Discrete Mathematics and Theoretical Computer Science DMTCS vol. **10:2**, 2008, 47–54

On the size of induced acyclic subgraphs in random digraphs

Joel Spencer[1] and C.R. Subramanian[2]

[1] *Courant Institute of Mathematical Sciences, New York University, New York, NY 10012, USA.*
email*: spencer@cs.nyu.edu.*

[2] *The Institute of Mathematical Sciences, Taramani, Chennai - 600 113, India.*
email*: crs@imsc.res.in*

received Feb 8, 2006 , revised Nov 1, 2007, accepted Jan 31, 2008.

Let $D \in \mathcal{D}(n,p)$ denote a simple random digraph obtained by choosing each of the $\binom{n}{2}$ undirected edges independently with probability $2p$ and then orienting each chosen edge independently in one of the two directions with equal probability $1/2$. Let $mas(D)$ denote the maximum size of an induced acyclic subgraph in D. We obtain tight concentration results on the size of $mas(D)$. Precisely, we show that

$$mas(D) \;\leq\; \frac{2}{\ln(1-p)^{-1}}\,(\ln np + 3e)$$

almost surely, provided $p \geq W/n$ for some fixed constant W. This combined with known and new lower bounds shows that (for p satisfying $p = \omega(1/n)$ and $p \leq 0.5$)

$$mas(D) \;=\; \frac{2(\ln np)}{\ln(1-p)^{-1}}\,(1 \pm o(1))\,.$$

This proves a conjecture stated by Subramanian in 2003 for those p such that $p = \omega(1/n)$. Our results are also valid for the random digraph obtained by choosing each of the $n(n-1)$ directed edges independently with probability p.

Keywords: Induced Acyclic Subgraphs, Random Digraphs

1 Introduction

Given a *simple* digraph $H = (V,F)$, let $mas(H)$ denote the maximum size ($|V'|$) of a subset $V' \subseteq V$ such that the induced subgraph $H[V']$ is acyclic. By "simple", we mean that there is at most one arc (directed edge) between any unordered pair of vertices. Self-loops are not allowed. Throughout, we mean vertex induced subgraphs whenever we use the term subgraphs. The associated algorithmic problem of deciding, given an arbitrary H and k, if $mas(H) \geq k$ is NP-complete [3]. The optimization version of this algorithmic problem is not polynomial-time approximable even within a multiplicative ratio of $O(n^\epsilon)$ for some $\epsilon > 0$ unless $P = NP$ [5].

Suppose D is a a random digraph drawn from the model $\mathcal{D}(n,p)$ defined below. Throughout, we assume that $V = \{1,\ldots,n\}$. Also, $p \leq 0.5$ is a positive real number.

Model $D \in \mathcal{D}(n,p)$ **:** Choose each *undirected* edge $\{u,v\}$ joining distinct elements of V independently with probability $2p$. For each chosen $\{u,v\}$, orient it in one of the two directions $\{u \to v, v \to u\}$ in D with equal probability (=0.5). The choices of arcs are independent for different chosen unordered pairs. This results in a simple digraph where each arc is chosen with probability p. When $p = 0.5$, there is an arc between every pair of vertices resulting in a random tournament.

The range of $mas(D)$ has been analyzed for $D \in \mathcal{D}(n,p)$ and the following upper and lower bounds were obtained by Subramanian in [7].

Theorem 1.1 *([7]) Let $D \in \mathcal{D}(n,p)$, $p \leq 0.5$. Then, almost surely,*

$$mas(D) \leq \min\left(\left\lceil \frac{2\,(\ln n)}{\ln(1-p)^{-1}} \right\rceil + 1, n\right).$$

Theorem 1.2 *([7]) Write $w = np$. For each fixed $\epsilon > 0$, there exists a sufficiently large positive constant w_ϵ such that : If p is such that $w_\epsilon \leq w = o(n)$, then, almost surely,*

$$mas(D) \ \geq \ \left(\frac{2}{p}\right)(\ln w - \ln\ln w - \ln 2 + 1 - 0.5\epsilon).$$

While the above-mentioned upper and lower bounds are asymptotically the same (upto constant multiplicative factors) for large values of p, they differ significantly for very small values of p. In particular, if $p = n^{-1+o(1)}$ and hence $\ln np = o(\ln n)$, then the lower bound $2p^{-1}(\ln np)$ is asymptotically much smaller than the upper bound $2p^{-1}(\ln n)$. In [7], it was conjectured that the upper bound can be improved further and also that lower and upper bounds are the same (upto negligible (asymptotically) additive terms). In this paper, we partially prove this by improving the upper bound and bring it closer to the lower bound and also show that they are asymptotically the same when p is not too small and obtain tight concentration results on the value of $mas(D)$. Formally, we obtain the following

Theorem 1.3 *Write $w = np$. There is a sufficiently large constant W such that : If p satisfies $w \geq W$, then, almost surely,*

$$mas(D) \ \leq \ \left(\frac{2}{\ln(1-p)^{-1}}\right)(\ln w + 3e)\,.$$

The lower bound of Theorem 1.2 requires that $p = o(1)$. One can obtain a lower bound for larger values of p also. This bound is precisely stated in Theorem 1.5 given below.

Let $G \in \mathcal{G}(n,p)$ denote the simple undirected random graph on V where each edge is chosen independently with probability p. Let $\alpha(G)$ denote the maximum size of an independent set in G. The following lemma (Lemma 2.1) of [7] relates lower bounds on $mas(D)$ and $\alpha(G)$.

Lemma 1.1 ([7]) *Let $D \in \mathcal{D}(n,p)$ and $G \in \mathcal{G}(n,p)$. For any positive integer b,*

$$\mathbf{Pr}(\ mas(D) \geq b\ ) \quad \geq \quad \mathbf{Pr}(\ \alpha(G) \geq b\ ).$$

For $p \geq (\ln n)^{-1}$, Bollobas and Erdős [1] and Matula [6] (see also Chapter 7 of [4]) obtained the following lower bound on $\alpha(G)$.

Theorem 1.4 ([1], [6]) *Write $w = np$. Let p be such that $(\ln n)^{-1} \leq p \leq 0.5$. Then, almost surely,*

$$\alpha(G) \ \geq \ \left(\frac{2}{\ln(1-p)^{-1}}\right)(\ln w - \ln\ln w - O(1))\,.$$

Applying Lemma 1.1 to the previous theorem, we obtain

Theorem 1.5 *Write $w = np$. Let p be such that $p \geq (\ln n)^{-1}$. Then, almost surely,*

$$mas(D) \ \geq \ \left(\frac{2}{\ln(1-p)^{-1}}\right)(\ln w - \ln\ln w - O(1))\,.$$

Combining the lower and upper bounds mentioned in Theorems 1.2, 1.5 and 1.3, we obtain the following corollary which provides a tight estimate of the range of $mas(D)$.

Corollary 1.1 *Write $w = np$. If p satisfies $w = \omega(1)$, then almost surely,*

$$mas(D) \ = \ \left(\frac{2(\ln w)}{\ln(1-p)^{-1}}\right)(1 \pm o(1))\,.$$

Here, $o(1)$ is with respect to n.

Proof: If $p \leq 1/(\ln n)$, then we apply the lower bound of Theorem 1.2 and also use the fact that $\ln(1-p)^{-1} \geq p$ for each $p < 1$ to get the desired result. Otherwise, we can apply the lower bound of Theorem 1.5 to obtain the corollary. □

Note 1 : The lower bound of Theorem 1.2 has a factor of p^{-1} which is better than the factor $1/(\ln(1-p)^{-1})$ used by the bound of Theorem 1.5. This improvement is marginal since Theorem 1.2 is applicable only for small ($p = o(1)$) values of p. But the improvement is significant for higher values of p, *if* possible. For example, when $p = 1/2$, $2p^{-1} = 4$ but $2/(\ln(1-p)^{-1}) = 2/(\ln 2) \approx 2.886$. But such an improvement for higher values of p is ruled out by the upperbound of Theorem 1.3.

Note 2 : By a constructive lower bound of $(1-\epsilon)(\ln w)/(\ln(1-p)^{-1})$ on the size of any maximal induced acyclic subgraph obtained by Subramanian in [7], we notice that, for any p, a simple greedy heuristic produces a solution whose size is almost surely within a ratio of $2+\delta$ for any positive δ. Thus, for random digraphs, one can obtain a constant factor approximation to the optimal solution.

The results (theorems and corollaries) stated in this section hold true if D is a random graph obtained by choosing each potential directed edge independently with probability p. See Section 3 for further details.

2 Proof of Theorem 1.3

We use q to denote the expression $\ln(1-p)^{-1}$. The following fact will be used in the proof. For $p = 0$, we take the limit.

$$\frac{q}{p} \ = \ \frac{\ln(1-p)^{-1}}{p} \ = \ \frac{-\ln(1-p)}{p} \ \leq \ 1.5 \quad \text{for } 0 \leq p \leq 0.5. \tag{1}$$

Let u be any positive integer and let $A \subseteq V$ be a set of size u such that $D[A]$ is acyclic. Then, there exists an ordering $\sigma : \{1, \ldots, u\} \rightarrow A$ of A such that each arc in $D[A]$ is of the form $\sigma(i) \rightarrow \sigma(j)$ for some $1 \leq i < j \leq u$. This implies that, for any $1 \leq k \leq u$, there is an ordered partition of A into k subsets $A_1, \ldots, A_k$ such that

- $|A_i| = u_i = \lceil u/k \rceil$ for $i \leq r$ and $|A_i| = u_i = \lfloor u/k \rfloor$ for $i > r$ where $r = u \bmod k$.
- For each $1 \leq i < j \leq k$, there is no arc in $D[A]$ from any element in A_j to any element in A_i.

Note 3: We shall carefully choose k below to give nearly optimal results. The case $k = 2$ was first explored by the authors. Already this gives $mas(D) = O(p^{-1} \ln(np))$ and the technical calculations below are considerably simpler.

Now, choose u to be the integer $u = q^{-1}(\ln w)(2 + \delta)$ where δ is determined by the requirement that $u = \lfloor q^{-1}(\ln w)(2 + 6e(\ln w)^{-1}) \rfloor$. It is sufficient to prove that, almost surely, the subgraph induced by any A of size u is not acyclic. Fix a subset A of size u. Using the above mentioned observation about such acyclic subsets and also the inequality of Lemma 2.1 (given later),

$$\begin{aligned}
\mathbf{Pr}(\ D[A] \ \text{ is acyclic }) &\leq \binom{u}{u_1, \ldots, u_k} (1-p)^{\sum_{i<j} u_i u_j} \\
&\leq \binom{u}{u_1, \ldots, u_k} (1-p)^{\binom{k}{2}(u/k)^2 - k/8} \\
&\leq k^u \, e^{-qu^2(k-1)/2k} (1-p)^{-k/8} \\
&= k^u \, e^{-u(\ln w)(2+\delta)(k-1)/2k} (1-p)^{-k/8} \\
&= \left(\frac{k}{e^{(\ln w)(2+\delta)(k-1)/2k}} \right)^u (1-p)^{-k/8}
\end{aligned}$$

This bound holds for any fixed A of size u. There are

$$\binom{n}{u} \leq \left(\frac{en}{u}\right)^u$$

sets of size u. This gives us

$$\begin{aligned}
\mathbf{Pr}(\ \exists A,\ |A| = u,\ D[A] \ \text{ is acyclic }) &\leq \left(\frac{enk}{ue^{(\ln w)(2+\delta)(k-1)/2k}} \right)^u (1-p)^{-k/8} \\
&= \left(\frac{ewk}{(2+\delta)(\ln w) w^{(2+\delta)(k-1)/2k}} \cdot \frac{q}{p} \right)^u (1-p)^{-k/8}
\end{aligned}$$

Now choose k to be the integer nearest to $2(\ln w)(3e)^{-1} + 2$. By choosing W sufficiently large, we can ensure that $k \leq 3(\ln w)(3e)^{-1}$. Thus, we have $2(\ln w)(3e)^{-1} + 1 < k \leq 3(\ln w)(3e)^{-1}$. It then follows that $2/(k-1) < (3e)(\ln w)^{-1}$ or equivalently that $2k/(k-1) = 2 + 2/(k-1) < 2 + (3e)(\ln w)^{-1}$. This implies that

$$\left(\frac{k-1}{2k}\right)(2+\delta) \geq \left(\frac{k-1}{2k}\right) \left(2 + (3e)(\ln w)^{-1}\right) > 1.$$

Also, since $p \le 0.5$, we have $(1-p)^{-1} \le 2$ and hence $(1-p)^{-k/8} \le e^{k/8} \le e^{3(\ln w)/8(3e)} = w^{3/8(3e)}$. Applying these in the probability estimate derived above, we get

$$\mathbf{Pr}(\ \exists A,\ |A| = u,\ D[A] \text{ is acyclic }) \quad \le \quad \left(\frac{1}{2} \cdot \frac{q}{p}\right)^{u} w^{3/8(3e)} \;=\; \left(\frac{4}{3}\right)^{-u} w^{1/8e}$$

Note that

$$FP = (4/3)^{-u} w^{1/8e} = w^{(1/8e)-(\ln 4/3)(2+\delta)/q} \le w^{\left(\frac{\ln 2}{8e} - (\ln 4/3)(2+\delta)\right)/q}$$

is inversely polynomial in w and is inversely exponential in $1/q$. Also, as $n \to \infty$, the quantity $\max\{1/q, w\}$ approaches ∞ for all choices of p. Hence, using our assumption $w \ge W$, we notice that $FP \to 0$ as $n \to \infty$. This establishes that, almost surely, $mas(D) \le 2q^{-1}(\ln w + 3e)$.

It remains to prove the following lemma.

Lemma 2.1 *Let u, k be two positive integers such that $u \ge k$. Let $r = u \bmod k$. Let $u = u_1 + \ldots + u_k$ with $u_i = \lceil u/k \rceil$ for $i \le r$ and $u_i = \lfloor u/k \rfloor$ for $i > r$, be an ordered partition of u into k positive parts. Then*

$$\sum_{1 \le i < j \le k} u_i u_j \;\ge\; \binom{k}{2}(u/k)^2 - k/8.$$

Proof: We have

$$\begin{aligned} \frac{u^2}{2} &= \binom{u}{2} + \frac{u}{2} = \binom{u_1 + \ldots + u_k}{2} + \frac{u}{2} \\ &= \sum_i \binom{u_i}{2} + \sum_{1 \le i < j \le k} u_i u_j + \sum_i \frac{u_i}{2} = \sum_i \frac{u_i^2}{2} + \sum_{1 \le i < j \le k} u_i u_j \end{aligned}$$

Similarly, we have

$$\frac{u^2}{2} = \left(\frac{k^2}{2}\right)\left(\frac{u}{k}\right)^2 = \left(\binom{k}{2} + \frac{k}{2}\right)\left(\frac{u}{k}\right)^2 = \binom{k}{2}\left(\frac{u}{k}\right)^2 + \left(\frac{k}{2}\right)\left(\frac{u}{k}\right)^2$$

Comparing the two equations, we get

$$\begin{aligned} \sum_{1 \le i < j \le k} u_i u_j &= \binom{k}{2}\left(\frac{u}{k}\right)^2 + \left(\frac{k}{2}\right)\left(\frac{u}{k}\right)^2 - \sum_i \frac{u_i^2}{2} \\ &= \binom{k}{2}\left(\frac{u}{k}\right)^2 - \left(\sum_i \left(u_i^2 - (u/k)^2\right)\right)/2 \end{aligned}$$

Let $u = lk + r$ where $0 \le r < k$. Define $a = u/k$ and $b = r/k$. Then, $u_i = a - b + 1$ for $i \le r$ and $u_i = a - b$ for $i > r$. It follows that

$$\sum_i (u_i^2 - (u/k)^2) \quad = \quad r\left((a-b+1)^2 - a^2\right) + (k-r)\left((a-b)^2 - a^2\right)$$

$$
\begin{aligned}
&= r\left((a-b+1)^2-(a-b)^2\right)+k\left((a-b)^2-a^2\right)\\
&= r\,(1+2a-2b)+k\left(b^2-2ab\right)\\
&= r\,(1+2a-2b)+r\,(b-2a)\\
&= r\,(1-b) \;=\; \frac{r(k-r)}{k} \;\le\; \frac{k^2}{4k} \;=\; \frac{k}{4}
\end{aligned}
$$

The lemma now follows. □

3 Non-simple random digraphs

Consider the other model 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.

Consider $D \in \mathcal{D}_2(n,p)$. For each unordered pair $\{u,v\}$ of distinct vertices, the 2-cycle $\langle u,v,u\rangle$ fails to be present in D with probability $1-p^2$. Hence,

$$\mathbf{Pr}(\ D \text{ has no 2-cycle }) \;=\; (1-p^2)^{\binom{n}{2}} \;\le\; e^{-n(n-1)p^2/2}$$

Hence if $np \to \infty$, then D has a 2-cycle almost surely. In fact, for any $f(n) = \omega(p^{-1})$, for each sufficiently large n, if A is any *fixed* subset of V of size $f(n)$, then the probability that A has a 2-cycle approaches 1. In particular, this is true for $f(n) = \Omega(p^{-1}(\ln np))$. Still, as stated in Theorem 3.1 given below, almost surely, there are induced acyclic subgraphs of size $f(n)$ in D and hence $mas(D) = \Omega(p^{-1}(\ln np))$. This may seem paradoxical but does not seem so once we notice that the number of subsets of size $f(n)$ is so large that some subset A manages to be devoid of not only 2-cycles but also of all cycles despite the low probability of this happening for any fixed set. Precisely, we obtain the following

Theorem 3.1 *Let $D \in \mathcal{D}_2(n,p)$ with $p \le 0.5$. The lower and upper bounds stated in Theorems 1.2, 1.3 and 1.5 hold true (with respective assumptions) for $D \in \mathcal{D}_2(n,p)$ also.*

Proof: One can verify from the proof of Lemma 1.1 that it holds true also when D is drawn using the $\mathcal{D}_2(n,p)$ model. Hence the lower bounds follow from applying this lemma to lower bounds on $\alpha(G)$ for $G \in \mathcal{G}(n,p)$. For the upper bounds, one can verify that the proof arguments that worked for $D \in \mathcal{D}(n,p)$ work for $D \in \mathcal{D}_2(n,p)$ also. □

As a corollary, we obtain

Corollary 3.1 *Let $D \in \mathcal{D}_2(n,p)$. Write $w = np$. If p satisfies $w = \omega(1)$, then almost surely,*

$$mas(D) \;=\; \left(\frac{2(\ln w)}{\ln(1-p)^{-1}}\right)(1 \pm o(1))\,.$$

Here, $o(1)$ is with respect to n.

Note 4: Note that the requirement $p \leq 0.5$ can be relaxed to p being upper limited by any positive constant strictly less than 1.

4 Further work

For p satisfying $w = np = \omega(1)$, Corollary 1.1 provides a tight estimate of $mas(D)$. However, the lower and upper bounds still differ by an additive term which is in $O(\frac{\ln \ln w}{\ln(1-p)^{-1}})$. We believe that the two bounds can be brought even closer to each other.

Similarly, in the case of $D \in \mathcal{D}_2(n, p)$, there is scope for tightening and reducing the gap of $O(\frac{\ln \ln w}{\ln(1-p)^{-1}})$ between the lower and upper bounds on $mas(D)$ derived in this paper. We believe this can also be reduced further and it is worth exploring.

Corollary 1.1 affirmatively settles the conjecture of [7] for $p = \omega(1/n)$. It would be interesting to extend the result to all $p \geq c/n$ for some absolute constant $c > 0$.

References

[1] B. Bollobas and P. Erdős, "Cliques in random graphs", *Math. Proc. Camb. Phil. Soc.* **80**, 419-427, 1988.

[2] A.M. Frieze, "On the independence number of random graphs", *Discrete Mathematics* **81**, 171-176, 1990.

[3] M.R. Garey and D.S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness*, W.H. Freemann, San Francisco, 1978.

[4] S. Janson, T. Luczak and A. Rucinski, *Random Graphs*, John Wiley & Sons, Inc. 2000.

[5] C. Lund and M. Yannakakis, "The Approximation of Maximum Subgraph Problems", *Proceedings of the 20th International Colloquium on Automata, Languages and Programming (ICALP'93)*, LNCS 700, pp. 40–51.

[6] D. Matula, "The largest clique size in a random graph", Tech. Rep., Dept. of Comp. Sci., Southern Methodist University, Dallas, Texas, 1976.

[7] C.R. Subramanian, "Finding induced acyclic subgraphs in random digraphs", *The Electronic Journal of Combinatorics*, 10 (2003), #R46.