Available online at www.sciencedirect.com

ScienceDirect

Electronic Notes in Theoretical Computer Science 346 (2019) 425–436

Electronic Notes in Theoretical Computer Science

www.elsevier.com/locate/entcs

On the Unavoidability of Oriented Trees

François Dross[1] and Frédéric Havet[2]

Projet Coati, I3S,
CNRS, INRIA et Université Côte d'Azur
Sophia-Antipolis, France

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 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 $(\frac{21}{8}(n-1))$-unavoidable. Finally, we prove that every oriented tree of order n with k leaves is $(n + 144k^2 - 280k + 124)$-unavoidable.

Keywords: Tournament, oriented trees, unavoidability, Ramsey theory.

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. The **unavoidability** of a 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 [14] which states that the unavoidabilty of the directed path on n vertices, is n. In 1971, Grünbaum studied the **antidirected paths** that are 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 [6] 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) or $n = 5$ (in

1 Email: francois.dross@inria.fr

2 Email: frederic.havet@inria.fr

https://doi.org/10.1016/j.entcs.2019.08.038
1571-0661/© 2019 Published by Elsevier B.V.

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). The same year, Rosenfeld [16] gave an easier proof and conjectured that there is a smallest integer $N_P > 7$ such that $\text{unvd}(P) = |P|$ for every oriented path of order at least N_P. The condition $N_P > 7$ results from Grünbaum's counterexamples. Several papers gave partial answers to this conjecture [1,5,17] until Rosenfeld's conjecture was verified by Thomason, who proved in [18] that N_P exists and is less than 2^{128}. Finally, Havet and Thomassé [11], showed that $\text{unvd}(P) = |P|$ for every oriented path P except the antidirected paths of order 3, 5, and 7.

Regarding oriented trees, Sumner (see [15]) made the following celebrated conjecture.

Conjecture 1.1 *Every oriented tree of order $n > 1$ is $(2n-2)$-unavoidable.*

The first linear bound was given by Häggkvist and Thomason [7]. Following improvements of Havet [8] and Havet and Thomassé [10], El Sahili [4] used the notion of median order, first used as a tool for Sumner's conjecture in [10], and proved that every oriented tree of order n ($n \geq 2$) is $(3n-3)$-unavoidable. Recently, Kühn, Mycroft and Osthus [13] 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 [12], in which an approximate version of the conjecture was proved. In [10], 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 [9]) made the following stronger conjecture than Sumner's one.

Conjecture 1.2 *Every oriented tree of order n with k leaves is $(n+k-1)$-unavoidable.*

As an evidence to this conjecture, Häggkvist and Thomason [7] proved 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 1.2 is true when $k = 2$ and Ceroi and Havet [3] showed that it holds for $k = 3$. Havet [9] also showed Conjecture 1.2 for a large class of trees.

1.1 Our results

In Section 3, we prove Conjecture 1.2 for arborescences.

Theorem 1.3 *Every arborescence of order n with k leaves is $(n+k-1)$-unavoidable.*

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

Theorem 1.4 *Every oriented tree of order n with k leaves is $(\frac{3}{2}n + \frac{3}{2}k - 2)$-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 1.5 *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 1.6 *Every oriented tree with $n \geq 3$ vertices and k leaves is $(\frac{9}{2}n - \frac{5}{2}k - \frac{9}{2})$-unavoidable.*

Theorems 1.4 and 1.6 yield the best bound towards Sumner's conjecture:

Corollary 1.7 *Every oriented tree of order $n \geq 2$ is $\left(\frac{21}{8}(n-1)\right)$-unavoidable.*

Proof. The value of $\min\left(\frac{3}{2}(n+k-1), \frac{9}{2}n - \frac{5}{2}k - \frac{9}{2}\right)$ is maximal when $\frac{3}{2}(n+k-1) = \frac{9}{2}n - \frac{5}{2}k - \frac{9}{2}$, that is when $k = \frac{3n-3}{4}$. In this case $\frac{3}{2}(n+k-1) = \frac{21}{8}(n-1)$. □

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

2 Definitions and preliminaries

Notation generally follows [2]. The digraphs have no parallel arcs and no loops.

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

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 **father** 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 father of v, then v is a **son** 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 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_iv_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. In other words, a median order is an ordering of the vertices such that the set of backward arcs is a minimum feedback arc set. Let us note basic properties of median orders of tournaments whose proofs are left to the reader.

Lemma 2.1 *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 $1 \le i < j \le n$:*

(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}.*

3 Unavoidability of arborescences

The aim of this section is to prove Theorem 1.3. We prove the following theorem which implies it directly by directional duality.

Theorem 3.1 *Every out-arborescence of order n with k out-leaves is $(n + k - 1)$-unavoidable.*

Proof. Let A be an out-arborescence with n nodes, k out-leaves and root r, and let T be a tournament on $m = n + k - 1$ vertices.

Let $\sigma = (v_1, v_2, \ldots, v_m)$ be a median order of T. For each node a of A, we fix an ordering $\mathcal{O}_a$ of the sons of A.

We now describe a procedure giving an embedding ϕ of A into T. We first set $\phi(r) = v_1$. We then consider the vertices of T one after another according to σ. 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 $a_j = \phi^{-1}(v_j)$. An **active node** is a node of A which is embedded and which has an unembedded son. The **active son** of an active node a is its minimum unembedded son according to $\mathcal{O}_a$. An **active vertex** is a vertex of T which is the image of an active node. For $2 \le i \le m$, we do the following. If v_i is dominated by an active vertex, then let j be the smallest index of an active vertex dominating v_i, and b be the active son of a_j. We set $\phi(b) = v_i$ and say that

v_i is **hit**. Otherwise v_i is dominated by no active vertex, and we assign no node to v_i, and say that v_i is **failed**.

To prove that this procedure yields an embedding of A into T, we shall prove that the set B of embedded nodes at the end of the procedure is $V(A)$ or equivalently that the set F of failed vertices has cardinality $k-1$.

Consider a vertex v_i in F. If at Step i, there was no active vertex, then all nodes of A are embedded and we have the result. So we may assume that there was an active vertex. Let ℓ_i be the largest index such that v_{ℓ_i} is active at Step i, and let $I_i = \{v_j \mid \ell_i < j \leq i\}$.

Since we only embed a node if we have embedded its father, $A\langle B\rangle$ 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 $|L| < k-1$.

Claim 3.2 *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 son of a_{ℓ_i} to it. Thus $I_i \cap F \subseteq I_i \cap N^-(v_{\ell_i})$ and so

$$|I_i \cap F| \leq |I_i \cap N^-(v_{\ell_i})|. \tag{1}$$

Let v_j be a hit vertex in I_i. By definition of ℓ_i, a_j is not active, so its sons (if any) are embedded, necessarily in $\{v_{j+1}, \dots, v_{i-1}\} \subseteq I_i$. Again, by definition of ℓ_i, all the sons of a_j are not active, and so their sons (if any) are embedded in I_i. And so on, all descendants of a_j are embedded in I_i and not active. We associate 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 above, they are hit, and the descendants of their pre-images are also embedded in I_i. Moreover, for each $v_j \in J$, the father of a_j is embedded in $\{v_1, \dots, v_{\ell_i}\}$ for otherwise, at Step j, the procedure would have assigned v_j to the active son of a_{ℓ_i} or another active node embedded in $\{v_1, \dots, v_{\ell_i}\}$. Hence no vertex of J is the image of an ancestor of another node embedded in J. Consequently, the out-leaves embedded inf J are all distinct. Thus

$$|I_i \cap N^+(v_{\ell_i})| \leq |I_i \cap \phi(L)|. \tag{2}$$

Now, by Lemma 2.1, $|I_i \cap N^-(v_{\ell_i})| \leq |I_i \cap N^+(v_{\ell_i})|$. Together with Equations (1) and (2), this proves the claim. $\diamondsuit$

Claim 3.3 *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 $\ell_i < \ell_j < i$. At Step i of the procedure, v_{ℓ_j} was not active, by definition of ℓ_i. But then it was still not active at Step j, a contradiction to the definition of ℓ_j. $\diamondsuit$

Now let M be the set of indices i such that $v_i \in F$ and I_i is maximal for inclusion. 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 3.3, the I_i,

$i \in M$, are pairwise disjoint. So $|F| = \sum_{i \in M} |I_i \cap F|$. By Claim 3.2, we obtain

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

This completes the proof. $\square$

Remark 3.4 With the embedding ϕ constructed in the above proof, there is an injection from the set F of failed vertices into $L^+(A)$ such that every failed 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 at Step i, there was some active vertex, then by Claim 3.2, $|I_i \cap F| \leq |I_i \cap \phi(L)|$. 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 failed vertex.

If at Step i, there was no active vertex, 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 failed vertex. Necessarily, $f(v_i)$ is embedded in a vertex preceding v_i in σ. $\square$

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. Observe that in the proof of Theorem 3.1, the root of the out-arborescence is embedded in the first vertex of the median order. By directional duality, the root of an in-arborescence is embedded in the last vertex of the median order. Hence one can easily derive the following corollary.

Corollary 3.5 *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 with root 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**. The set of components of the upward (resp. downward) forest is denoted by $\mathcal{C}_r^{\uparrow}(A)$ (resp. $\mathcal{C}_r^{\downarrow}(A)$), or simply $\mathcal{C}_r^{\uparrow}$ (resp. $\mathcal{C}_r^{\downarrow}$) when A is clear from the context. Set $\gamma_r^{\uparrow} = \sum_{C \in \mathcal{C}_r^{\uparrow}} (|V(C)| + |L^+(C)| - 2)$ and $\gamma_r^{\downarrow} = \sum_{C \in \mathcal{C}_r^{\downarrow}} (|V(C)| + |L^-(C)| - 2)$. Observe that each component of the upward (resp. downward) forest contains an arc and thus at least two vertices and one out-leaf (resp. in-leaf). Hence $|V(C)| + |L^+(C)| - 2 > 0$ for all $C \in \mathcal{C}_r^{\uparrow}$ and $|V(C)| + |L^-(C)| - 2 > 0$ for all $C \in \mathcal{C}_r^{\downarrow}$.

Lemma 4.1 *Let A be a rooted tree with n nodes and k leaves such that the root r of A 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, 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. 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.

Note that the embedding procedure presented in the proof of Theorem 3.1 could be replaced by the following one, which yields the same bijection.

Set $\phi(r) = v_1$. For $i = 1$ to m, if v_i is hit, take the $|N^+(a_i)|$ first out-neighbours of v_i in $\{v_{i+1}, \ldots, v_m\}$ that are not yet hit and assign them to the sons of a_i increasingly according to $\mathcal{O}_{a_i}$.

The interest of this procedure is that it considers the nodes of the tree in a different order. The main property we are going to use is that the images of all of the sons of a given node a are fixed at the same time, and we do not need to fix the order $\mathcal{O}_a$ before the set of those images is known. It means that we can choose the order $\mathcal{O}_a$ after we know which vertices are the images of the sons of a. Thus we can effectively choose which son of a is embedded to which vertex with the knowledge of the set of the images of the sons of a.

Now we build an arborescence A' from the rooted tree A, which we call the **equivalent arborescence** of A. For all $i \in \{1, ..., j\}$, do the following. Let f_i be the father 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 a rooted tree with the same root as A and since we removed the downward arcs and added only upward arcs, A' is even an out-arborescence. 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 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.

Therefore by Theorem 3.1, there is an embedding ϕ of A' into T. We build it according to the procedure presented in the beginning of this proof. Let $i \in \{1, ..., j\}$, and consider $S_i = V(C_i) \cup N_i$. This is a set of $n_i + k_i - 1$ sons 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 3.1, 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 are in A', and preserves the downward arcs since each downward arc of A is in some C_i. Therefore ψ is an embedding of A into T. $\square$

We are now able to prove Theorem 1.4 which states that every oriented tree of order n with k leaves is $(\frac{3}{2}(n+k)-2)$-unavoidable.

Proof of Theorem 1.4. Let T be a tournament on $\frac{3}{2}(n+k)-2$ vertices. Let A be an oriented tree with n nodes an 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 $\gamma_r^{\downarrow}$.

Since $\gamma_r^{\downarrow} \leq \gamma_r^{\uparrow}$, we have $\gamma_r^{\downarrow} \leq \frac{1}{2}(\gamma_r^{\uparrow} + \gamma_r^{\downarrow})$. For a rooted tree A, let $L'(A)$ be the set of leaves of A distinct from the root. Note that if A_1 and A_2 are two rooted trees that are disjoint except in one vertex which is the root of A_2, then $A_1 \cup A_2$ is a tree, and if we root it at the root of A_1, then $|L'(A_1 \cup A_2)| \geq |L'(A_1)| + |L'(A_2)| - 1$. By applying that successively for all the components of the upward and downward forests of (A, r), we get that $\gamma_r^{\uparrow} + \gamma_r^{\downarrow} \leq n+k-2$, and thus $\gamma_r^{\downarrow} \leq \frac{1}{2}(n+k)-1$. Hence, T has at least $n+k-1+\gamma_r^{\downarrow}$ vertices.

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 sr and possibly s or r if they become isolated. All components of F_r not containing sr are also components of F_s and the component C_0 of F_r containing sr either disappears (when sr 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. Lemma 4.1 finishes the proof. □

5 Unavoidability of trees with many leaves

The aim of this section is to establish Theorem 1.6, which states that every oriented tree with $n \geq 3$ vertices and k leaves is $(\frac{9}{2}n - \frac{5}{2}k - \frac{9}{2})$-unavoidable.

Proof. Let $m = \lceil \frac{9}{2}n - \frac{5}{2}k - \frac{9}{2} \rceil$. 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 3.5. Henceforth, we assume that A is not a bi-arborescence. In particular, $k < n-1$.

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 a neighbour in S^-, since otherwise it would be in S^+. Similarly, each in-leaf of H has a neighbour in S^+. In particular, $|S^-| \geq |L^+(H)|$ and $|S^+| \geq |L^-(H)|$.

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 vertex.

Let $\sigma = (v_1, v_2, \dots, v_m)$ be a median order of T. We distinguish between two cases: (1) one of S^- or S^+ is empty, and (2) both S^- and S^+ are non-empty. We only present here case (2) which is the most complicated one. A similar (and easier) argument works for case (1).

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 $\beta_r^{\downarrow} = \sum_{C \in \mathcal{C}_r^{\downarrow}} (3|V(C)| - 3) + 2|L^-(H)|$ and $\beta_r^{\uparrow} = \sum_{C \in \mathcal{C}_r^{\uparrow}} (3|V(C)| - 3) + 2|L^+(H)|$. 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, $\beta_r^{\downarrow} \leq \frac{1}{2}\beta_r^{\downarrow} + \frac{1}{2}\beta_r^{\uparrow} = \frac{3}{2}n_H + k_H - \frac{3}{2}$. 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 $p = \ell + n_H + k_H - 1 + \gamma_r^{\downarrow}$.

Phase 1: We embed H in $T\langle\{v_{\ell+1}, \dots, v_p\}\rangle$ using the procedure of Lemma 4.1 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.

Phase 2: While there is an unembedded vertex 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 $\{v_{i+1}, \dots, v_m\}$ that is not yet hit and assign it to an unembedded out-neighbour in S^+.

Unembed all vertices of $H' - H$.

Phase 3: While there is an unembedded vertex 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 $\{v_1, \dots, v_{i-1}\}$ 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 4.1, all vertices of H are embedded in Phase 1.

Let us now prove that all vertices 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

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

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

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

The two previous equations 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 $m - \ell \ge 2n_H + 2 \sum_{C \in \mathcal{C}_r^{\downarrow}} (|L^-(C)| - 1) + 2|S^+| - 2 \ge 2|B'| - 2$.

At each addition of a vertex 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 $\{v_{i+1}, \dots, v_m\}$ that is not yet hit and assigns it to an unembedded out-neighbour of $\phi^{-1}(v_i)$. Therefore, at each step, ϕ is an embedding of B'', the so far constructed sub-out-arborescence of B' into $T\langle\{v_{\ell+1}, \dots, v_{\ell+2|B''|-2}\}\rangle$, such that in each end-interval $\{v_j, \dots, v_{\ell+2|B''|-2}\}$ at most $\frac{\ell+2|B''|-2-j+1}{2}$ vertices are in $\phi(B'')$. Thus by Lemma 2.1 (M2), every vertex of v_i as an out-neighbour in $\{v_{i+1}, \dots, v_{\ell+2|B''|}\} \setminus \phi(B'')$ and the procedure can continue. Hence, B' can be embedded into $T\langle\{v_{\ell+1}, \dots, v_m\}\rangle$.

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 $\{v_1, \dots, v_p\}$.

Consider the moment when we try to embed the in-neighbours of b during Phase 3. Let hit be the number of vertices of $\{v_1, \dots, v_{i-1}\}$ that are hit at this moment. Since a is not embedded, we have $\text{hit} \ge |N^-(v_i) \cap \{v_1, \dots, v_{i-1}\}| - |N_A^-(b)| + 1$. By Lemma 2.1, $|N^-(v_i) \cap \{v_1, \dots, v_{i-1}\}| \ge \frac{i-1}{2}$. So

$$\text{hit} \ge \frac{i-1}{2} - |N_A^-(b)| + 1. \tag{3}$$

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 $\text{hit} = \text{hit}_2 + \text{hit}_3$, where hit_2 (resp. hit_3) is the number of vertices of $\{v_1, \dots, v_{i-1}\}$ 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

$$\text{hit}_3 \le |S^-| - |O_{<i}| - |N_A^-(b)|. \tag{4}$$

Consider now the embedding of H. It is made using the procedure of Lemma 4.1, which applies the procedure of Theorem 3.1 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 Remark 3.4, we can map each failed vertex to an out-leaf of H' whose image precedes the failed vertex in σ. But there are $k_{H'} - 1$ failed vertices, therefore each out-leaf of $O_{\geq i}$ except at most one corresponds to two vertices v_l with $l \geq i$. Thus, there are at least $2|O_{\geq i}| - 1$ vertices in $\{v_i, \dots, v_p\}$. Moreover, $\sum_{C \in \mathcal{C}_r^{\downarrow}} (|L(C)| - 1)$ vertices of $H' - H$ are unembedded at the end of Phase 2 and were neither out-leaves of H nor failed vertices. Hence,

$$\mathrm{hit}_2 \leq p - \ell - \sum_{C \in \mathcal{C}_r^{\downarrow}} (|L(C)| - 1) - 2|O_{\geq i}| + 1 = n_H + k_H + \sum_{C \in \mathcal{C}_r^{\downarrow}} (|V(C)| - 1) - 2|O_{\geq i}|. \quad (5)$$

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

$$\mathrm{hit}_2 \leq i - \ell - 1 \quad (6)$$

Summing 2 Eq. (3) + 2 Eq. (4) + Eq. (5) + Eq. (6) yields the following contradiction.

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

□

6 Unavoidability of trees with very few leaves

Due to lack of space, we just present here a very rough sketch of the proof of Theorem 1.8 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 with degree at least 3 and a **flat node** is a node with degree 2. A **segment** 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 segment is an **inner segment**; otherwise it is an **outer segment**. The **opposite** of an inner segment S, denoted by $\overline{S}$ is the inner segment with origin the terminus of S and terminus the origin of S.

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

(i) every inner segment 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 its blocks has length 1.

(ii) every outer segment has length 1.

Our proof of Theorem 1.8 involves two steps. We first prove the following lemma, which shows that it is sufficient to concentrate on stubs.

Lemma 6.1 *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 6.2 *Every stub with n nodes and $k \geq 6$ leaves is $(n + 36k^2 - 140k + 124)$-unavoidable.*

Theorem 1.8 follows directly from Lemmas 6.1 and 6.2 whose proofs are omitted due to lack of space.

References

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

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

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

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

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

[6] Grünbaum, B., *Antidirected Hamiltonian paths in tournaments*, J. Combin. Theory Ser. B **11** (1971), pp. 249–257.

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

[8] Havet, F., *Trees in tournaments*, Discrete Mathematics **243** (2002), pp. 121 – 134.

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

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

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

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

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

[14] Rédei, L., *Ein kombinatorischer Satz*, Acta. Litt. Sci. Szeged **7** (1934), pp. 39–43.

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

[16] Rosenfeld, M., *Antidirected Hamiltonian paths in tournaments.*, J. Combin. Theory Ser. B **12** (1971), pp. 93–99.

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

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