Discrete Mathematics 46 (1983) 145–150
North-Holland

TREES IN RANDOM GRAPHS

P. ERDÖS

Mathematics Institute of the Hungarian Academy of Sciences, Budapest, Hungary

Z. PALKA

Adam Mickiewicz University, Poznań, Poland

Received 7 June 1982
Revised 7 July 1982

We show that for every $\varepsilon>0$ almost every graph $G\in\mathcal{G}(n,p)$ is such that if

$$(1+\varepsilon)\frac{\log n}{\log d}<r<(2-\varepsilon)\frac{\log n}{\log d}$$

where $d=1/q$, then G contains a maximal induced tree of order r.

1. Introduction

Let us consider the probability space $\mathcal{G}(n,p)$ consisting of all graphs on n labeled vertices where each edge occurs with probability $p=1-q$, independently of all other edges. The aim of this note is to find such natural numbers which are likely to occur as orders of maximal induced trees contained in a graph $G\in\mathcal{G}(n,p)$ when $0<p<1$ is fixed. By a maximal induced tree we mean an induced tree which is not properly contained in any other tree.

A similar problem devoted to maximal complete subgraphs of G was considered by Bollobás and Erdös [2], who showed that for every $\varepsilon>0$ almost every (a.e.) graph $G\in\mathcal{G}(n,p)$ is such that if

$$(1+\varepsilon)\frac{\log n}{\log b}<r<(2-\varepsilon)\frac{\log n}{\log b}$$

where $b=1/p$, then G contains a clique of order r. The largest integer r for which a.e. graph $G\in\mathcal{G}(n,p)$ contains a topological complete r-graph was derived by Bollobás and Catlin [1]. Let us remark that some bounds of the orders of maximal induced trees in a graph $G\in\mathcal{G}(n,p)$ for $p>0.06$ have already been given by Karoński and Palka (see [4, 5]).

In Section 2 we give an upper bound for the order of an induced star in a random graph. This result (which may have interest on its own) is used in proving the main theorems presented in Section 3. An open problem with a discussion is given in the last part of this paper.

2. A lemma

Here we will consider the existence of an induced $(1, r)$-tree in a graph $G \in \mathcal{G}(n, p)$. By a $(1, r)$-tree we mean a complete bipartite graph $K_{1,r}$ which has two vertex classes of 1 and r vertices, respectively (such a graph is often called a star). Let the vertex from the first class be called the root of the star. To simplify the notation we shall put $b = 1/p$ and $d = 1/q$. The following lemma will be useful in proving our main results given in Section 3.

Lemma. *For every $\varepsilon > 0$ and $2 \leq r \leq (2-\varepsilon)(\log n)/(\log d)$ a.e. graph $G \in \mathcal{G}(n, p)$ contains an induced $(1, r)$-tree.*

Proof. Let X_r denote the number of induced $(1, r)$-trees in a graph $G \in \mathcal{G}(n, p)$. The expectation of X_r is

$$E_r = E(X_r) = n\binom{n-1}{r} p^r q^{\binom{r}{2}}.$$

To find the second moment of X_r, which is the sum of the probabilities of ordered pairs of $K_{1,r}$ we have to consider two different situations. First let us assume that two $K_{1,r}$'s have the same root and vertices from the second classes have l $(0 \leq l \leq r)$ common elements. The probability of such event is

$$p_1(l) = p^{2r-l} q^{2\binom{r}{2} - \binom{l}{2}}.$$

Further, one can choose

$$a_1(l) = n\binom{n-1}{r}\binom{r}{l}\binom{n-1-r}{r-l}$$

ordered pairs of such $K_{1,r}$'s. Secondly, two $(1, r)$-trees can have different roots. Then the following three possibilities should be taken into the consideration:

(i) The roots are not connected by an edge and vertices from the second classes have l $(0 \leq l \leq r)$ common elements; there are

$$a_2(l) = 2\binom{n}{2}\binom{n-2}{r}\binom{r}{l}\binom{n-2-r}{r-l}$$

ordered pairs of such $K_{1,r}$'s and the probability of each is

$$p_2(l) = p^{2r} q^{2\binom{r}{2} - \binom{l}{2}}.$$

(ii) The roots are connected and the edge joining them belongs to one of $K_{1,r}$'s; there are

$$a_3 = 2\binom{n}{2}\binom{n-2}{r}\binom{n-2-r}{r-1}$$

ordered pairs of such $K_{1,r}$'s and the probability of each is

$$p_3 = p^{2r} q^{2\binom{r}{2}}.$$

(iii) The roots are connected and the edge joining them belongs to both $K_{1,r}$'s; there are

$$a_4 = 2\binom{n}{2}\binom{n-2}{r-1}\binom{n-1-r}{r-1}$$

ordered pairs of such $K_{1,r}$'s and the probability of each is

$$p_4 = p^{2r-1} q^{2\binom{r}{2}}.$$

Therefore

$$E(X_r^2) = a_3 p_3 + a_4 p_4 + \sum_{l=0}^{r} [a_1(l)p_1(l) + a_2(l)p_2(l)]$$

$$\leq a_2(0)p_2(0)\left[1+\mathrm{O}\left(\frac{1}{n}\right)\right] + \sum_{l=1}^{r} a_1(l)p_2(l)[b^l + n].$$

Thus, denoting the variance of X_r by σ_r^2 we have for sufficiently large n

$$\frac{\sigma_r^2}{E_r^2} = \frac{E(X_r^2)}{E_r^2} - 1 \leq \mathrm{o}(1) + \sum_{l=1}^{r} \frac{\binom{r}{l}\binom{n-1-r}{r-l}}{\binom{n-1}{r}} d^{l(l-1)/2}(b^l n^{-1} + 1)$$

$$\leq \mathrm{o}(1) + \sum_{l=1}^{r} r^{2l} n^{-l} b^l d^{l(l-1)/2} = \mathrm{o}(1) + \sum_{l=1}^{r} F_l.$$

Now if n is sufficiently large and $2 \leq l \leq r-1$, then

$$F_l < F_2 + F_{r-1}.$$

Consequently

$$\mathrm{Prob}(X_r = 0) < \sigma_r^2 / E_r^2 < F_1 + F_r + r(F_2 + F_{r-1}) = \mathrm{o}(1)$$

for all $2 \leq r \leq (2-\varepsilon)(\log n)/(\log d)$ and large n. This completes the proof of the lemma.

Let us see that only one more step is necessary to show that the largest order of an induced star in a.e. graph $G \in \mathscr{G}(n, p)$ is

$$2\frac{\log n}{\log d} + \mathrm{o}(\log n).$$

As a matter of fact,

$$\mathrm{Prob}(X_r \geq 1) \leq E(X_r) = \mathrm{o}(1) \quad \text{for all } r \geq (2+\varepsilon)\frac{\log n}{\log d}.$$

3. Main results

Let $t(G)$ denote the order of the smallest maximal induced tree of a graph G.

Theorem 1. *For every $\varepsilon>0$ a.e. graph $G\in\mathcal{G}(n,p)$ satisfies*

$$(1-\varepsilon)\frac{\log n}{\log d}<t(G)<(1+\varepsilon)\frac{\log n}{\log d}.$$

Proof. Let Y_i denote the number of maximal induced trees of order i in a graph $G\in\mathcal{G}(n,p)$. Let

$$k=(1-\varepsilon)\frac{\log n}{\log d}.$$

Then

$$\mathrm{Prob}\left\{t(G)\leqslant(1-\varepsilon)\frac{\log n}{\log d}\right\}=\mathrm{Prob}\left\{\bigcup_{i=1}^{k}(Y_i>0)\right\}\leqslant\sum_{i=1}^{k}E(Y_i).$$

Now, for any $1\leqslant i\leqslant k$ and sufficiently large n we have

$$\begin{aligned}E(Y_i)&=\binom{n}{i}(1-ipq^{i-1})^{n-i}i^{i-2}p^{i-1}q^{(i-1)(i-2)/2}\\&\leqslant\frac{n^i}{i!}\exp[-(n-i)ipq^{i-1}]i^i\\&\leqslant\{n\exp[-npq^{i-1}+ipq^{i-1}+1]\}^i\\&\leqslant\{n\exp[-npq^{k-1}+2]\}^i<n^{-\varepsilon i}.\end{aligned}$$

Thus

$$\mathrm{Prob}\left\{t(G)\leqslant(1-\varepsilon)\frac{\log n}{\log d}\right\}=o(1)$$

which proves the left hand side of the desired inequality. Now we show that a.e. graph $G\in\mathcal{G}(n,p)$ contains a maximal induced tree of order less than $(1+\varepsilon)$ $(\log n)/(\log d)$. From our Lemma we can deduce that a.e. graph $G\in\mathcal{G}(n,p)$ contains at least one induced $(1,r)$-tree, where

$$r=\frac{\log n}{\log d}+\frac{(1+\gamma)\log\log n}{\log d}, \tag{1}$$

and $\gamma>0$ is a constant. It is easy to see that this tree is the maximal tree. As a matter of fact, the probability that there is a vertex in the graph G connected with exactly one vertex belonging to the tree is at least

$$(n-r-1)(r+1)pq^r\leqslant(\log n)^{-\gamma}(1+o(1)),$$

when r is given by (1). This completes the proof of the theorem.

Now, let $T(G)$ denote the order of the largest induced tree of a graph G. Then the following result holds.

Theorem 2. *For every $\varepsilon > 0$ a.e. graph $G \in \mathscr{G}(n, p)$ satisfies*

$$(2-\varepsilon)\frac{\log n}{\log d} < T(G) < (2+\varepsilon)\frac{\log n}{\log d}.$$

Proof. The left hand side of above inequality follows immediately from our Lemma. Now let Z_k denote the number of induced trees of order k. Let us take

$$k = (2+\varepsilon)\frac{\log n}{\log d}. \tag{2}$$

Then

$$E(Z_k) = \binom{n}{k} k^{k-2} p^{k-1} q^{(k-1)(k-2)/2}$$
$$\leq n^k e^k p^{k-1} q^{(k-1)(k-2)/2} < (c\, n^{-\varepsilon/2})^k$$

where c is a constant. Thus a.e. graph $G \in \mathscr{G}(n, p)$ contains no induced tree of order k given by (2).

Since the largest tree is at the same time the largest maximal tree, so we can formulate the following corollary of Theorems 1 and 2.

Corollary. *Given $\varepsilon > 0$ a.e. graph $G \in \mathscr{G}(n, p)$ is such that if*

$$(1+\varepsilon)\frac{\log n}{\log d} < r < (2-\varepsilon)\frac{\log n}{\log d},$$

then G contains a maximal induced tree of order r, but G does not contain a maximal induced tree of order less than $(1-\varepsilon)(\log n)/(\log d)$ or greater than $(2+\varepsilon)(\log n)/(\log d)$.

4. An open problem

Up to now the edge probability p was fixed. Now, let p be a function on n, i.e., $p = p(n)$ and tends to zero as $n \to \infty$. The following open problem is worth considering.

Problem. Find such a value of the edge probability p for which a graph $G \in \mathscr{G}(n, p)$ has the greatest induced tree.

As a comment to this problem let us notice that Erdös and Rényi have shown [3] that if Δ denotes the number of vertices of the greatest tree contained in a graph $G \in \mathcal{G}(n, p)$, then for $p = 1/n$

$$\lim_{n \to \infty} \mathrm{Prob}\{\Delta \geqslant n^{2/3}\omega(n)\} = 0$$

and

$$\lim_{n \to \infty} \mathrm{Prob}\{\Delta \geqslant n^{2/3}/\omega(n)\} = 1$$

where $\omega(n)$ is a sequence tending arbitrarily slowly to infinity. We are sure that for $p = c/n$, where $c > 1$ is a constant, a graph $G \in \mathcal{G}(n, p)$ contains a tree of order $n^{1-\varepsilon}$ ($\varepsilon > 0$ is a constant) but we also conjecture more, namely that $G \in \mathcal{G}(n, c/n)$ contains a tree of order $\gamma(c)n$, where $\gamma(c)$ depends only on c.

References

[1] B. Bollobás and P. Catlin, Topological cliques of random graphs, J. Combin. Theory (Ser. B) (1981) 223–227.

[2] B. Bollobás and P. Erdös, Cliques in random graphs, Math. Proc. Cambridge Philos. Soc. 80 (1976) 419–427.

[3] P. Erdös and A. Rényi, On the evolution of random graphs, Publ. Math. Inst. Hung. Acad. Sci. 5A (1960) 17–61.

[4] M. Karoński and Z. Palka, On the size of a maximal induced tree in a random graph, Math. Slovaca 30 (1980) 151–155.

[5] M. Karoński and Z. Palka, Addendum and erratum to the paper "On the size of a maximal induced tree in a random graph", Math. Slovaca 31 (1981) 107–108.