COLLOQUIUM MATHEMATICUM

VOL. XLV 1981 FASC. 1

ON PENDANT VERTICES IN RANDOM GRAPHS

BY

ZBIGNIEW PALKA (POZNAŃ)

1. Introduction. Consider a random undirected graph $G_{n,p}$ on n labeled vertices with no loops and multiple edges, in which each of $\binom{n}{2}$ edges occurs with a prescribed probability $p = 1-q$ $(0 < p < 1)$ independently of all other edges. The aim of this paper is to give some results about the distribution of the number of pendant vertices in $G_{n,p}$. A vertex x is called *pendant* if it is adjacent to exactly one other vertex, i.e. the degree of x is equal to one.

Some authors have considered similar problems for random graphs of other kind (see Meir and Moon [6], Na and Rapoport [7], Rényi [8]). The formula for the probability distribution of the number of isolated vertices in $G_{n,p}$, i.e. vertices of degree equal to zero, is given by Frank [3].

2. Pendant vertices of connected graphs. Let us denote by P_n the probability that a random graph $G_{n,p}$ is connected, i.e. for every pair x, y of distinct vertices of $G_{n,p}$ there exists an (x, y)-path. It is known that P_n may be computed according to the recurrence relation

$$(1)\qquad P_n = 1 - \sum_{s=1}^{n-1} \binom{n-1}{s-1} P_s q^{s(n-s)}, \quad n \geqslant 2 \text{ and } P_1 = 1,$$

obtained by Gilbert [5]. This formula follows from the fact that, for a random graph $G_{n,p}$, one and only one of the n following events for $s = 1, 2, \ldots, n$ is true: The vertex 1 is connected to $s-1$ other vertices and no one of these s connected vertices has any edges to other $n-s$ vertices.

Now denote by $\Pr\{V(n, k)\}$ the probability that $G_{n,p}$ is a connected graph with exactly k pendant vertices, $k = 0, 1, \ldots, n$. We prove the following result:

THEOREM 1. *Let $n \geqslant 3$. Then for $k = 0, 1, \ldots, n-1$*

$$\Pr\{V(n, k)\} = \sum_{m=k}^{n-1} (-1)^{m+k} \binom{m}{k} S_{n,m},$$

where

$$S_{n,m} = \binom{n}{m} \{(n-m) p q^{(2n-m-3)/2}\}^m P_{n-m}, \tag{2}$$

and P_{n-m} is given by (1).

Proof. Let $R_{n,m}$ be the probability that $G_{n,p}$ is a connected graph in which m $(1 \leqslant m \leqslant n-1)$ fixed vertices are pendant, i.e. each of them is an endpoint of exactly one edge. From the condition that the graph $G_{n,p}$ should be connected it follows that the other endpoints of these m edges are chosen from $n-m$ remaining vertices. The number of all such connections is equal to $(n-m)^m$. Since each edge occurs with the same probability $p = 1-q$, independently of all other edges, we have

$$R_{n,m} = (n-m)^m p^m \exp\left\{\left[m(n-m-1) + \binom{m}{2}\right] \log q\right\} P_{n-m}$$
$$= \{(n-m) p q^{(2n-m-3)/2}\}^m P_{n-m},$$

where P_{n-m} is the probability that a subgraph on $n-m$ vertices is connected, given by (1). The probabilities $R_{n,m}$ are equal for all possible m-subsets of vertices, so the sum $S_{n,m}$ of $R_{n,m}$ over all such subsets is given by (2). Put $S_{n,0} = P_n$. Then, by the application of the principle of inclusion and exclusion (see, e.g., [2], ch. 4), we get the probability of the existence of a connected graph with exactly k $(0 \leqslant k \leqslant n-1)$ pendant vertices.

Let us notice that, for $n \geqslant 3$, $\Pr\{V(n, n)\} = 0$, since it is impossible that a connected graph of order $n \geqslant 3$ has all pendant vertices.

We have computed numerical values of $\Pr\{V(20, k)\}$ which appear in Table 1. We give these probabilities up to $k = 6$, since, for successive

Table 1. The numerical values of $\Pr\{V(20, k)\}$ for some k and p

k	The edge probability p				
	0.10	0.15	0.20	0.25	0.30
0	0.000259	0.025705	0.214445	0.553471	0.815896
1	0.001487	0.068650	0.256991	0.271272	0.144591
2	0.004202	0.093481	0.164269	0.075070	0.015487
3	0.007709	0.085817	0.073997	0.015473	0.001321
4	0.010212	0.059146	0.026152	0.002240	0.000099
5	0.010255	0.032230	0.007640	0.000393	0.000007
6	0.007968	0.014289	0.001892	0.000052	0.0000005
P_{20}	0.050061	0.386284	0.745872	0.918378	0.977402

$k \geqslant 7$, $\Pr\{V(20, k)\}$ tends rapidly to 0. The last row of Table 1 contains the probability that $G_{20,p}$ is connected, which is the sum of $\Pr\{V(20, k)\}$ over $k = 0, 1, \ldots, 19$.

From Theorem 1 we shall obtain the probability $\Pr\{W(n, n+l, k)\}$ that $G_{n,p}$ is a connected $(n, n+l)$ graph with k pendant vertices, where by an (n, m) ***graph*** we mean a graph which has n labeled vertices, m edges and no loops or multiple edges. For this purpose we need the number $f(n, m)$ of connected (n, m) graphs. It is trivial that $f(n, n+l) = 0$ if $l < -1$. Cayley [1] proved that

$$f(n, n-1) = n^{n-2}, \tag{3}$$

and Rényi [9] found the formula for $f(n, n)$, i.e.

$$f(n, n) = \frac{1}{2}\sum_{s=3}^{n} s!\binom{n}{s} n^{n-s-1}. \tag{4}$$

Recently, Wright [11] derived the recurrence formula for $f(n, n+l)$ for successive l and n, namely

$$2(n+l+1)f(n, n+l+1) = 2\left(\binom{n}{2} - n - l\right) f(n, n+l) + \\ + \sum_{s=1}^{n-1}\binom{n}{s} s(n-s) \sum_{h=-1}^{l+1} f(s, s+h) f(n-s, n-s+l-h).$$

Using the exponential generating function of $f(n, n+l)$ Wright found also the exact formulae for $f(n, n)$, $f(n, n+1)$ and $f(n, n+2)$ which depend only on powers of n and on the number

$$h(n) = \sum_{s=1}^{n-1}\binom{n}{s} s^s (n-s)^{n-s},$$

and are of the forms

$$2f(n, n) = \big(h(n)/n\big) - n^{n-2}(n-1), \tag{5}$$

$$24 f(n, n+1) = n^{n-2}(n-1)(5n^2+3n+2) - 14h(n) \tag{6}$$

and

$$1152 f(n, n+2) \\ = (45n^2+386n+312)h(n) - 4n^{n-2}(n-1)(55n^3+36n^2+18n+12),$$

respectively. Now we can formulate the following result:

COROLLARY 1. *Let $n \geqslant 3$. Then for $k = 0, 1, \ldots, n-1$ and $l = -1, 0, 1, \ldots, n(n-3)/2$*

$$\Pr\{W(n, n+l, k)\} = g(n, n+l, k) p^{n+l} q^{(n^2-3n-2l)/2}, \tag{7}$$

where

$$(8)\qquad g(n, n+l, k) = \sum_{m=k}^{n-1} (-1)^{m+k}\binom{n}{m}\binom{m}{k}(n-m)^m f(n-m, n-m+l)$$

is the number of $(n, n+l)$ connected graphs with k pendant vertices.

Proof. In Theorem 1 we put

$$(9)\qquad P_{n-m} = f(n-m, n-m+l)p^{n-m+l}\exp\left\{\left[\binom{n-m}{2}-(n-m+l)\right]\log q\right\}.$$

According to formula (7) and to relations (3), (5) (or (4)) and (6) by the application of a computer calculations, we get the numerical values of $\Pr\{W(6, 6+l, k)\}$ which appear in Table 2.

Table 2. The numerical values of $\Pr\{W(6, 6+l, k)\}$ for some l, k and p

p	k	$l = -1$	$l = 0$	$l = 1$
0.3	1	—	0.03177	0.04123
	2	0.02471	0.05825	0.01815
	3	0.04942	0.01589	—
	4	0.01441	—	—
	5	0.00041	—	—
0.4	1	—	0.04458	0.08999
	2	0.02229	0.08173	0.03963
	3	0.04458	0.02229	—
	4	0.01300	—	—
	5	0.00037	—	—
0.5	1	—	0.03296	0.09979
	2	0.01099	0.06042	0.04395
	3	0.02197	0.01648	—
	4	0.00641	—	—
	5	0.00018	—	—

Remark 1. For fixed values of n, l and k it is easy to see that $\Pr\{W(n, n+l, k)\}$ assumes the maximal value for the edge probability $p = (n+l)/\binom{n}{2}$. On the other hand, the number of edges in $G_{n,p}$ is a random variable with the expectation equal to $\binom{n}{2}p$; thus to obtain a random graph $G_{n,p}$ having on the average $n+l$ edges we have to choose the value of p equal to $(n+l)/\binom{n}{2}$.

Remark 2. Rényi [8] found the formula for the number of trees with k pendant vertices, i.e.

$$(10)\qquad g(n, n-1, k) = \frac{n!}{k!}\mathfrak{S}_{n-2}^{n-k},$$

where $\mathfrak{S}_n^m$ is the Stirling number of the second kind. The equivalence of (10) and (8) for $l = -1$ follows from the fact that

$$(11)\qquad m!\,\mathfrak{S}_n^m = \sum_{s=0}^{m-1}(-1)^s\binom{m}{s}(m-s)^n$$

(see, e.g., [10], ch. 4). As a matter of fact, putting in (8) $l = -1, f(n, n-1) = n^{n-2}$, $m = k+s$ and next applying (11) we get (10).

Let v_n be a random variable denoting the number of pendant vertices of a random graph $G_{n,p}$. It is known (see [4]) that

$$(12)\qquad \mu_{[m]} = m!\,S_{n,m},$$

where $\mu_{[m]}$ is the m-th factorial moment of the distribution. So putting in (9) $l = -1$ and then setting it to (2) we obtain, according to (12), the following

COROLLARY 2. *If $G_{n,p}$ is a tree, then the first and the second moments of the random variable v_n are*

$$(13)\qquad \mathrm{E}\{v_n\} = n\left(1-\frac{1}{n}\right)^{n-2}Q_n = AQ_n$$

and

$$(14)\qquad \mathrm{E}\{v_n^2\} = \left\{n(n-1)\left(1-\frac{2}{n}\right)^{n-2}+n\left(1-\frac{1}{n}\right)^{n-2}\right\}Q_n = BQ_n,$$

respectively, where $Q_n = n^{n-2}p^{n-1}q^{(n-1)(n-2)/2}$ is the probability of appearance of a tree on n vertices.

Remark 3. If a random tree T_n of order n means a randomly chosen tree from the whole collection of n^{n-2} equiprobable trees and u_n denotes the number of pendant vertices of T_n, then according to Rényi's results (see [8]) we have $\mathrm{E}\{u_n\} = A$ and $\mathrm{E}\{u_n^2\} = B$, where A and B are defined by (13) and (14), respectively.

The following corollary states that for a large value of n and for every fixed edge probability $p > 0$ the random graph $G_{n,p}$ contains no vertex of degree one.

COROLLARY 3. *For every fixed $p > 0$*

$$\Pr\{V(n, 0)\} \to 1 \quad \text{as } n \to \infty.$$

Proof. From the inequality of Bonferroni (see, e.g., [2], ch. 4) and (2) we have

$$P_n - n(n-1)pq^{n-2}P_{n-1} \leqslant \Pr\{V(n,0)\} \leqslant P_n.$$

Since, for every fixed $p > 0$, $P_n \to 1$ as $n \to \infty$ (see [5]), we obtain the assertion.

3. Pendant vertices of arbitrary graphs. Here we derive a formula for the probability distribution of the number of pendant vertices in a random graph $G_{n,p}$ but no matter whether or not it is connected. Denote by $\Pr\{U(n,k)\}$ the probability that $G_{n,p}$ has exactly k pendant vertices, where $k = 0, 1, \ldots, n$. As usual, for every x and every natural m we set

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

and let $[x]$ denote the greatest integer not greater than x. Applying an analogous method of proof as in Theorem 1 we will show the following

THEOREM 2. *Let $n \geqslant 1$. Then for $k = 0, 1, \ldots, n$*

$$\Pr\{U(n,k)\} = \sum_{m=k}^{n} (-1)^{m+k} \binom{m}{k} S_{n,m},$$

where

$$S_{n,m} = \binom{n}{m} \sum_{i=0}^{[m/2]} \frac{m!(n-m)^{m-2i}}{i!\big((m-2i)!\big)\cdot 2^i} p^{m-i} q^{i+m(2n-m-3)/2} \quad \textit{if } 0 \leqslant m \leqslant n-1, \tag{15}$$

and

$$S_{n,n} = \begin{cases} 0 & \textit{if } n \textit{ is odd}, \\ (n)_{n/2}(pq^{n-2}/2)^{n/2} & \textit{if } n \textit{ is even}. \end{cases}$$

Proof. Let $R_{n,m}$ be the probability that m $(1 \leqslant m \leqslant n-1)$ fixed vertices are pendant. Since $G_{n,p}$ is not neccessarily connected, i $(i = 0, 1, \ldots, [m/2])$ pairs of these m vertices can form connected components, each of size two, which can be done in

$$\binom{m}{2}\binom{m-2}{2}\ldots\binom{m-2(i-1)}{2}\Big/ i! = \frac{m!}{i!\big((m-2i)!\big)\cdot 2^i}$$

ways, and other $m-2i$ vertices are joined to some vertices chosen from $n-m$ remaining vertices; the number of all such connections is equal to $(n-m)^{m-2i}$. Thus

$$R_{n,m} = \sum_{i=0}^{[m/2]} \frac{m!(n-m)^{m-2i}}{i!\big((m-2i)!\big)\cdot 2^i} p^{m-i} q^{i+m(2n-m-3)/2},$$

and the sum $S_{n,m}$ of these probabilities over all m-subsets is given by (15). It is evident that $S_{n,n} = 0$ if n is odd since exactly one vertex remains always isolated. If n is even, then $S_{n,n}$ is the probability that $G_{n,p}$ is a forest in which each of the $n/2$ components has size two. So

$$S_{n,n} = (n)_{n/2}(pq^{n-2}/2)^{n/2}.$$

Put $S_{n,0} = 1$. According to the principle of inclusion and exclusion we get our assertion.

Table 3. The numerical values of $\Pr\{U(20, k)\}$ for some k and p

k	The edge probability p				
	0.01	0.05	0.10	0.20	0.30
0	0.148346	0.000133	0.002450	0.272765	0.833700
1	0.000127	0.000725	0.015109	0.336408	0.148620
2	0.346934	0.006661	0.046621	0.224277	0.016148
3	0.005136	0.016079	0.095563	0.107144	0.001412
4	0.311052	0.058325	0.145750	0.041012	0.000111
5	0.006594	0.081873	0.175382	0.013330	0.000008
6	0.139152	0.169489	0.172633	0.003805	0.0000006
7	0.003117	0.150462	0.141580	0.000974	0.00000004
8	0.033851	0.204657	0.098550	0.000226	0.000000003
9	0.000685	0.117510	0.057965	0.000048	0.0000000002
10	0.004576	0.113012	0.029481	0.000009	0.00000000001

The numerical values of $\Pr\{U(20, k)\}$ appear in Table 3. This example shows us a rather surprising behaviour of $\Pr\{U(20, k)\}$ with respect to changes of the edge probability p. For example, if $p = 0.01$, then

$$\Pr\{U(20, 2k-1)\} < \Pr\{U(20, 2k)\}, \qquad k = 1, 2, \ldots, 10,$$

while for $p \geqslant 0.2$

$$\Pr\{U(20, k+1)\} < \Pr\{U(20, k)\}, \qquad k = 1, 2, \ldots, 19.$$

Comparing the probabilities $\Pr\{U(20, k)\}$ with $\Pr\{V(20, k)\}$ one can see a very small difference between these values when the edge probability p is greater than or equal to 0.3.

We show now that, for $p = p(n) = 1/(n-1)$, $G_{n,p}$ has on the average approximately n/e pendant vertices and that the variance of v_n, i.e. the number of pendant vertices of $G_{n,p}$, is asymptotically equal to $n(e-1)/e^2$.

COROLLARY 4. *Let* $p = p(n) = 1/(n-1)$. *Then*

$$\lim_{n\to\infty} \frac{\mathrm{E}\{v_n\}}{n} = \frac{1}{e}$$

and

$$\lim_{n\to\infty} \frac{\mathrm{Var}\{v_n\}}{n} = \frac{e-1}{e^2}.$$

Proof. Let $n \geqslant 3$. Then from (12) and (15) we obtain

$$\mathrm{E}\{v_n\} = \mu_{[1]} = n(n-1)pq^{n-2}$$

and

$$\begin{aligned}\mathrm{Var}\{v_n\} &= \mu_{[2]} + \mu_{[1]} - \mu_{[1]}^2 \\ &= n(n-1)pq^{n-2}\{(n-2)^2pq^{n-3} + q^{n-2} + 1\} - n^2(n-1)^2p^2q^{2n-4}.\end{aligned}$$

Now, setting $p = 1/(n-1)$, by a routine calculation we get

$$\mathrm{E}\{v_n\} = n\left(1 - \frac{1}{n}\right)^{n-2}$$

and

$$\mathrm{Var}\{v_n\} = n\left\{\left(1 - \frac{1}{n-1}\right)^{n-2} - \left(1 - \frac{1}{n-1}\right)^{2n-4}\right\},$$

whence we obtain the required asymptotic relations.

Let us notice that for fixed values of n the expectation of v_n takes the maximal value for $p = 1/(n-1)$. One can also see that for such a p a random graph $G_{n,p}$ has on the average $n/2$ edges.

Remark 4. Rényi [8] has shown that the number u_n of pendant vertices of a random tree T_n (for the definition of T_n see Remark 3) satisfies

$$\lim_{n\to\infty} \frac{\mathrm{E}\{u_n\}}{n} = \frac{1}{e} \quad \text{and} \quad \lim_{n\to\infty} \frac{\mathrm{Var}\{u_n\}}{n} = \frac{e-2}{e^2}.$$

Finally, we give the asymptotical property of $\Pr\{U(n, 0)\}$. We have

COROLLARY 5. *For every fixed* $p > 0$

$$\Pr\{U(n, 0)\} \to 1 \quad \textit{as } n \to \infty.$$

Proof. From the Bonferroni inequality we obtain

$$1 - n(n-1)pq^{n-2} \leqslant \Pr\{U(n, 0)\} \leqslant 1,$$

so if $n \to \infty$, we get the assertion.

REFERENCES

[1] A. Cayley, *A theorem on trees*, The Quarterly Journal of Mathematics 23 (1969), p. 376-378.

[2] W. Feller, *An introduction to probability theory and its applications*, Vol. 1, 2nd edition, New York 1961.

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

[4] M. Fréchet, *Les probabilités associées à un système d'événements compatibles et dépendants*, Actualités scientifiques et industrielles, Paris 1940-1943.

[5] E. N. Gilbert, *Random graphs*, Annals of Mathematical Statistics 30 (1959), p. 1141-1144.
[6] A. Meir and J. W. Moon, *On nodes of degree two in random trees*, Mathematika 15 (1968), p. 188-192.
[7] N. S. Na and A. Rapoport, *Distribution of nodes of a tree by degree*, Mathematical Biosciences 6 (1970), p. 313-329.
[8] A. Rényi, *Some remarks on the theory of trees*, Publications of the Mathematical Institute of the Hungarian Academy of Sciences 4 (1959), p. 73-85.
[9] — *On connected graphs I*, ibidem 4 (1959), p. 385-388.
[10] I. Tomescu, *Introducere in combinatorica*, Bucharest 1972. English edition: *Introduction to combinatorics*, Collet's (Publishers) Ltd., 1975.
[11] E. M. Wright, *The number of connected sparsely-edged graphs*, Journal of Graph Theory 1 (1977), p. 317-330.

INSTITUTE OF MATHEMATICS
A. MICKIEWICZ UNIVERSITY
POZNAŃ

Reçu par la Rédaction le 22. 9. 1978