COMMENTATIONES MATHEMATICAE UNIVERSITATIS CAROLINAE

28,1 (1987)

LARGE TREES IN RANDOM GRAPHS

Luděk KUČERA and Vojtěch RÖDL

Abstract: The aim of this note is to estimate the size of the largest tree which occurs as an induced subgraph in a random graph G(n,p). We give some upper bounds and lower bounds for all values of p ($p \gtrsim c/n$, $c>1$) and in particular we give a positive answer to a question of P. Erdös and Z. Palka showing that if $p \sim c/n$ for some fixed $c>1$ then the size of the largest tree is at least qn where q is a positive constant depending on c only.

Key words: Random graphs, induced tree.

Classification: 05C80

Introduction. Consider the probability space $\mathcal{G}(n,p)$ consisting of all graphs on n labelled vertices, where each edge is chosen with probability p, independently of all others. In this note we investigate how large trees occur as induced subgraphs in almost all $G \in \mathcal{G}(n,p)$. In the paragraph 1 we investigate the size of maximal trees (i.e. such trees that are not contained in any larger induced tree). We show that maximal trees have to be either quite small or very large, i.e. there is an interval in which the size of the maximal tree will almost never occur. In the paragraph 3 we first prove (for $p=o(1)$ the existence of some large induced trees in $\mathcal{G}(n,p)$ (Lemma 3.2) and this together with the result of the paragraph 2 implies the main result of the paragraph 3 - Theorem 3.3 giving the lower bound for the cardinality of the tree that almost surely occurs as an induced subgraph of $\mathcal{G}(n,p)$.

In the paragraph 3 we use the Chernoff inequality ([1], see also [3]) which we state here:

Let m be a positive integer and $1>q>0$, $1>\delta>0$ two reals. Let $k \le qm(1-\delta)$ then

$$\sum_{j \le k} \binom{m}{j} q^j (1-q)^{m-j} \le \exp\{m[(1-q+\delta q)\log\frac{1-q}{1-q+\delta q} + q(1-\delta)\log\frac{1}{1-\delta}]\}.$$

Easy calculations give that if $\delta < 1/2$ then we have also

$$(0)\quad \sum_{j \le k} \binom{m}{j} q^j (1-q)^{m-j} \le \exp\{-mq\delta^2(\tfrac{q}{2}+\delta)\} \le \exp\{-mw\delta^3\}.$$

1. Upper bound

Theorem 1.1. Let $\varepsilon > 0$ be a fixed real number. If n,k are natural numbers and p is a real number such that $1/n < p < 1$ and

$$(1)\quad k \ge \frac{2(\log(nep))}{\log \frac{1}{1-p}} + (1+\varepsilon)\frac{\log n}{\log(nep)} + 3$$

then $\mathcal{G}(n,p)$ contains almost surely no induced tree on k vertices.

Proof: Let A be a fixed k-element subset of the vertex set U. Denote by $\tilde{A}$ the event that the subgraph of $\mathcal{G}(n,p)$ induced on A is a tree. Then clearly

$$\mathrm{Prob}(\tilde{A}) = k^{k-2}p^{k-1}(1-p)^{\binom{k}{2}-(k-1)}$$

and thus

$$\mathrm{Prob}(\tilde{A} \text{ holds for some } A \subset V,\ |A|=k) \le \binom{n}{k}k^{k-2}p^{k-1}(1-p)^{\binom{k}{2}-(k-1)} <$$

$$< \frac{1}{p}\binom{n}{k}k^k p^k (1-p)^{\frac{(k-3)k}{2}} \le n[nep(1-p)^{(k-3)/2}]^k.$$

If k satisfies (1) then

$$\frac{k-3}{2} \ge \frac{\log(nep)}{\log\frac{1}{1-p}} + \frac{(1+\varepsilon)}{2}\frac{\log n}{\log(nep)}$$

and thus

$$(1-p)^{(k-3)/2} \le \exp\left[-\log(nep) - \frac{1+\varepsilon}{2}\,\frac{\log n}{\log(nep)}\log\frac{1}{1-p}\right].$$

Hence

$$n[nep(1-p)^{(k-3)/2}]^k \le n\exp\left(-\frac{1+\varepsilon}{2}\,k\,\frac{\log n}{\log(nep)}\log\frac{1}{1-p}\right) \le$$

$$\le n\exp(-(1+\varepsilon)\log n) = n^{-\varepsilon}.$$ Q.E.D.

Elementary calculation immediately gives:

Corollary 1.2. Let $\delta > 0$, then $\mathcal{G}(n,p)$ contains almost surely no induced subtree with more than $\frac{2(1+\delta)\log(nep)}{p}$ vertices.

Note that the bound of Theorem 1.1 is not satisfactory if p is small, the most interesting such case is if $p=c/n$ and $c>1$ is a small constant. The slight modification of the above proof however gives the following

Theorem 1.3. Let c, α be real numbers such that $0<\alpha<1<c$ and

(2) $$\frac{c\exp(-c\alpha/2)}{(1-\alpha)^{(1-\alpha)/\alpha}}<1 .$$

If n is a natural number and $p=c/n$ then $\mathcal{G}(n,p)$ contains almost surely no induced tree with $k \geq \alpha n$ vertices.

Proof: Set γ for the LHS of (2). Then, similarly as above the required probability is bounded by

$$\binom{n}{k} k^{k-2} p^{k-1} (1-p)^{\binom{k-1}{2}} \leq \frac{1}{k^2 p} \frac{n^k n^{n-k}}{k^k (n-k)^{n-k}} k^k p^k (1-p)^{k(k-3)/2} \leq$$

$$\leq \frac{1}{\alpha^2 cn} \left[\frac{c\exp(-\alpha c/2)(1-p)^{-3/2}}{(1-\alpha)^{(1-\alpha)/\alpha}}\right]^{\alpha n} \leq (\gamma(1-p)^{-3/2})^{\alpha n} \longrightarrow 0,$$

if $\gamma<1$. Q.E.D.

Note that

(3) $$\frac{c\exp(-\alpha c/2)}{(1-\alpha)^{(1-\alpha)/\alpha}} \leq \frac{2}{\alpha(1-\alpha)^{(1-\alpha)/\alpha}} \frac{c\alpha}{2} \exp\left(-\frac{c\alpha}{2}\right) \leq \frac{2}{e\alpha(1-\alpha)^{(1-\alpha)/\alpha}} .$$

The RHS of (3) is a decreasing function of α and tends to $2/e$ with $\alpha \longrightarrow 1$. This means that if α_0 is the (only) root of the equation $e\alpha(1-\alpha)^{(1-\alpha)/\alpha}=2$ then there is no induced tree of $k=n$ vertices if $\mathcal{G}(n,\frac{c}{n})$, for any $\alpha>\alpha_0$ and $c>1$.

2. Lower bound

Theorem 2.1. Let $c>1$ and $\delta>0$ be fixed real numbers. If n is a natural number and μ, ε, p are real numbers such that $p \geq c/n$, $0<\varepsilon<\mu-1$ and $\log np > \mu \log\log(nep)$ then the size k of an arbitrary maximal induced subtree of $\mathcal{G}(n,p)$ satisfies almost surely one of the following conditions:

either

$$k<\frac{(2+\delta)\log n}{\varepsilon \log pn} = k_0$$

or

(4) $$\operatorname{Min}(k_1,k_2) \leq k \leq \frac{2(\log(nep))}{\log\frac{1}{1-p}} + (1+\delta)\frac{\log n}{\log(nep)} + 3$$

where

(5) $\quad k_1 = \dfrac{\log np - \mu \log\log np}{\log \frac{1}{1-p}}$

and

(6) $\quad k_2 = \dfrac{(\mu - 1 - \varepsilon)\log np}{1 + \mu \log np}\, n$.

Proof: Let A be a set of k vertices of $\mathcal{G}(n,p)$ and let $v \notin A$. The probability that v is joined to exactly one vertex of A is $kp(1-p)^{k-1}$ and therefore the probability that no such vertex exists is $(1-kp(1-p)^{k-1})^{n-k}$.
Hence, the probability that A spans a maximal induced subtree is exactly

$p_A = k^{k-2}p^{k-1}(1-p)^{\binom{k}{2}-(k-1)}(1-kp(1-p)^{k-1})^{n-k}$.

The probability P_k that there exists a k-element set A that spans a maximal induced tree is therefore bounded from above by

$P_k \leq \sum_A p_A = \binom{n}{k}k^{k-2}p^{k-1}(1-p)^{\binom{k}{2}-(k-1)}(1-kp(1-p)^{k-1})^{n-k} \leq$

$\leq \frac{1}{p}(\frac{ne}{k})^k k^k p^k \exp(-(n-k)pk(1-p)^k) \leq$

(7) $\leq n\,[(nep)\exp(-(n-k)p(1-p)^k)]^k$.

If $k \leq \dfrac{\log np - \mu \log\log(nep)}{\log \frac{1}{1-p}}$ then $(1-p)^k \geq \exp(-\log(np) +$
$+ \mu \log\log(nep))$ and hence

$\exp(-(n-k)p(1-p)^k) \leq \exp(-(n-k)p(np)^{-1}(\log(nep))^{\mu}) \leq$

$\leq \exp(-(1-k/n)(1+\mu \log np))$.

If now

$$k \leq \frac{n(\mu - 1 - \varepsilon)\log np}{1 + \mu \log np}$$

then the RHS of (7) can be further bounded from above by

$n[np \exp(-(1+\varepsilon)\log np)]^k \leq n(np)^{-\varepsilon k}$.

If $k \geq \dfrac{(2+\delta)\log n}{\varepsilon \log np}$ then $P_k \leq n^{-(1+\delta)}$.

As $\sum P_k \leq n^{-\delta} = o(1)$ the theorem is proved. Q.E.D.

Now we give two consequences of Theorem 2.1.

<u>Theorem 2.2</u>. Let $f=f(x)$ be a function that tends to the infinity as $x \to \infty$. Given constants $\varepsilon, \delta > 0$, a natural number

n and a real number p such that $p \geq f(n)/n$, then the size k of any maximal induced subtree of $\mathcal{G}(n,p)$ satisfies almost surely one of these conditions:

either $k < \frac{(2+\delta)\log n}{\varepsilon \log np}$

or $\quad k > \frac{\log np-(1+\varepsilon)\log\log np}{\log \frac{1}{1-p}}$.

Proof: Since np tends to the infinity, $k_1/n \leq (\log np)/np \longrightarrow 0$, but k_2/n tends to $(\mu-1-\varepsilon)/\mu > 0$. It follows that $k_1 < k_2$ for large n.

Theorem 2.3. Given constants $\mu > 1$, $r > 0$, a natural number n and a real number $p \geq n^{r-1}$ then the size of any maximal induced subtree of $\mathcal{G}(n,p)$ is almost surely at least

$$\frac{\log np-\mu\log\log np}{\log \frac{1}{1-p}} .$$

Proof: k_0 is less than $(2+\delta)/\varepsilon r$ and since $(\log np)/\log\log(nep)$ tends to the infinity, we can choose ε so that $k_0 < 1$.

3. Large trees. Let k_0, k_1, k_2 be the numbers from Theorem 2.1. Results of the previous paragraph show that the existence of an induced subtree of the size k_0 implies almost surely the existence of an induced tree of the size $\mathrm{Min}(k_1,k_2)$, (Note that the interval between k_0 and $\mathrm{Min}(k_1,k_2)$ is non-empty for all choices of p.)

Theorem 2.3 shows that the proof of the existence of a tree of the size greater than k_0 is trivial, provided p is not too small. Now we are going to prove that if $pn > 1$ and, say, $p = O(n^{-1/2})$ then the random graph $\mathcal{G}(n,p)$ contains almost surely an induced subtree of the size $\delta p^{-1/4}$ for some $\delta > 0$.

Suppose that the vertices of $\mathcal{G}(n,p)$ are $v_1,\dots,v_n$. Define sets $U_1, V_1, \dots, U_k, V_k$ of vertices and sets $T_1,\dots,T_k$ of pairs of vertices as follows:

$V_1 = 0$,

$$U_i = \begin{cases} V_i & \text{if } V_i \neq 0 \\ \{v_m\} & \text{otherwise, where } m \text{ is the smallest integer such that} \\ & v_m \notin \bigcup_{j=1}^{i-1} U_j \end{cases}$$

V_{i+1} = the set of all vertices $v \notin \bigcup_{j=1}^{i-1} U_j$ adjacent to some $u \in U_i$,

T_i = the set of all pairs $\{v_s,u\}$, where $v_s \in U_i$, $u \notin \bigcup_{j=1}^{i} U_j$ and

u is adjacent to no $v_q \in U_i$, $q<s$.

k is defined by the condition

$$\sum_{j=1}^{k-1} |U_j| < p^{-1/4} \leq \sum_{j=1}^{k} |U_j| .$$

Note that each set U_i is contained in some component of the graph.

In order to show the existence of large trees it is sufficient to prove the next two lemmas:

Lemma 3.1. Let $c>1$ be a constant. If $np>c$ then the set U_k contains almost surely at least $\frac{c-1}{c+1} p^{-1/4}$ vertices.

Proof: To construct the sets U_i and V_i, it is sufficient to check which elements of T_i are edges of the graph $\mathcal{G}(n,p)$. Let $\varepsilon>0$ be a constant and suppose that the size of U_k is less than $\frac{c-1}{c+1} p^{-1/4}$. Denote $T=T_0 \cup \dots \cup T_{k-1}$, $t=|T|$ and $u=\sum_{j=1}^{k} |U_i|$. It is $u-|U_k| \geq p^{-1/4} - \frac{c-1}{c+1} p^{-1/4} = \frac{2}{c+1} p^{-1/4}$

$t \geq (u-|U_k|((n-u) \geq (u-|U_k|)(n-2p^{-1/4}) \geq (u-|U_k|)(1-\varepsilon)n$

for sufficiently large n.

The probability that less than $(1-\varepsilon)pt$ elements of T are edges is at most $\exp(-\varepsilon^3 pt)$.

The number of edges of T is equal to the size of the set $V_1 \cup \dots \cup V_k$ and therefore u is almost surely at least $(1-\varepsilon)pt$. If

$u \geq (1-\varepsilon)pt \geq (1-\varepsilon)p(u-|U_k|)(1-\varepsilon)n$

then

$$|U_k| \geq u\left(1-\frac{1}{(1-\varepsilon)^2 pn}\right) \geq p^{-1/4} \frac{c-1}{c+1}$$

if ε is chosen so that $(1-\varepsilon)^2 pn \geq (1+c)/2$.

Lemma 3.2. Let $c>1$ be a constant. If $c \leq p < n^{-1/2}$ then the random graph $\mathcal{G}(n,p)$ contains almost surely an induced tree of the size at least $\frac{c-1}{c+1}\, p^{-1/4}$

Proof. Let i be the largest number such that $V_i=0$. It follows from Lemma 3.1 that the set $U=U_i \cup \ldots \cup U_k$ together with the edges of $\mathcal{G}(n,p)$, which are elements of $T_i \cup \ldots \cup T_{k-1}$ is almost surely a tree of the size at least $\frac{c-1}{c+1}\, p^{-1/4}$. Without loss of generality, we can suppose that $|U| \leq p^{-1/4}$, otherwise we would use a sufficiently small set of the form $U_i \cup \ldots \cup U_{k-1} \cup \bar{U}$, where $\bar{U} \subset U_k$, instead of U.

If no pair $\{x,y\}$ such that $x,y \in U$, $\{x,y\} \notin T$ is an edge of $\mathcal{G}(n,p)$ then the set U is an induced subtree of $\mathcal{G}(n,p)$. The probability of this event is at least

$$1-(1-p)^{\binom{|U|}{2}} \gtrsim 1-\exp(-p\,\frac{|U|^2}{2}) \longrightarrow 1 \text{ because } p|U|^2 \leq p\ \ p^{-1/2} = p^{1/2} \longrightarrow 0.$$

The problem of existence of large trees in $\mathcal{G}(n,p)$ for p fixed was solved by P. Erdös and Z. Palka [3] who proved that almost every graph $G \in \mathcal{G}(n,p)$ contains a tree of size $(2-\varepsilon)\frac{\log n}{\log \frac{1}{1-p}}$, where $\varepsilon > 0$ is arbitrarily small positive real.

In [3] they also raised the question what is the largest tree in $\mathcal{G}(n,p)$ if $p \sim c/n$. The following theorem gives a linear (in n) lower bound.

Theorem 3.3. Let $c > 1$ and $\sigma > 0$ be fixed real numbers. If n is a natural number and μ, ε, p are real numbers such that $c/n \leq p$, $0 < \varepsilon < \mu-1$ and $\log np > \mu \log\log(nep)$ then the size k of the largest induced subtree satisfies almost surely the inequality $\mathrm{Min}(k_1,k_2) \leq k \leq \frac{2(\log(nep))}{\log \frac{1}{1-p}} + (1+\varepsilon)\frac{\log n}{\log(nep)} + 3$

where k_1 and k_2 are defined by (5) and (6) of Theorem 2.1.

Proof: The theorem follows immediately from Theorem 2.1 and Lemma 3.2, as one can easily verify that

$\frac{(2+\varepsilon)\log n}{\varepsilon \log np} < \frac{c-1}{c+1}\, p^{-1/4}$ for any n sufficiently large.

Corollary 3.5. Suppose that $p=o(n)$ and $pn \longrightarrow \infty$ as $n \longrightarrow \infty$ then the size of the largest induced subtree of $\mathcal{G}(n,p)$ satisfies almost surely the inequality

$$\frac{\log np}{p}(1+o(1))\leq k\leq 2\,\frac{\log np}{p}(1+o(1)).$$

References

[1] H. CHERNOFF: A measure of asymptotic efficiency for test of a hypothesis based on the sum of observations, Ann.Math.Stat. 23(1952), 493-509.

[2] P. ERDÖS, Z. PALKA: Trees in random graphs, Discrete Mathematics 46(1983), 145-150.

[3] P. ERDÖS, J. SPENCER: Probabilistic Methods in Combinatorics, Akademiai Kiadó, Budapest 1974.

KAM MFF UK, Charles University
Prague, Malostranské nám.25
11800 Praha 1
Czechoslovakia

FJFI ČVUT, Czech Technical
University, Husova 4
11600 Praha 1
Czechoslovakia

(Oblatum 20.5. 1986)