Combinatorics, Probability and Computing (2007) 16, 239–249. © 2006 Cambridge University Press
doi:10.1017/S0963548306008170 First published online 3 November 2006 Printed in the United Kingdom

Hamiltonian Cycles in Regular Tournaments

BILL CUCKLER

Department of Mathematics, Rutgers University, Piscataway, NJ08854, USA
(e-mail: cuckler@math.rutgers.edu)

Received 8 December 2004; *revised 3 July 2006*

We show that every regular tournament on n vertices has at least $n!/(2+o(1))^n$ Hamiltonian cycles, thus answering a question of Thomassen [17] and providing a partial answer to a question of Friedgut and Kahn [7]. This compares to an upper bound of about $O(n^{0.25}n!/2^n)$ for arbitrary tournaments due to Friedgut and Kahn (somewhat improving Alon's bound of $O(n^{0.5}n!/2^n)$). A key ingredient of the proof is a martingale analysis of self-avoiding walks on a regular tournament.

1. Introduction

A *tournament* is an oriented complete graph. (For background on tournaments, see [12]. Some notation is given in Section 2.) A tournament is called *regular* if for any vertex v the number of outneighbours of v is the same as the number of inneighbours. Equivalently, a tournament is regular if the outdegrees of all the vertices are the same (note this implies n is odd).

In what is usually considered the first use of the probabilistic method in combinatorics, Szele [16] proved that there exists a tournament with at least $n!/2^{n-1}$ Hamiltonian paths. He proved this by observing that the expected number of Hamiltonian paths in a random tournament is $n!/2^{n-1}$. The same method shows that there exists a tournament with at least $(n-1)!/2^n$ Hamiltonian cycles. Later, Alon [2] observed that Brégman's theorem on the permanent of $\{0,1\}$ matrices [5] implies an upper bound of $O(n^{0.5}n!/2^n)$ on the number of Hamiltonian cycles in an arbitrary tournament. He used this to derive an upper bound of $O(n^{1.5}n!/2^n)$ on the number of Hamiltonian paths, thus proving a strong form of a conjecture of Szele. Friedgut and Kahn [7], extending an entropy proof of Brégman's theorem due to J. Radhakrishnan [13], improved the upper bound to about $n^{1/2-0.2507}n!/2^n$. They asked for a *lower* bound on the number of Hamiltonian cycles in a *regular* tournament of $n^{n-o(n)}$, and speculated that it might even be that the number of Hamiltonian cycles in *any* regular tournament is at least $\Omega((n-1)!/2^n)$, so of the same order of magnitude as the expected number in a random tournament.

https://doi.org/10.1017/S0963548306008170 Published online by Cambridge University Press

Earlier, C. Thomassen [17] established a lower bound on the number of Hamiltonian cycles in a tournament as a function of the minimum outdegree. He remarked that his result gives a lower bound of

$$\frac{\left(\frac{1}{2}(n-1)\right)!}{\left[\frac{1}{6}(n+15)\right]!}$$

Hamiltonian cycles in a regular tournament, and asked if there exists a positive constant δ such that every regular tournament has at least $\delta^n n!$ Hamiltonian cycles. An analogous question for Hamiltonian *paths* was independently suggested in D. West's open problems column [19], where it was attributed to C. Grinstead.

Adler, Alon and Ross [1] improved on the above-mentioned analogue of Szele's theorem for cycles, showing that the expected number of Hamiltonian cycles in a random (uniform) member of a suitable set of regular tournaments is $(e-o(1))(n-1)!/2^{(n-1)}$. Wormald [20] showed the same asymptotic value for the expected number of Hamiltonian cycles in a random (uniform) regular tournament and exhibited a family of regular tournaments for which the expected number of Hamiltonian cycles in a uniform member is about $2.85584n!/2^{n-1}$.

Here we prove a fairly similar *lower* bound on the number of Hamiltonian cycles in an *arbitrary* regular tournament.

Theorem 1.1. *For every $\epsilon > 0$, there exists an N such that for any $n > N$, any regular tournament T on n vertices has at least $n!/(2+\epsilon)^n$ Hamiltonian cycles.*

In particular this answers Thomassen's question in the affirmative, and in fact gives essentially the best possible value of δ. We make no attempt to optimize $N(\epsilon)$.

To prove the theorem, we work with the *self-avoiding random walk* on T. This is the walk $X = (X_0, X_1, \ldots, X_L)$, where X_0 is some fixed vertex and for $0 \leqslant i \leqslant L-1$, X_{i+1} is chosen uniformly from $N^+(X_i) \setminus \{X_0, \ldots, X_{i-1}\}$. This is not quite a proper definition, since we may have $N^+(X_i) \subseteq \{X_0, X_1, \ldots X_{i-1}\}$, but we will show that getting stuck in this way is unlikely and not something we need to worry about. We stop the walk at time $L = n - n/\log^2(n)$ (following a standard abuse we treat large numbers as if they were integers) and would like to show that there are many possibilities for X, and that most of these can be completed to Hamiltonian cycles. Defining T_k to be the subtournament on the vertices $V \setminus \{X_0, \ldots, X_{k-1}\}$, most of our work in proving Theorem 1.1 will be devoted to showing that T_k is usually 'fairly regular' for any $k \leqslant L$.

The following definitions will be useful for describing when T_k is 'fairly regular'. We say $A \subseteq V$ is *relevant* if it is $N^+(v)$ for some $v \in V$. For a fixed set B of $(n-1)/2$ vertices, we will say a walk $X_a, X_{a+1}, \ldots, X_b$ has *error less than δ in B* if $|\{X_a, \ldots, X_b\} \cap B| \in \left(\left(\frac{1}{2} - \delta\right)(b-a+1), \left(\frac{1}{2} + \delta\right)(b-a+1)\right)$ (δ will depend on n). We say the walk $X_a, X_{a+1}, \ldots, X_b$ has *error less than δ* if it has error less than δ in B for all relevant sets B. We say X has *error less than δ at a* if $X_0, \ldots, X_a$ has error less than δ. The point of this definition is that if X has error less than δ at k for a small δ, then T_k is 'fairly regular'.

Call $k \leqslant L$ a *milestone* if $k = rn^{0.8}$ with r an integer. We always take log to mean the natural logarithm (base e). Throughout the paper, we will assume that n is large enough to support our inequalities. The key to the proof of Theorem 1.1 is the following lemma.

https://doi.org/10.1017/S0963548306008170 Published online by Cambridge University Press

Lemma 1.2. *If T is a regular tournament, then w.h.p. X has error less than $n^{-0.05}$ at every milestone.*

The analysis for this is based in part on a martingale similar to one used in [9]. We will prove Theorem 1.1 (assuming Lemma 1.2) in Section 3, and Lemma 1.2 in Section 4. First we give some background.

2. Notation and background

For a tournament T, let $V(T)$ be the set of vertices of T and $E(T)$ the set of arcs of T. We will use V for $V(T)$ and n for $|V|$. We write $N^+(v)$ for the set of outneighbours of v, $N^-(v)$ for the set of inneighbours, and $d(v) = |N^+(v)|$, the *outdegree* of v. Given a tournament T, we write T^r for the tournament in which every arc of T is reversed. We say an m-vertex tournament is α-*regular* if $d^+(v) \in (\frac{m}{2} - \alpha m, \frac{m}{2} + \alpha m)$ for all $v \in V$. A tournament T is called *Hamiltonian connected* if any pair of vertices can be joined by a Hamiltonian path in either direction.

A *Markov random walk* on a tournament is $Y_0, Y_1, \ldots$ where Y_0 is an arbitrary but fixed vertex, while Y_{i+1} is chosen uniformly from $N^+(Y_i)$ for $i \geqslant 0$. This is a Markov chain, and for all tournaments we will consider, the chain will have a unique stationary distribution (see [10] for Markov chain background). Given a probability distribution λ on V and $A \subseteq V$, let $\lambda(A) = \sum_{v \in A} \lambda(v)$. Given two distributions γ and λ on V, let $\|\gamma - \lambda\| = \max_{A \subseteq V} |\gamma(A) - \lambda(A)|$ (the *variation distance* between γ and λ).

For general basic probability theory, see, *e.g.*, [6]. We briefly recall the definition of conditional expectation in finite settings. Throughout this paragraph, Ω will be a finite underlying set of some probability space. Given a partition σ of Ω and $\omega, \omega' \in \Omega$, we say $\omega \equiv_\sigma \omega'$ if ω and ω' belong to the same block of σ. Given a partition σ of Ω and $\Psi : \Omega \to R$, $E[\Psi \mid \sigma]$ (the *conditional expectation of* Ψ *given* σ) is the function from Ω to R given by $E[\Psi \mid \sigma](\omega) = E[\Psi(\omega') \mid \omega' \equiv_\sigma \omega]$ (so $E[\Psi \mid \sigma]$ is constant on each block of σ). For random variables $\Psi_0, \ldots, \Psi_k$ defined on Ω, and $\Psi : \Omega \to R$, $E[\Psi \mid \Psi_0, \ldots, \Psi_k]$ is just $E[\Psi \mid \sigma]$, where σ is the partition of Ω generated by $\Psi_0, \ldots, \Psi_k$, that is, the partition given by $\omega \equiv_\sigma \omega' \iff \Psi_m(\omega) = \Psi_m(\omega')$ for all $m \leqslant k$. Given two partitions σ_1, σ_2 of Ω, write $\sigma_1 \preceq \sigma_2$ if for all $A \in \sigma_1$, there exists $B \in \sigma_2$ such that $A \subseteq B$ (σ_1 is a refinement of σ_2). Supposing $\sigma_1 \preceq \sigma_2$, then

$$E\{E[\Psi \mid \sigma_1] \mid \sigma_2\} = E[\Psi \mid \sigma_2] = E\{E[\Psi \mid \sigma_2] \mid \sigma_1\}. \tag{2.1}$$

A *coupling* of two Ω-valued random variables Z_1 and Z_2 with distributions μ_1 and μ_2, respectively, is a distribution μ on $\Omega \times \Omega$ such that, for all $y \in \Omega$, $\mu(\{y\} \times \Omega) = \mu_1(y)$ and $\mu((\Omega \times \{y\}) = \mu_2(y)$. The variation distance of μ_1 and μ_2 turns out to be the minimum over couplings μ (of μ_1 and μ_2) of $\mu(Z_1 \neq Z_2)$; and, in particular, for any such μ we have $\|\mu_1 - \mu_2\| \leqslant \mu(Z_1 \neq Z_2)$ [8]. We will refer to this inequality as the *coupling inequality*.

3. Proof of Theorem 1.1

Here we prove Theorem 1.1 assuming Lemma 1.2.

https://doi.org/10.1017/S0963548306008170 Published online by Cambridge University Press

Proof of Theorem 1.1. Let $X = (v = X_0, \dots, X_L)$ be the self-avoiding random walk of length L $(= n - n/\log^2(n))$ started at some vertex v. Let $\mathscr{P}$ be the set of those paths of length L starting from v which have error less than $n^{-0.05}$ at every milestone. By Lemma 1.2, $X \in \mathscr{P}$ w.h.p. Recall that T_k is the subtournament on the vertices $V \setminus \{X_0, \dots, X_{k-1}\}$. For $X \in \mathscr{P}$, $|N^+(X_i) \cap T_i| \geqslant n/2 - i(0.5 + n^{-0.05}) - n^{0.8}$ for all $i \leqslant L$. Consequently, for any $P \in \mathscr{P}$,

$$\Pr(X = P) \leqslant \prod_{i=1}^{L} \frac{1}{n/2 - i(0.5 + n^{-0.05}) - n^{0.8}}.$$

It follows that

$$\begin{aligned} |\mathscr{P}| &\geqslant (1 - o(1)) \prod_{i=1}^{L} \left(n/2 - i(0.5 + n^{-0.05}) - n^{0.8}\right) \\ &\geqslant \prod_{i=1}^{L} \left((n/2 - i/2)(1 - o(1))\right) \\ &\geqslant n!/(2 + o(1))^n. \end{aligned}$$

We complete the proof by showing

$$\text{any } X \in \mathscr{P} \text{ can be completed to a Hamiltonian cycle.} \tag{3.1}$$

To verify (3.1), observe that $X \in \mathscr{P}$ implies that the subtournament of T on $V \setminus \{X_1, \dots, X_{L-1}\}$ is $(n^{-0.05} \cdot n/(n - L))$-regular. We finish using a theorem of Thomassen [18] (Theorem 4.5 on p. 153), which implies (in our terminology) that a $(1/10 - 9/n)$-regular tournament is Hamiltonian connected. □

4. Proof of Lemma 1.2

Before proving Lemma 1.2, we need a few additional lemmas.

For $A \subseteq V$, set $S(A) = \{z \in V \mid d_A^+(z) > 0.09n\}$ and $R(A) = \{z \in V \mid d_A^-(z) > 0.09n\}$.

Lemma 4.1. *Let T be a (0.01)-regular tournament on n vertices. Then, for any $B \subseteq V$ with $|B| > 0.49n$ and any $x \in V$, $|N^+(x) \cap S(B)| > 0.08n$ and $|N^-(x) \cap R(B)| > 0.08n$.*

Proof. We prove only the first assertion since the second assertion is just the first applied to T^r. Let $K = V \setminus S(B)$, $K_1 = K \cap B$, $K_2 = K \setminus B$. Since (for any x) $|N^+(x)| > 0.49n$, it is enough to prove that $|K| = |K_1| + |K_2| < 0.41n$. There is some $v \in K_1$ with $d_{K_1}^+(v) \geqslant (|K_1| - 1)/2$; so, since $v \in K$, $(|K_1| - 1)/2 < 0.09n$. Therefore, $|K_1| < 0.185n$. Also, there is some $w \in K_2$ with $d_{K_2}^-(w) \geqslant (|K_2| - 1)/2$. Since $w \in K_2$, $0.51n > d^-(w) > (0.49 - 0.09)n + \frac{|K_2|-1}{2}$. It follows that $|K_2| < 0.225n$. □

From now on, π will be the stationary distribution for the Markov random walk on T. A crucial point, formalized in the next lemma, is that if T is close to regular then π is close to uniform.

Lemma 4.2. *Let T be a tournament with n vertices in which $d(v) \in (\frac{n}{2}(1 - \delta), \frac{n}{2}(1 + \delta))$ for all v, where $\delta < 0.00001$. Then, for all $v \in V$, $\pi(v) \in (\frac{1-300\delta}{n}, \frac{1+300\delta}{n})$.*

https://doi.org/10.1017/S0963548306008170 Published online by Cambridge University Press

Proof. Let $U = \{v \in V \mid \pi(v) < \frac{1}{n}\}$, and $W = V \setminus U$. Let u be a vertex with $\pi(u)$ minimum and w a vertex with $\pi(w)$ maximum. We just do this when $|W| \geqslant 0.5n$, the case $|U| \geqslant 0.5n$ being similar. We will show that $\pi(u) \geqslant \frac{1-200\delta}{n}$.

Since $|W| > 0.49n$ and for all v we have $0.49n < d(v) < 0.51n$, Lemma 4.1 implies that $|N^-(u) \cap R(W)| > 0.08n$. Therefore, defining $P = \{(y,x) \in E(T) \mid y \in W, x \in N^-(u)\}$, we have $|P| > 0.007n^2$.

Now consider

$$\pi(u) = \sum_{v \to u} \frac{\pi(v)}{d(v)} \geqslant \frac{2}{n(1+\delta)} \sum_{v \to u} \pi(v) = \frac{2}{n(1+\delta)} \sum_{v \to u} \sum_{x \to v} \pi(x)/d(x)$$
$$\geqslant \frac{4}{n^2(1+\delta)^2} \sum_{v \to u} \sum_{x \to v} \pi(x).$$

There are at least $\frac{n^2}{4}(1-\delta)^2$ summands, each of which is at least $\pi(u)$ by definition of u. Furthermore, at least $0.007n^2$ of the summands are at least $\frac{1}{n}$ (since $|P| > 0.007n^2$). Therefore,

$$\frac{n^2}{4}(1+\delta)^2\pi(u) \geqslant \left(0.243n^2 - \frac{\delta n^2}{2}\right)\pi(u) + 0.007n^2\frac{1}{n}.$$

This implies that $\pi(u) \geqslant \frac{1}{n(1+150\delta)}$, which by our assumption on δ implies that $\pi(u) \geqslant \frac{1-200\delta}{n}$.

Now, still assuming that $|W| \geqslant 0.5n$, we will show that $\pi(w) < \frac{1+300\delta}{n}$. We have already argued that $\pi(v) > \frac{1-200\delta}{n}$ for any $v \in V$. Therefore, since $|U| < 0.5n$, we have

$$\sum_{x \in W} (\pi(x) - 1/n) = \sum_{x \in U} (1/n - \pi(x)) < 100\delta.$$

It follows that for any set of vertices S,

$$\sum_{v \in S} (\pi(v) - 1/n) < 100\delta. \tag{4.1}$$

Therefore, we have

$$\pi(w) < \frac{2}{(1-\delta)n} \sum_{v \to w} \pi(v) \leqslant \frac{2}{(1-\delta)n}\left(\frac{d^-(w)}{n} + 100\delta\right)$$
$$< \frac{2}{(1-\delta)n}\left(\frac{(1+\delta)n/2}{n} + 100\delta\right) < \frac{1+300\delta}{n}.$$ □

In the following lemma, we will compare the Markov walk $Y_0, Y_1, \ldots$ to the self-avoiding walk $X_0, X_1, \ldots$ assuming that $X_0 = Y_0$, and show under a mild regularity assumption that the distributions of the random variables X_q and Y_q are close for small q.

Lemma 4.3. *Let T be a tournament with n vertices such that $d(v) > \frac{n}{3}$ for all $v \in V$. Then if $X_0 = Y_0$ and $q \in N$, $\|X_q - Y_q\| < 3q^2/n$.*

Proof. Let μ_1 be the law of $X_0, X_1, \ldots, X_q$ and μ_2 that of $Y_0, Y_1, \ldots, Y_q$. We will define a coupling μ of μ_1 and μ_2 for which $\mu(X_q \neq Y_q) < 3q^2/n$, and then the coupling inequality (see the end of Section 2) establishes the lemma. We choose μ as follows: for any $i < q$, if

https://doi.org/10.1017/S0963548306008170 Published online by Cambridge University Press

$X_i = Y_i$, then we choose μ to maximize the probability that $X_{i+1} = Y_{i+1}$. Otherwise, μ is arbitrary.

Claim. For $i < q$, $\mu(X_{i+1} \neq Y_{i+1} \mid X_0 = Y_0, X_1 = Y_1, \ldots, X_i = Y_i) < 3q/n$.

This is enough because

$$\mu(X_q \neq Y_q) \leqslant \sum_{k=1}^{q} \mu(X_k \neq Y_k \mid X_0 = Y_0, \ldots X_{k-1} = Y_{k-1}).$$

To prove the claim, we observe that $|N^+(Y_i) \cap \{Y_0, Y_1, \ldots, Y_{i-1}\}| < q$ and there are at least $n/3$ possibilities for Y_{i+1}, so $\Pr(Y_{i+1} \in \{Y_0, Y_1, \ldots, Y_i\}) < 3q/n$. Therefore, if μ is defined so that $X_{i+1} = Y_{i+1}$ whenever $X_0 = Y_0, \ldots, X_i = Y_i$ and $Y_{i+1} \notin \{Y_0, Y_1, \ldots, Y_i\}$, then $\mu(X_{i+1} \neq Y_{i+1} \mid X_0 = Y_0, X_1 = Y_1, \ldots, X_i = Y_i) < 3q/n$. □

The next lemma says that the Markov random walk on T does not take too long to get close to its stationary distribution.

Lemma 4.4. *If T is a (0.01)-regular tournament then $\|\pi - Y_C\| < \frac{1}{n}$ whenever $C > 100\log(n)$.*

Proof. We first define another Markov random walk on T which we denote by $Z_0, \ldots, Z_C$. Let Z_0 be chosen according to π, and let $Z_0, \ldots, Z_C$ be the C-step Markov random walk with first vertex Z_0. The distribution of Z_i is π for all i. Let μ_1 be the law of $Y_0, Y_1, \ldots, Y_C$, and μ_2 that of $Z_0, Z_1, \ldots, Z_C$. We will define a coupling μ of μ_1 and μ_2 for which $\mu(Y_C \neq Z_C) < \frac{1}{n}$; the lemma then follows via the coupling inequality.

We choose μ as follows. For any i, if $Y_i = Z_i$, then $Y_{i+1} = Z_{i+1}$. Otherwise, if i is even, we choose Y_{i+1} and Z_{i+1} independently (according to μ_1 and μ_2), and if i is odd we choose μ to maximize the probability that $Y_{i+1} = Z_{i+1}$. It is then enough to show the following.

Claim. For even i and arbitrary $y, z \in V$, $\mu(Y_{i+2} = Z_{i+2} \mid Y_i = y, Z_i = z) > 0.0225$.

This implies that for i even, $\mu(Y_{i+2} = Z_{i+2} \mid Y_0 \neq Z_0, \ldots, Y_i \neq Z_i) > 0.0225$, so

$$\begin{aligned}\mu(Y_C \neq Z_C) &= \mu(Y_0 \neq Z_0, \ldots, Y_C \neq Z_C)\\ &\leqslant \Pi_{k=1}^{C/2} \mu(Y_{2k} \neq Z_{2k} \mid Y_0 \neq Z_0, \ldots, Y_{2k-2} \neq Z_{2k-2})\\ &\leqslant (1 - 0.0225)^{50\log(n)} < \frac{1}{n},\end{aligned}$$

as desired.

To prove the claim, we note that by Lemma 4.1,

$$\mu(Y_{i+1} \in S(N^+(Z_{i+1})) > \frac{0.08n}{d(y)} > 0.15,$$

and that *if* $Y_{i+1} \in S(N^+(Z_{i+1}))$, the probability (under μ) that $Y_{i+2} = Z_{i+2}$ is at least

$$\frac{|N^+(Y_{i+1}) \cap N^+(Z_{i+1})|}{\max\{d(Y_{i+1}), d(Z_{i+1})\}} > \frac{0.09n}{0.51n} > 0.15.$$ □

https://doi.org/10.1017/S0963548306008170 Published online by Cambridge University Press

A key point in our argument is the use of martingales to establish the following lemma. For background on the use of martingales and Azuma's inequality in combinatorial problems, see [3], [4], or [11]; we briefly recall a few basics. A sequence of random variables $\Psi_1, \ldots, \Psi_k$ is called a *martingale difference sequence* if, for all $i \leqslant k$, $E[\Psi_i \mid \Psi_1, \ldots, \Psi_{i-1}] = 0$. Azuma's inequality states that if $\Psi_1, \ldots, \Psi_k$ is a martingale difference sequence and $|\Psi_i| \leqslant 1$ for all $i \leqslant k$, then for $t > 0$

$$\Pr\left(\left| \sum_{i=1}^{k} \Psi_i \right| > t\sqrt{l} \right) \leqslant 2e^{-t^2/2}. \tag{4.2}$$

We will use this to show that the number of times a self-avoiding random walk is in a set $B \subseteq V$ is unlikely to be far from a 'dynamic average' of the number of occurrences of the random walk in B. For $i < q$, we will take $\Pr[X_i \in B \mid X_0, X_1, \ldots, X_{i-q}]$ to mean simply $\Pr[X_i \in B \mid X_0]$.

Lemma 4.5. *Let $(v = X_0, X_1, \ldots X_l)$ be the self-avoiding random walk of length l on an arbitrary tournament T. Then, for all $t > 0$, $q \in \mathbf{N}$ with $q < l$ and $B \subseteq V$,*

$$\Pr\left(\left| \sum_{i=1}^{l} \Pr[X_i \in B \mid X_0, X_1, \ldots, X_{i-q}] - |\{X_1, X_2, \ldots, X_l\} \cap B| \right| > tq\sqrt{l} \right) < 2qe^{-t^2/2}. \tag{4.3}$$

The lemma (and proof) are valid in more general settings, *i.e.*, not just for the self-avoiding random walk.

Proof. We will write the summand in (4.3) as a telescoping sum and apply Azuma's inequality.

For $i \in [l]$, set

$$\Psi(i,k) = \Pr[X_i \in B \mid X_0, \ldots, X_k] - \Pr[X_i \in B \mid X_0, \ldots, X_{k+1}].$$

Note this means that $\Psi(i, i-1) = \Pr[X_i \in B \mid X_0, \ldots, X_{i-1}] - \mathbf{1}_{\{X_i \in B\}}$, and for $k < 0$, $\Psi(i,k) = 0$. We have

$$\begin{aligned} &\left| \sum_{i=1}^{l} \Pr[X_i \in B \mid X_0, \ldots, X_{i-q}] - |\{X_1, X_2, \ldots, X_l\} \cap B| \right| \\ &\quad = \left| \sum_{i=1}^{l} \left(\Pr[X_i \in B \mid X_0, \ldots, X_{i-q}] - \mathbf{1}_{\{X_i \in B\}} \right) \right| \\ &\quad = \left| \sum_{i=1}^{l} \sum_{m=1}^{q} \Psi(i, i-m) \right| \leqslant \sum_{m=1}^{q} \left| \sum_{i=1}^{l} \Psi(i, i-m) \right|. \end{aligned} \tag{4.4}$$

Now, fix m and let $\Psi_i = \Psi(i, i-m)$. Clearly, $|\Psi_i| \leqslant 1$ for all i. Also, $\{\Psi_i\}$ is a martingale difference sequence; in fact, we have the slightly stronger $E[\Psi_i \mid X_1, \ldots, X_{i-m}] = 0$ (it is stronger because $\Psi_1, \ldots, \Psi_{i-1}$ are functions of $X_1, \ldots, X_{i-m}$, so they generate a smaller

https://doi.org/10.1017/S0963548306008170 Published online by Cambridge University Press

partition):

$$\begin{aligned}&E[\Psi_i \mid X_0, \dots, X_{i-m}]\\&= E\left[\Pr\left[X_i \in B \mid X_0, \dots X_{i-m}\right] - \Pr\left[X_i \in B \mid X_0, \dots, X_{i-m+1}\right] \mid X_0, \dots, X_{i-m}\right]\\&= \Pr[X_i \in B \mid X_0, \dots X_{i-m}] - \Pr[X_i \in B \mid X_0, \dots, X_{i-m}] = 0,\end{aligned}$$

where we used (2.1) for the second equality.

Therefore, we may apply (4.2) to obtain

$$\Pr\left(\left|\sum_{i=1}^{l} \Psi(i, i-m) > t\sqrt{l}\right|\right) < 2e^{-t^2/2}.$$

The inequality in the lemma now follows from (4.4) and

$$\Pr\left(\sum_{m=1}^{q}\left|\sum_{i=1}^{l} \Psi(i, i-m)\right| > tq\sqrt{l}\right) \leqslant \sum_{m=1}^{q} \Pr\left(\left|\sum_{i=1}^{l} \Psi(i, i-m)\right| > t\sqrt{l}\right) \leqslant 2qe^{-t^2/2}. \quad \square$$

Define a (finite) sequence of milestones $a_0, \dots$ by setting $a_0 = 0$ and, for $i \geqslant 1$, taking a_i to be the smallest milestone which is at least $a_{i-1} + \frac{n-a_{i-1}}{603}$. We use $X(a, b) = (X_a, \dots, X_b)$ and $X(i) = (X_0, \dots, X_i)$. The ith *phase* of the walk X is $X(a_i + 1, a_{i+1})$. Given $A \subseteq V$, say a subwalk of phase m is *good in* A if it has error less than $n^{-0.1}\log^{3m}(n)$ in A and *bad in* A otherwise. Say a subwalk of phase m is *good* if it has error less than $n^{-0.1}\log^{3m}(n)$ and *bad* otherwise. A *basic segment* is a subwalk $X(a, b)$ with a, b consecutive milestones. We say that $X(i)$ is *good* (or X is *good at* i) if each basic segment contained in $X(i)$ is good.

Our main task in proving Lemma 1.2 will be to show the following.

Lemma 4.6. *If $j' < j$ are consecutive milestones such that X is good at j', then, with probability at least $1 - n^{-1}$, X is good at j.*

We now prove Lemma 1.2 using Lemma 4.6.

Proof of Lemma 1.2. Since there are $O(n^{0.2})$ milestones, X is good at every milestone with probability at least $(1 - n^{-1})^{O(n^{0.2})} = 1 - o(1)$. Also, since there are $O(\log\log(n))$ phases, any basic segment will have error less than $n^{-0.1}(\log(n))^{O(\log\log(n))} < n^{-0.05}$. $\square$

We now turn to the proof of Lemma 4.6, which will occupy us for the remainder of the paper.

Proof of Lemma 4.6. Let $Q = X(j' + 1, j)$ and let m be the phase of Q. Define α and δ by $a_m = \alpha n$ and $j' = \alpha n + \delta n$. Let $I = \{i \in N \mid j' + 100\log(n) < i \leqslant j\}$. Note that for $i \in I$, $X(i - 100\log n)$ determines whether $X(j')$ is good. Let $i \in I$ be fixed and let $v = i - 100\log(n)$. Given any $S \subseteq V$ and $p < L$, we let $S_p = S \setminus \{X_0, \dots, X_{p-1}\}$. Recall that T_p is the subtournament of T on V_p. Let $g'(n) = n^{-0.1}\log^{3(m-1)}(n)$ and $g(n) = n^{-0.1}\log^{3m}(n)$. Let μ be the law of the Markov random walk $X_v = Y_0, Y_1, \dots$ on T_v and let π be the stationary distribution of this walk.

https://doi.org/10.1017/S0963548306008170 Published online by Cambridge University Press

Because $v < L$,

$$|T_v| > n/\log^2(n). \tag{4.5}$$

Since $X(j')$ is good, we have for all relevant B

$$n/2 - \alpha n(0.5 + g'(n)) - \delta n(0.5 + g(n)) - n^{0.8} < |B_v|$$

$$< n/2 - \alpha n(0.5 - g'(n)) - \delta n(0.5 - g(n)), \tag{4.6}$$

implying, in particular (using (4.5)),

$$(0.5 - o(1))|T_v| < |B_v| < (0.5 + o(1))|T_v|. \tag{4.7}$$

That $1 - \alpha > 1/\log^2(n)$, $\delta < (1-\alpha)/602$, and $|T_v| = (1 - \alpha - \delta)n - O(n^{0.8})$, together with (4.6), imply that for all relevant B

$$0.5 - \frac{1}{600}g(n) < \frac{|B_v|}{|T_v|} < 0.5 + \frac{1}{600}g(n). \tag{4.8}$$

It was for this inequality that we needed the careful definition of good, the precise requirement being that it give $\frac{g(n)}{g'(n)} \gg \log^2(n)$.

Let A be a fixed relevant set. We are done if we can show that Q is bad in A with probability at most n^{-2}. We will use π_v for the stationary distribution of the Markov random walk on T_v; thus π_v depends not only on i, but on $X_0, \ldots, X_v$.

In view of (4.5) and (4.7), Lemma 4.3 implies that

$$|\Pr[X_i \in A \mid X(v)] - \mu(Y_{100\log(n)} \in A_v)| < O\left(\frac{\log^4(n)}{n}\right), \tag{4.9}$$

while Lemma 4.4 gives

$$|\mu(Y_{100\log(n)} \in A_v) - \pi_v(A_v)| < \frac{\log^2(n)}{n},$$

and combining these (and using the triangle inequality) we have

$$|\Pr[X_i \in A \mid X(v)] - \pi_v(A_v)| < O\left(\frac{\log^4(n)}{n}\right). \tag{4.10}$$

In view of (4.8), Lemma 4.2 implies

$$\pi_v(v) \in \left(\frac{1}{|T_v|}(1 - g(n)), \frac{1}{|T_v|}(1 + g(n))\right) \tag{4.11}$$

for all $v \in V_v$.

We just estimate the probability that

$$|\{X_{j'+1}, \ldots, X_j\} \cap A| > (0.5 + g(n))n^{0.8} \tag{4.12}$$

($j - j' = n^{0.8}$), the case $|\{X_{j'+1}, \ldots, X_j\} \cap A| < (0.5 - g(n))n^{0.8}$ being similar.

By (4.8), $|A_v| < |T_v|(0.5 + g(n)/600)$. Therefore, using (4.11),

$$\pi_v(A_v) < |T_v|(0.5 + g(n)/600)\frac{1}{|T_v|}(1 + g(n)) < 0.5 + 0.51g(n).$$

https://doi.org/10.1017/S0963548306008170 Published online by Cambridge University Press

We now let i range over I, and continue to write v $(= v(i)$ for $i - 100\log(n))$. We have

$$\sum_{i\in I} \pi_v(A_v) < n^{0.8}(0.5 + 0.51g(n))$$

and

$$\sum_{i\in I} |\Pr[X_i \in A \mid X(v)] - \pi_v(A_v)| < (n^{0.8})O\left(\frac{\log^4(n)}{n}\right),$$

so (*e.g.*)

$$\sum_{i\in I} \Pr[X_i \in A \mid X(v)] < n^{0.8}(0.5 + 0.52g(n)). \tag{4.13}$$

In view of this, the event (4.12) is contained in

$$\left\{ \left| \sum_{i\in I} \Pr[X_i \in A \mid X(v)] - |\{X_{j'+100\log n}, \ldots, X_j\} \cap A| \right| > 0.4n^{0.7} \right\}.$$

But according to Lemma 4.5, the probability of this last event is less than $e^{-n^{0.59}}$. Therefore, the probability of (4.12) is less than $e^{-n^{0.59}}$, so the probability that Q is bad in A is less than n^{-2}. □

Acknowledgement

I would like to thank Jeff Kahn for mentioning the problem and providing helpful suggestions, and also for careful editing.

References

[1] Adler, I., Alon, N. and Ross, S. M. (2001) On the maximum number of Hamiltonian paths in tournaments. *Random Struct. Alg.* **18** 291–296.
[2] Alon, N. (1990) The maximum number of Hamiltonian paths in tournaments. *Combinatorica* **10** 319–324.
[3] Alon, N. and Spencer, J. (2000) *The Probabilistic Method*, 2nd edn, Wiley-Interscience, New York.
[4] Bollobás, B. (1988) Martingales, isoperimetric inequalities and random graphs. In *Combinatorics* (A. Hajnal, L. Lovász and V. T. Sós, eds), Vol. 52 of *Colloq. Math. Soc. János Bolyai*, North Holland.
[5] Brégman, L. (1973) Some properties of nonnegative matrices and their permanents. *Math. Doklady* **14** 945–949.
[6] Feller, W. (1968) *An Introduction to Probability Theory and its Applications*, Vol. 1, Wiley, New York.
[7] Friedgut, E. and Kahn, J. (2005) On the number of Hamiltonian cycles in a tournament. *Combin. Probab. Comput.* **14** 769–781.
[8] Grimmett, G. (1999) *Percolation*, 2nd edn, Springer, New York.
[9] Kahn, J. and Kim, J. H. (1998) Random matchings in regular graphs. *Combinatorica* **8** 201–226.
[10] Kemeny, J. and Snell, J. (1960) *Finite Markov Chains*, Van Nostrand Reinhold, New York.
[11] McDiarmid, C. J. H. (1989) On the method of bounded differences. In *Surveys in Combinatorics 1989: Invited Papers at the 12th British Combinatorial Conference* (J. Siemons, ed.), Cambridge University Press, pp. 148–188.

https://doi.org/10.1017/S0963548306008170 Published online by Cambridge University Press

[12] Moon, J. (1968) *Topics on Tournaments*, Holt, Rinehart and Winston, New York.
[13] Radhakrishnan, J. (1997) An entropy proof of Brégman's Theorem. *J. Combin. Theory Ser. A* **77** 161–164.
[14] Radhakrishnan, J. Entropy and counting, *November 2001 Survey*.
[15] Schrijver, A. (1978) A short proof of Minc's conjecture. *J. Combin. Theory Ser. A* **25** 80–83.
[16] Szele, T. (1943) Kombinatorikai vizsgalatok az iranyitott teljes graffal. *kapcsolatban, Mt. Fiz. Lapok* **50** 223–256.
[17] Thomassen, C. (1985) Hamilton circuits in regular tournaments. *Annals Discrete Math.* **27** 159–162.
[18] Thomassen, C. (1980) Hamiltonian-connected tournaments *J. Combin. Theory Ser. B* **28** 142–163.
[19] West, D. (1991) Open Problem Column 3. *SIAM Activity Group Newsletter in Discrete Mathematics*, Summer 1991.
[20] Wormald, N. Tournaments with many Hamiltonian cycles. Preprint.

https://doi.org/10.1017/S0963548306008170 Published online by Cambridge University Press