On the unavoidability of oriented trees

François Dross, Frédéric Havet

On the unavoidability of oriented trees

François Dross[1] and Frédéric Havet[2]

[1]CNRS, Université Côte d'Azur, Inria, I3S, France
[2]Université Côte d'Azur, CNRS, Inria, I3S, France

March 15, 2021

Abstract

A digraph is *n-unavoidable* if it is contained in every tournament of order n. We first prove that every arborescence of order n with k leaves is $(n + k - 1)$-unavoidable. We then prove that every oriented tree of order n ($n \geq 2$) with k leaves is $(\frac{3}{2}n + \frac{3}{2}k - 2)$-unavoidable and $(\frac{9}{2}n - \frac{5}{2}k - \frac{9}{2})$-unavoidable, and thus $(\frac{21}{8}n - \frac{47}{16})$-unavoidable. Finally, we prove that every oriented tree of order n with k leaves is $(n + 144k^2 - 280k + 124)$-unavoidable.

1   Introduction

A **tournament** is an orientation of a complete graph. A digraph is *n*-**unavoidable** if it is contained (as a subdigraph) in every tournament of order n. Because of transitive tournaments, which are acyclic, n-unavoidable digraphs are acyclic. The **unavoidability** of an acyclic digraph D, denoted by $\mathrm{unvd}(D)$, is the minimum integer n such that D is n-unavoidable. It is well-known that the transitive tournament of order n is 2^{n-1}-unavoidable and thus every acyclic digraph of order n is 2^{n-1}-unavoidable. However, for acyclic digraphs with few arcs better bounds are expected. Special attention has been devoted to **oriented paths** and **oriented trees**, which are orientations of paths and trees respectively.

It started with Rédei's theorem [16] which states that the unavoidability of $\vec{P}_n$, the directed path on n vertices, is n: $\mathrm{unvd}(\vec{P}_n) = n$. In 1971, Grünbaum studied the **antidirected paths**, that is, the oriented paths in which every vertex has either in-degree 0 or out-degree 0 (in other words, two consecutive edges are oriented in opposite ways). He proved [7] that the unavoidability of an antidirected path of order n is n unless $n = 3$ (in which case it is not contained in the directed 3-cycle $\vec{C}_3$) or $n = 5$ (in which case it is not contained in the regular tournament of order 5) or $n = 7$ (in which case it is not contained in the Paley tournament of order 7). In the same year, Rosenfeld [18] gave an easier proof and conjectured that there is a smallest integer $N > 7$ such that $\mathrm{unvd}(P) = |P|$ for every oriented path of order at least N. The condition $N > 7$ results from Grünbaum's counterexamples. Several papers gave partial answers to this conjecture [1, 6, 19] until Rosenfeld's conjecture was verified by Thomason, who proved in [20] that N exists and is less than 2^{128}. Finally, Havet and Thomassé [12], showed that $\mathrm{unvd}(P) = |P|$ for every oriented path P except the antidirected paths of order 3, 5, and 7.

Regarding oriented trees, Sumner (see [17]) made the following celebrated conjecture.

Conjecture 1. *Every oriented tree of order $n > 1$ is $(2n - 2)$-unavoidable.*

The first linear bound was given by Häggkvist and Thomason [8]. Following improvements of Havet [9] and Havet and Thomassé [11], El Sahili [5] used the notion of median order, first used as a tool for Sumner's conjecture in [11], and proved that every oriented tree of order n ($n \geq 2$) is $(3n - 3)$-unavoidable. Recently, Kühn, Mycroft and Osthus [14] proved that Sumner's conjecture is true for all sufficiently large n. Their complicated proof makes use of the directed version of the Regularity Lemma and of results and ideas from a recent paper by the same authors [13], in which an approximate version of the conjecture was proved. In [11], Havet and Thomassé also proved that Sumner's conjecture holds for arborescences. An **in-arborescence**, (resp. **out-arborescence**) is an oriented tree in which all arcs are oriented towards (resp. away from) a fixed vertex called the root. An **arborescence** is either an in-arborescence or an out-arborescence.

If true, Sumner's conjecture would be tight. Indeed, the **out-star** S_n^+, which is the digraph on n vertices consisting of a vertex dominating the $n - 1$ others, is not contained in the regular tournaments of order $2n - 3$. However, such digraphs have many leaves. Therefore Havet and Thomassé (see [10]) made the following stronger conjecture than Sumner's one.

Conjecture 2. *Every oriented tree of order n with k leaves is $(n + k - 1)$-unavoidable.*

This conjecture is sustained by an earlier result of Häggkvist and Thomason [8] establishing the existence of a minimal function $g(k) \leq 2^{512k^3}$ such that every tree of order n with k leaves is $(n + g(k))$-unavoidable. Trees with two leaves are paths, so the above-mentioned results imply that Conjecture 2 is true when $k = 2$ and Ceroi and Havet [4] showed that it holds for $k = 3$. Havet [10] also settled Conjecture 2 for a large class of trees. Finally, Mycroft and Naia [15] proved that almost every tree of order n is n-unavoidable – thus settling Conjecture 2 for almost every tree.

1.1 Our results

In Section 3, we settle Conjecture 2 for arborescences.

Theorem 3. *Every arborescence of order n with k leaves is $(n + k - 1)$-unavoidable.*

Using this result, in Section 4, we derive the following.

Theorem 4. *Every oriented tree of order n with k leaves is $\left\lfloor \frac{3}{2}n + \frac{3}{2}k - 2 \right\rfloor$-unavoidable.*

This result gives us a good bound for trees with few leaves. In particular, it implies Sumner's conjecture for trees in which at most one third of the vertices are leaves.

Corollary 5. *Every oriented tree of order n with at most $\frac{n}{3}$ leaves is $(2n - 2)$-unavoidable.*

Then, in Section 5, we give the following upper bound on the unvoidability of trees, which is good for trees with many leaves.

Theorem 6. *Every oriented tree with $n \geq 3$ vertices and k leaves is $\left\lceil \frac{9}{2}n - \frac{5}{2}k - \frac{9}{2} \right\rceil$-unavoidable.*

Theorems 4 and 6 yield the best bound towards Sumner's conjecture for trees of 'small' order:

Corollary 7. *Every oriented tree of order $n \geq 2$ is $\left\lceil \frac{21}{8}n - \frac{47}{16} \right\rceil$-unavoidable.*

Proof. The value of $\min\left(\frac{3}{2}n + \frac{3}{2}k - 2, \frac{9}{2}n - \frac{5}{2}k - \frac{9}{2} \right)$ is maximal when $\frac{3}{2}n + \frac{3}{2}k - 2 = \frac{9}{2}n - \frac{5}{2}k - \frac{9}{2}$, that is when $k = \frac{6n-5}{8}$. In this case $\frac{3}{2}n + \frac{3}{2}k - 2 = \frac{21}{8}n - \frac{47}{16}$. □

Finally, in Section 6, we dramatically decrease the upper bound on the function $g(k)$ such that every tree of order n with k leaves is $(n + g(k))$-unavoidable by showing the following.

Theorem 8. *Every oriented tree with n nodes ($n \geq 2$) and k leaves is $(n + 144k^2 - 280k + 124)$-unavoidable.*

The above results rely on the notion of local median order (see below). Since a local median order can easily be constructed in polynomial time, all our proofs can be transformed into polynomial-time algorithms for finding an arborescence or an oriented tree in a tournament of the size indicated in the statement.

2 Definitions and preliminaries

Notation generally follows [2]. The digraphs have no parallel arcs and no loops. We denote by $[n]$ the set of integers $\{1, \ldots, n\}$.

Let D be a digraph. If (u, v) is an arc, we say that u **dominates** v and write $u \to v$. The **tail** of (u, v) is u and its **head** is v. For any $W \subseteq V(D)$, we denote by $D\langle W \rangle$ the subdigraph induced by W in D, and we denote by $D - W$ the digraph $D\langle V(D) \setminus W \rangle$.

The disjoint union of p digraphs $D_1, \ldots, D_p$ is denoted by $D_1 + \cdots + D_p$.

Let v be a vertex of D. The **out-neighbourhood** of v, denoted by $N_D^+(v)$, is the set of vertices w such that $v \to w$. The **in-neighbourhood** of v, denoted by $N_D^-(v)$, is the set of vertices w such that $w \to v$. The **out-degree** $d_D^+(x)$ (resp. the **in-degree** $d_D^-(x)$) is $|N_D^+(v)|$ (resp. $|N_D^-(v)|$).

Let A be an oriented tree. The **leaves** of A are the vertices adjacent to (at most) one vertex in D. There are two kinds of leaves: **in-leaves** which have out-degree 1 and in-degree 0 and **out-leaves** which have out-degree 0 and in-degree 1. The set of leaves (resp. in-leaves, out-leaves) of A is denoted by $L(A)$ (resp. $L^-(A)$, $L^+(A)$). Trivially, $L(A) = L^+(A) \cup L^-(A)$.

A **rooted tree** is an oriented tree with a specified vertex called the **root**. If A is a tree and r a vertex of A, we denote by (A, r) the tree A rooted at r. Let A be a rooted tree with root r. The **parent** of a node v in $V(A) \setminus \{r\}$ is the node adjacent to v in the unique path from r to v in A. If u is the parent of v, then v is a **child** of u. If w is on the path from r to v in A, we say that w is an **ancestor** of v and that v is a **descendant** of w.

For the sake of clarity, the vertices of a tree are called **nodes**.

Let $\sigma = (v_1, v_2, \ldots, v_n)$ be an ordering of the vertices of D. An arc (v_i, v_j) is **forward** (according to σ) if $i < j$ and **backward** (according to σ) if $j < i$. A **median order** of D is an ordering of the vertices of D with the maximum number of forward arcs, or equivalently the minimum number of backward arcs. Let us note basic properties of median orders of tournaments whose proofs are left to the reader.

Lemma 9. *Let T be a tournament and $(v_1, v_2, \ldots, v_n)$ a median order of T. Then, for any two indices i, j with $1 \leq i < j \leq n$:*

(M1) $(v_i, v_{i+1}, \ldots, v_j)$ *is a median order of the induced subtournament* $T\langle\{v_i, v_{i+1}, \ldots, v_j\}\rangle$.

(M2) *vertex* v_i *dominates at least half of the vertices* $v_{i+1}, v_{i+2}, \ldots, v_j$, *and vertex* v_j *is dominated by at least half of the vertices* $v_i, v_{i+1}, \ldots, v_{j-1}$. *In particular, each vertex* v_i, $1 \le i < n$, *dominates its successor* v_{i+1}.

A **local median order** is an ordering of the vertices of D that satisfies property (M2).

Let $\sigma = (v_1, \ldots, v_m)$ be a local median order of a tournament T. Let ϕ be an embedding of a tree A in T. It is σ-**forward** if for every terminal interval $I = \{v_i, \ldots, v_m\}$, $|\phi(A) \cap I| < \frac{1}{2}|I| + 1$; it is σ-**backward** if for every initial interval $I = \{v_1, \ldots, v_i\}$, $|\phi(A) \cap I| < \frac{1}{2}|I| + 1$; it is σ-**nice** if it is is both a σ-forward and a σ-backward. For all i and j in $\{1, \ldots, m\}$ with $i < j$ and $\sigma' = (v_i, \ldots, v_j)$, if ϕ' is a σ'-forward, σ'-backward or σ'-nice embedding of a tree A' into $T' = T\langle\{v_i, \ldots, v_j\}\rangle$, we also call it respectively a σ-forward, σ-backward, or σ-nice embedding of A' into T'.

Havet and Thomassé proved the following lemma which, with an easy induction, implies that every out-arborescence of order n is $(2n - 2)$-unavoidable.

Lemma 10 (Havet and Thomassé [11])**.** *Let A be a tree with an out-leaf a. Let T be a tournament and let $\sigma = (v_1, \ldots, v_p)$ be a local median order of T. Every σ-forward embedding of $A - a$ in $T - \{v_{p-1}, v_p\}$ can be extended to a σ-forward embedding of A in T.*

Because we shall employ the idea used to prove it, we give the proof of Lemma 10.

Proof. Assume there exists a σ-forward embedding ϕ of $A - a$ in $T - \{v_{p-1}, v_p\}$. Let b be the in-neighbour of a in A, and let $v_i = \phi(b)$. Since ϕ is σ-forward,

$$|\phi(V(A - a)) \cap \{v_{i+1}, \ldots, v_{p-2}\}| < \frac{1}{2}(p - 2 - i) + 1 = \frac{1}{2}(p - i).$$

Now, by (M2), v_i has at least $\frac{1}{2}(p - i)$ out-neighbours in $\{v_{i+1}, \ldots, v_p\}$. Hence, v_i has an out-neighbour v_j in $\{v_{i+1}, \ldots, v_p\} \setminus \phi(V(A - a))$. Set $\phi(a) = v_j$. One easily checks that ϕ is a σ-forward embedding of A in T. $\qquad\square$

Let $\sigma = (v_1, \ldots, v_m)$ be a local median order of a tournament T. Let F be a set of vertices of T. An embedding ϕ of a tree A in T is σ-F-**nice** if for all terminal interval $I = \{v_i, \ldots, v_m\}$, $|\phi(A) \cap I| < \max(\frac{1}{2}|I| - |F \cap I| + 1, 1)$ and for all initial interval $I = \{v_1, \ldots, v_i\}$, $|\phi(A) \cap I| < \max(\frac{1}{2}|I| - |F \cap I| + 1, 1)$. Note that if $F = \emptyset$, a σ-F-nice embedding is exactly a σ-nice embedding.

Lemma 11. *Let f be a positive integer and A be a tree of order n with root r. Let T be tournament of order $4n + 4f - 3$ with a set F of at most f vertices and let $(v_{-2n-2f+1}, \ldots, v_{2n+2f-1})$ be a local median order of T such that $v_0 \notin F$. There is a σ-F-nice embedding ϕ of A in T such that $\phi(r) = v_0$, and such that for all $a \in V(A)$, $\phi(a) \notin F$.*

Proof. We prove by induction on n, the result holding trivially when $n = 1$.

Assume now that $n \ge 2$. Let a be a leaf of A. By directional duality, we may assume that a is an out-leaf. Let b be the in-neighbour of a in A. Set $p = 2n + 2f - 1$. Let p' be the smallest integer such that $p' = 2n + 2|F \cap \{v_{-p'}, \ldots, v_{p'}\}| - 3$. Note that p' can be obtained by starting with $p' = 2n - 3$ and repeatedly replacing the value of p' by the value

of $2n + 2|F \cap \{v_{-p'}, \ldots, v_{p'}\}| - 3$ until the value of p' remains stable. This process makes p' increase at each step, and since we always have $p' \leq p$ (since $|F| \leq f$), the process terminates. Note that

$$p = p' + 2 + 2|F \cap \{v_{-p}, \ldots, v_{-p'-1}\}| + 2|F \cap \{v_{p'+1}, \ldots, v_p\}| \tag{1}$$

Let $T' = T\langle\{v_{-p'}, \ldots, v_{p'}\}\rangle$. By definition, $\sigma' = (v_{-p'}, \ldots, v_{p'})$ is a local median order of T'. Let $F' = F \cap V(T')$.

By the induction hypothesis, there exists a σ'-F'-nice embedding ϕ of $A - a$ in T' such that $\phi(r) = v_0$ and for all $a' \in V(A - a)$, $\phi(a') \notin F'$. Let $v_i = \phi(b)$. Since ϕ is σ'-F'-nice,

$$|\phi(V(A - a)) \cap \{v_{i+1}, \ldots, v_{p'}\}| < \max\left(\frac{1}{2}(p' - i) - |F \cap \{v_{i+1}, \ldots, v_{p'}\}| + 1, 1\right)$$

Note that for all final interval $I' = \{v_i, \ldots, v_{p'}\}$ of $(v_{-p'}, \ldots, v_{p'})$, we have $|I'| \geq 2|F \cap I'|$, by construction of p' (we add 2 to p' whenever we meet an additional vertex of F). Therefore the maximum in the previous inequation is attained by $\frac{1}{2}(p' - i) - |F \cap \{v_{i+1}, \ldots, v_{p'}\}| + 1$. Thus

$$|\phi(V(A - a)) \cap \{v_{i+1}, \ldots, v_{p'}\}| < \frac{1}{2}(p' - i) - |F \cap \{v_{i+1}, \ldots, v_{p'}\}| + 1$$

Hence, by Equation (1), we get

$$\begin{aligned} |\phi(V(A - a)) \cap \{v_{i+1}, \ldots, v_{p'}\}| \quad < \quad & \frac{1}{2}(p - i) - |F \cap \{v_{i+1}, \ldots, v_{p'}\}| + 1 \\ & -|F \cap \{v_{-p}, \ldots, v_{-p'-1}\}| - |F \cap \{v_{p'+1}, \ldots, v_p\}| - 1 \end{aligned}$$

So $|\phi(V(A - a)) \cap \{v_{i+1}, \ldots, v_{p'}\}| < \frac{1}{2}(p - i) - |F \cap \{v_{i+1}, \ldots, v_p\}|$.

Now, by (M2), v_i has at least $\frac{1}{2}(p - i)$ out-neighbours in $\{v_{i+1}, \ldots, v_p\}$, so at least $\frac{1}{2}(p - i) - |F \cap \{v_{i+1}, \ldots, v_p\}|$ out-neighbours in $\{v_{i+1}, \ldots, v_p\} \setminus F$. Hence, v_i has an out-neighbour v_j in $\{v_{i+1}, \ldots, v_p\} \setminus (\phi(V(A - a)) \cup F)$. Set $\phi(a) = v_j$. One easily checks that ϕ is a σ-F-nice embedding of A in T. $\qquad\square$

3 Unavoidability of arborescences

The aim of this section is to prove Theorem 3. We prove the following theorem which implies it directly by directional duality.

Theorem 12. *Let A be an out-arborescence with n nodes, k out-leaves and root r, let T be a tournament on $m = n + k - 1$ vertices, and let $\sigma = (v_1, v_2, \ldots, v_m)$ be a local median order of T. There is an embedding ϕ of A in T such that $\phi(r) = v_1$.*

Proof. Let us describe a greedy procedure giving an embedding ϕ of A into T. For each node a of A, we fix an ordering $\mathcal{O}_a$ of the children of a. If a vertex v_j of T is the image of a node, we say that it is **hit** and denote its pre-image by a_j; in symbols $a_j = \phi^{-1}(v_j)$.

- Set $\phi(r) = v_1$.

- For $i = 1$ to m, do

  - if v_i is not hit, then skip; we say that v_i has been **skipped**.

– if v_i is hit, then assign the $|N^+(a_i)|$ first not yet hit out-neighbours of v_i in $\{v_{i+1}, \ldots, v_m\}$ to the children of a_i (in increasing order according to $\mathcal{O}_a$).

Assume for a contradiction that this procedure does not yield an embedding of A into T. Then the set F of skipped vertices has cardinality at least k. Let B be the set of embedded nodes at the end of the procedure. Since we only embed a node after its parent, $A\langle B \rangle$ is an out-arborescence. Let L be the set of out-leaves of A that are in B. Since $A\langle B \rangle$ is a sub-arborescence of A, we have $|L| \leq k - 1$.

A node a is said to be **active for** i if $\phi(a) \in \{v_1, \ldots, v_i\}$ and it has a child b that is not embedded in $\{v_1, \ldots, v_i\}$ (i.e. either b is not embedded or $\phi(b) \in \{v_{i+1}, \ldots, v_m\}$).

Consider a vertex v_i in F. There is an active node for i, for otherwise all nodes of A would be embedded (in $\{v_1, \ldots, v_i\}$). Let ℓ_i be the largest index such that a_{ℓ_i} is active for i. Note that by definition of active node $\ell_i < i$. Set $I_i = \{v_j \mid \ell_i < j \leq i\}$.

Claim 12.1. *If $v_i \in F$, then $|I_i \cap F| \leq |I_i \cap \phi(L)|$.*

Subproof. Each out-neighbour of v_{ℓ_i} in I_i is hit for otherwise the procedure would have assigned a child of a_{ℓ_i} to it. Thus $I_i \cap F \subseteq I_i \cap N^-(v_{\ell_i})$ and so

$$|I_i \cap F| \leq |I_i \cap N^-(v_{\ell_i})|. \tag{2}$$

Let v_j be a hit vertex in I_i. By definition of ℓ_i, a_j is not active for i, so its children (if any) are embedded in $\{v_{j+1}, \ldots, v_{i-1}\} \subseteq I_i$. Again, by definition of ℓ_i, all the children of a_j are not active, and so their children (if any) are embedded in I_i. And so on, all descendants of a_j are embedded in I_i and not active. We assign to v_j an out-leaf w_j of A which is a descendant of a_j. We just showed that $\phi(w_j) \in I_i$.

Consider now the vertices of $J = I_i \cap N^+(v_{\ell_i})$. As seen before, they are hit, and the descendants of their pre-images are also embedded in I_i. Moreover, for each $v_j \in J$, the parent of a_j is embedded in $\{v_1, \ldots, v_{\ell_i}\}$ for otherwise, at Step ℓ_i, the procedure would have assigned v_j to an out-neighbbour of a_{ℓ_i} or another active node for i. Hence no vertex of J is the image of an ancestor of another node embedded in J. Consequently, the out-leaves embedded in J are all distinct. Thus

$$|I_i \cap N^+(v_{\ell_i})| \leq |I_i \cap \phi(L)|. \tag{3}$$

Now, by (M2), $|I_i \cap N^-(v_{\ell_i})| \leq |I_i \cap N^+(v_{\ell_i})|$. Together with Equations (2) and (3), this proves the claim. $\diamond$

Claim 12.2. *If $v_i \in F$ and $v_j \in F$, then either $I_i \cap I_j = \emptyset$, or $I_i \subseteq I_j$, or $I_j \subseteq I_i$.*

Subproof. Let $v_i, v_j \in F$ with $i < j$. Assume for a contradiction that $I_i \cap I_j \neq \emptyset$, $I_i \not\subseteq I_j$, and $I_j \not\subseteq I_i$. Then $\ell_i < \ell_j < i$. By definition of ℓ_i, a_{ℓ_j} is not active for i. Thus all its children are embedded in $\{v_1, \ldots, v_i\}$. Since $\{v_1, \ldots, v_i\} \subseteq \{v_1, \ldots, v_j\}$, a_{ℓ_j} is not active for j, a contradiction to the definition of ℓ_j. $\diamond$

Now let M be the set of indices i such that $v_i \in F$ and I_i is maximal for inclusion (i.e. for which there is no j such that $I_i \subseteq I_j$). Since $v_i \in I_i$ for all $v_i \in F$, we have $F \subseteq \bigcup_{i \in M} I_i$. Moreover, by Claim 12.2, the I_i, $i \in M$, are pairwise disjoint. So $|F| = \sum_{i \in M} |I_i \cap F|$. By Claim 12.1, we obtain

$$|F| = \sum_{i \in M} |I_i \cap F| \leq \sum_{i \in M} |I_i \cap \phi(L)| \leq |\phi(L)| = |L| \leq k - 1,$$

a contradiction. This completes the proof. $\qquad\square$

Observation 13. *With the embedding ϕ constructed in the above proof, there is an injection from the set F of skipped vertices into $L^+(A)$ such that every skipped vertex v_i is mapped to an out-leaf whose image precedes v_i in σ.*

Proof. We map the vertices v_i of F to an out-leaf in increasing order according to σ.

If there is an active vertex for i, then by Claim 12.1, $|I_i \cap F| \leq |I_i \cap \phi(L)|$. Hence, there is an out-leaf $f(v_i)$ of A with image in I_i (and thus preceding v_i in σ) that was not assigned earlier to a skipped vertex.

If there is no active vertex for i, then all nodes of A are embedded (in vertices preceding v_i in σ). Since $|F| \leq k - 1$, there exists an out-leaf $f(v_i)$ which is not yet assigned to any skipped vertex. Necessarily, $f(v_i)$ is embedded in a vertex preceding v_i in σ. $\qquad\square$

A **bi-arborescence** is a rooted tree A that is the union of an in-arborescence and an out-arborescence that are disjoint except in their common root, which is also the root of A. Theorem 12 directly implies the following corollary.

Corollary 14. *Let A be a bi-arborescence of order n with k leaves. If A has at least one in-leaf and at least one out-leaf, then A is $(n + k - 2)$-unavoidable. Otherwise A is $(n + k - 1)$-unavoidable.*

4   Unavoidability of trees with few leaves

For any rooted tree (A, r), we partition the arcs into the **upward arcs** (the ones directed away from the root) and the **downward arcs** (the ones directed towards the root). The subdigraph composed only of the upward arcs and the nodes that are in an upward arc is called the **upward forest**, and the subdigraph composed only of the downward arcs and the nodes that are in a downward arc is called the **downward forest**. See Figure 1. Note that neither the upward forest nor the downward forest contain isolated vertices. Moreover, the upward (resp. downward) forest is the disjoint union of out-arborescences (resp. in-arborescences).
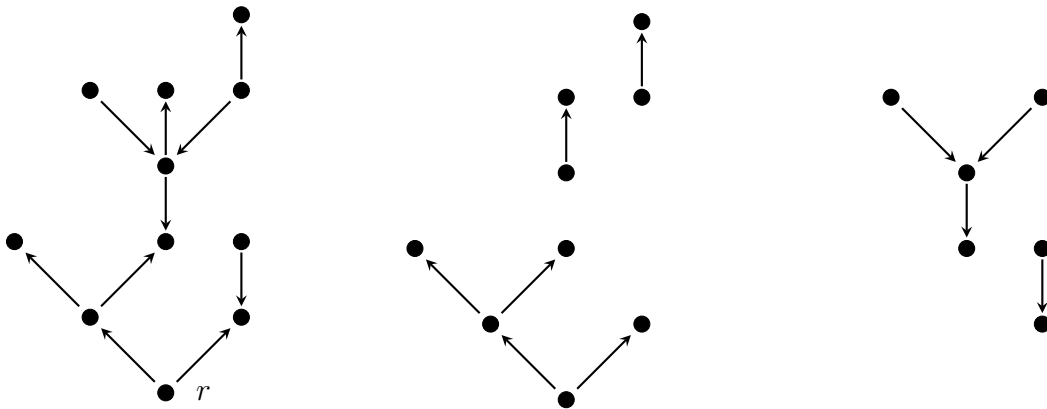

Figure 1: A rooted tree with root r (left), its upward forest (in mid), and its downward forest (right).

The set of components of the upward (resp. downward) forest is denoted by $\mathcal{C}_r^\uparrow(A)$ (resp. $\mathcal{C}_r^\downarrow(A)$). Set $\gamma_r^\uparrow(A) = \sum_{C \in \mathcal{C}_r^\uparrow(A)}(|V(C)| + |L^+(C)| - 2)$ and $\gamma_r^\downarrow(A) = \sum_{C \in \mathcal{C}_r^\downarrow(A)}(|V(C)| + |L^-(C)| - 2)$. When the tree A is clear from the context, we abbreviate $\mathcal{C}_r^\uparrow(A)$ (resp. $\mathcal{C}_r^\downarrow(A)$,

$\gamma_r^\uparrow(A)$, $\gamma_r^\downarrow(A)$) to $\mathcal{C}_r^\uparrow$ (resp. $\mathcal{C}_r^\downarrow$, $\gamma_r^\uparrow$, $\gamma_r^\downarrow$). Since the each component of the upward (resp. downward) forest has at least two vertices and one out-leaf (resp. in-leaf), we have $|V(C)| + |L^+(C)| - 2 > 0$ for all $C \in \mathcal{C}_r^\uparrow$ and $|V(C)| + |L^-(C)| - 2 > 0$ for all $C \in \mathcal{C}_r^\downarrow$.

Proposition 15. *Let (A, r) be a rooted tree of order n with k leaves. Then $\gamma_r^\uparrow + \gamma_r^\downarrow \leq n + k - 2$.*

Proof. By induction on n, the result holding trivially when $n = 1$. Assume now that $n \geq 1$. By directional duality, we may assume that there is a component C^* of the upward forest such that all its out-leaves are leaves of A. Let r^* be the root of C^* (as an out-arborescence). Let $A' = A - (V(C^* - r^*))$ and let n' and k' be the number of nodes and leaves, respectively, of A'. We have $n = n' + |V(C^*)| - 1$. Moreover, all the out-leaves of C^* are leaves of A but not of A^* and r^* is the unique vertex that can be a leaf of A' but not a leaf of A. Hence $k \geq k' + |L^+(C^*)| - 1$. One easily sees that the downward forest of (A', r) is the same as the downward forest of (A', r) and that the upward forest of (A, r) is the disjoint union of the upward forest of A' and C^*. Hence $\gamma_r^\uparrow(A) = \gamma_r^\uparrow(A') + |V(C^*)| + |L^+(C^*)| - 2$ and $\gamma_r^\downarrow(A) = \gamma_r^\downarrow(A')$. By the induction hypothesis, we have $\gamma_r^\uparrow(A') + \gamma_r^\downarrow(A') \leq n' + k' - 2$. All the abobve inequalities yield

$$
\begin{aligned}
\gamma_r^\uparrow(A) + \gamma_r^\downarrow(A) &= \gamma_r^\uparrow(A') + \gamma_r^\downarrow(A') + |V(C^*)| + |L^+(C^*)| - 2 \\
&\leq n' + k' - 2 + |V(C^*)| + |L^+(C^*)| - 2 \\
&\leq n + k - 2.
\end{aligned}
$$

$\square$

Let (A, r) be a rooted tree such that r has in-degree 0. Let us describe how to build an arborescence A' from the rooted tree (A, r), which we call the **equivalent arborescence** of A. Let $C_1, \ldots, C_j$ be the components of the downward forest of (A, r), and for $1 \leq i \leq j$, let n_i be the number of nodes and k_i the number of in-leaves of C_i. For all $i \in \{1, ..., j\}$, do the following. Let f_i be the parent of the root of C_i. Note that f_i exists since the root of A has in-degree 0, and thus is not in the downward forest. Remove all the arcs of C_i, add a set N_i of $k_i - 1$ new nodes, and put an arc from f_i to each new node and to each node of C_i (except to the root of C_i, since that arc already exists). Observe that A' is an out-arborescence rooted in r and since we removed the downward arcs and added only upward arcs. See Figure 2. By construction A' has $n + \sum_{i=1}^j (k_i - 1)$ nodes. Let $i \in \{1, ..., j\}$. The nodes of C_i that are tail of an upward arc in (A, r) are tail of the same upward arc in A', thus they are not leaves in A'. Hence, each in-leaf of C_i either is an in-leaf in A (if it is the tail of no upward arc), or is not an out-leaf in A'. Therefore, in C_i, there are at most $n_i - k_i$ out-leaves of A' that are not in-leaves in A. Recall that the new nodes are also out-leaves. Therefore A' has at most $k + \sum_{i=1}^j (n_i - 1)$ out-leaves.

Lemma 16. *Let (A, r) be a rooted tree with n nodes and k leaves such that r has in-degree 0. Then A is $(n + k - 1 + \gamma_r^\downarrow)$-unavoidable.*

Proof. Let $C_1, \ldots, C_j$ be the components of the downward forest of (A, r), and for $1 \leq i \leq j$, let n_i be the number of nodes and k_i the number of in-leaves of C_i. Let A' be the arborescence equivalent to A. As noted previously, A' has $n + \sum_{i=1}^j (k_i - 1)$ nodes, and at most $k + \sum_{i=1}^j (n_i - 1)$ out-leaves. Moreover, by definition, $\gamma_r^\downarrow = \sum_{i=1}^j (n_i + k_i - 2)$.

Let T be a tournament on $n + k - 1 + \gamma_r^\downarrow$ vertices. By Theorem 12, there is an embedding ϕ of A' into T.
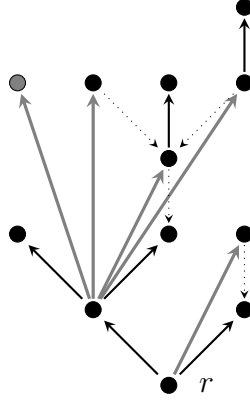
Figure 2: The arborescence equivalent to the rooted tree depicted in Figure 1. Arcs that are removed are dotted, and the arcs and vertices that are added are in gray.

Observe that in the greedy procedure described in the proof of Theorem 12, we do not need to fix the order $\mathcal{O}_a$ before the set of images of the children of a is known. Thus we can effectively choose which child of a is embedded to which vertex with the knowledge of the set of the images of the children of a. So we take advantage of this. Let $i \in \{1, ..., j\}$, and consider $S_i = V(C_i) \cup N_i$. This is a set of $n_i + k_i - 1$ children of f_i in A'. As argued previously, we can know $\phi(S_i)$ before we choose which node of S_i is embedded to which vertex. By Theorem 12, there is an embedding ϕ_i from C_i into $T\langle\phi(S_i)\rangle$. Now for each node a in C_i, we choose $\phi_i(a)$ as its image by ϕ.

Consider now ψ the restriction of the resulting embedding ϕ to $V(A)$. For all $i \in \{1, ..., j\}$, ψ coincides with ϕ_i. Hence ψ preserves the upward arcs since all the upward arcs of (A, r) are in A', and preserves the downward arcs since each downward arc of (A, r) is in some C_i. Therefore ψ is an embedding of A into T. $\square$

We are now able to prove Theorem 4 which states that every oriented tree of order n with k leaves is $(\lfloor \frac{3}{2}n + \frac{3}{2}k \rfloor - 2)$-unavoidable.

Proof of Theorem 4. Let T be a tournament on $\lfloor \frac{3}{2}(n + k) \rfloor - 2$ vertices. Let A be an oriented tree with n nodes and k leaves. Pick a root r such that $\min(\gamma_r^\uparrow, \gamma_r^\downarrow)$ is minimum. By directional duality, we may assume that this minimum is attained by $\gamma_r^\downarrow$.

Suppose for a contradiction that r has an in-neighbour s. The downward forest F_s of (A, s) is obtained from the downward forest F_r of (A, r) by removing the arc (s, r) and possibly s or r if they become isolated. All components of F_r not containing (s, r) are also components of F_s and the component C_0 of F_r containing (s, r) either disappears (when (s, r) is the sole arc of C_0), or loses one vertex (when r or s is a leaf of C_0), or is split into two components having in total as many vertices as C_0 and at most one more in-leaf than C_0. In any case, $\gamma_s^\downarrow < \gamma_r^\downarrow$, a contradiction. Consequently r has in-degree 0.

Since $\gamma_r^\downarrow \leq \gamma_r^\uparrow$, we have $\gamma_r^\downarrow \leq \lfloor \frac{1}{2}(\gamma_r^\uparrow + \gamma_r^\downarrow) \rfloor$. By Proposition 15, $\gamma_r^\uparrow + \gamma_r^\downarrow \leq n + k - 2$ and thus $\gamma_r^\downarrow \leq \lfloor \frac{1}{2}(n + k) \rfloor - 1$. Hence, T has at least $n + k - 1 + \gamma_r^\downarrow$ vertices.

Lemma 16 finishes the proof. $\square$

5 Unavoidability of trees with many leaves

The aim of this section is to establish Theorem 6, which we recall.

Theorem 6. Every oriented tree with $n \geq 3$ vertices and k leaves is $\left\lceil \frac{9}{2}n - \frac{5}{2}k - \frac{9}{2} \right\rceil$-unavoidable.

Proof. Set $m = \left\lceil \frac{9}{2}n - \frac{5}{2}k - \frac{9}{2} \right\rceil$. Let T be a tournament on m vertices. Let A be an oriented tree with n nodes and k leaves. If A is a bi-arborescence, then we have the result by Corollary 14. Henceforth, we assume that A is not a bi-arborescence.

The **out-leaf cluster** of A, denoted by S^+, is the set of nodes of A defined recursively as follows. Each out-leaf A is in S^+; if a is a node with exactly one in-neighbour and all its out-neighbours are in S^+, then a is also in S^+. We similarly define the **in-leaf cluster** S^- of A. Note that $A\langle S^- \rangle$ is a forest of in-arborescences, and $A\langle S^+ \rangle$ is a forest of out-arborescences. Moreover, $S^- \cap S^+ = \emptyset$ because A is not a bi-arborescence.

The **heart** of A, denoted by H, is the tree $A - (S^- \cup S^+)$. Set $n_H = |V(H)|$ and $k_H = |L(H)|$. We first note that each out-leaf of H has an in-neighbour in S^-, since otherwise it would be in S^+. Similarly, each in-leaf of H has an out-neighbour in S^+. In particular, $|S^-| \geq |L^+(H)|$ and $|S^+| \geq |L^-(H)|$.

We first establish a few inequalities. Note that

$$m - (4n - 2k - 3) = \left\lceil \frac{1}{2}(n - k - 1) \right\rceil - 1 \geq 0. \tag{4}$$

Since A is not a bi-arborescence, it follows that A is not a star. Hence $k \leq n - 2$, so

$$m = \left\lceil \frac{9}{2}n - \frac{5}{2}k - \frac{9}{2} \right\rceil \geq \left\lceil 2n + \frac{1}{2} \right\rceil = 2n + 1. \tag{5}$$

Moreover, we may assume that $k > (6n - 5)/8$, since otherwise Theorem 6 follows by Theorem 4. Note that every leaf of A lies in either S^- or in S^+, so $k \leq |S^-| + |S^+|$ and thus

$$n_H = n - |S^-| - |S^+| \leq n - k < n - \frac{6n - 5}{8} \leq \left\lfloor \frac{n}{4} - \frac{5}{8} \right\rfloor. \tag{6}$$

Inequalities (5) and (6) imply that

$$4n_H - 3 < 4\left\lfloor \frac{n}{4} - \frac{5}{8} \right\rfloor - 3 \leq n - 5 \leq m \tag{7}$$

We now describe an algorithm yielding an embedding ϕ of the tree A into T. It proceeds in three phases: in the first phase, we embed the heart of A, in the second phase we embed the out-leaf cluster, and in the third phase we embed the in-leaf cluster. At each step, a node of A is **embedded** if it already has an image by ϕ, and **unembedded** otherwise. If a vertex v_j of T is the image of a node, we denote this node by a_j; in symbols $a_j = \phi^{-1}(v_j)$. We say that a vertex is **hit** if it is the image of a node.

Let $\sigma = (v_1, v_2, \ldots, v_m)$ be a local median order of T. Our algorithm heavily relies on σ to embed A in T. It distinguishes two cases. We first deal with the easier case when one of S^-, S^+ is empty. By directional duality, we may assume that $S^- = \emptyset$. In that case, we proceed only in two phases. (A third phase to embed S^- is useless.) Set $s^+ = |S^+|$. First, by Lemma 11, we find a σ-nice embedding of H in $T\langle \{v_1, \ldots, v_{4n_H - 3}\} \rangle$, which is possible by (7). Then we take an ordering $(s_1, \ldots, s_{s^+})$ of S^+ such that, for every $i \in [s^+]$, s_i is an

out-leaf of $A\langle V(H) \cup \{s_1, \ldots, s_i\}\rangle$. Now for $i = 1$ to s^+, by induction, Lemma 10 yields a σ-forward embedding of $A\langle V(H) \cup \{s_1, \ldots, s_i\}\rangle$ into $T\langle\{v_1, \ldots, v_{4n_H-3+2i}\}\rangle$. At the end, we obtain an embedding of A into $T\langle\{v_1, \ldots, v_{4n_H+2s^+-3}\}\rangle$. We have $4n_H + 2s^+ - 3 = 4n - 2s^+ - 3 \leq 4n - 2k - 3 \leq m$ (by (4)).

Let us now deal with the more complicated case when both S^- and S^+ are non-empty. We first have to find an adequate root of A to start.

Let $\mathcal{C}_r^{\downarrow} = \mathcal{C}_r^{\downarrow}(H)$ and $\mathcal{C}_r^{\uparrow} = \mathcal{C}_r^{\uparrow}(H)$. For a root r of H, let $\beta_r^{\downarrow} = 2|L^-(H)| + 3\sum_{C \in \mathcal{C}_r^{\downarrow}}(|E(C)|)$ and $\beta_r^{\uparrow} = 2|L^+(H)| + 3\sum_{C \in \mathcal{C}_r^{\uparrow}}(|E(C)|)$. Let r be a root that minimizes $\min\{\beta_r^{\downarrow}, \beta_r^{\uparrow}\}$. By directional duality, we may assume that $\beta_r^{\downarrow} = \min\{\beta_r^{\downarrow}, \beta_r^{\uparrow}\}$. Therefore, $\beta_r^{\downarrow} \leq \frac{1}{2}\beta_r^{\downarrow} + \frac{1}{2}\beta_r^{\uparrow} = \frac{3}{2}n_H + k_H - \frac{3}{2}$, i.e.

$$2|L^-(H)| + 3\sum_{C \in \mathcal{C}_r^{\downarrow}}(|E(C)|) \leq \frac{3}{2}n_H + k_H - \frac{3}{2}. \tag{8}$$

We can assume that r has in-degree 0, since each in-neighbour s of r satisfies $\beta_s^{\downarrow} \leq \beta_r^{\downarrow}$.

Let us now detail our algorithm. Let $\ell = n_H - k_H - 1 + \sum_{C \in \mathcal{C}_r^{\downarrow}}(|V(C)| - 1) + 2|L^-(H)| + 2|S^-|$. Note that $\ell \geq 1$ because $|S^-| \geq 1$. Let $p = \ell + n_H + k_H - 1 + \gamma_r^{\downarrow}$.

> <u>Phase 1</u>: We embed H in $T\langle\{v_{\ell+1}, \ldots, v_p\}\rangle$ using the procedure of Lemma 16 for H. Note that in this procedure, we embed the equivalent arborescence H' of H which is bigger than H. Here we keep all vertices of H' embedded until the end of Phase 2.

> <u>Phase 2</u>: While there is an unembedded node in S^+, let i be the smallest integer such that $\phi^{-1}(v_i)$ has an unembedded out-neighbour in S^+, and take the first (i.e. with lowest index) out-neighbour of v_i in $\{v_{i+1}, \ldots, v_m\}$ that is not yet hit and assign it to an unembedded out-neighbour in S^+.

> Unembed all vertices of $V(H') \setminus V(H)$.

> <u>Phase 3</u>: While there is an unembedded node in S^-, let i be the largest integer such that $\phi^{-1}(v_i)$ has an unembedded in-neighbour in S^-, and take the last (i.e. with highest index) in-neighbour of v_i in $\{v_1, \ldots, v_{i-1}\}$ that is not yet hit and assign it to an unembedded in-neighbour in S^-.

Let us prove that this algorithm embeds all nodes of A. First, by Lemma 16, all nodes of H are embedded in Phase 1.

Let us now prove that all nodes of S^+ are embedded in Phase 2. Let B be the subtree of A induced by $V(H) \cup S^+$ and let B' be the out-arborescence obtained from B by replacing H by the equivalent arborescence H'. Observe that Phase 1 and Phase 2, may be seen as embedding B' and extracting a copy of B from B' at the same time. Let us show that our algorithm embeds the whole B' (and thus the whole B) in Phases 1 and 2. The equivalent arborescence H' has $n_H + \sum_{C \in \mathcal{C}_r^{\downarrow}}(|L^-(C)| - 1)$ nodes. Thus B' has $n_H + \sum_{C \in \mathcal{C}_r^{\downarrow}}(|L^-(C)| - 1) + |S^+|$ nodes.

All the leaves of A are either in S^- or in S^+, thus $k \leq |S^-| + |S^+|$. Therefore

$$m \geq \frac{9}{2}n - \frac{5}{2}k - \frac{9}{2} \geq \frac{9}{2}n_H + 2|S^-| + 2|S^+| - \frac{9}{2}. \tag{9}$$

For all $C \in \mathcal{C}_r^{\downarrow}$, we have $|L^-(C)| \leq |V(C)|$, thus

$$\sum_{C \in \mathcal{C}_r^{\downarrow}}(|V(C)| + 2|L^-(C)| - 3) + 2|L^-(H)| \leq \sum_{C \in \mathcal{C}_r^{\downarrow}}(3|V(C)| - 3) + 2|L^-(H)|. \tag{10}$$

Inequalities (8) and (10) yield

$$\sum_{C \in \mathcal{C}_r^{\downarrow}} (|V(C)| + 2|L^-(C)| - 3) + 2|L^-(H)| \le \frac{3}{2} n_H + k_H - \frac{3}{2}. \qquad (11)$$

Inequalities (8) and (11) yield

$$m \ge 3n_H - k_H - 3 + \sum_{C \in \mathcal{C}_r^{\downarrow}} (|V(C)| + 2|L^-(C)| - 3) + 2|L^-(H)| + 2|S^-| + 2|S^+|\,,$$

so with the definitions of ℓ and B', we have $m - \ell \ge 2n_H + 2 \sum\limits_{C \in \mathcal{C}_r^{\downarrow}} (|L^-(C)| - 1) + 2|S^+| - 2 \ge$

$2|B'| - 2$.

Each time we embed a node of B' during Phases 1 and 2, our procedure takes the first (i.e. with lowest index) out-neighbour of a vertex v_i in $\{v_{i+1}, \ldots, v_m\}$ that is not yet hit and assigns it to an unembedded out-neighbour of $\phi^{-1}(v_i)$. Therefore, at each step, ϕ is a σ-forward embedding of B'', the so far constructed sub-out-arborescence of B', into $T\langle\{v_{\ell+1}, ..., v_{\ell+2|B''|-2}\}\rangle$. Thus, as in Lemma 10, every vertex v_i has an out-neighbour in $\{v_{i+1}, ..., v_{\ell+2|B''|}\} \setminus \phi(B'')$ and the procedure can continue. Hence, B' can be embedded into $T\langle\{v_{\ell+1}, ..., v_m\}\rangle$.

Assume for a contradiction that the algorithm fails in Phase 3, which means a node a in S^- is not embedded. We can choose such a node a whose out-neighbour b is embedded. Let v_i be the image of b. Observe that b is in $S^- \cup V(H)$, so it has been embedded in Phase 1 or 3, and necessarily must be in $\{v_1, \ldots, v_p\}$.

Consider the moment when we try to embed the in-neighbours of b during Phase 3. Let hit be the number of vertices of $\{v_1, \ldots, v_{i-1}\}$ that are hit at this moment. Since a is not embedded, we have hit $\ge |N^-(v_i) \cap \{v_1, \ldots, v_{i-1}\}| - |N_A^-(b)| + 1$. By (M2), $|N^-(v_i) \cap \{v_1, \ldots, v_{i-1}\}| \ge \frac{i-1}{2}$. So

$$\text{hit} \ge \frac{i-1}{2} - |N_A^-(b)| + 1. \qquad (12)$$

Let us give some upper bounds on hit. Let $O_{<i}$ be the set of out-leaves of H embedded at some v_j with $\ell + 1 \le j < i$ and let $O_{\ge i}$ be the set of out-leaves of H embedded at some v_j with $i \le j \le p$. We have hit $= \text{hit}_2 + \text{hit}_3$, where hit_2 (resp. hit_3) is the number of vertices of $\{v_1, \ldots, v_{i-1}\}$ that are hit in Phase 1 and 2 (resp. Phase 3 until the considered moment).

At the considered moment, the algorithm has yet to embed the in-neighbours of b and the in-neighbours in S^- of the nodes embedded at each v_j for $j < i$. As noted previously, each out-leaf of H has an in-neighbour in S^-. Therefore, each out-leaf of $O_{<i}$ has an in-neighbour in S^- that is not yet embedded. Hence

$$\text{hit}_3 \le |S^-| - |O_{<i}| - |N_A^-(b)|. \qquad (13)$$

Consider now the embedding of H. It is made using the procedure of Lemma 16, which applies the procedure of Theorem 12 on H'. Let $k_{H'}$ be the number of out-leaves of H'. All the out-leaves of H are also out-leaves in H'. Moreover, by Observation 13, we can map each skipped vertex to an out-leaf of H' whose image precedes the skipped vertex in σ. But there are $k_{H'} - 1$ skipped vertices, therefore each out-leaf of $O_{\ge i}$ except at most one corresponds to two vertices v_j with $j \ge i$. Thus, there are at least $2|O_{\ge i}| - 1$ vertices

in $\{v_i, \dots, v_p\}$. Moreover, $\sum_{C \in \mathcal{C}_r^\downarrow}(|L(C)| - 1)$ vertices of $H' - V(H)$ are unembedded at the end of Phase 2 and were neither out-leaves of H nor skipped vertices. Hence,

$$\text{hit}_2 \leq p - \ell - \sum_{C \in \mathcal{C}_r^\downarrow}(|L(C)| - 1) - 2|O_{\geq i}| + 1 = n_H + k_H + \sum_{C \in \mathcal{C}_r^\downarrow}(|V(C)| - 1) - 2|O_{\geq i}|. \quad (14)$$

Since all vertices hit in Phases 1 and 2 are in $\{v_{\ell+1}, \dots, v_m\}$, we trivially have

$$\text{hit}_2 \leq i - \ell - 1 \quad (15)$$

Summing 2 Eq. (12) + 2 Eq. (13) + Eq. (14) + Eq. (15) yields:

$$\begin{aligned}
\ell &\leq n_H + k_H + \sum_{C \in \mathcal{C}_r^\downarrow}(|V(C)| - 1)) - 2|O_{\geq i}| + 2|S^-| - 2|O_{<i}| - 2 \\
&\leq n_H + k_H - 2L^+(H) + \sum_{C \in \mathcal{C}_r^\downarrow}(|V(C)| - 1)) + 2|S^-| - 2 \\
&= n_H - k_H - 2 + 2(L^-(H)) + \sum_{C \in \mathcal{C}_r^\downarrow}(|V(C)| - 1)) + 2|S^-| \\
&= \ell - 1,
\end{aligned}$$

a contradiction. $\qquad\square$

6 Unavoidability of trees with very few leaves

The aim of this section it to prove Theorem 8 which states that every oriented tree with n nodes and k leaves is $(n + 144k^2 - 280k + 124)$-unavoidable. Since the result holds for paths, we shall only consider trees that are not paths.

Let A be a tree which is not a path. A **branch-node** of A is a node with degree at least 3 and a **flat node** is a node with degree 2. A **bare path** in A is a subpath whose origin is a branch-node, whose terminus is either a branch-node or a leaf, and whose internal nodes are flat nodes. If its terminus is a branch-vertex, then the bare path is an **inner path**; otherwise it is an **outer path**. The **opposite** of an inner path S, denoted by $\overline{S}$, is the inner path with origin the terminus of S and terminus the origin of S. The **blocks** of a directed path P are the maximal directed subpaths of P.

A **stub** is a tree such that :

(i) every inner path has at most three blocks; moreover, if it has three blocks then its first and third block have length 1, and if it has two blocks then one of them has length 1.

(ii) every outer path has length 1.

Our proof of Theorem 8 involves two steps. We first prove the following lemma, which shows that it is sufficient to concentrate on stubs.

Lemma 17. *If there exists a function f such that every stub of order n and $k \geq 6$ leaves is $(n + f(k))$-unavoidable, then every tree of order n with $k \geq 3$ leaves is $(n + \max\{f(2k - 2b) + b \mid 0 \leq b \leq k - 3\})$-unavoidable.*

We then prove the following result on the unavoidability of stubs.

Lemma 18. *Every stub with n nodes and $k \geq 6$ leaves is $(n + 36k^2 - 140k + 124)$-unavoidable.*

Theorem 8 follows directly from Lemmas 17 and 18.

6.1 Reducing to stubs

6.1.1 Toolbox

Let $P = (x_1, \ldots, x_n)$ be a path. We say that x_1 is the **origin** of P and x_n is the **terminus** of P; x_1 and x_n are the **ends** of P. If $x_1 \to x_2$, P is an **out-path**, otherwise P is an **in-path**. The **directed out-path** of order n is the path $P = (x_1, \ldots, x_n)$ in which $x_i \to x_{i+1}$ for all $i \in \{1, \ldots, n-1\}$; the dual notion is that of a **directed in-path**. The **length** of a path is its number of edges. An ℓ-**out-path** (resp. ℓ-**in-path**) is an out-path (resp. in-path) of length ℓ. We denote the path $(x_2, \ldots, x_n)$ by $^*P$.

We enumerate the blocks of P from the origin to the terminus. The first block of P is denoted by $B_1(P)$ and its length by $b_1(P)$. Likewise, the ith block of P is denoted by $B_i(P)$ and its length by $b_i(P)$. The path P is uniquely specified by the signed sequence $sgn(P)(b_1(P), b_2(P), \ldots, b_k(P))$, called its **type**, where k is the number of blocks of P and $sgn(P) = +$ if P is an out-path and $sgn(P) = -$ if P is an in-path.

Thomason [20] proved the following two theorems. See also [12] for a short proof of the first one.

Theorem 19 (Thomason [20]). *Let P be an oriented path of order n. Let T be a tournament of order $n+1$ and X a set of $b_1(P) + 1$ vertices. There exists a copy of P in T with origin in X.*

Theorem 20 (Thomason, [20]). *Let P be a non-directed path of order n with first and last block of length 1. Let T be a tournament of order $n+2$ and X and Y be two disjoint subsets of T of order at least 2.*
If $P \neq \pm(1, 1, 1)$, then there is a copy of P in T with origin in X and terminus in Y.

The idea to find a tree A in a tournament T is to break some bare paths S, that is to remove the arcs and internal vertices of some subpaths R_S satisfying the hypothesis of Theorem 20 if S is an inner path and of Theorem 19 if S is an outer path. Then we find the resulting forest A' in T. Finally, we reconstruct the broken bare paths using Theorems 19 and 20. However, those theorems prescribe the origin and terminus not in a vertex but in a set of two vertices. Therefore, we need to have a little more than the paths of $S - R_S$ to reconstruct S. This is captured by the notion of fork.

The **fork** F of type $\tau = sgn(F)(b_1(F), b_2(F), \cdots, b_k(F))$ is the tree with vertex set $\{x_1, \ldots, x_{n-2}, p_1, p_2\}$ such that $(x_1, \ldots, x_{n-2}, p_1)$ and $(x_1, \ldots, x_{n-2}, p_2)$ are paths of type τ. The vertex x_1 is the **origin** and p_1 and p_2 are the **points** of the fork.

Let P be a path of length at least 2. Its **stump type** is

(i) $sgn(P)(b_1(P) - 1)$ if $b_1(P) \geq 2$, and the type of P is not $+(p, 1, q)$ (with $q \geq 2$),

(ii) $sgn(P)(b_1(P))$ if P is of type $+(p, 1, q)$ (with $q \geq 2$),

(iii) $sgn(P)(1)$ if $b_1(P) = b_2(P) = 1$ and P is not of type $\pm(1, 1, 1, p)$ (with $p \geq 2$) or $+(1, 1, 1, 1, 1)$,

(iv) $sgn(P)(1, 1)$ if P is of type $\pm(1, 1, 1, p)$ (with $p \geq 2$) or $+(1, 1, 1, 1, 1)$,

(v) $sgn(P)(1, b_2(P) - 1)$ if $b_1(P) = 1$ and $b_2(P) \geq 2$.

6.1.2   Proof of Lemma 17

Let A be a tree of order n with k leaves. An inner path is **unbreakable** if either it is directed, or it has precisely two blocks at least one of which has length 1, or it has precisely three blocks with the first and the third of length 1. Otherwise it is **breakable**. We construct an oriented forest B from A by applying the following two operations.

1. For each outer path S of A of length at least 2 and with origin x, we replace S by a fork F_S with origin x whose type is the stump type of S. The **remainder** of S, denoted by R_S, is the path obtained from S by removing the $|F_S| - 1$ first vertices of S. Note that R_S has first block of length 1. See Figure 3.
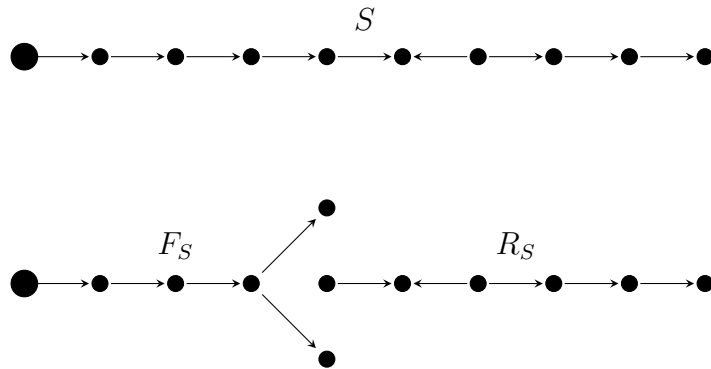

Figure 3: An outer path S of length at least 2, the fork F_S, and the remainder R_S.

2. For each breakable inner path S with origin x and terminus y, replace S by a fork F_S with origin x whose type is the stump type of S, and a fork $F_{\overline{S}}$ with origin y whose type is the stump type of $\overline{S}$. The **remainder** of S, denoted by R_S, is the path obtained from S by removing the $|F_S| - 2$ first and the $|F_{\overline{S}}| - 2$ last vertices of S. Note that R_S is not directed and has first and last blocks of length 1. See Figure 4.
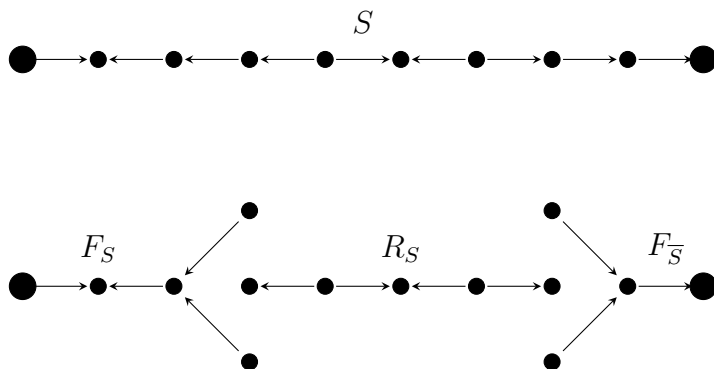

Figure 4: A breakable inner path S, the forks F_S and $F_{\overline{S}}$, and the remainder R_S.

Let b be the number of breakable inner paths and a be the number of outer paths of length 1. Note that in a tree with at least three leaves and no vertices of degree 2, the number of edges non-incident to a leaf is at most three less than the number of leaves.

Therefore $0 \leq b \leq k - 3$. The forest B has $b + 1$ components. By construction, B has at most $2k + 4b - a$ leaves, because Operation 1 replaces the leaf of each outer path of length at least 2 by the two leaves of a fork, and Operation 2 creates four leaves when a breakable inner path is broken by Operation 2. Moreover $|B| \leq |A| + b$, because $|F_S| \leq |S|$ for each outer path S of length at least 2, and $|F_S| + |F_{\overline{S}}| \leq |S| + 1$ for every breakable inner path. Finally, observe that every component C of B has at least $6 - a_C$ leaves, where a_C is the number of outer paths of length 1 of A contained in C. Thus a component has at most $2k - 2b$ leaves.

Let A_i, $i \in [b + 1]$, be the components of B. Note that the forks that are added in B are of specific types. Specifically, they are composed of a path with at most two blocks, at most one of which has length more than 1, and two paths of length 1. Hence from A, every breakable inner path, i.e. every inner path that would have broken Property (i) of the definition of a stub was removed and replaced by parts that do not break it. Therefore every A_i verifies Property (i) of the definition of a stub. Moreover, in every A_i, every outer path either was of length 1 in A, or is the result of the addition of a fork, in which case it is also of length 1. Therefore each A_i is a stub.

Thus, by hypothesis, each A_i is $(|A_i| + f(2k - 2b))$-unavoidable, and so B is $(|B| + f(2k - 2b))$-unavoidable. Let T be a tournament of order $n + \max\{f(2k - 2b) + b \mid 0 \leq b \leq k - 3\}$. T contains B as a subdigraph. We can now transform B into A as follows.

0. Initialize A^* to B.

1. For each breakable inner path S, let U be a set of $|R_S| - 2$ vertices in $T - A^*$. Let X_S be the set of points of F_S and $X_{\overline{S}}$ be the set of points of $\overline{S}$. Note that the stump types are defined in such a way that R_S is not of type $\pm(1, 1, 1)$. Since R_S has first and last block of length 1, by Theorem 20, in $T\langle U \cup X_S \cup X_{\overline{S}} \rangle$, there is a copy R_S^* of R_S with origin in X_S and terminus in $X_{\overline{S}}$. Remove from A^* the point of F_S which is not the origin of R_S^*, and the point of $F_{\overline{S}}$ which is not the terminus of R_S^*; add the path R_S^* to A^*.

2. For each outer path S, let U be a set of $|R_S| - 1$ vertices in $T - A^*$. Let X_S be the set of points of F_S. Since R_S has first block of length 1, by Theorem 19, in $T\langle U \cup X_S \rangle$, there is a copy R_S^* of R_S with origin in X_S. Remove from A^* the point of F_S which is not the origin of R_S^*, and add the path R_S^* to A^*.

Step 1 of this procedure reconstructs the breakable inner paths (which were broken) and Step 2 completes the outer paths of length at least 2. Therefore at the end of the procedure A^* is the desired tree A. This completes the proof of Lemma 17.

6.2 Unavoidability of stubs

The aim of this subsection is to prove Lemma 18.

Informally, the strategy is to treat separately the parts with vertices of degree at least 3, that we will call *islands*, which are connected by directed bare paths. We are going to embed the islands in some specific parts of the tournament, and make sure that we can connect them with the directed bare paths. To embed a directed bare path we only need as many vertices as there are nodes in the path. Moreover, the number of vertices of degree at least three is upper bounded in terms of the number of leaves in the stub, therefore we are allowed to miss a portion of the vertices when we embed the islands.

We need some preliminary results. On the one hand, we need to make sure that, while embedding a directed bare path, we can skip a specific part of the tournament (that will be used later to embed some other parts of the stub) without missing too many vertices. This will be done thanks to Lemma 22. On the other hand, we need to be able to embed several directed bare paths, that come from the same island and go towards the same direction, without missing any vertex. That will be done thanks to Lemma 23. We also note that we will need to avoid some specific vertices that may already be the image of a node by the embedding. Those will correspond to the internal vertices in previous calls to Lemma 22.

6.2.1 Toolbox

We first need a lemma to ensure that, in a tournament, we can find a directed out-path of length 2 from a vertex to another vertex within a certain range, while avoiding some specific vertices. The possibility of avoiding some specific vertices will be done by insuring that there are enough internally disjoint such paths with distinct termini.

Lemma 21. *Let k be a positive integer. Let T be tournament of order $m \geq 4k$ and let $(v_1, \ldots, v_m)$ be a local median order of T. There are at least k internally disjoint directed 2-out-paths with origin v_1 and terminus in $\{v_{m-4k+2}, \ldots, v_m\}$.*

Proof. Let $S = \{v_{m-4k+2}, \ldots, v_{m-1}\}$, and $M = \{v_2, \ldots, v_{m-4k+1}\}$. If v_1 dominates at least k vertices in S, then the paths (v_1, v_i, v_{i+1}) for $v_i \in N^+(v_1) \cap S$ give the result. Therefore, we may assume that v_1 dominates at most $k-1$ vertices in S. Now by (M2), v_1 dominates at least $\frac{m}{2} - 1$ vertices of $\{v_2, \ldots, v_{m-1}\}$, and so at least $\frac{m}{2} - k$ in M. Again by (M2), v_{m-4k+2} is dominated by at least $\frac{m-4k}{2}$ vertices of $\{v_2, \ldots, v_{m-4k+1}\}$. Thus $A = N^+(v_1) \cap N^-(v_{m-4k+2}) \cap M$ has cardinality at least $\frac{m}{2} - k + \frac{m-4k}{2} - (m - 4k) = k$. Hence the paths (v_1, v_i, v_{m-4k+1}) for $v_i \in A$ give the result. $\square$

In the previous lemma, the termini of the paths may be the same (namely, v_{m-4k+2}). However, by applying it successively for k' from 1 to k, we directly obtain the stronger following lemma:

Lemma 22. *Let k be a positive integer. Let T be tournament of order $m \geq 4k$ and let $(v_1, \ldots, v_m)$ be a local median order of T. There are at least k internally disjoint directed 2-out-paths with origin v_1 and distinct termini in $\{v_{m-4k+2}, \ldots, v_m\}$.*

Now we need a lemma that enables us to embed several directed out-paths with specific origins and such that the number of missed vertices does not depend on the length of the paths. It is expressed more generally in terms of arborescences instead of directed paths.

Lemma 23. *Let f, p, s be positive integers with $s > p$. For $q \in [p]$, let A_q be an out-arborescence with n_q nodes and k_q out-leaves. Let T be a tournament of order $m = s + \sum_{q \in [p]} (n_q + k_q - 1) + 2f - 1$, $\sigma = (v_1, \ldots, v_m)$ be a local median order of T and F a set of at most f vertices of T. If there are indices $1 \leq i_1 < \cdots < i_p \leq s$ such that $v_{i_q} \notin F$ for all $q \in [p]$, then there is an embedding ϕ of $B = A_1 + \cdots + A_p$ in T such that the root of A_q is embedded at v_{i_q} for all $q \in [p]$, and $\phi(a) \notin F$ for all $a \in V(B)$.*

Proof. Let us build an arborescence A' on which to apply Theorem 12. Add to B a node a with an out-going arc to each of the roots of the A_q's, and another node b with an out-going arc to a and to $s - p$ new nodes, $a_1, \ldots, a_{s-p}$. Order the children of b in the order $(a, a_1, \ldots, a_{s-p})$.

Now let T' be the tournament obtained from T by adding a transitive tournament S on $s-p+2$ vertices $v_{p-s-1}, v_{p-s}, \ldots, v_0$ in the transitive order. Let all these vertices except v_{p-s} dominate all vertices of T, and let v_{p-s} dominate $\{v_{i_q} \mid q \in [p]\} \cup \{v_i \mid s+1 \le i \le m\}$ and be dominated by the $s-p$ other vertices.

Note that A' has $n' = s-p+2+\sum_{q\in[p]} n_q$ nodes and $k' = s-p+\sum_{q\in[p]} k_q$ out-leaves, and that T' has $m+s-p+2 = n'+k'-1+2f$ vertices. The idea is to embed A' into T' using Theorem 12 with the ordering $\sigma' = (v_{p-s-1}, \ldots, v_m)$. We thus need to show that σ' is a local median order.

Claim 23.1. σ' *is a local median order.*

Subproof. We need to prove that σ' has property $(M2)$. Let $i < j$ be two integers in $\{v_{p-s-1}, \ldots, v_m\}$.

Let us first show that v_i dominates at least half of the vertices $v_{i+1}, \ldots, v_j$. If $i > 0$, it follows from the fact that σ is a local median order. If $i \le 0$ and $i \ne p-s$, then v_i dominates all the vertices $v_{i+1}, \ldots, v_j$ by construction. If $i = p-s$, it holds because v_{p-s} has $s-p$ out-neighbours in $\{v_{p-s+1}, \ldots, v_0\}$ and at most $s-p$ in-neighbours with positive index.

Let us first show that v_j is dominated by at least half of the vertices $v_i, \ldots, v_{j-1}$. If $j \le 0$, then by construction it is dominated by all vertices $v_i, \ldots, v_{j-1}$. Assume now that $j > 0$. If $i > 0$, it follows from the fact that σ is a local median order. If $i \le 0$, then v_j is dominated by at least half the vertices of $\{v_1, \ldots, v_{j-1}\}$ because σ is a local median order, it is dominated by v_0, and dominates at most one vertex (namely v_{p-s}) with non-positive index. Therefore if dominates at least half the vertices of $\{v_i, \ldots, v_{j-1}\}$. $\diamond$

Consequently, following the procedure of Theorem 12, we get an embedding ϕ of A' into T'. This embedding may however embed some nodes at vertices of F. For each vertex $r \in F$ in order, do the following: if a node c of A' is embedded at r, let d be the parent of c in A'; add a leaf in A' with parent d, and put it just before c in the order of the children of d; reapply Theorem 12 to obtain an embedding of the new version of A'. Note that in this construction, we add at most f vertices and f leaves to A'. Therefore the new arborescence has at most $n'+f$ vertices and at most $k'+f$ leaves. Recall that T' has $n'+k'-1+2f$ vertices. Thus Theorem 12 is still applicable, and in the resulting embedding, the only nodes that are embedded at vertices of F are not in A. With the right order on the neighbours of b (namely $(a, a_1, \ldots, a_{s-j})$), the algorithm of Theorem 12 maps b to v_{p-s-1}, a to v_{p-s}, a_i to v_{p-s+i} for all $i \in [s-p]$, and the root of A_q to v_{i_q} for all $q \in [p]$. Hence, the embedding ϕ restricted to the vertices of B is the desired embedding of B in T. $\square$

Observation 24. *Note that the knowledge that a vertex belongs to F in the previous lemma is only needed when we reach it. So the set F does not need to be fixed at the beginning of the procedure but can be exposed gradually.*

6.2.2 Proof of Lemma 18

Let A be a stub with n nodes and k leaves.

Let B be the forest obtained from A by removing the arcs and the internal vertices of the maximal directed paths of length at least 3 contained in its bare paths. The components of B are called the **islands** of A. Note that each island of A contains at least one branch-node of A. See Figure 5. Note moreover that there are at most $k-2$ islands
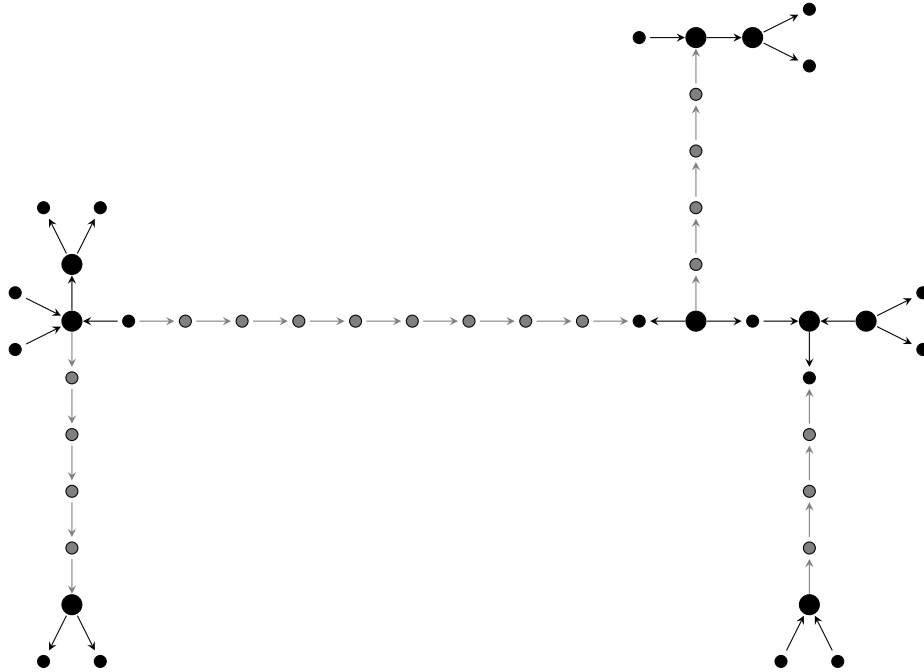
Figure 5: A stub: its islands are in black; the arcs and internal vertices of the maximal directed paths of length at least 3 in bare paths are in gray.

in A (since $k \geq 6$). Let $\widehat{B}$ be the digraph whose vertices are the islands of A, and such that there is an arc from C to C' in $\widehat{B}$ if and only if in A there is a directed bare out-path with origin in C and terminus in C'. For all arc e of $\widehat{B}$, we denote that directed bare out-path by $P(e)$. Observe that B is a forest and $\widehat{B}$ is a tree. See Figure 6.

Choose an island C_1 that has indegree 0. Take C_1 as the root of $\widehat{B}$. The notions of parent and child in the rooted tree $\widehat{B}$ are the classical ones. There is an ordering $(C_1, \ldots, C_r)$ of the islands of A such that

  (i) if $C_p \to C_q$ then $p \leq q$;

 (ii) for each island C, there exist p_C and q_C such that an island C_p is a descendant of C in $\widehat{B}$ if and only if it verifies $p_C \leq p \leq q_C$ (where each vertex is a descendant of itself).
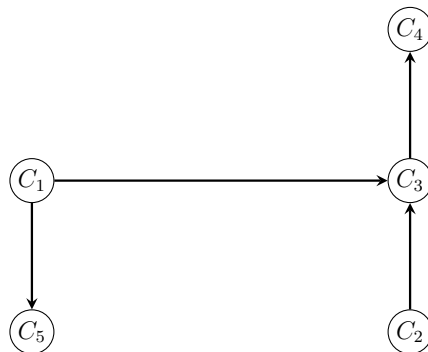

Figure 6: The digraph $\widehat{B}$ associated to the stub of Figure 5.

19

For all $p \in [r]$, let $E^-(C_p)$ be the set of the downward arcs of $\widehat{B}$ with head C_p that link C_p to one of its children, and let $E^+(C_p)$ be the set of the upward arcs of $\widehat{B}$ with tail C_p that leads C_p to one of its children. For an arc $e \in E^+(C_p)$, we let $Q(e)$ be the path obtained from $P(e)$ by removing its last two vertices (taking the path $P(e)$ from parent to child in $\widehat{B}$). Similarly, for an arc $e \in E^-(C_p)$, we let $Q(e)$ be the path obtained from $\overline{P}(e)$ by removing its last two vertices (taking the path $P(e)$ from parent to child in $\widehat{B}$). For all $e \in E(\widehat{B})$, we let $^*Q(e)$ be the path $Q(e)$ with the first vertex removed.

For all $p \in [r]$, the **space** of C_p is

$$\text{spc}(C_p) = 12|C_p| + 36k - 124 + \sum_{e \in E^-(C_p) \cup E^+(C_p)} (|Q(e)| + 1).$$

By definition we have

$$\sum_{p \in [r]} \left(|C_p| + \sum_{e \in E^-(C_p) \cup E^+(C_p)} |Q(e)| \right) = n.$$

Now $|E^-(C_p)| + |E^+(C_p)|$ is the number of arcs between C_p and its children in $\widehat{B}$, so

$$\sum_{p \in [r]} (|E^-(C_p)| + |E^+(C_p)|) \leq r - 1 \leq k - 3.$$

Since A is a stub, all its outer paths have length 1 and so remain in B. Moreover, an inner path of A is either directed, or has two blocks with one of length 1, or has three blocks with the first and the last of length 1. Therefore, at most three of its internal vertices remain in B.

Thus $\sum_{p \in [r]} |C_p| = |B| \leq k + (k - 2) + 3(k - 3) = 5k - 11$. Consequently,

$$\begin{aligned} \sum_{p \in [r]} \text{spc}(C_p) &\leq& n + 11(5k - 11) + (k - 3) + (k - 2)(36k - 124) \\ &\leq& n + 36k^2 - 140k + 124. \end{aligned}$$

Let T be a tournament of order $m = n + 36k^2 - 140k + 124$, and let $(v_1, \ldots, v_m)$ be a local median order of T. Now for all $p = 1$ to r, reserve the first $\text{spc}(C_p)$ unreserved vertices of T for C_p. Therefore the set of vertices reserved for C_p is

$$R_p = \left\{ v_i \mid \sum_{q<p} \text{spc}(C_q) + 1 \leq i \leq \sum_{q \leq p} \text{spc}(C_q) \right\}.$$

Set $\alpha_p = \displaystyle\sum_{e \in E^-(C_p)} (|Q(e)| + 1) + 6|C_p| + 10k - 29$. We partition R_p into three sets. The **middle** of C_p is the set

$$M_p = \{ v_i \mid \sum_{q<p} \text{spc}(C_q) + \alpha_p + 1 \leq i \leq \sum_{q<p} \text{spc}(C_q) + \alpha_p + 16k - 58 \},$$

the **left margin** of C_p is the set

$$M_p^- = \{ v_i \mid \sum_{q<p} \text{spc}(C_q) + 1 \leq i \leq \sum_{q<p} \text{spc}(C_q) + \alpha_p \},$$

and the **right margin** of C_p is the set

$$M_p^+ = \{\ v_i\ \mid\ \sum_{q<p} \operatorname{spc}(C_q) + \alpha_p + 16k - 57 \leq i \leq \sum_{q\leq p} \operatorname{spc}(C_q).$$

Informally, the middle is the range in which we are going to choose where to embed the root of C_p, and the left and right margin give us enough space to embed the island as well as the $Q(e)$ for $e \in E^-(C_p) \cup E^+(C_p)$.

We are going to build an embedding ϕ of A into T. Run a Breadth-First Search algorithm on $\widehat{B}$, and let Π be the resulting ordering. The ordering Π corresponds to a permutation π of $[r]$: $\Pi = (C_{\pi(1)}, C_{\pi(2)}, \ldots, C_{\pi(r)})$. The idea is to embed the islands in increasing order according of π so that each island is treated before its children. When an island C_p is considered, we embed all the vertices of $A_p = C_p \cup \bigcup_{e \in E^-(C_p) \cup E^+(C_p)} Q(e)$ in R_p. For each $e = C_pC_q$ in $E^+(C_p)$ and $E^-(C_p)$, we also embed the path of length 2 between the terminus of $Q(e)$, which is the penultimate node of $P(e)$ (in the parent→child order), and the terminus of $P(e)$ in M_q (this vertex is the root of C_q) using Lemma 22. When using this lemma, the internal vertex of this path is embedded in some vertex that must be forbidden for the others. Therefore, we need to keep track of these forbidden vertices in a set F.

Let us define formally the root a_p of C_p. Pick any node a_1 of C_1 as its root. For all $p \in \{2, \ldots, r\}$, let C_q be the parent of C_p in $\widehat{B}$. There is an arc e between C_p and C_q. The root a_p is the end of $P(e)$ which is in C_p.

A vertex is **free** if it is not yet the image of a node.

Let us now describe the algorithm in detail. It keeps track of a set F of at most $k-3$ vertices (at most one for each arc between two islands in $\widehat{B}$). To start, we set $F = \emptyset$, and we embed a_1 at v_{α_1+1}.

Then for $t = 1$ to r do the following:

0. Set $p = \pi^{-1}(t)$. The root a_p of C_p is already embedded at some vertex v_i.

1. Set $I_p = \{v_j \mid i - 2|C_p| - 2k + 5 \leq j \leq i + 2|C_p| + 2k - 5\}$. Embed C_p in $T\langle I_p \rangle$ thanks to Lemma 11, avoiding the vertices that are in $F \cap I_p$.

2. For each $e \in E^+(C_p)$, consider $P(e) = (x_{e,1}, \ldots, x_{e,\ell_e})$ from C_p to one of its children C_q. Note that $x_{e,1} = a_p$ is already embedded, and that $x_{e,\ell_e} = a_q$. Consider the lowest integer $j \geq i + 2|C_p| + 2k - 4$ such that $\phi(x_{e,1}) \to v_j$ and v_j is free. Embed $x_{e,2}$ at v_j.

   Proceed symmetrically, for the arcs in $E^-(C_p)$.

3. Apply Lemma 23 on the paths $^*Q(e)$ for $e \in E^+(C_p)$. As noted previously, in Lemma 23 we only need to know the vertices of F when we reach them in the construction (since we can just reapply the algorithm on a slightly different arborescence when we meet a vertex in F). We consider the construction of these paths in order, and when the last node of $Q(e)$ is reached, do the following: apply Lemma 22 to get $4k - 15$ internally disjoint directed 2-out-paths from $\phi(x_{e,\ell_e-2})$ to disjoint vertices in M_q (with C_q the head of e in $\widehat{B}$); pick one such path that does not use any vertex of F (here $|F| \leq k - 4$), nor any of the images of the roots of the $C_{q'}$ for $q' \in [r] \setminus \{p, q\}$ which are already embedded (there are at most $k - 4$ of these); put its second vertex in F, embed the penultimate node of $P(e)$ at its second vertex, and embed the root of C_q (which is also the terminus of $P(e)$) at its terminus. The

vertex that was added to F has a larger index than the vertices we are at in the application of Lemma 23, so we can make sure to not embed another node at it.

Do the symmetric on the paths $^*Q(e)$ for $e \in E^-(C_p)$.

Let us now prove that this algorithm results in an embedding of A into T.

Let us first prove that every vertex is mapped to a vertex and that every vertex of A_p is mapped into R_p.

At Step 1, we only embed the nodes of C_p in I_p, which is in an interval of $4|C_p|+4k-9$ vertices centered at some index i in the middle M_p.

At Step 2, we hit at most $|E^+(C_p)|$ out-neighbours of vertices that belong to I_p. Let $w_h = v_{i+2|C_p|+2k-5-h}$ be a vertex of I_p (hence $0 \le h \le 4|C_p|+4k-10$). It has at most h out-neigbours in $I(w_h) = \{v_j \mid i+2|C_p|+2k-4-h \le j \le i+2|C_p|+2k-5\}$. Set

$$J_p^+ = \{v_j \mid i+2|C_p|+2k-4 \le j \le i+6|C_p|+2|E^+(C_p)|+8k-21\}.$$

Note that $|J_p^+| = 2|E^+(C_p)|+4|C_p|+6k-16$. By (M2), w_h has at least $\frac{1}{2}(h+|J_p^+|) = \frac{1}{2}(h+2|E^+(C_p)|+4|C_p|+6k-16) \ge h+|E^+(C_p)|+k-3$ out-neighbours in $I(w_h) \cup J_p^+$ and so at least $|E^+(C_p)|+k-3$ out-neighbours in J_p^+. Hence the vertices of I_p have enough out-neighbours in J_p^+ to choose the $|E^+(C_p)|$ out-neighbours among vertices that are not in F.

Similarly, we hit at most $|E^-(C_p)|$ vertices that belong to the set

$$J_p^- = \{v_j \mid i-6|C_p|-2|E^-(C_p)|-8k+21 \le j \le i-2|C_p|-2k+4\}.$$

At Step 3, we only need to ensure that the conditions of Lemma 23 can be verified whithin R_p. Hence we only need to check that in R_p there are $\sum_{e \in E^+(C_p)}(|Q(e)|-1)+2|F|$ vertices after the last vertex of J_p^+ and $\sum_{e \in E^-(C_p)}(|Q(e)|-1)+2|F|$ vertices before the first vertex of J_p^-. The vertex v_i is in M_p, so $i \ge \sum_{q<p} \mathrm{spc}(C_q) + \alpha_p + 1$. Hence, there are at least $\alpha_p - (6|C_p|+2|E^-(C_p)|+8k-21) = \sum_{e \in E^-(C_p)}(|Q(e)|-1)+2k-8 \ge \sum_{e \in E^-(C_p)}(|Q(e)|-1)+2|F|$ vertices before v_i. (Recall that $|F| \le k-4$ when doing this). Furthermore, $i \le \sum_{q<p} \mathrm{spc}(C_q) + \alpha_p + 16k-58$. So, in R_p, there are at least $\mathrm{spc}(C_p) - (\alpha_p + 16k-58+6|C_p|+2|E^+(C_p)|+8k-29) = \sum_{e \in E^+(C_p)}(|Q(e)|-1)+2k-8$ vertices after v_i. This proves that every vertex is mapped to a vertex and that every vertex of A_p is mapped into R_p.

Finally, let us now show that two nodes are never mapped to a same vertex.

Observe first that at each iteration, we map nodes on distinct vertices. At Step 1, we map nodes into different vertices of I_p. Then at Step 2, we embed nodes into distinct vertices of J_p^+ and J_p^-, and the three sets I_p, J_p^+ and J_p^-, are pairwise disjoint. Finally, at Step 3, using Lemmas 22 and 23, we finish embedding the P_e for $e \in E^+(C_p)$ into M_p^+ and embedding the P_e for $e \in E^+(C_p)$ into M_p^-. We take care of adding the second vertex of the 2-paths in F each time we apply Lemma 22, and that we always avoid embedding vertices in F. Hence two nodes embedded during the same iteration are mapped to different vertices.

In addition, all vertices hit at the iteration t are in R_p (with $p = \pi^{-1}(t)$) except the ones hit when applying Lemma 22. Since the R_p's are pairwise disjoint, we only need to check that when applying Lemma 22, we do not map a node to a vertex onto which another vertex was or will be mapped. Property (ii) of the ordering $(C_1, \ldots, C_r)$ implies that, when applying Lemma 22, nodes are all mapped onto vertices in some R_q such that

C_q is a descendant of C_p. So the only possible conflicts are with other vertices hit when applying this lemma. But we take care of avoiding those vertices (the second vertices of the paths of length 2 are added to F and we specifically avoid the root of R_q for all $q \neq p$).

This completes the proof of Lemma 18.

7 Conclusion and further research

7.1 Towards Conjecture 2 and beyond

The bound $\frac{3}{2}n + \frac{3}{2}k - 2$ of Theorem 4 can be replaced by $n + k - 1 + \min_{r \in V(A)} \min(\gamma_r^\uparrow, \gamma_r^\downarrow)$. However, for any **antidirected tree** A, that is an oriented tree in which every node has either in-degree 0 or out-degree 0, we have $\min_{r \in V(A)} \min(\gamma_r^\uparrow, \gamma_r^\downarrow) = \frac{1}{2}n + \frac{1}{2}k - 1$.

Another step towards Conjecture 2 would be to prove it for antidirected trees.

When proving Theorem 8, we try to keep the proof as simple as possible and made no attempt to get the smallest upper bound on $g(k)$. For example, we can improve on the bound $n + 36k^2 - 140k + 124$ of Lemma 18 by studying more carefully the size of $|F|$ at each iteration. Likewise, we can slightly improve Lemma 17. Doing so, we can get a somewhat better upper bound on $g(k)$ than $114k^2 - 280k + 124$. However, all such bounds are quadratic in k, i.e. $\Omega(k^2)$. A next step towards Conjecture 2 would then be to prove that $g(k) \leq o(k^2)$ (that is every oriented tree of order n with k leaves is $(n + o(k^2))$-unavoidable), and ideally that $g(k) \leq \alpha \cdot k$ for some absolute constant α.

Conjecture 2 is tight because of out-stars and in-stars. But those trees have few nodes: just one more than leaves. In the same way, we believe that all the trees with n nodes and k leaves that are not $(n + k - 2)$-unavoidable have n small compared to k.

Conjecture 25. *For every fixed integer k, there is an integer n_k such that every oriented tree of order $n \geq n_k$ with k leaves is $(n + k - 2)$-unavoidable.*

This conjecture holds for $k = 2$ by a result of Havet and Thomassé [12], and for $k = 3$ as shown by Ceroi and Havet [4]. It is also supported by the result of Mycroft and Naia [15] stating that almost every tree of order n is n-unavoidable.

7.2 Generalisation to k-chromatic digraphs

A **proper k-colouring** of a digraph is a mapping c from its vertex into $\{1, \ldots, k\}$ such that $c(u) \neq c(v)$ for every arc (u, v). A digraph is **k-colourable** if it admits a proper k-colouring. The **chromatic number** of a digraph D, denoted $\chi(D)$, is the least integer k such that D is k-colourable. A digraph is *k-chromatic* if its chromatic number equals k.

The complete graph on n-vertices is the simplest n-chromatic graph, and so tournaments on n vertices are the simplest k-chromatic digraphs. The notion of unavoidability generalizes to the one of universality. A digraph F is **k-universal** if it is contained in every digraph with chromatic number k.

Burr [3] generalizes Sumner's conjecture to universality.

Conjecture 26 (Burr [3])**.** *Every every oriented tree of order n is $(2n - 2)$-universal.*

We also conjecture that Conjecture 2 extends to universality.

Conjecture 27. *Every oriented tree of order n with k leaves is $(n + k - 1)$-universal.*

Acknowledgements

References

[1] B. Alspach and M. Rosenfeld. Realization of certain generalized paths in tournaments. *Discrete Math.*, 34:199–202, 1981.

[2] J. Bondy and U. Murty. *Graph Theory*. Springer Publishing Company, Incorporated, 1st edition, 2008.

[3] S. A. Burr. Subtrees of directed graphs and hypergraphs. In *Proceedings of the Eleventh Southeastern Conference on Combinatorics, Graph Theory and Computing (Florida Atlantic Univ., Boca Raton, Fla., 1980), Vol. I.*

[4] S. Ceroi and F. Havet. Trees with three leaves are (n+1)-unavoidable. *Discrete Applied Mathematics*, 141(1-3):19–39, 2004.

[5] A. El Sahili. Trees in tournaments. *J. Combin. Theory Ser. B*, 92(1):183–187, 2004.

[6] R. Forcade. Parity of paths and circuits in tournaments. *Discrete Math.*, 6:115–118, 1973.

[7] B. Grünbaum. Antidirected Hamiltonian paths in tournaments. *J. Combin. Theory Ser. B*, 11:249–257, 1971.

[8] R. Häggkvist and A. Thomason. Trees in tournaments. *Combinatorica*, 11(2):123–130, 1991.

[9] F. Havet. Trees in tournaments. *Discrete Mathematics*, 243(1):121 – 134, 2002.

[10] F. Havet. On unavoidability of trees with k leaves. *Graphs and Combinatorics*, 19(1):101–110, 2003.

[11] F. Havet and S. Thomassé. Median orders of tournaments: a tool for the second neighborhood problem and Sumner's conjecture. *J. Graph Theory*, 35(4):244–256, 2000.

[12] F. Havet and S. Thomassé. Oriented hamiltonian paths in tournaments: a proof of Rosenfeld's conjecture. *J. Combin. Theory Ser. B*, 78:243–273, 2000.

[13] D. Kühn, R. Mycroft, and D. Osthus. A proof of Sumner's universal tournament conjecture for large tournaments. *Proceedings London Math. Soc.*, 102:731–766, 2011.

[14] D. Kühn, R. Mycroft, and D. Osthus. An approximate version of Sumner's universal tournament conjecture. *J. Combin. Theory, Ser. B*, 101:415–447, 2011.

[15] R. Mycroft and T. Naia. Unavoidable trees in tournaments. *Random Structures & Algorithms*, 53(2):352–385, 2018.

[16] L. Rédei. Ein kombinatorischer Satz. *Acta. Litt. Sci. Szeged*, 7:39–43, 1934.

[17] K. Reid and N. C. Wormald. Embedding oriented n-trees in tournaments. *Studia Scientiarum Mathematicarum Hungarica*, 18(2-4):377–387, 1983.

[18] M. Rosenfeld. Antidirected Hamiltonian paths in tournaments. *J. Combin. Theory Ser. B*, 12:93–99, 1971.

[19] H. Straight. The existence of certain type of semi-walks in tournaments. *Congr. Numer.*, 29:901–908, 1980.

[20] A. Thomason. Paths and cycles in tournaments. *Trans. Amer. Math. Soc.*, 296(1):167–180, 1986.