

Covering small subgraphs of (hyper)tournaments with spanning acyclic subgraphs

Raphael Yuster*

Department of Mathematics
University of Haifa
Haifa 3498838, Israel
raphael.yuster@gmail.com

Submitted: Feb 2, 2020; Accepted: Oct 5, 2020; Published: Oct 16, 2020

© The author. Released under the CC BY-ND license (International 4.0).

Abstract

While the edges of every tournament can be covered with two spanning acyclic subgraphs, this is not so if we set out to cover all acyclic H -subgraphs of a tournament with spanning acyclic subgraphs, even for very simple H such as the 2-edge directed path or the 2-edge out-star. We prove new bounds for the minimum number of elements in such coverings and for some H our bounds determine the exact order of magnitude.

A k -tournament is an orientation of the complete k -graph, where each k -set is given a total order (so tournaments are 2-tournaments). As opposed to tournaments, already covering the edges of a 3-tournament with the minimum number of spanning acyclic subhypergraphs is a nontrivial problem. We prove a new lower bound for this problem which asymptotically matches the known lower bound of covering all ordered triples of a set.

Mathematics Subject Classifications: 05C20, 05C35, 05C70

1 Introduction

Our main objects of study are tournaments and, more generally, directed graphs and directed hypergraphs. A *tournament* with n vertices is obtained by assigning an orientation to each edge of the complete graph K_n . For a tournament $T = ([n], E)$ we have $|E| = \binom{n}{2}$ where for any two distinct vertices i, j , either $(i, j) \in E$ or $(j, i) \in E$, but not both. Every n -vertex tournament is, in particular, a spanning subgraph of the complete directed n -vertex graph $D(n)$ which consists of all possible $n(n - 1)$ edges.

*Supported by Israel Science Foundation (grant No. 1082/16).

It is straightforward that the edge set of $D(n)$ is the disjoint union of two acyclic (i.e. transitive) tournaments. Indeed, consider some permutation $\pi \in S_n$ of the vertices. Let T_L be the tournament where $(i, j) \in E(T_L)$ if and only if $\pi(i) < \pi(j)$ and let T_R be the tournament where $(i, j) \in E(T_R)$ if and only if $\pi(i) > \pi(j)$. Then $E(T_R) \cup E(T_L) = E(D(n))$. Observe that T_R and T_L are both transitive tournaments. In particular as every n -vertex tournament T is a subgraph of $D(n)$, we can cover the edges of T using just two acyclic subgraphs of T (hereafter we use the terminology *acyclic graph* to refer to a directed acyclic graph).

However, the aforementioned edge-covering observation becomes significantly more involved if instead of just covering edges, we set out to cover all fixed H -subgraphs of a tournament with as few as possible acyclic subgraphs. Of course, for this to be meaningful we assume that H itself is an acyclic graph. We next state the problem formally.

Let $T = ([n], E)$ be a tournament and let $\pi \in S_n$ be a permutation of its vertices. The spanning acyclic subgraph of T corresponding to π , denoted by $L_\pi(T)$, consists of all the edges $(i, j) \in E(T)$ with $\pi(i) < \pi(j)$. We can visualize all the edges of $L_\pi(T)$ as going from “left to right”. We say that a subgraph of T is *covered* by the permutation π if it is a subgraph of $L_\pi(T)$.

Let H be a fixed acyclic graph with at least two edges (hence at least three vertices; we also assume that H has no isolated vertices as these can be discarded in our problem). A subgraph of T isomorphic to H is called an H -copy of T . What will then be the minimum number of permutations required to cover all H -copies of T ? Thus, we seek the smallest integer t such that for every tournament T with n vertices the following holds: There are permutations $\pi_1, \dots, \pi_t \in S_n$ such that each H -copy of T is covered by at least one of the π_i . We denote this t by $\tau_H(n)$. Trivially, $\tau_H(n)$ exists as we can just consider all $n!$ permutations and use the fact that each H -copy, being an acyclic graph, has a topological ordering. Hence our main problem is the following.

Problem 1. Let H be an acyclic graph with at least two edges. Determine $\tau_H(n)$.

1.1 Sequence covering arrays as upper bounds for $\tau_H(n)$

Reasonable upper bounds for $\tau_H(n)$ can rather easily be obtained by considering a related well-studied problem in the setting of permutations. Let T_k denote the transitive tournament on k vertices. An (n, k) -sequence covering array (SCA) is a set of permutations $X \subseteq S_n$ such that each T_k -copy of $D(n)$ is covered by at least one of the elements of X . Let $s(k, n)$ denote the smallest size of an (n, k) -SCA. Stated otherwise, we wish to find the smallest set of permutations of $[n]$ such that each sequence of k distinct elements of $[n]$ is a subsequence of at least one of the permutations.

While trivially $s(2, n) = 2$ by the earlier observation regarding $D(n)$ (taking any permutation and its reverse), even the asymptotic value of $s(3, n)$, though well-studied, is not known. A-priori, for a constant k , it is not entirely obvious that $s(k, n)$ grows to infinity with n , as each permutation contains $\binom{n}{k}$ sequences of order k and there are only $k! \binom{n}{k}$ possible sequences to cover. However, more is known than this trivial $k!$ lower bound for $s(k, n)$. The first to provide nontrivial bounds for $s(k, n)$ was Spencer [14]

and various improvements on the upper and lower bounds were sequentially obtained by Ishigami [9, 10], Füredi [7], Radhakrishnan [13], and Tarui [16]. See also the paper [4] for further results and references to many applications. The (asymptotic) state of the art regarding $s(3, n)$ is the upper bound of Tarui [16] and the lower bound of Füredi [7]:

$$\frac{2}{\log e} \log n \leq s(3, n) \leq (1 + o_n(1))2 \log n .^1 \quad (1)$$

We note that the limit $s(3, n)/\log n$ exists [7, 16], but its value is not known. For general fixed k , the best asymptotic upper and lower bounds are that of Spencer [14] and Radhakrishnan [13], respectively:

$$(1 - o_n(1)) \frac{(k-1)!}{\log e} \log n \leq s(k, n) \leq \frac{k}{\log(\frac{k!}{k!-1})} \log n . \quad (2)$$

It is immediate to see that if H has k vertices, then $\tau_H(n) \leq s(k, n)$, hence (1) and (2) serve as upper bounds for $\tau_H(n)$ when H has three vertices or, respectively, k vertices.

Another known result is an upper bound for arguably the simplest nontrivial case $H = S_2^+$, the directed out-star on two edges. Namely, H has vertices a, b, c and edges $(a, b), (a, c)$ (by symmetry, the same result holds for the in-star on two edges S_2^-). This nontrivial upper bound, proved in [8, 14] is

$$\tau_{S_2^+}(n) \leq \log \log n + \left(\frac{1}{2} + o_n(1) \right) \log \log \log n . \quad (3)$$

In fact, their construction is universal in the sense that the same construction holds for all tournaments, as it even holds for covering all the S_2^+ of $D(n)$. The permutations in the construction are such that for any $a \in [n]$ and any two distinct elements $b, c \in [n] \setminus \{a\}$ there is a permutation in which a appears before b and before c . Their construction is asymptotically tight for this latter requirement.

1.2 New results on $\tau_H(n)$

Our first main result consists of lower bounds for $\tau_H(n)$ for each acyclic H on three vertices. Thus we consider $\tau_H(n)$ for each $H \in \{S_2^+, S_2^-, P_3, T_3\}$ where P_3 is the directed path on two edges and T_3 is the transitive tournament on three vertices.

Theorem 2. *The following holds for all sufficiently large n .*

1. For $H \in \{S_2^+, S_2^-\}$ we have that $\log \log n - \log \log \log n - 2 \leq \tau_H(n)$ so with (3) we obtain:

$$\log \log n - \log \log \log n - 2 \leq \tau_H(n) \leq \log \log n + \left(\frac{1}{2} + o_n(1) \right) \log \log \log n .$$

¹Unless stated otherwise, all logarithms are in base 2.

2. For $H \in \{P_3, T_3\}$ we have that $\log n / (3 + \log \log n) \leq \tau_H(n)$ so with (1) we obtain:

$$\frac{\log n}{3 + \log \log n} \leq \tau_H(n) \leq (1 + o_n(1))2 \log n .$$

Observe that for $H \in \{S_2^+, S_2^-\}$ the upper and lower bounds stated in Theorem 2 are tight up to the triply logarithmic error term. For $H \in \{P_3, T_3\}$ the upper and lower bounds in Theorem 2 are close, but not tight: the ratio between them is only doubly logarithmic. Compare this to the upper and lower bounds for $s(3, n)$ given in (1). The latter are also not tight but the ratio there is a constant. Indeed, it is plausible that covering all T_3 (or all P_3) of any given n -vertex tournament requires significantly fewer elements than the number of elements required to cover all T_3 of $D(n)$.

Theorem 2 shows that $\tau_H(n)$ exhibits a dichotomy for acyclic graphs with at most three vertices. It is either doubly logarithmic in the case of $\{S_2^+, S_2^-\}$ or else it is essentially logarithmic (meaning it is $\Omega(\log n / \log \log n)$ and $O(\log n)$) in the case of $\{P_3, T_3\}$ or else, if H is just a single edge, then it is a constant.

We next consider the general case of arbitrary fixed H . First notice that if H' is a (not necessarily induced) subgraph of H , the inequality $\tau_{H'}(n) \leq \tau_H(n)$ is not obvious and may not be true. Nevertheless, we can prove the following lower bound based on whether H has a subgraph consisting of two edges or a subgraph which is a directed path with two edges.

Theorem 3. *Let H be an acyclic graph with at least two edges. Then, for all n sufficiently large, it holds that $\tau_H(n) \geq \log \log n - \log \log \log n - 2$. Furthermore, if H contains a directed path on two edges, then $\tau_H(n) \geq \log n / (3 + \log \log n)$.*

So, we see from Theorem 3 and from (2) that $\tau_H(n)$ is essentially logarithmic if H contains a directed path on two edges. In any case, already the existence of two edges (even if H itself consists of just two disjoint edges) implies a doubly logarithmic lower bound. Regarding graphs H which contain at least two edges and no directed path on two edges (notice that such graphs are always bipartite), we do not know whether it is always true that $\tau_H(n)$ is just doubly logarithmic, except for the following case. Let S_k^+ (S_k^-) denote the directed out-star (in-star) on k edges. Recall that we have proved the doubly logarithmic behavior of S_2^+ and S_2^- in Theorem 2. Generalizing (3), it has been proved by Hajnal and Spencer [14] that $\tau_{S_k^+}(n) = \tau_{S_k^-}(n) = O(\log \log n)$. So, together with Theorem 3, we have that $\tau_{S_k^+}(n) = \tau_{S_k^-}(n) = \Theta(\log \log n)$. The next theorem exhibits another basic family of acyclic graphs for which the right order of magnitude of $\tau_H(n)$ is determined. Let P_k denote the directed path on k vertices.

Theorem 4. *For all $k \geq 4$, $\tau_{P_k}(n) = \Theta(\log n)$.*

The theorems stated in this subsection are proved in Section 2.

1.3 3-tournaments

Recall that K_n^k , the complete k -graph on n vertices, consists of vertex set $[n]$ and edge set $\binom{[n]}{k}$. Just as tournaments are orientations of K_n , k -tournaments are orientations of K_n^k . That is, each edge of K_n^k is given a unique total order. An oriented k -graph is a subgraph of a k -tournament. We refer to [2] for more information and references on k -tournaments. Let $D(n, k)$ denote the complete directed k -graph on n vertices, namely each k -set of $[n]$ corresponds to $k!$ edges in $D(n, k)$, one edge for each possible total order. Observe that $D(n) = D(n, 2)$ and that every k -tournament with n vertices is a spanning sub(hyper)graph of $D(n, k)$. As the edge set of $D(n, k)$ is the set of all sequences of k distinct elements of $[n]$ (and hence corresponds bijectively to the T_k -copies of $D(n)$) we have that $s(k, n)$ can be defined equivalently as the smallest number of permutations that cover all edges of $D(n, k)$ (here an edge is covered by a permutation if it is a subsequence of that permutation). Let $\tau(k, n)$ be the smallest integer t such that the edges of each k -tournament with n vertices can be covered with t permutations (or, equivalently, covered by t oriented spanning acyclic k -graphs²). Hence, $\tau(k, n) \leq s(k, n)$ so (1) and (2) respectively serve as upper bounds for $\tau(3, n)$ and $\tau(k, n)$. But is $\tau(k, n)$ significantly smaller than $s(k, n)$? We conjecture it is not.

Conjecture 5. There is an absolute constant C_k such that $s(k, n) \leq \tau(k, n) + C_k$.

At least for $k = 3$, the stipulated closeness of both parameters is supported by the following result, which shows that the lower bound for $s(3, n)$ given in (1) holds asymptotically in a much more restrictive setting, and can be generalized to most oriented 3-graphs. To state the result in its general form, we consider the analogous model of the Erdős-Rényi random graph to oriented k -graphs. Let $\vec{G}_k(n, p)$ be the probability space of oriented k -graphs on vertex set $[n]$ where each k -set of $[n]$ is chosen independently to be an edge with probability p and, once chosen, its orientation is chosen uniformly among all possible $k!$ total orders. Notice that when $p = 1$ this amounts to the uniform distribution on labeled n -vertex k -tournaments. For an oriented k -graph, let $\tau(G)$ denote the smallest number of permutations that cover all the edges of G .

Theorem 6. Fix $0 \leq \alpha < 1$ and let $G \sim \vec{G}_3(n, n^{-\alpha})$.

$$\Pr \left[\tau(G) \leq (1 - \alpha) \frac{2}{\log e} \log n - 9 \log \log n \right] = o_n(1) .$$

In particular, for all sufficiently large n ,

$$\tau(3, n) \geq \frac{2}{\log e} \log n - 9 \log \log n .$$

Theorem 6 is proved in Section 3. The final section of the paper consists of some open problems and concluding remarks among which is a simple improvement of the upper bound for $s(k, n)$ given in (2).

²Here we say that an oriented k -graph is acyclic if there is a total order of its vertex set which is respected by each edge; we note that there are other notions of acyclicity for oriented k -graphs.

2 Covering the acyclic subgraphs of a tournament

2.1 Covering acyclic graphs on three vertices

As mentioned in the introduction, the upper bounds stated in Theorem 2 are known. The following three lemmata together yield the required lower bounds.

Lemma 7. $\log \log n - \log \log \log n - 2 \leq \tau_{S_2^+}(n)$.

Proof. Let $\vec{G}(n)$ be the symmetric probability space of all labeled tournaments on vertex set $[n]$ and let $G = ([n], E) \sim \vec{G}(n)$.

Consider some $X \subset [n]$ of $3k$ vertices, and let its vertices be v_1, \dots, v_{3k} where $v_i < v_{i+1}$ for $i = 1, \dots, 3k - 1$. We compute the probability of the event B_X , that $G[X]$ (the subtournament of G induced by X) has no three vertices v_a, v_b, v_c with $a < b < c$ such that $(v_b, v_a) \in E$ and $(v_b, v_c) \in E$. In other words, for B_X to occur, $G[X]$ must have no subgraph isomorphic to S_2^+ where the root vertex is smaller than one leaf and larger than another leaf. To estimate B_X , let $B_{i,j}$ denote the following event, defined for each $i = 1, \dots, k$ and for each $j = 1, \dots, k$: $B_{i,j}$ occurs if $(v_{k+j}, v_i) \in E$ and $(v_{k+j}, v_{2k+i}) \in E$. Clearly $\Pr[B_{i,j}] = \frac{1}{4}$ and the k^2 events $B_{i,j}$ are independent since they involve distinct edges. Finally, if B_X occurs, then none of the $B_{i,j}$ occur. Hence,

$$\Pr[B_X] \leq \left(\frac{3}{4}\right)^{k^2}.$$

In particular, choosing $k = \lfloor 8 \log n \rfloor$

$$\Pr \left[\bigcup_{X \subset [n], |X|=3k} B_X \right] \leq \binom{n}{3k} \left(\frac{3}{4}\right)^{k^2} < n^{3k} \left(\frac{3}{4}\right)^{k^2} < 1$$

where we have used the fact that $\lfloor 8 \log n \rfloor > \frac{3}{\log(4/3)} \log n$.

As this shows, we can now fix a tournament $G = ([n], E)$ with the property \mathcal{P} that for each $X \subset [n]$ of size $\lfloor 24 \log n \rfloor$, the subtournament $G[X]$ has an S_2^+ where the root vertex is smaller than one leaf and larger than another leaf.

Assume that n is a double exponent of 2, that is, $n = 2^{2^t}$ for some nonnegative integer t . We claim that we need at least $t - \log(t+5)$ permutations of S_n to cover all copies of S_2^+ of G . In order to see this, suppose π_1, \dots, π_m are permutations with $m \leq t - \log(t+5)$. By a theorem of Erdős and Szekeres [6], π_1 has a monotone subsequence of length $\sqrt{n} = 2^{2^{t-1}}$. As observed by de Bruijn (see [12]), by repeatedly applying this theorem, there is a subset X of $2^{2^{t-m}}$ vertices that appears as a monotone sequence in *each* of π_1, \dots, π_m . Of course, in some of the permutations it may be monotone increasing while in others it may be monotone decreasing. But now, by our assumption on m ,

$$2^{2^{t-m}} \geq 2^{2^{\log(t+5)}} = 2^{t+5} = 32 \cdot 2^t = 32 \log n > 24 \log n.$$

So, by the property \mathcal{P} of G , there is an S_2^+ in $G[X]$ that is not covered by any of the permutations π_1, \dots, π_m . Hence, we see that $\tau_{S_2^+}(n) \geq t - \log(t+5) \geq \log \log n - \log(\log \log n + 5)$.

If we omit the assumption that n is a double exponent of 2 we can use the largest number smaller than n which is a double exponent of 2 and consequently obtain that $\tau_{S_2^+}(n) \geq \log \log n - \log \log \log n - 2$. \square

Lemma 8. $\log n / (3 + \log \log n) \leq \tau_{T_3}(n)$.

Proof. As in the previous lemma, let $G = ([n], E) \sim \vec{G}(n)$. We prove that

$$\Pr \left[\tau_{T_3}(G) \leq \frac{\log n}{3 + \log \log n} \right] = o_n(1)$$

where $\tau_H(G)$ is the number of permutations needed to cover all H -copies of G . Let $X \subset [n]$ be a set of vertices. We say that X has the *midpoint property*, if for any two vertices $u, v \in [n] \setminus X$ there exists some vertex $x \in X$ such that $\{u, x, v\}$ induce in G a T_3 in which x is the middle vertex. Equivalently, if $(u, v) \in E$, then $(u, x), (x, v) \in E$ else if $(v, u) \in E$, then $(v, x), (x, u) \in E$.

We claim that with high probability there exists X with the midpoint property and with $|X| = \lfloor 5 \log n \rfloor - 1$. Consider two vertices $u, v \in [n] \setminus X$. Because of symmetry, one can assume $(u, v) \in E$. For a fixed $x \in X$, the probability that $(u, x) \notin E$ or $(x, v) \notin E$ is $\frac{3}{4}$. Hence, the probability that there does not exist a vertex $x \in X$ such that $\{u, x, v\}$ induce a transitive triple in which x is the middle vertex is $(3/4)^{|X|}$. Here we use the fact that the probability of this event is the product of the probabilities of $|X|$ mutually independent events, each with probability $\frac{3}{4}$. As there are less than $\binom{n}{2}$ pairs u, v to consider, we have that the probability that X does not have the midpoint property is at most

$$\binom{n}{2} \left(\frac{3}{4} \right)^{|X|}.$$

So when $|X| = \lfloor 5 \log n \rfloor - 1$, the latter probability is $o_n(1)$.

We may now fix X with $|X| = \lfloor 5 \log n \rfloor - 1$ satisfying the midpoint property and set $r = |X| + 1 = \lfloor 5 \log n \rfloor$. It remains to prove that $\tau_{T_3}(G) \geq \frac{\log n}{3 + \log \log n}$. Let π_1, \dots, π_m be m permutations of S_n covering all T_3 -copies of G . To each $v \in [n] \setminus X$ we assign an m -vector $\mathbf{v} \in [r]^m$ as follows. Let $p_{i,1} < p_{i,2} < \dots < p_{i,r-1}$ be the $r - 1$ locations of the elements of X in π_i , for $i = 1 \dots, m$. Also define $p_{i,0} = 0$ and $p_{i,r} = n + 1$. Now, set the i th coordinate of \mathbf{v} as follows: $\mathbf{v}(i) = t$ if the location of v in π_i is before $p_{i,t}$ and after $p_{i,t-1}$. This defines \mathbf{v} uniquely in $[r]^m$.

We next observe that if $u, v \in [n] \setminus X$ are two distinct vertices, then it must be that $\mathbf{u} \neq \mathbf{v}$. Indeed, since X has the midpoint property, there is some $x \in X$ such that the triangle induced by $\{u, x, v\}$ is a T_3 and x is its midpoint vertex. But since each such triangle is covered, this transitive triple is covered by some permutation π_i . Hence, in π_i , the location of x is somewhere between the locations of u and v . In particular, $\mathbf{v}(i) \neq \mathbf{u}(i)$.

But now, the number of possible vectors is at most r^m . As each vertex of $[n] \setminus X$ is assigned to a distinct vector, we have that

$$(5 \log n)^m \geq r^m \geq n - |X| \geq \frac{n}{2}.$$

Thus, we must have $m \geq \log n / (3 + \log \log n)$. It follows that $\tau_{T_3}(n) \geq \log n / (3 + \log \log n)$. \square

Lemma 9. $\log n / (3 + \log \log n) \leq \tau_{P_3}(n)$.

Proof. The proof is almost identical to the proof of Lemma 8. The only difference is that for the midpoint property to hold with respect to P_3 instead of T_3 , the respective probability is $\frac{1}{2}$ instead of $\frac{3}{4}$, so we may take $|X|$ to be even slightly smaller than $\lfloor 5 \log n \rfloor - 1$. The rest of the proof carries through identically. \square

Theorem 2 now follows by Lemmas 7, 8, 9. \square

It is worth noting that in all the above lemmas, the lower bounds were obtained by considering random tournaments. Those have the property of being almost regular tournaments since the in-degree and out-degree of each vertex are almost the same. One might stipulate that for every regular or almost regular tournament, the number of permutations required to cover its 3-vertex acyclic subgraphs grows with n (observe that the number of each of T_3, P_3, S_2^+, S_2^- in a tournament is determined only by the out-degree sequence of the tournament, and in particular, for a regular tournament, it is $\Theta(n^3)$ for each). This, however, is false as there are arbitrarily large regular tournaments for which $\tau_{T_3}(G) = 2$. Indeed, suppose n is odd and consider the regular tournament R_n on vertex set $[n]$ where (i, j) is an edge for $j = i + 1, \dots, i + (n - 1)/2$ (indices taken modulo n). It is immediate to verify that the two permutations $1, 2, \dots, n$ and $(n + 1)/2, \dots, n, 1, \dots, (n - 1)/2$ cover all the T_3 of R_n , so $\tau_{T_3}(R_n) = 2$.

Another interesting family of tournaments that contain some almost regular tournaments but for which $\tau_{P_3}(G)$ is bounded by a constant are *r-majority* tournaments. These tournaments, which have applications in social choice theory, were studied by several researchers (see [1] and the references therein). For an odd positive integer r , consider a set of r permutations of $[n]$. The *r-majority* tournament on vertex set $[n]$ generated by the given set of permutations is defined as follows. Set (i, j) as an edge if i appears before j in more than half of the permutations. Now, it is easy to verify that if G is an *r-majority* tournament, then $\tau_{P_3}(G) \leq r$, and the permutations proving this upper bound are the permutations that generate G . Indeed suppose (i, j) and (j, k) are two edges of G . Then in at least $(r + 1)/2$ permutations i appears before j and in a least $(r + 1)/2$ permutations j appears before k , so in at least one permutation i appears before j and j appears before k . Hence, every P_3 is covered by some permutation.

2.2 Larger H

Here we prove Theorem 3. To this end, the following lemma is useful.

Lemma 10. *Let H be a given acyclic graph and let S be an induced subgraph of H . Then the probability that in a random n -vertex tournament each copy of S is contained in some copy of H is $1 - o_n(1)$. Hence, with probability $1 - o_n(1)$ it holds that $\tau_H(G) \geq \tau_S(G)$.*

Proof. Let r be a fixed positive integer. Consider a random tournament $G = ([n], E) \sim \vec{G}(n)$. For any two disjoint (possibly empty) sets R and Q of $|R| + |Q| \leq r$ vertices of G

there are, with probability $1 - o_n(1)$, $\Theta(n)$ vertices $k \in [n] \setminus (R \cup Q)$ such that $(i, k) \in E$ for each $i \in R$, and $(k, j) \in E$ for each $j \in Q$. Indeed, fixing R and Q , the number of such vertices is the binomial distribution $\mathcal{B}(n - |R| - |Q|, (1/2)^{|R|+|Q|})$, and there are only $O(n^r)$ choices of pairs (R, Q) to consider. By Chernoff's large deviation approximation of the Binomial distribution applied to the union of $O(n^r)$ events, we have with probability $1 - o_n(1)$, for any such sets R and Q , there are at least such $2^r + r$ vertices k so that $(i, k) \in E$ for each $i \in R$, and $(k, j) \in E$ for each $j \in Q$. Hence, from now on we will assume that G has this latter property, denoting it \mathcal{P} .

Let r denote the number of vertices of H . As H is acyclic, it has a total ordering of its vertices, say h_1, \dots, h_r , where if (h_i, h_j) is an edge of H , then $i < j$. Since S is a subgraph of H with s vertices, the vertex set of S is some subsequence h_{S_1}, \dots, h_{S_s} with the property that if (h_{S_i}, h_{S_j}) is an edge of S , then $S_i < S_j$. Suppose that X is a subgraph of G that is isomorphic to S , on vertex set $\{x_1, \dots, x_s\}$, such that the mapping $h_{S_i} \rightarrow x_i$ is an embedding of S . We now construct a copy of H in G that contains X .

Consider first the vertex x_1 and recall that $h_{S_1} \rightarrow x_1$. Pick an arbitrary set K of at least 2^{S_1-2} vertices that are in-neighbors to all the vertices $\{x_1, \dots, x_s\}$. We know such a K exists as we may use property \mathcal{P} with $R = \emptyset$ and $Q = \{x_1, \dots, x_s\}$. It is well-known [15] that in any tournament, every set of 2^{t-1} vertices contains a transitive tournament on t vertices. Hence, we can pick $S_1 - 1$ vertices of K that form a transitive tournament on $S_1 - 1$ vertices. For each j , we let the j th vertex of this transitive tournament play the role of vertex h_j of H . We now continue the embedding by doing the following procedure for $t = 2, \dots, s$. Consider vertex x_t . Recall that $h_{S_t} \rightarrow x_t$. Now, if $S_t = S_{t-1} + 1$, there is nothing to do, as there are no vertices to place between x_{t-1} and x_t . Otherwise, pick an arbitrary set K of at least $2^{S_t-S_{t-1}-2}$ vertices of G such that each of them is an out-neighbor of each of the previously embedded vertices (corresponding to $h_1, \dots, h_{S_{t-1}}$) and an in-neighbor of each vertex in $\{x_t, x_{t+1}, \dots, x_s\}$, and such that the picked vertices have not been used in prior stages. Since $S_t - S_{t-1} < r$ and since there were at most S_{t-1} previously embedded vertices used in the procedure, Property \mathcal{P} of G ensures that K exists. As before, we can find a transitive tournament of size $S_t - S_{t-1} - 1$ vertices within K . So, we may assign the j th vertex of this transitive tournament to play the role of vertex $h_{S_{t-1}+j}$ of H . Once we have completed our procedure, we still have to take care of embedding the vertices that appear after h_{S_s} (if there are any). Recall that $h_{S_s} \rightarrow x_s$. Pick an arbitrary set K of at least 2^{r-S_s-1} vertices that are out-neighbors of all previously embedded vertices. Hence, we can pick $r - S_s$ vertices of K that form a transitive tournament on $r - S_s$ vertices. We let the j th vertex of this transitive tournament play the role of vertex h_{S_s+j} of H . By construction, this embedding contains a copy of H in G which contains X . \square

Notice that the requirement that S is an induced subgraph of H is important since we do not have to worry about the directions of edges connecting vertices of X . For example, if $H = T_3$ and $S = P_3$, then P_3 is a subgraph of T_3 , but not an induced one. But it is certainly false that every P_3 in a random graph is contained in some T_3 , since it may be that the P_3 induces a cyclic triangle in G .

Let M denote the graph consisting of two independent directed edges, say $(a, b), (c, d)$.

Computing $\tau_M(n)$ is closely related to the problem of computing the separation dimension of the complete graph (see [3]). It is easy to adjust the lower bound proof for S_2^+ given in Lemma 7 to the case of M and obtain using a similar proof that $\tau_M(n) \geq \log \log n - \log \log \log n - 2$. We state this and sketch the proof in the following lemma.

Lemma 11. $\log \log n - \log \log \log n - 2 \leq \tau_M(n)$.

Proof (sketch). Let $G = ([n], E) \sim \vec{G}(n)$ and consider some $X \subset [n]$ of $2k$ vertices consisting of v_1, \dots, v_{2k} where $v_i < v_{i+1}$ for $i = 1, \dots, 2k - 1$. We compute the probability of the event B_X , that $G[X]$ has no two disjoint edges $(v_a, v_b), (v_d, v_c)$ with $a < b$ and $c < d$. In particular, for B_X to occur it must be that all $\binom{k-2}{2}$ edges induced by $\{v_3, \dots, v_k\}$ are either all from the lower indexed vertex to the higher, or all from the higher to the lower. Hence,

$$\Pr[B_X] \leq \left(\frac{1}{2}\right)^{\binom{k-2}{2}}.$$

In particular, choosing $k = \lfloor 8 \log n \rfloor$

$$\Pr\left[\bigcup_{X \subset [n], |X|=2k} B_X\right] \leq \binom{n}{2k} \left(\frac{1}{2}\right)^{k^2/3} < n^{2k} \left(\frac{1}{2}\right)^{k^2/3} < 1.$$

We can now fix a tournament $G = ([n], E)$ with the property that for each $X \subset [n]$ of size $\lfloor 16 \log n \rfloor$, the subtournament $G[X]$ has two disjoint edges, where one edge is from a smaller vertex to a larger and the other edge is from a larger vertex to a smaller. The remainder of the proof is essentially identical to the argument in Lemma 7. \square

By Lemma 10, and by the random tournament constructions in Lemmas 7,8,9,11, we have the following corollary, which is a restatement of Theorem 3.

Corollary 12. *Let H be an acyclic graph with at least two edges. Then, for all n sufficiently large, it holds that $\tau_H(n) \geq \log \log n - \log \log \log n - 2$. Furthermore, if H contains a directed path on at least two edges, then $\tau_H(n) \geq \log n / (3 + \log \log n)$.*

Proof. Since H is an acyclic graph with at least two edges, it contains as an induced subgraph some graph $K \in \{S_2^+, S_2^-, M, P_3, T_3\}$. By Lemma 10, if n is sufficiently large, then a random graph G almost surely satisfies $\tau_H(G) \geq \tau_K(G)$. But by the proofs of Lemma 7 for $K \in \{S_2^+, S_2^-\}$, Lemma 8 for $K = T_3$, Lemma 9 for $K = P_3$, and Lemma 11 for $K = M$, we have that almost surely $\tau_K(G) \geq \log \log n - \log \log \log n - 2$ in the case where $K \in \{S_2^+, S_2^-, M\}$ and $\tau_K(G) \geq \log n / (3 + \log \log n)$ in the case where $K \in \{P_3, T_3\}$. Notice that the latter case holds whenever H contains a directed path on at least two edges. \square

2.3 Paths

Proof of Theorem 4. Since $\tau_{P_k}(n) \leq s(k, n)$, the upper bound in (2) implies $\tau_{P_k}(n) = O(\log n)$. The lower bound construction proceeds as follows. For simplicity, we shall

assume that n is an odd multiple of 3, as this does not affect the asymptotic claim since $\tau_{P_k}(n)$ is monotone in n . Recall again the regular tournament R_n on vertex set $[n]$ where (i, j) is an edge for $j = i + 1, \dots, i + (n - 1)/2$ (indices taken modulo n). Let $u = n/3$, $v = 2n/3$ and $w = n$. We claim that for any two vertices $x, y \in [n] \setminus \{u, v, w\}$ there is a P_k in R_n containing both x and y , and the sub-path between x and y passes through at least one of u, v, w . Indeed, we can assume without loss of generality that $1 \leq x < n/3$ by the cyclic structure of R_n and by the equidistant choice of u, v, w . Suppose first that $n/3 < y < 2n/3$. Then take the path (x, u, y) which is a P_3 . We can make it into a P_k by arbitrarily continuing the path from y with an additional path of $k - 3$ vertices (this is trivially possible if, say, $n > 3k$). Suppose next that $2n/3 < y < n$. Then take the path (y, w, x) which is a P_3 and continue as in the previous case to construct a P_k . So, we remain with the case where $1 < y < n/3$. Assume without loss of generality that $x < y$. If $y > n/6$ then take the path (y, v, w, x) which is a P_4 and continue if necessary to create a P_k . If $y < n/6$ then take the path (y, u, v, x) which is a P_4 and continue if necessary to create a P_k .

We can now use a similar idea as in the proof of Lemma 8. Let π_1, \dots, π_m be m permutations of S_n covering all the P_k subgraphs of R_n . To each $x \in [n] \setminus \{u, v, w\}$ we assign an m -vector $\mathbf{x} \in [4]^m$ as follows. Let $p_{i,1} < p_{i,2} < p_{i,3}$ be the three locations of u, v, w in π_i , for $i = 1 \dots, m$. Also define $p_{i,0} = 0$ and $p_{i,4} = n + 1$. Now, set the i th coordinate of \mathbf{x} as follows: $\mathbf{x}(i) = t$ if the location of x in π_i is before $p_{i,t}$ and after $p_{i,t-1}$. This defines \mathbf{x} uniquely in $[4]^m$.

We next observe that if $x, y \in [n] \setminus \{u, v, w\}$ are two distinct vertices, then it must be that $\mathbf{x} \neq \mathbf{y}$. Indeed, there is a P_k in R_n which contains x, y and the sub-path between x and y passes through at least one of u, v, w . But since each P_k is covered, then this particular P_k is covered by some permutation π_i . Hence, in π_i , the location of at least one of $\{u, v, w\}$ is somewhere between the locations of x and y . In particular, $\mathbf{x}(i) \neq \mathbf{y}(i)$.

But now, the number of possible vectors is at most 4^m . As each vertex of $[n] \setminus \{u, v, w\}$ is assigned to a distinct vector, we have that $4^m \geq n - 3$. Thus, we must have $m \geq \log(n - 3)/2$. It follows that $\tau_{P_k}(n) \geq \log(n - 3)/2$. \square

3 Covering the edges of 3-tournaments and random oriented 3-graphs

In this section we prove Theorem 6. Our main ingredient will be the method of hypergraph entropy used by Füredi [7] to prove his lower bound for $s(3, n)$. However, while the proof for $s(3, n)$ uses the fact that for each vertex x and for any two additional vertices u, v , there is an edge of $D(n, 3)$ consisting of u, v, x where x is the middle vertex, this is of course not so if we consider elements of $\vec{G}_3(n, 1)$ (3-tournaments) moreover elements of $\vec{G}_3(n, n^{-\alpha})$.

To overcome this obstacle, we first need to prove that a certain property holds for $G \sim \vec{G}_3(n, n^{-\alpha})$ with high probability. Recall that G is an oriented 3-graph so every 3-set of $[n]$ either does not induce any edge or induces a single directed edge. We say that a

vertex $x \in [n]$ has the k -midpoint property if for any $K \subset [n] \setminus \{x\}$ with $|K| = k$, there exists an edge of G of the form (u, x, v) where $u, v \in K$.

We now claim that $G \sim \vec{G}_3(n, n^{-\alpha})$ has the property that every vertex of G has the k -midpoint property for $k = \lfloor 6n^\alpha \log n \rfloor$, with high probability. Indeed, let $x \in [n]$ be an arbitrary vertex and let $K \subset [n] \setminus \{x\}$ be an arbitrary set of size k . What is the probability of having no edge of the form (u, x, v) where $u, v \in K$? For two given vertices $u, v \in K$, the probability of having one of the edges (u, x, v) or (v, x, u) in G is precisely $n^{-\alpha}/3$, since we must first make sure that the 3-set u, x, v induces an edge, and conditioning on that, we must ensure that the direction of that edge puts x in the middle. So, the probability of having no edge of the form (u, x, v) where $u, v \in K$ is $(1 - n^{-\alpha}/3)^{\binom{k}{2}}$. Hence, the probability that every vertex of G has the k -midpoint property is at least

$$1 - n \cdot \binom{n-1}{k} \left(1 - \frac{n^{-\alpha}}{3}\right)^{\binom{k}{2}} > 1 - \frac{1}{n}.$$

Indeed, the last inequality holds for $k = \lfloor 6n^\alpha \log n \rfloor$ as for this value we have that

$$\begin{aligned} n^2 \cdot \binom{n-1}{k} \left(1 - \frac{n^{-\alpha}}{3}\right)^{\binom{k}{2}} &< n^{k+2} \left(1 - \frac{n^{-\alpha}}{3}\right)^{\binom{k}{2}} \\ &< n^{k+2} e^{-\frac{k(k-1)}{6n^\alpha}} \\ &< \left(2^{1.1 \log n - \frac{k-1}{5n^\alpha}}\right)^k \\ &< (2^0)^k = 1. \end{aligned}$$

So, for the remainder of the proof we may consider a given oriented 3-graph G with the property that every vertex of G has the k -midpoint property for $k = \lfloor 6n^\alpha \log n \rfloor$. We prove that for such a graph,

$$\tau(G) \geq (1 - \alpha) \frac{2}{\log e} \log n - 9 \log \log n.$$

Consider the entropy function $H : [0, 1] \rightarrow [0, 1]$ defined as $H(y) = -y \log(y) - (1 - y) \log(1 - y)$ for $y \in (0, 1)$ and $H(0) = H(1) = 0$. A multihypergraph \mathcal{F} is a collection of (possibly empty) subsets of a vertex set V where with each subset (edge) there is associated a positive integer which is its multiplicity (i.e. the number of times the edge appears). Let $\mu(\mathcal{F})$ denote the maximum multiplicity of an edge of \mathcal{F} . The following lemma was proved for $\mu(\mathcal{F}) = 1$ by Kleitman, Shearer, and Sturtevant [11], and as observed by Füredi in [7], the same proof holds for larger μ .

Lemma 13. [11, 7] *Let \mathcal{F} be a multihypergraph with m edges on vertex set V . For each $x \in V$, let $p(x)$ denote the fraction of the edges that contain x . Then, $\sum_{x \in V} H(p(x)) \geq \log(m/\mu(\mathcal{F}))$.*

The argument in Füredi's lower bound proof for $s(3, n)$ proceeds as follows. Suppose π_1, \dots, π_r are permutations of $[n]$. For each $x \in [n]$, define a multihypergraph \mathcal{F}_x on vertex

set $[r]$ as follows. For each $v \in [n] \setminus \{x\}$ there is an edge of \mathcal{F}_x consisting of all indices i such that $\pi_i(v) < \pi_i(x)$. Observe that this hypergraph has $m = n - 1$ edges. For each $i \in [r]$, the number of edges containing i is precisely $\pi_i(x) - 1$. Hence, $p(i) = (\pi_i(x) - 1)/(n - 1)$. By Lemma 13 (see [7]) we have that

$$\log \frac{n - 1}{\mu(\mathcal{F}_x)} \leq \sum_{i=1}^r H\left(\frac{\pi_i(x) - 1}{n - 1}\right).$$

Now, let $\ell = \max_{x \in [n]} \mu(\mathcal{F}_x)$. So the last inequality remains valid if we replace $\mu(\mathcal{F}_x)$ with ℓ . Summing up this revised inequality over all vertices $x \in [n]$ and using some calculus, Füredi showed that

$$n \log \frac{n - 1}{\ell} < r(n - 1) \frac{\log e}{2}$$

which in turn implies that $\ell > (n - 1) \exp(-r(n - 1)/(2n))$.

But what is ℓ in our case? Suppose that the π_1, \dots, π_r cover all edges of G . We claim that for each vertex x we have $\mu(\mathcal{F}_x) \leq \lfloor 6n^\alpha \log n \rfloor$. Indeed, recall that each edge of \mathcal{F}_x corresponds to some $v \in [n] \setminus \{x\}$ and consists of all permutations in which v precedes x . So having $\mu(\mathcal{F}_x)$ multiple edges means that there are $\mu(\mathcal{F}_x)$ vertices of G that appear in each permutation either all before x or all after x . But since we assume that x has the $\lfloor 6n^\alpha \log n \rfloor$ -midpoint property, we must have that $\mu(\mathcal{F}_x) \leq \lfloor 6n^\alpha \log n \rfloor$. In particular, $\ell \leq 6n^\alpha \log n$. Thus,

$$6n^\alpha \log n > (n - 1)e^{-\frac{r(n-1)}{2n}} \geq ne^{-\frac{r}{2}}.$$

It follows that $r > (1 - \alpha) \frac{2}{\log e} \log n - 9 \log \log n$. □

4 Concluding remarks and open problems

There are a few natural open problems that may be more accessible than Problem 1 in its full general form.

Problem 14. Determine a function $f(n)$ such that $\tau_{P_3}(n) = \Theta(f(n))$.

Theorem 2 shows that $f(n) = \Omega(\log n / \log \log n)$ and (1) shows that $f(n) = O(\log n)$.

Problem 15. Determine a function $f_H(n)$ such that $\tau_H(n) = \Theta(f_H(n))$ for acyclic graphs H with at least two edges that do not contain P_3 as a subgraph. Can $f_H(n)$ be the same function for all such H (with the possible exception of the directed in-star and directed out-star)?

Currently, we know the answer to Problem 15 only for the directed out-star and directed in-star, in which case it is doubly logarithmic. Also recall that by Theorem 3 we must have $f_H(n) = \Omega(\log \log n)$ for any H other than the single edge.

We end this paper with a simple improvement of the upper bound for $s(k, n)$. It follows from an application of the symmetric case of the Lovász Local Lemma and can be presented in a more general form as follows. Suppose that G is a spanning subhypergraph

of $D(n, k)$. Namely, it only contains some ordered k -sets of $[n]$, not necessarily all of them. The *degree* of a vertex of G is the number of edges of G containing that vertex. As a special case, notice that the degree of every vertex of $D(n, k)$ is $k! \binom{n-1}{k-1} = \Theta(n^{k-1})$.

Proposition 16. *Let G be a directed k -graph with n vertices and with maximum degree at most d . Then $\tau(G) \leq \ln 2 \cdot k! \log(edk)$. In particular, $s(k, n) \leq C_k + \ln 2 \cdot k!(k-1) \log n$ where C_k is a constant that only depends on k .*

Proof. Let G be a directed k -graph on vertex set $[n]$ and with maximum degree d . Consider a set \mathcal{S} of m independently chosen random permutations of S_n (choices made uniformly with repetition). For an edge e of G , let A_e be the event that e is not covered by \mathcal{S} . Then, $\Pr[A_e] = (1 - 1/k!)^m$. Now, as each $v \in [n]$ has degree at most d , the event A_e is independent of all other events but at most kd . By the Lovász Local Lemma [5], if it holds that

$$\left(1 - \frac{1}{k!}\right)^m \cdot e \cdot kd < 1 \tag{4}$$

then with positive probability, \mathcal{S} is such that none of the events hold. In other words, there exists a set of m elements of S_n that covers all edges of G , showing that $\tau(G) \leq m$. Now, when $m \geq \ln 2 \cdot k! \log(edk)$, then (4) holds. Therefore, $\tau(G) \leq \ln 2 \cdot k! \log(edk)$. Observing that for $D(n, k)$ we have $d = k! \binom{n-1}{k-1} < kn^{k-1}$ we have that $s(k, n) \leq \ln 2 \cdot k!(k-1) \log n + C_k$ where $C_k \leq \ln 2 \cdot k! \log(ek^2)$. \square

For every fixed $k \geq 3$, the upper bound for $s(k, n)$ given in Proposition 16 is better than the upper bound in (2) by a constant factor depending on k (but notice that for $k = 3$ we still have a better upper bound in (1)). For example, if $k = 4$, the bound for $s(k, n)$ attained by Proposition 16, divided by $\log n$, approaches $72 \ln 2 \approx 49.9$. Comparatively, the constant attained by the union bound probabilistic argument in [14] giving the upper bound in (2) is $4 / \log(24/23) \approx 65.14$.

Acknowledgment

The author thanks Majdi Saeed for useful discussions.

References

- [1] N. Alon, G. Brightwell, H. Kierstead, A. Kostochka, and P. Winkler. Dominating sets in k -majority tournaments. *Journal of Combinatorial Theory, Series B*, 96(3):374–387, 2006.
- [2] Jørgen Bang-Jensen and Gregory Z. Gutin. *Digraphs: Theory, Algorithms and Applications*. Springer Science & Business Media, 2008.
- [3] M. Basavaraju, S. Chandran, M.C. Golumbic, R. Mathew, and D. Rajendraprasad. Boxicity and separation dimension. In *International Workshop on Graph-Theoretic Concepts in Computer Science (WG)*, pages 81–92, 2014.

- [4] Y. Chee, C. Colbourn, D. Horsley, and J. Zhou. Sequence covering arrays. *SIAM Journal on Discrete Mathematics*, 27(4):1844–1861, 2013.
- [5] P. Erdős and L. Lovász. Problems and results on 3-chromatic hypergraphs and some related questions. In A. Hajnal, R. Rado, and V.T. Sós, editors, *Infinite and Finite Sets*, volume 10, pages 609–627. North-Holland, Amsterdam, 1975.
- [6] P. Erdős and G. Szekeres. A combinatorial problem in geometry. *Compositio Mathematica*, 2:463–470, 1935.
- [7] Z. Füredi. Scrambling permutations and entropy of hypergraphs. *Random Structures & Algorithms*, 8(2):97–104, 1996.
- [8] Z. Füredi, P. Hajnal, V. Rödl, and W.T. Trotter. Interval orders and shift graphs. In A. Hajnal and V. T. Sós, editors, *Sets, Graphs, and Numbers*, volume 60, pages 297–313. North-Holland Publishing Co.; János Bolyai Mathematical Society, 1992.
- [9] Y. Ishigami. Containment problems in high-dimensional spaces. *Graphs and Combinatorics*, 11(4):327–335, 1995.
- [10] Y. Ishigami. An extremal problem of d permutations containing every permutation of every t elements. *Discrete Mathematics*, 159(1-3):279–283, 1996.
- [11] D. J. Kleitman, J. Shearer, and D. Sturtevant. Intersections of k -element sets. *Combinatorica*, 1(4):381–384, 1981.
- [12] J. Kruskal. Monotonic subsequences. *Proceedings of the American Mathematical Society*, 4(2):264–274, 1953.
- [13] J. Radhakrishnan. A note on scrambling permutations. *Random Structures & Algorithms*, 22(4):435–439, 2003.
- [14] J. Spencer. Minimal scrambling sets of simple orders. *Acta Mathematica Hungarica*, 22(3-4):349–353, 1972.
- [15] R. Stearns. The voting problem. *The American Mathematical Monthly*, 66(9):761–763, 1959.
- [16] J. Tarui. On the minimum number of completely 3-scrambling permutations. *Discrete Mathematics*, 308(8):1350–1354, 2008.