Proof of an entropy conjecture of Leighton and Moitra

Hüseyin Acan [*†]
huseyin.acan@rutgers.edu

Pat Devlin [*‡]
prd41@math.rutgers.edu

Jeff Kahn [*‡]
jkahn@math.rutgers.edu

Abstract

We prove the following conjecture of Leighton and Moitra. Let T be a tournament on $[n]$ and $\mathfrak{S}_n$ the set of permutations of $[n]$. For an arc uv of T, let $A_{uv} = \{\sigma \in \mathfrak{S}_n : \sigma(u) < \sigma(v)\}$.

Theorem. For a fixed $\varepsilon > 0$, if $\mathbb{P}$ is a probability distribution on $\mathfrak{S}_n$ such that $\mathbb{P}(A_{uv}) > 1/2 + \varepsilon$ for every arc uv of T, then the binary entropy of $\mathbb{P}$ is at most $(1 - \vartheta_\varepsilon)\log_2 n!$ for some (fixed) positive ϑ_ε.

When T is transitive the theorem is due to Leighton and Moitra; for this case we give a short proof with a better ϑ_ε.

1 Introduction

In what follows we use log for $\log_2$ and $H(\cdot)$ for binary entropy. The purpose of this note is to prove the following natural statement, which was conjectured by Tom Leighton and Ankur Moitra [6] (and told to the third author by Moitra in 2008).

Theorem 1. *Let T be a tournament on $[n]$ and σ a random (not necessarily uniform) permutation of $[n]$ satisfying:*

$$\text{for each arc } uv \text{ of } T,\ \ \mathbb{P}(\sigma(u) < \sigma(v)) > 1/2 + \varepsilon. \tag{1}$$

AMS 2010 subject classification: 05C20, 05D40, 94A17, 06A07

Key words and phrases: entropy, permutations, tournaments, regularity

[*]Department of Mathematics, Rutgers University

[†]Supported by National Science Foundation Fellowship (Award No. 1502650).

[‡]Supported by NSF grant DMS1501962.

Then

$$H(\sigma) \le (1 - \vartheta) \log n!, \tag{2}$$

where $\vartheta > 0$ *depends only on* ε.

(We will usually think of permutations as bijections $\sigma : [n] \to [n]$). The original motivation for Leighton and Moitra came mostly from questions about sorting partially ordered sets; see [6] for more on this.

For the special case of *transitive* T, Theorem 1 was proved in [6] with $\vartheta_\varepsilon = C\varepsilon^4$. Note that for a *typical* (a.k.a. *random*) T, the conjecture's hypothesis is unachievable, since, as shown long ago by Erdős and Moon [2], no σ agrees with T on more than a $(1/2+o(1))$-fraction of its arcs. In fact, it seems natural to expect that transitive tournaments are the *worst* instances, being the ones for which the hypothesized agreement is easiest to achieve. From this standpoint, what we do here may be considered somewhat unsatisfactory, as our ϑ's (for general T) are quite a bit worse than what [6] gives for transitive T. In this special case it's easy to see [6, Claim 4.14] that one can't take ϑ greater than 2ε, which seems likely to be close to the truth. We make some progress on this, giving a surprisingly simple proof of the following improvement of [6].

Theorem 2. *For* T, $\mathbb{P}$, σ *as Theorem 1 with* T *transitive,*

$$H(\sigma) \le (1 - \varepsilon^2/8) n \log n.$$

The proof of Theorem 1 is given in Section 3 following brief preliminaries in Section 2. The underlying idea is similar to that of [6], which in turn was based on the beautiful tournament ranking bound of W. Fernandez de la Vega [1]; see Section 3 (end of "Sketch") for an indication of the relation to [6]. Theorem 2 is proved in Section 4.

2 Preliminaries

Usage

In what follows we assume n is large enough to support our arguments and pretend all large numbers are integers.

As usual $G[X]$ is the subgraph of G induced by X; we use $G[X, Y]$ for the bipartite subgraph induced (in the obvious sense) by disjoint X and Y. For a digraph D, $D[X]$ and $D[X, Y]$ are used analogously. For both graphs and digraphs, we use $|\cdot|$ for number of edges (or arcs).

Also as usual, the *density* of a pair (X, Y) of disjoint subsets of $V(G)$ is $d(X,Y) = d_G(X,Y) = |G[X,Y]|/(|X||Y|)$, and we extend this to bipartite digraphs D in which

$$\text{at most one of } D \cap (X \times Y),\ D \cap (Y \times X) \text{ is nonempty.} \tag{3}$$

For a digraph D, D^r is the digraph gotten from D by reversing its arcs.

Write $\mathfrak{S}_n$ for the set of permutations of $[n]$. For $\sigma \in \mathfrak{S}_n$, we use T_σ for the corresponding (transitive) tournament on $[n]$ (that is, $uv \in T_\sigma$ iff $\sigma(u) < \sigma(v)$) and for a digraph D (on $[n]$) define

$$\mathrm{fit}(\sigma, D) = |D \cap T_\sigma| - |D^r \cap T_\sigma|$$

(e.g. when D is a tournament, this is a measure of the quality of σ as a ranking of D).

Regularity

Here we need just Szemerédi's basic notion [7] of a regular pair and a very weak version (Lemma 3) of his Regularity Lemma. As usual a bipartite graph H on disjoint $X \cup Y$ is *δ-regular* if

$$|d_H(X', Y') - d_H(X, Y)| < \delta$$

whenever $X' \subseteq X$, $Y' \subseteq Y$, $|X'| > \delta|X|$ and $|Y'| > \delta|Y|$, and we extend this in the obvious way to the situation in (3). It is easy to see that if a bigraph H is δ-regular then its bipartite complement is as well; this implies that for a tournament T on $[n]$ and X, Y disjoint subsets of $[n]$,

$$T \cap (X \times Y) \text{ is } \delta\text{-regular if and only if } T \cap (Y \times X) \text{ is.} \tag{4}$$

The following statement should perhaps be considered folklore, though similar results were proved by János Komlós, circa 1991 (see [5, Sec. 7.3]).

Lemma 3. *For each $\delta > 0$ there is a $\beta > 2^{-\delta^{-O(1)}}$ such that for any bigraph H on $X \cup Y$ with $|X|, |Y| \geq n$, there is a δ-regular pair (X', Y') with $X' \subseteq X, Y' \subseteq Y$ and each of $|X'|, |Y'|$ at least βn.*

Corollary 4. *For each $\delta > 0$, β as in Lemma 3 and digraph $G = (V, E)$, there is a partition $L \cup R \cup W$ of V such that $E \cap (L \times R)$ is δ-regular and* $\min\{|L|, |R|\} \geq \beta|V|/2$.

Proof. Let $X \cup Y$ be an (arbitrary) equipartition of V and apply Lemma 3 to the undirected graph H underlying the digraph $G \cap (X \times Y)$. $\square$

3 Proof of Theorem 1

We now assume that σ drawn from the probability distribution $\mathbb{P}$ on $\mathfrak{S}_n$ satisfies (1) and try to show (2) (with ϑ TBA). We use $\mathbb{E}$ for expectation w.r.t. $\mathbb{P}$ and μ for uniform distribution on $\mathfrak{S}_n$.

Sketch and connection with [6]

We will produce $S_1, \ldots, S_m \subseteq T$ with $S_i \subseteq L_i \times R_i$ for some disjoint $L_i, R_i \subseteq [n]$, satisfying:

(i) with $\|S_i\| := \min\{|L_i|, |R_i|\}$, $\sum \|S_i\| = \Omega(n \log n)$ (where the implied constant depends on ε);

(ii) each S_i is δ-regular (with $\delta = \delta_\varepsilon$ TBA);

(iii) for all $i < j$, either $(L_i \cup R_i) \cap (L_j \cup R_j) = \varnothing$ or $L_j \cup R_j$ is contained in one of L_i, R_i (note this implies the S_i's are disjoint).

Let $A_i = \{\mathrm{fit}(\sigma, S_i) > \varepsilon|S_i|\}$ and $Q = \{\sum \mathbb{1}_{A_i} \cdot ||S_i|| = \Omega(n \log n)\}$ (where we follow the standard use of curly brackets to denote events). The main points are then:

(a) $\mathbb{P}(Q)$ is bounded below by a positive function of ε. (This is just (i) together with a couple applications of Markov's Inequality.)

(b) Regularity of S_i implies $\mu(A_i) \leq \exp[-\Omega(\|S_i\|)]$.

(c) Under (iii), for any $I \subseteq [m]$,

$$\mu(\cap_{i \in I} A_i) < \exp[-\textstyle\sum_{i \in I} \Omega(\|S_i\|)]$$

(a weak version of independence of the A_i's under μ).

And these points easily combine to give (2) (see (6) and (8)).

For the transitive case in [6] most of this argument is unnecessary; in particular, regularity disappears and there is a natural *decomposition* of T into S_i's: Supposing $T = \{ab : a < b\}$ and (for simplicity) $n = 2^k$, we may take the S_i's to be the sets $L_i \times R_i$ with (L_i, R_i) running over pairs

$$([(2s-2)2^{-j}n + 1, (2s-1)2^{-j}n], [(2s-1)2^{-j}n + 1, 2s2^{-j}n]), \qquad (5)$$

with $j \in [k]$ and $s \in [2^{j-1}]$. (As mentioned earlier, this decomposition of the (identity) permutation $(1, \ldots, n)$ also provides the framework for [1].)

After some translation, our argument (really, a fairly small subset thereof) then specializes to essentially what's done in [6]. $\square$

Set $\delta = .03\varepsilon$ and let β be half the β of Lemma 3 and Corollary 4. We use the corollary to find a rooted tree $\mathcal{T}$ each of whose internal nodes has degree (number of children) 2 or 3, together with disjoint subsets $S_1, S_2, \ldots, S_m$ of (the arc set of) T, corresponding to the internal nodes of $\mathcal{T}$. The nodes of $\mathcal{T}$ will be subsets of $[n]$ (so the *size,* $|U|$, of a node U is its size as a set).

To construct $\mathcal{T}$, start with root $V_1 = [n]$ and repeat the following for $k = 1, \ldots$ until each unprocessed node has size less than (say) $t := \sqrt{n}$. Let V_k be an unprocessed node of size at least t and apply Corollary 4 to $T[V_k]$ to produce a partition $V_k = L_k \cup R_k \cup W_k$, with $|L_k|, |R_k| > \beta|V_k|$ and $S_k := T \cap (L_k \times R_k)$ δ-regular of density at least 1/2. (Note (4) says we can reverse the roles of L_k and R_k if the density of $T \cap (L_k \times R_k)$ is less than 1/2.) Add L_k, R_k, W_k to $\mathcal{T}$ as the children of V_k and mark V_k "processed." (Note the V_k's are the *internal* nodes of $\mathcal{T}$; nodes of size less then t are not processed and are automatically leaves. Note also that there is no restriction on $|W_k|$ and that, for $k > 1$, V_k is equal to one of L_i, R_i, W_i for some $i < k$.)

Let m be the number of internal nodes of $\mathcal{T}$ (the final tree). Note that the leaves of $\mathcal{T}$ have size at most t and that the S_i's satisfy (ii) and (iii) of the proof sketch; that they also satisfy (i) is shown by the next lemma.

Set
$$\Lambda = \textstyle\sum_{i=1}^m |V_i|;$$
this quantity will play a central role in what follows.

Lemma 5. $\Lambda \geq \frac{1}{2} n \log_3 n$

Proof. This will follow easily from the next general (presumably known) observation, for which we assume $\mathcal{T}$ is a tree satisfying:

- the nodes of $\mathcal{T}$ are subsets of S, an s-set which is also the root of $\mathcal{T}$;
- the children of each internal node U of $\mathcal{T}$ form a partition of U with at most b blocks;
- the leaves of $\mathcal{T}$ are $U_1, \ldots, U_r$, with $|U_i| = u_i \leq t$ (any t) and depth d_i.

Lemma 6. *With the setup above,* $\sum u_i d_i \geq s \log_b (s/t)$.

(Of course this is exact if $\mathcal{T}$ is the complete b-ary tree of depth d and all leaves have size $2^{-b}s$).

Proof. Recall that the *relative entropy* between probability distributions p and q on $[r]$ is

$$D(p\|q) = \sum p_i \log(p_i/q_i) \geq 0$$

(the inequality given by the concavity of the logarithm). We apply this with $p_i = u_i/s$ and q_i the probability that the ordinary random walk down the tree ends at u_i. In particular $q_i \geq b^{-d_i}$, which, with nonnegativity of $D(p\|q)$ and the assumption $u_i \leq t$, gives

$$\begin{aligned}\sum (u_i/s) d_i \log b &\geq \sum (u_i/s)\log(1/q_i)\\ &\geq \sum (u_i/s)\log(s/u_i) \geq \log(s/t).\end{aligned}$$

The lemma follows. □

This gives Lemma 5 since $\sum |V_i| = \sum_U |U| d(U)$, with U ranging over leaves of $\mathcal{T}$ (and $d(\cdot)$ again denoting depth). □

Lemma 7. *The number m of internal nodes of $\mathcal{T}$ is less than n.*

Proof. A straightforward induction shows that the number of leaves of a rooted tree is $1 + \sum(b(w) - 1)$, where w ranges over internal nodes and b denotes number of children. The lemma follows since here the number of leaves is at most n (actually at most $3\sqrt{n}$) and each $d(w)$ is at least 2. □

Recalling that $A_i = \{\sigma \in \mathfrak{S}_n : \mathrm{fit}(\sigma, S_i) \geq \varepsilon |S_i|\}$ and that $\mathbb{E}$ refers to $\mathbb{P}$, we have $\mathbb{E}[\mathrm{fit}(\sigma, S_i)] \geq 2\varepsilon |S_i|$, which with

$$\mathbb{E}[\mathrm{fit}(\sigma, S_i)] \leq \mathbb{P}(A_i)|S_i| + (1 - \mathbb{P}(A_i))\varepsilon |S_i| \leq (\mathbb{P}(A_i) + \varepsilon)|S_i|$$

gives $\mathbb{P}(A_i) \geq \varepsilon$ (essentially Markov's Inequality applied to $|S_i| - \mathrm{fit}(\sigma, S_i)$).

Set $\xi_i = |V_i| \mathbf{1}_{A_i}$ and $\xi = \sum_i \xi_i$. Recall $\Lambda = \sum_{i=1}^m |V_i|$ and let Q be the event $\{\xi \geq \varepsilon\Lambda/2\}$. Then $\mathbb{E}[\xi_i] = |V_i|\mathbb{P}(A_i) \geq \varepsilon |V_i|$, implying $\mathbb{E}[\xi] = \sum \mathbb{E}[\xi_i] \geq \varepsilon\Lambda$, and (since $\xi_i \leq |V_i|$) $\xi \leq \Lambda$; so using Markov's Inequality as above gives $\mathbb{P}(Q) \geq \varepsilon/2$.

Thus, with σ chosen from $\mathfrak{S}_n$ according to $\mathbb{P}$, we have

$$\begin{aligned}H(\sigma) &\leq H(\mathbb{P}(Q)) + (1 - \mathbb{P}(Q))\log n! + \mathbb{P}(Q)\log|Q|\\ &\leq\ 1 + \log n! + \mathbb{P}(Q)\log\mu(Q) \leq 1 + \log n! + (\varepsilon/2)\log\mu(Q) \qquad (6)\end{aligned}$$

(recall μ is uniform measure on $\mathfrak{S}_n$).

Let

$$\mathcal{J} = \{I \subseteq [m] : \sum_{i \in I} |V_i| \geq \varepsilon\Lambda/2\}$$

and, for $I \in \mathcal{J}$, let $A_I = \cap_{i \in I} A_i$. Set

$$b = \varepsilon^2 \delta \beta^3 / 33 \tag{7}$$

(see (12) for the reason for the choice of b). We will show, for each $I \in \mathcal{J}$,

$$\mu(A_I) \leq e^{-b\varepsilon\Lambda/2}, \tag{8}$$

which implies

$$\log \mu(Q) = \log \mu(\cup_{I \in \mathcal{J}} A_I) \leq \log |\mathcal{J}| - (b\varepsilon\Lambda \log e)/2 \leq n - (b\varepsilon\Lambda \log e)/2,$$

the second inequality following from $|\mathcal{J}| \leq 2^m$ together with Lemma 7. With $c = \varepsilon^3 \delta \beta^3 / 150 < (b\varepsilon \log_3 e)/4$, this bounds (for large n) the r.h.s. of (6) by

$$(1 - \varepsilon c/2) \log n!,$$

which proves Theorem 1 with $\vartheta = \varepsilon^4 \delta \beta^3 / 300 = \exp[-\varepsilon^{-O(1)}]$. □

The rest of our discussion is devoted to the proof of (8). For a digraph $D \subseteq L \times R$ with L, R disjoint subsets of V, say a pair (X, Y) of disjoint subsets of $[n]$ with $|X| = |L|$, $|Y| = |R|$ is *safe* for D if

$$\mathrm{fit}(\tau, D) < \varepsilon |L||R|/4 \tag{9}$$

for every bijection $\tau : L \cup R \to X \cup Y$ with $\tau(L) = X$ (where $\mathrm{fit}(\tau, D)$ has the obvious meaning). We also say $\sigma \in \mathfrak{S}_n$ is *safe* for D if $(\sigma(L), \sigma(R))$ is. Note that since S_i has density at least $1/2$ in $L_i \times R_i$, the σ's in A_i are unsafe for S_i.

Lemma 8. *Assume the above setup with $|L| + |R| = l$ and $|L| = \gamma l$, and set $\lambda = 2\delta$ and $\zeta = \varepsilon\delta\gamma(1-\gamma)/4$. Let $I_1 \cup \cdots \cup I_r$ be the natural partition of $X \cup Y$ into intervals of size λl. If D is δ-regular and*

$$|X \cap I_j| = (\gamma\lambda \pm \zeta) l \quad \forall j \in [r], \tag{10}$$

then (X, Y) is safe for D.

(Of course an *interval* of $Z = \{i_1 < \cdots < i_u\}$ is one of the sets $\{i_s, \ldots, i_{s+t}\}$.)

Proof. For τ as in the line after (9), let $L_j = L\cap\tau^{-1}(I_j)$ and $R_j = R\cap\tau^{-1}(I_j)$ ($j \in [r]$). Then

$$|\mathrm{fit}(\tau, D)| \le \sum_{1\le i<j\le r} ||D\cap(L_i\times R_j)| - |D\cap(L_j\times R_i)|| + \gamma(1-\gamma)\lambda l^2. \quad (11)$$

Here the last term is an upper bound on the contribution of pairs contained in the I_j's: if $|L_j| = \gamma_j|I_j| = \gamma_j\lambda l$ (so $|R_j| = (1-\gamma_j)\lambda l$ and $\sum\gamma_j = \gamma/\lambda$), then

$$\sum\gamma_j(1-\gamma_j) \le \sum\gamma_j - (\sum\gamma_j)^2/r = (\gamma-\gamma^2)/\lambda$$

gives

$$\sum|L_j||R_j| = \sum\gamma_j(1-\gamma_j)\lambda^2 l^2 \le \gamma(1-\gamma)\lambda l^2.$$

On the other hand, regularity and (10) (which implies $|L_i| > \delta|L|$ $(=\delta\gamma l)$ since $\gamma\lambda - \zeta > \gamma\delta$, and similarly $|R_i| > \delta|R|$) give, for all $i \ne j$,

$$|D\cap(L_i\times R_j)| = (d\pm\delta)|L_i||R_j|,$$

where d is the density of D. Combining this with (10) bounds each of the summands in (11) by

$$[(d+\delta)(\gamma\lambda+\zeta)((1-\gamma)\lambda+\zeta) - (d-\delta)(\gamma\lambda-\zeta)((1-\gamma)\lambda-\zeta)]l^2$$
$$= 2[\lambda\zeta d + \delta(\gamma(1-\gamma)\lambda^2+\zeta^2)]l^2$$

and the r.h.s. of (11) by

$$\left\{2\tbinom{r}{2}[\lambda\zeta d + \delta(\gamma(1-\gamma)\lambda^2+\zeta^2)] + \gamma(1-\gamma)\lambda\right\} l^2 < \varepsilon\gamma(1-\gamma)l^2/4.$$

(The main term on the l.h.s. is the one with $\lambda\zeta d$, which, since $r^{-1} = \lambda = 2\delta$, is less than half the r.h.s. The second and third terms are much smaller (the second since δ is much smaller than ε).) □

Corollary 9. *For D and parameters as in Lemma 8, and σ uniform from $\mathfrak{S}_n$,*

$$\Pr(\sigma \textit{ is unsafe for } D) < 2r\exp[-2\zeta^2 l/\lambda].$$

Proof. Let $(X, Y) = (\sigma(L), \sigma(R))$. Once we've chosen $X\cup Y$ (determining $I_1,\ldots,I_r$), $2\exp[-2\zeta^2 l/\lambda]$ is the usual Hoeffding bound [3, Eq. (2.3)] on the probability that X violates (10) for a given j. (The bound may be more familiar when elements of $X\cup Y$ are in X independently, but also applies to the *hypergeometric* r.v. $|X\cap I_j|$; see e.g. [4, Thm. 2.10 and (2.12)].) □

Proof of (8). Let

$$B_i = \{\sigma \in \mathfrak{S}_n : \sigma \text{ is unsafe for } S_i\}$$

and $B_I = \cap_{i \in I} B_i$. Then $A_i \subseteq B_i$ (as noted above) and (therefore) $A_I \subseteq B_I$. Moreover—perhaps the central point—the B_i's are independent, since B_i depends only on the relative positions of $\sigma(L_i)$ and $\sigma(R_i)$ within $\sigma(V_i)$.

On the other hand, Corollary 9, applied with $D = S_i$ (so $L = L_i$, $R = R_i$, $l = |L_i| + |R_i|$ and $\gamma = |L_i|/l \in (\beta, 1-\beta)$) gives

$$\begin{aligned} \Pr(B_i) &< 2r \exp[-2\zeta^2 l/\lambda] < 2r \exp[-\varepsilon^2 \delta \beta^2 l/64] \\ &< 2r \exp[-\varepsilon^2 \delta \beta^3 |V_i|/32] < e^{-b|V_i|}. \end{aligned} \tag{12}$$

(Recall b was defined in (7); since we assume $|V_i|$ is large ($|V_i| > t = \sqrt{n}$), the choice leaves a little room to absorb the $2r$.) And of course (12) and the independence of the B_i's give (8). □

4 Back to the transitive case

Theorem 2 is an easy consequence of the next observation.

Lemma 10. *Let* $\mathbf{Y}$ *a random* m*-subset of* $[2m]$ *satisfying*

$$\mathbb{E}|\{(a,b) : a < b, a \in [2m] \setminus \mathbf{Y}, b \in \mathbf{Y}\}| > (\tfrac{1}{2} + \varepsilon)m^2. \tag{13}$$

Then $H(\mathbf{Y}) < (1 - \varepsilon^2/8)2m$*.*

To get Theorem 2 from this, let $T = \{ab : a < b\}$ and, for simplicity, $n = 2^k$, and decompose $T = \bigcup (L_i \times R_i)$ as in (5). For each i, say with $|L_i|$ $(= |R_i|) = m_i$, let $\mathbf{Y}_i \subseteq [2m_i]$ consist of the indices of positions within $\sigma(L_i \cup R_i)$ occupied by $\sigma(R_i)$; that is, if $\sigma(L_i \cup R_i) = \{j_1 < \cdots < j_{2m_i}\}$, then $\mathbf{Y}_i = \{l : j_l \in \sigma(R_i)\}$. Then Lemma 10 (its hypothesis provided by (1)) gives

$$H(\mathbf{Y}_i) \leq (1 - \varepsilon^2/8)2m_i;$$

so, since σ is determined by the $\mathbf{Y}_i$'s, we have

$$H(\sigma) \leq \sum H(\mathbf{Y}_i) \leq (1 - \varepsilon^2/8) \sum (2m_i) = (1 - \varepsilon^2/8) n \log n. \qquad \square$$

Remark. Note that the $\Omega(\varepsilon^2)$ of Theorem 2 is the best one can do without using (1) for more than its implication of (13) for the (L_i, R_i, Y_i)'s, which is all we are getting from it here.

Proof of Lemma 10. For $a \in [2m]$, set $\mathbb{P}(a \in \mathbf{Y}) = 1/2 + \delta_a$. Then

$$H(\mathbf{Y}) \le \sum_a H(1/2 + \delta_a) \le \sum_a (1 - 2\delta_a^2)$$

(where the 2 could actually be $2 \log e$); so it is enough to show

$$\sum \delta_a^2 \ge \varepsilon^2 m/8.$$

For a given m-subset Y of $[2m]$, we have

$$\begin{aligned} f(Y) :&= |\{(a,b) : a < b, a \in [2m] \setminus Y, b \in Y\}| \\ &= \sum_{b \in Y} (b-1) - \binom{m}{2} = \sum_{b \in Y} b - \binom{m+1}{2}. \end{aligned}$$

(the first sum counts pairs (a,b) with $a < b$ and $b \in Y$, and $\binom{m}{2}$ is the number of such pairs with a also in Y); so we have

$$(\tfrac{1}{2} + \varepsilon)m^2 < \mathbb{E} f(\mathbf{Y}) = \sum (\tfrac{1}{2} + \delta_b) b - \binom{m+1}{2} = \sum \delta_b b + m^2/2,$$

implying $\sum \delta_b b > \varepsilon m^2$. Combining this with $2m \sum_{\delta_b > 0} \delta_b \ge \sum \delta_b b$, we have $\sum_{\delta_b > 0} \delta_b > \varepsilon m/2$ and then, using Cauchy-Schwarz,

$$\sum \delta_b^2 \ge \sum_{\delta_b > 0} \delta_b^2 \ge \frac{1}{2m} (\varepsilon m/2)^2 = \varepsilon^2 m/8. \qquad \square$$

References

[1] W. Fernandez de la Vega, On the maximal cardinality of a consistent set of arcs in a random tournament, *J. Comb. Th. Series B* (1983), 328-332.

[2] P. Erdős and J. Moon, On sets of consistent arcs in a tournament, *Canad. Math. Bull.* **8** (1965), 269-271.

[3] W. Hoeffding, Probability inequalities for sums of bounded random variables, *J. Amer. Statistical Assoc.* **58** (1963), 13-30.

[4] S. Janson, T. Łuczak and A. Ruciński, *Random Graphs*, Wiley, New York, 2000.

[5] J. Komlós and M. Simonovits, Szemerédi's regularity lemma and its applications in graph theory, *Combinatorics, Paul Erdős is eighty, Vol. 2 (Keszthely, 1993)*, 295-352, Bolyai Soc. Math. Stud. 2, János Bolyai Math. Soc., Budapest, 1996.

[6] T. Leighton and A. Moitra, On Entropy and Extensions of Posets, manuscript 2011. http://people.csail.mit.edu/moitra/docs/poset.pdf.

[7] E. Szemerédi, Regular Partitions of Graphs, pp. 399-401 in *Problémes Combinatoires et Théorie des Graphes (Colloq. Internat. CNRS, Univ. Orsay, Orsay, 1976)*, Paris: Éditions du Centre National de la Recherche Scientifique (CNRS), 1978.