Discussiones Mathematicae
Graph Theory 34 (2014) 467–495
doi:10.7151/dmgt.1758

INDUCED ACYCLIC TOURNAMENTS IN RANDOM DIGRAPHS: SHARP CONCENTRATION, THRESHOLDS AND ALGORITHMS[1]

Kunal Dutta and C.R. Subramanian

The Institute of Mathematical Sciences
Taramani, Chennai–600113, India

e-mail: kdutta@imsc.res.in
crs@imsc.res.in

Abstract

Given a simple directed graph $D = (V, A)$, let the size of the largest induced acyclic tournament be denoted by $mat(D)$. Let $D \in \mathcal{D}(n,p)$ (with $p = p(n)$) be a *random* instance, obtained by randomly orienting each edge of a random graph drawn from $\mathcal{G}(n, 2p)$. We show that $mat(D)$ is asymptotically almost surely (a.a.s.) one of only 2 possible values, namely either b^* or $b^* + 1$, where $b^* = \lfloor 2(\log_r n) + 0.5 \rfloor$ and $r = p^{-1}$.

It is also shown that if, asymptotically, $2(\log_r n) + 1$ is not within a distance of $w(n)/(\ln n)$ (for any sufficiently slow $w(n) \to \infty$) from an integer, then $mat(D)$ is $\lfloor 2(\log_r n) + 1 \rfloor$ a.a.s. As a consequence, it is shown that $mat(D)$ is 1-point concentrated for all n belonging to a subset of positive integers of density 1 if p is independent of n. It is also shown that there are functions $p = p(n)$ for which $mat(D)$ is provably *not* concentrated in a single value. We also establish thresholds (on p) for the existence of induced acyclic tournaments of size i which are sharp for $i = i(n) \to \infty$.

We also analyze a polynomial time heuristic and show that it produces a solution whose size is at least $\log_r n + \Theta(\sqrt{\log_r n})$. Our results are valid as long as $p \geq 1/n$. All of these results also carry over (with some slight changes) to a related model which allows 2-cycles.

Keywords: random digraphs, tournaments, concentration, thresholds, algorithms.

2010 Mathematics Subject Classification: 05C80.

[1]A preliminary version of parts of this work appeared as an extended abstract in LATIN, 2010, Oaxaca, Mexico.

1. Introduction

By a simple directed graph, we mean a directed graph having no 2-cycles. Throughout the paper, we assume, w.l.o.g., that $V = \{1, 2, \ldots, n\}$. Given a directed graph $D = (V, A)$, we want to find the maximum size of an induced acyclic tournament in D, denoted by $mat(D)$. A *tournament* is a simple directed graph whose underlying undirected graph is a complete graph. A tournament is *acyclic* if and only if it is transitive. In this paper, we study the problem of determining $mat(D)$ for random digraphs both from an analytical and an algorithmic point of view.

We study the following model of a simple random digraph introduced in [23]. In what follows, *a.a.s.* refers to 'asymptotically almost surely'; and $p \leq 0.5$ is a real number. Throughout this paper, r denotes p^{-1}.

Model $\mathcal{D}(n,p)$: Let the vertex set be $V = \{1, 2, \ldots, n\}$. Choose each undirected edge joining distinct elements of V independently with probability $2p$. For each chosen $\{u, v\}$, independently orient it in one of the two directions $\{u \to v, v \to u\}$ in D with equal probability $= 1/2$. The resulting directed graph is an orientation of a simple graph, i.e., there are no 2-cycles.

1.1. Analytical aspects

Subramanian [23] first studied the related problem of determining $mas(D)$, the size of a largest induced acyclic subgraph in a random digraph $D = (V, E)$, and later Spencer and Subramanian [22] obtained the following result.

Theorem 1.1 [22]**.** *Let $D \in \mathcal{D}(n,p)$ and $w = np$. There is a sufficiently large constant W such that: If p satisfies $w \geq W$, then, a.a.s,*

$$mas(D) \in \left[\left(\frac{2}{\ln q} \right) (\ln w - \ln \ln w - O(1)), \left(\frac{2}{\ln q} \right) (\ln w + 3e) \right],$$

where $q = (1-p)^{-1}$.

Thus, with probability $1 - o(1)$, $mas(D)$ lies in an integer band of width $O\left(\frac{\ln \ln w}{\ln q}\right)$. But this upper bound on width is asymptotically $\Theta(r \ln \ln w)$, and hence can become large for small values of p. However, if we focus on more restricted subgraphs, namely, induced acyclic tournaments, then the optimum size can be shown (see Theorem 1.2 below) to be one of two consecutive values a.a.s. In other words, we obtain a 2-point concentration for $mat(D)$. This is one of our main results in this paper.

Theorem 1.2. *Let $\{\mathcal{D}(n,p) : p = p(n), n \geq 1\}$ be an infinite sequence of probability distributions. Let $w = w(n)$ be any sufficiently slowly growing function* *of n (say, any w with $w < \sqrt{n}$ always) such that $w \to \infty$ as $n \to \infty$. Let $D \in \mathcal{D}(n,p)$. Then, a.a.s., the following holds:*

(i) *Suppose $p \geq 1/n$. Define*

$$d = 2\log_r n + 1 = \frac{2(\ln n)}{\ln r} + 1; \quad b^* = \lfloor d - 1/2 \rfloor = \lfloor 2(\log_r n) + 0.5 \rfloor.$$

Then, $mat(D)$ is either b^* or $b^* + 1$.

(ii) $mat(D) \in \{2,3\}$ if $1/(wn) \leq p < 1/n$.

(iii) $mat(D) = 2$ if $wn^{-2} \leq p < 1/(wn)$.

(iv) $mat(D) \leq 2$ if $1/(wn^2) \leq p < wn^{-2}$.

(v) $mat(D) = 1$ if $p < (wn^2)^{-1}$.

Similar two-point concentration results are known for maximum clique size $\omega(G)$ of a random undirected graph $G \in \mathcal{G}(n,p)$ for $p \leq 0.5$ (see [5, 4, 13]). The chromatic number $\chi(G)$ is another parameter which has been shown to be 2-point concentrated for sparse random undirected graphs (see [16, 2, 1]). However, unlike the case of $mat(D)$, there is no explicit closed form expression for $\omega(G)$. With some assumptions about $p = p(n)$, one can also prove (proof presented in Section 3) a stronger one-point concentration (Theorem 1.3 below) on $mat(D)$ for all large values of n.

Theorem 1.3. *Let $\mathcal{D}(n,p)$, d be as defined in Theorem 1.2. Let $w = w(n)$ be any function so that as $n \to \infty$, $w(n) \leq 0.5(\ln n)$ and $w \to \infty$. If $p \geq 1/n$ is such that d satisfies $\frac{w}{\ln n} \leq \lceil d \rceil - d \leq 1 - \frac{w}{\ln n}$ for all large values of n, then a.a.s. $mat(D) = \lfloor d \rfloor$.*

As a consequence, we also obtain the following concentration result. For any choice of $p = p(n)$ and any given definition of $f(n) = 1 - o(1)$, let $N_{f,p}$ denote the set of natural numbers n such that $mat(D)$ takes a specific value with probability at least $f(n)$. Let us call $p = p(n)$ a *constant function* if, for some $a \in [0, 0.5]$, $p(n) = a$ for every n. Then,

Corollary 1.4. *For every constant function $p = p(n)$, there exists a function $f = f(n) = 1 - o(1)$ such that the set $N_{f,p}$ is a subset of natural numbers having density 1.*

Our proof (presented in Subsection 3.1) of the above corollary is direct and does not take recourse to the Borel-Cantelli Lemma which is applied in similar one-point concentration proofs. Perhaps, similar direct proofs are possible in other cases where the Borel-Cantelli Lemma has been used, as for example, in proving a one-point concentration result for the clique number $\omega(G)$ of random undirected graphs.

It is interesting to note that the bounds on $\lceil d \rceil - d$ assumed in Theorem 1.3 are essentially tight. We give an example of a function $p = p(n)$ such that the assumptions in Theorem 1.3 do not hold, and prove that $mat(D)$ is **not** 1-point concentrated.

Theorem 1.5. *For any fixed $j \in \mathbb{Z}^+$ (with $j \geq 3$) and $c \in \mathcal{R}^+$, let $D \in \mathcal{D}(n,p)$, $p = n^{-2/(j-1+\frac{c}{\ln n})}$. Then, for every sufficiently large n, each of the two events*
(i) $mat(D) = j - 1$ *and*
(ii) $mat(D) = j$
occurs with probability lower bounded by a positive constant.

The proof of this theorem is provided in Subsection 3.2 and is based on applying Lovász Local Lemma and Paley-Zygmund Inequality.

We also establish a threshold (on p) for the existence of induced acyclic tournaments of size i. For every fixed i, the threshold is coarse and is a sharp one if $i = i(n)$ varies with n and is any suitably growing function which goes to ∞ as $n \to \infty$. These are stated in the following theorem whose proof is presented in Subsection 3.3.

Theorem 1.6. *For every (positive) integer valued function $i = i(n)$ such that $i(n) \in \{1, \ldots, \lfloor 2\log_2 n \rfloor\}$ for every n, there exist functions $p_i = p_i(n) \in [0,1]$ and $q_i = q_i(n) \in [0,1]$ such that: If $D \in \mathcal{D}(n,p)$ with $1/n \leq p = p(n) \leq 0.5$, then a.a.s. the following holds:*
(a) *if $p \geq p_i + q_i$, then $mat(D) \geq i$.*
(b) *if $p \leq p_i - q_i$, then $mat(D) < i$.*
Also, if $i = i(n) \to \infty$ is a growing function of n, then the threshold $p_i(n)$ is a sharp threshold in the sense that $q_i(n) = o(p_i(n))$.

The proof of Theorem 1.7 (see Subsection 1.2) suggests a correspondence between cliques in arbitrary undirected graphs and acyclic tournaments in specific orientations of these graphs. A quantitative statement of this relationship can be obtained when random graphs are compared to random digraphs. See Lemma 9.1 of Subsection 9.1 for the statement and its proof.

Outline: The presentation of the results is organized as follows: In Section 2, we provide the proof of Theorem 1.2. In Section 3, the proofs of Theorem 1.3, Corollary 1.4, Theorem 1.5 and Theorem 1.6 are presented. The proofs of the Theorems 1.2 and 1.3 are based on the Second Moment Method.

1.2. Algorithmic aspects

By MAT(D,k), we denote the following computational problem: Given a simple directed graph $D = (V, A)$ and k, determine if $mat(D) \geq k$. By MAT(D), we denote its optimization version. That is, given D, find an induced acyclic tournament of maximum size.

The MAT(D, k) problem, when D is a tournament, is the complement of the (Directed) Feedback Vertex Set problem, in which a minimum subset of vertices has to be removed to make the remaining digraph acyclic. These problems come up in various algorithms in computer science, such as in proving partial correctness of programs [9], in deadlock recovery in operating systems [6], and in VLSI design. They have been widely studied by approximation algorithmists also, e.g. [18]. As such, it is a natural generalization to consider MAT(D, k) for arbitrary digraphs, and is of importance in algorithm design. However, MAT(D, k) is known to be NP-complete [11], even if D is restricted to be a tournament [21]. Also, MAT(D) is known to be hard to approximate [17] when the input is an arbitrary digraph: For some $\epsilon > 0$, a polynomial-time approximation algorithm with an approximation ratio of $O(n^{\epsilon})$ is not possible unless $P = NP$.

Below we strengthen both of these results as follows. We show that MAT(D) is hard and inapproximable even when D is restricted to be acyclic (a dag), as shown in Theorem 1.7. The proof is given in the Appendix.

Theorem 1.7. MAT(D, k) *is NP-complete when D is restricted to be acyclic. Also, for every fixed $\epsilon > 0$, the optimization problem* MAT(D) *is not efficiently approximable with an approximation ratio of $n^{1-\epsilon}$, even if D is restricted to be acyclic, unless for every problem in NP there is a probabilistic algorithm that runs in expected polynomial time, and never makes an error* (*i.e., only the running time is stochastic*)[2].

Therefore, it seems hopeless to find polynomial-time algorithms for MAT(D, k), even if we allow randomized or approximation algorithms. However, the *average* case version of the problem—finding $mat(D)$ for a random digraph D—offers some hope. In this version, we seek to design efficient algorithms for computing an optimal solution which succeed a.a.s. over a random digraph. We use the model $\mathcal{D}(n, p)$ defined before for studying random digraphs.

We show (see Theorem 4.1) that a.a.s. *every* maximal induced acyclic tournament is of size which is at least nearly half of the optimal size. Hence any greedy heuristic obtains a solution whose approximation ratio is a.a.s. 2. This is similar to the case of cliques in undirected random graphs (see e.g. [4]).

We also study another heuristic which combines greedy and brute-force approaches as follows. We first apply the greedy heuristic to get a partial solution whose size is nearly $\log_r n - c\sqrt{\log_r n}$ for some arbitrary constant c. Amongst the remaining vertices, let C be the set of vertices such that each vertex in C can be individually and "safely" added to the partial solution. Then, in the subgraph

[2]This hypothesis is known as $NP \neq ZPP$. The faith in this hypothesis is almost as strong as that in $NP \neq P$ [12].

induced by C we find an optimal solution by brute-force and combine it with the partial solution. It is shown in Theorem 5.1 that this modified approach produces a solution whose size is at least $\log_r n + c\sqrt{\log_r n}$. This results in an additive improvement of $\Theta(\sqrt{\log_r n})$ over the simple greedy approach. The improvement is due to the fact we stop using greedy heuristic at a point where it is possible to apply brute-force efficiently. This approach is similar to (and was motivated by) the one used in [15] for finding large independent sets in $\mathcal{G}(n, 1/2)$.

As a consequence, we see that the problem of finding an optimal induced acyclic tournament can be approximated within a ratio of $2 - o(1)$ a.a.s. for random digraphs. This is in sharp contrast to the worst-case version where, by Theorem 1.7, it is very unlikely to be approximable even with a large multiplicative ratio.

Outline: The presentation of the algorithmic results is as follows. Theorem 4.1 is stated and proved in Section 4. Theorem 5.1 is stated and proved in Section 5.

1.3. Non-simple random digraphs

Each of the concentration and algorithmic results mentioned before also carry over (with some slight changes) to a related random model $\mathcal{D}_2(n, p)$ where we allow 2-cycles to be present and each of the potential arcs is chosen independently. These are presented in Section 6. In Section 7, we present some observations on the concentration of the maximum size of an induced tournament (not necessarily acyclic) for the two models of random directed graphs. Finally, in Section 8, we conclude with a summary and some open problems.

Throughout, we use standard notation. $\Re^+$ denotes the positive real numbers, $\mathcal{N}$ denotes the set of natural numbers. We often use the short notations $p = p(n)$, $w = w(n)$ to denote functions (real or integer valued) over $\mathcal{N}$. We also use standard notations like $O(\cdot)$, $\Omega(\cdot)$, $o(\cdot)$ and $\omega(\cdot)$ with usual meanings.

2. Analysis of $\mathcal{D}(n, p)$

Let U be any fixed subset of V of size b. The following two easy-to-verify claims play a role in the analysis. The proof of Claim 2.1 is provided in the Appendix.

Claim 2.1. *A directed acyclic graph $H = (U, A)$ has at most one (directed) Hamilton path.*

Claim 2.2. *For any $p = (n)$ with $p \le 1/2$,* $\mathbf{Pr}\,[D[U]$ *is an acyclic tournament*$]$ $= b!\, p^{\binom{b}{2}}$

Proof. Let $\mathcal{E}(U)$ denote the event that $D[U]$ is an acyclic tournament. Any acyclic tournament on U is characterized by a unique linear ordering $\sigma = (\sigma_1, \ldots,$

σ_b) of U with every ordered pair (σ_i, σ_j) $(i < j)$ of vertices joined by a forward arc $\sigma_i \to \sigma_j$. For any fixed linear ordering σ of U, let $\mathcal{E}(U, \sigma)$ denote the event that $D[U]$ is an acyclic tournament characterized by σ. Any such event $\mathcal{E}(U, \sigma)$ is the conjunction of $\binom{b}{2}$ identical and independent events, since the linear ordering σ forces each (i, j) edge to be present and also determines their orientation. Hence, we have

$$\mathbf{Pr}(\mathcal{E}(U, \sigma)) \;=\; p^{\binom{b}{2}}.$$

Now considering all $\sigma \in Perm(U)$, where $Perm(U)$ is set of all linear orderings of the elements of U, we get

$$\mathbf{Pr}(\mathcal{E}(U)) \;=\; \mathbf{Pr}\left(\bigcup\nolimits_{\sigma \in Perm(U)} \mathcal{E}(U, \sigma)\right).$$

Also, there are exactly $b!$ choices for σ, and these choices are mutually exclusive, since by Claim 2.1, the linear ordering σ is unique for a given acyclic tournament. Hence,

$$\mathbf{Pr}[D[U] \text{ is an acyclic tournament}] \;=\; \sum\nolimits_{\sigma} \mathbf{Pr}(\mathcal{E}(u, \sigma)) \;=\; b!\, p^{\binom{b}{2}}$$

which completes the proof. ∎

Before we proceed further, we introduce some notations which play an important role in the analysis. Define $\delta = \lceil d \rceil - d$. Then, it follows that

$$b^* = \begin{cases} d - 2 + \delta & \text{if } \delta > 1/2; \\ d - 1 + \delta & \text{if } \delta \le 1/2. \end{cases}$$

For a given b, let $m = \binom{n}{b}$ and let $(A_1, \ldots, A_m)$ denote any but fixed ordering of the set of all b-sized subsets of V. For $i \in [m]$, let X_i denote the indicator random variable whose value is 1 if $D[A_i]$ induces an acyclic tournament and is 0 otherwise. Let $X(b) = X(n, b)$ denote the number of induced acyclic tournaments of size b in D. Since there are $\binom{n}{b}$ sets of size b, it follows by Linearity of Expectation that

$$E[X(n, b)] = \sum\nolimits_i E[X_i] = \binom{n}{b}\, b!\, p^{\binom{b}{2}}.$$

We are only interested in the behavior of $E[X(n, b)]$ for $b \in [1, b^* + 2]$. From the definition of b^*, it follows that $b^* + 2 \le \lceil d \rceil + 1 \le \frac{2(\ln n)}{\ln r} + 3 \le 3(\ln n)$ for sufficiently large n since $p \le 1/2$. As a result, we have

$$(A) \qquad [1 - o(1)] \cdot f(n, p, b)^b \;\le\; E[X(n, b)] \;\le\; f(n, p, b)^b,$$

where $f : (\Re^+)^3 \to \Re^+$ such that $f(n, p, b) \;=\; n\, p^{(b-1)/2}$.

Setting $f(n, p, b) = 1$ and solving for b, we see that

$$f(n, p, b) > 1 \text{ if } b < d; \;\; f(n, p, d) = 1; \;\; f(n, p, b) < 1 \text{ if } b > d.$$

2.1. Proof of $mat(D) \leq b^* + 1$

First, we focus on proving the upper bound of Theorem 1.2. This is done by proving that

$$\mathbf{Pr}(X(b^*+2) > 0) \;\leq\; E[X(b^*+2)] \;=\; o(1).$$

Recall that b^* can be expressed in terms of d and δ in two different ways depending on the value of δ.

Case I. $\delta > 1/2$

$$\begin{aligned} E[X(b^*+2)] &= E[X(d+\delta)] \\ &\leq f(n,p,d+\delta)^{d+\delta} \\ &= \left(f(n,p,d)\cdot p^{\delta/2}\right)^{d+\delta} \\ &= p^{\delta(d+\delta)/2} \;=\; p^{\delta d/2}\cdot p^{\delta^2/2} \\ &= n^{-\delta}\cdot p^{\delta(1+\delta)/2} \;\leq\; n^{-\delta} \text{ (since } p\leq 1 \text{ and } \delta\geq 0) \\ &\leq n^{-1/2} \;=\; o(1). \end{aligned}$$

Case II. $\delta \leq 1/2$

$$\begin{aligned} E[X(b^*+2)] &= E[X(d+1+\delta)] \\ &\leq f(n,p,d+1+\delta)^{d+1+\delta} \\ &= \left(f(n,p,d)\cdot p^{(1+\delta)/2}\right)^{d+1+\delta} \\ &= p^{(1+\delta)(d+1+\delta)/2} \;=\; p^{(1+\delta)d/2}\cdot p^{(1+\delta)^2/2} \\ &= n^{-(1+\delta)}\cdot p^{(1+\delta)(2+\delta)/2} \;\leq\; n^{-(1+\delta)} \text{(since } p\leq 1 \text{ and } \delta\geq 0) \\ &\leq n^{-1} \;=\; o(1). \end{aligned}$$

This establishes the upper bound.

2.2. Proof of $mat(D) \geq b^*$

Next, we focus on proving the lower bound of Theorem 1.2. For this, we first show that $E[X(b^*)] \to \infty$ as $n \to \infty$.

Case I. $\delta > 1/2$

$$\begin{aligned} E[X(b^*)] &= E[X(d-2+\delta)] \\ &\geq [1-o(1)]\cdot f(n,p,d-2+\delta)^{d-2+\delta} \\ &= [1-o(1)]\cdot \left(f(n,p,d)\cdot p^{(-2+\delta)/2}\right)^{d-2+\delta} \\ &= [1-o(1)]\cdot p^{(-2+\delta)(d-2+\delta)/2} \;=\; p^{(-2+\delta)d/2}\cdot p^{(-2+\delta)^2/2} \end{aligned}$$

$$
\begin{aligned}
&= [1 - o(1)] \cdot n^{2-\delta} \cdot p^{(2-\delta)(1-\delta)/2} \\
&\geq [1 - o(1)] \cdot n^{2-\delta} \cdot p^{3/8} \ \ (\text{since } p \leq 1 \ \text{ and } \delta > 1/2) \\
&\geq n^{1/2} \ \rightarrow \ \infty \quad \text{as} \quad n \ \rightarrow \ \infty.
\end{aligned}
$$

Case II. $\delta \leq 1/2$

$$
\begin{aligned}
E[X(b^*)] &= E[X(d-1+\delta)] \\
&\geq \ [1 - o(1)] \cdot f(n,p,d-1+\delta)^{d-1+\delta} \\
&= [1 - o(1)] \cdot \Big(f(n,p,d) \cdot p^{(-1+\delta)/2} \Big)^{d-1+\delta} \\
&= [1 - o(1)] \cdot p^{(-1+\delta)(d-1+\delta)/2} \ = \ p^{(-1+\delta)d/2} \cdot p^{(-1+\delta)^2/2} \\
&= [1 - o(1)] \cdot n^{1-\delta} \cdot p^{(-1+\delta)\delta/2} \\
&\geq [1 - o(1)] \cdot n^{1-\delta} (\text{since } p \leq 1 \ \text{ and } \delta \leq 1/2) \\
&\geq n^{1/2} \ \rightarrow \ \infty \quad \text{as} \quad n \ \rightarrow \ \infty.
\end{aligned}
$$

For the sake of notational simplicity, we use X to denote $X(b^*)$ and use b to denote b^* for the rest of this section. Now, we need to show that $X > 0$ with high probability. We use the well-known Second Moment Method to establish this. Let $Var(X)$ denote the variance of X.

Recall that X_i denotes the indicator random variable for the i-th b-size subset of V. Using standard arguments (see [3]), it can be seen that

$$
Var(X) \leq E[X] + \sum\nolimits_{i \neq j} COV(X_i, X_j), \tag{1}
$$

where the second sum is over ordered pairs and $COV(X_i, X_j) = E(X_i X_j) - E(X_i)E(X_j)$ is the covariance between X_i and X_j. Note that X_i and X_j are independent whenever $|A_i \cap A_j| \leq 1$ and in that case $COV(X_i, X_j) = 0$. Otherwise, with $|A_i \cap A_j| = l$, we have

$$
\begin{aligned}
COV(X_i, X_j) &\leq E(X_i X_j) = E(X_i)E(X_j | X_i = 1) \\
&= b! p^{\binom{b}{2}} \cdot (b!/l!) \cdot p^{\binom{b}{2} - \binom{l}{2}},
\end{aligned} \tag{2}
$$

where the last equality follows from Claim 2.2. Also, for any fixed i, the number of b-sized subsets A_j such that $|A_i \cap A_j| = l$ is exactly $\binom{b}{l}\binom{n-b}{b-l}$. As a result,

$$
\begin{aligned}
\sum\nolimits_{i\neq j} COV(X_i,X_j) &= \sum\nolimits_i \sum\nolimits_{j\,:\,2\le|A_i\cap A_j|\le b-1} COV(X_i,X_j)\\
&\le \sum\nolimits_i b!p^{\binom{b}{2}}\cdot\left(\sum\nolimits_{2\le l\le b-1}\binom{b}{l}\binom{n-b}{b-l}\left(\frac{b!}{l!}\right)\cdot p^{\binom{b}{2}-\binom{l}{2}}\right)\\
&= E[X]\cdot\left(\sum\nolimits_{2\le l\le b-1}\binom{b}{l}\binom{n-b}{b-l}\left(\frac{b!}{l!}\right)\cdot p^{\binom{b}{2}-\binom{l}{2}}\right) \qquad (3)\\
&= E[X]^2\cdot\left(\sum\nolimits_{2\le l\le b-1}\frac{(b)_l}{(l!)^2}\binom{n-b}{b-l}\binom{n}{b}^{-1}\cdot p^{-\binom{l}{2}}\right)\\
&= E[X]^2\cdot M,
\end{aligned}
$$

where $M = M(n,p,b)$ is as defined above. Applying Chebyshev's Inequality and (1), it follows that

$$
\begin{aligned}
\mathbf{Pr}[X=0] &\le Var(X)(E[X])^{-2}\\
&\le \left(E[X]+\sum\nolimits_{i\neq j} COV(X_i,X_j)\right)(E[X])^{-2}. \qquad (4)
\end{aligned}
$$

Combining (4) and (3), we notice that

$$
\mathbf{Pr}(X=0) \le (E[X])^{-1} + M \;\; = \;\; o(1) \qquad (5)
$$

provided $M = M(n,b) = o(1)$ since it has already been shown that $E[X]\to\infty$. Thus, we only need to show that $M = o(1)$ to complete the arguments.

Now, we focus on showing that $M = o(1)$. Notice that

$$
\begin{aligned}
M &= \sum\nolimits_{2\le l\le b-1}\frac{(b)_l}{(l!)^2}\cdot\binom{n-b}{b-l}\cdot\binom{n}{b}^{-1}\cdot p^{-\binom{l}{2}}\\
&\le \sum\nolimits_{2\le l\le b-1}\binom{b}{l}^2\cdot\frac{1}{(n)_l}\cdot p^{-\binom{l}{2}}\\
&= \sum\nolimits_{2\le l\le b-1}\binom{b}{l}^2\cdot\frac{1+o(1)}{n^l}\cdot p^{-\binom{l}{2}}\\
&= (1+o(1))\sum_{2\le l\le b-1}\binom{b}{l}^2 p^{l(d-l)/2}\\
&= (1+o(1))\sum\nolimits_{2\le l\le b-1} F_l,
\end{aligned}
$$

where the last-but-one equality follows using $f(n,p,d) = 1$.

Let t_l be the ratio between successive terms: $t_l = F_{l+1}/F_l$. Now take the ratio of ratios: $s_l = t_{l+1}/t_l$.

$$t_l = F_{l+1}/F_l = \frac{\binom{b}{l+1}^2 p^{(l+1)(d-l-1)/2}}{\binom{b}{l}^2 p^{l(d-l)/2}}, = \left(\frac{b-l}{l+1}\right)^2 p^{-l+(d-1)/2}$$

$$s_l = \left(\left(\frac{b-l-1}{b-l}\right)\left(\frac{l+1}{l+2}\right)\right)^2 p^{-1}.$$

First we state the following easy-to-prove fact regarding any sequence of positive real numbers.

Observation 2.3. *For a sequence of positive real numbers $a_1, \ldots, a_n$, if $s_i = a_{i+2}a_i/a_{i+1}^2 \geq 1$ for all $1 \leq i \leq n-2$, then for all $i \in [n]$, we have $a_i \leq \max\{a_1, a_n\}$.*

Claim 2.4. (i) *If $p \leq 1/4$, then $s_l \geq 1$ for every b with $2 \leq l \leq b-3$.*
(ii) *If $p > 1/4$, then $s_l \geq 1$ for every b with $2 \leq l \leq b-4$ and also $t_{b-2} > 1$.*

From the above (Observation 2.3 and Claim 2.4), the proof of Theorem 1.2 follows easily, as we get that for all l: $2 \leq l \leq b-1$, for all $p \geq 1/n$, $F_l \leq \max\{F_2, F_{b-1}\}$. Now $F_2 = \binom{2}{2}^2 p^{2(d-2)/2} = p^{2\left(\frac{d-1}{2}-\frac{1}{2}\right)} = \frac{p^{-1}}{n^2} = O(1/n)$ and $F_{b-1} = \binom{b}{b-1}^2 p^{(b-1)(d-b+1)/2} \leq b^2\left(p^{(b-1)/2}\right) = b^2\left(\frac{p^{(b-d)/2}}{n}\right) \leq b^2\left(\frac{r^{(d-b)/2}}{n}\right) \leq b^2\left(\frac{r^{3/4}}{n}\right) = O\left(\frac{(\ln n)^2}{n^{1/4}}\right)$. Therefore, $M = (1+o(1)) \cdot \sum_{l=2}^{b-1} F_l = O\left(\frac{(\ln n)^3}{n^{1/4}}\right) = o(1)$.

Proof of Claim 2.4. *Case* (i). Assume that $p \leq 1/4$ and $2 \leq l \leq b-3$. Then, we have $(b-l-1)/(b-l) \geq 2/3$ and $(l+1)/(l+2) \geq 3/4$. This implies that $s_l \geq p^{-1}/4 \geq 1$.

Case (ii). Assume that $p > 1/4$ and $l \leq b-4$. It can be verified that the square term in s_l is at least 1/2 and $p^{-1} \geq 2$, so $s_l \geq 1$. Now $t_{b-2} = (2/(b-1))^2 p^{-(b-2)+(d-1)/2} \geq (4r^b p^2)/(nb^2) \geq \frac{4np^{2.5}}{b^2} \to \infty$, using our assumption that $p > 1/4$. ∎

We have thus completely established that $M = o(1)$ for all $p \geq 1/n$, thereby establishing that $\mathbf{Pr}(X = 0) = o(1)$. Hence, a.a.s., $mat(D) \in \{b^*, b^*+1\}$ for the stated range of p. The remaining parts of Theorem 1.2 are straightforward to derive and are given in Subsection 9.4 for the sake of completeness. This completes the proof of Theorem 1.2.

3. One-point Concentration and Threshold Results

Recall the definition of d, δ from the proof of Theorem 1.2. The proof of Theorem 1.3 proceeds by considering the following 2 cases:

Case I. $0 < w/\ln n \leq \delta \leq 1/2$. In this case, $\lfloor d \rfloor = b^*$. From Theorem 1.2, it only remains to show that $\mathbf{Pr}[mat(D) \geq b^* + 1] \to 0$ as $n \to \infty$. We again use the first moment method to show that $E[X(b^*+1)] = o(1)$.

By our assumption about p, $\delta \leq 1/2$. Hence, by definition, $b^* + 1 = d + \delta$. Thus,

$$\begin{aligned} E[X(b^*+1)] &\leq f(n,p,d+\delta)^{d+\delta} = (p^{\delta/2})^{d+\delta} = p^{\delta(d+\delta)/2} = n^{-\delta} \cdot p^{\delta(1+\delta)/2} \\ &\leq n^{-\delta} \ \leq \ n^{-w/\ln n} = e^{-w} \ \to \ 0 \quad \text{as } n \to \infty. \end{aligned}$$

Case II. $1/2 < \delta \leq 1 - w/\ln n < 1$. Here, $\lfloor d \rfloor = b^* + 1$. The proof proceeds by verifying that $\mathbf{Pr}[X(b^*+1) = 0] = o(1)$, and hence, $mat(D) \geq b^* + 1$ a.a.s. Together with the upper bound on $mat(D)$ when $p \geq 1/n$ in Theorem 1.2, this gives the desired result. Briefly, this can be seen as follows. From (5), it suffices to show that

(i) $E[X(n, b^*+1)] \to \infty$ as $n \to \infty$, and

(ii) $M = M(n,p,b^*+1) = o(1)$.

To prove (i), we notice that

$$\begin{aligned} E[X(b^*+1)] &\geq [1-o(1)] \cdot f(n,p,d+\delta-1)^{d+\delta-1} \\ &= [1-o(1)] \cdot (p^{(\delta-1)/2})^{d+\delta-1} = [1-o(1)] \cdot p^{(\delta-1)(d+\delta-1)/2} \\ &= [1-o(1)] \cdot n^{1-\delta} \cdot p^{-\delta(1-\delta)/2} \geq [1-o(1)] \cdot n^{w/\ln n} \\ &= [1-o(1)] \cdot e^{w} \ \to \ \infty \quad \text{as } n \to \infty. \end{aligned}$$

To prove (ii), we need to go along the proof of Theorem 1.2, and evaluate $M(n,p,b^*+1)$.

An easy check reveals that $M(n,p,b^*+1) = M(n,p,b^*) \cdot O((\ln n)^3) + T_{b^*}$, where

$$T_{b^*} = \binom{b^*+1}{b^*}^2 \left((n)_{b^*} p^{\binom{b^*}{2}} \right)^{-1} \ \leq \ 2\binom{b^*+1}{b^*}^2 \left(np^{(b^*-1)/2} \right)^{-b^*}.$$

Now, from the proof of Theorem 1.2, we have that $M(n,p,b^*) = O\left(\frac{(\ln n)^3}{n^{1/4}}\right)$. Therefore $M(n,p,b^*+1) = O\left(\frac{(\ln n)^6}{n^{1/4}}\right) + T_{b^*}$.

Next, using the definition of b^* when $\delta > 1/2$, we have

$$\begin{aligned} T_{b^*} &\le 2(b^*+1)^2(np^{(b^*-1)/2})^{-b^*} = 2(b^*+1)^2(np^{(d-3+\delta)/2})^{-b^*} \\ &= 2(b^*+1)^2(p^{-1+\delta/2})^{-b^*} = 2(b^*+1)^2 p^{b^*(1-\delta/2)} = o(1). \end{aligned}$$

Thus it is verified that both $M \cdot O((\ln n)^3)$ and T_{b^*}, and hence their sum, are $o(1)$.

3.1. Proof of Corollary 1.4

Let p be fixed but arbitrary. It follows from the definitions of d and δ, that for every n, we have $n = r^{\frac{k-(1+\delta)}{2}}$ for some nonnegative integer k. Also, it follows from Theorem 1.3 that for every sufficiently large n, $mat(D)$ is concentrated on one value if $\frac{w}{\ln n} \le \delta \le 1 - \frac{w}{\ln n}$. Hence, for every such n, we must have

$$r^{\frac{k-2}{2}+\frac{w}{2\ln n}} \le n \le r^{\frac{k-1}{2}-\frac{w}{2\ln n}}.$$

For every $k \ge 2$, we define two values as follows.

$$m_{k,l} = \min\left\{n : n \ge r^{\frac{k-2}{2}+\frac{w}{2\ln n}}\right\};\ \ m_{k,h} = \max\left\{n : n \le r^{\frac{k-1}{2}-\frac{w}{2\ln n}}\right\}.$$

It follows that $mat(D)$ is just one value for every *sufficiently large* $n \in R$ where $R = \bigcup_{k\ge 2} R_k$ and $R_k = \{n : m_{k,l} \le n \le m_{k,h}\}$. Hence it suffices to show that R is a subset of density 1 of the set $\mathcal{N}$ of positive integers. Now, $\mathcal{N} - R = \bigcup_{k\ge 3} S_k$ where $S_k = \{n \in \mathcal{N} : m_{k-1,h} < n < m_{k,l}\}$.

For every $k \ge 3$,

$$|R_k| \approx r^{\frac{k-2}{2}}\left(r^{\frac{1}{2}-\frac{w}{2\ln m_{k,h}}} - r^{\frac{w}{2\ln m_{k,l}}}\right) \text{ and } |S_k| \approx r^{\frac{k-2}{2}}\left(r^{\frac{w}{2\ln m_{k,l}}} - r^{-\frac{w}{2\ln m_{k-1,h}}}\right).$$

By choosing w suitably, we can ensure that $w/\ln n \to 0$ as $n \to \infty$. Also, $m_{k,h}$ and $m_{k,l}$ grow exponentially in k. Hence, for every sufficiently large k,

$$|R_k| \approx r^{\frac{k-2}{2}}\left(r^{\frac{1}{2}} - 1\right) \text{ and } |S_k| = O\left(r^{\frac{k-2}{2}}\left(\frac{w(r^{k/2})}{k\ln t}\right)\right) = o\left(r^{\frac{k-2}{2}}\right).$$

Thus, for all sufficiently large k, we have $\sum_{j\le k+1} |S_j| = O(|S_{k+1}|) = o(|R_k|)$. As a result, we have

$$\frac{|R\cap[n]|}{n} \to 1 \text{ as } n \to \infty.$$

This shows that R has density 1 as a subset of $\mathcal{N}$. This completes the proof of Corollary 1.4.

3.2. Proof of Theorem 1.5

The proof is based on an application of Lovász Local Lemma stated below.

Lemma 3.1. *Let $\mathcal{A} = \{E_1, E_2, \ldots, E_m\}$ be a collection of events over a probability space such that each E_i is totally independent of all but the events in $\mathcal{D}_i \subseteq \mathcal{A} \backslash \{E_i\}$. If there exists a real sequence $\{x_i\}_{i=1}^m$, $x_i \in [0,1)$, such that*

$$\text{For all } i \in [m],\ \mathbf{Pr}[E_i] \le x_i \prod_{j:E_j \in \mathcal{D}_i} (1-x_j), \text{ then}$$
$$\mathbf{Pr}\left[\bigcap_{i=1}^m \overline{E}_i\right] \ge \prod_{i=1}^m (1-x_i) > 0.$$

Hence, with positive probability, none of the events occur.

First, notice that for the given value of p, $d = (j + c/(\ln n))$, $b := b^* = j-1$ and $\delta = 1 - c/\ln n > 1/2$, for sufficiently large n. By Theorem 1.2, we know that for the given probability $p = p(n)$, $mat(D) \in \{b, b+1\}$ a.a.s. Therefore to prove Theorem 1.5, it suffices to show that there exist constants $0 < c_1 \le c_2 < 1$ such that $c_1 \le \mathbf{Pr}(mat(D) = b+1) \le c_2$ for all sufficiently large n. This is proved below. For various symbols like, d, δ and b^*, we use the same meanings used in the proof of Theorem 1.2.

Consider the expected number of acyclic tournaments of size $b+1$:

$$\begin{aligned}
E[X_{b+1}] &\approx \left(np^{b/2}\right)^{b+1} = \left(np^{(d+\delta-2)/2}\right)^{b+1} = \left(p^{(-1+\delta)/2}\right)^{b+1} \\
&= \left(p^{-c/2(\ln n)}\right)^{d-1+\delta} = \left(r^{c/2(\ln n)}\right)^{d-1+\delta} = \left(e^{c/2(\log_r n)}\right)^{d-1+\delta} \\
&= (e^c)^{1+(\delta/2(\log_r n))} \quad \approx \quad e^{c'},
\end{aligned}$$

for some constant $c' > 0$. If the expectation had been a constant less than 1, a simple application of Markov's inequality would have established the upper bound on the probability.

Case I. Proof of $\mathbf{Pr}(mat(D) = b+1) \le c_2$. We apply Lovász Local Lemma 3.1 to prove this claim. For every i, $1 \le i \le N := \binom{n}{b+1}$, define E_i to be the event that A_i induces an acyclic tournament, where A_i is the i-th $(b+1)$-set in some fixed ordering of all $(b+1)$-subsets of V. For every i, $\mathbf{Pr}(E_i) = q := (b+1)!p^{\binom{b+1}{2}} = o(1)$. Choose $x_i = x = 25q$ for each i. Construct the dependency graph on N events by joining E_i and E_j if $|A_i \cap A_j| \ge 2$. It can be seen that each E_i is totally independent of all other E_j's which are not adjacent to E_i. Note that the dependency graph is regular with the uniform degree of any E_i being given by $deg(E_i) = \sum_{2 \le k \le b} \binom{b+1}{k}\binom{n-b-1}{b+1-k}$. It is easy to see that $deg(E_i) \le \binom{b+1}{2}\binom{n-2}{b-1} \approx Y$ where $Y := N(b^4/2n^2)$. Using $q = o(1)$ and $Nq \approx e^{c'}$, it follows that $Yq = Nqb^4/(2n^2) = o(1)$ and also that $\ln(1-25q) = -25q[1+o(1)]$. To apply the Local Lemma, it suffices to prove that

$$q \;\leq\; 25q(1-25q)^Y.$$

Equivalently, it suffices to prove that

$$1 \leq 25e^{Y(\ln(1-25q))} \;=\; 25e^{-25Yq[1+o(1)]} = 25[1-o(1)].$$

The above clearly holds true. Now applying 3.1, one gets that

$$\mathbf{Pr}[\cap_i \overline{E_i}] \geq \prod_{i=1}^{N}(1-x) \;=\; (1-25q)^N \;\approx\; e^{-25Nq} \geq e^{-25e^{c'}}.$$

Therefore, $\mathbf{Pr}(mat(D) = b+1) \leq \mathbf{Pr}[mat(D) \geq b+1] = \mathbf{Pr}[\cup_i E_i] \leq c_2$, where $c_2 := 1 - e^{-25e^{c'}}$.

Case II. Proof of $\mathbf{Pr}(mat(D) = b+1) \geq c_1$. To prove this, we use the following version of Paley-Zygmund Inequality (see [10])

$$\mathbf{Pr}[X_{b+1} > 0] \geq E[X_{b+1}]^2/E[X_{b+1}^2]. \tag{6}$$

Notice that the RHS of the previous inequality (6) is exactly $1/(1+z)$, where $z = Var(X_{b+1})/E[X_{b+1}]^2$. As in the proof of Theorem 1.3, $z \leq E[X_{b+1}]^{-1} + M(n,p,b+1)$, and $M(n,p,b+1) \leq M(n,p,b) \cdot (\ln n)^3 + T_b$. Now, $M(n,p,b) = O\left(\frac{(\ln n)^6}{n^{1/4}}\right) = o(1)$, and it was shown that $T_b = o(1)$. Therefore, we get that $z \leq e^{-c'} + o(1) \approx e^{-c'}$, and therefore $1/(1+z)$ in (6) is at least c_1' where c_1 is the constant defined by $c_1 = 1/(1+e^{-c'})$. This proves that $\mathbf{Pr}(mat(D) \geq b+1) \geq c_1$. As a result, $\mathbf{Pr}(mat(D) = b+1) = \mathbf{Pr}(mat(D) \geq b+1) - \mathbf{Pr}(mat(D) \geq b+2) \geq c_1 - o(1) \approx c_1$.

Hence there exist constants $c_1, c_2 \in (0,1)$ such that,

$$c_1 \leq \mathbf{Pr}[mat(D) = b+1] \leq c_2, \quad \text{and hence}$$
$$1 - o(1) - c_2 \leq \mathbf{Pr}[mat(D) = b] \leq 1 - o(1) - c_1.$$

Thus, $mat(D)$ is **not** concentrated at any *single* point.

3.3. Proof of Theorem 1.6

First, we prove the following lemma, from which the theorem follows as an easy consequence.

Lemma 3.2. *Let $i = i(n) \in \{1, \ldots, \lfloor 2\log_2 n \rfloor\}$ be any fixed function of n. Let $D \in \mathcal{D}(n,p)$ and let $w = w(n)$ be any function of n so that as $n \to \infty$, $w \to \infty$ and $w \leq (0.5 \ln n)$. Then, asymptotically almost surely, the following are true:*

(i) *If* $p \geq n^{-2/\left(i-1+\frac{w}{\ln n}\right)}$, *then* $mat(D) \geq i$.

(ii) *If* $p \leq n^{-2/\left(i-1-\frac{w}{\ln n}\right)}$, *then* $mat(D) < i$.

Proof. The probability of having an induced acyclic tournament of size i only increases with increasing p. From the one-point concentration result of Theorem 1.3, it follows that if p is such that d (defined before) satisfies $d \geq i + \frac{w}{\ln n}$, then a.a.s. $mat(D) \geq i$. Similarly, if p is such that d satisfies $d \leq i - \frac{w}{\ln n}$, then a.a.s. $mat(D) < i$. However,

$$\begin{aligned} d \geq i + \frac{w}{\ln n} \; &\Leftrightarrow \; \log_r n \geq \frac{i-1}{2} + \frac{w}{2\ln n} \\ &\Leftrightarrow \; n \geq p^{-\left(\frac{i-1}{2}+\frac{w}{2\ln n}\right)} \\ &\Leftrightarrow \; p \geq n^{-2/\left(i-1+\frac{w}{\ln n}\right)}. \end{aligned}$$

Similarly, we have

$$d \leq i - \frac{w}{\ln n} \Leftrightarrow p \leq n^{-2/\left(i-1-\frac{w}{\ln n}\right)}.$$

This completes the proof of the lemma. ∎

From the above lemma, Theorem 1.6 can be proved as follows. We choose $w(n) = \sqrt{\ln n}$ and it satisfies the conditions of the lemma above. We set $lb_i(n) = n^{-2/\left(i-1-\frac{w}{\ln n}\right)}$ and $ub_i(n) = n^{-2/\left(i-1+\frac{w}{\ln n}\right)}$, and define $p_i(n) = (ub_i(n)+lb_i(n))/2$, and $q_i(n) = (ub_i(n) - lb_i(n))/2$.

If $i(n) \to \infty$, we choose $w(n) = i(n)/4$ so that $w(n) \to \infty$. Also, it can be verified that $lb_i(n) = ub_i(n)[1 - o(1)]$ and hence $q_i(n) = o(p_i(n))$, so we have a sharp threshold for such $i = i(n)$.

Remark. In the above proof, notice that the ratio $\frac{q_i}{p_i} \leq \frac{ub_i - lb_i}{ub_i}$ which is

$$1 - e^{-\frac{w}{(i-1)^2-(w/\ln n)^2}} = 1 - \left(1 - O\left(\frac{w}{(i-1)^2 - (w/\ln n)^2}\right)\right) = O(1/i)$$

for $w = i/4$.

4. Finding an Induced Acyclic Tournament

In this section, we obtain a lower bound (see Theorem 4.1 below) on the size of any maximal induced acyclic tournament. As a consequence, any simple heuristic which builds a maximal solution of a given random digraph, a.a.s. produces an acyclic tournament of size within a muliplicative factor ($\approx 1/2$) of the optimal.

In what follows, we assume that $p \geq n^{-1/4}$ mainly to focus on the interesting range of p. If p is smaller, then $mat(D) \leq 9$ almost surely and one can find provably optimal solutions in polynomial time.

Theorem 4.1. *Given $D \in \mathcal{D}(n,p)$ with $p \geq n^{-1/4}$ and any $w = w(n)$ such that $w(n) \to \infty$ as $n \to \infty$, with probability $1 - o(1)$, every maximal induced acyclic tournament is of size at least $\lceil \delta \log_r n \rceil$, where $\delta = 1 - \frac{\ln(\ln np + w)}{\ln n}$.*

Proof. Without loss of generality, we assume that $p \geq n^{-1/4}$ so that $\log_r n \geq 4$. Hence $d = \delta(\log_r n) > 3$ and $\lfloor d \rfloor \geq 3$.

For any induced acyclic tournament $D[A]$ of size $|A| = b$, $b \leq d = \delta(\log_r n)$, and any vertex $u \in V \setminus A$, the probability that u can be added to A is given by

$$\mathbf{Pr}[D[A \cup \{u\}] \text{ is an acyclic tournament}] = (b+1)p^b.$$

The above equality is true since $D[A \cup \{u\}]$ induces an acyclic tournament if and only if u can be added to any of the $b+1$ positions in the unique Hamilton path of $D[A]$ in such a way that each of the edges joining u with vertices in A is present and is oriented in the proper direction. Also, this probability decreases with increasing b.

This event depends only on the edges joining u with the vertices in A, and hence, is independent of events corresponding to other vertices in $V \setminus A$. Therefore, the probability that $D[A]$ is a maximal acyclic tournament is given by

$$\begin{aligned}\mathbf{Pr}(D[A]) \text{ is maximal } &= \mathbf{Pr}[\forall u \in V \setminus A,\ u \text{ cannot be added to } A] \\ &= \left(1 - (b+1)p^b\right)^{n-b}.\end{aligned}$$

As b increases, this probability increases and hence achieves its maximum (for $b \leq d$) at $b = \lfloor d \rfloor$. Hence, for an induced acyclic tournament $D[A]$ of size $\lfloor d \rfloor$, we have (using $(d+1)(n-d) \geq nd$):

$$\begin{aligned}\mathbf{Pr}(D[A] \text{ is maximal}) &\leq \left(1 - (\lfloor d \rfloor + 1)p^{\frac{\delta(\ln n)}{\ln r}}\right)^{n - \lfloor d \rfloor} \leq \left(1 - \frac{d+1}{n^{\delta}}\right)^{n-d} \\ &\leq e^{-\frac{(d+1)(n-d)}{n^{\delta}}} \leq e^{-dn^{1-\delta}}.\end{aligned}$$

For any *fixed* set A of size $b \leq d$, let $\mathcal{E}(A)$ denote the event that $D[A]$ is a maximal induced acyclic tournament.

$$\mathbf{Pr}(\mathcal{E}(A)) \quad \leq \quad b! p^{\binom{b}{2}} e^{-dn^{1-\delta}}.$$

Thus,

$$\begin{aligned}\mathbf{Pr}(\exists A, |A| = b : \mathcal{E}(A)) &\leq \binom{n}{b} b! p^{\binom{b}{2}} e^{-dn^{1-\delta}} \\ &\leq \left(np^{(b-1)/2}\right)^b e^{-dn^{1-\delta}} = (f(n,p,b))^b e^{-dn^{1-\delta}},\end{aligned} \tag{7}$$

where we recall that $f(n,p,b) = np^{(b-1)/2}$.

Note that for each $b \le \lfloor d \rfloor$,

$$\frac{f(n,p,b+1)^{b+1}}{(f(n,p,b))^b} = f(n,p,b+1)p^{b/2} = np^b \ge np^d = n^{1-\delta} = (\ln np) + w.$$

Hence $\sum_{b \le d}(f(n,p,b))^b \ \le\ (f(n,p,\lfloor d \rfloor))^{\lfloor d \rfloor} \sum_{b \le \lfloor d \rfloor} (\ln np + w)^{-(\lfloor d \rfloor - b)}$ $= 2(f(n,p,\lfloor d \rfloor))^{\lfloor d \rfloor}$. As a result, taking the union bound over all choices of A, we see that (using $\lfloor d \rfloor \ge 3$)

$$\mathbf{Pr}(\exists A, |A| \le \lfloor d \rfloor : \mathcal{E}(A)) \le 2\,(f(n,p,\lfloor d \rfloor))^{\lfloor d \rfloor}\, e^{-dn^{1-\delta}} \ \le\ 2\,(np)^{\lfloor d \rfloor}\, e^{-dn^{1-\delta}}.$$

For $\delta = 1 - \frac{\ln(\ln np + w)}{\ln n}$ this probability is

$$\mathbf{Pr}[\exists A,\ |A| \le d : \mathcal{E}(A)] \le 2 \cdot e^{d(\ln np - (\ln np + w))} = 2 \cdot e^{-dw} \le 2e^{-w} \to 0 \text{ as } n \to \infty.$$

Hence, every maximal induced acyclic tournament is of size at least $\lceil d \rceil$. ■

5. Another Efficient Heuristic with Improved Guarantee

We present below another efficient heuristic which will be analyzed and be shown to have an additive improvement of $\Theta\left(\sqrt{\log_r n}\right)$ over the guarantee given (in Section 4) on the size of any maximal solution. It is similar to a heuristic presented in [15] for finding large independent sets in $G \in \mathcal{G}(n, 1/2)$. We show that, for every large constant $c > 0$, one can find in polynomial time an acyclic tournament of size at least $\lfloor \log_r n + c\sqrt{\log_r n} \rfloor$.

The idea is to construct greedily a solution A of size $g(n,p,c) = \lceil \log_r n - c\sqrt{\log_r n} \rceil$ and then add an optimal solution (found by an exhaustive search) in the subgraph induced by those vertices each of which can be safely and individually added to A to get a bigger solution. We will show that exhaustive search can be done in polynomial time and yields (a.a.s.) a solution of size $2c\sqrt{\log_r n}$. As a result, we finally get a solution of the stated size. The algorithm is described below.

ACYTOUR$(D = (V,E), p, c)$

1. Choose and fix a linear ordering σ of V.
2. $c' = 1.2c$; $A = \emptyset$; $B = V$.
3. **while** $B \ne \emptyset$ and $|A| < g(n/2, p, c')$ **do**
4. Let u be the σ-smallest vertex in B.

5. **if** $D[A \cup \{u\}]$ induces an acyclic tournament **then** add u to A.

6. Remove u from B. **endwhile**

7. **if** $|A| < g(n/2, p, c')$ **or** $|B| < n/2$, **then** Return FAIL and halt.

8. $C = \{u \in B : \forall v \in A,\ v \to u \in E\}$; $r = p^{-1}$; $\mu = |B|p^{|A|}$.

9. **if** $|C| \notin [(0.9)\mu, (1.1)\mu]$ **then** Return FAIL.

10. **for each** $X \subset C :\ |X| = \lfloor 2c'\sqrt{\log_r n/2} \rfloor - 1$ **do**

11. **if** $D[X]$ is an acylic tournament **then** Return $D[A \cup X]$ and halt. **endfor**

12. Return FAIL.

We analyze the above algorithm and obtain the following result.

Theorem 5.1. *Let $D \in \mathcal{D}(n,p)$. For every sufficiently large constant $c \geq 1$, if p is such that $n^{-1/c^2} \leq p \leq 0.5$, then, with probability $1 - o(1)$, ACYTOUR(D) will output an induced acyclic tournament of size at least $b' = \lfloor (1+\epsilon') \log_r n \rfloor$, where $\epsilon' = c/\sqrt{\log_r n}$.*

Proof. Recall our assumption that c is sufficiently large.

Correctness. First, we prove the correctness. Note that $D[A]$ is always an induced acyclic tournament. Also, each $u \in C$ is such that $D[A \cup \{u\}]$ is an acyclic tournament with u as the unique sink vertex (having zero out-degree). Hence, any acyclic tournament $D[X]$ present as a subgraph in $D[C]$ can be safely added to A so that $D[A \cup X]$ also induces an acyclic tournament.

Analysis. Consider the following events defined as

Failure at step 7 :
$$\mathcal{E}_1 := |A| < g(n/2, p, c') \textbf{ or } |B| < n/2;$$
Failure at step 9 :
$$\mathcal{E}_2 := \overline{\mathcal{E}_1} \cap \mathcal{E}_2', \text{ where } \mathcal{E}_2' := |C| \notin [(0.9)\mu, (1.1)\mu];$$
Failure at step 12 :
$$\mathcal{E}_3 := \overline{\mathcal{E}_1} \cap \overline{\mathcal{E}_2} \cap \mathcal{E}_3', \text{ where}$$
$$\mathcal{E}_3' := mat(D[C]) < \lfloor 2c'\sqrt{\log_r n/2} \rfloor - 1.$$

If none of the events $\{\mathcal{E}_1, \mathcal{E}_2, \mathcal{E}_3\}$ holds, then the algorithm will succeed and output a solution whose size is

$$\begin{aligned} |A \cup X| &\geq \log_r(n/2) - c'\sqrt{\log_r(n/2)} + 2c'\sqrt{\log_r n/2} - 2 \\ &\geq (1+\epsilon')(\log_r n) + (c'-c)\sqrt{\log_r n/2} - 2.5 - \log_r 2 \\ &\geq (1+\epsilon')(\log_r n) + (0.2c)\sqrt{\log_r n/2} - 3.5 \\ &\geq (1+\epsilon')(\log_r n). \end{aligned}$$

The probability of this happening is

$$\mathbf{Pr}(\overline{\mathcal{E}_1} \cap \overline{\mathcal{E}_2} \cap \overline{\mathcal{E}_3}) = 1 - \mathbf{Pr}(\mathcal{E}_1 \cup \mathcal{E}_2 \cup \mathcal{E}_3).$$

The events $\mathcal{E}_1, \mathcal{E}_2'$ and $\mathcal{E}_3'$ are totally independent since they are determined by pairwise disjoint sets of potential edges. Also, the events $\mathcal{E}_1, \mathcal{E}_2$ and $\mathcal{E}_3$ are mutually exclusive and hence

$$\begin{aligned} \mathbf{Pr}(\mathcal{E}_1 \cup \mathcal{E}_2 \cup \mathcal{E}_3) &= \mathbf{Pr}(\mathcal{E}_1) + \mathbf{Pr}(\mathcal{E}_2) + \mathbf{Pr}(\mathcal{E}_3) \\ &\leq \mathbf{Pr}(\mathcal{E}_1) + \mathbf{Pr}(\mathcal{E}_2' \mid \overline{\mathcal{E}_1}) + \mathbf{Pr}(\mathcal{E}_3' \mid \overline{\mathcal{E}_1} \cap \overline{\mathcal{E}_2}). \end{aligned}$$

Let V_1 denote the set of first $n/2$ vertices of σ. Then, by Theorem 4.1, any maximal tournament in $D[V_1])$ is of size at least $\log_r(n/2) - \log_r(\ln(n/2) + \ln\ln(n/2)) \geq g(n/2, p, c') = \lceil \log_r(n/2) - c'\sqrt{\log_r(n/2)} \rceil$, with probability $1 - o(1)$. Hence, $\mathbf{Pr}(\mathcal{E}_1) = o(1)$.
For any fixed vertex $u \in B$,

$$\mathbf{Pr}(u \in C) \ = \ \mathbf{Pr}(\forall\, v \in A,\ (v,u) \in E) \ = \ p^{|A|}.$$

Hence

$$\mu \ = \ E[|C|] \ = \ |B| \cdot p^{|A|}.$$

Since $|C|$ is the sum of $|B|$ identical and independent indicator random variables, by applying Chernoff-Hoeffding bounds (see [19, 3]), we get that

$$\mathbf{Pr}(|C| \notin [(0.9)\mu, (1.1)\mu]) \ \leq \ 2e^{-\mu/300}.$$

Since $|A| = g(n/2, p, c')$, we deduce that

$$\mu \ \approx \ |B| \cdot 2r^{c'\sqrt{\log_r n/2}}/n,$$

after justifiably ignoring the effect of the ceiling function used in the definition of $g(n/2, p, c)$. More precisely, since we know that $\mathcal{E}_1$ has not occured, $|B| \geq n/2$ and hence

$$r^{c'\sqrt{\log_r n/2}} \ \leq \ \mu \ \leq \ 2r^{c'\sqrt{\log_r n/2}}. \tag{8}$$

It is easy to verify that $\mu \to \infty$ as $n \to \infty$. Hence $\mathbf{Pr}(\mathcal{E}_2' \mid \overline{\mathcal{E}_1}) = o(1)$.

Given that neither of $\mathcal{E}_1$ and $\mathcal{E}_2$ holds, it follows that $|C| \geq (0.9)\mu \approx (0.9) \cdot r^{c'\sqrt{\log_r n/2}}$. Hence, using $r \geq 2$ and applying Theorem 1.2,

$$mat(D[C]) \geq \left\lfloor 2c'\sqrt{\log_r n/2} + 0.5 + 2\log_r 0.9 \right\rfloor \geq \left\lfloor 2c'\sqrt{\log_r n/2} \right\rfloor - 1$$

with probability $1 - o(1)$. This establishes that $\mathbf{Pr}(\mathcal{E}_3' \mid \overline{\mathcal{E}_1} \cap \overline{\mathcal{E}_2}) = o(1)$. It then follows from (8) that ACYTOUR(D) outputs a solution of required size with probability $1 - o(1)$.

Time Complexity. It is easy to see that the running time is polynomial except for the **for** loop of lines 10 and 11. The maximum number of iterations of the **for** loop is at most

$$\binom{|C|}{|X|} \leq \binom{(1.1)\mu}{\left\lfloor 2c'\sqrt{\log_r(n/2)} \right\rfloor} \leq \left(\binom{(1.1)\cdot 2r^{c'\sqrt{\log_r(n/2)}}}{\left\lfloor 2c'\sqrt{\log_r(n/2)} \right\rfloor}\right)$$
$$= O\left(r^{4c'^2(\log_r n)}\right) = O\left(n^{O(1)}\right),$$

where the upper bound on μ is the one obtained in (8). Since each iteration takes polynomial time, the algorithm finishes in polynomial time always. ∎

Remark 5.2. In Theorem 5.1, we assume that $p \geq n^{-1/c^2}$. This is because if $p \leq n^{-1/c^2}$, then $mat(D) \leq \lceil 2c^2 + 1 \rceil$ a.a.s. and hence even a provably optimal solution can be found in polynomial time a.a.s.

6. $mat(D)$ for Non-simple Random Digraphs

We also consider another model introduced in [22] which does not force the random digraph to be simple and allows cycles of length 2.

Model $D \in \mathcal{D}_2(n,p)$: Choose each *directed* edge $u \to v$ joining distinct elements of V independently with probability p.

Note that if $D \in \mathcal{D}_2(n,p)$ and $D' \in \mathcal{D}_2(n, 1-p)$, then for every b, we have

$$\mathbf{Pr}(mat(D) = b) = \mathbf{Pr}(mat(D') = b).$$

Hence, for the rest of this section, without loss of generality, we assume that $p \leq 0.5$ and use q to denote $1 - p$.

The maximum size of any induced acyclic tournament is determined by those unordered pairs $\{u, v\}$ such that exactly one arc between u and v is present. Hence, if $D \in \mathcal{D}_2(n,p)$ and $D' \in \mathcal{D}(n, pq)$, then for every b, we have

$$\mathbf{Pr}(mat(D) = b) = \mathbf{Pr}(mat(D') = b).$$

Hence, we can obtain the following analogues of Lemma 9.1, Theorems 1.2, 1.3, 1.6, 4.1, 5.1 and Corollary 1.4.

Lemma 6.1. *For any positive integer b, for a random digraph $D \in \mathcal{D}_2(n,p)$,*

$$\mathbf{Pr}[mat(D) \geq b] \quad \geq \quad \mathbf{Pr}[\omega(G) \geq b],$$

where $G \in \mathcal{G}(n,pq)$.

Theorem 6.2. *Let $D \in \mathcal{D}_2(n,p)$ with $p \geq 1/n$. Define*

$$d = 2\log_{(pq)^{-1}} n + 1 = \frac{2(\ln n)}{\ln(pq)^{-1}} + 1; \quad b^* \; = \; \lfloor d - 1/2 \rfloor.$$

Then, a.a.s. as $n \to \infty$, $mat(D)$ is either b^* or $b^* + 1$.

Theorem 6.3. *Let $D \in \mathcal{D}_2(n,p)$. Let $w = w(n)$ be any function so that as $n \to \infty$, $w \leq 0.5(\ln n)$ and $w \to \infty$. If $p = p(n)$, $p \geq 1/n$, is such that d (defined in Theorem 6.2) satisfies $\frac{w}{\ln n} \leq \lceil d \rceil - d \leq 1 - \frac{w}{\ln n}$ for all large values of n, then $mat(D)$ is a.a.s equal to $\lfloor d \rfloor$.*

Corollary 6.4. *Let $D \in \mathcal{D}_2(n,p)$. For every constant function $p = p(n)$, there exists a function $f = f(n) = 1 - o(1)$ such that the set $N_{f,p}$ is a subset of natural numbers having density 1.*

Our goal is to obtain a threshold statement in terms of $p = p(n)$. First, observe that Theorem 1.6 can be applied straightaway to get a threshold statement (for $\mathcal{D}_2(n,p)$ model) in terms of the parameter pq. However, to get a threshold in terms of p more work needs to be done. Before stating the analogue of the threshold theorem, we need some definitions.

Let $w = w(n)$ be a sufficiently slow-growing function of n, such that $w = \omega(1)$ and $w = o(\ln n)$. Let $i = i(n)$ be a suitably growing function which goes to ∞ as $n \to \infty$. Define $a = n^{-2/(i-1+\frac{w}{\ln n})}$, and $b = n^{-2/(i-1-\frac{w}{\ln n})}$. Let $f(x,y)$ denote the function $x^2 - x + y$.

Theorem 6.5. *Let $i = i(n) \in \{1, \ldots, \lfloor \log_4 n \rfloor\}$ (for every n) be any fixed function of n. Then, there exist functions $c = c_i(n) \in [0,1]$ and $d = d_i(n) \in [0,1]$ such that: if $D \in \mathcal{D}_2(n,p)$ with $p \geq 1/n$, then, asymptotically almost surely, the following are true:*

(i) *If $p \geq c$, then $pq \geq a$ and hence $mat(D) \geq i$.*

(ii) *If $p \leq d$, then $pq \leq b$ and hence $mat(D) < i$,*

where c, d are the real positive roots in the range $[0, 1/2]$ of the quadratic equations $f(x,a) = 0$ and $f(y,b) = 0$ respectively. Also, if $i = i(n)$ is a growing function, then $c - d = o(c)$.

Hence it follows that we obtain thresholds (sharp if $i = i(n)$ increases) for the existence of induced acyclic tournaments of size i.

***Proof of Theorem* 6.5.** Notice that $pq = p(1-p) = p-p^2$ and hence if $pq = y$, $y \in \Re^+$, then $p^2 - p + y = 0$, i.e., $f(p,y) = 0$. Now, taking y to be a and b respectively, we get that if $p = c$, then $pq = a$; if $p = d$, then $pq = b$. Also, since pq is increasing when $x \in [0,1/2]$, $p \geq c$ implies $pq \geq a$, and $p \leq d$ implies $pq \leq b$. The Claims (i) and (ii) now follow by applying Lemma 3.2. It is easy to check that for each $y = a, b$, $f(x,y) = 0$ has 2 positive real roots only one of which lies in the range $[0,1/2]$.

Now, for a sharp threshold we need to show that $(c-d) = o(c)$, i.e., $1-d/c = o(1)$. This is proved as follows: If $d/c \geq (1-1/\sqrt{i})$, then we are done, since $1-d/c \leq 1/\sqrt{i} = o(1)$. Therefore, assume that $d/c \leq (1-1/\sqrt{i})$. Now, $c \in [0,1/2]$ and hence $c \leq 1/2$. By our assumption, $d \leq (1-1/\sqrt{i})c \leq \frac{1}{2}(1-1/\sqrt{i})$. Hence, $c+d \leq 1-\frac{1}{2\sqrt{i}}$. Since c and d satisfy $f(c,a) = 0$ and $f(d,b) = 0$, after subtracting, we get $f(c,a)-f(d,b) = (c-d)(c+d-1)+a-b = 0$. Therefore, $c-d = \frac{a-b}{1-c-d}$. Now using the upper bound on $c+d$, we get $c-d \leq \frac{a-b}{1/(2\sqrt{i})} = 2(a-b)\sqrt{i}$. Observe that $a \leq c \leq 2a$, since $a = c-c^2$ and $c \in [0,1/2]$. Therefore, $(c-d)/c \leq (c-d)/a \leq 2(a-b)\sqrt{i}/a$. But from the remark following the proof of Theorem 1.6, we have that $(a-b)/a = O(1/i)$. Therefore $(c-d)/c = O(\sqrt{i}/i) = O(1/\sqrt{i}) = o(1)$. Thus in this case too, the threshold is seen to be sharp. ∎

Theorem 6.6. *Given $D \in \mathcal{D}_2(n,p)$ with $pq \geq n^{-1/4}$ and any $w = w(n)$ such that $w(n) \to \infty$ as $n \to \infty$, with probability $1-o(1)$, every maximal induced acyclic tournament is of size at least $d = \lfloor \delta \log_{(pq)^{-1}} n \rfloor$, where $\delta = 1 - \frac{\ln(\ln(npq)+w)}{\ln n}$.*

Theorem 6.7. *Let $D \in \mathcal{D}_2(n,p)$. For every sufficiently large constant $c \geq 1$, if $p \leq 0.5$ is such that $n^{-1/c^2} \leq pq \leq 0.25$, then, with probability $1-o(1)$, $ACYTOUR(D)$ will output an induced acyclic tournament of size at least $b' = \lfloor (1+\epsilon') \log_{(pq)^{-1}} n \rfloor$, where $\epsilon' = c/\sqrt{\log_{(pq)^{-1}} n}$.*

Remark 6.8. However, in the case of $\mathcal{D}_2(n,p)$ model, we need to slightly modify the description of ACYTOUR(D) as follows: In the definition of C (Line 8), we also need to require that $(u,v) \notin E$ for each $v \in A$.

7. On the Maximum Size of Induced Tournaments

Suppose we drop the requirement of acyclicity of the induced tournament. It then reduces to the clique problem as follows. Let us first recall some basic facts about the distributions of $\omega(G)$ and $\alpha(G)$ for $G \in \mathcal{G}(n,p)$. $\omega(G)$ ($\alpha(G)$) denotes the maximum size of a clique (an independent set) in G. It is easy to verify that $\omega(G)$

for $G \in \mathcal{G}(n,p)$ and $\alpha(G)$ for $G \in \mathcal{G}(n,1-p)$ are identically distributed. Also, by the classical results of Bollobás and Erdős [5], and Grimmett and McDiarmid (see e.g. [4], Chapter 11), $\omega(G)$ is a.a.s. concentrated in just two values for every $p = p(n) \leq 1 - n^{-\epsilon}$ for some suitably small constant $\epsilon > 0$. But it does not seem to exhibit such sharp concentration behavior for larger values of p. In particular, if p is such that $p = 1 - n^{-2/3}$, $\omega(G)$ is only known (see [10]) to be concentrated in a band of $\Theta(n^{2/3})$.

This has implications to the concentration of the maximum size of an induced (need not be acyclic) tournament in a random digraph. We use $\omega(D)$ to denote the maximum size of an induced tournament in D. It is clear that $\omega(D)$ for $D \in \mathcal{D}(n,p)$ and $\omega(G)$ for $G \in \mathcal{G}(n,2p)$ are identically distributed for every $p = p(n) \leq 0.5$. Similarly, $\omega(D)$ for $D \in \mathcal{D}_2(n,p)$ and $\omega(G)$ for $G \in \mathcal{G}(n,2p(1-p))$ are identically distributed for every $p = p(n) \leq 1$.

But, unlike the case of $mat(D)$, the concentration of $\omega(D)$ is quite different between the two models $\mathcal{D}(n,p)$ and $\mathcal{D}_2(n,p)$. First, we focus on the model $\mathcal{D}_2(n,p)$. Since $2p(1-p) \leq 0.5$ for any $0 \leq p \leq 1$, $\omega(G)$ is 2-point concentrated for $G \in \mathcal{G}(n,2p(1-p))$, and hence we notice that $\omega(D)$ is always concentrated in just two values for any p.

If $D \in \mathcal{D}(n,p)$, then $\omega(D)$ is concentrated in just two values a.a.s. for any $p = p(n) \leq 0.25$. However, for $0.25 < p \leq 0.5$, $\omega(D)$ is not tightly concentrated and has the same distribution and concentration behavior as $\omega(G)$ for certain ranges of $p \geq 0.5$ (see the discussion before).

8. Summary

The problem of determining the size of the largest induced acyclic tournament $mat(D)$ in a random digraph was studied. We showed that a.a.s. $mat(D)$ takes one of only two possible values. The result is valid for all ranges of the arc probability p. The value of $mat(D)$ also has an explicit closed form expression (for all ranges of p) which does not seem to exist for clique number $\omega(G)$ of a random graph.

The results of this paper and those of [23], [22] and [8] show that $mat(D)$ of a random digraph behaves like the clique number $\omega(G)$ of a random graph and maximum induced acyclic subgraph size $mas(D)$ behaves like the independence number $\alpha(G)$ of a random graph (see also the discussion above in Section 7).

We then showed that a.a.s. every maximal acyclic tournament is of a size which is at least nearly half of the optimal size. As a result, one immediately gets an efficient approximation algorithm whose approximation ratio is bounded by $2 + O((\ln\ln n)/(\ln n))$. We also considered and analyzed another efficient heuristic whose approximation ratio was shown to be $2 - O(1/\sqrt{\log_r n})$.

An interesting and natural open problem that comes to mind is the following.

Open Problem: Let p be a constant such that $0 < p \leq 0.5$. Design a polynomial time algorithm which, given $D \in \mathcal{D}(n,p)$, a.a.s. finds an induced acyclic tournament of size at least $(1+\epsilon)\log_r n$ for some positive constant ϵ.

Solving this problem could turn out to be as hard as designing an efficient algorithm which finds, given $G \in \mathcal{G}(n,1/2)$, a clique of size $(1+\epsilon)\log_2 n$ and the latter problem has remained open for more than three decades.

Unlike the case of $mat(D)$, the gap between lower and upper bounds on $mas(D)$ obtained in [23, 22] is not very sharp. However, further progress has been made on shortening this gap and the details appear in the extended abstract [8].

Acknowledgements

We sincerely thank the anonymous referees whose detailed review and comments led to a considerable improvement of our presentation. In particular, we were motivated to obtain Theorem 1.5 by a query posed by one of the reviewers.

References

[1] D. Achlioptas and A. Naor, *The two possible values of the chromatic number of a random graph*, Ann. of Math. **162** (2005) 1333–1349. doi:10.4007/annals.2005.162.1335

[2] N. Alon and M. Krivelevich, *The concentration of the chromatic number of random graphs*, Combinatorica **17** (1997) 303–313. doi:10.1007/BF01215914

[3] N. Alon and J.H. Spencer, The Probabilistic Method (Wiley International, 2001).

[4] B. Bollobás, Random Graphs (2nd Edition, Cambdrige University Press, 2001). doi:10.1017/CBO9780511814068

[5] B. Bollobás and P. Erdős, *Cliques in random graphs*, Math. Proc. Cambridge Philos. Soc. **80** (1988) 419–427. doi:10.1017/S0305004100053056

[6] B.K. Rosen, *Robust linear algorithms for cutsets*, J. Algorithms **3** (1982) 205–217. doi:10.1016/0196-6774(82)90020-7

[7] K. Dutta and C.R. Subramanian, *Largest induced acyclic tournament in random digraphs: A 2-point concentration*, in: Proceedings of LATIN-2010 (9th Latin American Theoretical Informatics Symposium), Oaxaca, Mexico, April 2010.

[8] K. Dutta and C.R. Subramanian, *Induced acyclic subgraphs in random digraphs: Improved bounds*, in: Proceedings of AofA-2010 (21st Internatioanl Meeting on Probabilistic and Asymptotic Methods for the Analysis of Algorithms), Vienna, Austria, June 2010, to appear.

[9] R.W. Floyd *Assigning meaning to programs*, Proc. Sympos. Appl. Math. **19** (1967) 19–32.

[10] A.M. Frieze, *On the independence number of random graphs*, Discrete Math. **81** (1990) 171–176.
doi:10.1016/0012-365X(90)90149-C

[11] M.R. Garey and D.S. Johnson, Computers and Intractability: A Guide To The Theory of NP-Completeness (W.H.Freeman, San Francisco, 1978).

[12] J. Hastad, *Clique is hard to approximate within* $n^{1-\varepsilon}$, Acta Math. **182** (1999) 105–142.
doi:10.1007/BF02392825

[13] S. Janson, T. Łuczak and A. Ruciński, Random Graphs (John Wiley & Sons, Inc., 2000).
doi:10.1002/9781118032718

[14] S. Khot, *Improved inapproximability results for maxclique, chromatic number and approximate graph coloring*, in: Proceedings of the 42nd IEEE Symp. Foundations of Computer Science (FOCS 2001) 600–609.

[15] M. Krivelevich and B. Sudakov, *Coloring random graph*, Inform. Process. Lett. **67** (1998) 71–74.
doi:10.1016/S0020-0190(98)00092-1

[16] T. Łuczak, *A note on the sharp concentration of the chromatic number of random graphs*, Combinatorica **11** (1991) 295–297.
doi:10.1007/BF01205080

[17] C. Lund and M. Yannakakis, *The approximation of maximum subgraph problems*, Proceedings of the 20th International Colloquium on Automata, Languages and Programming (ICALP'93), Lecture Notes in Comput. Sci. **700** (1993) 40–51.

[18] M. Cai, X. Deng and W. Zang, *An approximation algorithm for feedback vertex set in tournaments* SIAM J. Comput. **30** (2001) 1993–2007.
doi:10.1137/S0097539798338163

[19] R. Motwani and P. Raghavan, Randomized Algorithms (Cambridge University Press, 1995).
doi:10.1017/CBO9780511814075

[20] C.H. Papadimitriou and M. Yannakakis *Optimization, approximation, and complexity classes*, J. Comput. System Sci. (Special issue for the 20th ACM Symposium on Theory of Computing) **43** (1991) 425–440.

[21] E. Speckenmeyer, *On feedback problems in digraphs*, Proceedings of the 15th International Workshop on Graph Theoretic Concepts in Computer Science (WG'89), Springer-Verlag, Lecture Notes in Comput. Sci. **411** (1990) 218–231.
doi:10.1007/3-540-52292-1_16

[22] J.H. Spencer and C.R. Subramanian, *On the size of induced acyclic subgraphs in random digraphs*, Discrete Math. Theor. Comput. Sci. **10** (2008) 47–54.

[23] C.R. Subramanian, *Finding induced acyclic subgraphs in random digraphs*, Electron. J. Combin. **10** (2003) #R46.

9. APPENDIX

9.1. $mat(D)$ versus $\omega(G)$

The following lemma relates the probabilities in the two models $\mathcal{D}(n,p)$ and $\mathcal{G}(n,p)$ for having, respectively, tournaments and cliques of specific sizes. Its proof is similar to the proof of an analogous relationship involving $mas(D)$ and $\alpha(G)$ (maximum size of an independent set in G) established in [23].

Lemma 9.1. *For any positive integer b, for a random digraph $D \in \mathcal{D}(n,p)$,*

$$\mathbf{Pr}[mat(D) \geq b] \geq \mathbf{Pr}[\omega(G) \geq b],$$

where $G \in \mathcal{G}(n,p)$.

Proof. Given a linear ordering σ of vertices of D and a subset A of size b, we say that $D[A]$ is consistent with σ if for every $\sigma_i, \sigma_j \in A$ with $i < j$, $D[A]$ has the arc (σ_i, σ_j).

Let τ denote an arbitrary but fixed ordering of V. Once we fix τ, the spanning subgraph of D formed by arcs of the form $(\tau(i), \tau(j))$ $(i < j)$ is having the same distribution as $\mathcal{G}(n,p)$. Hence, for any A, the event of $D[A]$ being consistent with τ is equivalent to the event of A inducing a clique in $\mathcal{G}(n,p)$. Hence,

$$\begin{aligned}
\mathbf{Pr}(\ mat(D) \geq b) &= \mathbf{Pr}(\ \exists A,\ |A| = b,\ D[A] \text{ is an acyclic tournament}) \\
&= \mathbf{Pr}(\ \exists A,\ |A| = b,\ \exists \sigma,\ D[A] \text{ is consistent with } \sigma) \\
&= \mathbf{Pr}(\ \exists \sigma,\ \exists A,\ |A| = b,\ D[A] \text{ is consistent with } \sigma) \\
&\geq \mathbf{Pr}(\ \exists A,\ |A| = b,\ D[A] \text{ is consistent with } \tau\ ) \\
&= \mathbf{Pr}(\ \omega(G) \geq b).
\end{aligned}$$

Hence it is natural that we have a bigger upper bound for $mat(D)$ than we have for $\omega(G)$. ■

Note: Recall that we first draw an undirected $G \in \mathcal{G}(n, 2p)$ and then choose uniformly randomly an orientation of $E(G)$. Hence, for any fixed $A \subseteq V$ of size b with $b = \omega(1)$,

$$\mathbf{Pr}(\ D[A] \text{ is an acyclic tournament } \mid G[A] \text{ induces a clique }) = \frac{b!}{2^{\binom{b}{2}}} = o(1).$$

However, there are so many cliques of size b in G that one of them manages to induce an acyclic tournament.

9.2. Proof of Theorem 1.7

We reduce the NP-complete Maximum Clique problem MC(G,k) to the MAT(D,k) problem as follows. Given an instance $(G=(V,E),k)$ of the first problem, compute an instance $f(G)=(G'=(V,A),k)$ in polynomial time where

$$A=\{(u,v): uv\in E, u<v\}.$$

Clearly, G' is a dag and it is easy to see that a set $V'\subseteq V$ induces a clique in G if and only if V' induces an acyclic tournament in G'. This establishes that MAT(D,k) is NP-hard even if D is restricted to be a dag.

The inapproximability of MAT(D) follows from the following observation. Note that the reduction $G\to f(G)$ is an L-reduction in the sense of [20], since $|f(G)|=|G|$ and $\omega(G)=mat(G')$. Hence, any inapproximability result on maximum clique in undirected graphs (for example [12, 14]), implies a similar inapproximability for the MAT(D) problem.

9.3. Proof of Claim 2.1

Order the vertices of U along a Hamilton path P (if any exists) of H. An arc $(u,v)\in A$ is a forward arc if u comes before v in P and is a backward arc otherwise. Since H is acyclic, any arc $(v,u)\in A$ must be a forward arc, since otherwise the segment of P from u to v along with (v,u) forms a cycle in H.

Now if there is another Hamilton path Q in H, $Q\neq P$, then walking along P, consider the first vertex a where Q differs from P. Then in the path Q, a is visited immediately after some vertex a' that comes after a in P. But this implies that (a',a) is a backward arc in H contradicting the observation earlier that H has no backward arc.

9.4. Remaining cases of Theorem 1.2

For $1/wn\le p<1/n$,

$$E[X(n,4)]=\binom{n}{4}\cdot 4!\cdot p^{\binom{4}{2}}\le n^4p^6 \ \le\ (1/n^2)=o(1).$$

Now, an acyclic tournament of size 2 is simply an edge which a.a.s. exists since:

$$\mathbf{Pr}[mat(D)<2]=\mathbf{Pr}[D \text{ is the empty graph}]=(1-2p)^{\binom{n}{2}}\le e^{-n(n-1)p}=o(1),$$

since $p\ge 1/wn\ge w/n^2$. Hence, when $1/wn\le p\le 1/n$, $mat(D)\in\{2,3\}$, a.a.s.

For $wn^{-2}\le p<1/wn$,

$$E[X(n,3)]=\binom{n}{3}\cdot 3!\cdot p^{\binom{3}{2}}\le n^3p^3=o(1) \text{ since } np=o(1).$$

The proof for $mat(D) \geq 2$ is the same as in the previous case, since $n^2p = \omega(1)$, and hence, at least one arc will exist, a.a.s. So when $w/n^2 \leq p \leq 1/wn$, $mat(D) = 2$, a.a.s.

For $(wn^2)^{-1} \leq p \leq w/n^2$, $E[X(n,3)] = o(1)$, as in the previous case, and so $mat(D) = 1$ or 2, a.a.s. When $p < (wn^2)^{-1}$, $mat(D) = 1$ since D a.a.s. has no directed edge.

Received 25 April 2011
Revised 29 April 2013
Accepted 29 April 2013