Strong Erdős–Hajnal properties in chordal graphs

Minho Cho[a] Andreas F. Holmsen[b,d] Jinha Kim[c,d] Minki Kim[e]

Submitted: Jun 2, 2023; Accepted: Mar 15, 2024; Published: May 31, 2024

Abstract

A graph class $\mathcal{G}$ has the strong Erdős–Hajnal property (SEH-property) if there is a constant $c = c(\mathcal{G}) > 0$ such that for every member G of $\mathcal{G}$, either G or its complement has $K_{m,m}$ as a subgraph where $m \geqslant \lfloor c|V(G)|\rfloor$. We prove that the class of chordal graphs satisfy SEH-property with constant $c = 2/9$.

On the other hand, a strengthening of SEH-property which we call the colorful Erdős–Hajnal property was discussed in geometric settings by Alon et al. (2005) and by Fox et al. (2012). Inspired by their results, we show that for every pair F_1, F_2 of subtree families of the same size in a tree T with k leaves, there exists subfamilies $F_1' \subseteq F_1$ and $F_2' \subseteq F_2$ of size $\theta\left(\frac{\ln k}{k}|F_1|\right)$ such that either every pair of representatives from distinct subfamilies intersect or every such pair do not intersect. Our results are asymptotically optimal.

Mathematics Subject Classifications: Primary 05C69; Secondary 05C05, 05C35

1 Introduction

Background.

A classical conjecture of Erdős and Hajnal [10] asserts that if G is a graph on n vertices which does not contain some fixed graph H as an induced subgraph, then G contains a clique or an independent set on at least $\lfloor n^\delta \rfloor$ vertices where $\delta > 0$ is a constant depending only on the graph H. In general we say that a graph class $\mathcal{G}$ has the *Erdős–Hajnal property* if there exists a constant $\delta = \delta(\mathcal{G})$ such that every graph in $\mathcal{G}$ on $n > 1$ vertices contains a clique or an independent set of size n^δ. (Here we use the term *graph class* to mean a family of graphs that is closed under taking induced subgraphs.) [5, 10, 14]

[a] Extremal Combinatorics and Probability Group, Institute for Basic Science (IBS), Daejeon, South Korea (`minhocho@ibs.re.kr`).
[b] Department of Mathematical Sciences, KAIST, Daejeon, South Korea (`andreash@kaist.edu`).
[c] Department of Mathematics, Chonnam National University, Gwangju, South Korea (`jinhakim@jnu.ac.kr`).
[d] Discrete Mathematics Group, Institute for Basic Science (IBS), Daejeon, South Korea.
[e] Division of Liberal Arts and Sciences, GIST, Gwangju, South Korea (`minkikim@gist.ac.kr`).

Rather than asking for a large clique or independent set, one variation of the Erdős–Hajnal problem, asks for a large *bi-clique* as a subgraph in G or in the complement of G. Here a *bi-clique of size* $2k$ is a complete bipartite graph whose vertex classes each consists of k vertices.

A graph class $\mathcal{G}$ is said to have the *strong Erdős–Hajnal property* (SEH-property) if there exists a constant $\varepsilon = \varepsilon(\mathcal{G}) > 0$ such that every graph $G \in \mathcal{G}$ on n vertices or its complement $\overline{G}$ contains bi-clique of size $2\lfloor \varepsilon n \rfloor$. It was shown in [1] that if a graph class has the SEH-property, then it also has the Erdős–Hajnal property.[1]

It is known that the class of all H-free graphs satisfies the SEH-property if and only if both H and its complement $\overline{H}$ are forest. (For a simple proof, see [8].) Actually, the latter condition does not hold for any H on more than four vertices, and this shows the restrictive aspect of the SEH-property in solving the Erdős-Hajnal conjecture. However, the above classification of classes of H-free graphs satisfying the SEH-property can be extended to find more graphs classes that are defined by a finite list of forbidden induced subgraphs. Given a collection F of graphs, G is called *F-free* if G does not contain any graph in F as an induced subgraph. In [7], it was revealed that if F is finite, then the collection of F-free graphs has the SEH-property if and only if F contains (possibly identical) two graphs H_1 and H_2 such that H_1 and $\overline{H_2}$ are forests.

If F is an infinite collection, then the situation becomes more subtle and not much is known in general. When F is the collection of all odd cycles of length at least five and their complements (also called *antiholes*), then by the strong perfect graph theorem [6], F-free graphs are precisely the perfect graphs, and they are known not to satisfy the SEH-property [12]. On the other hand, excluding all subdivisions of a single graph H and their complements guarantees the SEH-property [8]. By noting that every antihole on at least six vertices contains C_4 as an induced subgraph, this result (see also [3]) immediately implies that the class of all chordal graphs have the strong Erdős-Hajnal property. Here, the graph is *chordal* if it contains no induced cycle of length 4 or greater. See [7, 8] for an overview on graph classes satisfying the SEH-property.

A number of graph classes arising from discrete geometry have been shown to have the SEH-property, most notably are the cases of semi-algebraic graphs [1] and intersection graphs of convex sets in the plane [15]. The goal of this paper is to study the SEH-property and related properties for some specific graph classes, with a focus on the class of chordal graphs, which is an extension of the class of interval graphs.

Our results.

The most general and powerful results regarding the SEH-property, typically do not give particular good bounds on the constant ε (nor do they aim to do so). One of the goals of this paper is to provide (asymptotically) optimal constants for the SEH-property with respect to the following graph classes:

[1]The reader should be warned that the name "strong Erdős–Hajnal property" appears in the literature in various contexts. Here we are using the terminology introduced in [15].

- *Interval graphs.* An interval graph is the intersection graph of a finite family of intervals on the real line. That is, each vertex can be represented by an interval and two vertices are adjacent if and only if the corresponding intervals intersect.
- *Cographs.* A cograph (complement-reducible graph) is a graph that can be obtained from a single vertex by complementation and disjoint union. Equivalently, it is a graph which does not contain the path on four vertices as an induced subgraph.
- *Chordal graphs.* A chordal graph is a graph in which every cycle on four or more vertices has a chord, that is, there are no induced cycle on four or more vertices. Equivalently, a chordal graph is the intersection graph of a finite family of subtrees of an ambient tree [16]. (This is called the subtree representation of the chordal graph.)

Theorem 1. *The following graph classes satisfy the strong Erdős–Hajnal property.*

(1) Interval graphs with constant $\varepsilon = 1/4$.

(2) Cographs with constant $\varepsilon = 1/4$.

(3) Chordal graphs with constant $\varepsilon = 2/9$.

We now turn our attention to a variation of the SEH-property. We say that a graph class $\mathcal{G}$ has the *colorful Erdős–Hajnal property* (CEH-property) if there exists a constant $\varepsilon_c = \varepsilon_c(\mathcal{G}) > 0$ such that for any graph $G \in \mathcal{G}$ on n vertices and *for any* partition of the vertex set $V(G)$ into parts of size $\lceil n/2 \rceil$ and $\lfloor n/2 \rfloor$, G or its complement $\overline{G}$ contains a bi-clique of size $2\left\lfloor \frac{\varepsilon_c n}{2} \right\rfloor$ whose vertex classes belong to different parts of the given partition of $V(G)$. In other words, we ask not only for a large bi-clique in G or its complement $\overline{G}$, but for one that respects an arbitrarily preassigned equipartition of the vertex set of G. See [1, 2, 11, 22] for more results related to the CEH-property.

It was shown in [1] that semi-algebraic graphs satisfy the CEH-property, but this general and powerful result does not give particularly good bounds on the constant involved. Our next goal is to determine (asymptotically) optimal constants for the CEH-property with respect to the same graph classes as in Theorem 1. However, it will be evident that the class of chordal graphs *does not* satisfy the CEH-property, and therefore we consider a refinement of this class.

Recall that the *leafage* of a chordal graph G, denoted by $\ell(G)$, is the minimum number of leaves of the ambient tree in a subtree representation of G. For an integer $k \geqslant 2$ let $\mathcal{T}_k$ denote the family of chordal graphs whose leafage is at most k. That is,

$$\mathcal{T}_k = \{G : G \text{ is chordal with } \ell(G) \leqslant k\}.$$

This gives us an infinite chain $\mathcal{T}_2 \subset \mathcal{T}_3 \subset \cdots \subset \mathcal{T}_\infty$ where $\mathcal{T}_2$ is the class of interval graphs and $\mathcal{T}_\infty$ is the class of chordal graphs.

Theorem 2. *The following graph classes satisfy the colorful Erdős–Hajnal property.*

(1) Interval graphs with constant $\varepsilon_c = 1/3$.

(2) Cographs with constant $\varepsilon_c = 1/4$.

(3) The class $\mathcal{T}_k$ *with constant* $\varepsilon_c = \frac{\ln k}{20k}$.

Basic terminology and notation

As usual, a graph G is an ordered pair $G = (V, E)$ consisting of a finite vertex set $V = V(G)$ and an edge set $E = E(G) \subset \binom{V}{2}$. In particular, all graphs in this paper are simple, having no loops and no parallel edges. The *complement graph* of a graph G is the graph $\overline{G} = (V, \binom{V}{2} - E(G))$. The disjoint union of two sets A and B is denoted by $A \mathbin{\dot\cup} B$, and the *disjoint union* of two given graphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ is the graph $(V_1 \mathbin{\dot\cup} V_2, E_1 \mathbin{\dot\cup} E_2)$. With a slight abuse of notation we denote this by $G_1 \mathbin{\dot\cup} G_2$.

A complete bipartite graph is simply called a *bi-clique.* A bi-clique $K_{m,n}$ is *balanced* if $m = n$. We define the *size* of bi-cliques only for balanced ones and the size of $K_{m,m}$ is $2m$.

Given a family F of nonempty sets, the *intersection graph* of F is a graph $G = (V, E)$ such that $V = F$ and two vertices A and B are adjacent in G if and only if $A \cap B \neq \emptyset$. Let G be the intersection graph of a family F of nonempty sets, and let F_1 and F_2 be two disjoint subfamilies of F. We say F_1 and F_2 *correspond to a bi-clique in* G if every member of F_1 intersects every member of F_2. Similarly, we say F_1 and F_2 *correspond to a bi-clique in* $\overline{G}$ if every member of F_1 is disjoint from every member of F_2.

For a vertex $v \in V(G)$, the *neighborhood* of v, denoted by $N(v)$, is the set of all vertices adjacent to v. The *closed neighborhood* of v is $N[v] := N(v) \cup \{v\}$. The *degree* of v, denoted by $\deg(v)$, is the number of adjacent vertices to v, and $\Delta(G)$ denotes the maximum degree over all vertices in G.

For a tree T, a *leaf* of T is a vertex of degree 1 in T. Given a pair of vertices $u, v \in V(T)$, we denote by $P_T(u, v)$ the unique path in T connecting u and v. More generally, for a vertex set $U = \{u_1, u_2, \ldots, u_n\} \subseteq V(T)$, the inclusion-minimal subtree of T that contains U is denoted by $\mathrm{Tree}_T(U)$ or $\mathrm{Tree}_T(u_1, u_2, \ldots, u_n)$. In other words, $\mathrm{Tree}_T(U) = \bigcup_{u,v \in U} P_T(u, v)$.

Outline of paper.

In Section 2 we provide examples guaranteeing that the constants in Theorems 1 and 2 cannot be increased (except for the class $\mathcal{T}_k$ whose construction will be given later). In Section 3 we prove two lemmas that will be useful in the proofs of both Theorems 1 and 2. The first one deals with "generic" subtree representations of chordal graphs and the other is a basic lemma on cographs. The proof of Theorem 1 is given in Section 4, and Section 5 contains the proof of Theorem 2 as well as a probabilistic construction that shows that our bound for the class $\mathcal{T}_k$ is asymptotically tight. We wrap up in Section 6 with some final remarks and open problems.

2 Optimality of constants in Theorem 1 and Theorem 2

Example. The constants in Theorem 1 can in general not be increased. Let us first consider the case of interval graphs. Let G_1 be the intersection graph of the intervals

$$I_1 = [0,1],\ I_2 = [1,2],\ I_3 = [2,3],\ I_4 = [3,4].$$

Note that G is a graph on four vertices and the largest bi-clique in G or $\overline{G}$ has size two. We can make arbitrarily large examples by taking k copies of each of the intervals. The resulting intersection graph has $4k$ vertices and the largest bi-clique in G or $\overline{G}$ has size $2k$.

Next we give a construction for the case of cographs. Obviously a complete graph is a cograph, and so the graph $G = \overline{(K_k \cup K_k \cup K_k)} \cup K_k$ is a cograph on $4k$ vertices and it is easily checked that the largest bi-clique in G or its complement $\overline{G}$ has size at most $2k$.

Finally, we give a construction for the case of chordal graphs by giving a subtree representation. Let $T = K_{1,3}$ and $V(T) = \{v, u_1, u_2, u_3\}$ where v is the vertex of degree 3. Let G be the intersection graph of the following nine subtrees of T:

$$\begin{aligned} &T_1 = T_2 = u_1, && T_3 = T_4 = u_2, && T_5 = T_6 = u_3, \\ &T_7 = P_T(u_1, u_2), && T_8 = P_T(u_1, u_3), && T_9 = P_T(u_2, u_3), \end{aligned}$$

where $P_T(u, v)$ denotes the unique path in T connecting vertices u and v. Thus G is a graph on nine vertices and it is easily checked that the largest bi-clique contained in G or $\overline{G}$ has size four. To obtain arbitrarily large examples simply take k copies of each subtree.

Example. The constants in Theorem 2 can in general not be increased. Here we give examples for the case of interval graphs and cographs. These are similar to the ones for the SEH-property. For the class $\mathcal{T}_k$ we will give an asymptotic matching bound, but the argument is a bit more involved and is given in section 5. Note that this implies that the class of chordal graphs (which is the class $\mathcal{T}_\infty$) does not satisfy the CEH-property for any fixed constant $\varepsilon_c > 0$.

Here is a construction for interval graphs. Let G be the intersection graph of the following intervals

$$\begin{aligned} &I_1 = [0,1], && I_2 = [2,3], && I_3 = [4,5], \\ &J_1 = [1,2], && J_2 = [3,4], && J_3 = [5,6], \end{aligned}$$

and consider the partition of the vertex set of G containing the I_m intervals in one part and the J_m intervals in the other. Thus G is a graph on six vertices and it is easily seen that the largest bi-clique in G that respects this given partition has size two. As before we can make arbitrarily large example by taking k copies of each of the intervals.

Here is a construction for cographs. Let H denote the bipartite graph in Figure 1. Note that the largest bi-clique that respects the vertex partition of H has size two.

Now consider the following cograph on vertices $\{v_1, \dots, v_8\}$. Let G_1 be the graph on $\{v_1, v_5\}$ without an edge, let G_2 be disjoint union of edges v_2v_6 and v_3v_7, and define

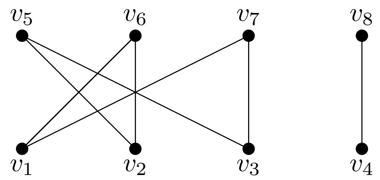

Figure 1: The bipartite graph H.

$G_3 = \overline{G_1} \cupdot \overline{G_2}$. Finally, let G_4 be the edge v_4v_8 and let $G_5 = \overline{G_3} \cupdot G_4$. It is easily seen that the induced bipartite subgraph between parts $\{v_1, \dots, v_4\}$ and $\{v_5, \dots, v_8\}$ is isomorphic to H.

For each $k \geqslant 1$, a general example G on $8k$ vertices can be made by replacing each v_i by *any* cograph H_i on k vertices and partitioning $U_1 = \bigcup_{i=1}^{4} V(H_i)$ and $U_2 = \bigcup_{i=5}^{8} V(H_i)$. Note that H_i can be any cograph because its edges disappear when we restrict G to the edges between two parts U_1 and U_2.

3 Auxiliary results

We start this section with a simple lemma regarding subtree representations of chordal graphs. This will be useful later on in the proofs of Theorems 1 and 2. Recall that $\mathcal{T}_k$ is the class of chordal graphs with leafage at most k.

Lemma 3. *Any graph $G \in \mathcal{T}_k$ has a subtree representation as the intersection graph of a family $\{T_i\}$ of subtrees of an ambient tree T where*

(i) The ambient tree T has k leaves and maximum degree 3.

(ii) No two subtrees T_i and T_j share a common leaf.

Proof. The fact that G has a subtree representation in an ambient tree T with at most k leaves follows from the well-known result of Gavril [16] and the definition of $\ell(G)$. We fix such a subtree representation $\{T_i\}$ and show how to modify the ambient tree T and the subtree representation without changing the intersection graph.

We first show how to reduce the maximum degree of the ambient tree T. Fix a vertex $v \in V(T)$ of degree $d \geqslant 4$. Let $C_1, C_2, \dots, C_d$ be the components of $T - v$, and let u_i be the unique neighbor of v in C_i for each $i \in [d]$. Introduce new vertices v_1, v_2, …, v_d connected by edges such that they form a path P.

We construct new tree T' on the vertex set $(V(T) - \{v\}) \cup \{v_1, v_2, \dots, v_d\}$. Edges in each C_i remain the same in T', each u_i is connected to v_i by an edge, and finally add the edges of P to T'. In other words,

$$E(T') = E(C_1) \cup E(C_2) \cup \cdots \cup E(C_d) \cup \{u_1v_1, \dots, u_dv_d\} \cup E(P).$$

It is obvious that T' is a tree. Now we construct the new subtree family $\{T_i'\}$. If the original subtree T_i contains the vertex v, then we set $V(T_i') = (V(T_i) - \{v\}) \cup \{v_1,$

…, $v_d\}$. Otherwise we set $V(T_i') = V(T_i)$. Finally let T_i' be the minimal subtree of T' that contains the vertices $V(T_i')$. It is easily seen that $\{T_i\}$ and $\{T_i'\}$ have isomorphic intersection graphs.

Note that the new vertices $v_1, \dots, v_d$ all have degree at most 3 in T', and that T' has the same number of leaves as T. Therefore repeating the process until there are no more vertices of degree greater than 3 proves claim (i).

To prove claim (ii), suppose $v \in V(T)$ is a common leaf of T_i and T_j. If v is a leaf of T, then we first modify T as follows; add a new vertex u to T and connect it to v. (We do not change any subtrees yet.) If v is not a leaf of T, then T remains unchanged.

Now let u be a neighbor of v which is not a vertex of T_i (which must exist after possibly making the change above). Subdivide the edge uv once; so the edge uv is replaced by the path uwv (in both T and every subtree containing the original edge uv), then add the edge vw to T_i. Note that these modifications maintain claim (i).

In effect, this reduces the total number of common leaves between subtrees, while the intersection graph remains the same. We repeat the same procedure until no two subtrees share a common leaf. □

The next lemma concerning cographs will be needed for the proofs of both Theorems 1 and 2. Consider a vertex v of a graph G and subset $W \subset V(G) \setminus \{v\}$. We say that W *conforms* to v if either $W \subset N(v)$ or $W \cap N(v) = \emptyset$.

Lemma 4. *Let G be a cograph on the vertex set V. For any nonempty vertex set $U \subseteq V$, there exists a subset $W \subseteq V$ such that*

(i) $\frac{1}{4}|U| \leqslant |U \cap W| \leqslant \max\{\frac{1}{2}|U|, 1\}$.

(ii) For every $v \in V - W$, W conforms to v.

Proof. Let $\mathcal{P}$ denote the class of all cographs. Recall the recursive definition of $\mathcal{P}$:

(i) The graph K_1 on one isolated vertex belongs to $\mathcal{P}$.

(ii) If $G \in \mathcal{P}$, then its complement $\overline{G}$ also belongs to $\mathcal{P}$.

(iii) If $G, H \in \mathcal{P}$, then their disjoint union $G \cup H$ also belongs to $\mathcal{P}$.

For our given graph $G \in \mathcal{P}$ we recursively define sequences of cographs $\{G_i\}$ and $\{H_i\}$ as follows. Start by setting $G_1 = G$. For G_i with $i \geqslant 1$ and $|V(G)| \geqslant 2$, either G_i or $\overline{G_i}$ equals the disjoint union of two cographs $G_{i+1} \cup H_{i+1}$ by the recursive construction of G_i. Select G_{i+1} so that $|U \cap V(G_{i+1})| \geqslant |U \cap V(H_{i+1})|$. Since the order of G_i is strictly decreasing in i, the sequence $\{G_i\}$ is finite and terminates when $G_m = K_1$ for some integer $m \geqslant 1$. This defines the two sequences $G_1, G_2, \dots, G_m$ and $H_2, H_3, \dots, H_m$.

Note that if $|U| = 1$, then we can prove the lemma simply by taking $W = V$. Thus we may assume $|U| \geqslant 2$.

If there exists an $i \geqslant 2$ such that $\frac{1}{4}|U| \leqslant |U \cap V(H_i)| \leqslant \frac{1}{2}|U|$, then pick the smallest such i and set $W = V(H_i)$.

If no such i exists, then we must have $|U \cap V(H_i)| < \frac{1}{4}|U|$ for every $2 \leqslant i \leqslant m$. The case that $|U \cap V(H_i)| > \frac{1}{2}|U|$ can not happen since this would contradict the disjointness of $U \cap V(G_i)$ and $U \cap V(H_i)$. We have $|U \cap V(G_1)| = |U|$ and $|U \cap V(G_m)| \leqslant 1$, so there is some i such that $|U \cap V(G_i)| > \frac{1}{2}|U|$ but $|U \cap V(G_{i+1})| \leqslant \frac{1}{2}|U|$. Now we know that

$$|U \cap V(G_{i+1})| = |U \cap V(G_i)| - |U \cap V(H_{i+1})| > \tfrac{1}{2}n - \tfrac{1}{4}n = \tfrac{1}{4}n,$$

and we set $W = V(G_{i+1})$.

It remains to show that for every $v \in V - W$, W conforms to v. First consider the case $W = V(G_i)$ for some i. By the construction of the sequence $\{G_i\}$, we have the series of inclusions $V(G_1) \supset V(G_2) \supset \cdots \supset V(G_{i-1}) \supset V(G_i)$. Let j be the largest index such that $v \in V(G_j)$. Obviously $1 \leqslant j < i$, and since $V(G_j) = V(G_{j+1}) \uplus V(H_{j+1})$, the choice of j implies that $v \in V(H_{j+1})$. As $G_{j+1} \uplus H_{j+1}$ equals G_j or $\overline{G_j}$, $V(G_{j+1})$ conforms to any vertex in $V(H_{j+1})$. This shows the desired property since $W = V(G_i) \subseteq V(G_{j+1})$.

The remaining case $W = V(H_i)$ for some i can be proved in a similar way, since every $v \in V - W$ is either in $V(G_i)$ or in $V(H_j)$ for some $1 \leqslant j < i$. □

4 The strong Erdős–Hajnal property

In this section we prove Theorem 1. The cases of interval graphs and cographs are simple and will be treated first. The case of chordal graphs is more involved and takes up the majority of the proof.

Proof of Theorem 1 for interval graphs. Let G be an interval graph on n vertices. We first show that it suffices to prove assuming that G has a representation as the intersection graph of a family $F = \{I_j\}_{j=1}^n$ of compact intervals on the real line such that no two intervals share a common endpoint. Let $F = \{I_j\}_{j=1}^n$ be a representation of G. For every nonempty intersection of intervals in F choose a point and let P be the collection of such points. Replacing each I_j by $\operatorname{conv}(I_j \cap P)$ which is a compact interval does not change the intersection graph of F. This yields a subtree representation of G where the ambient tree is the path on P joining all consecutive pairs by size and each I_j is represented as the induced path on $I_j \cap P$. By Lemma 3 we have a subtree representation of G without any pair of subtrees sharing a common leaf. Moreover the maximum degree of ambient tree remains the same, thus the subtrees are subpaths of a long path. Embedding this representation into the real line yields the desired representation by intervals.

For a point $x \in \mathbb{R}$, let $L(x)$ denote the number of intervals I_j whose rightmost endpoint is strictly less than x, and let $R(x)$ denote the number of intervals whose leftmost endpoint is strictly greater than x. Observe that for all sufficiently small x we have $L(x) = 0$ and $R(x) = n$, and for all sufficiently large x we have $L(x) = n$ and $R(x) = 0$. Since the intervals all have distinct endpoints it follows that $L(x)$ is weakly increasing and changes in increments of 1, while $R(x)$ is weakly decreasing and changes in increments of -1. Moreover these changes happen at distinct x-values. Therefore there exists a point x_0 such that $M = L(x_0) = R(x_0)$. If $M \geqslant n/4$ then there is a bi-clique in $\overline{G}$ of size $2\lfloor n/4 \rfloor$.

If $M < n/4$ then G contains a clique of size $n/2$ and consequently G contains a bi-clique of size $2\lfloor n/4 \rfloor$. □

Proof of Theorem 1 for cographs. We apply Lemma 4 to $U = V(G)$. This gives us a subset $W \subseteq V$ with $\frac{1}{4}n \leqslant |W| \leqslant \frac{1}{2}n$ such that W conforms to every $v \in V - W$. Note that $|V - W| \geqslant \frac{1}{2}n$, therefore there exists a subset $X \subset V - W$ with $|X| \geqslant n/4$ such that either $W \subset N(x)$ for every $x \in X$, or $W \cap N(x) = \emptyset$ for every $x \in X$. This implies G or $\overline{G}$ contains a bi-clique of desired size. □

Proof of Theorem 1 for chordal graphs. Let G be a chordal graph on n vertices, and for contradiction, we assume that neither G nor $\overline{G}$ contains a bi-clique of size $2\left\lfloor \frac{2}{9}n \right\rfloor$. By Lemma 3 we may assume that G has a subtree representation as an intersection graph of a family of subtrees $F = \{T_i\}_{i=1}^n$ of an ambient tree T where $\Delta(T) \leqslant 3$ and no two subtrees T_i and T_j share a common leaf. We may assume the maximum degree $\Delta(T) = 3$, otherwise T is a path (or possibly a single vertex) and therefore G is an interval graph which was treated above.

For each $v \in V(T)$, let $F_v \subset F$ be the collection of subtrees that contain the vertex v. If $|F_v| \geqslant \frac{4}{9}n \geqslant 2\left\lfloor \frac{2}{9}n \right\rfloor$ for some v, then the members of F_v form a clique in G and we are done. Therefore assume $|F_v| < \frac{4}{9}n$ for every v.

For any vertex $v \in V(T)$, let $C_1(v), C_2(v), C_3(v)$ be the components of $T - v$, and let $F_i(v) \subset F$ be the family of subtrees contained in $C_i(v)$. (Note that we allow for the possibility that some of the $C_i(v)$ and/or $F_i(v)$ are empty.) Clearly, for every vertex $v \in V(T)$ we have $F = F_v \uplus F_1(v) \uplus F_2(v) \uplus F_3(v)$.

Claim 5. *There exists a degree 3 vertex $v \in V(T)$ such that $\frac{1}{9}n \leqslant |F_i(v)| \leqslant \frac{2}{9}n$ for every $i = 1, 2, 3$.*

Proof of Claim 5. For every vertex v let us label the components $C_i(v)$ such that $|F_1(v)| \geqslant |F_2(v)| \geqslant |F_3(v)|$. We first show if there is no vertex that satisfies the claim, then we have $|F_1(v)| \geqslant \frac{2}{9}n > |F_2(v)| \geqslant |F_3(v)|$ for every vertex $v \in V(T)$. To see why, assume there is a vertex v such that $|F_1(v)| < \frac{2}{9}n$. If $\deg(v) < 3$, then

$$|F| = |F_v| + |F_1(v)| + |F_2(v)| + |F_3(v)| \leqslant \tfrac{4}{9}n + \tfrac{2}{9}n + \tfrac{2}{9}n + 0 < n = |F|,$$

which is a contradiction. Therefore we have $\deg(v) = 3$, and for every $i = 1, 2, 3$, we get

$$|F_i(v)| \geqslant |F_3(v)| = |F| - |F_v| - |F_1(v)| - |F_2(v)| > n - \tfrac{4}{9}n - \tfrac{2}{9}n - \tfrac{2}{9}n = \tfrac{1}{9}n,$$

but then v is a vertex satisfying the claim. Consequently we must have $|F_1(v)| \geqslant \frac{2}{9}n$. If also $|F_2(v)| \geqslant \frac{2}{9}n$, then $F_1(v)$ and $F_2(v)$ correspond to a bi-clique in $\overline{G}$ of the desired size, and therefore $|F_1(v)| \geqslant \frac{2}{9}n > |F_2(v)| \geqslant |F_3(v)|$.

Now consider the following orientation of the edges of T. For any given $v \in V(T)$, let u be the (unique) neighbor of v contained in $C_1(v)$ that is adjacent v, and assign the orientation $\overrightarrow{vu}$. By the observations in the previous paragraph, every vertex has a *unique* outgoing edge. Furthermore we claim that every edge will be assigned a *unique* orientation. This is because if an edge $uv \in E(T)$ is assigned either no orientation or both orientations, then $C_1(v)$ and $C_1(u)$ are disjoint, which implies that $F_1(v)$ and $F_1(u)$ correspond to a bi-clique in $\overline{G}$ of the desired size.

Thus, if there is no vertex satisfying the claim, then we obtain an orientation of the edges of T in which each vertex has a unique outgoing edge, which is impossible. ■

Now fix a vertex $v \in V(T)$ satisfying the condition in Claim 5, and let u_i denote the unique neighbor of v in the component $C_i(v)$. For every $w \in V(T) - \{v\}$, we label the components $C_i(w)$ of $T - w$ (some of which may be empty) such that $C_1(w)$ contains the vertex v, and define $C_{23}(w)$ as the induced subgraph of T on $V(T) - V(C_1(w))$. Define $\Gamma(w)$ to be the collection of subtrees in F which are contained in either $C_2(w)$ or $C_3(w)$, that is, $\Gamma(w) = F_2(w) \cup F_3(w)$, and let $\Gamma^+(w)$ be the collection of subtrees in F contained in $C_{23}(w)$. Note that according to this new notation, we have $\Gamma^+(u_i) = F_i(v)$ for every $i = 1, 2, 3$.

Claim 6. *For every $i = 1, 2, 3$, there is a vertex $w_i \in V(C_i(v))$ that satisfies the following:*

(i) $|\Gamma^+(w_i)| \geqslant \frac{1}{9}n$

(ii) $|F_2(w_i)|, |F_3(w_i)| < \frac{1}{9}n$.

Proof of Claim 6. Fix an $i \in \{1, 2, 3\}$ and set $v_1 := u_i$. By the choice of v we have $\frac{1}{9}n \leqslant |\Gamma^+(v_1)| \leqslant \frac{2}{9}n$, and we are done if both $|F_2(v_1)|$ and $|F_3(v_1)|$ are strictly less than $\frac{1}{9}n$. Otherwise, we may assume that $|F_2(v_1)| > \frac{1}{9}n > |F_3(v_1)|$ and we set v_2 as the unique neighbor of v_1 in the component $C_2(v_1)$. We repeat the same argument to define $v_{j+1} \in C_2(v_j)$ adjacent to v_j whenever $|F_2(v_j)| > \frac{1}{9}n$. Note that $F_2(v_j) = \Gamma^+(v_{j+1}) \supseteq F_2(v_{j+1})$ if both v_j and v_{j+1} are defined and the distance from v to v_j is j. By the definition of $F_2(\cdot)$, when $u \in V(T)$ is a leaf, we have $F_2(u) = \emptyset$ and $|F_2(u)| = 0 < \frac{1}{9}n$. Therefore, the process terminates at some v_m with $|F_2(v_m)| < \frac{1}{9}n \leqslant |F_2(v_{m-1})|$ and $w_i = v_{m-1}$ is the desired vertex. ■

For each $i = 1, 2, 3$, let $G_i \subset F_i(v)$ be the collection of subtrees that intersect the path $P_T(v, w_i)$. Let $H_i \subset F_i(v)$ be the collection of subtrees that are disjoint from the subgraph $C_{23}(w_i)$ and from the path $P_T(v, w_i)$. Equivalently, $H_i = (F_i(v) - G_i) - \Gamma^+(w_i)$. Note that $|H_i| \leqslant \frac{1}{9}n$ since H_i and $\Gamma^+(w_i)$ are disjoint subfamilies of $F_i(v)$ and $\frac{1}{9}n \leqslant |\Gamma^+(w_i)| \leqslant |F_i(v)| \leqslant \frac{2}{9}n$.

Finally, we define some additional subfamilies of F_v. Let $X_\emptyset \subset F_v$ be the collection of subtrees that contains none of w_1, w_2, and w_3. Let $X_i \subset F_v$ be the collection of subtrees containing only w_i but *not* the other w_j's. For every $1 \leqslant i < j \leqslant 3$, let $Y_{ij} \subset F_v$ be the collection of subtrees containing both w_i and w_j. Note that a member of F that contains w_1, w_2, and w_3, belongs to Y_{12}, Y_{13}, and Y_{23}.

The following observations identify certain bi-cliques in G or $\overline{G}$ which allow us to bound the sizes of the various subfamilies we have defined. This will eventually lead us to the existence of a bi-clique of size $2\left\lfloor \frac{2}{9}n \right\rfloor$ in G or $\overline{G}$.

Observation 1. Every member of $X_\emptyset \cup X_1 \cup F_1(v)$ is disjoint from every member of $\Gamma^+(w_2) \cup \Gamma^+(w_3)$, and so the two subfamilies correspond to a bi-clique in $\overline{G}$.

The same obviously holds for the symmetric cases as well. So by the assumption that G or $\overline{G}$ contains no bi-clique of size $2\left\lfloor\frac{2}{9}n\right\rfloor$, and since $\Gamma^+(w_i) \geqslant \frac{1}{9}n$, we must have

$$|X_\emptyset| + |X_i| + |F_i(v)| < \tfrac{2}{9}n,$$

for every $i = 1, 2, 3$. We now have the following inequality.

$$\begin{aligned}|X_\emptyset| + |X_1| + |X_2| + |X_3| &\leqslant 3\,|X_\emptyset| + |X_1| + |X_2| + |X_3| \\ &< \left(\tfrac{2}{9}n - |F_1(v)|\right) + \left(\tfrac{2}{9}n - |F_2(v)|\right) + \left(\tfrac{2}{9}n - |F_3(v)|\right) \\ &= \tfrac{2}{3}n - (|F_1(v)| + |F_2(v)| + |F_3(v)|).\end{aligned}$$

Note that $F_v = X_\emptyset \mathbin{\dot\cup} X_1 \mathbin{\dot\cup} X_2 \mathbin{\dot\cup} X_3 \mathbin{\dot\cup} (Y_{12} \cup Y_{13} \cup Y_{23})$, and so we have

$$\begin{aligned}|Y_{12} \cup Y_{13} \cup Y_{23}| &= |F_v| - (|X_\emptyset| + |X_1| + |X_2| + |X_3|) \\ &= n - (|F_1(v)| + |F_2(v)| + |F_3(v)|) - (|X_\emptyset| + |X_1| + |X_2| + |X_3|) \\ &> \tfrac{1}{3}n.\end{aligned}$$

By double-counting, one of $|Y_{12} \cup Y_{13}|$, $|Y_{12} \cup Y_{23}|$, or $|Y_{13} \cup Y_{23}|$ is strictly greater than $\frac{2}{9}n$, and without loss generality we may assume that $|Y_{12} \cup Y_{13}| > \frac{2}{9}n$. Choose a subset $Y \subseteq Y_{12} \cup Y_{13}$ of size $\lfloor\frac{2}{9}n\rfloor$.

Observation 2. Every member of Y intersects every member of $(F_v - Y) \cup G_1$, and so the two subfamilies correspond to a bi-clique in G.

We may therefore assume that $|F_v - Y| + |G_1| < \frac{2}{9}n$, which gives us

$$\begin{aligned}\tfrac{2}{9}n &> |F_v - Y| + |G_1| = |F_v| - \lfloor\tfrac{2}{9}n\rfloor + |G_1| \\ &\geqslant n - (|F_1(v)| + |F_2(v)| + |F_3(v)|) - \tfrac{2}{9}n + |G_1| \\ &= \tfrac{7}{9}n - (|F_2(v)| + |F_3(v)|) - |F_1(v)| + |G_1|\,.\end{aligned}$$

From this we can conclude that

$$\begin{aligned}|H_1| + |F_2(w_1)| + |F_3(w_1)| &= |F_1(v)| - |G_1| \\ &> \tfrac{5}{9}n - (|F_2(v)| + |F_3(v)|),\end{aligned}$$

and therefore

$$|H_1| + |F_2(w_1)| + |F_3(w_1)| + |F_2(v)| + |F_3(v)| > \tfrac{5}{9}n.$$

Observation 3. Any two members taken from distinct families among H_1, $F_2(w_1)$, $F_3(w_1)$, $F_2(v)$, $F_3(v)$ are pairwise disjoint.

In particular, if we partition $S = \{H_1, F_2(w_1), F_3(w_1), F_2(v), F_3(v)\}$ into two parts $S = S_1 \mathbin{\dot\cup} S_2$, then this corresponds to a bi-clique in $\overline{G}$. Our final goal is to divide S evenly so that $\bigcup S_1 := \bigcup_{G \in S_1} G$ and $\bigcup S_2 := \bigcup_{G \in S_2} G$ each contain at least $\frac{2}{9}n$ subtrees.

Recall $\frac{1}{9}n \leqslant |F_2(v)|, |F_3(v)| \leqslant \frac{2}{9}n$ and the three subfamilies H_1, $F_2(w_1)$, and $F_3(w_1)$ each have size at most $\frac{1}{9}n$. Now we describe how to split S evenly. Start with $S_1 = \{F_2(v)\}$ and $S_2 = \{F_3(v)\}$. Next, take one of the remaining subfamilies in $S - (S_1 \cup S_2)$ and add it to the part S_i which contains the fewest subtrees. Repeat this until the three subfamilies H_1, $F_2(w_1)$, $F_3(w_1)$ have been distributed. Then the resulting S_1 and S_2 satisfy $||\bigcup S_1| - |\bigcup S_2|| \leqslant \frac{1}{9}n$, since $||F_2(v)| - |F_3(v)|| \leqslant \frac{1}{9}n$ and as we distribute the remaining subfamilies, the difference $|\bigcup S_1| - |\bigcup S_2|$ changes by at most $\frac{1}{9}n$ in each step. Because $|\bigcup S_1| + |\bigcup S_2| > \frac{5}{9}n$, we have that $\bigcup S_1$ and $\bigcup S_2$ each contain at least $\frac{2}{9}n$ subtrees, which completes the proof. □

5 The colorful Erdős–Hajnal property

In this section we prove Theorem 2. As in the previous section, the cases of interval graphs and cographs are simple and will be treated first. Finally we deal with the case of the graph class $\mathcal{T}_k$, where we give asymptotically matching upper and lower bounds.

Proof of Theorem 2 for interval graphs. Let G be an interval graph on n vertices. As in the proof of Theorem 1 for interval graphs, we assume that G has a representation as the intersection graph of a family F of n compact intervals on the real line such that no two intervals share a common endpoint.

Our goal is to show that for any partition $F = F_1 \cup F_2$ such that $|F_1| = \lceil \frac{n}{2} \rceil$ and $|F_2| = \lfloor \frac{n}{2} \rfloor$ there are subfamilies of $H_1 \subset F_1$ and $H_2 \subset F_2$, each of size at least $\lfloor \frac{n}{6} \rfloor$, such that either every member of H_1 intersects every member of H_2, or every member of H_1 is disjoint from every member of H_2.

For each $i = 1, 2$, let a_i be the smallest real number such that at least one third of the members of F_i are contained in the half-line $(-\infty, a_i]$:

$$a_i := \min \left\{ a \in \mathbb{R} : |\{ I \in F_i : I \subseteq (-\infty, a] \}| \geqslant \tfrac{1}{3}|F_i| \right\}.$$

Similarly, define b_i as the largest real number such that at least one third of elements of F_i are contained in the half-line $[b_i, \infty)$:

$$b_i := \max \left\{ b \in \mathbb{R} : |\{ I \in F_i : I \subseteq [b, \infty) \}| \geqslant \tfrac{1}{3}|F_i| \right\}.$$

For an interval $J \subset \mathbb{R}$, let $F_i|_J \subset F_i$ denote the collection of intervals of F_i that are contained in J. Note that both $F_i|_{(-\infty,a_i]}$ and $F_i|_{[b_i,\infty)}$ have size exactly $\lceil \frac{1}{3}|F_i| \rceil$, and both $F_i|_{(-\infty,a_i)}$ and $F_i|_{(b_i,\infty)}$ have size exactly $\lceil \frac{1}{3}|F_i| \rceil - 1$.

We divide cases according to the relative order between a_1, a_2, b_1 and b_2.

Case 1. $a_1 < b_2$ or $a_2 < b_1$: If $a_1 < b_2$, then we set $H_1 = F_1|_{(-\infty,a_1]}$ and $H_2 = F_2|_{[b_2,\infty)}$ to obtain the desired subfamilies corresponding to a bi-clique in $\overline{G}$. The case $a_2 < b_1$ is symmetric handled in the same way.

For the rest of proof, assume $b_1 \leqslant a_2$ and $b_2 \leqslant a_1$. Note that these conditions imply that we must have $b_1 \leqslant a_1$ or $b_2 \leqslant a_2$; if $a_1 < b_1$, then we have $b_2 \leqslant a_1 < b_1 \leqslant a_2$. By symmetry we may assume that $b_2 \leqslant a_2$ holds.

Case 2. $a_1 < b_1$ and $b_2 \leqslant a_2$: We have $b_2 \leqslant a_1 < b_1 \leqslant a_2$. In this case we set $H_1 = F_1 - \left(F_1|_{(-\infty,a_1)} \cup F_1|_{(b_1,\infty)} \right)$ and $H_2 = F_2 - \left(F_2|_{(-\infty,a_2)} \cup F_2|_{(b_2,\infty)} \right)$. Then

$|H_i| \geqslant |F_i| - 2\left\lceil \frac{1}{3}|F_i| \right\rceil + 2 \geqslant \left\lceil \frac{1}{3}|F_i| \right\rceil \geqslant \left\lfloor \frac{n}{6} \right\rfloor$. Observe that every interval $I \in H_1$ must *intersect* the interval $[a_1, b_1]$, and that every interval in H_2 must *contain* the interval $[b_2, a_2]$. Since $[a_1, b_1] \subset [b_2, a_2]$ it follows that the families H_1 and H_2 are the desired subfamilies corresponding to a bi-clique in G.

Case 3. $b_1 \leqslant a_1$ and $b_2 \leqslant a_2$: We may assume the intervals $[b_1, a_1]$ and $[b_2, a_2]$ intersect, otherwise it is either $a_1 < b_2$ or $a_2 < b_1$ which is covered by Case 1. Here we set $H_i = F_i - \left(F_i|_{(-\infty, a_i)} \cup F_i|_{(b_i, \infty)} \right)$ for both $i = 1, 2$. As in Case 2, we have $|H_i| \geqslant \left\lceil \frac{1}{3}|F_i| \right\rceil \geqslant \left\lfloor \frac{n}{6} \right\rfloor$ and every member in H_i must contain the interval $[b_i, a_i]$, and consequently the families H_1 and H_2 correspond to a bi-clique in G. □

Proof of Theorem 2 for cographs. Let G be a cograph on the vertex set V with a partition $V = V_1 \cup V_2$ where $|V_1| = \lceil |V|/2 \rceil$ and $|V_2| = \lfloor |V|/2 \rfloor$. Our goal is to find subsets $U_i \subset V_i$ of size at least $|V_i|/4$ for each $i = 1, 2$ such that either every vertex in U_1 is adjacent to all vertices of U_2 or there is no adjacent pair of $u_1 \in U_1, u_2 \in U_2$. This implies the existence of a bi-clique of desired size in G or $\overline{G}$.

Applying Lemma 4 with $U = V_1$, we get a subset $W \subseteq V$ such that $\frac{1}{4}|V_1| \leqslant |V_1 \cap W| \leqslant \frac{1}{2}|V_1|$ and W conforms to every $v \in V - W$. Now define subsets $X(W), Y(W) \subseteq V - W$ by setting

$$\begin{aligned} X(W) &:= \{v \in V - W : W \subset N(v)\}, \\ Y(W) &:= \{v \in V - W : W \cap N(v) = \emptyset\}. \end{aligned}$$

By the choice of W using Lemma 4, we have $X(W) \uplus Y(W) = V - W$. We now distinguish two cases.

Case 1. $|V_2 \cap W| < \frac{1}{2}|V_2|$: Note that $|V_2 \cap (V - W)| \geqslant \frac{1}{2}|V_2|$. Then one of the sets $V_2 \cap X(W)$ or $V_2 \cap Y(W)$ has cardinality at least $\frac{1}{4}|V_2|$. We define U_2 to be the set of larger cardinality, and define $U_1 = V_1 \cap W$.

Case 2. $|V_2 \cap W| \geqslant \frac{1}{2}|V_2|$: Note that $|V_1 \cap (V - W)| \geqslant \frac{1}{2}|V_1|$. Then one of the sets $V_1 \cap X(W)$ or $V_1 \cap Y(W)$ has cardinality at least $\frac{1}{4}|V_1|$. We define U_1 to be the set of larger cardinality, and defined by $U_2 = V_2 \cap W$.

In both cases, we have $|U_i| \geqslant |V_i|/4$ for $i = 1, 2$, and this completes the proof. □

Proof of Theorem 2 for chordal graphs. The case $k = 2$ is covered by results on intervals. We therefore assume that G is a chordal graph on n vertices, with leafage $\ell(G) = k \geqslant 3$ and we are given a partition of the vertices into subsets V_1 and V_2 whose sizes differ by at most one.

By Lemma 3 we may fix a subtree representation of G as an intersection graph of a family F of n subtrees of an ambient tree T where T has k leaves, $\Delta(T) = 3$, and where no two subtrees share a common leaf. The vertex partition $V_1 \cup V_2$ corresponds to a partition $F = F_1 \cup F_2$.

Let $L(T) = \{v_1, v_2, \ldots, v_k\}$ be the set of k leaves of T. For each $i \in [k]$, define $L_i = L(T) - \{v_i\}$ and let $T_i = \mathrm{Tree}_T(L_i)$ which is T with one leaf removed. Let r_i be the closest degree 3 vertex to v_i so that $E(T) - E(T_i) = E(P_T(v_i, r_i))$. Let s_i be the unique neighbor of r_i in $P_T(v_i, r_i)$. Note that $s_i = v_i$ is possible when v_i is adjacent to r_i.

First we show the CEH-property of $\mathcal{T}_k$ with a smaller constant including the case $k = 2$:

Claim 7. *For every $k \geqslant 2$, $\mathcal{T}_k$ satisfies the CEH-property with constant $\varepsilon_c = \frac{1}{3(k-1)}$.*

Proof of Claim 7. We proceed to induction on k. The base case $k = 2$ is covered by results on intervals. Assume that $k \geqslant 3$ and let T, F_1, F_2 and T_i's be as above.

If there is some i such that T_i intersects both at least $\left\lceil \frac{k-2}{k-1} |F_1| \right\rceil$ members of F_1 and $\left\lceil \frac{k-2}{k-1} |F_2| \right\rceil$ members of F_2, then the result follows by applying the inductive assumption on families

$$F_j(T_i) := \{X \cap T_i : X \in F_j\}$$

with $j = 1, 2$.

Therefore, for each $i \in [k]$ the path $P_T(v_i, s_i)$ contains at least $\left\lceil \frac{1}{k-1} |F_1| \right\rceil$ members of F_1 or $\left\lceil \frac{1}{k-1} |F_2| \right\rceil$ members of F_2. Assume that $P_T(v_1, s_1)$ contains at least $\left\lceil \frac{1}{k-1} |F_1| \right\rceil$ members of F_1. If T_1 contains at least $\left\lceil \frac{1}{k-1} |F_2| \right\rceil$ members of F_2, then the two subfamilies correspond a bi-clique of desired size in G.

Otherwise, at least $\left\lceil \frac{k-2}{k-1} |F_2| \right\rceil$ members of F_2 make nonempty intersections with $P_T(v_1, s_1)$. Consider the intersections as subpaths of $P_T(v_1, s_1)$. Together with $\left\lceil \frac{1}{k-1} |F_1| \right\rceil$ members of F_1 contained in the same path, by Theorem 2 (1), there exists a bi-clique of desired size in G or $\overline{G}$. ■

Now we prove the CEH-property for $\mathcal{T}_k$ with the promised constant $\varepsilon_c = \frac{\ln k}{20k}$. Let T and $r_1, r_2, \ldots, r_k$ be as above. We define $\mathrm{Tree}_T(r_1, r_2, \ldots, r_k)$ as the *trunk* of T, denoted by $\mathrm{Trunk}(T)$.

A key observation is that the trunk of a tree is again a tree with fewer leaves:

Observation 4. $\mathrm{Trunk}(T)$ is a tree with at most $\left\lfloor \frac{k}{2} \right\rfloor$ leaves.

Proof of Observation 4. Let ℓ be a leaf of $\mathrm{Trunk}(T)$. We claim that $\ell = r_i$ for at least two indices $i \in [k]$ which is sufficient to prove the observation.

First we show that $\ell = r_i$ for some i. Note that $\mathrm{Trunk}(T) = \mathrm{Tree}_T(r_1, r_2, \ldots, r_k) = \bigcup_{i,j \in [k]} P_T(r_i, r_j)$. Hence $\ell \in P_T(r_i, r_j)$ for some $i, j \in [k]$. However the degree of $\deg(\ell) = 1$ so it cannot be an interior vertex of the path. Thus ℓ is either r_i or r_j.

Next, we show that $\ell = r_i$ for at least two values of i. If $\mathrm{Trunk}(T)$ has no edge then T is a subdivision of star $K_{1,k}$ and we are done. Now consider the case where $\mathrm{Trunk}(T)$ has an edge. Let w be the unique neighbor of ℓ in $\mathrm{Trunk}(T)$. Since $\ell = r_i$, it has another neighbor s_i in T, which is the unique neighbor of ℓ in $P_T(v_i, \ell) = P_T(v_i, r_i)$. Since $\deg(r_i) = 3$ in T, $\ell = r_i$ has the third neighbor u in T other than w and s_i.

Consider a leaf v_j of T such that $P_T(v_j, \ell)$ contains u. Note that it must be $j \neq i$, and we finish the proof by showing $\ell = r_j$. Assume not. Then the path $P_{\mathrm{Trunk}(T)}(\ell, r_j)$ contains w since w is a unique neighbor of ℓ in $\mathrm{Trunk}(T)$. However it implies $w \in P_T(\ell, r_j)$. On the other hand, $P_T(v_j, \ell)$ and $P_T(\ell, r_j)$ are edge disjoint hence $P_T(v_j, \ell) \cup P_T(\ell, r_j)$ is a path from v_j to r_j in T. Since T is tree, it is also the unique path between v_j and r_j. But this contradicts to that r_j is the closest vertex of degree 3 to v_j. ■

Let $R = \text{Trunk}(T) \subset T$. For each $i = 1, 2$, define $(F_i)_R$ to be the collection of subtrees in F_i that intersect R:

$$(F_i)_R := \{X \in F_i : X \cap R \neq \emptyset\}.$$

The complement of $(F_i)_R$, which is the collection of subtrees in F_i that are disjoint from R is denoted by $\overline{(F_i)_R}$:

$$\overline{(F_i)_R} := \{X \in F_i : X \cap R = \emptyset\}.$$

Thus, each member of $\overline{(F_i)_R}$ can be viewed as a subpath of $P_T(v_j, s_j)$ for some $j \in [k]$, where s_j is the unique neighbor of r_j in $P_T(v_j, r_j)$.

We will prove the theorem using induction on $k \geqslant 2$. First, note that the constant $\frac{1}{3(k-1)}$ from Claim 7 is greater than $\varepsilon_c = \frac{\ln k}{20k}$ for small k, say when $2 \leqslant k \leqslant e^{20/3}$. Therefore it is sufficient to prove the theorem when k is large enough, say $k > e^{20/3}$.

We divide into cases according to the size of $(F_1)_R$ and $(F_2)_R$.

First, assume that both $(F_1)_R$ and $(F_2)_R$ are big, say $|(F_i)_R| \geqslant \frac{2}{3}|F_i|$ for each $i = 1, 2$. Define subtree families $F_i(R) := \{X \cap R : X \in (F_i)_R\}$ of R as a multiset. By the induction hypothesis there exists subfamilies $(F_1)'_R \subseteq F_1(R)$ and $(F_2)'_R \subseteq F_2(R)$ corresponding to a bi-clique in G or $\overline{G}$, with size:

$$\left|(F_i)'_R\right| \geqslant \frac{1}{20} \frac{\ln\lfloor k/2 \rfloor}{\lfloor k/2 \rfloor} |F_i(R)| \geqslant \frac{1}{20} \frac{\ln\lfloor k/2 \rfloor}{\lfloor k/2 \rfloor} \frac{2}{3}|F_i| \geqslant \frac{1}{20} \frac{\ln k}{k}|F_i|,$$

where the first inequality comes once we think of R as a tree of at most $\lfloor k/2 \rfloor$ leaves.

Next, consider the case where both $(F_1)_R$ and $(F_2)_R$ are small, meaning $|(F_i)_R| < \frac{2}{3}|F_i|$ for each i. Then we have $|\overline{(F_i)_R}| \geqslant \frac{1}{3}|F_i|$. Recall that each element of $\overline{(F_i)_R}$ is a subpath of some path $P_T(v_j, s_j)$. Therefore we may view each family $\overline{(F_i)_R}$ as a family of intervals contained in the open interval $(i-1, i) \subseteq \mathbb{R}$, and Theorem 2 for interval graphs guarantees the existence of subfamilies $F'_i \subseteq \overline{(F_i)_R}$ of size $|F'_i| \geqslant \frac{1}{3}|\overline{(F_i)_R}| \geqslant \frac{1}{9}|F_i|$.

Finally, consider the last case where only one of $(F_1)_R$ or $(F_2)_R$ is big. Without loss of generality assume $(F_1)_R$ is big so that $|(F_1)_R| \geqslant \frac{2}{3}|F_1|$ and $|\overline{(F_2)_R}| \geqslant \frac{1}{3}|F_2|$. For each $j \in [k]$, define the family $H_j \subset \overline{(F_2)_R}$ of subtrees contained in $P_T(v_j, s_j)$. Note that $\{H_1, H_2, \dots, H_k\}$ form a partition of $\overline{(F_2)_R}$. Assume the size of parts are in decreasing order so that $\sum_{i \leqslant m} |H_i| \geqslant \frac{m}{k}|\overline{(F_2)_R}|$.

We will choose two sequences of subfamilies $(F_1)_R = F^{(0)} \supseteq F^{(1)} \supseteq \dots \supseteq F^{(k)}$ and $H'_1 \subseteq H_1, H'_2 \subseteq H_2, \dots, H'_k \subseteq H_k$ satisfying the following three conditions for every $j \in [k]$:

(i) $\left|F^{(j)}\right| \geqslant \frac{1}{2}\left|F^{(j-1)}\right| \geqslant \frac{1}{2^j}\left|(F_1)_R\right|$.

(ii) $\left|H'_j\right| \geqslant \frac{1}{2}\left|H_j\right|$.

(iii) $F^{(j)}$ and H'_j correspond to a bi-clique in G or $\overline{G}$.

Let us first show how the existence of such subfamilies yields the conclusion of theorem. Note that for every $j \leqslant m \leqslant k$, $F^{(m)} \subseteq F^{(j)}$ and H'_j also correspond to bi-cliques. For a fixed $m \in [k]$ we produce a partition of $[m]$ as follows: Let $I \subset [m]$ denote the set of indices $i \in [m]$ such that every member of $F^{(m)}$ intersects every member of H'_i. Similarly, let $J \subset [m]$ denote the set of indices $j \in [m]$ such that every member of $F^{(m)}$ is disjoint from every member of H'_j. Note that $H'_I = \bigcup_{i\in I} H'_i$ and $H'_J = \bigcup_{j\in J} H'_j$ are disjoint, hence one of them has size at least one half of $|\bigcup_{i\in[m]} H'_i|$. Assume that $|H'_I| \geqslant \frac{1}{2}|\bigcup_{i\in[m]} H'_i|$, where the opposite case $|H'_J| \geqslant \frac{1}{2}\left|\bigcup_{i\in[m]} H'_i\right|$ is handled in the same way.

Now $F^{(m)}$ and H'_I correspond to a bi-clique in G, and their sizes are bounded below by

$$|F^{(m)}| \geqslant \tfrac{1}{2^m}\,|(F_1)_R| \geqslant \tfrac{1}{3\cdot 2^{m-1}}|F_1|$$

and

$$|H'_I| \geqslant \tfrac{1}{2}\left|\bigcup\nolimits_{i\in[m]} H'_i\right| \geqslant \tfrac{1}{4}\left|\bigcup\nolimits_{i\in[m]} H_i\right| \geqslant \tfrac{m}{4k}|\overline{(F_2)_R}| \geqslant \tfrac{m}{12k}|F_2|.$$

We now take $m = \lfloor \log_2 k - \log_2 \ln k\rfloor \geqslant \frac{1}{2}\log_2 k$, which yields $|F^{(m)}| \geqslant \frac{2}{3}\frac{\ln k}{k}|F_1|$ and $|H'_I| \geqslant \frac{1}{20}\frac{\ln k}{k}|F_2|$, which produces the desired bi-clique in G.

We now show how to construct the promised subfamilies $\{F^{(j)}\}_{j\in[k]}$ and $\{H'_j\}_{j\in[k]}$. As stated above, let $F^{(0)} = (F_1)_R$. Fix $j \in [k]$, and assume that $F^{(i)}$ and H'_i's are recursively constructed for every $i < j$. Consider the path $P_T(v_j, s_j)$, and take the vertex a_j on it which is closest to v_j and satisfies that the subpath $P_T(v_j, a_j)$ contains at least half of the members of F_j. Note that there is a unique member $X_j \in F_j$ which is contained in $P_T(v_j, a_j)$ and has a_j as an endpoint.

We distinguish two cases: either at least half of members of $F^{(j-1)}$ contain a_j, or less than half of them contain a_j.

In the former case, let $F^{(j)}$ be those members of $F^{(j-1)}$ containing a_j and let H'_j consist of X_j and the collection of members of H_j which are not fully contained in $P_T(v_j, a_j)$. Note that the two new subfamilies $F^{(j)}$ and H'_j satisfy conditions (i)-(iii) above, and that $F^{(j)}$ and H'_j correspond to a bi-clique in G.

In the latter case, let $F^{(j)}$ be those members of $F^{(j-1)}$ that do not contain a_j and let H'_j be the collection of members of H_j which are fully contained in $P_T(v_j, a_j)$. Again we note that these new subfamilies satisfy all three conditions, and that $F^{(j)}$ and H'_j correspond to a bi-clique in $\overline{G}$. This finishes the inductive step of construction and concludes the proof. □

The asymptotically matching lower bound for the case of chordal graphs in Theorem 2 is a consequence of the following.

Theorem 8. *Let $k \geqslant 17$ be an integer and let T be a tree with k leaves. There exist two subtree families F_1, F_2 of T with the following property: If $H_1 \subset F_1$ and $H_2 \subset F_2$ are such that either every member of H_1 intersects every member of H_2, or every member of H_1 is disjoint from every member of H_2, then $|H_i| \leqslant \frac{2\ln k}{k\ln 2}\,|F_i|$ for some $i = 1, 2$.*

Remark. To prove Theorem 8, we give a construction, where F_1 and F_2 have different sizes. By duplicating vertices, we can construct F_1' and F_2' with equal size so that they satisfy the statement of Theorem 8. Let F_1 and F_2 be any families that satisfy Theorem 8, say $|F_1| = n$ and $|F_2| = m$ for some positive integers n and m. For any integer $t \geqslant 1$, we can take F_1' as the multiset having mt copies of each element of F_1 and F_2' as be the multiset having nt copies of each element of F_2. Then $|F_1'| = |F_2'|$ and definitely F_1' and F_2' also satisfy Theorem 8.

We split the proof of Theorem 8 into two steps. First, we show that every bipartite graph can be "realized" as an intersection graph between two subtree families of some tree. Then we complete the proof by showing the existence of a bipartite graph G without a large bi-clique in G or in $\overline{G}$, which is realized as subtree families of a tree with at most k leaves.

Lemma 9. *Let $G \subseteq K_{k,n}$ be a bipartite graph with $3 \leqslant k \leqslant n$. For every tree T with k leaves, there exist two subtree families F_1 and F_2 of T such that the intersection graph between F_1 and F_2 is isomorphic to G.*

Proof. Let $V = V_1 \cup V_2$ be the vertex partition of $K_{k,n}$ and say $V_1 = \{w_1, w_2, \dots, w_k\}$. Let $v_1, v_2, \dots, v_k$ be the leaves of T and fix a vertex $u \in \mathrm{Trunk}(T)$. Note that u is distinct from all v_i's since $\mathrm{Trunk}(T) = \bigcup_{i,j\in[k]} P_T(r_i, r_j)$ and no leaf of T lies on any $P_T(r_i, r_j)$. For each vertex $x \in V_2$, let G_x be the subtree of T defined by $G_x := \mathrm{Tree}_T(\{u\} \cup \{v_i : w_i x \in G\})$. Define the first subtree family F_1 as

$$F_1 = \{G_x : x \in V_2\}.$$

For each $i \in [m]$, let H_i be the tree on $\{v_i\}$ with no edges. The second family consists of all such "singletons" H_i:

$$F_2 = \{H_i : i \in [m]\}.$$

Two trees G_x and H_i intersect if and only if $w_i x \in G$, showing that the intersection graph between F_1 and F_2 is isomorphic to G. □

From now on, let $k \geqslant 17$ be a fixed integer. For simplicity let $c = c(k) = \frac{2}{\ln 2}\frac{\ln k}{k}$. Note that $c(k) < \frac{1}{2}$ for every $k \geqslant 17$.

Let $n \geqslant k$ be an integer and consider a random graph $G \subseteq K_{k,n}$ formed by independently choosing each edge of $K_{k,n}$ with probability $\frac{1}{2}$. Let $a = \lceil ck \rceil$ and $b = \lceil cn \rceil$ be integers. Let X be the total number of copies of $K_{a,b}$ in G. We show $2E[X] < 1$ for sufficiently large n so that there is some G that $K_{a,b}$ is contained in neither of G nor $\overline{G}$ as a subgraph. Then the subtree representation of G by the subtree families F_1 and F_2 of $T = K_{1,k}$ provided by Lemma 9 satisfies Theorem 8.

By linearity of expectation, we have

$$E[X] = \binom{k}{a}\binom{n}{b} 2^{1-ab}.$$

In order to estimate $E[X]$, we need the following lemma for binomial coefficients.

Lemma 10. *Let $r \in (0,1)$ be a rational number. Let d be a real number such that $\frac{1}{r^r(1-r)^{1-r}} < d$. For every sufficiently large n such that rn is an integer, it holds that $\binom{n}{rn} < d^n$.*

Proof. By Stirling's approximation, we have $\lim_{n\to\infty} \sqrt[n]{n!} = \frac{n}{e}$. Using this formula one can easily show that $\lim_{n\to\infty} \binom{n}{rn}^{1/n} = \frac{1}{r^r(1-r)^{1-r}}$. □

Let $r \in (c, \frac{1}{2})$ be a rational number slightly larger than $c(k)$. For sufficiently large n such that rn is an integer and $rn \geqslant b$, we bound $E[X]$ from above as:

$$E[X] = \binom{k}{a}\binom{n}{b}2^{1-ab} \leqslant 2^k \left(\frac{1}{r^r(1-r)^{1-r}}\right)^n 2^{1-c^2nk} = 2^{k+1}\left(\frac{1}{2^{c^2k}r^r(1-r)^{1-r}}\right)^n$$

Now our goal is to show $2^{c^2k}r^r(1-r)^{1-r} > 1$ so that $E[X] < 1$ for sufficiently large n. Taking logarithm, the inequality is equivalent with:

$$0 < c^2k\ln 2 + r\ln r + (1-r)\ln(1-r).$$

Putting $ck = \frac{2\ln k}{\ln 2}$ yields

$$\begin{aligned}0 &< 2c\ln k + r\ln r + (1-r)\ln(1-r) = r\left(\frac{2c}{r}\ln k + \ln r\right) + (1-r)\ln(1-r)\\ &= r\ln k^{2c/r}r + (1-r)\ln(1-r).\end{aligned}$$

One can easily check that $r\ln\frac{1}{r} + (1-r)\ln(1-r) > 0$ for every $r \in (0, \frac{1}{2})$. Thus it is enough to show that $k^{2c/r}r \geqslant \frac{1}{r}$ or equivalently $k^{2c/r}r^2 \geqslant 1$ for our choices of k, c and r. However, this easily follows from the continuity of an auxiliary function $f(x) = k^{2c/x}x^2$ at $x = c$ and the fact that $f(c) = k^2c^2 = \left(\frac{2\ln k}{\ln 2}\right)^2 > 1$.

6 Concluding remarks

Recall that intersection graphs of planar convex sets have the SEH-property. On the other hand, they do not enjoy the CEH-property. This can be easily seen by Theorem 8 and the following Lemma [17].

Lemma 11. *Let T be a tree and F be a family of subtrees of T. There is a family C of convex sets in $\mathbb{R}^2$ such that C and F have isomorphic nerve complexes.*

It is natural to ask whether intersection graphs of convex sets in higher dimensions satisfy SEH- or CEH-properties. However, it is already pointless to consider those properties in dimension three since every graph can be realized as the intersection graph of some convex sets in $\mathbb{R}^3$ [23]. Another direction is to consider Erdős-Hajnal type properties in the class of *intersection hypergraphs*. We conjecture that intersection 3-uniform hypergraphs of planar convex sets satisfy the following generalization of SEH-property.

Conjecture 12. There exists a constant $c > 0$ for which the following holds. For every finite family F of convex sets in $\mathbb{R}^2$ (or in $\mathbb{R}^3$), we can find pairwise disjoint subfamilies F_1, F_2, $F_3 \subseteq F$ of size $|F_i| \geqslant c\,|F|$ for every $i = 1, 2, 3$ such that either every rainbow triple of F_1, F_2, F_3 intersect or every such rainbow triple do not intersect.

Acknowledgements

M. Cho and A. F. Holmsen were supported by the National Research Foundation of Korea (NRF) grants funded by the Ministry of Science and ICT (NRF-2020R1F1A1A01048490). M. Cho was also supported by the Institute for Basic Science (IBS-R029-C4). A. F. Holmsen and J. Kim were supported by Institute for Basic Science (IBS-R029-C1). M. Kim was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2022R1F1A1063424) and by GIST Research Project grant funded by the GIST in 2023. This work was conducted when the first author was a Ph.D student at KAIST and included in his dissertation [4].

The authors thank an anonymous referee for helpful comments that improved the readability of the manuscript.

References

[1] N. Alon, J. Pach, R. Pinchasi, R. Radoičić, and M. Sharir. Crossing patterns of semi-algebraic sets. *J. Combin. Theory Ser. A*, 111(2):310–326, 2005.

[2] M. Axenovich, C. Tompkins, and L. Weber. Large homogeneous subgraphs in bipartite graphs with forbidden induced subgraphs. *J. Graph Theory*, 97(1):34–46, 2021.

[3] M. Bonamy, N. Bousquet and S. Thomassé. The Erdős-Hajnal conjecture for long holes and antiholes. *SIAM J. Discrete Math.*, 30(2):1159–1164, 2016.

[4] M. Cho. Extremal results on hypergraphs arising from geometric objects. Doctoral dissertation, Korea Adv. Inst. Science. Techn., Daejeon, Republic of Korea, 2023.

[5] M. Chudnovsky. The Erdős-Hajnal conjecture—a survey. *J. Graph Theory*, 75(2): 178–190, 2014.

[6] M. Chudnovsky, N. Robertson, P. Seymour, and R. Thomas. The strong perfect graph theorem. *Ann. of Math. (2)*, 164(1):51–229, 2006.

[7] M. Chudnovsky, A. Scott, P. Seymour, and S. Spirkl. Pure pairs. I. Trees and linear anticomplete pairs. *Adv. Math.*, 375:107396, 2020.

[8] M. Chudnovsky, A. Scott, P. Seymour, and S. Spirkl. Pure pairs. II. Excluding all subdivisions of a graph. *Combinatorica*, 41(3):379–405, 2021.

[9] P. Erdős. Some remarks on the theory of graphs. *Bull. Amer. Math. Soc.*, 53:292–294, 1947.

[10] P. Erdős and A. Hajnal. Ramsey-type theorems. *Discrete Appl. Math.* 25:37–52, 1989.

[11] P. Erdős, A. Hajnal, and J. Pach. A Ramsey-type theorem for bipartite graphs. *Geombinatorics*, 10(2):64–68, 2000.

[12] J. Fox. A bipartite analogue of Dilworth's theorem. *Order*, 23:197–209, 2006.

[13] J. Fox, M. Gromov, V. Lafforgue, A. Naor, and J. Pach. Overlap properties of geometric expanders. *J. Reine Angew. Math.*, 671:49–83, 2012.

[14] J. Fox and J. Pach. Erdős-Hajnal-type results on intersection patterns of geometric objects. In *Horizons of combinatorics*, vol. 17 of *Bolyai Soc. Math. Stud.*, 79–103. Springer, 2008.

[15] J. Fox, J. Pach, and C. D. Tóth. Turán-type results for partial orders and intersection graphs of convex sets. *Israel J. Math*, 178:29–50, 2010.

[16] F. Gavril. The intersection graphs of subtrees in trees are exactly the chordal graphs. *J. Combin. Theory Ser. B*, 16:47–56, 1974.

[17] M. Kim. Intersection patterns of subtree families and colorful fractional Helly theorems. Master's thesis, Korea Adv. Inst. Science. Techn., Daejeon, Republic of Korea, 2014.

[18] D. Larman, J. Matoušek, J. Pach, and J. Törőcsik. A Ramsey-type result for convex sets. *Bull. London Math. Soc.*, 26(2):132–136, 1994.

[19] C. G. Lekkerkerker and J. Ch. Boland. Representation of a finite graph by a set of intervals on the real line. *Fund. Math.*, 51:45–64, 1962/63.

[20] J. Pach and J. Solymosi. Crossing patterns of segments. *J. Combin. Theory Ser. A*, 96(2):316–325, 2001.

[21] J. Pach and J. Solymosi. Structure theorems for systems of segments. In *Discrete and Computational Geometry*, vol. 2098 of *Lecture Notes in Comput. Sci.*, 308–317. Springer, 2009.

[22] A. Scott, P. Seymour, and S. Spirkl. Pure pairs. IV. Trees in bipartite graphs. *J. Combin. Theory Ser. B*, 161:120–146, 2023.

[23] H. Tietze. Über das Problem der Nachbargebiete im Raum. *Monatsh. Math. Phys.*, 16(1):211–216, 1905.