of P . This proves (1).

If m is the largest element of P , then for any coloring c of $2^{[n]}$ without a rainbow string copy of P , pick one set from every color class of c except for the color class of $[n]$. The family obtained must be $(P \setminus \{m\})$ -free, as such a copy can be extended by $[n]$ to obtain a rainbow strong copy of P . This proves (2). \square

As a corollary, we obtain that if P does not contain a weak copy of $2C_2$ nor a copy of C_3 , then $\text{ar}(n, P) = |P| - 1$. Indeed, then $\mathcal{P}^-(P)$ contains the antichain A_k on $k = |P| - 1$ elements, and any k sets form a weak copy of A_k , so $\text{La}(n, A_k) = k - 1$.

All weak $\{2C_2, C_3\}$ -free posets P are of the form $P' + A_k$, where $+$ denotes the union of an incomparable pair of copies of P' and A_k with P' being either a fork \vee_s ($a < b_1, b_2, \dots, b_s$) or a broom \wedge_s ($c_1, \dots, c_s < d$).

Proposition 1.3.

- (1) $\text{ar}^*(n, \vee_s) = \text{ar}^*(n, \wedge_s) = (s - 1)(n - 1) + 2$.
- (2) $\text{ar}^*(n, A_k) = 3 + (k - 2)(n - 1)$ if $n \geq 2k$.
- (3) $\text{ar}^*(n, \wedge_s + A_k) = \text{ar}^*(n, \vee_s + A_k) = O_{s,k}(n^2)$.

For weak copies of P , the lower bound of Proposition 1.1 and the upper bound of Proposition 1.2 differ only by 1 and the possible difference of $\text{La}_{\text{con}}(n, P)$ and $\text{La}(n, P)$. In most cases, these are negligible. For example, if $P = \diamond$ is the diamond poset on four elements a, b, c, d with $a < b, c < d$ being all relations, then $\mathcal{P}^-(\diamond) = \{\vee, \wedge\}$. A result of Katona and Tarján [14] states that $\text{La}(n, \{\vee, \wedge\}) = 2 \binom{n-1}{\lfloor \frac{n-1}{2} \rfloor}$ and an extremal family is $\binom{[n-1]}{\lfloor \frac{n-1}{2} \rfloor} \cup \{F \cup \{n\} : F \in \binom{[n-1]}{\lfloor \frac{n-1}{2} \rfloor}\}$, which is convex. (Here and throughout the paper,

$\binom{S}{k}$ denotes the family of all k -subsets of S , and $\binom{[n]}{k}$ is often referred to as the k th layer of $2^{[n]}$.) So, $\text{La}(n, \mathcal{P}^-(\diamond)) = \text{La}_{\text{con}}(n, \mathcal{P}^-(\diamond))$. Finally, if in a coloring c of $2^{[n]}$ the empty set and $[n]$ receives distinct colors, then to avoid a rainbow weak copy of \diamond , one can only use at most one further colors. Therefore, $\text{ar}(n, \diamond) = 1 + \text{La}(n, \mathcal{P}^-(\diamond)) = 1 + \text{La}_{\text{con}}(n, \mathcal{P}^-(\diamond)) = 1 + 2^{\binom{n-1}{\lfloor \frac{n-1}{2} \rfloor}}$.

How far apart can $\text{La}(n, \mathcal{P}^-(P))$ and $\text{La}_{\text{con}}(n, \mathcal{P}^-(P))$ be? It was conjectured [3, 11] that for any set \mathcal{P} of posets, $\text{La}(n, \mathcal{P})$ ($\text{La}^*(n, \mathcal{P})$) is asymptotically the size of most number of middle layers of $2^{[n]}$ that one can have without containing of weak (strong) copy of a member of \mathcal{P} . Note that the union of consecutive layers is always a convex family. The conjecture was disproved by Ellis, Ivan, and Leader [5], who showed that $\text{La}^*(n, B_d), \text{La}(n, B_d) \geq (d + \varepsilon - o(1)) \binom{n}{\lfloor \frac{n}{2} \rfloor}$ for $d \geq 4$, where B_d is the Boolean poset on $2^{[d]}$ ordered by inclusion. Although the construction of [5] for B_d can be easily turned into a construction for $B_d \setminus \{\emptyset\}$ or for $B_d \setminus \{[d]\}$, we do not see how it could be turned into a construction that avoids both of them, i.e. a construction that is $\mathcal{P}^-(B_d)$ -free.

Problem 1.4. Is there a poset P and a positive real ε such that $\text{La}(n, \mathcal{P}^-(P)) \geq (1 + \varepsilon) \text{La}_{\text{con}}(n, \mathcal{P}^-(P))$?

1.1. Main results. We will focus on forbidding strong copies of P and on cases not covered by Proposition 1.2 (2), i.e. on posets having neither largest nor smallest elements.

The oriented Hasse diagram $\vec{H}(P)$ of a poset P is an oriented graph with vertex set P where \vec{pq} is an arc if q covers p , i.e. $p \leq_P q$ and there is no $z \neq p, q$ with $p \leq_P z \leq_P q$. We obtain the (unoriented) Hasse diagram $H(P)$ of P from $\vec{H}(P)$ by dropping the orientation of all arcs. A poset is a *tree poset* if $H(P)$ is a tree.

Bukh proved [3] $\text{La}(n, T) = (h(T) - 1 + o(1)) \binom{n}{\lfloor n/2 \rfloor}$ for any tree poset T , where $h(T)$ denotes the *height* of T , the length of the longest chain in T . This was strengthened by Boehnlein and Jiang [2] to $\text{La}^*(n, T) = (h(T) - 1 + o(1)) \binom{n}{\lfloor n/2 \rfloor}$. As the construction showing the lower bound in these results, consists of the middle $h(T) - 1$ layers, which is a convex family, we obtain $\text{La}_{\text{con}}(n, T) = (1 + o(1)) \text{La}(n, T) = (1 + o(1)) \text{La}^*_{\text{con}}(n, T) = (1 + o(1)) \text{La}^*(n, T)$ for all tree posets. Our first theorem determines $\text{ar}^*(n, T)$ asymptotically for all tree posets.

Note that if for all maximal or minimal elements of T we have $h(T \setminus \{t\}) = h(T)$, then the result follows from $\text{ar}^*(n, T) \leq \text{La}^*(n, T)$ of (1) and $1 + \text{La}^*(n, \mathcal{P}^-(T)) \leq \text{ar}^*(n, T)$ of Proposition 1.1 (2). Also, if T has a largest or a smallest element, then the theorem follows from Proposition 1.1 (2) and Proposition 1.2 (2). So the hard part of the result is about tree posets T containing a minimal or maximal element m that is contained in all longest chains of T and yet m is neither smallest nor largest element of T .

Theorem 1.5. *For any tree poset T , we have $\text{ar}^*(n, T) = (1 + o(1)) \text{La}^*(n, \mathcal{P}^-(T))$.*

The *crown poset* O_{2k} is a poset with oriented Hasse diagram an antirected cycle, i.e. O_{2k} has elements $a_1, a_2, \dots, a_k, b_1, b_2, \dots, b_k$ with $a_i, a_{i+1} < b_i$ for all $i = 1, 2, \dots, k - 1$ and $a_k, a_1 < b_k$. O_4 is mostly called the butterfly poset and is denoted by \bowtie based on the drawing of its Hasse diagram. $\text{La}(n, \bowtie) = \binom{n}{\lfloor \frac{n}{2} \rfloor} + \binom{n}{\lfloor \frac{n}{2} \rfloor + 1}$ was proved first by DeBonis, Katona, and Swanepoel [4]. As \bowtie is a strong subposet of the tree poset X on elements a_1, a_2, b_1, b_2, c with $a_1, a_2 < c < b_1, b_2$, Boehnlein and Jiang's result implies $\text{La}^*(n, \bowtie) = (2 + o(1)) \binom{n}{\lfloor \frac{n}{2} \rfloor}$. For even $k \geq 4$, $\text{La}(n, O_{2k}) = (1 + o(1)) \binom{n}{\lfloor \frac{n}{2} \rfloor}$ was proved by Griggs and Lu [11], and the same result for odd $k \geq 7$ was obtained by Lu [16]. The value of $\text{La}^*(n, O_{2k})$ for $k \geq 3$ has not yet been addressed.

As $\mathcal{P}^-(O_{2k})$ consists of two tree posets both of height 2, Proposition 1.1 (1) and Proposition 1.2 (1) imply $\text{ar}(n, O_{2k}) = (1 + o(1)) \binom{n}{\lfloor n/2 \rfloor}$ for all $k \geq 2$. By (1), we have $\text{ar}^*(n, \bowtie) \leq \text{La}^*(n, \bowtie) = (2 + o(1)) \binom{n}{\lfloor \frac{n}{2} \rfloor}$. It is easy to see that $\text{ar}^*(n, \bowtie) \geq \binom{n}{\lfloor \frac{n}{2} \rfloor} + \binom{n}{\lfloor \frac{n}{2} \rfloor + 1} + 1$. Indeed, color sets of size $\lfloor \frac{n}{2} \rfloor$ and $\lfloor \frac{n}{2} \rfloor + 1$ by their own color, and color all other sets white. As the intersection of two unrelated sets of different colors has size at most $\lfloor \frac{n}{2} \rfloor$ and the union of two unrelated sets of different colors has size at least $\lfloor \frac{n}{2} \rfloor + 1$, this coloring does not admit a strong rainbow copy of \bowtie . This shows $\text{ar}^*(n, \bowtie) = (2 + o(1)) \binom{n}{\lfloor \frac{n}{2} \rfloor}$. As $\mathcal{P}^-(\bowtie) = \{\vee, \wedge\}$,

this also shows an example when the lower bound of Proposition 1.1 (2) and the trivial upper bound (1) differ asymptotically and the true asymptotics of $\text{ar}^*(n, P)$ is given by (1). For all values $k \geq 3$ we show the following result.

Theorem 1.6. *For any $k \geq 3$, we have $\text{ar}^*(n, O_{2k}) = (1 + o(1)) \binom{n}{\lfloor \frac{n}{2} \rfloor}$.*

2. PROOFS

In the next subsection, we state some embedding results for tree posets and how they imply Theorem 1.5 and Theorem 1.6. Then in Section 2.2, we prove the embedding results. Finally, Section 2.3 contains the proof of Proposition 1.3.

2.1. Statements of embedding results and their implications. To be able to state our results, we need to introduce some notation.

Let $\tilde{B}_n = \{F \in 2^{[n]} : ||F| - n/2| \leq 2\sqrt{n \ln n}\}$ and $\tilde{B}_n^T = \{F \in 2^{[n]} : ||F| - n/2| \leq 4|T|\sqrt{n \ln n}\}$. Note that by Chernoff's inequality, $|2^{[n]} \setminus \tilde{B}_n^T| \leq |2^{[n]} \setminus \tilde{B}_n| \leq \frac{1}{n^2} \binom{n}{\lfloor \frac{n}{2} \rfloor}$ holds for large enough n .

For any poset P and $p \in P$, we set $\mathcal{D}_P(p) = \{q \in P : q \leq_P p\}$. Similarly, for any family $\mathcal{F} \subseteq 2^{[n]}$ of sets and $G \in 2^{[n]}$, we write $\mathcal{D}_{\mathcal{F}}(G) = \{F \in \mathcal{F} : F \subseteq G\}$ and $\mathcal{D}(G) = \{F : F \subseteq G\}$, $\mathcal{U}_n(G) = \{F \in 2^{[n]} : F \supseteq G\}$ and we omit n from the subscript when it is clear from context.

Finally, if \mathcal{G} is a strong copy of a poset P shown by the bijection $f : P \rightarrow \mathcal{G}$, then the set $f(p) \in \mathcal{G}$ will be denoted by G_p .

Theorem 2.1. *Let T be a tree poset of height $k+1$ that contains a maximal element m such that all chains of length $k+1$ in T contain m . Then for any positive real ε , there exists $n_0 = n_0(\varepsilon)$ such that if $n \geq n_0$ and $\mathcal{F} \subseteq \tilde{B}_n$ with $|\mathcal{F}| = (k-1 + \varepsilon) \binom{n}{\lfloor n/2 \rfloor}$, then \mathcal{F} contains a strong copy \mathcal{G} of $T \setminus \{m\}$ such that for any $p \in T \setminus \mathcal{D}_T(m)$ the image G_p is not a subset of $\cup_{d \in \mathcal{D}_T(m) \setminus \{m\}} G_d$.*

Let P_{2k-1} denote the poset that we obtain from O_{2k} by removing a maximal element, i.e. P_k has $2k-1$ elements $a_1, a_2, \dots, a_k, b_1, b_2, \dots, b_{k-1}$ with $a_i, a_{i+1} < b_i$.

Theorem 2.2. *For any $k \geq 3$ and positive real ε , there exists $n_0 = n_0(\varepsilon)$ such that if $n \geq n_0$ and $\mathcal{F} \subseteq \tilde{B}_n$ with $|\mathcal{F}| = (1 + \varepsilon) \binom{n}{\lfloor n/2 \rfloor}$, then \mathcal{F} contains a strong copy \mathcal{G} of P_{2k-1} such that $G_{a_i} \not\subseteq G_{a_1} \cup G_{a_k}$ for all $i = 2, \dots, k-1$.*

Proof of Theorem 1.5 and Theorem 1.6. The lower bounds of both theorems follow from Proposition 1.1 (2).

For the upper bounds let c be any coloring of $2^{[n]}$. Let \mathcal{F}_c be a family that we obtain by picking one set from each color class that does not contain any set from $2^{[n]} \setminus \tilde{B}_n$. Clearly, the number of colors used by c is at most $|2^{[n]} \setminus \tilde{B}_n| + |\mathcal{F}_c| \leq \frac{1}{n^2} \binom{n}{\lfloor \frac{n}{2} \rfloor} + |\mathcal{F}_c|$.

For Theorem 1.6: if $|\mathcal{F}_c| \leq (1 + o(1)) \binom{n}{\lfloor \frac{n}{2} \rfloor}$, then we are done. Otherwise, by Theorem 2.2, we obtain a strong copy $A_1, A_2, \dots, A_k, B_1, \dots, B_{k-1}$ of P_{2k-1} such that $A_1 \cup A_k$ does not contain any of the A_i s ($i = 2, 3, \dots, k-1$). Let B_k be any superset of $A_1 \cup A_k$ of size at least $\frac{n}{2} + 2\sqrt{n \ln n}$ with $A_i \not\subseteq B_k$ for all $i = 2, 3, \dots, k-1$. Such a set exists as there is $a_i \in A_i \setminus (A_1 \cup A_k)$ for all i , so $[n] \setminus \{a_2, \dots, a_{k-1}\}$ can play the role of B_k . Then $A_1, A_2, \dots, A_k, B_1, \dots, B_{k-1}, B_k$ form a strong copy of O_{2k} . Also, it is rainbow as the A_i s, B_i s represent different color classes of c , and these color classes do not contain sets from $2^{[n]} \setminus \tilde{B}_n$, so their colors are distinct from that of B_k .

For Theorem 1.5: as observed in the introduction, it is enough to consider tree posets T of height $k+1$ with a maximal or minimal element m that is contained in all chains of length $k+1$ in T . Without loss of generality, we may assume that m is maximal. If $|\mathcal{F}_c| \leq (k-1 + o(1)) \binom{n}{\lfloor \frac{n}{2} \rfloor}$, then we are done. Otherwise, by Theorem 2.1, \mathcal{F} contains a strong copy \mathcal{G} of $T \setminus \{m\}$ such that $G := \cup_{d \in \mathcal{D}_T(m) \setminus \{m\}} G_d$ does not contain any G_p with $p \notin \mathcal{D}_T(m)$. Then a superset G' of G of size at least $\frac{n}{2} + 2\sqrt{n \ln n}$ still not containing G_p for

any $p \notin \mathcal{D}_T(m)$ together with \mathcal{G} form a strong copy of T . The existence of G' and the rainbow property of $\mathcal{G} \cup \{G'\}$ follows as in the previous paragraph. \square

2.2. Proofs of embedding results. There have been several papers on embedding tree posets into set families starting with the seminal paper of Bukh [3] determining $\text{La}(n, T)$ for any tree poset T . This was extended by Boehnlein and Jiang [2] to strong copies of T . The proof of 2.1 will closely follow the reasoning of [2]. Supersaturation and counting results on weak and strong copies of tree posets (for all or for some special classes) were obtained in [20] and then in [1] and [12]. The proof of 2.2 uses some ideas from [20].

We will first prove Theorem 2.2. We start by introducing some notation and preliminaries. The *Lubell-mass* of a family $\mathcal{F} \subseteq 2^{[n]}$ is defined as

$$\lambda_n(\mathcal{F}) = \sum_{F \in \mathcal{F}} \frac{1}{\binom{n}{|F|}} = \frac{1}{n!} \sum_{\mathcal{C} \in \mathbf{C}_n} |\mathcal{C} \cap \mathcal{F}|,$$

the average number of sets in \mathcal{F} that a maximal chain contains chosen uniformly at random from \mathbf{C}_n , the set of all maximal chains in $2^{[n]}$. We will consider the max-partition of \mathbf{C}_n according to \mathcal{F} , that is we write $\mathbf{C}_F = \{\mathcal{C} \in \mathbf{C}_n : F \text{ is largest in } \mathcal{F} \cap \mathcal{C}\}$. Then $\frac{1}{|\mathbf{C}_F|} \sum_{\mathcal{C} \in \mathbf{C}_F} |\mathcal{C} \cap \mathcal{F}| = \lambda_{|F|}(\mathcal{D}_{\mathcal{F}}(F))$ and $\lambda_n(\mathcal{F})$ is a weighted average of $\lambda_{|F|}(\mathcal{D}_{\mathcal{F}}(F))$ and so we have the following.

Observation 2.3. If $\lambda_{|F|}(\mathcal{D}_{\mathcal{F}}(F)) \leq B$ for all F , then $\lambda_n(\mathcal{F}) \leq B$ holds.

Lemma 2.4 (Griggs, Li, Lu, Lemma 3.2 in [10]). *For any family $\mathcal{F} \subseteq 2^{[n]}$, we have $|\mathcal{F}| \leq \lambda_n(\mathcal{F}) \binom{n}{\lfloor n/2 \rfloor}$.*

We will need the following folklore observation on minimum and average degree of graphs.

Observation 2.5. If G is a graph with average degree d , then G contains a subgraph G' with minimum degree at least $d/2$.

In the proof of Theorem 2.2, we will be looking for a copy of a special subposet with the help of which we will find the copy of P_{2k-1} with the extra properties.

Definition 2.6. *Let $S^{k,\ell}$ denote the height 2 poset on $\ell \cdot k + 1$ elements such that $H(S^{k,\ell})$ is a spider with ℓ legs each of length k . As the height is 2, there exist two such posets depending on whether the leaves are maximal or minimal elements. We define $S^{k,\ell}$ to be the one where the leaves are maximal elements.*

We move some of the calculations of the proof of Theorem 2.2 to here.

Lemma 2.7. *For k fixed, $100k \leq j \leq 4\sqrt{n \ln n}$ and n large enough, we have*

(1)

$$\frac{\binom{n+2\sqrt{n \ln n}}{j}}{\binom{n-2\sqrt{n \ln n}}{j}} \leq n^{20}$$

(2)

$$\sum_{i=j/k}^j \binom{kn^{1/3}j}{i} \binom{n/2 + 2\sqrt{n \ln n}}{j-i} = o\left(\frac{1}{\sqrt{n \ln n}} \binom{n/2 - 2n\sqrt{n \ln n}}{j-22}\right)$$

(3)

$$\binom{n^{2/3} + kn^{1/3}j}{j} = o\left(\frac{1}{\sqrt{n \ln n}} \binom{n/2 - 2n\sqrt{n \ln n}}{j-22}\right)$$

Proof. Using $1 + x \leq e^x$, we obtain

$$\frac{\binom{n+2\sqrt{n \ln n}}{j}}{\binom{n-2\sqrt{n \ln n}}{j}} \leq \left(\frac{n - 2\sqrt{n \ln n}}{n - 6\sqrt{n \ln n}} \right)^{4\sqrt{n \ln n}} \leq e^{\frac{5\sqrt{n \ln n}}{n} \cdot 4\sqrt{n \ln n}} \leq n^{20}$$

which proves (1).

Using (1), $\left(\frac{n}{k}\right)^k \leq \binom{n}{k}$, $\frac{\binom{n}{k}}{\binom{n}{k'}} \leq \left(\frac{k'}{k}\right)^{k'-k}$ for $k < k'$ and $i \geq j/k \geq 1000$, $j \leq \sqrt{n \ln n}$, we obtain

$$\frac{\binom{kn^{1/3}j}{i} \binom{n/2+2\sqrt{n \ln n}}{j-i}}{\binom{n/2-2\sqrt{n \ln n}}{j-22}} \leq \frac{n^{20} \binom{kn^{1/3}j}{i} \binom{n/2+2\sqrt{n \ln n}}{j-i}}{\binom{n/2+2\sqrt{n \ln n}}{j-22}} \leq \frac{n^{20} (k^2 n^{1/3})^i j^{i-22}}{n^{i-22}} \leq k^2 n^{42} \left(\frac{j}{n^{2/3}}\right)^i = o\left(\frac{1}{n^2}\right).$$

As there are at most $\sqrt{n \ln n}$ terms in its left hand side, we obtain (2).

Using $\left(\frac{n}{k}\right)^k \leq \binom{n}{k} \leq \left(\frac{en}{k}\right)^k$, we have $\binom{n^{2/3}+kn^{1/3}j}{j} \leq (n^{2/3})^j$ and $\binom{n/2-2\sqrt{n \ln n}}{j-22} \geq \left(\frac{n}{10\sqrt{n \ln n}}\right)^{j-22}$. As $j \geq 1000k$, we have $(n^{2/3})^j < n^{0.99j-22}$ and (3) follows. \square

After all these preparations, we are ready to prove Theorem 2.2.

Proof of Theorem 2.2. Let $\mathcal{F} \subseteq \tilde{B}_n$ be a set family with $|\mathcal{F}| = (1+\varepsilon)\binom{n}{\lfloor n/2 \rfloor}$. For any $F \in \mathcal{F}$ and j positive integer, let us define

$$S_j(F) = \{G \in \mathcal{F} : G \subseteq F, |G| = |F| - j\}.$$

Then we partition \mathcal{F} into $\mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3$ with

$$\mathcal{F}_1 = \left\{ F \in \mathcal{F} : \exists j : 1 \leq j \leq 1000k, |S_j(F)| \geq \frac{\varepsilon}{10000k} \binom{|F|}{j} \right\},$$

$$\mathcal{F}_2 = \left\{ F \in \mathcal{F} \setminus \mathcal{F}_1 : \exists j \geq 1000k + 1, |S_j(F)| \geq \binom{|F|}{j-22} \right\}, \quad \text{and} \quad \mathcal{F}_3 = \mathcal{F} \setminus (\mathcal{F}_1 \cup \mathcal{F}_2).$$

We obtain the following bound on $\lambda(\mathcal{F}_3)$ and thus, by Lemma 2.4, on $|\mathcal{F}_3|$.

Claim 2.8. *If n is large enough, then $\lambda_n(\mathcal{F}_3) < 1 + \varepsilon/2$.*

Proof of Claim. By definition, for any $F \in \mathcal{F}_3$

$$\lambda_{|F|}(\mathcal{D}_{\mathcal{F}_3}(F)) \leq \sum_{j=0}^{4 \ln n \sqrt{n}} \frac{|S_j(F)|}{\binom{|F|}{j}} \leq 1 + 1000k \cdot \frac{\varepsilon}{10000k} + 4 \ln n \sqrt{n} \cdot \left(\frac{4 \ln n \sqrt{n}}{n}\right)^{22} \leq 1 + \varepsilon/2,$$

if n is large enough. The claim follows from Observation 2.3. \square

By Lemma 2.4, $|\mathcal{F}_3| \leq (1+\varepsilon/2)\binom{n}{\lfloor n/2 \rfloor}$, so either \mathcal{F}_1 or \mathcal{F}_2 has size at least $\frac{\varepsilon}{4}\binom{n}{\lfloor n/2 \rfloor}$. On many occasions, we will see that a subfamily of \mathcal{F}_1 or \mathcal{F}_2 will have size $f(\varepsilon)\binom{n}{\lfloor n/2 \rfloor}$ or $g(\varepsilon)\binom{n}{j}$. If it is not important, then we will not state what the precise function f is, just write $\varepsilon', \varepsilon''$, etc.

CASE I $|\mathcal{F}_1| \geq \frac{\varepsilon}{4}\binom{n}{\lfloor n/2 \rfloor}$

By definition of \mathcal{F}_1 , for any $F \in \mathcal{F}_1$, there exists $1 \leq j(F) \leq 1000k$ such that $|S_{j(F)}(F)| \geq \frac{\varepsilon}{10000k} \binom{|F|}{j}$. So for some $1 \leq j \leq 1000k$, the family $\mathcal{F}_1^j = \{F \in \mathcal{F}_1 : j(F) = j\}$ has size at least $\frac{\varepsilon}{4000k} \binom{n}{\lfloor n/2 \rfloor}$. Consider the bipartite graph B with parts \mathcal{F} and \mathcal{F}_1^j and $F \in \mathcal{F}, F_1 \in \mathcal{F}_1^j$ joined by an edge if and only if $F \subseteq F_1$, $|F| = |F_1| - j$, and we say that this edge has color set $F_1 \setminus F$. By definition of \mathcal{F}_1 , the number of edges in B is at least $\varepsilon' \binom{n/2-2\ln n \sqrt{n}}{j} \binom{n}{\lfloor n/2 \rfloor}$ and so the average degree of B is at least $\varepsilon'' \binom{n/2-2\ln n \sqrt{n}}{j}$. By Observation 2.5, B contains a subgraph B' with minimum degree at least $\varepsilon''' \binom{n/2-2\ln n \sqrt{n}}{j}$.

We plan to obtain a copy of $S^{k-2,\ell}$ in \mathcal{F} where $\ell = \ell(k, \varepsilon)$ is a large constant independent of n . If the element of $S^{k-2,\ell}$ corresponding to the center of the spider is a minimal element, then we start with an arbitrary set in the \mathcal{F} part of B' , if it is a maximal element, then we start with an arbitrary element of the \mathcal{F}_1 part of B' . Then we proceed as follows: if some connected part of $S^{k-2,\ell}$ is already defined and we want to add a neighbor of F or F_1 , then we pick a neighbor such that the color set of the connecting edge should be disjoint with the color set of every previously defined edge of the spider. This is possible, as the union of all previous color sets has size at most $jk\ell$, and so the number of neighbors that do not satisfy this requirement is at most $jk\ell \binom{n+2\sqrt{n \ln n}}{j-1}$ which is smaller than the minimum degree $\varepsilon''' \binom{n/2-2 \ln n \sqrt{n}}{j}$ if n is large enough (here we use that in Case I, j is a fixed constant at most $1000k$). By definition, this family of sets forms a weak copy of $S^{k-2,\ell}$. We claim that due to the disjointness of the color sets of the edges, they form a strong copy of $S^{k-2,\ell}$. Suppose first that $F \in \mathcal{F}, F_1 \in \mathcal{F}_1$ are on the same leg of the spider, but are not consecutive elements. If F is closer to the center of the spider, then the leg continues as FF'_1F' and the color set of F'_1F' is removed and never put back, so F_1 cannot contain F . Similarly, if F_1 is closer to the center and the leg continues as $F_1F'_1F'$, then the color set of F'_1F' is added and never later removed, so F_1 cannot contain F . If F, F_1 are on different legs of the spider, then if the center F_0 is in the \mathcal{F} part, then the color set of $F_0F'_1$ is contained in F where F'_1 is the first element of the spider after the center, and so F_1 cannot contain F . Finally, if F_0 is in the \mathcal{F}_1 part, then the color set of F_0F' is removed and is never put back, and so is not contained in F_1 which therefore cannot contain F . Here F' is the first element of the leg of F_1 .

Let G denote the intersection of all sets of the spider defined above. Observe that $|G| \geq n/2 - 3\sqrt{n \ln n}$ if n is large enough as every set in the spider can remove at most $1000k$ elements from the intersection. Let F^1, F^2, \dots, F^ℓ be the sets corresponding to the leaves of the spider and let x_i be an element of F^i from the color set of the edge incident to F^i . If we find $1 \leq a < b \leq \ell$ and $y \in G$ such that F^a and F^b have neighbors $G^a \subseteq F^a \setminus \{y\}$ and $G^b \subseteq F^b \setminus \{y\}$ with $x_a \in F^a, x_b \in F^b$, then these two legs of the spider together with G^a and G^b form the desired copy of P_{2k-1} .

The number of subsets of F^a of size $|F^a| - j$ not containing x_a is at most $\binom{|F^a|-1}{j-1}$, so if n is large enough, then the number of subsets of F^a of size $|F^a| - j$ containing x_a is at least $\varepsilon'''' \binom{|F^a|}{j}$. Therefore, for any F^a and x_a , there exist $\alpha(\varepsilon)n$ elements $y \in G$ such that there is an $(|F^a| - j)$ -subset $G_y^i \in \mathcal{F}$ with $x_i \in G_y^i, y \notin G_y^i$. Taking $\ell = \ell(\varepsilon)$ to be larger than $\lceil \alpha(\varepsilon)^{-1} \rceil$, we obtain a pair x_a, x_b with a common y . This finishes the proof of this case.

CASE II $|\mathcal{F}_2| \geq \frac{\varepsilon}{4} \binom{n}{\lfloor n/2 \rfloor}$

By definition of \mathcal{F}_2 , for any $F \in \mathcal{F}_2$, there exists $1000k < j(F) \leq 4 \ln n \sqrt{n}$ such that $|S_{j(F)}(F)| \geq \binom{|F|}{j-22}$. So for some $1000k < j \leq 4\sqrt{n \ln n}$, the family $\mathcal{F}_2^j = \{F \in \mathcal{F}_2 : j(F) = j\}$ has size at least $\frac{\varepsilon}{4 \ln n \sqrt{n}} \binom{n}{\lfloor n/2 \rfloor}$. Consider the bipartite graph B with parts \mathcal{F} and \mathcal{F}_2^j and $F \in \mathcal{F}, F_2 \in \mathcal{F}_2^j$ joined by an edge if and only if $F \subseteq F_2, |F| = |F_2| - j$ and let $F_2 \setminus F$ be the color set of this edge. By definition of \mathcal{F}_2 , the number of edges in B is at least $\frac{\varepsilon}{4 \ln n \sqrt{n}} \binom{n/2-2\sqrt{n \ln n}}{j-22} \binom{n}{\lfloor n/2 \rfloor}$ and the average degree is at least $\frac{\varepsilon}{8\sqrt{n \ln n}} \binom{n/2-2\sqrt{n \ln n}}{j-22}$ in B . By Observation 2.5, B contains a subgraph B' with minimum degree at least $\frac{\varepsilon}{16\sqrt{n \ln n}} \binom{n/2-2n\sqrt{n \ln n}}{j-22}$.

Our strategy is as in Case I, we want to find a spider in B' that corresponds to an $S^{k-2,\ell}$ in \mathcal{F} but now $\ell = \ell(n) = n^{1/3}$. This time, we require a little less from the color sets of the edges used: we want that for any edge FF_2 out of the j colors, less than j/k should appear in color sets of previously defined edges. The total number of colors used is at most $kn^{1/3}j$. So the number of sets having not enough new elements is at most

$$\sum_{i=j/k}^j \binom{kn^{1/3}j}{i} \binom{n/2 + 2\sqrt{n \ln n}}{j-i}$$

which is, by Lemma 2.7 (2), negligible compared to the minimum degree of B' . So it possible to pick the edges of the spider as required. As in the previous case, the spider is a weak copy of $S^{k-2,\ell}$ by definition of B . We claim that it is a strong copy of $S^{k-2,\ell}$. Suppose F is in the \mathcal{F} -part of B and F_2 is in the \mathcal{F}_2 part of B and they both belong to the spider, but FF_2 is not an edge of the spider. If F_2 is closer to the center on the same leg as F or if they are on distinct legs, then there is a step after F_2 is defined when $j - j/k$ new colors are added to the leg of F , and later at most j/k old colors are removed in each step and there are at most $k/2 - 1$ steps when we remove colors, so F will contain at least $j/2$ new colors compared to F_2 . Similarly, if F is closer to the center of the spider with F_2 on the same leg, then after adding F , there is a step when $j - j/k$ colors are removed, and then colors are added in at most $k/2 - 1$ steps, at most j/k old colors in each such step, so F again will have at least $j/2$ colors not in F_2 .

Let again F^1, F^2, \dots, F^ℓ be the sets corresponding to the leaves of the $S^{k-2,\ell}$. By the argument above, for any $1 \leq a, b \leq \ell$, if C^a denotes the color set of the edge incident to F^a , then $|C^a \setminus F^b| \geq j/2$. The number of subsets of F^a of size $|F^a| - j$ that contain more than j/k elements of C^a is

$$\sum_{i=j/k}^j \binom{|C^a|}{i} \binom{|F^a \setminus C^a|}{j-i} \leq 2 \binom{j}{j/k} \binom{n/2 + 2\sqrt{n \ln n}}{j - j/k},$$

which by the calculation of Lemma 2.7 (2), has a lower order of magnitude compared to the minimum degree in B . So every F^a has at least $\frac{\varepsilon'}{\sqrt{n \ln n}} \binom{n/2 - 2\sqrt{n \ln n}}{j-22}$ incident edges in B with at most j/k colors used in the spider. As in Case I, let G denote the intersection of all sets in the spider. Clearly, $|G| \geq n/2 - n^{1/3} \sqrt{n \ln n}$, and so by Lemma 2.7 (3), for every a there exists a set $Y^a \subset G$ of $n^{2/3}$ elements that appear as colors on some edge incident to F^a . As $\ell = n^{1/3}$ there exist a, b and y with $y \in Y^a \cap Y^b$. Thus there exist G^a, G^b with $y \notin G^a \subseteq F^a, y \notin G^b \subseteq F^b, |C^a \cap G^a|, |C^b \cap G^b| \leq j/k$. Then G^a, G^b and the legs of F^a and F^b form the needed copy of P_{2k-1} . \square

We now turn our attention to Theorem 2.1. As the proof very closely follows the proof in [2], let us briefly sketch their approach to embed a tree poset T of height k into any family $\mathcal{F} \subseteq \tilde{B}_n$ of size at least $(k - 1 + \varepsilon) \binom{n}{\lfloor n/2 \rfloor}$. First, following [3], they define a sequence $T \subseteq T_1 \supset T_2 \supset \dots \supset T_\ell$ of tree posets such that T_ℓ is a chain and $T_i \setminus T_{i+1}$ is a chain interval (see the definition below). Then a sequence of $\mathcal{G}_1 \supset \mathcal{G}_2 \supset \dots \supset \mathcal{G}_\ell$ of families of *good* k -chains of \mathcal{F} is defined, where good approximately means extendable for any small obstacles. Finally, the chain T_ℓ is embedded into \mathcal{F} using a k -chain from \mathcal{G}_ℓ and any $T_i \setminus T_{i+1}$ is embedded using a k -chain from \mathcal{G}_i as \mathcal{G}_i is good thus extendable with respect to \mathcal{G}_{i+1} . For weak copies of T , the good property is not necessary, but when one looks for strong copies of T one has to make sure that sets F_t from newly used chains are not in relation with sets $F_{t'}$ of already used chains if t, t' are incomparable in T . As the number of already embedded elements (the potential obstacles) is at most $|T| - 1$, the goodness condition ensures that one can extend the embedding for a strong copy of T . In the proof of Theorem 2.1, we will apply the same procedure, but with one extra obstacle and this new obstacle might belong to $\tilde{B}_n^T \setminus \tilde{B}_n$. So the change is not very substantial for this part of the proof, but we need to be a bit more careful, when defining the sequence $T_1 \supset \dots \supset T_\ell$.

We start describing the preliminaries by addressing T . A poset P is *k-saturated* if $h(P) = k$ and all maximal chains have length k .

Lemma 2.9 (Bukh [3]). *Any poset P of height h is a strong subposet of an h -saturated poset P' . Moreover, any tree poset T of height h is a strong subposet of an h -saturated tree poset T' .*

Sketch of the proof. Consider the canonical decomposition $\cup_{i=1}^h A_i = P$ that we obtain by setting A_1 to be the antichain of all minimal elements of P , and for any $1 < j \leq h$ we let A_j to be the set of minimal elements of $P \setminus \cup_{i=1}^{j-1} A_i$.

- If $q \in A_j$ is a maximal element of P , then add $h - j$ elements q_1, q_2, \dots, q_{h-j} forming a chain $q \leq q_1 \leq q_2 \leq \dots \leq q_{h-j}$.
- If \overrightarrow{pq} is an arc of the oriented Hasse diagram of P with $p \in A_i, q \in A_j$, then add $j - i - 1$ elements $r_1, r_2, \dots, r_{j-i-1}$ forming a chain $p \leq r_1 \leq \dots \leq r_{j-i-1} \leq q$.

□

If $p \leq_P q$, then the interval $[p, q]$ is $\{z : p \leq_P z \leq_P q\}$. A *chain interval* is an interval that induces a chain in P . The *poset distance* of p, q in P is the minimum k such that there exist vertices $p = v_0, v_1, v_2, \dots, v_k = q$ such that v_i and v_{i+1} are comparable in P for all $i = 0, 1, \dots, k-1$. The next lemma is from [3], the moreover part is not explicit there, but comes from its proof.

Lemma 2.10 (Bukh [3]). *If T is a k -saturated tree poset that is not a chain, then there exist a leaf v and a chain interval $I = [v, u]$ or $I = [u, v]$ such that $T \setminus I'$ is a k -saturated tree poset, where $I' = I \setminus \{u\}$. Moreover, v is a leaf such that there exists $w \in T$ with vw maximizing poset distance.*

We need one more lemma on the possible size of unions and intersections of sets that are used when embedding T . These are the possible extra obstacles mentioned earlier.

Lemma 2.11. *Suppose $G_1, G_2, \dots, G_h \in \tilde{B}_n$ with $h \leq |T|$ are such that their comparability graph is connected. Then $\cup_{i=1}^h G_i \in \tilde{B}_n^T$.*

Proof. We can assume that the indexing of the G_i s is such that for any $1 \leq j \leq h$ the comparability graph of G_1, G_2, \dots, G_j is connected. Therefore there exists $i < j$ such that G_i, G_j are comparable. If $G_j \subseteq G_i$, then G_j does not add new element to the union of the G_i s, while if $G_i \subseteq G_j$, then G_j adds at most $|G_j \setminus G_i| \leq 4\sqrt{n \ln n}$ elements to the union of the G_i . Thus $|\cup_{i=1}^h G_i| \leq n/2 + 2\sqrt{n \ln n} + 4(h-1)\sqrt{n \ln n} \leq n/2 + 4|T|\sqrt{n \ln n}$ as claimed. □

The following definitions are all from [2]. For a family $\mathcal{G} \subseteq 2^{[n]}$ of sets, we define $\mathcal{D}(\mathcal{G}) = \cup_{G \in \mathcal{G}} \mathcal{D}(G)$ and $\mathcal{U}(\mathcal{G}) = \cup_{G \in \mathcal{G}} \mathcal{U}(G)$. For a set $G \in \tilde{B}_n$ and a family $\mathcal{S} \subseteq \tilde{B}_n^T$ with $\mathcal{U}(G) \cap \mathcal{S} = \emptyset$ (in [2] the family of obstacles satisfied $\mathcal{S} \subseteq \tilde{B}_n$, but for the proof of Theorem 2.1, we will have to allow $\mathcal{S} \subseteq \tilde{B}_n^T$), we define the *forbidden neighborhood below and above G* as

$$\mathcal{D}^*(G, \mathcal{S}) = [(\mathcal{D}(G) \setminus \{G\}) \cap (\mathcal{U}(\mathcal{S}) \cup \mathcal{D}(\mathcal{S}))] \cap \tilde{B}_n$$

and

$$\mathcal{U}^*(G, \mathcal{S}) = [(\mathcal{U}(G) \setminus \{G\}) \cap (\mathcal{U}(\mathcal{S}) \cup \mathcal{D}(\mathcal{S}))] \cap \tilde{B}_n.$$

Let \mathbf{C}_n denote the set of all $n!$ maximal chains in $[n]$. A pair $(\mathcal{C}, \mathcal{Q})$ is a *k -marked chain* with markers in \mathcal{F} if $\mathcal{C} \in \mathbf{C}_n$, $\mathcal{Q} \subseteq \mathcal{C} \cap \mathcal{F}$ and $|\mathcal{Q}| = k$. For a set \mathcal{L} of k -marked chains, a set G , and an integer $1 \leq d \leq k$, we write $\mathcal{L}(G, d) = \{(\mathcal{C}, \mathcal{Q}) \in \mathcal{L} : G \text{ is the } d\text{th member of } \mathcal{Q} \text{ from top}\}$.

The next definitions are a little altered compared to their original version in [2]. There, the witness family \mathcal{S} was from \tilde{B}_n , while here it is from \tilde{B}_n^T . This is because our additional obstacle is a union of some of the sets of a copy of T in \tilde{B}_n^T , and then Lemma 2.11 guarantees that this new obstacle lies within \tilde{B}_n^T .

A set $G \in \tilde{B}_n$ is *(d, T) -lower bad relative to \mathcal{L}* if there exists a family $\mathcal{S} \subseteq \tilde{B}_n^T$ with $\mathcal{U}(G) \cap \mathcal{S} = \emptyset$ and $|\mathcal{S}| \leq |T|$ such that

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

Similarly, $G \in \tilde{B}_n$ is *(d, T) -upper bad relative to \mathcal{L}* if there exists a family $\mathcal{S} \subseteq \tilde{B}_n^T$ with $\mathcal{D}(G) \cap \mathcal{S} = \emptyset$ and $|\mathcal{S}| \leq |T|$ such that

$$\mathcal{L}(G, d) \neq \emptyset \quad \text{and} \quad \forall (\mathcal{C}, \mathcal{Q}) \in \mathcal{L}(G, d), \quad \mathcal{Q} \cap \mathcal{U}^*(G, \mathcal{S}) \neq \emptyset.$$

The family \mathcal{S} in the above definitions is called a *(d, T) -lower (upper) witness of G* .

We say that G is (d, T) -lower-bad relative to \mathcal{C} and \mathcal{L} if G is (d, T) -lower-bad relative to \mathcal{L} and there exists at least one \mathcal{Q} such that $(\mathcal{C}, \mathcal{Q}) \in \mathcal{L}(G, d)$. We say that G is (d, T) -upper-bad relative to \mathcal{C} and \mathcal{L} if G is (d, T) -upper-bad relative to \mathcal{L} and there exists at least one \mathcal{Q} such that $(\mathcal{C}, \mathcal{Q}) \in \mathcal{L}(G, d)$. A k -marked chain $(\mathcal{C}, \mathcal{Q})$ is good relative to \mathcal{L} if \mathcal{Q} does not contain a vertex G that is either (d, T) -lower-bad or (d, T) -upper-bad relative to \mathcal{C} and \mathcal{L} for any $1 \leq d \leq k$.

Proposition 2.12 (Equivalent of Proposition 4.1. in [2]). *Let $(\mathcal{C}, \mathcal{Q})$ be a member of \mathcal{L} that is good relative to \mathcal{L} , and let $G \in \mathcal{Q}$. Suppose G is the d th set of \mathcal{Q} . Then for any family \mathcal{S} of at most $|T|$ sets of \tilde{B}_n^T , where $\mathcal{S} \cap \mathcal{U}(G) = \emptyset$, there exists a member $(\mathcal{C}, \mathcal{Q}) \in \mathcal{L}(G, d)$ such that \mathcal{C} is disjoint from $\mathcal{D}^*(G, \mathcal{S})$. For any family \mathcal{S} of at most $|T|$ sets of \tilde{B}_n^T , where $\mathcal{S} \cap \mathcal{D}(G) = \emptyset$, there exists a member $(\mathcal{C}, \mathcal{Q}) \in \mathcal{L}(G, d)$ such that \mathcal{C} is disjoint from $\mathcal{U}^*(G, \mathcal{S})$.*

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

Theorem 2.13 (Equivalent of Theorem 5.1. in [2]). *Let $k, t \geq 2$ be integers and let ε be a fixed positive real. Suppose that $\mathcal{F} \subseteq \tilde{B}_n$ is a family with $|\mathcal{F}| \geq (k - 1 + \varepsilon) \binom{n}{\lfloor n/2 \rfloor}$. For each $\mathcal{C} \in \mathbf{C}_n$, let $Y(\mathcal{C})$ denote the set of members of \mathcal{F} contained in \mathcal{C} . If n is large enough with respect to k, t , and ε , then there exist functions X_1, \dots, X_t from \mathbf{C}_n to $2^{\mathcal{F}}$ satisfying the following.*

- (1) For all $\mathcal{C} \in \mathbf{C}_n$, $X_1(\mathcal{C}) = Y(\mathcal{C})$.
- (2) For all $1 \leq i \leq t - 1$ and $\mathcal{C} \in \mathbf{C}_n$, $X_{i+1}(\mathcal{C}) \subseteq X_i(\mathcal{C})$, and if $X_{i+1}(\mathcal{C}) \neq \emptyset$ then

$$\frac{|X_{i+1}(\mathcal{C})|}{|X_i(\mathcal{C})|} \geq 1 - \frac{1}{4kt}.$$

- (3) For all $1 \leq i \leq t$, the family of k -marked chains \mathcal{L}_i with markers in \mathcal{F} , defined by $\mathcal{L}_i = \{(\mathcal{C}, \mathcal{Q}) : \mathcal{C} \in \mathbf{C}_n, \mathcal{Q} \in \binom{X_i(\mathcal{C})}{k}\}$, satisfies

$$|\mathcal{L}_i| \geq \varepsilon(n!/k)(1 - \frac{i}{2t}).$$

- (4) For all $1 \leq i \leq t - 1$, every member of \mathcal{L}_{i+1} is good relative to \mathcal{L}_i .

The proof of Theorem 2.13 is given in the appendix. With all the above preparation, we are ready to prove Theorem 2.1.

Proof of Theorem 2.1. Let T be a tree poset of height $k + 1$ such that m is a maximal element of T with all chains of length $k + 1$ in T containing m . Let m_1, m_2, \dots, m_a denote the neighbors of m in $H(T)$. Let T_0 be the tree obtained from T by reversing all arcs $\overrightarrow{m_i m}$ in $\vec{H}(T)$ and adding vertices v_1, v_2, \dots, v_{k-2} forming a chain $v_1 \leq v_2 \leq \dots \leq v_{k-2} \leq m$. Clearly, $h(T_0) = k$ and $T \setminus \{m\} = T_0 \setminus \{m, v_1, v_2, \dots, v_{k-2}\}$. Let T'_0 be the k -saturated tree poset obtained from T_0 by Lemma 2.9. Observe that for any pair $p, q \in \mathcal{D} := \cup_{i=1}^a \mathcal{D}_{T'_0}(m_i)$ the poset distance in T'_0 is at most 4, and if $a \geq 2$, then for any $v \in P \setminus \mathcal{D}$ there exists $p \in \mathcal{D}$ such that their poset distance is larger than 4. If $a = 1$, then the poset distance of pairs in \mathcal{D} is at most 2. In both cases, one can apply Lemma 2.10 to obtain a sequence $T'_0, T'_1, \dots, T'_\ell$ such that

- all T'_i are k -saturated,
- $T'_{i-1} \setminus T'_i$ is $[u_i, v_i] \setminus \{u_i\}$ or $[u_i, v_i] \setminus \{v_i\}$ where $[u_i, v_i]$ is a chain interval,
- T'_ℓ is a chain,
- and $T'_{i_0} = \mathcal{D}$ for some i_0 .

Let $\mathcal{F} \subseteq \tilde{B}_n$ be a family of size at least $(k - 1 + \varepsilon) \binom{n}{\lfloor n/2 \rfloor}$. Applying Theorem 2.13, we obtain a sequence of nested k -marked chains $\mathcal{L}_0, \mathcal{L}_1, \dots, \mathcal{L}_\ell$. By property (3) of Theorem 2.13, \mathcal{L}_ℓ is non-empty, so taking the \mathcal{Q} -part of any k -marked chain $(\mathcal{C}, \mathcal{Q})$ embeds T'_ℓ into \tilde{B}_n and for any $Q \in \mathcal{Q}$ at the d -level of T'_ℓ we

have $\mathcal{L}_\ell(Q, d) \neq \emptyset$. As in [2], we proceed by backward induction to embed T'_i for all $i < \ell$ with the extra conditions that if Q plays the role of an element of T'_i at level d from top, then $\mathcal{L}_i(Q, d) \neq \emptyset$ and also that if $p \in T'_0 \setminus \mathcal{D}$, then G_p is not comparable to $G = \cup_{j=1}^a G_{m_j}$. Note that if $p \in T'_0 \setminus \mathcal{D}$, then $p \in T'_i \setminus T'_{i+1}$ for some $i < i_0$ and so m_1, m_2, \dots, m_a are already embedded. Suppose we have managed to do this for T'_{i+1}, \dots, T'_ℓ and let $\mathcal{G}_j \subseteq \mathcal{F}$ be the copy of T'_j obtained in the process. We know $T'_{i+1} = T'_i \setminus (I_i \setminus u_i)$ where $I_i = [u_i, v_i]$ or $I_i = [v_i, u_i]$ v_i being at level 1 or k from top. Without loss of generality, we may assume the former case and assume u_i is at level d from top. Then by the inductive hypothesis $\mathcal{L}_{i+1}(Q, d) \neq \emptyset$ for the set $Q \in \mathcal{F}$ playing the role of u_i . If $i \geq i_0$, then let $\mathcal{S}_i = \mathcal{G}_i \setminus \mathcal{D}_{\mathcal{G}_i}(Q)$, while if $i < i_0$, then set $\mathcal{S}_i = \mathcal{G}_i \setminus \mathcal{D}_{\mathcal{G}_i}(Q) \cup \{G\}$. As $T'_0 \setminus T'_1 \neq \emptyset$, we have $|\mathcal{S}_i| \leq |T'_0| = t$ in both cases. By Proposition 2.12, there exists a member $(\mathcal{C}_i, \mathcal{Q}_i) \in \mathcal{L}_i(Q, d)$ with \mathcal{C} being disjoint with $\mathcal{U}^*(Q, \mathcal{S}_i)$. We can embed $I \setminus \{u_i\}$ with the part of \mathcal{Q}_i below Q . As these newly added sets are not in the forbidden neighborhood $\mathcal{U}^*(Q, \mathcal{S}_i)$, together with \mathcal{G}_{i+1} , they form a copy \mathcal{G}_i of T'_i such that the extra condition of not being comparable to G hold. Also, the \mathcal{L}_i s are nested, $(\mathcal{C}_i, \mathcal{Q}_i)$ shows that the other extra condition hold. \square

2.3. Posets not containing weak copies of $2C_2$.

Proof of Proposition 1.3. For the lower bound of (1), consider the following construction: for j with $2 \leq j \leq s-1$ take the chain $\mathcal{C}_j = \{\{j\}, [j, j+1], [j, j+2], \dots, [j, j-3], [j, j-2]\}$, where $[a, b]$ with $b < a$ denotes the set $[a, n] \cup [b]$. Also, let \mathcal{C}_1 be the chain $\{\{1\}, [2], \dots, [s-2], [s-2] \cup \{s\}, [s-2] \cup [s, s+1], \dots, [s-2] \cup [s, n]\}$. Color all sets in $\{\emptyset, [n]\} \cup \cup_{i=1}^{s-1} \mathcal{C}_i$ with distinct colors and color all remaining sets with the color of $[n]$. This coloring uses $(s-1)(n-1) + 2$ colors and does not admit a strong rainbow copy of \wedge_s as \emptyset is not a member of any strong copy. Also, by construction, any set $C \in \mathcal{C}_i$ does not contain any set $C' \in \mathcal{C}_{i-1}$ and so the maximum rainbow antichain that C contains has size $s-1$: $s-2$ sets from \mathcal{C}_j s plus one colored as $[n]$. Finally, any set colored as $[n]$ contains rainbow antichains only from $\cup_{i=1}^{s-1} \mathcal{C}_i$, so at most $s-1$ sets. This shows $\text{ar}^*(n, \wedge_s) \geq (s-1)(n-1) + 2$. For the upper bound, observe that if a coloring of $2^{\lfloor n \rfloor}$ uses more than $(s-1)(n-1) + 2$ colors, then, by Dilworth's theorem, there is a rainbow antichain of size s that uses colors differing from that of \emptyset and that of $[n]$.

For the lower bound of (2) consider $k-2$ pairwise internally disjoint maximal chains and color all sets in their union by distinct colors and use PURPLE for all other sets. As any antichain can contain at most one set from each chain, the maximum size of a rainbow antichain is $k-1$. This proves $\text{ar}^*(n, A_k) \geq 3 + (k-2)(n-1)$. For the upper bound, observe that any level $\binom{[n]}{j}$ can use at most $k-1$ colors apart from that of $[n]$ (and that of \emptyset) as otherwise would have a rainbow A_k . Also, any level $\binom{[n]}{j}$ can use at most $k-2$ colors not used in $\binom{[n]}{k}$. Indeed, if $k < j$ and F_1, F_2, \dots, F_{k-1} are sets of size j using colors unused by any set of size k , then pick $x_i \notin F_i$ for all $1 \leq i \leq k-1$ and a k -set containing all x_i s (they may coincide) forms a rainbow A_k with the F_i s. If $j < k$, then for such F_i s consider $y_i \in F_i$ and a k -subset F of $[n] \setminus \{y_1, y_2, \dots, y_{k-1}\}$ (here we use $k \leq n-k$). So the total number of colors used is $2 + (k-1) + (n-2)(k-2)$ (two for the colors of $[n], \emptyset$, then $k-1$ for the colors used in $\binom{[n]}{k}$, and $k-2$ additional colors for all other $n-2$ levels). This completes the proof.

To see the upper bound of (3), observe that for any $j \leq n-2$, by counting pairs (F, x) $x \notin F$, if $\mathcal{F} \subseteq \binom{[n]}{j}$ contains more than $\frac{sn}{2}$ sets, then there exists a subfamily $\mathcal{F}' \subseteq \mathcal{F}$ such that $|\mathcal{F}'| \geq s+1$ and $|\cup_{F \in \mathcal{F}'} F| < n$. Let m be the size of the smallest possible $|\cup_{F \in \mathcal{F}'} F|$ with this property. Using the minimality of m , the same reasoning shows that $|\mathcal{F}'| \leq \frac{sm}{2}$. So if $|\mathcal{F}| > \frac{sn}{2} + k+1$, then there exist \mathcal{F}' and \mathcal{F}'' with $|\mathcal{F}'| = s+1, |\mathcal{F}''| = k+1$ such that $F'' \not\subseteq \cup_{F' \in \mathcal{F}'} F'$ for all $F'' \in \mathcal{F}''$. Let us apply this statement to the family \mathcal{F} that we obtain by picking a set from each color meeting $\binom{[n]}{j}$. Now, adding $M = \cup_{F' \in \mathcal{F}'} F'$ and possibly removing the j -set in $\mathcal{F}' \cup \mathcal{F}''$ with the color of M yields a rainbow copy of $\wedge_s + A_k$. So, if a coloring avoids such a copy, then every level uses at most $\frac{sn}{2} + k+1$ colors. As there are $n-1$ levels and $\emptyset, [n]$, the result follows. \square

REFERENCES

- [1] J. Balogh, R.I. Garcia, M.C. Wıgal, On the number of P -free set systems for tree posets P . *Order* (2025), 1-13.
- [2] E. Boehnlein, T. Jiang, Set families with a forbidden induced subposet. *Combinatorics, Probability and Computing*, 21(4) (2012), 496-511.
- [3] B. Bukh, Set families with a forbidden subposet. *the Electronic Journal of Combinatorics*, 16 (2009), R142.
- [4] A. De Bonis, G.O.H. Katona, K.J. Swanepoel, Largest family without $A \cup B \subset C \cap D$, *J. Combin. Theory Ser. A* 111 (2005) 331-336.
- [5] D. Ellis, M.R. Ivan, I. Leader, Turán densities for daisies and hypercubes. *Bulletin of the London Mathematical Society*, 56(12) (2024), 3838-3853.
- [6] P. Erdős, On a lemma of Littlewood and Offord, *Bull. Amer. Math. Soc.* 51 (1945), 898-902.
- [7] P. Erdős, M. Simonovits, V. T. Sós: Anti-Ramsey theorems, *Infinite and finite sets* (Colloq., Keszthely, 1973; dedicated to P. Erdős on his 60th birthday), Vol. II; *Colloq. Math. Soc. János Bolyai*, Vol. 10, pp. 633-643, North-Holland, Amsterdam, 1975.
- [8] D. Gerbner, B. Patkós, *Extremal finite set theory*. (2018) Chapman and Hall/CRC.
- [9] J.R. Griggs, W.T. Li, Progress on poset-free families of subsets. In *Recent trends in combinatorics* (2016) pp. 317-338. Cham: Springer International Publishing.
- [10] J.R. Griggs, W.T. Li, L. Lu, Diamond-free families. *Journal of Combinatorial Theory, Series A*, 119(2) (2012), 310-322.
- [11] J.R. Griggs, L. Lu, On families of subsets with a forbidden subposet. *Combinatorics, Probability and Computing*, 18(5) (2009), 731-748.
- [12] T. Jiang, S. Longbrake, S. Spiro, L. Yepremyan, Tree posets: Supersaturation, enumeration, and randomness. *Canadian Journal of Mathematics*. Published online 2025:1-32. doi:10.4153/S0008414X2510165X
- [13] G.O.H. Katona, Y. Mao, K. Ozeki, Z. Wang, G. Yang, Boolean lattice without small rainbow subposets, arxiv manuscript, arXiv:2602.00680
- [14] G.O.H. Katona, T. Tarján, Extremal problems with excluded subgraphs in the n -cube, *Graph Theory, Lagow*, 1981, Lecture Notes in Math. 1018 (Springer-Verlag, Berlin, 1983) 084-093.
- [15] P. Keevash, D. Mubayi, B. Sudakov, J. Verstraëte, Rainbow Turán problems. *Combinatorics, Probability and Computing*, 16(1) (2007), 109-126.
- [16] L. Lu, On crown-free families of subsets. *Journal of Combinatorial Theory, Series A*, 126 (2014), 216-231.
- [17] A. Methuku, D. Pálvölgyi, Forbidden hypermatrices imply general bounds on induced forbidden subposet problems. *Combinatorics, Probability and Computing*, 26(4) (2017), 593-602.
- [18] B. Patkós, Induced and Non-induced Forbidden Subposet Problems, *The Electronic Journal of Combinatorics* (2015), P1.30.
- [19] B. Patkós, Rainbow Turán problems for forbidden subposets, (2025) arXiv preprint arXiv:2511.14298.
- [20] B. Patkós, A. Treglown, On Some Extremal and Probabilistic Questions for Tree Posets. *The Electronic Journal of Combinatorics*, (2024) P1-19.

3. APPENDIX

We include a proof of Theorem 2.13 which is basically identical to that of Theorem 5.1. in [2].

Lemma 3.1 (Equivalent of Lemma 3.1. in [2]). *Let $F \in \tilde{B}_n, \mathcal{S} \subseteq \tilde{B}_n^T$ with $\mathcal{S} \cap \mathcal{U}(\mathcal{S}) = \emptyset$ and $|\mathcal{S}| \leq n/6$. If \mathcal{C} is a uniformly chosen random maximal chain in $\mathcal{D}(F)$, then*

$$\mathbb{P}(\mathcal{C} \cap \mathcal{D}^*(F, \mathcal{S}) \neq \emptyset) \leq \frac{12(|T| + 4)|\mathcal{S}| \sqrt{n \ln n}}{n}$$

if n is large enough.

Proof. The probability that $S \in \mathcal{S} \cap \mathcal{D}(F) \cap \tilde{B}_n$ belongs to \mathcal{C} is $\binom{|F|}{|S|}^{-1} \leq \frac{1}{|F|} \leq \frac{3}{n}$ and so by the union bound, writing $|\mathcal{S}| = s$, we have

$$(2) \quad \mathbb{P}(\mathcal{C} \cap \mathcal{S} \cap \mathcal{D}(F) \cap \tilde{B}_n) \leq \frac{3s}{n}.$$

Let S_1, S_2, \dots, S_p with $p \leq s$ be the sets of \mathcal{S} in $\mathcal{D}(F)$. We have $|F| - |S_i| \leq 4(|T| + 1)\sqrt{\ln n}$ and thus $|\cap_{i=1}^p S_i| \geq |F| - p4(|T| + 1)\sqrt{\ln n}$. If \mathcal{C} 's first set below F is $F \setminus \{a\}$ for some $a \in \cap_{i=1}^p S_i$, then \mathcal{C} is disjoint

with $\mathcal{U}(\mathcal{S}) \cap (\mathcal{D}(F) \setminus \{F\})$, and so using $|F| \geq n/3$ we have

$$(3) \quad \mathbb{P}(\mathcal{C} \cap \mathcal{U}(\mathcal{S}) \cap (\mathcal{D}(F) \setminus \{F\}) \neq \emptyset) \leq \frac{12s(|T| + 1)\sqrt{n \ln n}}{n}.$$

By the assumption, $\mathcal{U}(F) \cap \mathcal{S} = \emptyset$, we have elements $x_i \in F \setminus S_i$ for all $1 \leq i \leq s$. If the set C of \mathcal{C} of size $n/2 - 2\sqrt{n \ln n}$ contains all x_i s, then C does not get into $\mathcal{D}(\mathcal{S}) \cap \tilde{B}_n$. So writing $\ell = |F| - (n/2 - 2\sqrt{n \ln n}) \leq 4\sqrt{n \ln n}$, using $|F| - \ell \geq n/3$ and $(1 - x)^\ell \geq 1 - 2x\ell$ for $x \leq 1/2$, we obtain

$$(4) \quad \mathbb{P}(\mathcal{C} \cap \mathcal{D}(\mathcal{S}) \cap \mathcal{D}(F) \cap \tilde{B}_n \neq \emptyset) \leq 1 - \frac{|F| - s}{|F|} \cdot \frac{|F| - s - 1}{|F| - 1} \cdots \frac{|F| - s - \ell}{|F| - \ell} \leq 1 - \left(1 - \frac{s}{|F| - \ell}\right)^\ell \leq \frac{24s\sqrt{n \ln n}}{n}.$$

Adding up (2), (3), and (4), we obtain the claim of the Lemma. \square

From here on, we use the notation of [2], so a maximal chain is denoted by M rather than \mathcal{C} sets in \mathbb{B}_n are denoted by u or v , and so on.

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

Given a sequence $J = (j_1, j_2, \dots, j_q)$ of numbers in $[n]$, where either $j_1 < j_2 < \dots < j_q$ or $j_1 > j_2 > \dots > j_q$, and a chain C in \mathbb{B}_n , let $C[J]$ denote the subchain of C consisting of the j_1 -th, j_2 -th, \dots , j_q -th members of \mathcal{F} on C (counted from the top). If C contains fewer than 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 $C[\{j\}]$.

As [2], $h < n/6$ is a bound on the size of all witnesses.

Lemma 3.2. *Let $d \in [k]$. Let p be a positive integer. Let J be an increasing sequence of $2p$ numbers in $[n]$. Let $v \in \tilde{\mathbb{B}}_n$. Let M be a uniformly chosen random full chain of $D(v)$. Then*

$$\Pr(M[J] \text{ forms a } (d, T)\text{-lower-bad string}) \leq \left(\frac{12(|T| + 4)h\sqrt{n \ln n}}{n}\right)^p.$$

Proof. Let $\gamma = \frac{12(|T| + 4)h\sqrt{n \ln n}}{n}$. We use induction on p . For fixed p , we prove the statement for all J with $2p$ numbers and all $v \in \tilde{\mathbb{B}}_n$. For the basis step, let $p = 1$. Suppose $J = (j_1, j_2)$, where $j_1 < j_2$. Let $v \in \tilde{\mathbb{B}}_n$ be given. Let M be a uniformly chosen random full chain of $D(v)$. We have

$$\Pr(M[J] \text{ is a } (d, T)\text{-lower-bad string}) \leq \sum_{u \in D(v)} \Pr(M[j_1] = u) \cdot \Pr(M[j_2] \in D^*(u, S_{u,d}) \mid M[j_1] = u)$$

Fix any $u \in D(v)$. The set of 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 $v \geq x \geq u$) that contain exactly j_1 members of \mathcal{F} and all full chains of $D(u)$. So, $\Pr(M[j_2] \in D^*(u, S_{u,d}) \mid M[j_1] = u)$ 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 $D^*(u, S_{u,d})$, which by Lemma 3.1, is at most γ . Hence,

$$\begin{aligned} \Pr(M[J] \text{ is a } d\text{-lower-bad string}) &\leq \sum_{u \in D(v)} [\Pr(M[j_1] = u) \cdot \gamma] \\ &= \gamma \cdot \sum_{u \in D(v)} \Pr(M[j_1] = u) \leq \gamma, \end{aligned}$$

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 $p \geq 2$. 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 $J' = (j_3 - j_2 + 1, j_4 - j_2 + 1, \dots, j_{2p} - j_2 + 1)$. We have

$$\begin{aligned} \Pr(M[J] \text{ is a } (d, T)\text{-lower-bad string}) &\leq \sum_{u \in D(v)} \sum_{y \in D^*(u, S_{u,d})} [\Pr(M[j_1] = u) \cdot \Pr(M[j_2] = y \mid M[j_1] = u) \\ &\quad \cdot \Pr(M_y[J'] \text{ is a } (d, T)\text{-lower-bad string} \mid M[j_1] = u, M[j_2] = y)] \end{aligned}$$

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, T) -lower-bad string is the same as the probability that given a random full chain M' of $D(y)$, $M'[J']$ forms a (d, T) -lower-bad string. By induction hypothesis, this is at most γ^{p-1} . So,

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

This completes the induction step and our proof. \square

We now show Theorem 2.13 (that we restate with the notation of [2]) that we can obtain a sequence of families of k -marked chains with markers in \mathcal{F} , $\mathcal{L}_1 \supseteq \mathcal{L}_2 \dots \supseteq \mathcal{L}_h$, such that for each $i \in [h]$, $|\mathcal{L}_i| \geq (\epsilon n! / k)(1 - \frac{i}{2k})$ and 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 .

First we state a lemma from [3].

Lemma 3.3. ([3]) *Let ϵ be a small positive real. Let n be a sufficiently large positive integer. Let $\mathcal{F} \subseteq \widetilde{\mathbb{B}}_n$. Let \mathcal{L} denote the family of all the k -marked chains with markers in \mathcal{F} . If $|\mathcal{F}| \geq (k-1+\epsilon)\binom{n}{\lfloor n/2 \rfloor}$, then*

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

Theorem 3.4. *Let k, h be positive integers and let ϵ be a fixed positive real. Let n be sufficiently large (as a function of k, h, ϵ). Let $\mathcal{F} \subseteq \widetilde{\mathbb{B}}_n$ be a family with $|\mathcal{F}| \geq (k-1+\epsilon)\binom{n}{\lfloor n/2 \rfloor}$. 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, \dots, X_h from $\mathcal{C}(\mathbb{B}_n)$ to $2^{\mathcal{F}}$ such that the following holds:*

- (1) $\forall M \in \mathcal{C}(\mathbb{B}_n), X_1(M) = Y(M)$.
- (2) $\forall i, 1 \leq i \leq h-1, \forall M \in \mathcal{C}(\mathbb{B}_n)$ if $X_{i+1}(M) \neq \emptyset$ then $\frac{|X_{i+1}(M)|}{|X_i(M)|} \geq 1 - \frac{1}{4kh}$.
- (3) $\forall i \in [h]$, 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

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

(4) $\forall i \in [h-1]$, every member of \mathcal{L}_{i+1} is good relative to \mathcal{L}_i (where good and bad are defined with respect to h , i.e. witnesses of being bad are of size at most h).

Proof. 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 3.3, we have

$$(5) \quad |\mathcal{L}_1| \geq (\epsilon/k)n!.$$

So item 3 holds. There is nothing else to prove. For the induction step, let $i \geq 1$ and suppose the functions X_1, \dots, 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, T) -lower-bad relative to \mathcal{L}_i , we fix a corresponding (d, T) -lower-witness $S_{v,d}$. For each $d \in [k]$ and each $v \in \widetilde{\mathbb{B}}_n$ that is (d, T) -upper-bad relative to \mathcal{L}_i , we fix a corresponding (d, T) -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 $\mathbb{B}_{i,d}^-(M)$ denote the set of vertices in $X_i(M)$ that are (d, T) -lower-bad relative to M and \mathcal{L}_i . Let $\mathbb{B}_i^-(M) = \bigcup_{d=1}^k \mathbb{B}_{i,d}^-(M)$. For each $d \in [k]$, let $\mathbb{B}_{i,d}^+(M)$ denote the set of vertices in $X_i(M)$ that are (d, T) -upper-bad relative to M and \mathcal{L}_i . Let $\mathbb{B}_i^+(M) = \bigcup_{d=1}^k \mathbb{B}_{i,d}^+(M)$. Let $\mathbb{B}_i(M) = \mathbb{B}_i^-(M) \cup \mathbb{B}_i^+(M)$. Let $x(M) = |X_i(M)|$ and let $b(M) = |\mathbb{B}_i(M)|$. Set $C = 4kh$. Let

$$\begin{aligned} \mathcal{C}_1 &= \{M \in \mathcal{C}(\mathbb{B}_n) : X_i(M) \neq \emptyset, \frac{b(M)}{x(M)} \leq \frac{1}{C}\} \\ \mathcal{C}_2 &= \{M \in \mathcal{C}(\mathbb{B}_n) : X_i(M) \neq \emptyset, \frac{b(M)}{x(M)} > \frac{1}{C}\} \end{aligned}$$

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

If $X_i(M) = \emptyset$ or $M \in \mathcal{C}_2$, then let $X_{i+1}(M) = \emptyset$

Otherwise, $M \in \mathcal{C}_1$, and we let $X_{i+1}(M) = X_i(M) \setminus \mathbb{B}_i(M)$

Claim 1. We have

- (1) $\forall M \in \mathcal{C}(\mathbb{B}_n)$, where $X_{i+1}(M) \neq \emptyset$, we have $|X_{i+1}(M)| \geq (1 - \frac{1}{C})|X_i(M)| \geq (1 - \frac{1}{4kh})|X_i(M)|$.
- (2) Each member of \mathcal{L}_{i+1} is good relative to \mathcal{L}_i .
- (3) $\sum_{M \in \mathcal{C}_1} \binom{|X_{i+1}(M)|}{k} \geq (1 - \frac{k}{C}) \sum_{M \in \mathcal{C}_1} \binom{|X_i(M)|}{k} \geq (1 - \frac{1}{4h}) \sum_{M \in \mathcal{C}_1} \binom{|X_i(M)|}{k}$.

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 $|\mathbb{B}_i(M)|/|X_i(M)| \leq 1/C$. Since $X_{i+1}(M) = X_i(M) \setminus \mathbb{B}_i(M)$, item 1 follows immediately. All 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 $X_{i+1}(M) = X_i(M) \setminus \mathbb{B}_i(M)$, and $Q \in \binom{X_{i+1}(M)}{k}$, Q contains no vertex that is either (d, T) -lower-bad or (d, T) -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 $b(M) = |\mathbb{B}_i(M)|$ and $x(M) = |X_i(M)|$. The number of k -subsets of $X_i(M)$ that contain a member of $\mathbb{B}_i(M)$ is certainly at most

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

Therefore, we have

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

So item 3 (of Claim 1) holds. \square

Claim 2. We have $\sum_{M \in \mathcal{C}_2} \binom{|X_i(M)|}{k} \leq \frac{4k}{n^{1/3}} \cdot n!$.

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 $|\mathbb{B}_i^-(M)| \geq |\mathbb{B}_i(M)|/2 = b(M)/2$ 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 $|\mathbb{B}_{i,d}^-(M)| \geq |\mathbb{B}_i^-(M)|/k$. Clearly, $\mathcal{C}_2^- = \bigcup_{d=1}^k \mathcal{C}_{2,d}^-$. For each $d \in [k]$, we first bound $\sum_{M \in \mathcal{C}_{2,d}^-} \binom{|X_i(M)|}{k}$.

For each $M \in \mathcal{C}_{2,d}^-$, we define a sequence $R_d^-(M)$, called the *greedy (d, T) -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 $\mathbb{B}_{i,d}^-(M)$ that we encounter. Recall that this means x_1 is (d, T) -lower-bad relative to M and \mathcal{L} and we have fixed a (d, T) -lower-witness $S_{x_1, d}$ of v (relative to \mathcal{L}_i) with $|S_{x_1, d}| \leq h$ 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 d -th one among them form a Q with $(M, Q) \in \mathcal{L}_i(x_1, d)$. Since x_1 is (d, T) -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, T) -lower-bad relative to M and \mathcal{L}_i , then let x_2 denote the next vertex in $X_i(M)$ that is (d, T) -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 $R_d^-(M) = (x_1, y_1, x_2, y_2, \dots, x_p, y_p)$. By our procedure, $p \geq \lceil |\mathbb{B}_{i,d}^-(M)|/k \rceil$. 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, T) -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,

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

So $|Y(M)| = |X_1(M)| \leq 2|X_i(M)| \leq 2|\mathbb{B}_i(M)|C \leq 4|\mathbb{B}_i^-(M)|C \leq 4k|\mathbb{B}_{i,d}^-(M)| \leq 4k^2pC$ (recall that $p \geq \frac{|\mathbb{B}_{i,d}^-(M)|}{k}$). Clearly the largest number in $P_d^-(M)$ is no more than $|Y(M)| \leq 4k^2pC$. 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, T) -lower-bad string relative to \mathcal{L}_i by the definition of $R_d^-(M)$. Thus

$$\begin{aligned} |\{M \in \mathcal{C}_{2,d}^- : P_d^-(M) = J\}| &\leq |\{M \in \mathcal{C}(\mathbb{B}_n) : M[J] \text{ forms a } (d, T)\text{-lower-bad string relative to } \mathcal{L}_i\}| \\ &\leq \left(\frac{12(|T| + 4)h\sqrt{n \ln n}}{n}\right)^p \cdot n! \quad (\text{by Lemma 3.2}) \end{aligned}$$

So

$$|\{M \in \mathcal{C}_{2,d}^- : |P_d^-(M)| = 2p\}| \leq \binom{4k^2pC}{2p} \cdot \left(\frac{12(|T| + 4)h\sqrt{n \ln n}}{n}\right)^p \cdot n! \leq 2^{4k^2pC} \left(\frac{12(|T| + 4)h\sqrt{n \ln n}}{n}\right)^p \cdot n!.$$

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

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

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

$$2^{4k^2pC} \cdot 2^{4k^2pC} \cdot \left(\frac{12(|T| + 4)h\sqrt{n \ln n}}{n}\right)^p \cdot n! \leq \left(\frac{1}{n^{1/3}}\right)^p \cdot n!,$$

for sufficiently large n . Summing over all $p \geq 1$, we get

$$\sum_{M \in \mathcal{C}_{2,d}^-} \binom{|X_i(M)|}{k} \leq \sum_{p=1}^{\infty} \left(\frac{1}{n^{1/3}}\right)^p \cdot n! \leq \frac{2}{n^{1/3}} \cdot n!,$$

for large n . Summing over all $d \in [k]$, we get

$$\sum_{M \in \mathcal{C}_2^-} \binom{|X_i(M)|}{k} \leq \frac{2k}{n^{1/3}} \cdot n!,$$

By a similar argument, we have

$$\sum_{M \in \mathcal{C}_2^+} \binom{|X_i(M)|}{k} \leq \frac{2k}{n^{1/3}} \cdot n!,$$

for large n . Hence

$$\sum_{M \in \mathcal{C}_2} \binom{|X_i(M)|}{k} \leq \frac{4k}{n^{1/3}} \cdot n!,$$

for large n . □

Claim 3. We have

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

Proof of Claim 3. By induction hypothesis,

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

By Claim 2, $\sum_{M \in \mathcal{C}_2} \binom{|X_i(M)|}{k} \leq \frac{4k}{n^{1/3}} \cdot n! \leq (\epsilon/k)n!(\frac{1}{4h})$, for large n . So

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

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

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

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

HUN-REN ALFRÉD RÉNYI INSTITUTE OF MATHEMATICS
 Email address: patkos@renyi.hu