Unavoidable subtrees

Maria Axenovich* and Georg Osang†

July 5, 2012

Abstract

Let $\mathcal{T}_k$ be a family of all k-vertex trees. For $\mathcal{T} \subseteq \mathcal{T}_k$ and a tree T, we write $T \to \mathcal{T}$ if T contains at least one of the trees from $\mathcal{T}$ as a subtree, we write $T \not\to \mathcal{T}$ otherwise. Let $\text{ex}(\mathcal{T})$ be the smallest integer n, if such exists, such that for any tree T on at least n vertices $T \to \mathcal{T}$. It is shown that $\min\{\text{ex}(\mathcal{T}) : \mathcal{T} \subseteq \mathcal{T}_k, |\mathcal{T}| = q\} = 2^{\Theta(k \log^{q-1} k)}$, where $\log^{q-1}$ is the $q-1$ times iterated logarithm. In addition, the bounds on $\text{ex}(\mathcal{T})$ for families $\mathcal{T}$ with a given number of spiders are given.

1 Introduction

For a finite family $\mathcal{T}$ of k-vertex trees and a tree T, we write $T \to \mathcal{T}$ if T contains at least one of the trees from $\mathcal{T}$ as a subtree, we write $T \not\to \mathcal{T}$ otherwise. We call a family $\mathcal{T}$ ***unavoidable*** if there is an integer n such that for any tree T on at least n vertices $T \to \mathcal{T}$. We denote the set of all n-vertex trees $\mathcal{T}_n$, an n-vertex star S_n and an n-vertex path P_n.

Clearly, not every family of trees is unavoidable. Observe that each unavoidable family $\mathcal{T}$ of trees must contain a path and a star, otherwise either an arbitrarily large star or an arbitrarily long path will avoid $\mathcal{T}$. There is an extensive literature on unavoidable trees in tournaments, see for example [7], unavoidable trees in topological graphs [6], and more. The universal trees containing all trees of given order have been studied as well, see [1, 2, 3]. The algorithmic aspect of finding a given tree as a subtree in a larger given tree has been addressed in [8]. However, the extremal problem of unavoidable subtrees in trees did not receive a due attention. Here, we investigate the smallest value of n such that any tree on n vertices contains a member of $\mathcal{T}$. Let, for an unavoidable family $\mathcal{T}$ of trees,

$$\text{ex}(\mathcal{T}) = \min\{n : \forall T \in \mathcal{T}_n,\ T \to \mathcal{T}\}.$$

In particular, for any tree T on at least $\text{ex}(\mathcal{T})$ vertices $T \to \mathcal{T}$, but there is a tree on $\text{ex}(\mathcal{T}) - 1$ vertices such that $T \not\to \mathcal{T}$. We will later prove the easy fact that

Proposition 1. $\text{ex}(\{S_{k+1}, P_{k+1}\}) = \begin{cases} 2 + \frac{k-1}{k-3}((k-2)^{(k-1)/2} - 1) & \textit{if } k \textit{ is odd} \\ 3 + 2\frac{k-2}{k-3}((k-2)^{(k-2)/2} - 1) & \textit{if } k \textit{ is even} \end{cases} = 2^{0.5k \log k(1+o(1))}$.

*Departments of Mathematics, Iowa State University, USA, and Karlsruhe Institute of Technology, Email: axenovic@iastate.edu. The research is supported in part by NSF-DMS grant 0901008.

†Karlsruhe Institute of Technology, Germany, Email: georg.osang@student.kit.edu

In addition, if $\mathcal{T}_1 \subseteq \mathcal{T}$ then $\mathrm{ex}(\mathcal{T}_1) \geq \mathrm{ex}(\mathcal{T})$. Since any unavoidable tree family $\mathcal{T}$ in $\mathcal{T}_k$ contains a star S_k and a path P_k, $\mathrm{ex}(\{S_k, P_k\}) \geq \mathrm{ex}(\mathcal{T})$. The main question we investigate here is how large the difference $\mathrm{ex}(\{S_k, P_k\}) - \mathrm{ex}(\mathcal{T})$ can be. Formally, let

$$f(k,q) = \min\{\mathrm{ex}(\mathcal{T}) : \ \mathcal{T} \subseteq \mathcal{T}_k, \ |\mathcal{T}| = q\},$$

if this minimum exists. We show that the behavior of $\mathrm{ex}(\mathcal{T})$ as a function depends heavily on the number of spiders $\mathcal{T}$ contains. A *spider* is a tree with at most one vertex of degree greater than 2. For this purpose we define the following function for non-negative integers p, q, k:

$$f(k,p,q) = \min\{\mathrm{ex}(\mathcal{T}) : \quad \mathcal{T} \subseteq \mathcal{T}_k \text{ is a union of } p+2 \text{ spiders and } q \text{ non-spiders}\}.$$

For a tree Q that is not a spider, let the *span* of Q, denoted $\mathrm{span}(Q)$, be the set of distances between vertices of degree at least 3. We denote the maximum degree of a graph and its diameter with Δ and diam respectively. In the following, the logarithms are base 2 and $\log^i x = \log\log\cdots\log x$, where log is iterated i times.

Theorem 1. $f(k,p,q) = 2^{\Theta(k\log^{p+1} k)}$*, specifically,*

$$2^{4^{-q-1}k\log^{p+1}k(1+o(1))} \leq f(k,p,q) \leq 2^{k\log^{p+1}k(1+o(1))}.$$

The upper bound is guaranteed by a family $\mathcal{T} = \{S_k, P_k, Q_1, \ldots, Q_p\}$*, where* Q_i *is a balanced spider of maximum degree* $\log^i k$ *for* $i = 1, \ldots, p$*.*

Corollary 1. $f(k,p) = f(k,p-2,0) = 2^{\Theta(k\log^{p-1} k)}$*, specifically,*

$$2^{0.25k\log^{p-1}k(1+o(1))} \leq f(k,p) \leq 2^{k\log^{p-1}k(1+o(1))}.$$

Corollary 2. $f(k,0,q) = 2^{\Theta(k\log k)}$*, specifically,*

$$2^{0.5\cdot 4^{-q}k\log k(1+o(1))} \leq f(k,0,q) \leq 2^{0.5k\log k}.$$

In order to establish the lower bounds, for each eligible family of trees we will construct a tree that avoids the family. For the upper bounds, we will find a specific family of trees and show that no tree of a certain size can avoid it.

2 Definitions, observations, constructions

Henceforth, sequences and vectors will be written in bold script, and their elements referred to in subscript, e.g., a sequence $\mathbf{b} = (b_0, \ldots)$, or a vector $\mathbf{u} = (u_0, u_1, \ldots, u_m)$ for some $m \geq 0$. We shall always index the elements of the vectors starting with 0. The *depth* of a rooted tree T, $\mathrm{depth}(T)$, is the largest distance between the root and any leaf. Most of our results use balanced (stable) rooted trees where the vertices at the same distance from the root have the same degree. Formally, for a vector $\mathbf{u}$, with $u_i \geq 2$, $i = 1, \ldots, m$ and $u_0 \geq 1$, let a *balanced tree with vector* $\mathbf{u}$, $B(\mathbf{u})$, be a rooted tree of depth $m+1$ in which all vertices at distance i from the root have degree u_i, $i = 0, \ldots, m$. The vertices of distance $m+1$ from the root are leaves. The diameter of $B(\mathbf{u})$ is $2m+2$. We have that

$$\begin{aligned} |V(B(\mathbf{u}))| &= 1 + u_0 + u_0(u_1 - 1) + \cdots + u_0(u_1-1)(u_2-1)\cdots(u_m - 1) \\ &\geq u_0 \prod_{i=1}^{m}(u_i - 1). \end{aligned} \tag{1}$$

A complete k-ary tree, T, of depth r is a balanced tree with vector $(k, k+1, \ldots, k+1)$, where $k+1$ is repeated $r-2$ times. For a spider that is not a path, the vertex of maximum degree is the *head* or *center* of the spider; the spider is a union of paths, called *legs*, with one endpoint a leaf and another the head. Observe that if $Q \subseteq T$ then $\text{span}(Q) \subseteq \text{span}(T)$, $\Delta(Q) \leq \Delta(T)$, $\text{diam}(Q) \leq \text{diam}(T)$. For all other standard graph theoretic definition, we refer the reader to [11]. In all the calculations we omit floors and ceilings when their usage is clear from the context.

For a tree T, and a root $r \in V(T)$, define a partial order on $V(T)$ naturally with $v' \leq v$ if v is on the v'-r-path. Intuitively, the closer to the root a vertex is, the greater it is with regards to this partial order. All subtrees are rooted respecting the original order. We say that a subtree is *inherited* by a vertex v if its vertex set consists of all vertices u such that $u \leq v$. The children of a vertex v are those vertices v' adjacent to v with $v' \leq v$. The parent of a vertex v, $v \neq r$ is a vertex u adjacent to v, $v \leq u$. The *inherited subtree depth* of a vertex v in a rooted tree is the largest distance to a vertex u with $u \leq v$, i.e., the depth of the tree inherited by v. For the balanced trees $B(\mathbf{u})$ we always assume that the partial order with respect to this root is implied.

For a given family $\mathcal{T}$ of k-vertex spiders of diameter at most $k/2$ with maximum degrees $\Delta_1, \ldots, \Delta_x$, for some x, define a ***reduced family of spiders*** $\mathcal{T}'' = \mathcal{T}''(\mathcal{T})$ to consist of x balanced spiders, where the i^{th} spider has maximum degree Δ_i and leg length $(k/2)/\Delta_i$, $i = 1, \ldots, x$.

Lemma 1. *For any balanced tree B of diameter at most $k/2$, and any family $\mathcal{T}$ of k-vertex spiders, $B \not\to \mathcal{T}''$ implies that $B \not\to \mathcal{T}$.*

Proof of Lemma 1. Let B be a balanced tree of diameter at most $k/2$. Assume that $Q \subseteq B$, for some $Q \in \mathcal{T}$ with Δ legs of lengths $\ell_1 \geq \ell_2 \geq \cdots \geq \ell_\Delta$. Since B is balanced, it contains a balanced spider with Δ legs of length ℓ_2. Since $\ell_1 \leq k/2$, $\ell_2 \geq (k - \ell_1)/(\Delta - 1) \geq (k/2)/\Delta =: \ell$. The balanced spider with Δ legs of length ℓ is in $\mathcal{T}''$. □

2.1 Construction of a tree avoiding a family of spiders

Let $\mathcal{T}$ be a given family of k-vertex spiders containing S_k and P_k. We will construct a tree of diameter $k/2$, so let $\mathcal{T}'$ be a subfamily of $\mathcal{T}$ consisting only of spiders of diameter at most $k/2$. Let $\mathcal{T}'' = \mathcal{T}''(\mathcal{T}')$ be the reduced family of spiders of maximum degrees Δ_i and leg-length ℓ_i, respectively, $i = 0, \ldots, x$, for some x, $x \leq |\mathcal{T}| - 2$. Let $k - 1 = \Delta_0 > \Delta_1 > \Delta_2 > \cdots > \Delta_x \geq 3$. Note that $1 = \ell_0 < \ell_1 < \ell_2 < \ldots < \ell_x$, $\ell_i = (k/2)/\Delta_i$, and define $\ell_{x+1} := k/4 - 1$. Let

$$\mathbf{u} = (\Delta_x - 1, \underbrace{\Delta_x - 1, \ldots, \Delta_x - 1}_{\ell_{x+1} - \ell_x}, \ldots, \underbrace{\Delta_i - 1, \ldots, \Delta_i - 1}_{\ell_{i+1} - \ell_i}, \ldots, \underbrace{\Delta_0 - 1, \ldots, \Delta_0 - 1}_{\ell_1 - \ell_0})$$

be a vector with a total of $k/4 - 1$ entries. Define the intervals $I_i := \{j \geq 1 \ : \ u_j = \Delta_i - 1\}$, i.e., the set of positions (except for position 0) occupied by $\Delta_i - 1$, for $i = 0, \ldots, x$. Let our desired tree to avoid the family of spiders $\mathcal{T}$ (including S_k and P_k) be

$$T = T_s(\mathcal{T}) := B(\mathbf{u}).$$

Note that, using (1), $|V(T)| \geq \prod_{i=0}^{x} (\Delta_i - 2)^{|I_i|} = \prod_{i=0}^{x} (\Delta_i - 2)^{\ell_{i+1} - \ell_i}$.

Any vertex of degree at least Δ_i in T has inherited subtree depth of less than ℓ_i in T, $i = 1, \ldots, q-2$, so $T \not\to \mathcal{T}''$. By Lemma 1 T avoids the trees of diameter at most $k/2$, and by the fact that T has diameter less than $k/2$, we have that $T \not\to \mathcal{T}$.

This concludes the construction. We shall analyze the size of T in Section 3.

Before beginning the next constuction, we need some more definitions and lemmas:

Let $\mathbf{b}$ be a binary sequence, and D a set of positive integers. We say that $\mathbf{b}$ ***avoids*** D if for any three indices x, y, z such that $b_x = b_y = b_z = 1$ (here x, y, z do not need to be distinct), $|y-x| + |z-y| \notin D$. If a set D consists just of one element d, instead of writing that $\mathbf{b}$ avoids $\{d\}$ we simply write that $\mathbf{b}$ avoids d. We define the *relative frequency* of `1`s of a binary sequence $\mathbf{b}$ in the interval $I = [s, t]$, or *frequency* $\mathrm{freq}(\mathbf{b}, I)$ for short, as the number of `1`s in the sequence $(b_s, b_{s+1}, \ldots, b_t)$, divided by the total amount of integers in the interval, i.e., by $t - s + 1$. For a binary sequence $\mathbf{b}$ and a vector $\mathbf{u}$, define $\mathbf{b} \circ \mathbf{u} := \mathbf{u}'$ as follows: $u'_i = \begin{cases} u_i & \text{if } b_i = 1, \\ 1 & \text{if } b_i = 0 \text{ and } i = 0, \\ 2 & \text{if } b_i = 0 \text{ and } i \neq 0. \end{cases}$

Lemma 2. *For any finite set of positive integers D, and any interval I, there is a binary sequence $\mathbf{b}$ avoiding D, with a frequency at least $4^{-|D|}$ in I.*

Lemma 3. *If a binary sequence $\mathbf{b}$ avoids d, then for any vector $\mathbf{u}$, $d \notin \mathrm{span}(B(\mathbf{b} \circ \mathbf{u}))$.*

We will prove these lemmas formally in the appendix.

2.2 Construction of a tree avoiding a general family of trees

Let $\mathcal{T} = \mathcal{Q}_s \cup \mathcal{Q}_n$, $|\mathcal{Q}_s| = p+2$, $|\mathcal{Q}_n| = q$ be a given family of k-vertex trees, where $\mathcal{Q}_s$ is a family of spiders containing P_k and S_k and $\mathcal{Q}_n$ is a family of non-spiders. Let $T_s = T_s(\mathcal{Q}_s)$ be the tree from Construction 2.1, i.e., a tree avoiding $\mathcal{Q}_s$. We have that T_s is a balanced tree $T = B(\mathbf{u})$, for some vector $\mathbf{u} = (u_0, \ldots, u_{k/4-2})$. We shall construct a tree avoiding $\mathcal{T}$ by trimming T_s in such a way that its span avoids some element of the span of each non-spider in $\mathcal{T}$. For that, we need parameters ι, D and $\mathbf{b}$.

- Choose ι from $i = 0, \ldots, x$ as the index for which the product of elements in the interval I_i, i.e., $\prod_{j \in I_i}(u_j - 1) = (\Delta_i - 2)^{\ell_{i+1} - \ell_i}$, is maximal.
- Let D be a set of representatives of spans of the trees from $\mathcal{Q}_n$, i.e., $|D \cap \mathrm{span}(Q)| \geq 1$ for each $Q \in \mathcal{Q}_n$, and $|D| \leq |\mathcal{Q}_n| = q$.
- Let $\mathbf{b}$ be a binary sequence avoiding D with frequency at least $4^{-|D|}$ in I_ι, guaranteed by Lemma 2.

Finally, let our desired tree be

$$T = T(\mathcal{Q}_s \cup \mathcal{Q}_n) := B(\mathbf{b} \circ \mathbf{u}).$$

By Lemma 3, $\mathrm{span}(T) \cap D = \emptyset$. With that, we get that for any $Q \in \mathcal{Q}_n$ there is a $d \in \mathrm{span}(Q)$ with $d \notin \mathrm{span}(T)$, and therefore T avoids all $Q \in \mathcal{Q}_n$. Since T is a subtree of T_s, and T_s avoids $\mathcal{Q}_s$, we have that T avoids $\mathcal{Q}_s$ too.

We shall analyze the size of T in Section 3.

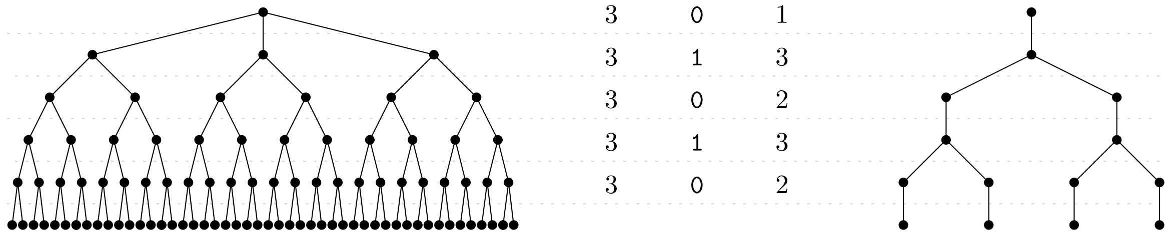

Figure 1: A tree $B(\mathbf{u})$, a sequence $\mathbf{u}$, a sequence $\mathbf{b}$, a sequence $\mathbf{b} \circ \mathbf{u}$, and a tree $B(\mathbf{b} \circ \mathbf{u})$, where the entries of the sequences correspond to the layers of the trees depicted on the same horizontal line. Here $\mathbf{u} = (3, 3, 3, 3, 3)$, $\mathbf{b} = (0, 1, 0, 1, 0)$, $\mathbf{b} \circ \mathbf{u} = (1, 3, 2, 3, 2)$.

3 Proofs

3.1 Proof of Proposition 1 (Star and Path)

Proof of Proposition 1. Observe first that if T is a tree of largest order avoiding S_{k+1} and P_{k+1}, then the longest path in G has length $k-1$, otherwise one can subdivide an edge in a longest path of T to obtain larger such a graph.

If k is odd, then $B(k-1, \ldots, k-1)$ of depth $(k-1)/2$ has diameter $k-1$, maximum degree $k-1$ and $1 + \frac{k-1}{k-3}((k-2)^{(k-1)/2} - 1)$ vertices. Consider some tree T avoiding P_{k+1} and S_{k+1}. Let $\{c\}$ be the center of a longest path P of length $k-1$. Any other vertex of T is at distance at most $(k-1)/2$ from c because otherwise a path of length at least $(k+1)/2$ from c and a sub-path of P with endpoint c together give a longer than P path. Thus T is a tree rooted at c with depth at most $(k-1)/2$ and maximum degree is at most $k-1$, so $|V(T)| \leq 1 + \frac{k-1}{k-3}((k-2)^{(k-1)/2} - 1)$.

If k is even, then two graphs isomorphic to $B(k-2, k-1, \ldots, k-1)$ of depth $(k-2)/2$, linked together at the roots by another edge, form a graph of diameter $k-1$, a maximum degree of $k-1$ and with $2 + 2\frac{k-2}{k-3}((k-2)^{(k-2)/2} - 1)$ vertices. Consider some tree T avoiding P_{k+1} and S_{k+1}. Let $\{r, l\}$ be the center of the longest path P of length at most $k-1$. Define L and R to be the trees rooted at r and l, respectively, and obtained from T by deleting the edge rl. As in the previous case, any $v \in V(L)$ has distance at most $(k-2)/2$ from l, and the degree of l in L is at most $k-2$ within L. So, L, and by symmetry, R are rooted trees of depth at most $(k-2)/2$, the degree of each vertex at most $k-1$ and the degree of the root at most $k-2$. So, $|V(T)| = |V(L)| + |V(R)| \leq 2 + 2\frac{k-2}{k-3}((k-2)^{(k-2)/2} - 1)$. □

3.2 Proof of Theorem 1 (Main Theorem)

Proof of the upper bound

Let $\mathcal{T} = \{S_k, P_k, Q_1, \ldots, Q_p\}$, where Q_i is a balanced spider of maxdegree $\log^i k$ for $i = 1, \ldots, p$. Let T be a tree that avoids all $\mathcal{T}$. We have to show that $|V(T)| \leq 2^{k \log^{p+1} k(1+o(1))}$. Observe first that $\Delta(T) \leq k-2$ and $\mathrm{diam}(T) \leq k-2$. Fix some vertex r to be the root of T and consider the partial order of vertices with respect to this root. We say that a vertex is ***i*-*small*** if its inherited subtree has depth at most $k/\log^{i+1} k$.

Claim: For $i = 0, \ldots, p-1$ a tree inherited by an i-small vertex in T has at most $s_i := 2^{(i+1)k(1+o(1))}$ vertices.

We proceed by induction on i. If v is a 0-small vertex and S is the tree inherited by v, then $\Delta(S) \leq \Delta(T) \leq k$ and $\mathrm{depth}(S) \leq k/\log k$ by definition. Thus $|V(S)| \leq k^{k/\log k} = 2^k$.

Consider an $(i+1)$-small vertex v and its inherited subtree S. Obtain S' from S by removing all i-small vertices. Observe that each of the remaining vertices has inherited subtree depth greater than $k/\log^{i+1} k$ in T. If there is a vertex u in S' of degree at least $\log^{i+1} k$ in S', then each child of u in S' has inherited subtree depth greater than $k/\log^{i+1} k$ in T due to being in S', so u is a center of a copy of Q_{i+1}, a contradiction. Thus $\Delta(S') < \log^{i+1} k$ in S'. As $\mathrm{depth}(S') \leq \mathrm{depth}(S) \leq k/\log^{i+2} k$ by definition, we get that $|V(S')| \leq (\log^{i+1} k)^{k/\log^{i+2} k} = 2^k$. We have that a tree S is a union of S' and i-small trees inherited by children of some vertices of S'. Each vertex in S' has at most k children in S, each of which inherits in S at most one such a subtree of size at most s_i. This yields

$$|V(S)| \leq |V(S')| k s_i \leq k \cdot 2^k \cdot 2^{(i+1)k(1+o(1))} = 2^{(i+2)k(1+o(1))} = s_{i+1},$$

and proves the claim.

Now, we shall consider a tree T and apply an argument almost identical to the one used in the Claim. Delete all $(p-1)$-small vertices from T. The resulting tree T' has maximum degree at most $\log^p k$ and it has depth at most k since $\mathrm{diam}(T) \leq k$. Thus $|V(T')| \leq (\log^p k)^k = 2^{k \log^{p+1} k}$. As before, each vertex of T' has at most k neighbors, each of which inherits in T at most one such a tree of size at most $s_{p-1} = 2^{pk(1+o(1))}$. Thus,

$$|V(T)| \leq |V(T')| k s_{p-1} \leq 2^{k \log^{p+1} k} \cdot k \cdot 2^{pk(1+o(1))} = 2^{k \log^{p+1} k (1+o(1))},$$

concluding the proof of the upper bound.

Proof of the lower bound for spiders

Let $\mathcal{T}$ be a family of $p+2$ spiders including P_k and S_k. Let $T = T_s(\mathcal{T})$ as in Construction 2.1. We know that $T \not\to \mathcal{T}$, so it only remains to analyze the size of T. Recall that the number of spiders in the reduced family x is at most p. Recall further that $\ell_{x+1} := k/4 - 1$ and formally define $\Delta_{x+1} := (k/2)/\Delta_{x+1}$ so the property $\ell_i = (k/2)/\Delta_i$ is fulfilled also for $i = x+1$. Recall that $T = B(\mathbf{u})$, where $\mathbf{u}$ has blocks of indices $I_x, \ldots, I_0$ with entries of $\Delta_i - 1$ in the block corresponding to I_i, $i = 0, \ldots, x$. We bound the number of vertices in T from below by the number of leaves:

$$\begin{aligned}
|V(B(\mathbf{u}))| &\overset{(1)}{\geq} u_0 \prod_{i=1}^{m} (u_i - 1) \\
&\geq \prod_{i=0}^{x} (\Delta_i - 2)^{|I_i|} \\
&= \prod_{i=0}^{x} (\Delta_i - 2)^{\ell_{i+1} - \ell_i} \\
&= \prod_{i=0}^{x} (\Delta_i - 2)^{\frac{k}{2}(\Delta_{i+1}^{-1} - \Delta_i^{-1})}. \qquad (2)
\end{aligned}$$

We compare two monotone sequences Δ_i and $f_i := \log^i k/\log^{i+1} k$, $i = 1, \ldots, x$ and define a corresponding index ι.

- Case 1. $\Delta_1 \leq f_1 = \frac{\log k}{\log\log k}$.

  Then the first term in the product of (2) is $(\Delta_0 - 2)^{\frac{k}{2}(\Delta_1^{-1} - \Delta_0^{-1})} \geq (k-3)^{\frac{k}{2}\left(\frac{\log\log k}{\log k} - \frac{1}{k-1}\right)} = 2^{\frac{k}{2}\log\log k(1+o(1))}$. Set $\iota := 0$.

- Case 2. $\Delta_x \geq f_x = \frac{\log^x k}{\log^{x+1} k}$.

  Then the last term in the product of (2) is $(\Delta_x - 2)^{\frac{k}{4} - 1 - \frac{k}{2}\Delta_x^{-1}} \geq \left(\frac{\log^x k}{\log^{x+1} k} - 2\right)^{\frac{k}{4} - 1 - \frac{k}{2}\frac{\log^{x+1} k}{\log^x k}} = 2^{(\log^{x+1} k - \log^{x+2} k)(\frac{k}{4} - \frac{k}{2}\frac{\log^{x+1} k}{\log^x k})(1+o(1))} = 2^{\frac{k}{4}\log^{x+1} k(1+o(1))}$. Set $\iota := x$.

- Case 3. There is some i, $1 \leq i \leq x-1$ with $\Delta_i \geq f_i$ and $\Delta_{i+1} \leq f_{i+1}$.
  In this case we bound the i^{th} term in the product of (2): $(\Delta_i - 2)^{\frac{k}{2}(\Delta_{i+1}^{-1} - \Delta_i^{-1})} \geq \left(\frac{\log^i k}{\log^{i+1} k} - 2\right)^{\frac{k}{2}\left(\frac{\log^{i+2} k}{\log^{i+1} k} - \frac{\log^{i+1} k}{\log^i k}\right)} = 2^{\frac{k}{2}\log^{i+2} k(1+o(1))}$. Set $\iota := i$.

So, not only we bound the number of vertices in T, thereby showing the lower bound for spiders, but more specifically, we show the following fact that we will need for the general lower bound:

$$\exists \iota \in \{0, 1, \ldots, x\} \qquad (\Delta_\iota - 2)^{|I_\iota|} \geq 2^{\frac{k}{4}\log^{x+1} k(1+o(1))} \geq 2^{\frac{k}{4}\log^{p+1} k(1+o(1))}. \tag{3}$$

Proof of the general lower bound

Start with the family of k-vertex trees $\mathcal{T} = \mathcal{Q}_s \cup \mathcal{Q}_n$, $|\mathcal{Q}_s| = p + 2$, $|\mathcal{Q}_n| = q$, where $\mathcal{Q}_s$ is a family of spiders including P_k and S_k and $\mathcal{Q}_n$ is a family of non-spiders. Consider a tree $T = T(\mathcal{Q}_s \cup \mathcal{Q}_n)$, a corresponding D, $\mathbf{b}$ and ι from Construction 2.2. We shall analyze the size of T. Recall that $\mathbf{u}' := \mathbf{b} \circ \mathbf{u}$. Split up I_ι into I' and I'', where $I' = \{i \in I_\iota \ : b_i = \mathtt{1}\}$ and $I'' = \{i \in I_\iota \ : b_i = \mathtt{0}\}$. From Lemma 2 we have that $\text{freq}(\mathbf{b}, I) \geq 4^{-|D|} \geq 4^{-q}$, so $|I'| \geq |I_\iota| 4^{-q}$. It follows from (3) that $(\Delta_\iota - 2)^{|I_\iota|} \geq 2^{\frac{k}{4}\log^{p+1} k(1+o(1))}$. Using this information, we get the following lower bound (note that $0 \notin I_\iota$ ensuring that the second product is not 0):

$$\begin{aligned}
|V(B(\mathbf{u}'))| &\overset{(1)}{\geq} u'_0 \prod_{i=1}^{m}(u'_i - 1) \\
&\geq \prod_{i \in I_\iota}(u'_i - 1) \\
&= \prod_{i \in I'}(\Delta_\iota - 2) \prod_{i \in I''} 1 \\
&= (\Delta_\iota - 2)^{|I'|} \\
&\geq \left((\Delta_\iota - 2)^{|I_\iota|}\right)^{4^{-q}} \\
&\geq 2^{\frac{k}{4}\log^{p+1} k(1+o(1))4^{-q}}.
\end{aligned}$$

Together with the upper bound from the first part of the proof, this proves the theorem. $\square$

4 Acknowledgements

The authors thank Ignaz Rutter for fruitful discussions and comments on the manuscript.

References

[1] Chung, F. R. K.; Graham, R. L.; Coppersmith, D., On trees containing all small trees. The theory and applications of graphs (Kalamazoo, Mich., 1980), pp. 265–272, Wiley, New York, 1981.

[2] Chung, F. R. K.; Graham, R. L., Universal caterpillars. J. Combin. Theory Ser. B 31 (1981), no. 3, 348–355.

[3] Goldberg, M. K.; Livsic, E. M. Minimal universal trees. (Russian) Mat. Zametki 4 1968 371–379.

[4] Meir, A.; Moon, J. W., On subtrees of certain families of rooted trees. Ars Combin. 16 (1983), B, 305–318.

[5] Milliken, K. R., A partition theorem for the infinite subtrees of a tree. Trans. Amer. Math. Soc. 263 (1981), no. 1, 137–148.

[6] Pach, J.; Solymosi, J.; Tóth, G., Unavoidable configurations in complete topological graphs. U.S.-Hungarian Workshops on Discrete Geometry and Convexity (Budapest, 1999/Auburn, AL, 2000). Discrete Comput. Geom. 30 (2003), no. 2, 311–320.

[7] Reid, K. B.; Wormald, N. C., Embedding oriented n-trees in tournaments. Studia Sci. Math. Hungar. 18 (1983), no. 2-4, 377–387.

[8] Shamir, R.; Tsur, D., Faster Subtree Isomorphism, Journal of Algorithms, Volume 33, Issue 2, (1999), 267280.

[9] Steel, M.; Székely, L., An improved bound on the maximum agreement subtree problem. (English summary) Appl. Math. Lett. 22 (2009), no. 11, 1778–1780.

[10] Su, C.; Miao, B.; Feng, Q., Subtrees of random binary search trees. (Chinese. English, Chinese summary) Chinese J. Appl. Probab. Statist. 22 (2006), no. 3, 304–310.

[11] West, D., Introduction to Graph Theory, Published by Prentice Hall 1996, 2001.

5 Appendix

Proof of Lemma 2. For two binary sequences $\mathbf{b}$ and $\mathbf{b}'$, let $\mathbf{b} \otimes \mathbf{b}'$ be the binary sequence whose i^{th} element is the product of the i^{th}s elements in $\mathbf{b}$ and $\mathbf{b}'$, i.e., element-wise logical "and". For a positive integer s, a shift, $\mathbf{b}^s$ of a sequence $\mathbf{b}$ is defined by $b^s_x := b_{x+s}$, $x = 0, 1, \ldots$. For a periodic binary sequence $\mathbf{b}$ with period p, we define the frequency $\text{freq}(\mathbf{b})$ as $\text{freq}(\mathbf{b}) := \text{freq}(\mathbf{b}, [0, p-1])$.

Recall that for a set of positive integers D, a binary sequence $\mathbf{b}$ avoids D if there are no three indices x, y, z with $b_x = b_y = b_z = \mathtt{1}$ such that $|x - y| + |y - z| = d$ for some $d \in D$. Note that if $\mathbf{b}$ avoids D, then a shift $\mathbf{b}^s$ avoids D as well for any s; if furthermore $\mathbf{b}'$ avoids D', then

$\mathbf{b} \otimes \mathbf{b}'$ avoids $D \cup D'$.

First we shall prove the following claim by induction on m.

Claim 1.
Let $\mathbf{b}_i$ be periodic binary sequences with period p_i and relative frequency $f_i = \text{freq}(\mathbf{b}_i)$, $i = 1, \ldots, m$. Then for every interval I, there are shift values $s_i \in [0, p_i - 1]$, such that $\mathbf{b} = \mathbf{b}_1^{s_1} \otimes \ldots \otimes \mathbf{b}_m^{s_m}$ has frequency $f := \text{freq}(\mathbf{b}, I) \geq \prod_{i=1}^{m} f_i$.

Let l be the number of integers in the interval.
Induction step: Let $\mathbf{b}'$ be the sequence obtained for $\mathbf{b}_1, \ldots, \mathbf{b}_m$, with the period p' and a frequency $f' = \text{freq}(\mathbf{b}', I) \geq \prod_{i=1}^{m-1} f_i$ in the interval I. We have that $\mathbf{b}'$ has $f' \cdot l$ entries `1` in the interval I, and $\mathbf{b}_m$ has $f_m \cdot p_m$ entries `1` in any interval of length p_m.

For each `1` of $\mathbf{b}'$ in I, there are $f_m \cdot p_m$ shifts $s \in [0, p_m - 1]$ such that the `1` from $\mathbf{b}'$ gets matched up with a `1` from $\mathbf{b}_m$. Summing up the amount of `1`s that $\mathbf{b}' \otimes \mathbf{b}_m^s$ has in I over all shifts $s \in [0, p_m - 1]$, we get $f_m \cdot p_m \cdot f' \cdot l$ `1`s in total. That means that on average over all s, there are $f_m \cdot f' \cdot l$ `1`s. So there is at least one shift value s_m such that the interval contains at least $f_m \cdot f' \cdot l$ entries of `1` for $\mathbf{b} := \mathbf{b}' \otimes \mathbf{b}_m^{s_m}$. As the length of the interval is l, this means that the $\text{freq}(\mathbf{b}, I)$ is at least $f_m \cdot f' \geq \prod_{i=1}^{m} f_i$.

Induction base: To get the statement for $m = 1$, apply the induction step to the sequence $\mathbf{b}'$ consisting only of `1`s, and note that $\mathbf{b}_1 = \mathbf{b}_1 \otimes \mathbf{b}'$. For this sequence $\mathbf{b}'$, the statement is obviously true. This proves the Claim 1.

Claim 2. For every d, there is a periodic binary sequence $\mathbf{b}$ with frequency at least $\frac{1}{4}$ and avoiding d.

For $d = 1, 3, 5$, the repeated sequence of `10` serves the purpose, as any two `1`s have a distance divisible by 2 from each other. For $d = 2, 4$, the repeated sequence of `100` serves the purpose, as any two `1`s have a distance divisible by 3 from each other. For $d \geq 6$, consider a periodic binary sequence $\mathbf{b}$ formed by a block of $\lfloor \frac{d-1}{2} \rfloor$ `1`s followed by d `0`s. We show that $\mathbf{b}$ avoids d by showing the contrary is false: Assume that there is a set of indices x, y, z such that $b_x = b_y = b_z = 1$ and $|x - y| + |z - y| = d$. Then we see that all three x, y and z must correspond to the same block of 1s. However, the difference in indices within such a block is at most $\lfloor \frac{d-1}{2} \rfloor$, so $|x - y| + |z - y| \leq 2\lfloor \frac{d-1}{2} \rfloor < d$. This concludes the proof of Claim 2.
Now the lemma follows from these two claims. □

Proof of Lemma 3. Let $B = B(\mathbf{b} \circ \mathbf{u})$. Assume $d \in \text{span}(B)$. Then there is a path P of length d in B with its endpoints x' and z' of degree at least 3. Let y' be the vertex in this path closest to the root of B. If y' is the root, then it must have degree at least 2. If y' is not the root, y' must have degree at least 3, as it is either x' or z', or has two vertices in P adjacent to it. Let x, y and z be the distances of x', y' and z' from the root respectively. It holds that $|x - y| + |y - z| = |x - z| = d$. However due to the degree of x', y' and z' we must have $b(x) = b(y) = b(z) = 1$, showing that b doesn't avoid d. □