HAMILTON CYCLES IN A RANDOM TOURNAMENT

Svante Janson

Abstract. The number of Hamilton cycles in a random tournament is asymptotically normally distributed.

1. Introduction

A tournament on n vertices is a directed graph obtained by orienting each edge of the complete undirected graph K_n in one of the two possible directions. We will here consider the random tournament T_n obtained by choosing uniformly among all tournaments, or equivalently, by orienting the edges independently with probability $1/2$ for each direction. For a background see [7], where also further references are given. (Other models of random tournaments, where the directions are assigned with unequal probabilities, have been considered by several authors, see e.g. [2], [9], [3], [6], but we will here only consider the simplest and most symmetric case.)

We will study the number of Hamilton cycles in T_n and our main result is the following. (We assume throughout $n \geq 3$.)

Theorem 1. *Let H_n be the number of directed Hamilton cycles in the random tournament T_n. Then*

$$\operatorname{E} H_n = (n-1)!\, 2^{-n}, \tag{1.1}$$

$$\operatorname{Var} H_n = \Big(\frac{2}{n} + O\big(\frac{1}{n^2}\big)\Big)(\operatorname{E} H_n)^2 \tag{1.2}$$

and H_n is asymptotically normally distributed:

$$\frac{H_n - \operatorname{E} H_n}{(\operatorname{Var} H_n)^{1/2}} \xrightarrow{d} \mathrm{N}(0,1) \qquad \text{as} \quad n \to \infty. \tag{1.3}$$

Using (1.1) and (1.2), we can rewrite (1.3) as

$$\sqrt{\frac{n}{2}}\Big(H_n/\big((n-1)!\,2^{-n}\big) - 1\Big) \xrightarrow{d} \mathrm{N}(0,1). \tag{1.4}$$

Remark. This result differs markedly from the corresponding results in [4] for the random undirected graph $G_{n,p}$ and the random directed graph $D_{n,p}$, with $p = 1/2$ (or any other fixed value); in those cases the number of Hamilton cycles is asymptotically *log-normal.* In contrast, the result for $G_{n,m}$, with for example $m = \frac{1}{2}\binom{n}{2}$, is very similar; furthermore, the projection method used below is very similar to the one used for $G_{n,m}$ in [4].

Acknowledgment. This research was mainly done during the conference Random Graphs '93 in Poznań, where this problem was put to me by Noga Alon. I thank him and Andrzej Ruciński for stimulating discussions.

Supported by the Göran Gustafsson Foundation for Research in Natural Sciences and Medicine

2. A projection method.

We will prove Theorem 1 by a method similar to the well-known 'method of first projection' for $G_{n,p}$, where one approximates the random variable of interest with a linear function on the number of edges in the random graph. For random tournaments (as for $G_{n,m}$), however, the number of edges is constant and we instead resort to the 'second projection', using the numbers of different types of subgraphs on three vertices. We begin by studying the numbers of such subgraphs; for completeness we do it in somewhat more detail than necessary for just the proof of Theorem 1.

There are $\frac{1}{2}(n)_3$ paths of length 2 in K_n. Each of them becomes in a tournament either a vertex with a pair of outgoing edges, a directed path, or a vertex with a pair of incoming edges. Let N_0, N_1 and N_2 be the numbers of times these possibilities happen in a given tournament on n vertices; thus $N_0+N_1+N_2 = \frac{1}{2}(n)_3$. Letting d_{ik} and d_{0k} be the in-degree and out-degree of vertex k, we have $N_0 = \sum_k \binom{d_{0k}}{2}$, $N_1 = \sum_k d_{0k}d_{ik}$ and $N_2 = \sum_k \binom{d_{ik}}{2}$. Hence

$$2N_0 + N_1 = \sum_k d_{0k}(d_{0k} - 1 + d_{ik}) = \sum_k d_{0k}(n-2) = \binom{n}{2}(n-2) = \frac{1}{2}(n)_3$$

since $d_{ik} + d_{0k} = n-1$ for each k and $\sum d_{0k}$ is just the total number of edges. Similarly, also $2N_2 + N_1 = \frac{1}{2}(n)_3$, and thus $N_0 = N_2$.

Another way to see this is to let N_3 be the number of directed cycles of length 3 in the tournament. Since any subset of three vertices in the tournament induces a subgraph of order three which is either a directed cycle or a triangle where the vertices have outdegrees 0,1 and 2, it follows that

$$N_0 = N_2 = \binom{n}{3} - N_3 \qquad \text{and} \qquad N_1 = 3N_3 + \binom{n}{3} - N_3 = \binom{n}{3} + 2N_3.$$

Hence any of the variables N_0, N_1, N_2, N_3 is a linear function of any of the others, and it suffices to consider one of them. We will actually find it convenient to use a fifth variable, viz. the linear combination $S = N_1 - N_0 - N_2 = 4N_3 - \binom{n}{3}$. (The formulae above yield $N_0 = N_2 = \frac{1}{8}(n)_3 - \frac{1}{4}S$, $N_1 = \frac{1}{4}(n)_3 + \frac{1}{2}S$ and $N_3 = \frac{1}{24}(n)_3 + \frac{1}{4}S$.)

Considering a random tournament again, these functionals become random variables. We may write $S = \sum J(P)$, summing over all paths P of length 2 in K_n, where

$$J(P) = \begin{cases} +1 & \text{if } P \text{ becomes a directed path in the tournament} \\ -1 & \text{otherwise.} \end{cases}$$

The random variables $J(P)$ have mean 0 and are pairwise independent; hence $\mathrm{E}\,S = 0$ and $\mathrm{E}\,S^2 = \sum_P \mathrm{E}\,J(P)^2 = \frac{1}{2}(n)_3$. Furthermore, as $n \to \infty$, S is asymptotically normal. This was proved already by Moran [8] (for the cycle count N_3, which is equivalent by the formulae above); it also follows directly from e.g. Janson and Nowicki [5, Theorem 4] or, after a simple fourth moment calculation, De Jong [1, Theorem 1], see also [3] and [10].

We collect these results as a lemma.

Lemma 1. $\mathrm{E}\,S = 0$, $\mathrm{Var}\,S = \frac{1}{2}(n)_3$ *and*

$$\frac{S - \mathrm{E}\,S}{(\mathrm{Var}\,S)^{1/2}} \xrightarrow{d} \mathrm{N}(0,1) \qquad \textit{as} \quad n \to \infty. \qquad \square$$

We state the projection method result in the following form.

Theorem 2. *Suppose that X is any functional of the random tournament T_n such that*

$$\frac{(\mathrm{E}\, XS)^2}{\operatorname{Var} X \operatorname{Var} S} \to 1 \qquad \text{as} \quad n \to \infty$$

(or, equivalently, $(\mathrm{E}\, XS)^2 \sim \frac{1}{2} n^3 \operatorname{Var} X$*). Then*

$$\frac{X - \mathrm{E}\, X}{(\operatorname{Var} X)^{1/2}} \xrightarrow{d} \mathrm{N}(0,1) \qquad \text{as} \quad n \to \infty.$$

Proof. Let $\sigma_X^2 = \operatorname{Var} X$, $\sigma_S^2 = \operatorname{Var} S$ and $\tau = \operatorname{sign}(\mathrm{E}\, XS)$. Then

$$\mathrm{E}\Big(\frac{X - \mathrm{E}\, X}{\sigma_X} - \tau \frac{S}{\sigma_S}\Big)^2 = \frac{\operatorname{Var} X}{\sigma_X^2} - 2\tau \frac{\mathrm{E}\, XS}{\sigma_X \sigma_S} + \frac{\operatorname{Var} S}{\sigma_S^2} = 2 - 2\frac{|\,\mathrm{E}\, XS|}{\sigma_X \sigma_S} \to 0$$

and thus $(X - \mathrm{E}\, X)/\sigma_X - \tau S/\sigma_S \xrightarrow{p} 0$ as $n \to \infty$. Since $S/\sigma_S \xrightarrow{d} \mathrm{N}(0,1)$ by Lemma 1, τ is non-random with $\tau = \pm 1$ and the normal distribution is symmetric, we have $\tau S/\sigma_S \xrightarrow{d} \mathrm{N}(0,1)$ and the result follows by Cramr's theorem. $\square$

Remark. The careful reader may observe that although we called this a projection method, we did not use the actual orthogonal projection $\frac{\mathrm{E}\, XS}{\sigma_S^2} S$ of X; this is only a matter of convenience in the proof.

Remark. Most other small subgraph counts for random tournaments are easily shown to be asymptotically normal using this projection result, but there are some exceptions. This is entirely similar to the situation for $G_{n,p}$ and $G_{n,m}$, but the exceptional graphs differ in the three cases. We hope to return to this in another paper.

3. Proof of Theorem 1.

We will use Theorem 2, and must therefore compute (at least approximately) $\mathrm{E}\, H_n S$ and $\operatorname{Var} H_n$. It will be convenient to regard directed graphs as sets of directed edges, in our case subsets of the edge set of the complete directed graph D_n (we will not consider any graphs with isolated vertices, so this is unambiguous). Hence, for example, $|A \cap B|$ denotes the number of common edges in A and B. If A is any directed graph, we let $\overline{A}$ denote the graph obtained by reversing the orientation of each edge, and let A^0 be the corresponding undirected graph.

Let $\mathcal{C}$ be the set of the $(n-1)!$ directed Hamilton cycles in D_n. Then

$$\mathrm{E}\, H_n = \mathrm{E} \sum_{C \in \mathcal{C}} I(C \subset T_n) = \sum_{C \in \mathcal{C}} \mathrm{P}(C \subset T_n) = (n-1)!\, 2^{-n}, \tag{3.1}$$

which verifies (1.1).

Lemma 2. $\mathrm{E}\, H_n S = n\, \mathrm{E}\, H_n$.

Proof. Let $\mathcal{P}$ be the set of all paths of length 2 in K_n. Then

$$\begin{aligned} \mathrm{E}(H_n S) &= \mathrm{E} \sum_{C \in \mathcal{C}} \sum_{P \in \mathcal{P}} I(C \subset T_n) J(P) \\ &= \mathrm{E} \sum_{C \in \mathcal{C}} \mathrm{P}(C \subset T_n) \sum_{P \in \mathcal{P}} \mathrm{E}\big(J(P) \mid C \subset T_n\big). \end{aligned} \tag{3.2}$$

For any fixed $C \in \mathcal{C}$, $\mathrm{E}\big(J(P) \mid C \subset T_n\big) = 0$ for every P that contains at least one edge not in C^0; there remain n choices of P with $P \subset C^0$, and for them $C \subset T_n$ implies $J(P) = 1$. Hence the final inner sum above equals n, and

$$\mathrm{E}(H_n S) = n \sum_{C \in \mathcal{C}} \mathrm{P}(C \subset T_n) = n\, \mathrm{E}\, H_n. \qquad \square$$

We next give an exact expression for $\operatorname{Var}(H_n)$.

Lemma 3.

$$\begin{aligned}\frac{\operatorname{Var} H_n}{(\operatorname{E} H_n)^2} &= \frac{\operatorname{E} H_n^2}{(\operatorname{E} H_n)^2} - 1 \\ &= \frac{1+(-1)^n}{(n-1)!} + \sum_{k=1}^{\lfloor (n-1)/2 \rfloor} \sum_{j=1}^{k \wedge (n-2k)} \frac{n(n-2k-1)!^2 (k-1)!\, 2^j}{(n-1)!\,(n-2k-j)!\,(k-j)!\,j!\,(j-1)!}.\end{aligned}$$

Proof.

$$\operatorname{E} H_n^2 = \sum_{C,D\in\mathcal{C}} \operatorname{P}(C,D \subset T_n) = \sum_{C\in\mathcal{C}} \operatorname{P}(C \subset T_n) \sum_{D\in\mathcal{C}} \operatorname{P}(D \subset T_n \mid C \subset T_n). \tag{3.3}$$

By symmetry, the inner sum is the same for every $C \in \mathcal{C}$. Moreover, $\operatorname{P}(D \subset T_n \mid C \subset T_n) = 0$ if $C \cap \overline{D} \neq \emptyset$, and otherwise $\operatorname{P}(D \subset T_n \mid C \subset T_n) = 2^{-|D\setminus C|} = 2^{|C\cap D|} \operatorname{P}(D \subset T_n)$. Hence, for any fixed $C \in \mathcal{C}$, using a type of inclusion–exclusion argument,

$$\begin{aligned}(n-1)!\,\frac{\operatorname{E} H_n^2}{(\operatorname{E} H_n)^2} &= 2^n \sum_{D\in\mathcal{C}} \operatorname{P}(D \subset T_n \mid C \subset T_n) \\ &= \sum_{D\in\mathcal{C}} \frac{\operatorname{P}(D \subset T_n \mid C \subset T_n)}{\operatorname{P}(D \subset T_n)} = \sum_{D\in\mathcal{C}} 2^{|C\cap D|}\, 0^{|C\cap \overline{D}|} \\ &= \sum_{D\in\mathcal{C}} (1+1)^{|C\cap D|} (1-1)^{|C\cap\overline{D}|} = \sum_{D\in\mathcal{C}} \sum_{A\subset C\cap D} \sum_{B\subset C\cap\overline{D}} (-1)^{|B|} \\ &= \sum_{A,B\subset C} \sum_{D\in\mathcal{C}: D\supset A, \overline{D}\supset B} (-1)^{|B|} \\ &= \sum_{A,B\subset C} (-1)^{|B|} |\{D \in \mathcal{C} : D \supset A \cup \overline{B}\}|. \end{aligned}\tag{3.4}$$

If A and B are subgraphs of C with a common edge or just a common vertex, then there are no directed Hamilton cycles $D \supset A \cup \overline{B}$. Hence, if we collect all terms in the last sum with a given $A \cup B$, then we only have to consider A (and B) which are unions of some weak components of $A \cup B$. If $A \cup B = C$, this gives the two cases $A = C$, $B = \emptyset$ and $A = \emptyset$, $B = C$, with a single Hamilton cycle $D \supset A \cup \overline{B}$ in each case; hence the contribution to (3.4) is $1 + (-1)^n$. If $A \cup B = \emptyset$ we have $A = B = \emptyset$ and a contribution to (3.4) of $|\mathcal{C}| = (n-1)!$. In the remaining cases, $A \cup B$ is a union of $j \geq 1$ vertex-disjoint paths. If these have the lengths $l_1, \dots, l_j$, and $l = \sum_i l_i = |A \cup B|$, there are $n - l - j$ remaining vertices. Thus there are $j + n - l - j = n - l$ paths and isolated vertices, and there are $(n-l-1)!$ Hamilton cycles $D \supset A \cup \overline{B}$, which are obtained by taking these paths and vertices in any order (starting with a fixed one), travelling along each of the paths in the right direction. The contribution to (3.4) for such an $A \cup B$ is therefore

$$\begin{aligned}\sum_B (-1)^{|B|} (n-l-1)! &= (n-l-1)! \prod_{i=1}^{j} \bigl(1 + (-1)^{l_i}\bigr) \\ &= \begin{cases} 2^j (n-l-1)! & \text{if all } l_i \text{ are even,} \\ 0 & \text{if some } l_i \text{ is odd.}\end{cases}\end{aligned}\tag{3.5}$$

It remains to calculate the number of such $A \cup B \subset C$ with all l_i even for given j and l. Since l has to be even, we write $l = 2k$, and let $M(j,k)$ be the sought number. We have to consider subsets of C that consist of $j \geq 1$ paths of $l_i = 2k_i \geq 2$ edges, $i = 1, \dots, j$, separated by gaps of lengths, say, $m_1, \dots, m_j \geq 1$, with $\sum_i k_i = k$. Fix a vertex $v \in C$. By symmetry, exactly $\frac{j}{n} M(j,k)$ of these subset have one of their j paths starting at v, we call these *special* subsets. Each special subset is uniquely determined by the sequences $k_1, \dots, k_j$ and $m_1, \dots, m_j$, read in order along C starting at v. These sequences of integers ≥ 1 are subject to the conditions $\sum_i k_i = k$ and $\sum_i m_i = n - l = n - 2k$ only; hence they may be chosen in $\binom{k-1}{j-1}$ and $\binom{n-2k-1}{j-1}$ ways respectively, which gives a total of $\binom{k-1}{j-1}\binom{n-2k-1}{j-1}$ special subsets. Since this equals $\frac{j}{n} M(j,k)$, we have

$$M(j,k) = \frac{n}{j}\binom{k-1}{j-1}\binom{n-2k-1}{j-1} = \frac{n(n-2k-1)!\,(k-1)!}{(n-2k-j)!\,(k-j)!\,j!\,(j-1)!}. \tag{3.6}$$

(In particular, $M(j,k) = 0$ unless $1 \leq j \leq k$ and $j \leq n-2k$.) This and (3.5) enable us to complete the evaluation of (3.4) as

$$\begin{aligned}(n-1)!\,&\frac{\operatorname{E} H_n^2}{(\operatorname{E} H_n)^2}\\ &= 1 + (-1)^n + (n-1)! + \sum_{k=1}^{\lfloor (n-1)/2 \rfloor} \sum_{j=1}^{k \wedge (n-2k)} \frac{n(n-2k-1)!^2 (k-1)!}{(n-2k-j)!\,(k-j)!\,j!\,(j-1)!} 2^j.\end{aligned}$$

The result follows by dividing through by $(n-1)!$ and subtracting 1. ☐

It remains to estimate the terms in the sum in Lemma 3. This can be done rather crudely as follows. First, for $k=1$ we have the single term with $j = k = 1$, which equals

$$\frac{n(n-3)!}{(n-1)!} \cdot 2 = \frac{2n}{(n-1)(n-2)} = \frac{2}{n} + O\big(\frac{1}{n^2}\big);$$

this is the term we want. Secondly, for all terms we have

$$\frac{(n-2k-1)!}{(n-2k-j)!} \leq \frac{(n-2k+j-2)!}{(n-2k-1)!}$$

and thus, using $j \leq k$ and $k \leq (n-1)/2$,

$$\begin{aligned}\frac{n(n-2k-1)!^2}{(n-1)!\,(n-2k-j)!} &\leq n\frac{(n-2k+j-2)!}{(n-1)!} \leq n\frac{(n-k-2)!}{(n-1)!}\\ &\leq n\frac{1}{(n-k-1)^{k+1}} \leq n\Big(\frac{2}{n-1}\Big)^{k+1} \leq n\Big(\frac{3}{n}\Big)^{k+1}.\end{aligned}$$

Hence the sum of the terms with $k \geq 2$ is at most

$$\sum_{k=2}^{n}\sum_{j=1}^{k} n\Big(\frac{3}{n}\Big)^{k+1} \frac{(k-1)!\,2^j}{(k-j)!\,j!\,(j-1)!} \leq \sum_{k=2}^{n}\sum_{j=0}^{k} 3\Big(\frac{3}{n}\Big)^{k}\binom{k}{j}2^j = \sum_{k=2}^{n} 3\Big(\frac{3}{n}\Big)^{k} 3^k = O(n^{-2}).$$

Consequently, Lemma 3 yields

$$\frac{\operatorname{Var} H_n}{(\operatorname{E} H_n)^2} = \frac{2}{n} + O\big(\frac{1}{n^2}\big),$$

which verifies (1.2). We also obtain from this, using Lemma 2,

$$(\operatorname{E} H_n S)^2 = n^2 (\operatorname{E} H_n)^2 \sim \frac{n^3}{2} \operatorname{Var} H_n,$$

and the asymptotic normality follows from Theorem 2. ☐

References

1. P. De Jong, A central limit theorem for generalized multilinear forms, *J. Multivariate Anal.* **34** (1990), 275–289.
2. O. Frank, Stochastic competition graphs, *Rev. Internat. Stat. Inst.* **36** (1971), 319–326.
3. J. Gruszka, *Cycles in tournaments and random tournaments*, Ph.D. dissertation, Adam Mickiewicz University, Poznań, 1990. (Polish)
4. S. Janson, The numbers of spanning trees, Hamilton cycles and perfect matchings in a random graph, *Combinatorics, Probability & Computing* (to appear).
5. S. Janson and K. Nowicki, The asymptotic distributions of generalized U-statistics with applications to random graphs, *Probab. Th. Rel. Fields* **90** (1991), 341–375.
6. T. Łuczak, A. Ruciński and J. Gruszka, On the evolution of a random tournament (to appear).
7. J.W. Moon, *Topics on tournaments*, Holt, Rinehart and Winston, New York, 1968.
8. P.A. Moran, On the method of paired comparisons, *Biometrica* **34** (1947), 363–365.
9. K. Nowicki, Asymptotic normality of triad counts in random digraphs, *Stochastic Models* **5** (1989), 163–180.
10. A. Ruciński, Proving normality in combinatorics, *Random graphs, vol. 2*, Wiley, New York, 1992, pp. 215–232.

Svante Janson, Department of Mathematics, Uppsala University, PO Box 480, S-751 06 Uppsala, Sweden

E-mail address: svante.janson@math.uu.se