Median orders of tournaments: a tool for the second neighbourhood problem and Sumner's conjecture.

Frédéric Havet
and
Stéphan Thomassé

Laboratoire LaPCS, UFR de Mathématiques,
Université Claude Bernard
43, Boulevard du 11 novembre 1918,
69622 Villeurbanne Cedex, France
email : havet@jonas.univ-lyon1.fr
thomasse@jonas.univ-lyon1.fr

Abstract

We give a short constructive proof of a theorem of Fisher: every tournament contains a vertex whose second outneighbourhood is as large as its first outneighbourhood. Moreover, we exhibit two such vertices provided that the tournament has no dominated vertex. The proof makes use of median orders. A second application of median orders is that every tournament of order $2n-2$ contains every arborescence of order $n>1$. This is a particular case of Sumner's conjecture: every tournament of order $2n-2$ contains every oriented tree of order $n>1$. Using our method, we prove that every tournament of order $(7n-5)/2$ contains every oriented tree of order n.

A *median order* of a tournament T is a linear extension of an acyclic subdigraph of T, maximal with respect to its number of arcs. This concept arises naturally in voting theory, and many articles deal with the computation of such orders. Determining a median order of a digraph is NP-hard, and the complexity for tournaments is still unknown (see [1]). Surprisingly, the notion of median order, well-studied for its own sake, has been seldom used as a tool in tournament theory. It appears that median orders provide a very powerful inductive method. In this paper, we apply them to two questions. The first one is finding a vertex whose first neighbourhood is no greater than its second neighbourhood. This was known as Dean's conjecture [2] until Fisher [3] proved it. We give, in Theorem 1, a short constructive proof of this fact. Actually, our method affords a slight extension (Theorem 2): if a tournament has no dominated vertex, there exist two vertices which satisfy Dean's property. The second question is Sumner's conjecture (see [8]), posed around 1972, asserting that for $n > 1$, every tournament of order $2n-2$ contains every oriented tree of order n. In 1982, Wormald [8] proved that, for $n \geq 4$, every tournament of order $n\log_2(2n/e)$ contains every oriented n-tree. A year later, Reid and Wormald [7] showed that every near-regular $(2n-2)$-tournament contains every oriented n-tree. In addition, they proved that every orientation of a caterpillar of order n and diameter at most 4 is contained in every $(2n-2)$-tournament. The first linear bound was given by Häggkvist and Thomason [5] in 1991. They obtained $12n$ in place of $2n-2$, and determined an asymptotic bound of $(4+o(1))n$. Their method, based on the notion of k-heart of a tree, was later used by Havet [6] to reduce the bound to $7.6n$. By means of median orders, we prove in Theorem 3 that Sumner's conjecture holds for arborescences (trees oriented from a root); also, by the same short argument, that the bound of $4n-6$ holds for all trees (Theorem 4). In the last section of this paper, we show in Theorem 5 that this bound can be improved to $(7n-5)/2$. But the calculation is more involved and the argument no longer simple.

1 The feedback property.

In this paper, digraphs are understood to be orientations of finite simple graphs, that is, loopless and without multiple arcs or circuits of length two. Let $D = (V, E)$ be a digraph with vertex set V and arc set E. The induced restriction of D to a subset S of V is denoted by $D_{|S}$. Let v be a vertex of D. The *outneighbourhood* of v in D is the set $N_D^+(v) = \{x \in V(D) \ : \ v \to x\}$ and the *second outneighbourhood* of v in D is the set $N_D^{++}(v) = (\cup_{x\in N_D^+(v)} N_D^+(x)) \setminus N_D^+(v)$. The *outdegree* of v is the number of elements of $N_D^+(v)$; we denote it by $d_D^+(v)$. The dual notions of *indegree*, *inneighbourhood* and *second inneighbourhood* are denoted by $d_D^-(v)$, $N_D^-(v)$ and $N_D^{--}(v)$, respectively. Since we always deal with a tournament T, the notations $N^+(v), d^+(v), \ldots$ refer to $N_T^+(v), d_T^+(v), \ldots$ A vertex v of T is *dominating* (resp. *dominated*) if $d^-(v) = 0$ (resp. $d^+(v) = 0$). Let $T = (V, E)$ be a finite tournament. An *order of* T is a total order $L = (V, E')$ of the vertices of T. We shall often regard the order L as an enumeration $(x_1, \ldots, x_n)$ of the vertices of T, or as a tournament on V with arc set $E' = \{x_i \to x_j \ : \ i < j\}$. The pair (T, L) always denotes a tournament T together with an order L of T. We denote by $T \cap L$ the acyclic directed graph $(V, E \cap E')$. The *interval* $[x_i, x_j]$, for $i \leq j$, of (T, L) is the subset of vertices $\{x_i, x_{i+1}, \ldots, x_j\}$. An order L of T which maximizes the number of arcs of $T \cap L$ is a *median order* of T. Note that every median order L of T satisfies the *feedback property* : for every i, j such that $1 \leq i \leq j \leq n$, both the outdegree of x_i and the indegree of x_j in $(T \cap L)_{|[x_i,x_j]}$ are at least $(j-i)/2$; that is:

$$d^+_{T_{|[x_i,x_j]}}(x_i) \geq d^-_{T_{|[x_i,x_j]}}(x_i) \text{ and } d^-_{T_{|[x_i,x_j]}}(x_j) \geq d^+_{T_{|[x_i,x_j]}}(x_j).$$

Indeed, assume for instance that this property does not hold for x_i. Then inserting x_i just after x_j would increase the number of arcs of $T \cap L$. A *local median order* of T is an order of T which satisfies the feedback property. Note that, by the feedback property, $x_1, \ldots, x_n$ is a Hamiltonian path whenever $L = (x_1, \ldots, x_n)$ is a local median order of T. Let T be a tournament of order n. A vertex v of T is a *feed vertex* (resp. a *back vertex*) if there exists a local median order L of T such that v is maximal in L (resp. minimal in L). We recall that a vertex x in a tournament T is a *king* if $\{x\} \cup N^+(x) \cup N^{++}(x) = V(T)$.

There is obviously a significant difference between median orders and local median orders since one can construct easily a local median order of a given tournament of order n (in time $\mathrm{O}(n^4)$, for instance, by means of a greedy algorithm), whereas finding a median order is NP-hard. The crucial property of (local) median orders is the following: if I is an interval of a (local) median order L of T, then $L_{|I}$ is a (local) median order of $T_{|I}$. This easy observation provides a very powerful inductive tool, as we shall see in the following sections. In order to introduce the notion of local median orders, we use it to prove the following classical (easy) result:

Proposition 1 *Every tournament has a king. Moreover, a tournament with no dominating vertex has at least three kings.*

Proof. We prove first that every back vertex of a tournament T is a king. Let x_1 be a back vertex of T and $L = (x_1, \ldots, x_n)$ be a local median order of T. Now pick any vertex x_i. By the feedback property, both the outdegree of x_1 and the indegree of x_i in $(T \cap L)_{|[x_1,x_i]}$ are at least $(i-1)/2$. So, either x_1 dominates x_i, or there is $1 < k < i$ for which x_1 dominates x_k, which in turn dominates x_i. Thus, x_1 is a king of T. Now suppose that T has no dominating vertex, let x_i be the inneighbour of x_1 which is minimal with respect to its index in L and let x_j be the inneighbour of x_i which is minimal with respect to its index in L. We claim that both x_i and x_j are kings of T. First, observe that x_j belongs to $[x_1, x_i]$. Now, x_i is a back vertex, hence a king, of $T_{|[x_i,x_n]}$, and, via x_1, is also a king of $T_{|[x_1,x_i]}$. Moreover, x_j is a back vertex, hence a king, of $T_{|[x_j,x_n]}$, and, via x_i, is at distance at most two from the vertices of $[x_1, x_{j-1}]$. $\square$

2 Median orders and second neighbourhoods.

One of the (apparently) simplest open questions concerning digraphs is Seymour's second neighbourhood conjecture, asserting that one can always find, in a finite digraph $D = (V, E)$, a vertex x such that $|N_D^+(x)| \leq |N_D^{++}(x)|$ (to avoid this notation, we will say that x has a *large second neighbourhood in* D). Very surprisingly, this question remained unsolved even for tournaments (it was known as Dean's conjecture [2]) until Fisher [3] proved the existence of such a vertex as follows: Fisher and Ryan [4] exhibited a probability distribution p on V such that $p(N^+(v))$ is greater than or equal to $p(N^-(v))$ for every vertex v. Subsequently, Fisher proved that, for tournaments, this probability also satisfies $p(N^-(v)) \leq p(N^{--}(v))$ for all vertices v. Thus, by an averaging argument and a sum inversion, at least one vertex x has a large second neighbourhood. In Theorem 1, we give an explicit construction of such a vertex x. We first need some definitions. Let $L = (x_1, \ldots, x_n)$ be a local median order of a tournament T. We distinguish two types of vertices of $N^-(x_n)$: a vertex $x_j \in N^-(x_n)$ is *good* if there exists $x_i \in N^+(x_n)$, with $i < j$, such that $x_i \to x_j$; otherwise x_j is *bad*. We denote the set of good vertices of (T, L) by G_L.

Theorem 1 *Every feed vertex of a tournament has a large second neighbourhood.*

Proof. We prove here a stronger result. Let $L = (x_1, \ldots, x_n)$ be a local median order of a tournament T. We prove by induction on n that x_n satisfies $|N^+(x_n)| \leq |G_L|$. The case $n = 1$ holds vacuously. Assume now that n is greater than one. If there is no bad vertex, we have $G_L = N^-(x_n)$. Moreover the feedback property ensures that $|N^+(x_n)| \leq |N^-(x_n)|$, so the conclusion holds. Now we assume that there exists a bad vertex x_i, and we choose it minimal with respect to its index i. Denote by G_L^u the set $G_L \cap [x_{i+1}, x_n]$, by G_L^d the set $G_L \cap [x_1, x_i]$, by $N^+(x_n)^u$ the set $N^+(x_n) \cap [x_{i+1}, x_n]$ and by $N^+(x_n)^d$ the set $N^+(x_n) \cap [x_1, x_i]$. Applying the induction hypothesis to the restriction of (T, L) to $[x_{i+1}, x_n]$ gives directly that $|G_L^u| \geq |N^+(x_n)^u|$, since every good vertex of this restriction is, *a fortiori*, a good vertex of (T, L). By the minimality of the index of x_i, every vertex of $[x_1, x_{i-1}]$ is either in G_L^d or in $N^+(x_n)^d$. Moreover, since x_i is bad, we have $N^+(x_n)^d \subseteq N^+(x_i) \cap [x_1, x_i]$ and (equivalently), $G_L^d \supseteq N^-(x_i) \cap [x_1, x_i]$. The feedback property applied to $[x_1, x_i]$ gives:

$$|G_L^d| \geq |N^-(x_i) \cap [x_1, x_i]| \geq |N^+(x_i) \cap [x_1, x_i]| \geq |N^+(x_n)^d|. \tag{1}$$

Thus $|G_L^d| \geq |N^+(x_n)^d|$ and $|G_L^u| \geq |N^+(x_n)^u|$, so our induction hypothesis holds for every non-negative integer n. $\square$

A natural question is to seek another vertex with large second neighbourhood. Obviously, this is not always possible: consider for instance a regular tournament dominating a single vertex, or simply, a transitive tournament. In both cases, the sole vertex with large second neighbourhood is the dominated vertex. We prove now that a tournament always has two vertices with large second neighbourhood, provided that every vertex has outdegree at least 1. The notion of local median orders turns out to be too weak for that purpose, so we use median orders.

We introduce the notion of *sedimentation* of a median order $L = (x_1, \ldots, x_n)$ of T, denoted by $Sed(L)$. We recall that, by the proof of Theorem 1, $|N^+(x_n)| \leq |G_L|$. If $|N^+(x_n)| < |G_L|$, then $Sed(L) = L$. If $|N^+(x_n)| = |G_L|$, we denote by $b_1, ..., b_k$ the bad vertices of (T, L) and by $v_1, ..., v_{n-1-k}$ the vertices of $N^+(x_n) \cup G_L$, both enumerated in increasing order with respect to their index in L. In this case, $Sed(L)$ is the order $(b_1, \ldots, b_k, x_n, v_1, \ldots, v_{n-1-k})$ of T.

Lemma 1 *The order $Sed(L)$ is a median order of T.*

Proof. If $Sed(L) = L$, there is nothing to prove. Otherwise, we assume that $|N^+(x_n)| = |G_L|$. The proof is by induction on k, the number of bad vertices. If $k = 0$, all the vertices are good or in $N^+(x_n)$, in particular $N^-(x_n) = G_L$. Thus, $|N^+(x_n)| = |N^-(x_n)|$ and the order $Sed(L) = (x_n, x_1, \ldots, x_{n-1})$ is a median order of T. (Note that this is not true for local median orders.) Now, assume that k is a positive integer. Let i be the index of the vertex b_1 in L (that is $b_1 = x_i$). As before, denote by G_L^u the set $G_L \cap [x_{i+1}, x_n]$, by G_L^d the set $G_L \cap [x_1, x_i]$, by $N^+(x_n)^u$ the set $N^+(x_n) \cap [x_{i+1}, x_n]$ and by $N^+(x_n)^d$ the set $N^+(x_n) \cap [x_1, x_i]$. By (1), $|G_L^d| \geq |N^+(x_n)^d|$. Since $|G_L^u| \geq |N^+(x_n)^u|$ and $|G_L^d| + |G_L^u| = |N^+(x_n)^d| + |N^+(x_n)^u|$, we have $|G_L^u| = |N^+(x_n)^u|$, $|G_L^d| = |N^+(x_n)^d|$ and again by (1) $|N^+(x_i) \cap [x_1, x_i]| = |N^-(x_i) \cap [x_1, x_i]|$; in particular $L' = (b_1, x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$ is a median order of T. Observe also that the bad vertices of (T, L') are exactly the bad vertices of (T, L). To conclude, apply the induction hypothesis to the restriction of (T, L) to $[x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n]$. $\square$

Define now inductively $Sed^0(L) = L$ and $Sed^{q+1}(L) = Sed(Sed^q(L))$. If the process reaches a rank q such that $Sed^q(L) = (y_1, \ldots, y_n)$ and $|N^+(y_n)| < |G_{Sed^q(L)}|$, call the order L *stable*. Otherwise, call L *periodic*.

Theorem 2 *A tournament with no dominated vertex has at least two vertices with large second neighbourhood.*

Proof. Let $L = (x_1, \ldots, x_n)$ be a median order of T. By Theorem 1, x_n has a large second neighbourhood, so we need to find another vertex with this property. Consider the restriction of (T, L) to the interval $[x_1, \ldots, x_{n-1}]$, and denote it by (T^d, L^d). Suppose first that L^d is stable, and consider an integer q for which $Sed^q(L^d) = (y_1, \ldots, y_{n-1})$

and $|N^+_{T^d}(y_{n-1})| < |G_{Sed^q(L^d)}|$. Note that $(y_1, \dots, y_{n-1}, x_n)$ is a median order of T, and consequently $y_{n-1} \to x_n$. Thus,

$$|N^+(y_{n-1})| = |N^+_{T^d}(y_{n-1})| + 1 \le |G_{Sed^q(L^d)}| \le |N^{++}(y_{n-1})|.$$

So y_{n-1} has a large second neighbourhood in T. Now assume that L^d is periodic. Since T has no dominated vertex, x_n has an outneighbour x_j. Note that, for every integer q, the feed vertex of $Sed^q(L^d)$ dominates x_n. So x_j is not the feed vertex of any $Sed^q(L^d)$. Observe also that, since L^d is periodic, x_j must be a bad vertex of some $Sed^q(L^d)$, otherwise the index of x_j would always increase during the sedimentation process. Now fix this value of q. Let $Sed^q(L^d) = (y_1, \dots, y_{n-1})$. We claim that y_{n-1} has a large second neighbourhood in T: on the one hand we have

$$|N^+(y_{n-1})| = |N^+_{T^d}(y_{n-1})| + 1 = |G_{Sed^q(L^d)}| + 1$$

and on the other hand we have $y_{n-1} \to x_n \to x_j$, so the second neighbourhood of y_{n-1} has at least $|G_{Sed^q(L^d)}| + 1$ elements. $\square$

It appears that the limitation of the use of median orders for the second neighbourhood conjecture are roughly the same as those of Fisher's proof. For instance, the following statement can easily be proved using both approaches. Here, a *quasi-transitive* digraph satisfies the property ($x \to y$ and $y \to z$)$\Rightarrow$($x \to z$ or $z \to x$).

Lemma 2 *Let $D = (V, E)$ be a quasi-transitive digraph and p be a probability distribution on V. There exists a vertex x of D such that $p(N^+_D(x)) \le p(N^{++}_D(x))$.*

Proof. Consider an order L on D which maximizes the sum of the probabilities of the arcs of $D \cap L$ (here the probability of an arc is the product of the probabilities of its two endvertices). The maximal vertex of L is the vertex x we are looking for. $\square$

However, the feedback method fails dramatically for digraphs in general, as one can check in the following example: consider the circuit on four elements $a \to b \to c \to d \to a$, and the probability distribution $p(a) = 3/10$, $p(b) = 2/10$, $p(c) = 4/10$ and $p(d) = 1/10$. Here the (weighted) median order is (a, b, c, d), but, alas, d does not have a large second neighbourhood.

3 Median orders and Sumner's conjecture.

An *oriented tree* (or simply *tree*) is an orientation of an acyclic connected graph. An *arborescence* is an oriented tree in which one vertex called the *root* has indegree zero and the remaining vertices have indegree one. An *outleaf* (resp. *inleaf*) of a tree A is a vertex

x such that $d_A^+(x) = 0$ and $d_A^-(x) = 1$ (resp. $d_A^-(x) = 0$ and $d_A^+(x) = 1$). Let A be a tree, T a tournament and L a local median order of T. An *embedding* of A into T is an injective mapping $f : V(A) \to V(T)$ such that $f(x) \to f(y)$ whenever $x \to y$. A digraph D is *m-unavoidable* if, for every tournament T of order m, there exists an embedding of D into T.

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

An embedding f of A into T is an *L-embedding* if, for every vertex $x \in T$, the following two conditions hold:

$$|N_L^+(x) \cap f(A)| \leq |N_L^+(x) \setminus f(A)| + 1$$
$$|N_L^-(x) \cap f(A)| \leq |N_L^-(x) \setminus f(A)| + 1.$$

If f only satisfies the first inequality (resp. the second inequality), we speak of *L-up-embedding* (resp. *L-down-embedding*). A tree is *L-embeddable* into T (resp. *L-up-embeddable* into T) if there exists an L-embedding of A into T (resp. *L-up-embedding* of A into T). A tree A is *m-well-embeddable* (resp. *m-well-up-embeddable*) if for every tournament T of order m and every local median order L of T, A is L-embeddable (resp. *L-up-embeddable*) into T.

Theorem 3 *Every arborescence of order $n > 1$ is $(2n-2)$-unavoidable.*

Proof. We prove by induction on n the following stronger statement: every arborescence A of order $n > 1$ is $(2n-2)$-well-up-embeddable. This is true if A is an arc. If $n > 2$, consider a tournament T on $2n-2$ vertices and $L = (x_1, \ldots, x_{2n-2})$ a local median order of T. Denote by (T', L') the restriction of (T, L) to $[x_1, x_{2n-4}]$. Let x be an outleaf of A, y the inneighbour of x and denote by A' the arborescence $A \setminus \{x\}$. By the induction hypothesis, there is an L'-up-embedding f of A' into T'. Denote by x_i the vertex $f(y)$. We have

$$|N_L^+(x_i) \cap f(A')| = |N_{L'}^+(x_i) \cap f(A')| \leq |N_{L'}^+(x_i) \setminus f(A')| + 1 = |N_L^+(x_i) \setminus f(A')| - 1.$$

In particular, $|N_L^+(x_i) \cap f(A')| < |N_L^+(x_i) \setminus f(A')|$. The feedback property applied to the interval $[x_i, x_{2n-2}]$ of L ensures that at least one vertex x_j of $N_L^+(x_i) \setminus f(A')$ belongs to $N^+(x_i)$. Extend now f by letting $f(x) = x_j$. It is routine to check that this extension of f is an L-up-embedding of A into T (indeed, we add two new vertices to the top of L' whereas $f(A')$ only increases by one vertex). $\square$

Observe that, for arborescences, the same proof gives a little more than Sumner's conjecture: in every tournament of order $2n-2$, there is a particular vertex x and an acyclic subdigraph D for which every arborescence on n vertices is contained in D and rooted at x. Consider for this any back vertex x of a local median order L and take $D = L \cap T$.

Theorem 4 *Every tree of order $n > 1$ is $(4n-6)$-unavoidable.*

Proof. We prove, again by induction on n, that every tree A of order $n > 1$ is $(4n-6)$-well-embeddable. This is true when A is an arc. If $n > 2$, consider T a tournament on $4n-6$ vertices and $L = (x_1, \dots, x_{4n-6})$ a local median order of T. Denote by (T', L') the restriction of (T, L) to $[x_3, x_{4n-8}]$. Let x be an outleaf, if any, of A. Let y be the inneighbour of x and denote by A' the tree $A \setminus \{x\}$. By the induction hypothesis, there is an L'-embedding f of A' into T'. Denote the vertex $f(y)$ by x_i. Note that $|N^+_{L'}(x_i) \cap f(A')| \leq |N^+_{L'}(x_i) \setminus f(A')| + 1$, and the vertices x_{4n-7}, x_{4n-6} ensure that $|N^+_L(x_i) \cap f(A')| < |N^+_L(x_i) \setminus f(A')|$. The feedback property applied to the interval $[x_i, x_{4n-6}]$ of L provides at least one vertex x_j in $(N^+_L(x_i) \setminus f(A')) \cap N^+(x_i)$. Now extend f by letting $f(x) = x_j$. It is routine to check that this extension of f is an L-embedding of A into T, indeed we add two new vertices to both ends of L'. A similar argument works for an inleaf of A. $\square$

Observe again that, given a tournament T of order $4n-6$, there exists an acyclic digraph D of T and a particular vertex x of T such that for every tree A of order n with a fixed vertex v, there is an embedding f of A into D such that $f(v) = x$. Unfortunately, we do not see how the proof of Theorem 4 could naturally be improved to give the bound $2n-2$. Indeed, $4n$ is really a critical value for this problem, and it is easy to understand: we have, *a priori*, no way to decide if a given vertex of T will be considered as an outleaf or as an inleaf in the inductive construction of the tree A. So, to be sure that we will complete the tree, we need to have available twice as many vertices as necessary. This is also the reason why the asymptotic bound obtained by Häggkvist and Thomason was $4n$. However, this is by no means the end of the road. We prove in the following section that the bound can be reduced to $7n/2$. But this entails case analysis and counting arguments.

4 Beyond $4n$, a bound in $7n/2$.

To obtain a better bound, we need a construction of trees which is a little bit more elaborate than just adding one leaf at a time. Indeed, the operations involved here use paths of length 3. To describe this method, let us introduce some notation: let A be an oriented tree and $a_1, \dots, a_k$ an oriented path. When we write $A \cup a_1, \dots, a_k$, it is implicitly assumed that the sole vertex of intersection of A and $a_1, \dots, a_k$ is a_1, and hence that, the resulting digraph is always a tree. Note that since the choice of a_1 in A is free, the tree is not unique.

Lemma 3 *If A is m-well-up-embeddable, then $A \cup a \to b$ is $(m+2)$-well-up-embeddable. If A is m-well-embeddable, then $A \cup a \to b$ is $(m+4)$-well-embeddable.*

The proof of Lemma 3 is contained in the proofs of Theorems 3 and 4.

Lemma 4 *Let A' be a tree of order n and $A = A' \cup a \to b \to c$. If A' is m-well-embeddable with $m \geq \frac{10n-12}{3}$, then A is $(m+6)$-well-embeddable.*

Proof. Let $L = (x_1, \ldots, x_{m+6})$ be a local median order of a tournament T. Denote by L' and T' the restrictions of L and T to the interval $[x_3, x_{m+2}]$. By the hypothesis of the lemma, there exists an L'-embedding f of A' into T'; let $x_i = f(a)$. By the feedback property and since f is an L'-embedding, at least one vertex x_j of $[x_{i+1}, x_{m+4}] \setminus f(A')$ is an outneighbour of x_i; we choose such an x_j with maximal index. Again note that x_j has an outneighbour x_k (also chosen with maximal index) in $[x_{j+1}, x_{m+6}] \setminus f(A')$. We now extend f by setting $f(b) = x_j$ and $f(c) = x_k$; for convenience we still call this extension f. We claim that f is an L-embedding of A into T. One part of the proof is routine: since we added four vertices to the top of L', f is clearly an L-up-embedding. The critical point is to prove that f is also an L-down-embedding. Assume, by way of contradiction, that this is not the case, so there exists a vertex x_l such that

$$|[x_1, x_l] \cap f(A)| > |[x_1, x_l] \setminus f(A)| + 1.$$

Recall that we added the vertices x_1 and x_2 to L', and by construction these vertices do not belong to $f(A)$. For this reason we have necessarily that $l \geq 3$. Note also that since $m \geq 2n-1$ (because A' is m-well-embeddable), we have $l \leq m+2$. Thus:

$$|[x_3, x_l] \cap f(A)| \geq |[x_3, x_l] \setminus f(A)| + 4.$$

However, since f is an L'-embedding of A' into T', we have:

$$|[x_3, x_l] \cap f(A')| \leq |[x_3, x_l] \setminus f(A')| + 1.$$

These two inequalities imply that $k \leq l$ and:

$$|[x_3, x_l] \cap f(A')| \geq |[x_3, x_l] \setminus f(A')|. \tag{2}$$

Set

$$\begin{aligned}
A_1 &= [x_3, x_i] \cap f(A') \quad , \quad D_1 = [x_3, x_i] \setminus f(A') \\
A_2 &= [x_{i+1}, x_j] \cap f(A') \quad , \quad D_2 = [x_{i+1}, x_j] \setminus f(A') \\
A_3 &= [x_{j+1}, x_l] \cap f(A') \quad , \quad D_3 = [x_{j+1}, x_l] \setminus f(A') \\
A_4 &= [x_{l+1}, x_{m+2}] \cap f(A') \quad , \quad D_4 = [x_{l+1}, x_{m+2}] \setminus f(A').
\end{aligned}$$

We write also $a_i = |A_i|$ and $d_i = |D_i|$ for $i \in \{1, 2, 3, 4\}$. In T, by the maximality of j and k, the vertex x_i is dominated by $D_3 \cup D_4 \cup \{x_{m+3}, x_{m+4}\}$ and x_j is dominated by $D_4 \cup \{x_{m+3}, x_{m+4}, x_{m+5}, x_{m+6}\}$.

In $T_{|[x_i,x_{m+4}]}$, the feedback property ensures that $d^+(x_i) \geq d^-(x_i)$. It follows that:

$$d_3 + d_4 + 2 \leq a_2 + a_3 + a_4 + d_2. \tag{3}$$

Since in $T_{|[x_j,x_{m+6}]}$, we have $d^+(x_j) \geq d^-(x_j)$,

$$d_4 + 4 \leq a_3 + a_4 + d_3. \tag{4}$$

By Inequality 2, we have:

$$d_1 + d_2 + d_3 \leq a_1 + a_2 + a_3. \tag{5}$$

And since $m \geq \frac{10n-12}{3}$, we have:

$$\frac{10(a_1 + a_2 + a_3 + a_4) - 12}{3} \leq a_1 + a_2 + a_3 + a_4 + d_1 + d_2 + d_3 + d_4$$

.

That is :

$$\frac{7}{3}(a_1 + a_2 + a_3 + a_4) \leq d_1 + d_2 + d_3 + d_4 + 4. \tag{6}$$

Inequalities 5 and 6 yield:

$$\frac{4}{3}(a_1 + a_2 + a_3) + \frac{7}{3}a_4 \leq d_4 + 4. \tag{7}$$

Inequality 3 and twice Inequality 4 yield:

$$3d_4 + 10 \leq a_2 + 3a_3 + 3a_4 + d_2 + d_3. \tag{8}$$

Combining Inequalities 5 and 8, it follows that:

$$3d_4 + 10 \leq a_1 + 2a_2 + 4a_3 + 3a_4. \tag{9}$$

Finally, Inequality 7 multiplied by 3 and Inequality 9 give that $3a_1 + 2a_2 + 4a_4 \leq 2$. This is a contradiction since $a_1 > 0$. □

This lemma and Lemma 3 yield the following corollaries:

Corollary 1 *Let A' be a tree of order n and $A = A' \cup a \to b \to c \to d$. If A' is m-well-embeddable with $m \geq \frac{10n-14}{3}$ then A is $(m + 10)$-well-embeddable.*

Corollary 2 *Let A' be a tree of order n and $A = A' \cup a \to b \to c \leftarrow d$. If A' is m-well-embeddable with $m \geq \frac{10n-12}{3}$ then A is $(m+10)$-well-embeddable.*

Corollary 3 *Let A' be a tree of order n and $A = A' \cup a \to b \leftarrow c \leftarrow d$. If A' is m-well-embeddable with $m \geq \frac{10n-14}{3}$ then A is $(m+10)$-well-embeddable.*

Up to this point, we are able to add all the paths of length three, except the alternating path. For this particular path, we have the following result.

Lemma 5 *Let A' be a tree of order n and $A = A' \cup a \to b \leftarrow c \to d$. If A' is m-well-embeddable with $m \geq 3n$ then A is $(m+10)$-well-embeddable.*

Proof. Let $L = (x_1, \ldots, x_{m+10})$ be a local median order of a tournament T. Denote by T' and L' the restrictions of T and L to $[x_5, x_{m+4}]$. There exists an L'-embedding f of A' into T'. Let $f(a) = x_i$. Denote by h the greatest index in $\{5, \ldots, m+4\}$, if it exists, such that $|[x_5, x_h] \cap f(A')| \geq |[x_5, x_h] \setminus f(A')|$. By the feedback property and since f is an L'-embedding, there exists x_j in $[x_i, x_{m+6}]$, chosen with maximal index, such that $x_i \to x_j$ and $x_j \notin f(A')$. Suppose first that $j > h$ or h does not exist. Setting $f(b) = x_j$, still f is an $L_{|[x_5, x_{m+6}]}$-embedding of $A' \cup a \to b$ into $T_{|[x_5, x_{m+6}]}$. Then by applying Lemma 3 twice, one can extend f to an L-embedding of A into T.

Suppose now that $j \leq h$. We prove that x_i dominates two vertices of $[x_i, x_h] \setminus f(A')$. Set $A_1 = [x_5, x_i] \cap f(A')$ and $D_1 = [x_5, x_i] \setminus f(A')$, $A_2 = [x_{i+1}, x_h] \cap f(A')$ and $D_2 = [x_{i+1}, x_h] \setminus f(A')$, $A_3 = [x_{h+1}, x_{m+4}] \cap f(A')$ and $D_3 = [x_{h+1}, x_{m+4}] \setminus f(A')$. We also let $a_i = |A_i|$ and $d_i = |D_i|$ for $i = 1, 2, 3$.

Suppose for contradiction that $|N^+(x_i) \cap D_2| \leq 1$. Using the feedback property of x_i in $[x_i, x_{m+6}]$, we obtain (since $D_3 \cup \{x_{m+5}, x_{m+6}\} \subseteq N^-(x_i)$):

$$d_2 + d_3 \leq a_2 + a_3. \tag{10}$$

By definition of h, we have:

$$d_1 + d_2 \leq a_1 + a_2. \tag{11}$$

Since $m \geq 3n$, we have:

$$2(a_1 + a_2 + a_3) \leq d_1 + d_2 + d_3. \tag{12}$$

Inequalities 11 and 12 yield:

$$a_1 + a_2 + 2a_3 \leq d_3. \tag{13}$$

Then, combining Inequalities 10 and 13 gives:

$$a_1 + d_2 + a_3 \leq 0.$$

And this is a contradiction since A_1 contains at least x_i. Now, pick two vertices x_j and x_k in $N^+(x_i) \cap D_2$. Without loss of generality we may suppose that $x_j \to x_k$. If $x_j \leftarrow x_{m+8}$, then let $f(b) = x_j$ and $f(c) = x_{m+8}$. One can easily check that f is an $L_{|[x_3,x_{m+8}]}$-embedding of $A' \cup a \to b \leftarrow c$ into $T_{|[x_3,x_{m+8}]}$. Thus, by Lemma 3, one can extend f into an L-embedding of A. If $x_j \to x_{m+8}$, we set $f(b) = x_k$, $f(c) = x_j$ and $f(d) = x_{m+8}$. Again, f is an L-embedding of A into T. $\square$

Note that an analogous proof gives an improvement of the bound of Corollary 2: $m \geq 3n$ in place of $m \geq \frac{10n-12}{3}$. To achieve the proof, we term *star with center* x a tree T with a particular vertex x such that every vertex of T distinct from x is a leaf. The class of trees $\mathcal{T}_3$ is defined inductively as follows: the singleton is in $\mathcal{T}_3$. If A is in $\mathcal{T}_3$ and P is a path of length 3, then $A \cup P$ is in $\mathcal{T}_3$.

Lemma 6 *Let A be a tree of $\mathcal{T}_3$. If the order of A is $3n+1$, then A is $(10n+1)$-well-embeddable.*

Proof. By induction on n. If $n = 0$, the statement is obviously true. If $n = 1$, this is a consequence of Theorem 4. So, assume that the conclusion holds for $n \geq 1$. Let $A' \in \mathcal{T}_3$ be a tree of order $3n+1$, P a path of length 3 and denote by A the tree $A' \cup P$. By the induction hypothesis, A' is $(10n+1)$-well-embeddable. Thus, by Lemma 5 and Corollaries 1, 2 and 3, A is $(10n+11)$-well-embeddable. $\square$

Theorem 5 *Every tree of order $n > 0$ is $(\frac{7n-5}{2})$-unavoidable.*

Proof. Let A be a tree of order n and A_1 be a maximal subtree of A which belong to $\mathcal{T}_3$. Denote the order of A_1 by n_1, by Lemma 6, A_1 is $(\frac{10n_1-7}{3})$-well-embeddable. The forest $A \setminus A_1$ is the union of l isolated vertices and p stars S_i $(1 \leq i \leq p)$, with respective centers x_i. Note that each x_i is connected to A_1 by an arc. Let A_2 be the subtree induced by A on $V(A_1) \cup \{x_1, \ldots, x_p\}$. By Lemma 3, A_2 is $(\frac{10n_1-7}{3} + 4p)$-well-embeddable and $A \setminus A_2$ is the union of $k \geq p + l$ isolated vertices. Let I be the set of vertices of $A \setminus A_2$ which are inleaves of A; we set $i = |I|$. By directional duality, we may suppose that $i \leq k/2$. Let A_3 be the subtree of A induced by the vertices of $V(A_2) \cup I$. By Lemma 3, A_3 is $(\frac{10n_1-7}{3} + 4(p+i))$-well-embeddable. Moreover, $A \setminus A_3$ is a subset of the outleaves of A. Thus by Lemma 3, A is $(\frac{10n_1-7}{3} + 4(p+i) + 2(k-i))$-well-up-embeddable. Since $i \leq k/2$, $k \geq p$ and $k + p = n - n_1$, we have:

$$4(p+i) + 2(k-i) = 4p + 2k + 2i \leq 4p + 3k \leq \frac{7}{2}(n - n_1).$$

These inequalities together yield:

$$\frac{10n_1-7}{3}+4(p+i)+2(k-i)\leq\frac{10n_1-7}{3}+\frac{7}{2}(n-n_1)\leq\frac{21n-n_1-14}{6}\leq\frac{7n-5}{2}.$$

So, the tree A is $(\frac{7n-5}{2})$-unavoidable. □

We gratefully thank J.A. Bondy for his help during the preparation of this paper.

References

[1] I. Charon, A. Guénoche, O. Hudry and F. Woirgard, New results on the computation of median orders, *Discrete Math.*, **165/166** (1997), 139–153.

[2] N. Dean and B. J. Latka, Squaring the tournament—an open problem. Proceedings of the Twenty-sixth Southeastern International Conference on Combinatorics, Graph Theory and Computing (Boca Raton, FL, 1995). *Congr. Numer.*, **109** (1995), 73–80.

[3] D. C. Fisher, Squaring a tournament: a proof of Dean's conjecture. *J. Graph Theory*, **23** (1996), 43–48.

[4] D. C. Fisher and J. Ryan, Tournament games and positive tournaments. *J. Graph Theory*, **19** (1995), 217–236.

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

[6] F. Havet, Trees in tournaments, preprint.

[7] K. B. Reid and N. C. Wormald, Embedding oriented n-trees in tournaments. *Studia Sci. Math. Hungar.*, **18** (1983), 377–387.

[8] N. C. Wormald, Subtrees of large tournaments, in Combinatorial Mathematics, X (Adelaide, 1982), *Lecture Notes in Math.*, **1036** (1983), 417–419.