k-Leaf Powers Cannot be Characterized by a Finite Set of Forbidden Induced Subgraphs for $k \geq 5$

Max Dupré la Tour* Manuel Lafond† Ndiamé Ndiaye‡ Adrian Vetta§

Abstract

A graph $G = (V, E)$ is a k-leaf power if there is a tree T whose leaves are the vertices of G with the property that a pair of leaves u and v induce an edge in G if and only if they are distance at most k apart in T. For $k \leq 4$, it is known that there exists a finite set $\mathscr{F}_k$ of graphs such that the class $\mathcal{L}(k)$ of k-leaf power graphs is characterized as the set of strongly chordal graphs that do not contain any graph in $\mathscr{F}_k$ as an induced subgraph. We prove no such characterization holds for $k \geq 5$. That is, for any $k \geq 5$, there is no finite set $\mathscr{F}_k$ of graphs such that $\mathcal{L}(k)$ is equivalent to the set of strongly chordal graphs that do not contain as an induced subgraph any graph in $\mathscr{F}_k$.

1 Introduction

A fundamental question in graph theory concerns whether or not a graph $G = (V, E)$ can be represented (or approximated) by a simpler graph, for instance a tree T, while preserving the desired information from the original graph. The pairwise distances of G often need to be summarized into sparser structures, with notable examples including *graph spanners* [10, 1, 14, 20] and *distance emulators* [27, 8, 28] which respectively ask for a subgraph of G or for another graph that approximates the distances of G. If the distance information to preserve only concerns "close together" versus "far apart" then this can take the following form: given a graph G and an integer k, does there exists a tree T whose leaves are the vertices of G, such that distinct vertices u and v are adjacent in G *if and only if* the distance $d_T(u, v)$ from u to v in T is at most k? If the answer is affirmative then G is dubbed a *k-leaf power* of T (and T is dubbed a *k-leaf root* of G).

The study of k-leaf powers and roots were instigated by Nishimura, Ragde and Thilikos [25]. On the applied side, these graphs are of significant interest in the field of computational biology with respect to *phylogenetic trees*, which aim to explain the distance relationships observed on available data between species, genes, or other types of taxa. Indeed, k-leaf powers can be used to represent and explain pairs of genes that underwent a bounded number of evolutionary events in their evolution [23, 15], or that have conserved closely related biological functions during evolution [19]. On the theory side, despite their simplicity, several fundamental graph theoretic problems concerning k-leaf powers remain open. The purpose of this research is to resolve one such long-standing open problem. Specifically, we prove that the class $\mathcal{L}(k)$ of k-leaf power graphs cannot be characterized via a finite set of forbidden induced subgraphs for $k \geq 5$. In contrast, for $k \leq 4$ such finite characterizations were previously shown to exist [11, 5, 3].

*McGill University: maxduprelatour@gmail.com
†Université de Sherbrooke: manuel.lafond@usherbrooke.ca
‡McGill University. ndiame.ndiaye@mail.mcgill.ca
§McGill University. adrian.vetta@mcgill.ca

1.1 Background

Let $\mathcal{L}(k)$ denote the class of all k-leaf power graphs, for $k \geq 2$. The class of all leaf power graphs is then denoted by $\mathcal{L} = \bigcup_k \mathcal{L}(k)$. The literature on leaf power graphs has primarily focused on two major themes. One, obtaining graphical characterizations for both the class $\mathcal{L}$ and the classes $\mathcal{L}(k)$, for fixed values of k. Two, designing efficient algorithms to recognize graphs that belong to these classes.

Let's begin with the former theme. Here important roles are played by chordal and strongly chordal graphs. A graph is *chordal* if every cycle of length four or more has a *chord*, an edge connecting two non-consecutive vertices of cycle. A graph is *strongly chordal* if it is chordal **and** all its even cycles of length 6 or more have an *odd chord*, a chord connecting two vertices an odd distance apart along the cycle. Now, it is known that every graph in $\mathcal{L}$ and $\mathcal{L}(k)$ is *strongly chordal*.[1] To see this, first note that a leaf power graph is an induced subgraph of a power of a tree. Second, note that trees are strongly chordal, and taking powers and induced subgraphs both preserve this property [26]). However, the reciprocal is not true: there exist strongly chordal graphs that are not leaf powers. The first such example was discovered by Brandstädt et al. [4]. Subsequently, six additional examples were identified by Nevries and Rosenke [24] who conjectured that any strongly chordal graph not containing any of these seven graphs as an induced subgraph is a leaf power. However, a weaker version of this conjecture, that there are only a finite number (rather than seven) of obstructions was was disproved by Lafond [17]. The author constructed an infinite family of *minimal* strongly chordal graphs that are not leaf powers (i.e., removing any vertex results in a leaf power).

For fixed k, the conjecture that $\mathcal{L}(k)$ may be characterized by a finite set of obstructions remained open. Indeed, for $k \leq 4$, the classes $\mathcal{L}(k)$ can be characterized as chordal graphs that do not contain any graph from $\mathscr{F}_k$ as induced subgraphs, where $\mathscr{F}_k$ is a finite set. Specifically:

- $k = 2$: A graph is in $\mathcal{L}(2)$ *if and only if* it is a disjoint union of cliques. That is, $\mathcal{L}(2)$ is precisely the set of graphs that forbid P_3, the chordless path with three vertices, as an induced subgraph. Thus $|\mathscr{F}_2| = 1$.

- $k = 3$: Dom et al. [11] gave the first characterization of $\mathcal{L}(3)$: a graph is in $\mathcal{L}(3)$ *if and only if* it is chordal and does not contain a bull, a dart or a gem as induced subgraph. Thus $|\mathscr{F}_3| = 3$. Other characterizations of $\mathcal{L}(3)$ were later discovered [5]

- $k = 4$: Brandstädt, Bang Le and Sritharan [3] proved that a graph is in $\mathcal{L}(4)$ *if and only if* it is chordal and does not contain as induced subgraph one of a finite set $\mathscr{F}_4$ of graphs[2].

Given this, the aforementioned conjecture naturally arose: for every k, is the class $\mathcal{L}(k)$ equivalent to the set of chordal graphs that do not contain as induced subgraphs any of a finite set $\mathscr{F}_k$ of graphs?

For $k = 5$, Brandstädt, Bang Le and Rautenbach [6] proved this is true for a special subclass of $\mathcal{L}(5)$. Specifically, the *distance hereditary*[3] 5-leaf power graphs are chordal graphs that do contain a set of 34 graphs as induced subgraphs. However, for the general case, they state

[1] In particular they do not contain, as induced subgraphs, chordless cycles of length greater than three, nor *sun graphs*.

[2] Formally, they show that the set of basic 4-leaf power, where no two leaves of the leaf root share a parent, can be characterized by chordal graphs which do not have one of 8 graphs as induced subgraphs. $\mathscr{F}_4$ can be deduced from this set.

[3] A graph G is distance hereditary if for all pairs of vertices (u, v) in all subgraphs of G either the distance is the same as in G or there is no path from u to v.

"*For $k \geq 5$, no characterization of k-leaf powers is known despite considerable effort. Even the characterization of 5-leaf powers appears to be a major open problem.*" [6]

The contribution of this paper is to disprove the conjecture: for all $k \geq 5$, it is impossible to characterize the set of k-leaf powers as the set of chordal graphs which are $\mathscr{F}_k$-free for $|\mathscr{F}_k|$ finite. In fact, we show that even for the more restrictive class of strongly chordal graphs it is impossible to characterize the set of k-leaf powers as the set of strongly chordal graphs which are $\mathscr{F}_k$-free for finite $|\mathscr{F}_k|$.

Let us conclude this section by discussing the second major theme in this area, namely, efficient recognition algorithms. The computational complexity of deciding whether or not a graph is in $\mathcal{L}$ is wide open. We remark, however, that some graphs in $\mathcal{L}$ have a *leaf rank* that is exponential in the number of their vertices, where the leaf rank of a graph G is the minimum k such that $G \in \mathcal{L}(k)$ [16]. The question of computing the leaf rank of subclasses of $\mathcal{L}$ in polynomial time was recently initiated in [21].

For fixed values of k, though, progress has been made in designing polynomial-time algorithms for the $\mathcal{L}(k)$ recognition problem. For $\mathcal{L}(2), \mathcal{L}(3)$ and $\mathcal{L}(4)$, this immediately follows from the above characterizations because $\mathscr{F}_2, \mathscr{F}_3$ and $\mathscr{F}_4$ are finite. In fact, all these three recognition problems can be solved in linear time; see [5, 3]. Using a dynamic programming approach, Chang and Ko [9] described a linear-time algorithm for the $\mathcal{L}(5)$ recognition problem, and Ducoffe [12] proposed a polynomial-time algorithm for the $\mathcal{L}(6)$ recognition problem. Recently, Lafond [18] designed a polynomial-time algorithm for the $\mathcal{L}(k)$ recognition problem, for any constant $k \geq 2$. The algorithm is theoretically efficient albeit completely impractical: the polynomial's exponent depends only on k but is $\Omega(k \uparrow\uparrow k)$, that is, a tower of exponents $k^{k^{\cdot^{\cdot^{\cdot^{k}}}}}$ of height k. We remark that the algorithm does not rely on specific characterizations of k-leaf power graphs aside from the fact that they are chordal. It appears difficult to significantly improve its running time without a better understanding of the graph theoretical structure of graphs in $\mathcal{L}(k)$. Our work assists in this regard by improving our knowledge of k-leaf powers in terms of forbidden induced subgraphs.

1.2 Overview and Results

We now present an overview of the paper and our results. In Section 2 we present our main theorem:

Theorem 1.1. *For $k \geq 5$, the set of k-leaf powers cannot be characterized as the set of strongly chordal graphs which are $\mathscr{F}_k$-free, where $\mathscr{F}_k$ is a finite set of graphs.*

There we discuss the three types of gadgets we need. These gadgets can be combined to form an infinite family of pairwise incomparable graphs which are not k-leaf powers. We prove the main theorem modulo three critical lemmas on the gadgets. In Section 3 we present proofs of the three critical lemmas. Finally, in Section 4 we show how to modify our proof to derive a similar theorem for *linear k-leaf powers*:

Theorem 1.2. *For $k \geq 5$, the set of linear k-leaf powers cannot be characterized as the set of strongly chordal graphs which are $\mathscr{F}_k$-free where $\mathscr{F}_k$ is a finite set of graphs.*

Here, a linear k-leaf power is a graph that has a k-leaf root which is the subdivision of a *caterpillar*. We remark that that the class of linear leaf powers can be recognised in linear time, as shown by Bergougnoux et al. [2].

2 The Proof Modulo Three Critical Lemmas

In this section we prove our main theorem, Theorem 1.1, assuming the validity of three critical lemmas. The proofs of these lemmas form the main technical contribution of the paper and are deferred to Section 3.

2.1 Preliminaries

Before presenting the proof of Theorem 1.1, we present necessary definitions and notations. Let's start with a formal definition of k-leaf powers. Let $G = (V, E)$ be a simple finite graph, and $k \geq 2$ be an integer. G is called a *k-leaf power* if there exists a tree T, known as a *k-leaf root of G*, with the following properties:

- V is the set of leaves of T.
- For any pair of vertices $u, v \in V$, there is an edge $uv \in E$ *if and only if* the $d_T(u, v) \leq k$.

Here d_T is the distance metric induced by the tree T when two adjacent vertices are a distance of 1 apart. To simplify the notation, we will use d instead of d_T when the context is clear. We will use the notation dist_G to denote distance within the graph G and thus distinguish it from the distance d_T induced by a leaf root T.

2.2 The Proof of the Main Theorem

To prove Theorem 1.1, for any $k \geq 5$, we will construct a collection of arbitrarily large strongly chordal graphs that are *minimal non k-leaf powers*. Specifically, these graphs have the property that any "strict" induced subgraph is a k-leaf power.

To accomplish this goal, we fix $k \geq 5$. We then begin by designing a graph H_n, for all $n \geq 0$, built using three gadget graphs joined in series. First will be the *top gadget* and last the *bottom gadget*. In between will be exactly n copies of the *interior gadget*. We denote these gadget graphs by $\mathrm{Top}, \mathrm{Bot}$ and I, respectively. These gadget graphs will satisfy a set of critical properties. To formalize these properties we require the following definition. Given a graph $G = (V, E)$ and T a k-leaf root of G. For $v \in V$, let $m_T(v) = \min_{u \in V \setminus \{v\}} d_T(u, v)$. That is, $m_T(v)$ is the shortest distance in the tree T from the leaf v to any other leaf u.

The aforementioned properties of $\mathrm{Top}, \mathrm{Bot}$ and I are stated in the subsequent three critical lemmas.

Lemma 2.1. *For all $k \geq 4$, there exists a gadget graph Top that contains a vertex $t \in V(Top)$ such that:*

1. *For any k-leaf root T of Top, $m_T(t) = 3$.*
2. *There exists a k-leaf root T_{Top} of Top.*

Lemma 2.2. *For all $k \geq 4$, there exists a gadget graph Bot that contains a vertex $b \in V(Bot)$ such that:*

1. *For any k-leaf root T of Bot, $m_T(b) \leq k - 1$.*
2. *There exists a k-leaf root T_{Bot} such that $m_{T_{Bot}}(b) = k - 1$*

Lemma 2.3. *For all $k \geq 5$, there exists a gadget graph I that contains two distinct vertices $t_I, b_I \in V(I)$ such that:*

1. *For all k-leaf roots T of I, $m_T(t_I) \geq k \implies m_T(b_I) = 3$.*
2. *There exists a k-leaf root T_I of I such that $m_{T_I}(t_I) = k$ and $m_{T_I}(b_I) = 3$.*
3. *There exists a k-leaf root R_I of I such that $m_{R_I}(t_I) = k - 1$ and $m_{R_I}(b_I) = 4$.*

We will prove the existence of gadget graphs $\mathrm{Top}, \mathrm{Bot}$ and I required to verify the three lemmas in Section 3. For the rest of the section, we will assume these lemmas and use them to prove our main result.

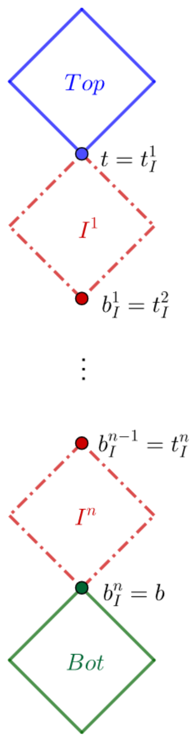


Figure 1: The construction of H_n

First, as alluded to above, we then combine our three gadgets to create an intermediary graph H_n. In particular, H_n is the graph obtained by connecting in series one copy of Top, then n copies of I: $I^1, \ldots, I^n$ and finally one copy of Bot. This construction is illustrated in Figure 1. The vertices t_I and b_I, mentioned in Lemma 2.3, of the j-th copy I^j are denoted t_I^j and b_I^j, respectively. Notice that to connect the gadgets within H_n, we identify the vertices described in Lemmas 2.1, 2.2, and 2.3 as follows. We identify t with t_I^1, for all $j < n$, b_I^j with t_I^{j+1}, and finally, b_I^n with b. As a special case when $n = 0$, the graph H_0 is obtained by taking Top and Bot and identifying t with b.

In order to prove Theorem 1.1 we must study the structure of H_n. We denote by $H_n - \mathrm{Top}$ (resp. $H_n - \mathrm{Bot}$) the graph obtained from H_n by deleting the top gadget Top (resp. the bottom gadget Bot), i.e. removing all vertices of Top (resp. Bot) except for the common vertex $t = t_I^1$ (resp. $b = b_I^n$). Of importance is the next lemma.

Lemma 2.4. *The graph H_n has the following properties:*

1. $\mathrm{dist}_{H_n}(b,t) \geq n$.

2. H_n *is strongly chordal.*

3. $H_n - Top$ *and* $H_n - Bot$ *are both* k*-leaf powers.*

4. H_n *is not a* k*-leaf power.*

Proof. In order to prove 1, the distance between t_I and b_I in I is at least 1 because the two vertices are distinct. Hence $\mathrm{dist}_{H_n}(b,t) \geq n$, because there are n copies of the interior gadget I.

For the proof of 2, the three gadgets are k-leaf powers and, therefore, are strongly chordal. The construction of H_n does not introduce additional cycles; thus, H_n remains strongly chordal.

To prove 3, we combine the leaf-root properties provided by the gadgets T_{Top}, T_{Bot}, T_I, and R_I, as illustrated in Figure 2.

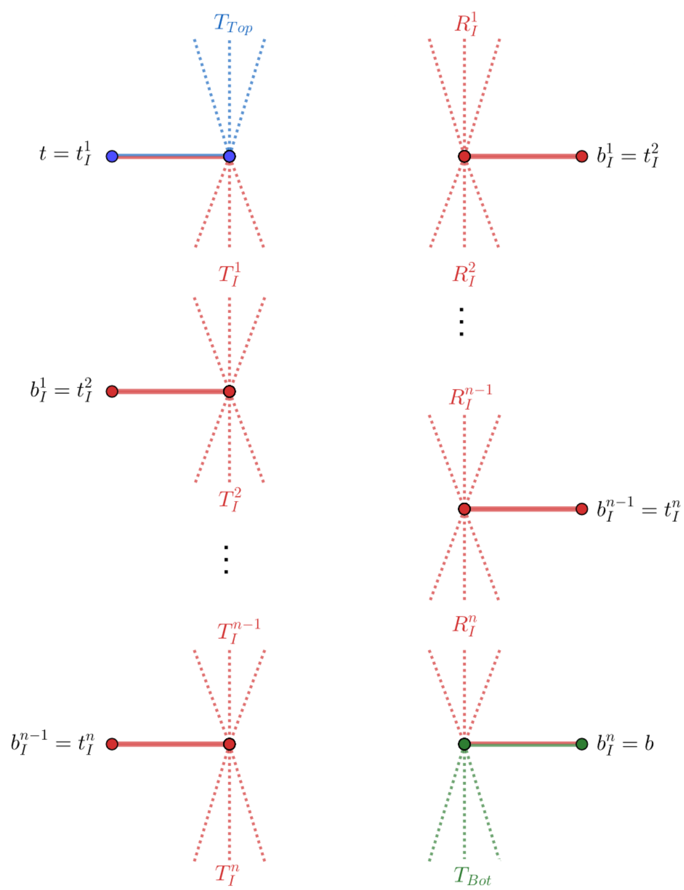


Figure 2: The k-leaf roots of $H_n - \text{Bot}$ and $H_n - \text{Top}$

The left tree in Figure 2 is obtained by merging T_{Top} with n copies of T_I, denoted as $T_I^1, \ldots, T_I^n$. We identify t with t_I^1, and we identify the of t in T_{Top} with t_I^1 and with the parent of t_I^1 in T_I^1. Similarly, for all $j \leq n-1$, we identify b_I^j and its parent with t_I^{j+1} and its parent, respectively. We now prove that the resulting tree is a k-leaf root of $H_n - \text{Top}$. Top and each copy of the interior gadget I are the k-leaf power of the corresponding subtree: T_{Top} for Top and T_I^j for the j-th copy of I. It remains to show that we do not introduce any additional, unwanted edges. If v_1 is a leaf of T_{Top} different from t, and v_2 is a leaf of T_I^1 different from t_I^1 then, using Lemma 2.1, we conclude $d(v_1, t) \geq m_{T_{\text{Top}}}(t) = 3$. Furthermore, using the second point of Lemma 2.3, we conclude $d(v_2, t_I^1) \geq m_{T_I}(t_I) = k$. Therefore, $d(v_1, v_2) \geq d(v_1, t) + d(v_2, t_I^1) - 2 \geq k + 3 - 2 = k + 1 > k$, and there is no edge between v_1 and v_2 in the k-th power of the tree. Similarly, if v_1 is a leaf of T_I^j different from b_I^j for some $j \leq n-1$ and v_2 is a leaf of T_I^{j+1} different from t_I^{j+1}, we have $d(v_1, b_I^j) \geq 3$ and $d(v_2, t_I^{j+1}) \geq k$. Thus, $d(v_1, v_2) \geq d(v_1, b_I^j) + d(v_2, t_I^{j+1}) - 2 \geq k + 3 - 2 = k + 1 > k$, and there is no edge between v_1 and v_2 in the k-th power of the tree.

The right tree in Figure 2 is formed by merging n copies of R_I, denoted $R_I^1, \ldots, R_I^n$, with T_{Bot}. For all $j \leq n - 1$, we identify b_I^j with t_I^{j+1}, and we identify the parent of b_I^j with the parent of t_I^{j+1}. Finally, we identify b_I^n and its parent in R_I^n with b and its parent in T_{Bot}, respectively. Similar to the left tree, we must prove that no additional, unwanted edges are created. If v_1 is a leaf of R_I^j different from b_I^j for some $j \leq n - 1$ and v_2 is a leaf of R_I^{j+1} different from t_I^{j+1} then, using Lemma 2.3, we conclude $d(v_1, b_I^j) \geq m_{R_I}(b_I) = 4$ and $d(v_2, t_I^{j+1}) \geq m_{R_I}(t_I) = k - 1$. Thus, $d(v_1, v_2) \geq d(v_1, b_I^j) + d(v_2, t_I^{j+1}) - 2 \geq k - 1 + 4 - 2 = k + 1 > k$, and there is no edge between v_1 and v_2 in the k-th power of the tree. Similarly if v_1 is a leaf of R_I^n different from b_I^n and v_2 is a leaf of T_{Bot} different from b then we have $d(v_1, b_I^n) \geq 4$ and, using Lemma 2.2, $d(v_2, b) \geq k - 1$. Therefore $d(v_1, v_2) \geq d(v_1, b_I^n) + d(v_2, b) - 2 \geq k + 1 > k$ and v_1,v_2 are not connected in the k-th power of the tree. This completes the proof of 3, the third point of the lemma.

It remains to prove 4, the final point of the lemma. We start by proving by induction that for any integer n, in any k-leaf root T of $H_n - \text{Bot}$, there is a leaf at distance 3 of b in T. When $n = 0$, there are no gadgets I between Bot and Top, so $b = t$ and the property holds by property 1 of Lemma 2.1. Turning to the induction step, assume that the property holds for some $n \geq 0$ and consider a k-leaf root T of $H_{n+1} - \text{Bot}$. Since $H_n - \text{Bot}$ is an induced subgraph of $H_{n+1} - \text{Bot}$, some induced subgraph of T is a k-leaf root of $H_n - \text{Bot}$. By the induction hypothesis, there exists a vertex v_1 in $H_n - \text{Bot}$ at a distance exactly 3 from $b_I^n = t_I^{n+1}$ in an induced subgraph of T. Adding vertices will not alter the distance, so $d(v_1, b_I^n) = 3$ in T. We claim that every vertex of I^{n+1} is at distance at least k from $b_I^n = t_I^{n+1}$ in T (except b_I^n itself). Assume, by contradiction, that there exists a vertex v_2 in the last copy I^{n+1}, distinct from b_I^n such that that $d(v_2, b_I^n) \leq k - 1$. This assumption would imply that $d(v_1, v_2) \leq d(v_1, b_I^n) + d(v_2, b_I^n) - 2 \leq 3 + (k - 1) - 2 = k$, meaning that v_1 and v_2 are connected in the k-th power of T, contradicting the fact that T is a k-leaf root of $H_{n+1} - \text{Bot}$. Therefore, all vertices in I^{n+1}, distinct from b_I^n, are at a distance of at least k from b_I^n in T. We can now apply property 1 of Lemma 2.3, which concludes the induction.

Now assume by contradiction that there exists a k-leaf root T of H_n. T induces a k-leaf root of $H_n - \text{Bot}$. In particular, there exists a vertex v_1 in $H_n - \text{Bot}$ such that $d(b_I^n, v_1) = 3$. Moreover, property 1 of Lemma 2.2 implies that there exists a vertex v_2 in Bot, distinct from b, such that $d(v_2, b) \leq k - 1$. Combining these equations, we get $d(v_1, v_2) \leq k - 1 + 3 - 2 = k$, contradicting the fact that there is no edge between v_1 and v_2. This proves that H_n is not a k-leaf power, as desired. □

As stated H_n is an intermediate graph in proving the main result. We will actually show the existence of an induced subgraph $G_{k,n}$ of H_n that is strongly chordal and minimal non k-leaf power. More precisely, we have the following lemma.

Lemma 2.5. *For all $k \geq 5$ and $n \geq 0$, there exists a graph $G_{k,n}$ such that:*

1. *$G_{k,n}$ is strongly chordal and contains at least n vertices.*
2. *$G_{k,n}$ is not a k-leaf power.*
3. *If $G \neq G_{k,n}$ is an induced subgraph of $G_{k,n}$ then G is a k-leaf power.*

Proof. Let $G_{k,n}$ be a minimal induced subgraph of H_n that is not a k-leaf power. It is strongly chordal because, by Lemma 2.4, H_n is strongly chordal. By definition, $G_{k,n}$ is not in $\mathcal{L}(k)$, but every induced subgraph G of $G_{k,n}$ is in $\mathcal{L}(k)$ if $G \neq G_{k,n}$. It remains to prove that $|G_{k,n}| \geq n$. By Lemma 2.4, both H_n − Bot and H_n − Top are in $\mathcal{L}(k)$. Therefore, $G_{k,n}$ must contain a vertex from Top and a vertex from Bot. Moreover, $G_{k,n}$ is connected by its minimality, since the disjoint union of k-leaf powers is a k-leaf power. Hence, $G_{k,n}$ must contain a path from Top to Bot. By Lemma 2.4, $\text{dist}_{H_n}(b,t) \geq n$, and therefore $|G_{k,n}| \geq n$. □

Our main result follows directly from Lemma 2.5

Proof of Theorem 1.1. If $\mathcal{L}(k)$ is the set of strongly chordal graphs which are $\mathscr{F}_k$-free, then $\mathscr{F}_k$ must contain $G_{k,n}$ for all n because it is strongly chordal and a minimal non k-leaf power by Lemma 2.5. But the set $\{G_{k,n}, n \geq 0\}$ is infinite because $G_{k,n}$ has more than n vertices. Therefore $\mathscr{F}_k$ must be infinite, which concludes the proof of Theorem 1.1. □

3 The Gadget Graphs

So we have proven the main theorem modulo the three critical lemmas. Recall to prove these lemmas we must construct the appropriate three gadget graphs, namely Top, Bot and I. We present these constructions and give formal proofs of Lemmas 2.1, 2.2 and 2.3 in this section.

We start with a general observation. In a tree T, if a pair of leaves are a distance of 2 apart, they share the same parent. Consequently, their distances to every other leaf are identical. A consequence of this is that if two vertices are not connected by an edge, or if they have different neighborhoods in a graph, they must be at a distance of at least 3 in any leaf root of that graph. In the gadgets we describe in this section, any two vertices connected by an edge always have distinct neighborhoods. Therefore, we assume that for any pair of vertices x and y and any leaf root T, we have $d_T(x,y) \geq 3$.

3.1 The Top Gadget

We begin by showing the existence of an appropriate top gadget, Top.

Lemma 2.1. *For all $k \geq 4$, there exists a gadget graph Top that contains a vertex $t \in V(Top)$ such that:*

1. *For any k-leaf root T of Top, $m_T(t) = 3$.*
2. *There exists a k-leaf root T_{Top} of Top.*

Proof. Let P be the path on the $2k-3$ vertices $v_1, \ldots, v_{2k-3}$. The top gadget Top is defined as P^{k-2} with $t = v_{k-2}$. Two examples of Top are shown in Figure 3, for the cases $k \in \{5,6\}$. First, note that Top is a k-leaf power. In particular, a k-leaf root of Top is the caterpillar T_{Top}, which is constructed by attaching a leaf to every vertex of P. For the first property, we use Lemma 2 from [29] which implies that in any k-leaf root T of Top, we have $d(v_{k-3}, v_{k-2}) \leq 3$. □

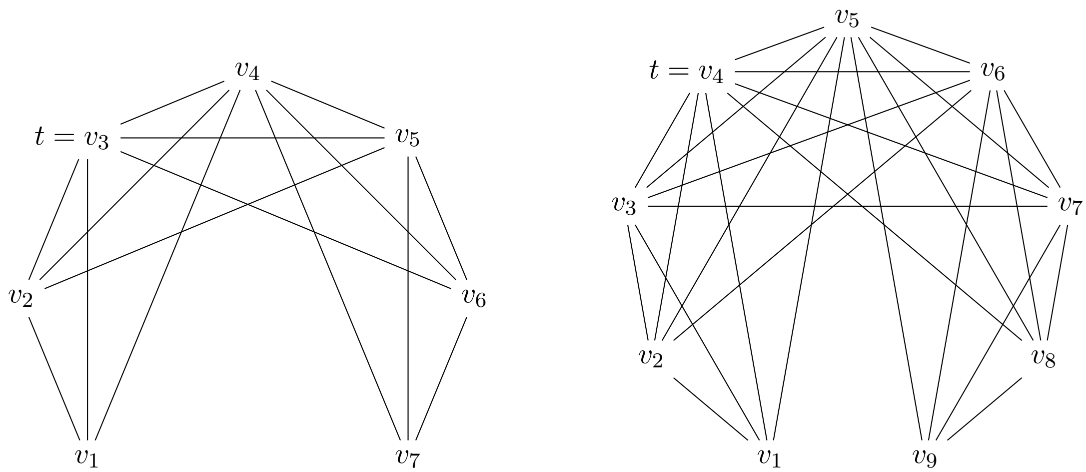


Figure 3: The Top Gadget for $k = 5$ and $k = 6$.

3.2 The Bottom Gadget

Next, we construct the bottom gadget, Bot. A key technical tool we require is the *4-Point Condition*. This is the following classical characterization of tree metrics.

Theorem 3.1 (4-Point Condition)**.** *[7] Let d be a distance on a finite set V, then there exists a tree T whose leaves are V such that $\forall u, v \in V$ $d_T(u, v) = d(u, v)$ if and only if the following condition is true for all $(u, v, w, t) \in V$:*

$$d(u, v) + d(w, t) \leq \max \{d(u, w) + d(v, t), d(v, w) + d(u, t)\}.$$

Our bottom gadget Bot will simply be a *diamond*, the complete graph on 4 vertices minus one edge. Consequently, we begin by proving the following corollary of the 4-Point Condition when applied to a diamond.

Corollary 3.2. *In any k-leaf root T of a diamond with vertex set $\{b, v_1, v_2, v_3\}$ where $(v_1, v_3) \notin E$, $d(b, v_2) \neq k$.*

Proof. Assume for contradiction that $d(b, v_2) = k$. Then since $d(v_1, v_3) > k$, we get $d(v_1, v_3) + d(b, v_2) > 2k$.

On the other hand, since we have a leaf root of the diamond, we must have:

$$\max \{d(v_1, v_2), d(v_2, v_3), d(v_3, b), d(b, v_1)\} \leq k.$$

This implies that $d(v_1, v_2) + d(v_3, b) \leq 2k$ and $d(v_2, v_3) + d(b, v_1) \leq 2k$.

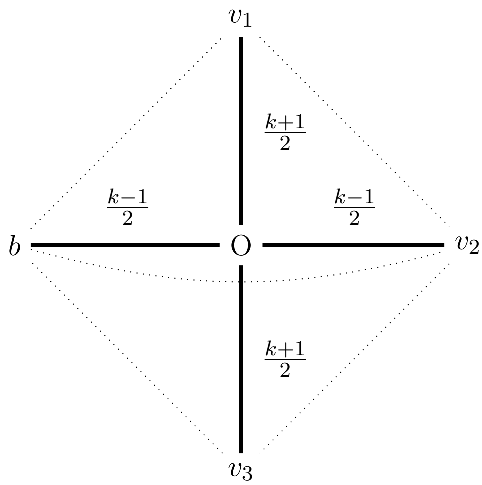
So, $d(v_1, v_3) + d(b, v_2) > \max \{d(v_1, v_2) + d(v_3, b), d(v_2, v_3) + d(b, v_1)\}$ which contradicts Theorem 3.1 □

Corollary 3.2 allows us to prove our critical lemma for the bottom gadget.

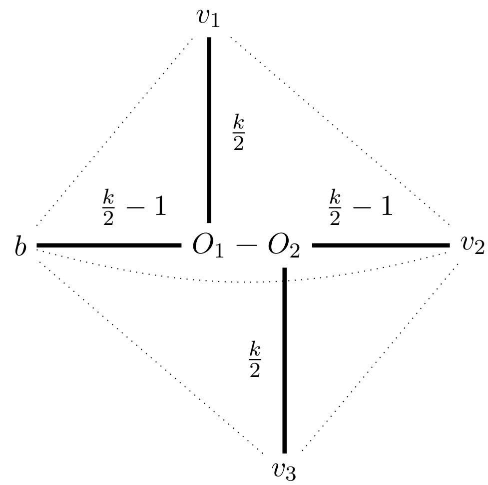
Lemma 2.2. *For all $k \geq 4$, there exists a gadget graph Bot that contains a vertex $b \in V(Bot)$ such that:*

1. *For any k-leaf root T of Bot, $m_T(b) \leq k - 1$.*
2. *There exists a k-leaf root T_{Bot} such that $m_{T_{Bot}}(b) = k - 1$*

Proof. Let the graph Bot be the diamond with vertex set $\{b, v_1, v_2, v_3\}$ and non-edge (v_1, v_3). By Corollary 3.2, $d(b, v_2) \leq k-1$ in any leaf root. Thus the first property holds. For the second property, there are two cases, illustrated in Figure 4, depending upon the parity of k.



(a) T_{Bot} for k odd



(b) T_{Bot} for k even

Figure 4: The k-leaf roots of the diamond with $\min_{v \in V(D)\setminus\{b\}} d(b, v) = k-1$. Here the bold edges denote paths of the described length; the dotted edges are the edges of the diamond.

- If k is odd, start with b and v_2 at distance $k-1$. Let O be the midpoint of the two at distance $\frac{k-1}{2}$ from both. Set v_1 and v_3 to be each at distance $\frac{k+1}{2}$ from O. Then b and v_2 will both be at distance exactly k from both v_1 and v_3, but v_1 and v_3 are at distance $k+1$ from each other. Thus, the only distance which is greater than k is $d(v_1, v_3)$ and the closest vertex to b is v_2, as desired.

- If k is even, start with b and v_2 at distance $k-1$. Set O_1 to be the point at distance $\frac{k}{2}-1$ from b and O_2 the point at distance $\frac{k}{2}-1$ from v_2. Add v_1 at distance $\frac{k}{2}$ from O_1 and v_3 at distance $\frac{k}{2}$ from O_2. Then b is at distance $k-1$ from v_1 and k from v_3, while v_2 is at distance k from v_1 and $k-1$ from v_3. Note that v_1 and v_3 are at distance $k+1$ from each other. Thus, the only distance which is greater than k is $d(v_1, v_3)$ and the closest vertex to b is v_2, as desired.

This lemma follows. □

3.3 The Interior Gadget

Lastly, we have the most complex construction, that of the interior gadget, I. Now we require the following lemma which, again, is a consequence of the 4-Point Condition.

Lemma 3.3. *If $d(t, x_1) \leq \min\{d(t, x_2), d(t, x_3)\}$ and $d(y, x_1) > \max\{d(y, x_2), d(y, x_3)\}$, then:*

$$d(t, x_1) + d(x_2, x_3) < d(t, x_2) + d(x_1, x_3) = d(t, x_3) + d(x_1, x_2)$$

Proof. Assume that $d(t, x_1) + d(x_2, x_3) \geq \max\{d(t, x_2) + d(x_1, x_3), d(t, x_3) + d(x_1, x_2)\}$. Then, by using this assumption and our bound on $d(t, x_1)$, we get:

$$\begin{aligned} d(x_2, x_3) &\geq \max\{d(x_1, x_3) + (d(t, x_2) - d(t, x_1)), d(x_1, x_2) + (d(t, x_3) - d(t, x_1))\} \\ &\geq \max\{d(x_1, x_3), d(x_1, x_2)\} \end{aligned}$$

Then, by combining this bound with our bound on $d(y, x_1)$, we get:

$$d(y, x_1) + d(x_2, x_3) > \max\{d(y, x_2) + d(x_1, x_3), d(y, x_3) + d(x_1, x_2)\}$$

This contradicts Theorem 3.1. This implies that the assumption is wrong, that is, we must have:

$$d(t, x_1) + d(x_2, x_3) < \max\{d(t, x_2) + d(x_1, x_3), d(t, x_3) + d(x_1, x_2)\}.$$

Now, assume without loss of generality that $d(t, x_2) + d(x_1, x_3) \geq d(t, x_3) + d(x_1, x_2)$. Then, by Theorem 3.1, we must have $d(t, x_2) + d(x_1, x_3) \leq \max\{d(t, x_1) + d(x_2, x_3), d(t, x_3) + d(x_1, x_2)\}$. Since $d(t, x_2) + d(x_1, x_3) > d(t, x_1) + d(x_2, x_3)$, this implies that $d(t, x_2) + d(x_1, x_3) \leq d(t, x_3) + d(x_1, x_2)$. So we get that $d(t, x_2) + d(x_1, x_3)$ is bounded above and below by $d(t, x_3) + d(x_1, x_2)$ so they must be equal.

So we get:

$$d(t, x_1) + d(x_2, x_3) < d(t, x_2) + d(x_1, x_3) = d(t, x_3) + d(x_1, x_2).$$

This completes the proof. □

We will also use the following simple lemma:

Lemma 3.4. *For any 3 leaves u, v, w of a tree, $d(u, v) + d(u, w) + d(v, w)$ is even.*

Proof. Since we have a tree, there is a unique vertex O which is simultaneously in the path from u to v, the path from v to w and the path from u to w.

Hence $d(u, v) + d(u, w) + d(v, w) = 2 \cdot (d(u, O) + d(v, O) + d(w, O))$ which must be even. □

We now have all the tools needed to prove our critical lemma for the interior gadget.

Lemma 2.3. *For all $k \geq 5$, there exists a gadget graph I that contains two distinct vertices $t_I, b_I \in V(I)$ such that:*

1. *For all k-leaf roots T of I, $m_T(t_I) \geq k \implies m_T(b_I) = 3$.*
2. *There exists a k-leaf root T_I of I such that $m_{T_I}(t_I) = k$ and $m_{T_I}(b_I) = 3$.*
3. *There exists a k-leaf root R_I of I such that $m_{R_I}(t_I) = k - 1$ and $m_{R_I}(b_I) = 4$.*

Before proving this lemma, let's discuss the requirement that $k \geq 5$. First observe that no such graph can exist for $k \leq 2$ because if $m_T(b_I) = 3$ then b_I is an isolated vertex in I. Thus its distance to other leaves does not matter as long as it's large enough, so the lemma could not hold. Similarly, if $k = 3$, the existence of a 3-leaf root T_I implies that b_I is not an isolated vertex in I. But the existence of a 3-leaf root R_I implies that b_I is an isolated vertex, a contradiction. Finally, for $k = 4$, while there is no direct simple proof that the statement does not hold for any graph, the existence of a characterization of 4-leaf powers implies that no such graph can exist.

Proof. For convenience, we denote t_I and b_I by t and b, respectively. To prove the lemma, we will explicitly construct I for any value of $k \geq 5$. In order to do so, we consider two cases depending, again, upon the parity of k.

For k odd, set $q = \frac{k-1}{2}$. In particular, for $k \geq 5$ we must have $q \geq 2$.

We construct the graph I using the following sets of vertices:

- t and b
- $X = \{x_1, \ldots, x_q\}$
- $Y = \{y_2, \ldots, y_q\}$

The edge set is defined as follows:

- For $i = 1, \ldots q$, (t, x_i) and (b, x_i) are edges. That is, t and b are adjacent to all vertices in X.
- For $i = 1, \ldots, q$, for $j = i+1, \ldots q$, (x_i, x_j) is an edge. That is, X forms a clique.
- For $i = 2 \ldots q$, for $j = i \ldots q$, (y_i, x_j) is an edge.
- (b, y_q) is an edge.

Equivalently, it will be helpful to define the set of edges using the neighborhood of each vertex:

- t is adjacent to $X = \{x_1, \ldots, x_q\}$.
- b is adjacent to X and to y_q.
- For $i = 1, \ldots, q$, x_i is adjacent to t, to b, to $X \setminus \{x_i\}$ and to y_j for $j = 2, \ldots, i$ (with x_1 having no neighbor in Y).
- For $i = 2, \ldots, q$, for $j = i, \ldots, q$, y_i is adjacent to x_j. If $i = q$, then y_q is also adjacent to b.

That is, we take $I = (V, E)$ to be defined by:

$$\begin{aligned} V &= \{t, b\} \cup \left(\bigcup_{i=1}^{q} \{x_i\} \right) \cup \left(\bigcup_{i=2}^{q} \{y_i\} \right) \\ E &= \left(\bigcup_{i=1}^{q} \{(t, x_i), (b, x_i)\} \right) \cup \left(\bigcup_{1 \leq i < j \leq q} \{(x_i, x_j)\} \right) \cup \left(\bigcup_{2 \leq i \leq j \leq q} \{(y_i, x_j)\} \right) \cup \{(b, y_q)\} \end{aligned} \tag{1}$$

This construction is illustrated in Figure 5. We remark that this construction only makes sense for $k \geq 5$ because if $k = 3$ or $k = 1$ then Y is not well defined.

We will prove that this graph satisfies the lemma by proving three claims.

Claim 3.5. *For I as defined in (1):*
For all k-leaf roots T of I, $m_T(t) = k \Longrightarrow m_T(b) = 3$.

Proof. Assume that $m_T(t) = k$, then $\forall i \in [q]$, $d(t, x_i) = k$. Then for all distinct $i, j \in [q]$, by Lemma 3.4 with t, x_i and x_j, $d(x_i, x_j)$ must be even (and, thus, at most $k-1$).

Assume $i < j < \ell$. Then $d(t, x_i) = k = \min\{d(t, x_j), d(t, x_\ell)\}$. Furthermore, $d(y_j, x_i) > k \geq \max\{d(y_j, x_j), d(y_j, x_\ell)\}$ because y_j is adjacent to x_j and x_ℓ but not x_i (nor t). So, by Lemma 3.3, we get:

$$d(t, x_i) + d(x_j, x_\ell) < d(t, x_\ell) + d(x_i, x_j) = d(t, x_j) + d(x_i, x_\ell)$$

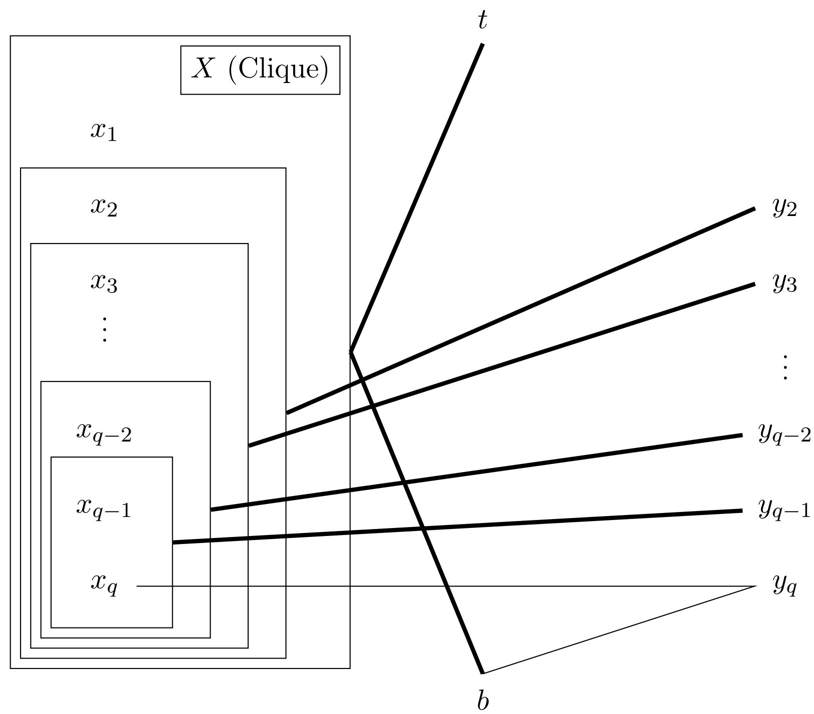


Figure 5: The interior gadget I for odd k. The bold edges signify all possible connections are made.

But because $d(t, x_i) = d(t, x_j) = d(t, x_\ell) = k$. This is true if and only if

$$d(x_j, x_\ell) < d(x_i, x_j) = d(x_i, x_\ell)$$

This implies that for every $i \in [q-1]$, there exists some integer λ_i such that $d(x_i, x_{i+1}) = d(x_i, x_{i+2}) = \cdots = d(x_i, x_q) = \lambda_i$. Moreover, as shown, the λ_i must be even. Thus $k - 1 \geq \lambda_1 > \cdots > \lambda_{q-1} > 2$. By definition, $2q = k - 1$. In particular, there are only $q - 1$ even numbers greater that 2 and at most $k - 1$. Therefore, $\lambda_i = k + 1 - 2i$. Specifically, we have shown that $d(x_i, x_j) = k + 1 - 2i$, for $1 \leq i < j \leq q$.

Recall $d(t, x_i) = k = \min\{d(t, x_q), d(t, b)\}$ and $d(y_q, x_i) > k \geq \max\{d(y_q, x_q), d(y_q, b)\}$, for $i < q$. So, by Lemma 3.3, we obtain:

$$d(x_q, b) + d(x_i, t) < d(x_q, x_i) + d(t, b) = d(x_q, t) + d(x_i, b).$$

Consider $i = 1$. Recall $k \geq 5$ and $q \geq 2$. So $q \neq 1$ implying that $x_q \neq x_1$. It follows that

$$\begin{aligned} & d(x_1, x_q) + d(t, b) = d(x_q, t) + d(x_1, b) \\ \Longrightarrow\ & k - 1 + d(t, b) = k + d(x_1, b) \\ \Longrightarrow\ & d(t, b) = 1 + d(x_1, b) \end{aligned}$$

However, we must have $d(t, b) > k$ and $d(x_1, b) \leq k$. So it must be that $d(t, b) = k + 1$ and $d(x_1, b) = k$.

Next consider $i = q - 1$. Because $q \geq 2$, we have $i \geq 1$ and so x_{q-1} exists. Then $d(x_{q-1}, x_q) =$

$k+1-2(q-1) = k+3-(k-1) = 4$. Therefore, because $d(x_i, t) = k$ for all $1 \leq i \leq q$, we have

$$\begin{aligned} & d(x_{q-1}, x_q) + d(t, b) = d(x_q, t) + d(x_{q-1}, b) \\ \implies & 4 + k + 1 = k + d(x_{q-1}, b) \\ \implies & d(x_{q-1}, b) = 5 \end{aligned}$$

Finally, recall that $d(x_q, b) + d(x_{q-1}, t) < d(x_q, x_{q-1}) + d(t, b)$. This implies that $d(x_q, b) + k < 4 + (k+1)$. In particular, $d(x_q, b) < 5$. Moreover, by Lemma 3.4, we know $d(x_q, b) + d(x_q, x_{q-1}) + d(x_{q-1}, b)$ is even. But $d(x_q, x_{q-1})$ is even and $d(x_{q-1}, b)$ is odd. Hence $d(x_q, b)$ must be odd. In particular, it must be odd **and** less than 5. Hence $d(x_q, b) = 3$, which is what we wanted to show. $\square$

Claim 3.6. *For I as defined in (1):*
There exists a k-leaf root T_I of I such that $m_{T_I}(t) = k$ and $m_{T_I}(b) = 3$.

Proof. We will prove this by explicitly constructing a leaf root. Recall k is odd and $k = 2q+1$.

1. Take a path of length $k+1 = 2q+2$ from t to b.
2. Label the vertices along the path from t to b which are at distance $q+i$ of t as O_i for $i = 1, \ldots, q+1$.
3. Add a path of length $q-i+1$ from O_i to x_i for $i = 1, \ldots, q$.
4. Add a path of length $k-2$ from O_{q+1} to y_q.
5. If $k \geq 7$, add a path of length $q+i$ from O_i to y_i for $i = 2, \ldots, q-1$. (For $k = 5$ we have $q = 2$, so these y_i do not exist.)

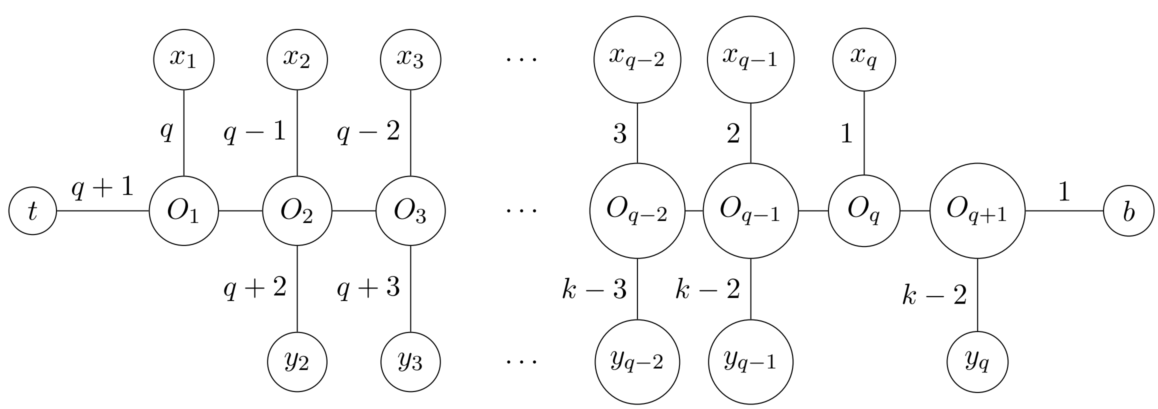


Figure 6: The k-leaf root T_I of the interior gadget for odd $k = 2q+1$. (Recall these will be connected in series below the leaf root of the top gadget; see Figure 2.)

This construction is shown in Figure 6. It remains to verify that this is a valid k-leaf root of I, that is, the leaf vertices at distance at most k in the tree are exactly the edges of I. The required case analysis follows.

- The path from t to b has length $k+1 > k$ and t and b are not neighbors, as desired.
- For $i = 1, \ldots, q$, the path from t to x_i goes through O_i. That is, it is the path from t to O_i which has length $q+i$ followed by the path from O_i to x_i which has length $q-i+1$. So, the path from t to x_i has length $(q+i)+(q-i+1) = k$.

- The path from t to y_q goes through O_{q+1}, so it has length $(2q+1)+(k-2) > k$.
- For $i = 1, \ldots, q$, the path from b to x_i goes through O_q and O_i so it has length $2 + (q-i) + (q-i+1) \leq k$. In particular, for $i = q$, the path from b to x_q has length 3.
- The path from b to y_q goes through O_{q+1} so it has length $1 + (k-2) \leq k$
- For $1 \leq i < j \leq q$, the path from x_i to x_j goes through O_i and O_j so it has length $(q-i+1) + (j-i) + (q-j+1) \leq k$
- For $i = 1, \ldots, q$ and $j = 2, \ldots, q-1$, the path from x_i to y_j goes through O_i and O_j (possibly the same vertex) so it has length $(q-i+1) + (|i-j|) + (q+j) = k + (j-i) + |i-j|$ which is at most k if and only if $i \geq j$.
- For $i = 1, \ldots, q$, the path from x_i to y_q goes through O_i and O_{q+1} so it has length $(q-i+1) + (q+1-i) + (k-2)$ which is equal to k if $i = q$ and strictly greater than k otherwise.

For $k = 5$ the case analysis is complete. If $k \geq 7$ then $q \geq 3$ and so $|Y| > 1$. Thus we have four more cases to verify:

- For $i = 2, \ldots, q-1$, the path from t to y_i goes through O_i, so it has length $(q+i)+(q+i) > k$.
- For $i = 2, \ldots, q-1$, the path from b to y_i goes through O_q and O_i so it has length $2 + (q-i) + (q+i) > k$.
- For $2 \leq i < j \leq q-1$, the path from y_i to y_j goes through O_i and O_j, so it has length $(q+i) + (j-i) + (q+j) > k$.
- For $i = 2, \ldots, q-1$, the path from y_i to y_q goes through O_i and O_{q+1} so it has length $(q+i) + (q+1-i) + (k-2) > k$.

Hence the desired k-leaf root exists. □

It remains to prove the final property to conclude that the claim holds when k is odd:

Claim 3.7. *For I as defined in (1):*
There exists a k-leaf root R_I of I such that $m_{R_I}(t) = k-1$ and $m_{R_I}(b) = 4$.

Proof. To construct R_I we make two minor modifications to the tree T_I used to prove Claim 3.6. First we start with a path of length $k+2$ from t to b. To do this we simply place b at distance 2 from O_{q+1} instead of 1. All the remaining vertices are then placed using the same process *except* for x_1 which is now at distance $q-1$ from O_1 instead of at distance q. The resultant tree is shown in Figure 7. It suffices to verify that all the vertices are still at a correct distance from x_1 and from b.

- The distance from x_1 to all other vertices except b has decreased by 1 so we need to verify that the y_i and y_q are still at distance at least $k+1$. For $i = 2, \ldots, q-1$, the distance from x_1 to y_i is $(q-1) + (i-1) + (q+i) > k$ (since $i \geq 2$). The distance from x_1 to y_q is $(q-1) + q + (k-2) = 2q + k - 3 > k$ (since $q \geq 2$).
- The distance from b to all other vertices except x_1 has increased by 1 so we need to verify that the x_i's and y_q are still at distance at most k. For $i \geq 2$, the distance from b to x_i is $3 + (q-i) + (q-i+1) \leq k$. The distance from b to y_q is $2 + (k-2) = k$. Moreover, all distances from b to an x_i vertex are now at least 4 and not 3 (including x_1, which is at distance $k = 2q+1$ from b). In particular, the distance between b and x_q is exactly 4.

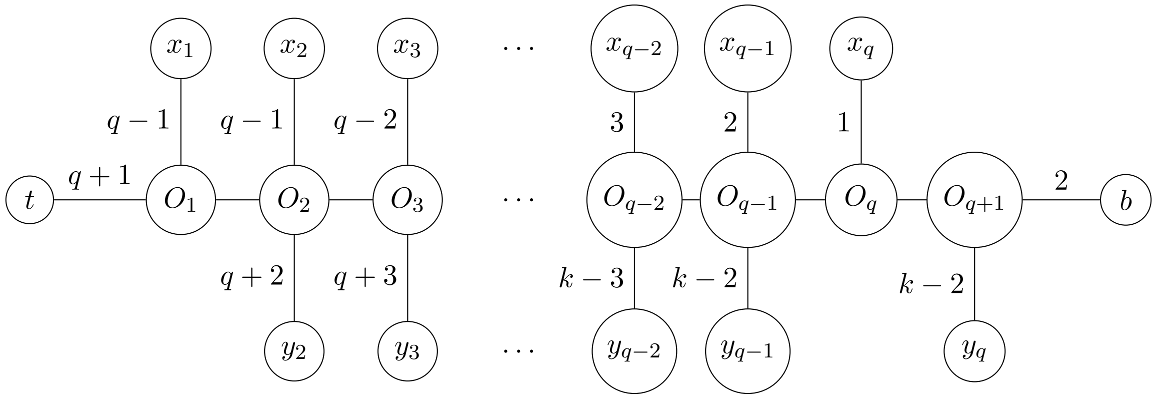


Figure 7: The k-leaf root R_I of the interior gadget for odd $k = 2q+1$. (Recall these will be connected in series above the leaf root of the bottom gadget; see Figure 2.)

Observing that t is now at distance $k-1$ from x_1 and is still at distance k to its other neighbors, we get $m_{R_I}(t) = k-1$ and $m_{R_I}(b) = 4$. Hence the desired k-leaf root exists. □

We have now proven the result holds for k odd. Let's now prove it for k even. The construction of the graph I for even values of k is very similar to the construction for odd values of k, but is slightly more intricate. Take $q = \frac{k}{2}$. The vertex set of G is then

- t and b
- $X = \{x_1, \ldots, x_q\}$
- $Y = \{y_2, \ldots, y_q\}$
- z_1 and z_2.

The edge set of G is defined as follows:

- (t, x_1) is an edge and $\forall i = 2, \ldots q$, (t, x_i), (b, x_i), (z_1, x_i) and (z_2, x_i) are all edges. That is, t is adjacent to all vertices in X while b, z_1 and z_2 are adjacent to all vertices in X except x_1.
- For $i = 1, \ldots, q$, for $j = i+1, \ldots q$, (x_i, x_j) is an edge. That is, X forms a clique.
- For $i = 2, \ldots, q$, for $j = i, \ldots, q$, (y_i, x_j) is an edge.
- (b, y_q) is an edge.
- (z_1, b) and (z_2, b) are both edges, in particular, for all $i \geq 2$, $\{z_1, z_2, b, x_i\}$ will form a diamond (with the (z_1, z_2) edge being missing).

Equivalently, it is again informative to define the set of edges using the neighborhoods of each vertex:

- t and x_1 are adjacent to each other and to X.
- b is adjacent to $X \setminus \{x_1\}$, to y_q and to z_1 and z_2.
- For $i = 2, \ldots q$, x_i is adjacent to t, b, z_1, z_2, $X \setminus \{x_i\}$ and to y_j for $j = 1, \ldots, i$.
- For $i = 2, \ldots, q$, for $j = i, \ldots, q$, y_i is adjacent to x_j.

- y_q is adjacent to x_q and b.
- z_1 and z_2 are adjacent to $X \setminus \{x_1\}$ and to b.

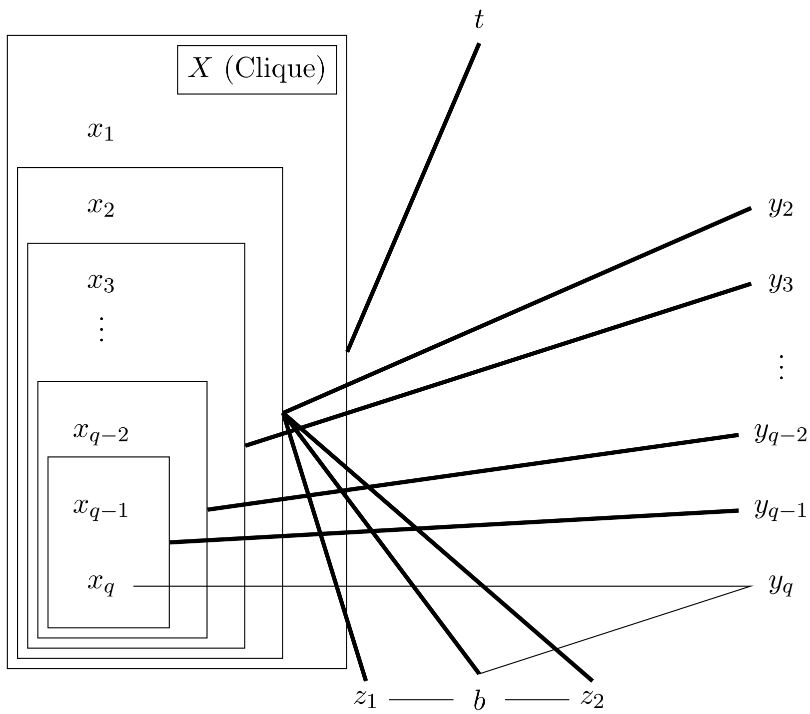


Figure 8: The interior gadget for even k.

That is, we take $I = (V, E)$ to be defined by:

$$
\begin{aligned}
V &= \{t, b, z_1, z_2\} \cup \left(\bigcup_{i=1}^{q} \{x_i\}\right) \cup \left(\bigcup_{i=2}^{q} \{y_i\}\right) \\
E &= \{(t, x_1)\} \cup \left(\left\{\bigcup_{i=2}^{q} \{(t, x_i), (b, x_i), (z_1, x_i), (z_2, x_i)\}\right\}\right) \\
&\quad \cup \left(\bigcup_{1 \le i < j \le q} \{(x_i, x_j)\}\right) \cup \left(\bigcup_{2 \le i \le j \le q} \{(y_i, x_j)\}\right) \cup \{(b, y_q), (z_1, b), (z_2, b)\}
\end{aligned}
\tag{2}
$$

This construction is illustrated in Figure 8. We remark that this construction only makes sense for $k \geq 4$. If $k = 2$ then Y is not well defined. But, while the construction makes sense for $k = 4$, we will show later where it fails to work.

Recall z_1 and z_2 form a diamond with b and x_i for any $i \geq 2$. Therefore, by Corollary 3.2, we have that $d(b, x_i) \neq k$ in any k-leaf root, in particular this is true for $i = 2$ so we get $d(b, x_2) \neq k$.

Claim 3.8. *For I as defined in (2):*
For all k-leaf roots T of I, $m_T(t) = k \Longrightarrow m_T(b) = 3$.

Proof. Most of the arguments used to prove Claim 3.5 are still valid. Using Lemma 3.3 we can show that $d(x_i, x_j) = k + 2 - 2i$. Furthermore, since $k \geq 6$, $q \geq 3$ and for $2 \leq i \leq q-1$, we have $d(x_q, b) + d(x_i, t) < d(x_i, x_q) + d(t, b) = d(x_q, t) + d(x_i, b)$.

However (x_1, b) is not an edge in this case so instead we must consider $i = 2$. Now we get:

$$\begin{aligned}
&d(x_q, x_2) + d(t, b) = d(x_q, t) + d(x_2, b) \\
\implies &k - 2 + d(t, b) = k + d(x_2, b) \\
\implies &d(t, b) = 2 + d(x_2, b)
\end{aligned}$$

Moreover, we must have $d(t, b) > k$ and $d(x_2, b) \leq k$. So we get either $d(t, b) = k+2$ and $d(x_2, b) = k$ or $d(t, b) = k+1$ and $d(x_2, b) = k-1$. But we cannot have $d(x_2, b) = k$ as this violates Corollary 3.2 when considering the diamond (b, z_1, x_2, z_2). Hence, we must have $d(x_2, b) = k-1$ and $d(t, b) = k+1$.

Next, as in Claim 3.5, consider $i = q-1$. Then as $d(x_{q-1}, x_q) = k+2-2(q-1) = k+2-(k-2) = 4$, we get

$$\begin{aligned}
&d(x_{q-1}, x_q) + d(t, b) = d(x_q, t) + d(x_{q-1}, b) \\
\implies &4 + k + 1 = k + d(x_{q-1}, b) \\
\implies &d(x_{q-1}, b) = 5
\end{aligned}$$

As before, we have shown that $d(x_q, b) + d(x_{q-1}, t) < d(x_q, x_{q-1}) + d(t, b)$. This implies $d(x_q, b) + k < 4 + (k+1)$ and so $d(x_q, b) < 5$. Moreover, by Lemma 3.4, $d(x_q, b) + d(x_q, x_{q-1}) + d(x_{q-1}, b)$ is even. Consequently $d(x_q, b)$ must be odd. But $d(x_q, x_{q-1}) = 4$ is even and $d(x_{q-1}, b) = 5$ is odd. Thus $d(x_q, b)$ must be odd **and** less than 5. Hence $d(x_q, b) = 3$, as desired. □

Claim 3.9. *For I as defined in 2:*
There exists a k-leaf root T_I of I such that $m_{T_I}(t) = k$ and $m_{T_I}(b) = 3$.

Proof. We construct I similarly to the proof of Claim 3.6. Recall that $k = 2q$.

1. Take a path of length $k + 1$ from t to b.
2. Label as O_i the vertex along the path from t to b at distance $q-1+i$ from t, for $i = 1, \ldots, q+1$.
3. Add a path of length $q - i + 1$ from O_i to x_i for $i = 1, \ldots, q$.
4. Add a path of length $q + i - 1$ from O_i to y_i for $i = 2, \ldots, q - 1$.
5. Add a path of length $k - 2$ from O_{q+1} to y_q.
6. Add a path of length q from O_2 to z_1 and a path of length q from O_3 to z_2. (Since $k \geq 6$ we have $q \geq 3$ and, so, both O_2 and O_3 exist.)

This graph T_I is shown in Figure 9. We must show T_I is a k-leaf root of I. The fact that these distances are valid follows from Claim 3.6 with the modification that $k = 2q$. It remains to verify that z_1 and z_2 also have the correct neighborhoods.

- For all $i = 2, \ldots, q$, the vertices t, x_1, y_q, and y_i are at distance at least $q + 1$ from O_2 and O_3. Thus, they are at distance at least $2q + 1 > k$ from z_1 and z_2.
- For $i = 2, \ldots, q$, the vertices x_i and b are both at distance at most q from O_2 and from O_3. Thus they are all within distance $2q = k$ from z_1 and z_2.
- The path from z_1 to z_2 goes through both O_2 and O_3 so it has length $q + 1 + q > k$.

Hence the desired k-leaf root T_I exists. □

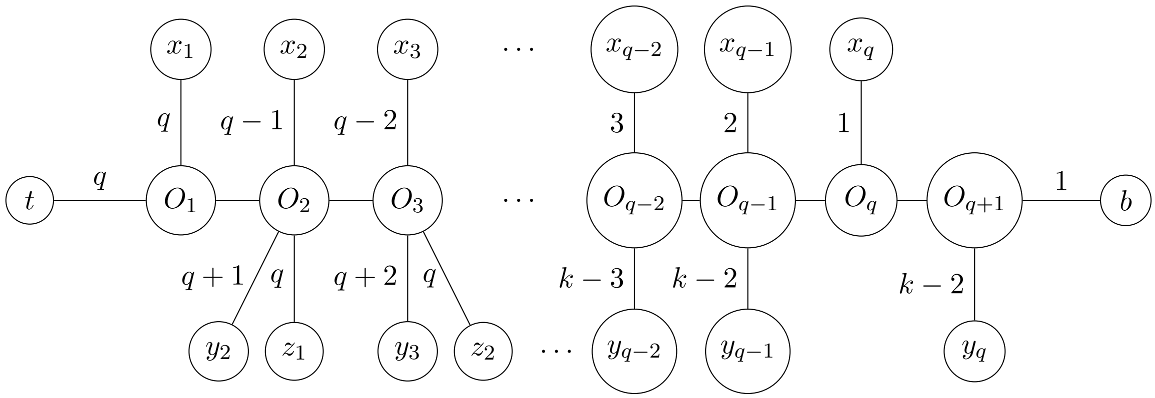


Figure 9: The k-leaf root R_I of the interior gadget for even $k = 2q$. (Recall these will be connected in series below the leaf root of the top gadget; see Figure 2.)

Claim 3.10. *For I as defined in 2:*
There exists a k-leaf root R_I of I such that $m_{R_I}(t) = k - 1$ and $m_{R_I}(b) = 4$.

Proof. To construct R_I we make four minor modifications to the tree T_I used to prove Claim 3.9. First we start with a path of length $k + 2$ from t to b. To do this we simply place b at distance 2 from O_{q+1} instead of 1. Second, we place x_2 at distance $q - 2$ from O_1 instead of at distance $q - 1$. Third and fourth, we place z_1 and z_2 at distance q from O_3 and O_4, respectively. The resultant tree is shown in Figure 10.

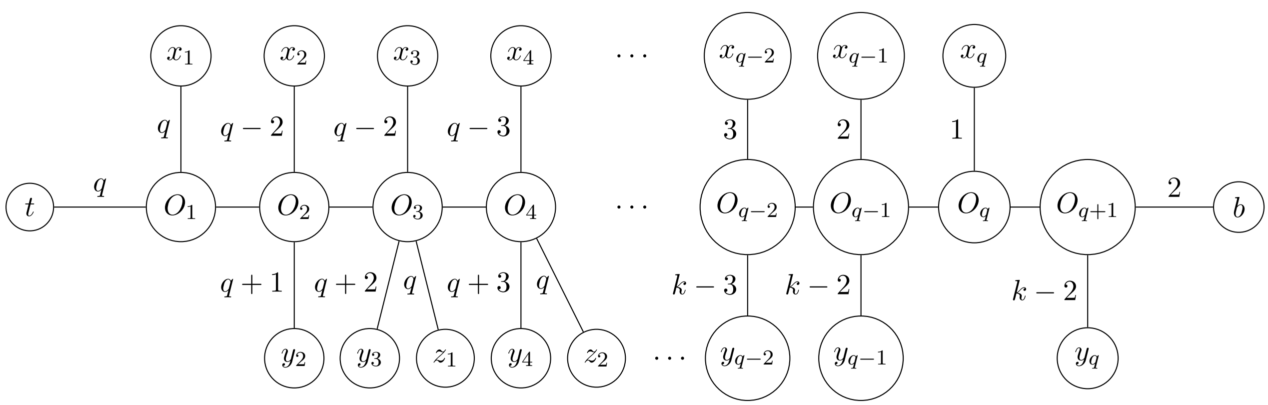


Figure 10: The k-leaf root R_I of the interior gadget for even $k = 2q$. (Recall these will be connected in series above the leaf root of the bottom gadget; see Figure 2.)

We emphasize that this construction is **NOT** possible for $k = 4$. This is because then $q = 2$ and O_4, the vertex O_4 does not exist. Thus this construction applies for even $k \geq 6$.

Finally, as in the proof of Claim 3.7, these changes do not change the neighborhood of the displaced vertices b and x_2, z_1 and z_2. Hence the desired k-leaf root exists. □

This completes the proof of the lemma. □

With our three critical lemmas proven, the main theorem now holds by the method shown in Section 2.

4 Linear Leaf Powers

A *caterpillar* is a graph which has a central path and a set of leaves whose neighbor is on the central path. A graph is said to be a *linear leaf power* if it has a leaf root which is the subdivision of a caterpillar. Such a leaf root is called a *linear leaf root* [2].

Our results apply not only to general leaf powers but also to this variant. Indeed, even if we restrict to having a subdivision of a caterpillar as a leaf root, it is impossible to get a simple forbidden subgraph characterization of linear k-leaf powers for $k \geq 5$.

Theorem 4.1. *For $k \geq 5$, the set of linear k-leaf powers cannot be written as the set of strongly chordal graphs which are $\mathscr{F}_k$-free where $\mathscr{F}_k$ is a finite set of graphs.*

As in the general case, we prove this using three gadgets. However, we must now add a condition to ensure that merging the gadgets preserves having a subdivision of a caterpillar as a leaf root.

Lemma 4.2. *For all $k \geq 5$ there exists a gadget graph Top that contains a vertex $t \in V(Top)$ such that:*

1. *For any linear k-leaf root T of Top, $m_T(t) = 3$.*
2. *There exists a linear k-leaf root T_{Top} of Top where t is a neighbor of the last node of the central path.*

Lemma 4.3. *For all $k \geq 5$ there exists a gadget graph Bot that contains a vertex $b \in V(Bot)$ such that:*

1. *For any linear k-leaf root T of Bot, $m_T(b) \leq k - 1$.*
2. *There exists a linear k-leaf root T_{Bot} such that $m_{T_{Bot}}(b) = k - 1$ where b is a neighbor of the last node of the central path.*

Lemma 4.4. *For all $k \geq 5$ there exists a gadget graph I that contains two distinct vertices $t_I, b_I \in V(I)$ such that:*

1. *For all linear k-leaf roots T of I, $m_T(t_I) \geq k \implies m_T(b_I) = 3$.*
2. *There exists a linear k-leaf root T_I of I such that $m_{T_I}(t_I) = k$ and $m_{T_I}(b_I) = 3$ where b and t are neighbors of the first and last node of the central path respectively.*
3. *There exists a linear k-leaf root R_I of I such that $m_{R_I}(t_I) = k - 1$ and $m_{R_I}(b_I) = 4$ where b and t are neighbors of the first and last node of the central path respectively.*

Merging the gadgets is identical to before except that we now use the condition that, in each gadget, t and/or b are neighbors of the extremal vertices of the central path. This means when merging the gadgets using the parent of these vertices, we merge the central paths by their endpoint to create a longer path, ensuring we produce another caterpillar subdivision.

We can verify that Lemma 4.3 follows from Lemma 2.2 and Lemma 4.4 follows from Lemma 2.3, as our constructions for the interior gadget and the Bottom Gadget used for the general case satisfy the properties needed, namely the leaf root used in the proofs are subdivisions of caterpillars with the required vertices connected to the extremal vertices of the central path. On the other hand, the graph used to construct the Top Gadget does not satisfy this property[4]; so we need to construct a new graph for Top.

[4] The Top Gadget for the case of general k-leaf powers is a caterpillar but the vertex t used is not connected to an extremal vertex of the central path.

Proof of Lemma 4.2. The gadget graph Top consists of a $k-1$ clique with vertices $X = \{x_1, \ldots, x_{k-1}\}$, to which we add vertices $Y = \{y_0, \ldots, y_k\}$ such that the neighborhood of y_i is $X \cap \{x_{i-1}, x_i, x_{i+1}\}$ (that is, y_i has 3 neighbors if $i = 2, \ldots, k-2$, y_1 and y_{k-1} have two neighbors and y_0 and y_k have one neighbor). This gadget is illustrated in Figure 11[5].

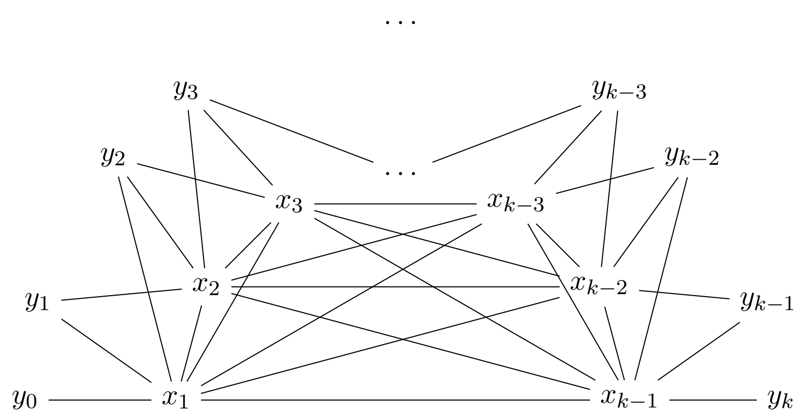


Figure 11: The top gadget for linear leaf roots.

Let T be any linear k-leaf root of Top. For $1 \leq i \leq k-1$, let O_i be the vertex on the central path closest to x_i in T. We wish to show that for $1 \leq i < j \leq k-1$, $O_i \neq O_j$. For a contradiction, assume that $O_i = O_j = O$, for some $i < j$. Now consider y_{i-1} and y_{j+1}. Since $i < j$, y_{i-1} is not a neighbor of x_j and y_{j+1} is not a neighbor of x_i in Top. In T, every path from a leaf to x_i or to x_j must go through O. In particular, for all leaves v of T we have $d(x_i, O) - d(x_j, O) = d(x_i, v) - d(x_j, v)$. However, we must have $d(x_i, y_{i-1}) < d(x_j, y_{i-1})$ and $d(x_i, y_{j+1}) > d(x_j, y_{j+1})$. This implies that $0 > d(x_i, y_{i-1}) - d(x_j, y_{i-1}) = d(x_i, O) - d(x_j, O) = d(x_i, y_{j+1}) - d(x_j, y_{j+1}) > 0$, a contradiction. Thus $O_i \neq O_j$.

Since X forms a clique, all of its vertices must be within distance k from one another in T. Moreover, they must also all be connected to a different vertex of the path. Let ℓ and r be the endpoints of the central path. Let O_ℓ be the O_i closest one to ℓ and O_r the closest one to r. (Note that O_ℓ and O_r are not necessarily ℓ and r since these only take into account X and not Y.)

Every O_i must be contained in the path from O_ℓ to O_r. In particular, the distance from x_ℓ to x_r is k if and only if there is no vertex in the path which is not O_i for any i, and both x_ℓ and x_r are at distance one from O_ℓ and O_r, respectively. In particular, there exists a permutation σ of $\{1, \ldots, k-1\}$ such that the path from O_ℓ to O_r is exactly $O_\ell = O_{\sigma(1)} \to O_{\sigma(2)} \to \cdots \to O_{\sigma(k-2)} \to O_{\sigma(k-1)} = O_r$.

Consider $x_{\sigma(1)} = x_\ell$ and $x_{\sigma(2)}$. By construction, one of $y_{\sigma(2)-1}$ or $y_{\sigma(2)+1}$ is not a neighbor of $x_{\sigma(1)}$ but is a neighbor of $x_{\sigma(2)}$. Let $y_{\sigma(2)\pm1}$ denote the one which satisfies this property (or either of them if both satisfy it). The path from $y_{\sigma(2)\pm1}$ to $x_{\sigma(1)}$ and the path from $y_{\sigma(2)\pm1}$ to $x_{\sigma(2)}$ must both go through $O_{\sigma(1)}$ or both go through $O_{\sigma(2)}$. Since $d(x_{\sigma(1)}, O_{\sigma(1)}) = 1$, $x_{\sigma(1)}$ is closer to $O_{\sigma(1)}$ than $x_{\sigma(2)}$, therefore both paths cannot go through $O_{\sigma(1)}$ otherwise $y_{\sigma(2)\pm1}$ would be closer to $x_{\sigma(1)}$ than $x_{\sigma(2)}$. Moreover, since $d(x_{\sigma(1)}, O_{\sigma(2)}) = d(x_{\sigma(1)}, O_{\sigma(1)}) + d(O_{\sigma(1)}, O_{\sigma(2)}) = 2$ and $y_{\sigma(2)\pm1}$ is closer to $O_{\sigma(2)}$ than to $O_{\sigma(1)}$, we must have $d(x_{\sigma(2)}, O_{\sigma(2)}) < 2$. Consequently, $d(x_{\sigma(2)}, O_{\sigma(2)}) = 1$. It immediately follows that $y_{\sigma(2)\pm1}$ is at distance k from $x_{\sigma(2)}$ and at distance $k+1$ from $x_{\sigma(1)}$.

Similarly, we can show that $x_{\sigma(1)}$ has at least one neighbor, denoted $y_{\sigma(1)\pm1}$, which is not a neighbor of $x_{\sigma(2)}$. This implies that its distance to $O_{\sigma(1)}$ is $k-1$. However, all x_i, for $i \neq \sigma(1)$, are at distance at least 2 from $O_{\sigma(1)}$. So there exists $y_{\sigma(1)\pm1}$ which has no neighbor in X except $x_{\sigma(1)}$.

[5] We remark that a more careful analysis of this family of graphs might also work for the general case.

The same argument shows that $y_{\sigma(k-1)\pm 1}$ has no neighbor in X except $x_{\sigma(k-1)}$. But the only vertices in Y of degree 1 are y_0 and y_k. Therefore, either $1 = \sigma(1)$ or $1 = \sigma(k-1)$. In either case x_1 is at distance exactly 3 from another (leaf) vertex, namely, either $x_{\sigma(2)}$ or $x_{\sigma(k-2)}$. Next observe that no two distinct leaves of T can be at distance 2. This is because no two vertices of Top have the same set of neighbors. Hence, taking $t = x_1$, we obtain $m_T(t) = 3$.

It remains to find a linear k-leaf root satisfying the second property in the lemma. We build this tree T_{Top} as follows:

1. Take a path of length k from O_1 to O_{k-1}.
2. Label the vertices along the path from x_1 to x_{k-1} which are at distance i from x_1 as O_i for $i = 1, \dots, k-1$.
3. Add a path of length 1 from O_i to x_i for $i = 2, \dots, k-2$.
4. Add a path of length $k-2$ from O_i to y_i for $i = 1, \dots, k-1$.
5. Add a path of length $k-1$ from O_1 to y_0 and from O_{k-1} to y_k.

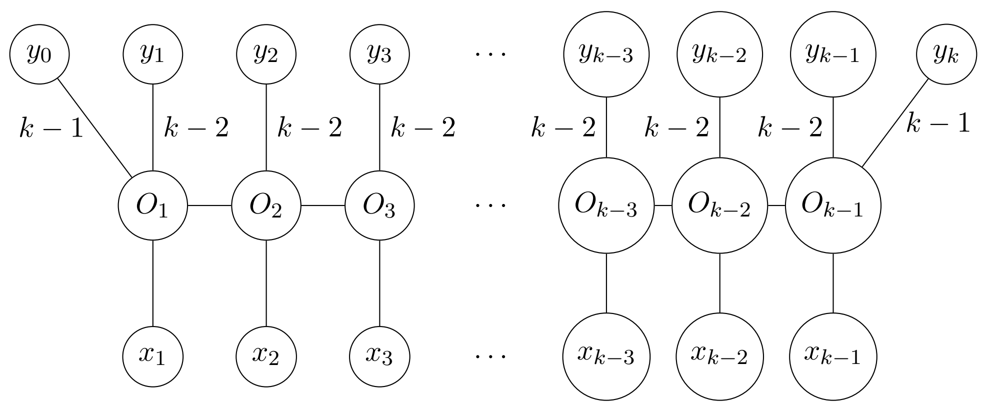


Figure 12: The linear k-leaf root T_{Top} for the top gadget.

This tree is shown in Figure 12. It is easy to verify that this tree does induce X to form a clique in its k-leaf power graph, and that y_i is only adjacent to $\{x_{i-1}, x_i, x_{i+1}\} \cap X$. Moreover, taking $t = x_1$ again yields the desired result. $\square$

We remark that this result does not immediately imply that there is no characterization for the entire class of linear leaf powers using chordal graphs and a finite number of forbidden induced subgraphs. We have proven that such a characterization is impossible for each $k \geq 5$, but not necessarily for the union over all k. It was proved by Bergougnoux et al. [2] that linear leaf powers are also co-threshold tolerance. Further work on this could include verifying whether co-threshold tolerance graphs can be characterized using a finite number of obstructions. Furthermore, it is worth noting that, despite Lemma 4.1, linear leaf powers are recognizable in polynomial time [2].

5 Acknowledgements

We would like to thank the Montreal Game Theory Workshop for helping bring together our research group.

6 Conclusion

We have shown that k-leaf powers require a deeper characterization than strong chordality with a finite set of forbidden induced subgraphs. Several directions to explore remain in order to gain a more comprehensive understanding of k-leaf power graphs. First, is it possible to construct and/or characterize minimal, strongly chordal graphs that are not k-leaf powers? We were able to construct graphs H_n that *contain* such minimal examples as induced subgraphs; but we did not construct those examples explicitly. Following this line of reasoning, it may be possible to characterize k-leaf powers as strongly chordal graphs that also forbid an additional infinite, but easy-to-describe family of forbidden subgraphs. A famous example of this are interval graphs, which are the chordal graphs containing no asteroidal triples [22].

Second, are there relevant subclasses of k-leaf powers that can be characterized by strong chordality and a finite set of forbidden induced subgraphs? For example, the k-leaf powers whose k-leaf roots admit a subdivision of a star should be easy enough to characterize. What about subdivisions of a tree with a small number, say two or three, of non-leaf vertices? One may also consider the k-leaf powers of caterpillars (not subdivided). Based on the midpoint arguments of Brandstädt et al. [4, Theorem 6], it would appear that, for even k, these coincide with the unit interval graphs whose intervals have length $k-2$ and integer endpoints. Such (twin-free) graphs were shown to admit a finite set of forbidden induced subgraphs in [13]. If this characterization extends to caterpillar k-leaf powers, this would show that taking subdivisions is necessary for our result on caterpillar graphs. It may also be interesting to characterize k-leaf powers for other graph classes that are known to be contained in $\mathcal{L}$; for instance, k-leaf powers that are also ptolemaic graphs, interval graphs, rooted directed path graphs, and others.

References

[1] R. Ahmed, G. Bodwin, F. Sahneh, K. Hamm, M. Jebelli, S. Kobourov, and R. Spence. Graph spanners: A tutorial review. *Computer Science Review*, 37:100253, 2020.

[2] B. Bergougnoux, S. Høgemo, J. Telle, and M. Vatshelle. Recognition of linear and star variants of leaf powers is in p. In Michael A. Bekos and Michael Kaufmann, editors, *Graph-Theoretic Concepts in Computer Science*, pages 70–83, Cham, 2022. Springer International Publishing.

[3] A. Brandstädt, V. Le, and R. Sritharan. Structure and linear-time recognition of 4-leaf powers. *ACM Trans. Algorithms*, 5(1), 12 2008.

[4] A. Brandstädt, C. Hundt, F. Mancini, and P. Wagner. Rooted directed path graphs are leaf powers. *Discrete Mathematics*, 310(4):897–910, 2010.

[5] A. Brandstädt and V. Le. Structure and linear time recognition of 3-leaf powers. *Information Processing Letters*, 98(4):133–138, 2006.

[6] A. Brandstädt, V. Le, and D. Rautenbach. A forbidden induced subgraph characterization of distance-hereditary 5-leaf powers. *Discrete Mathematics*, 309(12):3843–3852, 2009.

[7] P. Buneman. A note on the metric properties of trees. *Journal of Combinatorial Theory, Series B*, 17(1):48–50, 1974.

[8] H-C. Chang, R. Krauthgamer, and Z. Tan. Almost-linear ε-emulators for planar graphs. In *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*, pages 1311–1324, 2022.

[9] M-S. Chang and M-T. Ko. The 3-steiner root problem. In Andreas Brandstädt, Dieter Kratsch, and Haiko Müller, editors, *Graph-Theoretic Concepts in Computer Science*, pages 109–120, Berlin, Heidelberg, 2007. Springer Berlin Heidelberg.

[10] V. Cohen-Addad, A. Filtser, P. Klein, and H. Le. On light spanners, low-treewidth embeddings and efficient traversing in minor-free graphs. In *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)*, pages 589–600. IEEE, 2020.

[11] M. Dom, J. Guo, F. Hüffner, and R. Niedermeier. Error compensation in leaf root problems. pages 389–401, 12 2004.

[12] G. Ducoffe. The 4-steiner root problem. In *International Workshop on Graph-Theoretic Concepts in Computer Science*, pages 14–26. Springer, 2019.

[13] G. Durán, F. Slezak, L. Grippo, F. Oliveira, and J. Szwarcfiter. On unit interval graphs with integer endpoints. *Electronic Notes in Discrete Mathematics*, 50:445–450, 2015.

[14] A. Filtser, M. Kapralov, and N. Nouri. Graph spanners by sketching in dynamic streams and the simultaneous communication model. In *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1894–1913. SIAM, 2021.

[15] M. Hellmuth, C. Seemann, and P. Stadler. Generalized fitch graphs ii: Sets of binary relations that are explained by edge-labeled trees. *Discrete Applied Mathematics*, 283:495–511, 2020.

[16] S. Høgemo. Lower bounds for leaf rank of leaf powers. *CoRR*, abs/2402.18245, 2024.

[17] M. Lafond. On strongly chordal graphs that are not leaf powers. In H. Bodlaender and G. Woeginger, editors, *Graph-Theoretic Concepts in Computer Science*, pages 386–398, Cham, 2017. Springer International Publishing.

[18] M. Lafond. Recognizing k-leaf powers in polynomial time, for constant k. *ACM Trans. Algorithms*, 19(4), 9 2023.

[19] M. Lafond and N. El-Mabrouk. Orthology and paralogy constraints: satisfiability and consistency. *BMC genomics*, 15:1–10, 2014.

[20] H. Le and S. Solomon. Near-optimal spanners for general graphs in (nearly) linear time. In *Proceedings of the 2022 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 3332–3361. SIAM, 2022.

[21] V. Le and C. Rosenke. Computing optimal leaf roots of chordal cographs in linear time. In *International Symposium on Fundamentals of Computation Theory*, pages 348–362. Springer, 2023.

[22] C. Lekkerkerker and J Boland. Representation of a finite graph by a set of intervals on the real line. *Fundamenta Mathematicae*, 51:45–64, 1962.

[23] Y. Long and P. Stadler. Exact-2-relation graphs. *Discrete Applied Mathematics*, 285:212–226, 2020.

[24] R. Nevries and C. Rosenke. Towards a characterization of leaf powers by clique arrangements. *Graphs and Combinatorics*, 32:2053–2077, 2016.

[25] N. Nishimura, P Ragde, and D. Thilikos. On graph powers for leaf-labeled trees. *Journal of Algorithms*, 42(1):69–108, 2002.

[26] A. Raychaudhuri. On powers of strongly chordal and circular arc graphs. *Ars Combin.*, 34:147–160, 1992.

[27] M. Thorup and U. Zwick. Spanners and emulators with sublinear distance errors. In *Proceedings of the seventeenth annual ACM-SIAM symposium on Discrete algorithm*, pages 802–809, 2006.

[28] J. Van Den Brand, S. Forster, and Y. Nazari. Fast deterministic fully dynamic distance approximation. In *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1011–1022. IEEE, 2022.

[29] P. Wagner and A. Brandstädt. The complete inclusion structure of leaf power classes. *Theoretical Computer Science*, 410(52):5505–5514, 2009. Combinatorial Optimization and Applications.