SIAM J. ALG. DISC. METH.
Vol. 5, No. 4, December 1984

006

ON THE MAXIMAL NUMBER OF STRONGLY INDEPENDENT VERTICES IN A RANDOM ACYCLIC DIRECTED GRAPH*

AMNON B. BARAK† AND PAUL ERDŐS‡

Abstract. Let $\mathscr{A}_n$ denote a random acyclic directed graph which is obtained from a random graph with vertex set $\{1, 2, \cdots, n\}$, such that each edge is present with a prescribed probability p and all the edges are directed from higher to lower indexed vertices. Define a subset of vertices in $\mathscr{A}_n$ to be *strongly independent* if there is no directed path between any pair of vertices in the subset. We show that the sequence $\mathscr{I}(\mathscr{A}_n)$, the number of vertices in the largest strongly independent vertex subset of $\mathscr{A}_n$ satisfies with probability tending to 1,

$$\frac{\mathscr{I}(\mathscr{A}_n)}{\sqrt{\log n}} \to \frac{\sqrt{2}}{\sqrt{\log 1/q}} \quad \text{as } n \to \infty,$$

where $q = 1 - p$.

Key words. acyclic directed graphs, random graphs, independent vertices

CR categories. 5.32, 5.5

1. Introduction. A random graph is a graph with vertex set $\mathbb{N}$, the set of natural numbers, such that each pair of vertices is joined by an edge with a prescribed probability p, independently of the presence or absence of any other edges. We assume no loops or multiple edges. A random acyclic directed graph is a random graph in which all the edges are directed such that there are no directed cycles.

In this paper we consider the class $\mathscr{A}$ of random acyclic directed graphs which are obtained from random graphs by directing all the edges from higher to lower indexed vertices. In other words the random variables e_{ij}, $1 \leqq j < i$, defined by

$$e_{ij} = \begin{cases} 1 & \text{if there is an edge from vertex } i \text{ to vertex } j \text{ in } \mathscr{A}, \\ 0 & \text{otherwise,} \end{cases}$$

are independent random variables with $P\{e_{ij} = 1\} = p$ and $P\{e_{ij} = 0\} = 1 - p = q$. Let $\mathscr{A}_n$ denote a subgraph of $\mathscr{A}$ spanned by the vertices $\{1, 2, \cdots, n\}$.

DEFINITION. Two vertices i, j of $\mathscr{A}_n$ are called *strongly independent* (independent) if there is no directed path (edge) from i to j $(i > j)$.

Notice that the transitive closure of our random graph is a partially ordered set (poset). Two vertices in the graph are strongly independent iff they are incomparable in this poset. A set of vertices which are pairwise strongly independent correspond to an antichain in the poset and vice versa.

Let $\mathscr{I}(\mathscr{A}_n)$ denote the number of vertices in the largest strongly independent subset of $\mathscr{A}_n$. Then in this paper we prove that with probability tending to 1, the sequence $\mathscr{I}(\mathscr{A}_n)$ satisfies:

$$\frac{\mathscr{I}(\mathscr{A}_n)}{\sqrt{\log n}} \to \frac{\sqrt{2}}{\sqrt{\log 1/q}} \quad \text{as } n \to \infty.$$

The applications of these results could be in the fields of operation research, scheduling theory and parallel computation, since several problems which may be formulated in terms of acyclic directed graphs have solutions which are specified by the maximal number of strongly independent vertices.

* Received by the editors April 9, 1981, and in final revised form September 6, 1983.

† Department of Computer Science, The Hebrew University of Jerusalem, Jerusalem 91904, Israel.

‡ Institute of Mathematics, Hungarian Academy of Sciences, Budapest, Hungary.

We note that random (undirected) graphs of the kind used in this paper were investigated in connection with cliques [1], coloring [3] and complete subgraphs [4]. Random graphs of a slightly different kind were investigated in detail in [2].

2. Strongly independent vertex sets. In this section we find lower and upper bounds for k, the number of vertices in strongly independent subsets of $\mathcal{A}_n$.

Lower bound. Consider the following subsets of k *consecutive* vertices in $\mathcal{A}_n$: $\{1, 2, \cdots, k\}$, $\{k+1, k+2, \cdots, 2k\}$, $\{2k+1, 2k+2, \cdots, 3k\}, \cdots$. Then the number of these subsets is $\lceil n/k \rceil$. The probability that a subset is independent is $q^{\binom{k}{2}}$. Note that in this case the subset is also strongly independent. The probability that a subset is not independent is $1-q^{\binom{k}{2}}$. Also, the probability that none of the subsets are independent is:

$$(1-q^{\binom{k}{2}})^{\lceil n/k \rceil} \approx (1-q^{\binom{k}{2}})^{n/k}.$$

Since $1-x \leqq e^{-x}$ if $x \geqq 0$, we have

$$(1-q^{\binom{k}{2}})^{n/k} \leqq \exp(-q^{\binom{k}{2}} n/k).$$

This probability tends to zero if

$$q^{\binom{k}{2}} n/k \to \infty \quad \text{as } n \to \infty,$$

which is implied if

$$\log n - \log k + \binom{k}{2} \log q \to \infty \quad \text{as } n \to \infty,$$

or

$$\log n - \log k - \frac{k(k-1)}{2} \log \frac{1}{q} \to \infty \quad \text{as } n \to \infty.$$

Let

$$k = \lfloor K_n - \varepsilon \rfloor,$$

where ε is a positive constant and

$$K_n = \sqrt{\frac{2 \log n}{\log(1/q)} + \frac{1}{4}} + \frac{1}{2}. \tag{2.1}$$

Then

$$\log n - \frac{K_n(K_n-1)}{2} \log \frac{1}{q} = 0,$$

therefore

$$\log n - \log k - \frac{k(k-1)}{2} \log \frac{1}{q} \geqq \varepsilon K_n \log \frac{1}{q} - \frac{\varepsilon^2+\varepsilon}{2} \log \frac{1}{q} - \log(K_n - \varepsilon) \to \infty \quad \text{as } n \to \infty,$$

for every fixed value $\varepsilon > 0$. We have proved:

THEOREM 1. *Let $\mathcal{A}_n$ be a random acyclic directed graph. Then the probability that $\mathcal{A}_n$ has no strongly independent vertex subset of size $k < K_n$ tends to zero as $n \to \infty$.*

Upper bound. Let $a_1 < a_2 < \cdots < a_k$ be a subset of k vertices in $\mathscr{A}_n$ and let $E_p(n, k)$ denote the expectation for the number of subsets with k strongly independent vertices in $\mathscr{A}_n$.

The strategy of the proof is to consider four different cases of distances between a_k and a_1. First, we assume $a_k - a_1 > k^4$; next we consider $Ck \log k \leqq a_k - a_1 \leqq k^4$ where C is a constant. In the third case we consider $a_k - a_1 < Ck \log k$ and finally in the fourth case $a_k - a_1 \leqq Mk$, where M is a positive constant. In each case we prove that the probability that there is a strongly independent vertex subset of size $k \geqq K_n$, tends to zero as $n \to \infty$.

Case 1. Let $a_k - a_1 > k^4$. In this case the number of subsets of k vertices in $\mathscr{A}_n$ is bounded by $\binom{n}{k}$. The probability that each subset is strongly independent is bounded by the product of the probability that there is no directed path of length 1 from a_k to a_1 and the probability that there is no directed paths of length 2 from a_k to a_1, through at least $k^4 - k$ vertices which are not in the subset. These probabilities are q and $1 - p^2$ respectively. Therefore,

$$E_p(n, k) \leqq \binom{n}{k} q(1-p^2)^{k^4-k} \leqq \frac{n^k}{k!} q(1-p^2)^{k^4-k} \leqq n^k (1-p^2)^{k^4-k}.$$

This expectation tends to zero as $n \to \infty$ if

$$k \log n + (k^4 - k) \log (1-p^2) \to -\infty \quad \text{as } n \to \infty.$$

Since

$$\log (1-p^2) = -|\log (1-p^2)|,$$

we must have

$$(k^3 - 1)|\log (1-p^2)| - \log n \to \infty \quad \text{as } n \to \infty,$$

which is satisfied if

$$k > \sqrt[3]{\frac{\log n}{|\log (1-p^2)|} + 1}.$$

Conclusion. If $a_k - a_1 > k^4$, then even for values of k which are smaller than K_n the probability that a_1 and a_k are strongly independent tends to zero as $n \to \infty$.

Case 2. Let $Ck \log k \leqq a_k - a_1 \leqq k^4$ where C is a constant. First, we find a bound for the possible number of different subsets of k vertices. Clearly, for each vertex of $\mathscr{A}_n$ there are at most k^4 different subsets, from which we can choose k vertices. As a result, the number of subsets with k verices in $\mathscr{A}_n$ is bounded by $\binom{k^4}{k} nk^4$.

Next, we find the probability that each subset is strongly independent. This probability is bounded by the product of the probability that the subset is independent and the probability that there is no directed path of length 2 from a_k to a_1, through any vertex j which is not in the subset $\{a_1, a_2, \cdots, a_k\}$, for $a_1 < j < a_k$.

The expectation for the number of strongly independent vertex subsets of size k is bounded in this case by:

$$E_p(n, k) \leqq \binom{k^4}{k} nk^4 q^{\binom{k}{2}} (1-p^2)^{Ck \log k - k}$$

$$\leqq k^{4k} n q^{k(k-1)/2} (1-p^2)^{Ck \log k - k}.$$

This expectation tends to zero as $n \to \infty$ if

$$4k \log k + \log n - \frac{k(k-1)}{2} \log \frac{1}{q} + (k - Ck \log k)|\log (1-p^2)| \to -\infty \quad \text{as } n \to \infty.$$

Let $k \geqq K_n$, where K_n is defined in (2.1). Then

$$\log n - \frac{k(k-1)}{2} \log \frac{1}{q} \leqq 0.$$

Thus

$$4k \log k + (k - Ck \log k)|\log (1-p^2)| \to -\infty \quad \text{as } n \to \infty,$$

provided that

$$C > \frac{4}{|\log (1-p^2)|}. \tag{2.2}$$

Conclusion. If $Ck \log k \leqq a_k - a_1 \leqq k^4$, where C is defined in (2.2), then the probability that there is a strongly independent vertex subset of size $k \geqq K_n$ tends to zero as $n \to \infty$.

Case 3. Let $a_k - a_1 < Ck \log k$, where $C > 4/|\log (1-p^2)|$.

(a) Suppose that the interval between the first r vertices and the last r vertices in the subset $\{a_1, a_2, \cdots, a_n\}$ includes at least $(1+\alpha)k$ vertices, where α is a positive constant. In other words, we assume that

$$a_{k-r+1} - a_r \geqq (1+\alpha)k,$$

of which clearly, at least αk vertices of $\mathscr{A}_n$ are not in the subset $\{a_1, a_2, \cdots, a_k\}$.

This subset is strongly independent if it is independent and for each pair (a_i, a_{k-i+1}) for $i = 1, 2, \cdots, r$, there is no directed path of length 2 from a_i to a_{k-i+1} through αk vertices j, $a_r < j < a_{k-r+1}$, which are not in the subset.

The expectation $E_p(n, k)$ is bounded in this case by:

$$E_p(n,k) \leqq \binom{Ck \log k}{k} nCk \log k \, q^{\binom{k}{2}} (1-p^2)^{\alpha r k}$$

$$\leqq (Ck \log k)^k n q^{k(k-1)/2} (1-p^2)^{\alpha r k}.$$

This expectation tends to zero as $n \to \infty$ if

$$k \log (Ck \log k) + \log n - \frac{k(k-1)}{2} \log \frac{1}{q} - \alpha r k |\log (1-p^2)| \to -\infty \quad \text{as } n \to \infty.$$

Suppose that we choose $k \geqq K_n$. Then

$$\log n - \frac{k(k-1)}{2} \log \frac{1}{q} \leqq 0,$$

and

$$k(\log C + \log k + \log \log k - \alpha r |\log (1-p^2)|) \to -\infty \quad \text{as } n \to \infty,$$

provided that

$$\alpha r |\log (1-p^2)| > \log k + \log \log k.$$

It is therefore sufficient to choose a value of r such that

$$r \geqq (\log k)^{1+\sigma}, \tag{2.3}$$

where σ is a positive constant.

(b) Suppose that the conditions of (a) are not satisfied i.e., if $r \geqq (\log k)^{1+\sigma}$, where σ is a positive constant, then

$$a_{k-r+1} - a_r \leqq (1+\alpha)k,$$

for every positive value of α. Suppose however, that $a_r - a_1 \geqq r + Qk$ or $a_k - a_{k-r+1} \geqq r + Qk$, where Q is a positive constant to be defined. Then the subset $\{a_1, a_2, \cdots, a_n\}$ is strongly independent if it is independent and there is no directed path of length 2 from a_k to a_1, through at least Qk vertices of $\mathscr{A}_n$ which are not in the subset.

The expectation $E_p(n, k)$ is bounded in this case by:

$$E_p(n, k) \leqq \binom{Ck \log k}{2r} \binom{(1+\alpha)k}{k} nq^{\binom{k}{2}} (1-p^2)^{Qk}.$$

Note that as $n \to \infty$, for every $0 < \alpha < 1$ we can choose a positive constant β such that

$$\binom{(1+\alpha)k}{k} \leqq (1+\beta)^k. \tag{2.4}$$

Thus

$$E_p(n, k) \leqq (Ck \log k)^{2r} (1+\beta)^k nq^{k(k-1)/2} (1-p^2)^{Qk}.$$

This expectation tends to zero as $n \to \infty$ if

$$2r \log (Ck \log k) + k \log (1+\beta) + \log n - \frac{k(k-1)}{2} \log \frac{1}{q} - Qk|\log (1-p^2)| \to -\infty.$$

Suppose that we choose $k \geqq K_n$. Then

$$Qk|\log (1-p^2)| - 2r \log (Ck \log k) - k \log (1+\beta) \to \infty \quad \text{as } n \to \infty,$$

provided that

$$Q > \frac{\log (1+\beta)}{|\log (1-p^2)|}, \tag{2.5}$$

where β is a given positive constant.

Conclusion. If $a_k - a_1 < Ck \log k$ where $C > 4/|\log (1-p^2)|$, and $a_{k-r+1} - a_r \geqq (1+\alpha)k$, where r is defined in (2.3) and α is a positive constant, or $a_{k-r+1} - a_r \leqq (1+\alpha)k$ and $a_r - a_1 \geqq r + Qk$ or $a_k - a_{k-r+1} \geqq r + Qk$, where Q is defined in (2.5), then the probability that there is a strongly independent vertex subset of size $k \geqq K_n$ tends to zero as $n \to \infty$.

Case 4. Let $a_k - a_1 \leqq Mk$, where M is a positive constant to be defined. Suppose that $a_r - a_1 \leqq r + Qk$, and $a_k - a_{k-r+1} \leqq r + Qk$, where r and Q are defined in (2.3) and (2.5) respectively and that $a_{k-r+1} - a_r \leqq (1+\alpha)k$ for every value of $\alpha > 0$. Then

$$a_k - a_1 \leqq 2r + 2Qk + (1+\alpha)k < k + 2Qk + (1+\alpha)k = k(2 + 2Q + \alpha).$$

Thus $M = 2 + 2Q + \alpha$.

Definition. Let $a_1 < a_2 < \cdots < a_k$ be a subset of k vertices in $\mathscr{A}_n$ such that $a_k - a_1 \leqq Mk$. Then the subset is called *nearly consecutive.*

THEOREM 2. *Let $\mathcal{A}_n$ be a random acyclic directed graph and let K_n be defined in* (2.1). *Then the probability that $\mathcal{A}_n$ has a strongly independent vertex subset of size $k \geqq K_n$ tends to zero as $n \to \infty$.*

Proof. First, note that by the previous conclusions, it is sufficient to consider only nearly consecutive subsets.

The expectation for the number of subsets with k strongly independent vertices in $\mathcal{A}_n$ is bounded in this case by

$$E_p(n, k) \leqq \binom{(1+\alpha)k}{k}\left[\binom{Qk+r}{r}\right]^2 nq^{\binom{k}{2}}.$$

Using (2.4) for some $\beta > 0$, we get

$$E_p(n, k) \leqq (1+\beta)^k (Qk+r)^{2r} nq^{k(k-1)/2}.$$

This expectation tends to zero as $n \to \infty$ if

$$k \log(1+\beta) + 2r \log(Qk+r) + \log n - \frac{k(k-1)}{2} \log \frac{1}{q} \to -\infty \quad \text{as } n \to \infty.$$

Let

$$k = \lceil K_n + \delta \rceil,$$

where δ is a positive constant. Then

$$\begin{aligned} &k \log(1+\beta) + 2r \log(Qk+r) + \log n - \frac{k(k-1)}{2} \log \frac{1}{q} \\ &\quad \leqq (K_n+\delta) \log(1+\beta) + 2r \log(QK_n + Q\delta + r) - \frac{\delta^2 - \delta}{2} \log \frac{1}{q} - \delta K_n \log \frac{1}{q}. \end{aligned}$$

Thus, the expectation tends to zero as $n \to \infty$ provided that

$$\delta \log \frac{1}{q} > \log(1+\beta).$$

Given $\delta > 0$, it is now sufficient to choose $\beta > 0$ so that

$$\delta > \frac{\log(1+\beta)}{\log 1/q}.$$

and the theorem is proved.

3. A maximal strongly independent vertex set. We now prove the main result of this paper.

THEOREM 3. *Let $\mathcal{A}_n$ be a random acyclic directed graph with vertex set $\{1, 2, \cdots, n\}$ and let $\mathcal{I}(\mathcal{A}_n)$ be the number of vertices in the largest (maximal) stongly independent subset of $\mathcal{A}_n$. Then with probability tending to 1, the sequence $\mathcal{I}(\mathcal{A}_n)$ satisfies:*

$$\frac{\mathcal{I}(\mathcal{A}_n)}{\sqrt{\log n}} \to \frac{\sqrt{2}}{\sqrt{\log 1/q}} \quad as\ n \to \infty.$$

Proof. Let

$$K_n = \sqrt{\frac{2 \log n}{\log 1/q} + \frac{1}{4}} + \frac{1}{2}.$$

Then by Theorem 1:

$$P\{\mathscr{I}(\mathscr{A}_n) < \lfloor K_n - \varepsilon \rfloor\} \to 0 \quad \text{as } n \to \infty,$$

or

$$P\{\mathscr{I}(\mathscr{A}_n) \geqq \lfloor K_n - \varepsilon \rfloor\} \to 1 \quad \text{as } n \to \infty, \tag{3.1}$$

for every $\varepsilon > 0$, and by Theorem 2,

$$P\{\mathscr{I}(\mathscr{A}_n) \geqq \lceil K_n + \delta \rceil\} \to 0 \quad \text{as } n \to \infty,$$

or

$$P\{\mathscr{I}(\mathscr{A}_n) < \lceil K_n + \delta \rceil\} \to 1 \quad \text{as } n \to \infty. \tag{3.2}$$

for every $\delta > 0$.

From (3.1), (3.2) and the Borel–Cantelli lemma follows that as $n \to \infty$

$$\limsup_{n\to\infty} \frac{\mathscr{I}(\mathscr{A}_n)}{\sqrt{\log n}} \leqq \frac{\sqrt{2}}{\sqrt{\log 1/q}},$$

and

$$\liminf_{n\to\infty} \frac{\mathscr{I}(\mathscr{A}_n)}{\sqrt{\log n}} \geqq \frac{\sqrt{2}}{\sqrt{\log 1/q}},$$

and the theorem follows.

COROLLARY. *Suppose that the interval $[K_n - \varepsilon, K_n + \varepsilon]$ does not include an integer, i.e., for every integer I, $|K_n - I| > \varepsilon$, for some $1 > \varepsilon > 0$.*

Then

$$P\{\mathscr{I}(\mathscr{A}_n) = \lfloor K_n \rfloor\} \to 1 \quad \text{as } n \to \infty.$$

We note that there is an absolute constant D $(0 < D < 1)$, so that the probability that one of the maximal strongly independent vertex subsets is consecutive, is greater than D. If K_n is not close to an integer, this probability tends to 1, but if K_n is close to an integer then this probability does not tend to 1.

Finally, it is interesting to note that although the obtained results are asymptotic, they hold even for small values of n. For example, for $p = 0.5$, $n = 10$, and $K_n = 3.13$, less than 8% of a sample of random acyclic directed graphs had a maximal strongly independent vertex subset larger than 3.

REFERENCES

[1] B. BOLLOBAS AND P. ERDÖS, *Cliques in random graphs*, Math. Proc. Camb. Phil. Soc., 80 (1976), pp. 419–427.

[2] P. ERDÖS AND A. RENYI, *On the evolution of random graphs*, Publ. Math. Inst. Hung. Acad. Sci., 5A (1960), pp. 17–61.

[3] G. R. GRIMMETT AND C. J. H. MCDIARMID, *On colouring random graphs*, Math. Proc. Comb. Phil. Soc., 77 (1975), pp. 313–324.

[4] D. W. MATULA, *On the complete subgraphs of a random graphs*, Proc. 2nd Conference on Combinatory Theory and Applications, Chapel Hill, NC 1970, pp. 356–369.