ZASTOSOWANIA MATEMATYKI
APPLICATIONES MATHEMATICAE
XVII, 2 (1982)

Z. PALKA (Poznań)

ISOLATED TREES IN A RANDOM GRAPH

1. Introduction. Let $G_{n,p}$ be a random graph with n labelled nodes, where each of the $\binom{n}{2}$ possible edges occurs with the same probability p $(0 < p < 1)$ independently of all other edges.

The main purpose of this note is to give the exact probability distribution of the number τ_k of isolated trees of order k in $G_{n,p}$, i.e. trees of order k which are isolated subgraphs of $G_{n,p}$. Some other properties of induced trees (not necessarily isolated) in a random graph have been considered in [7], where bounds on the size of the maximal induced tree in $G_{n,p}$, i.e. such a tree which is not properly contained in any other tree, are established. For a review of the results on random graphs see [6].

Erdös and Rényi [1] have considered the asymptotic properties of the random variable τ_k, but for a random graph of different kind, which is defined as follows.

Assume that n labelled nodes are given. Let us choose at random N edges among the $n_1 = \binom{n}{2}$ possible edges, so that each of the $\binom{(n_1)}{N}$ possible choices of these edges is equiprobable and let us denote such a random graph by $G_{n,N}$. It is evident that the number of edges in $G_{n,N}$ is known, whereas the number of edges in $G_{n,p}$ is a random variable with expectation $\binom{n}{2}p$.

It is also well known (see, e.g., [9]) that asymptotically, i.e. as $n \to \infty$, in most cases there is no essential difference between $G_{n,N}$ and $G_{n,p}$ if we put

$$p = N/\binom{n}{2}.$$

This follows from the fact that if n is large, then the number of edges in $G_{n,p}$ with such a p does not differ significantly from N. All theorems about the random variable τ_k in [1] are of asymptotic character. We shall show that some of these theorems follow immediately from our main result as well.

Let $\boldsymbol{P}_{n,p}(A)$ denote the probability that a random graph $G_{n,p}$ has some property A. As usual, for every x and every natural m we set

$$(x)_m = x(x-1)\dots(x-m+1), \qquad (x)_0 = 1,$$

and let $[x]$ denote the greatest integer not greater than x.

2. Exact results. The following theorem is valid:

THEOREM 1. *If $m = [n/k]$ and τ_k is the number of isolated trees of order k ($k \geqslant 1$) in $G_{n,p}$, then*

$$(1) \qquad \boldsymbol{P}_{n,p}(\tau_k = i) = \sum_{j=0}^{m-i} (-1)^j \binom{i+j}{i} S_{i+j} \qquad (i = 0, 1, \dots, m),$$

where $S_0 = 1$,

$$(2) \qquad S_r = \frac{(n)_{rk}}{r!}\left(\frac{t_k}{k!}\right)^r q^s, \quad s = rkn - \binom{r+1}{2}k^2 \quad (r = 1, 2, \dots, m),$$

and t_k is the probability that a set of k labelled nodes spans a tree, given by

$$(3) \qquad t_k = k^{k-2}p^{k-1}q^{(k-1)(k-2)/2}.$$

Proof. Let n and k be fixed. Denote by A_j ($1 \leqslant j \leqslant m$) the event that the j-th k-element subset of the n-element set of nodes is an isolated tree of $G_{n,p}$, and let B_i be the event that exactly i events occur among the events $A_1, A_2, \dots, A_m$. It is clear that

$$(4) \qquad \boldsymbol{P}_{n,p}(\tau_k = i) = \Pr(B_i).$$

Now, if $j_1 \neq j_2 \neq \dots \neq j_i$, then

$$\Pr(A_{j_1}A_{j_2}\dots A_{j_i}) = t_k^i q^{s'},$$

where

$$s' = ik(n-ik) + \binom{i}{2}k^2 = ikn - \binom{i+1}{2}k^2.$$

Thus it is easily checked that for $i = 1, 2, \dots, m$ we obtain

$$S_i = \sum \Pr(A_{j_1}A_{j_2}\dots A_{j_i}) = \frac{(n)_{ik}}{i!(k!)^i} t_k^i q^{s'},$$

where the summation is extended over all i-tuples of pairwise node-disjoint trees of order k which can be formed using n labelled nodes. Now, putting $S_0 = 1$ and taking into account equality (4), we obtain our thesis from the well-known Jordan's theorem (see, e.g., [4]).

From the result above we get

COROLLARY. *The first and the second moments of the random variable* τ_k *are given by*

$$\mathrm{E}(\tau_k) = \binom{n}{k} t_k q^{k(n-k)} \tag{5}$$

and

$$\mathrm{E}(\tau_k^2) = \mathrm{E}(\tau_k)\left[1+\binom{n-k}{k} t_k q^{k(n-2k)}\right], \tag{6}$$

respectively.

Proof. Formulas (5) and (6) follow from (2) and from the relation (see, e.g., [3]) $\mu_{(i)} = i!S_i$, where $\mu_{(i)}$ is the i-th factorial moment of the random variable τ_k.

Remark 1. For $k = 1$, i.e. for the number of isolated nodes, formulas (1), (5) and (6) have been obtained by Frank ([2], Chapter 7).

The numerical values of $\boldsymbol{P}_{25,p}(\tau_k = i)$ appear in Table 1, from which one can see how the edge probability p influences $\boldsymbol{P}_{n,p}(\tau_k = i)$. It should be noted that for $p < 0.005$ for $p > 0.1$ the values of $\boldsymbol{P}_{25,p}(\tau_k = 0)$ are near 1 for all $k \geqslant 2$.

TABLE 1

k	i	Edge probability p			
		0.005	0.02	0.05	0.1
2	0	0.27104	0.07231	0.25247	0.79596
	1	0.38981	0.20946	0.33636	0.17606
	2	0.23836	0.28355	0.23582	0.02473
	3	0.08126	0.23506	0.11485	0.00290
	4	0.01703	0.13132	0.04319	0.00031
	5	0.00228	0.05126	0.01317	0.00003
3	0	0.88193	0.46388	0.58050	0.94358
	1	0.11294	0.38453	0.30853	0.05369
	2	0.00503	0.12833	0.08977	0.00259
	3	0.00010	0.02138	0.01821	0.00013
4	0	0.98371	0.74541	0.74985	0.97940
	1	0.01622	0.23079	0.21242	0.02007
	2	0.00007	0.02298	0.03402	0.00051
5	0	0.99756	0.87991	0.83577	0.99080
	1	0.00244	0.11737	0.14857	0.00903

If n is large, the number of terms in (1) increases rapidly, so it may be convenient to use the following recurrence formula:

PROPERTY 1. *If* $m = [n/k]$, *then*

$$\boldsymbol{P}_{n,p}(\tau_k = 0) = 1 - \sum_{i=1}^{m} S_i \boldsymbol{P}_{n-ik,p}(\tau_k = 0),$$

where S_i *is given by* (2).

Proof. This follows from the fact that for $i = 0, 1, \ldots, m$ the formula

$$\boldsymbol{P}_{n,p}(\tau_k = i) = S_i \boldsymbol{P}_{n-ik,p}(\tau_k = 0)$$

holds, since the probability that in $G_{n,p}$ there exist exactly i isolated trees each of order k can be expressed by the probability that ik specified nodes span a forest and no isolated trees of order k exist among the remaining $n-ik$ nodes.

Now we derive the lower and upper bounds for $\boldsymbol{P}_{n,p}(\tau_k \geqslant 1)$. Using the inequalities (see, e.g., [4])

$$\sum_{j=0}^{2s+1} (-1)^j \binom{i+j}{i} S_{i+j} \leqslant \boldsymbol{P}_{n,p}(\tau_k = i) \leqslant \sum_{j=0}^{2s} (-1)^j \binom{i+j}{i} S_{i+j}, \tag{7}$$

valid for any integer $s \geqslant 0$, one can obtain

$$S_1 - S_2 \leqslant \boldsymbol{P}_{n,p}(\tau_k \geqslant 1) \leqslant S_1.$$

Thus

PROPERTY 2. *We have*

$$\boldsymbol{P}_{n,p}(\tau_k \geqslant 1) \geqslant \binom{n}{k} t_k q^{k(n-k)} \left[1 - \frac{1}{2}\binom{n-k}{k} t_k q^{k(n-2k)}\right] \tag{8}$$

and

$$\boldsymbol{P}_{n,p}(\tau_k \geqslant 1) \leqslant \binom{n}{k} t_k q^{k(n-k)}, \tag{9}$$

where t_k *is given by* (3).

Remark 2. If $\boldsymbol{P}_{n,p}(C)$ denotes the probability of connectedness of $G_{n,p}$, then the lower bound on $1-\boldsymbol{P}_{n,p}(C)$ is the probability that at least one of the nodes $1, 2, \ldots, n$ is connected to no other node, i.e. $\boldsymbol{P}_{n,p}(\tau_1 \geqslant 1) \leqslant 1-\boldsymbol{P}_{n,p}(C)$. So putting $k = 1$ into (8) we obtain the lower bound on $1-\boldsymbol{P}_{n,p}(C)$, which has been first derived by Gilbert [5].

Now we give another lower bound for the probability $\boldsymbol{P}_{n,p}(\tau_k \geqslant 1)$, which follows from the so-called strong second moment inequality (see Matula [8]):

If ξ is a non-negative integer-valued random variable, then

$$\Pr(\xi \geqslant 1) \geqslant \frac{\mathrm{E}^2(\xi)}{\mathrm{E}(\xi^2)}.$$

Thus, taking into account formulas (5) and (6), one can obtain

PROPERTY 3. *We have*

$$(10)\qquad \boldsymbol{P}_{n,p}(\tau_k \geqslant 1) \geqslant \frac{\binom{n}{k} t_k q^{k(n-k)}}{1+\binom{n-k}{k} t_k q^{k(n-2k)}},$$

where t_k is given by (3).

It is easily checked that (10) is better than (8) only if

$$(11)\qquad \binom{n-k}{k} t_k q^{k(n-2k)} > 1$$

holds.

Let us show also that Properties 2 and 3 can be used for the investigation of the size γ of the largest isolated tree in $G_{n,p}$. We have evidently

$$\boldsymbol{P}_{n,p}(\gamma \geqslant z) = \boldsymbol{P}_{n,p}\Big(\bigcup_{k\geqslant z} (\tau_k \geqslant 1)\Big) \leqslant \sum_{k\geqslant z} \boldsymbol{P}_{n,p}(\tau_k \geqslant 1).$$

Thus from (9) we get

$$\boldsymbol{P}_{n,p}(\gamma \geqslant z) \leqslant \sum_{k\geqslant z} \binom{n}{k} t_k q^{k(n-k)}.$$

In order to obtain the lower bound it is sufficient only to notice that

$$\boldsymbol{P}_{n,p}(\gamma \geqslant z) \geqslant \boldsymbol{P}_{n,p}(\tau_z \geqslant 1),$$

and according to condition (11) one can estimate $\boldsymbol{P}_{n,p}(\tau_z \geqslant 1)$ using either (8) or (10).

To illustrate the considerations above we have computed the following values of the lower and upper bounds on $\boldsymbol{P}_{n,p}(\gamma \geqslant z)$ for $n = 15$ and $p = 0.05$:

$$0.8935 \leqslant \boldsymbol{P}_{15,0.05}(\gamma \geqslant 1) \leqslant 1,$$
$$0.6117 \leqslant \boldsymbol{P}_{15,0.05}(\gamma \geqslant 2) \leqslant 1,$$
$$0.4111 \leqslant \boldsymbol{P}_{15,0.05}(\gamma \geqslant 3) \leqslant 1,$$
$$0.2285 \leqslant \boldsymbol{P}_{15,0.05}(\gamma \geqslant 4) \leqslant 0.5496,$$
$$0.1300 \leqslant \boldsymbol{P}_{15,0.05}(\gamma \geqslant 5) \leqslant 0.3046,$$
$$0.0758 \leqslant \boldsymbol{P}_{15,0.05}(\gamma \geqslant 6) \leqslant 0.1719,$$
$$0.0443 \leqslant \boldsymbol{P}_{15,0.05}(\gamma \geqslant 7) \leqslant 0.0958.$$

The lower bounds on $\boldsymbol{P}_{15,0.05}(\gamma \geqslant 1)$ and $\boldsymbol{P}_{15,0.05}(\gamma \geqslant 2)$ have been computed using formula (10). In all other cases we have used Property 2.

3. Asymptotic properties. Now let us turn to the case where the number of nodes of $G_{n,p}$ tends to infinity. If the edge probability p is a constant, then from Property 2 for any $k \geqslant 1$ it follows that $\boldsymbol{P}_{n,p}(\tau_k = 0) \to 1$ as $n \to \infty$. Thus a random graph $G_{n,p}$ has, with probability 1, no isolated trees. The situation changes if the edge probability p depends on n, i.e. $p = p(n)$, and tends to zero as $n \to \infty$. As a matter of fact, after Erdös and Rényi [1] we can formulate theorems describing properties of τ_k, but with respect to a random graph $G_{n,p}$.

THEOREM 2 (Erdös and Rényi). *If*

$$\lim_{n\to\infty} p n^{k/(k-1)} = \varrho \quad (0 < \varrho < \infty), \tag{12}$$

then

$$\lim_{n\to\infty} \boldsymbol{P}_{n,p}(\tau_k = i) = \frac{\lambda^i e^{-\lambda}}{i!} \quad (i = 0, 1, \ldots),$$

where

$$\lambda = \frac{\varrho^{k-1} k^{k-2}}{k!}. \tag{13}$$

THEOREM 3 (Erdös and Rényi). *If*

$$np = \frac{1}{k}\log n + \frac{k-1}{k}\log\log n + y + o(1), \tag{14}$$

where $-\infty < y < +\infty$, *then*

$$\lim_{n\to\infty} \boldsymbol{P}_{n,p}(\tau_k = i) = \frac{\mu^i e^{-\mu}}{i!} \quad (i = 0, 1, \ldots),$$

where

$$\mu = \frac{e^{-ky}}{kk!}. \tag{15}$$

In other words, in both cases the number of isolated trees of order k contained in $G_{n,p}$ has in the limit for $n \to \infty$ the Poisson distribution with expectations λ and μ, respectively. Now we give a little different proofs of Theorems 2 and 3 which follow immediately from Theorem 1.

Proof of Theorem 2. Taking into account

$$(n)_{ik} = \big(1+o(1)\big)n^{ik} \quad \text{and} \quad 1-p = \exp\big(-p+O(p^2)\big),$$

we infer from (2) and (3) that

$$S_i = \frac{1}{i!}\left(\frac{k^{k-2}}{k!} n^k p^{k-1} e^{-knp}\right)^i \big(1+O(iknp^2)\big). \tag{16}$$

Further, by assumption (12), we have $pn \to 0$ as $n \to \infty$, so finally

$$\lim_{i\to\infty} S_i = \frac{\lambda^i}{i!},$$

where λ is given by (13). Now, according to inequalities (7), for any fixed value $s \geqslant 0$ we have

$$\frac{\lambda^i}{i!}\sum_{j=0}^{2s+1}(-1)^j\frac{\lambda^j}{j!} \leqslant \liminf_{n\to\infty} \boldsymbol{P}_{n,p}(\tau_k = i)$$
$$\leqslant \limsup_{n\to\infty} \boldsymbol{P}_{n,p}(\tau_k = i) \leqslant \frac{\lambda^i}{i!}\sum_{j=0}^{2s}(-1)^j\frac{\lambda^j}{j!}.$$

Since s can be chosen arbitrarily large, we infer that the limit

$$\lim_{n\to\infty} \boldsymbol{P}_{n,p}(\tau_k = i)$$

exists and equals

$$\lim_{n\to\infty} \boldsymbol{P}_{n,p}(\tau_k = i) = \frac{\lambda^i}{i!}\sum_{j=0}^{\infty}(-1)^j\frac{\lambda^j}{j!} = \frac{\lambda^i e^{-\lambda}}{i!}.$$

Thus Theorem 2 is proved.

Proof of Theorem 3. From (14) we get

$$knp = \big(1+o(1)\big)\log n,$$

and, consequently,

$$\frac{k^{k-2}}{k!} n^k p^{k-1}\exp(-knp) \sim \frac{1}{kk!}\exp\big(\log n + (k-1)\log\log n - knp\big)$$
$$\sim \frac{1}{kk!}\exp(-ky).$$

Setting this into (16) we have

$$\lim_{n\to\infty} S_i = \frac{\mu^i}{i!},$$

where μ is given by (15), and the proof is completed by the use of inequalities (7) exactly as in the proof of Theorem 2.

References

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

[2] O. Frank, *Statistical inference in graphs*, Swedish Research Institute of National Defence, Stockholm 1971.

[3] M. Fréchet, *Les probabilités associées à un système d'événements compatibles et dépendant*, Actualités Sci. Indust., Paris 1940, 1943.
[4] I. Galambos, *On the sieve methods in probability theory I*, Studia Sci. Math. Hungar. 1 (1960), p. 39-50.
[5] E. N. Gilbert, *Random graphs*, Ann. Math. Statist. 30 (1959), p. 1141-1144.
[6] M. Karoński, *A review of random graphs*, Technical Report, Depart. of Math., A. Mickiewicz Univ., Poznań 1978.
[7] — and Z. Palka, *On the size of a maximal induced tree in a random graph*, Math. Slovaca 30 (1980), p. 151-155.
[8] D. W. Matula, *The largest clique size in a random graph*, Technical Report, Depart. of Computer Science, Southern Methodist Univ., Dallas 1976.
[9] V. E. Stepanov (В. Е. Степанов), *Случайные графы*, Вопросы кибернет. (1973), p. 164-185.

INSTITUTE OF MATHEMATICS
ADAM MICKIEWICZ UNIVERSITY
60-769 POZNAŃ

Received on 21. 7. 1978

Z. PALKA (Poznań)

IZOLOWANE DRZEWA W GRAFIE LOSOWYM

STRESZCZENIE

Oznaczmy przez $G_{n,p}$, $0 < p < 1$, graf losowy określony na zbiorze wierzchołków $\{1, 2, \ldots, n\}$, w którym każda z $\binom{n}{2}$ możliwych krawędzi występuje z prawdopodobieństwem p niezależnie jedna od drugiej. W pracy podano wyniki dotyczące rozkładu prawdopodobieństwa liczby izolowanych drzew o tym samym wymiarze w $G_{n,p}$. Wyznaczono również oszacowanie dolne i górne dla prawdopodobieństwa, iż wymiar największego izolowanego drzewa w grafie losowym jest co najmniej równy danej wielkości. Podano przykłady ilustrujące omawiane zagadnienia.