COLLOQUIUM MATHEMATICUM

VOL. XLVII 1982 FASC. 1

ISOLATED TREES IN A BICHROMATIC RANDOM GRAPH

BY

ZBIGNIEW PALKA (POZNAŃ)

1. Introduction. Assume that the bichromatic random graph $G_{m,n,p}$ has m given labelled points $P_1, P_2, \ldots, P_m$ of one colour (say red), n given labelled points $Q_1, Q_2, \ldots, Q_n$ of another colour (say blue) and let each of mn possible edges connecting only a red point with a blue point occur with a prescribed probability $p = 1-q$ $(0 \leqslant p \leqslant 1)$ independently of all other edges. A connected bichromatic graph which has k red points, l blue points, and $k+l-1$ edges is called a (k, l)-*tree*. Let the random variable $\tau_{k,l}$ $(k \geqslant 1,\ l \geqslant 1)$ denote the number of isolated (k, l)-trees in the bichromatic random graph $G_{m,n,p}$, i.e., (k, l)-trees which are isolated subgraphs of $G_{m,n,p}$. In this paper, we give first the exact probability distribution of the random variable $\tau_{k,l}$. A similar result concerning the distribution of the number of isolated trees of order k in a usual random graph $G_{n,p}$ is presented in [11]. Next, we formulate theorems describing the asymptotic character of behaviour of $\tau_{k,l}$ as well as the size of the greatest isolated (k, k)-tree in $G_{m,n,p}$. These theorems are analogous to the ones of Erdös and Rényi [3], concerning the asymptotic probability distributions of the number of isolated trees of order k and the size of the greatest isolated tree in the random graph $G_{n,N}$ which has n labelled points and N edges randomly chosen from among the $\binom{n}{2}$ possible edges. We would like to remark that the methods used for the proofs of Theorems 3, 5, and 7 are mainly those of Erdös and Rényi [3]. For a review of the results on random graphs, we refer the reader to [6].

We denote by $\mathbf{P}_{m,n,p}(A)$ the probability that the graph $G_{m,n,p}$ has the property A. Next, $\mathrm{E}(\xi) = \mathrm{E}\xi$ and $\mathrm{D}^2(\xi)$ mean the expectation and the variance of the random variable ξ, respectively. As usual, for every x and every natural number m we set

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

and let $[x]$ denote the integer part of x.

2. The exact result. In this section we give the exact probability distribution of the number $\tau_{k,l}$ of isolated (k, l)-trees in the bichromatic random graph $G_{m,n,p}$.

THEOREM 1. *If* $h = \min([m/k], [n/l])$, *then*

$$P_{m,n,p}(\tau_{k,l} = i) = \sum_{j=0}^{h-i} (-1)^j \binom{i+j}{i} S_{i+j} \qquad (i = 0, 1, \ldots, h),$$

where

$$S_i = \frac{(m)_{ik}(n)_{il}}{i!} \left(\frac{t_{k,l}}{k!\,l!} q^{kn+lm-(i+1)kl} \right)^i \tag{1}$$

and $t_{k,l}$ *is the probability of occurrence of a* (k, l)*-tree, i.e.,*

$$t_{k,l} = k^{l-1} l^{k-1} p^{k+l-1} q^{kl-(k+l-1)}. \tag{2}$$

Proof. Let A_j $(1 \leqslant j \leqslant h)$ denote the event that the j-th $(k+l)$-element subset of the $(m+n)$-element set of points is an isolated (k, l)-tree of $G_{m,n,p}$ and let B_i be the event that exactly i among the events $A_1, A_2, \ldots, A_h$ occur. It is clear that

$$P_{m,n,p}(\tau_{k,l} = i) = \Pr(B_i). \tag{3}$$

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

$$\Pr(A_{j_1} A_{j_2} \ldots A_{j_i}) = t_{k,l}^i q^{ik(n-il)+il(m-ik)+i(i-1)kl} = t_{k,l}^i q^{i(kn+lm-(i+1)kl)},$$

where $t_{k,l}$ is given by (2). Formula (2) follows from the fact that the number of bichromatic (k, l)-trees is equal to $k^{l-1} l^{k-1}$ (see [1]). Thus it is easily checked that for $i = 1, 2, \ldots, h$

$$S_i = \sum \Pr(A_{j_1} A_{j_2} \ldots A_{j_i}) = \frac{(m)_{ik}(n)_{il}}{i!(k!)^i(l!)^i} (t_{k,l} q^{kn+lm-(i+1)kl})^i, \tag{4}$$

where the summation is extended over all i-tuples of pairwise point-disjoint (k, l)-trees which can be formed by using m labelled red points and n labelled blue points. Let $S_0 = 1$. Then by (3) we obtain our thesis from the well-known Jordan's theorem (see, e.g., [4]).

3. Asymptotic distributions. We deal now with the case of $m \sim cn$, where $c > 0$ does not depend on m and n. Using the Bonferroni inequality

$$1 - S_1 \leqslant P_{m,n,p}(\tau_{k,l} = 0) \leqslant 1$$

and applying (1) we have

COROLLARY 1. *If* $m \sim cn$, *then for each fixed* $k, l \geqslant 1$ *and* $p > 0$

$$\lim_{n\to\infty} \boldsymbol{P}_{m,n,p}(\tau_{k,l} = 0) = 1.$$

This property follows also from another known result of random graph theory. Indeed, if p is constant, then the mean number of edges in $G_{m,n,p}$ ($m \sim cn$) has the order of magnitude n^2, and the bichromatic random graph $G_{m,n,p}$ is connected with probability tending to 1 as $n \to \infty$ (see [10]). Thus, $G_{m,n,p}$ does not contain any isolated tree. This situation changes if the edge probability p depends on m and n, i.e., $p = p(m, n)$, and $p \to 0$ as $m, n \to \infty$. As a matter of fact, the following theorem is valid:

THEOREM 2 (Palásti [8]). *If* $m \sim n$ *and* $p = p(m, n) \to 0$ *as* $m, n \to \infty$ *in such a way that*

$$\lim_{n\to\infty} pn^{(k+l)/(k+l-1)} = \varrho \quad (0 < \varrho < \infty), \tag{5}$$

then

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

where

$$\lambda = \frac{\varrho^{k+l-1} k^{l-1} l^{k-1}}{k!\,l!}. \tag{6}$$

In other words, the number of isolated (k, l)-trees contained in $G_{m,n,p}$ ($m \sim n$) has in the limit for $n \to \infty$ the Poisson distribution with expectation λ. Now we give a little different proof of Theorem 2, which follows from Theorem 1.

Proof. Since

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

from (1) and (2) we obtain

$$S_i = \frac{1}{i!}\left(\frac{k^{l-1}l^{k-1}}{k!\,l!} m^k n^l p^{k+l-1} e^{-(kn+lm)p}\right)^i \{1+O(i(kn+lm)p^2)\}. \tag{7}$$

Further, it is easily seen that, by (5), $pn \to 0$ as $n \to \infty$, and because $m \sim n$, we have

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

where λ is given by (6). Now, if $\mu_{[i]}$ denotes the i-th factorial moment of the random variable $\tau_{k,l}$, then (see, e.g., [2], p. 71) $\mu_{[i]} = i!\,S_i$. Thus

$$\lim_{n\to\infty} \mu_{[i]} = \lambda^i.$$

On the other hand, the i-th factorial moment of the Poisson distribution with expectation λ is equal to λ^i (see, e.g., [2], p. 77). Since the Poisson distribution is uniquely determined by its moments, by the second limit theorem (see, e.g., [7]), we obtain our thesis.

Now let us pay our attention to the following substantial fact. If (5) holds, then a random graph $G_{m,n,p}$ ($m \sim n$) contains only such (k,l)-trees which are isolated. Indeed, if a (k,l)-tree were a subgraph but not an isolated subgraph of $G_{m,n,p}$, then $G_{m,n,p}$ would have a connected subgraph consisting of $k+l+1$ points and $k+l$ edges. We will show that it is impossible. Let $\mathscr{B}_{k,l,v}$ denote any non-empty class of connected balanced bichromatic random graphs which contain k red points, l blue points, and v edges (connecting points of different colours), where $k, l \geqslant 1$ and $k+l-1 \leqslant v \leqslant kl$. Then the threshold function concerning the property B that $G_{m,n,p}$ ($m \sim n$) contains a subgraph isomorphic to some element of $\mathscr{B}_{k,l,v}$ is equal to $n^{2-(k+l)/v}$ (see [9]), i.e.,

$$\lim_{n\to\infty} \boldsymbol{P}_{m,n,p}(B) = \begin{cases} 0 & \text{if } \lim\limits_{n\to\infty} pn^{2-(k+l)/v} = 0, \\ 1 & \text{if } \lim\limits_{n\to\infty} pn^{2-(k+l)/v} = \infty. \end{cases} \tag{8}$$

Since, in our case, p satisfies (5), we have

$$\lim_{n\to\infty} pn^{2-(k+l+1)/(k+l)} = 0$$

and, by (8), the probability that $G_{m,n,p}$ contains a connected subgraph consisting of $k+l+1$ points and $k+l$ edges tends to zero as $n\to\infty$.

Now, if A denotes the property that the bichromatic random graph $G_{m,n,p}$ ($m \sim n$) contains some (k,l)-tree (not necessarily isolated), then by (8) we have

$$\lim_{n\to\infty} \boldsymbol{P}_{m,n,p}(A) = \begin{cases} 0 & \text{if } \lim\limits_{n\to\infty} pn^{(k+l)/(k+l-1)} = 0, \\ 1 & \text{if } \lim\limits_{n\to\infty} pn^{(k+l)/(k+l-1)} = \infty. \end{cases} \tag{9}$$

Now, let us consider the asymptotic distribution of the random variable $\tau_{k,l}$ when

$$\lim_{n\to\infty} pn^{(k+l)/(k+l-1)} = \infty. \tag{10}$$

THEOREM 3. *If $m \sim n$ and* (10) *holds, but*

$$\lim_{n\to\infty} \left(np - \frac{1}{k+l}\log n - \frac{k+l-1}{k+l}\log\log n \right) = -\infty, \tag{11}$$

then for $-\infty < x < +\infty$

$$\lim_{n\to\infty} P_{m,n,p}\left(\frac{\tau_{k,l}-M_{n,p}}{\sqrt{M_{n,p}}} < x\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp(-u^2/2)\,du,$$

where

$$M_{n,p} = n\,\frac{k^{l-1}l^{k-1}}{k!\,l!}\,(np)^{k+l-1}e^{-(k+l)np}. \tag{12}$$

Proof. Note first that the two conditions (10) and (11) are equivalent to the condition

$$\lim_{n\to\infty} M_{n,p} = \infty. \tag{13}$$

Since the normal distribution $N(0,1)$ is uniquely determined by its moments, it suffices to show that

$$\lim_{n\to\infty} \mathrm{E}\left(\frac{\tau_{k,l}-M_{n,p}}{\sqrt{M_{n,p}}}\right)^r = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} u^r \exp(-u^2/2)\,du. \tag{14}$$

Taking into account the relation between moments and factorial moments (see [5], p. 77, and [2], p. 71), we have

$$\mathrm{E}(\tau_{k,l}^r) = \sum_{i=1}^{r} \sigma_r^{(i)}\, i!\, S_i,$$

where $\sigma_r^{(i)}$ are the Stirling numbers of the second kind. Thus from (7) we get

$$\mathrm{E}(\tau_{k,l}^r) = \{1+O(r(k+l)np^2)\} \sum_{i=1}^{r} \sigma_r^{(i)} M_{n,p}^i, \tag{15}$$

where $M_{n,p}$ is given by (12). Now using the identity (see [3], p. 33)

$$\sum_{i=1}^{r} \sigma_r^{(i)} \lambda^i = \sum_{i=0}^{\infty} \frac{\lambda^i}{i!}\, e^{-\lambda} i^r \qquad (r = 1, 2, \ldots),$$

we obtain

$$\mathrm{E}\left(\frac{\tau_{k,l}-M_{n,p}}{\sqrt{M_{n,p}}}\right)^r \tag{16}$$

$$= \frac{1}{M_{n,p}^{r/2}} \sum_{i=0}^{\infty} \frac{M_{n,p}^i}{i!}\,[\exp(-M_{n,p})](i-M_{n,p})^r\{1+O(r(k+l)np^2)\}.$$

Since in our case $M_{n,p} \to \infty$ by (13), the right-hand side in (16) converges to the r-th moment of $N(0,1)$ (see [3], p. 33), and (14) is proved.

Let us notice now that from (15) we have

$$\mathrm{E}(\tau_{k,l}) \sim M_{n,p},$$

and as $\lim_{n\to\infty} M_{n,p} = \infty$, this means that the assertion of Theorem 3 can be expressed as follows: the number of isolated (k, l)-trees in $G_{m,n,p}$, where $m \sim n$, is always asymptotically normally distributed if $n \to \infty$ and $p(m,n) \to 0$, so that the mean number of such trees tends to infinity.

From Theorems 2, 3 and from (9) we get the following asymptotic property of $\tau_{k,l}$:

COROLLARY 2. *If $m \sim n$ and $p(m,n) \to 0$ in such a way that* (5) *holds, then the number of isolated (k, l)-trees in $G_{m,n,p}$ is finite and has the Poisson distribution, whereas the number of isolated trees of order less than $k+l$ has a normal distribution and there are no isolated trees of order greater than $k+l$.*

Finally, let us turn to the case where

$$(17)\qquad (k+l)np = \log n + (k+l-1)\log\log n + y + o(1),\qquad -\infty < y < +\infty.$$

Then the mean number of isolated (k, l)-trees is again finite and the following theorem is valid:

THEOREM 4. *If* (17) *holds and $m \sim n$, then*

$$\lim_{n\to\infty} \boldsymbol{P}_{m,n,p}(\tau_{k,l} = i) = \frac{\lambda^i e^{-\lambda}}{i!} \qquad (i = 0, 1, \ldots),$$

where

$$(18)\qquad \lambda = \frac{k^{l-1} l^{k-1}}{k!\,l!} (k+l)^{-(k+l-1)} e^{-y}.$$

Proof. From (17) we obtain

$$(k+l)np = \big(1+o(1)\big)\log n,$$

so

$$\begin{aligned} n(np)^{k+l-1} & \exp\big(-(k+l)np\big) \\ & \sim \exp\Big(\log n + (k+l-1)\big(\log\log n - \log(k+l)\big) - (k+l)np\Big) \\ & \sim (k+l)^{-(k+l-1)} e^{-y}, \end{aligned}$$

and by (7) we get

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

where λ is given by (18). Thus the proof is completed by the use of factorial moments exactly as in the proof of Theorem 2.

4. The size of the greatest isolated tree. We investigate here the size of the greatest bichromatic isolated (k, l)-tree of $G_{m,n,p}$ in a particular case where $k = l$ and $m \sim n$. Let γ denote the number of points of the greatest (k, k)-tree which is a component of $G_{m,n,p}$. Let $p \sim c/n$, where $c > 0$ does not depend on n. We have

THEOREM 5. *Suppose $p \sim c/n$ with $c \neq 1$. Let ω_n be a sequence tending to $+\infty$ arbitrarily slow. Then*

$$\lim_{n\to\infty} P_{m,n,p}\big(\gamma \geqslant a^{-1}(\log n - 3\log\log n) + \omega_n\big) = 0 \tag{19}$$

and

$$\lim_{n\to\infty} P_{m,n,p}\big(\gamma \geqslant a^{-1}(\log n - 3\log\log n) - \omega_n\big) = 1, \tag{20}$$

where

$$e^{-a} = ce^{1-c} \quad (i.e.,\ a = c-1-\log c > 0). \tag{21}$$

Proof. Let us notice that

$$P_{m,n,p}(\gamma \geqslant z) = P_{m,n,p}\Big(\bigcup_{k\geqslant z/2} (\tau_{k,k} \geqslant 1)\Big) \leqslant \sum_{k\geqslant z/2} \mathrm{E}(\tau_{k,k}).$$

Since $m \sim n$, by (15) and (12) we have

$$\mathrm{E}(\tau_{k,k}) = n\left(\frac{k^{k-1}}{k!}\right)^2 c^{2k-1} e^{-2kc}\big(1 + O(kcp)\big).$$

Using the Stirling formula and taking into account (21) we have further

$$P_{m,n,p}(\gamma \geqslant z) = O\left(\frac{ne^{-az}}{z^3}\right).$$

Now, if $z_1 = a^{-1}(\log n - 3\log\log n) + \omega_n$, then

$$P_{m,n,p}(\gamma \geqslant z_1) = O\big(\exp(-a\omega_n)\big),$$

which proves (19). To prove (20) we have to estimate the expectation and the variance of $\tau_{t,t}$, where $2t = z_2 = a^{-1}(\log n - 3\log\log n) - \omega_n$. By (15) we have

$$\mathrm{E}(\tau_{t,t}) \sim \frac{1}{2c\pi}\,\frac{ne^{-2at}}{t^3} \sim \frac{4a^3}{c\pi}\exp(a\omega_n). \tag{22}$$

and

$$\mathrm{D}^2(\tau_{t,t}) = O\big(\mathrm{E}(\tau_{t,t})\big). \tag{23}$$

Now we have clearly

$$P_{m,n,p}(\gamma \geqslant z_2) \geqslant P_{m,n,p}(\tau_{t,t} \geqslant 1) = 1 - P_{m,n,p}(\tau_{t,t} = 0).$$

Using the Chebyshev inequality and taking into account the fact that the random variable $\tau_{t,t}$ is nonnegative, we infer from (22) and (23) that

$$P_{m,n,p}(\tau_{t,t} = 0) = O\big(\exp(-a\omega_n)\big).$$

Thus

$$P_{m,n,p}(\gamma \geqslant z_2) \geqslant 1 - O\big(\exp(-a\omega_n)\big)$$

and (20) is proved.

In other words, Theorem 5 states that if $p \sim c/n$ ($c \neq 1$), then the size γ of the greatest isolated (k, k)-tree satisfies

$$|\gamma - a^{-1}(\log n - 3\log\log n)| \leqslant \omega_n$$

almost surely. Now, let us investigate the number of isolated (k, k)-trees when

$$2k = h = a^{-1}(\log n - 3\log\log n) + l, \tag{24}$$

where l is an arbitrary real number such that h is a positive integer. We have

THEOREM 6. *If $m \sim n$, $p \sim c/n$ with $c \neq 1$, and $a = c - 1 - \log c$, then the number of isolated trees of size h, where h is given by* (24), *contained in $G_{m,n,p}$ has for $n \to \infty$ the Poisson distribution with expectation $\lambda = 4a^3 e^{-la}/(c\pi)$.*

Proof. Since $k = l$, $m \sim n$, and $np \sim c$, we have from (7)

$$S_i \sim \frac{1}{i!}\left\{\frac{n}{c}\left(\frac{k^{k-1}}{k!}\right)^2 (ce^{-c})^{2k}\right\}^i.$$

Using the Stirling formula and taking into account (21) and (24) we have further

$$S_i \sim \frac{1}{i!}\left(\frac{4}{c\pi}\,\frac{n}{(2k)^3}\,e^{-2ka}\right)^i \sim \frac{1}{i!}\left(\frac{4}{c\pi}\,a^3 e^{-la}\right)^i,$$

and the proof is completed by the use of factorial moments exactly as in the proof of Theorem 2.

Now, let $c = 1$. In this case, the size of the greatest isolated (k, k)-tree contained in $G_{m,n,p}$ ($m \sim n$) satisfies

$$\omega_n^{-1} \leqslant \gamma n^{-1/3} \leqslant \omega_n$$

almost surely. As a matter of fact, the following result is valid:

THEOREM 7. *If* $p \sim 1/n$ *and* ω_n *is a sequence tending to* $+\infty$ *arbitrarily slow, then*

$$\lim_{n\to\infty} \boldsymbol{P}_{m,n,p}(\gamma \geqslant n^{1/3}\omega_n) = 0 \tag{25}$$

and

$$\lim_{n\to\infty} \boldsymbol{P}_{m,n,p}(\gamma \geqslant n^{1/3}/\omega_n) = 1. \tag{26}$$

Proof. Proceeding analogously as in the proof of Theorem 5 we have

$$\boldsymbol{P}_{m,n,p}(\gamma \geqslant z) = O(n/z^3).$$

Thus, if $z_1 = n^{1/3}\omega_n$, then

$$\boldsymbol{P}_{m,n,p}(\gamma \geqslant z_1) = O(\omega_n^{-3}),$$

and so (25) holds. Now, if $2t = z_2 = n^{1/3}/\omega_n$, then

$$\mathrm{E}(\tau_{t,t}) \sim \frac{4}{\pi}\omega_n^3 \quad \text{and} \quad \mathrm{D}^2(\tau_{t,t}) = O(\omega_n^3),$$

so (26) follows by using again the Chebyshev inequality.

5. Remark (added on February 14, 1980). It should be mentioned that K. Schürger in his unpublished Ph. D. thesis [12] has considered, among other problems, the probability distributions of the number of isolated k^*-trees (where $k^* = (k_1, k_2, \ldots, k_r)$) in the r-partite random graph $G_{n^*,N}$, $n^* = (n_1, n_2, \ldots, n_r)$, which has N edges randomly chosen from $\sum_{1\leqslant i<j\leqslant r} n_i n_j$ possible edges. It appears that Theorems 2-4 of our paper are special cases of more general theorems proved by Schürger. We would like to remark also that the method used for the proofs of Theorems 2 and 4 differs from that applied by Schürger.

REFERENCES

[1] T. L. Austin, *The enumeration of point labelled chromatic graphs and trees*, Canadian Journal of Mathematics 12 (1960), p. 535-545.

[2] F. N. David and D. E. Barton, *Combinatorial chance*, London 1962.

[3] P. Erdös and A. Rényi, *On the evolution of random graphs*, A Magyar Tudományos Akadémia Matematikai Kutató Intézetének Közleményei 5 (1960), p. 17-61.

[4] I. Galambos, *On the sieve methods in probability theory. I*, Studia Scientiarum Mathematicarum Hungarica 1 (1966), p. 39-50.

[5] K. Jordan, *Chapters on the classical calculus of probability*, Budapest 1972.

[6] M. Karoński, *A review of random graphs*, Technical Report, Department of Mathematics, A. Mickiewicz University, Poznań 1978.
[7] M. G. Kendall and A. Stuart, *The advanced theory of statistics*, Vol. 1, *Distribution theory*, 2nd edition, London 1962.
[8] I. Palásti, *On the distribution of the number of trees which are isolated subgraphs of a chromatic random graph*, A Magyar Tudományos Akadémia Matematikai Kutató Intézetének Közleményei 6 (1961), p. 405-409.
[9] — *Threshold function for subgraphs of given type of the bichromatic random graph*, ibidem 7 (1962), p. 215-221.
[10] — *On the connectedness of bichromatic random graphs*, ibidem 8 (1963), p. 431-440.
[11] Z. Palka, *Isolated trees in a random graph*, Zastosowania Matematyki 17 (1982), p. 309-316.
[12] K. Schürger, *Über die Entwicklung zufälliger p-partiter Graphen*, Universität Heidelberg, 1974 (unpublished).

INSTITUTE OF MATHEMATICS
A. MICKIEWICZ UNIVERSITY
POZNAŃ

Reçu par la Rédaction le 9. 6. 1979;
en version modifiée le 15. 11. 1980