ON THE EXISTENCE OF A FACTOR OF DEGREE ONE OF A CONNECTED RANDOM GRAPH

By

P. ERDŐS and A. RÉNYI (Budapest), members of the Academy

§ 0. Introduction

In a series of papers (see [1], [2], [3]) we have considered the structure of a random graph $\Gamma_{n,N}$ obtained as follows: we select at random N edges among the $\binom{n}{2}$ possible edges connecting n given points so that each of the $\binom{\binom{n}{2}}{N}$ possible choices has the same probability $\binom{\binom{n}{2}}{N}^{-1}$. In [1] we have proved that if n and N tend both to $+\infty$ so that

$$N = \tfrac{1}{2}\, n \log n + cn + o(n) \tag{0. 1}$$

then denoting by C_0 the class of connected graphs and by $P_{n,N}(A)$ the probability that the graph $\Gamma_{n,N}$ belongs to the class A, we have

$$\lim_{\substack{n\to\infty \\ N=\frac{1}{2}n\log n+cn+o(n)}} P_{n,N}(C_0) = e^{-e^{-2c}}. \tag{0. 2}$$

It follows from (0. 2) that

$$\lim_{n\to\infty} P_{n,N}(C_0) = \begin{cases} 1 & \text{if} \quad n\to\infty \quad \text{and} \quad \dfrac{N-\frac{1}{2}n\log n}{n} \to +\infty \\ 0 & \text{if} \quad n\to\infty \quad \text{and} \quad \dfrac{N-\frac{1}{2}n\log n}{n} \to -\infty. \end{cases} \tag{0. 3}$$

We have proved also that if C_k denotes the class of all graphs consisting of a connected component and k isolated points, we have

$$\lim_{\substack{n\to\infty \\ N=\frac{1}{2}n\log n+cn+o(n)}} P_{n,N}(C_k) = \frac{\exp(-2kc - e^{-2c})}{k!} \qquad (k=1, 2, \ldots). \tag{0. 4}$$

Note that the limits on the right of (0. 4) are the terms of a Poisson distribution with mean value e^{-2c}, and thus their sum for $k=0, 1, \ldots$ equals 1. Thus in case $N=\frac{1}{2}\, n \log n + O(n)$ the random graph $\Gamma_{n,N}$ consists with probability tending to 1 of a connected component and a certain number of isolated points.

A graph G is said to have *a factor of degree one* if one can select such a subset S of the edges of G that each point P of G is the endpoint of one and only one edge belonging to the set S. Clearly a trivial necessary condition for the existence of a factor of degree one of a graph G is that the number of points of G should be even.

According to a celebrated theorem of TUTTE [4] a graph has a factor of degree one if and only if deleting arbitrarily r points of G ($r=0, 1, 2, \ldots$) among the connected components of the remaining graph G^* the number of connected components consisting of an odd number of points is less than $r+1$.

Note that if G has n points and n is even, and G^* is obtained by deleting r points of G (of course the edges at least one of the endpoints of which is deleted are also deleted) and if G^* contains t odd components then $t \equiv r \bmod 2$; thus the theorem of Tutte can be formulated also as follows: A graph G has a factor of degree one if and only if the number of points of G is even and if deleting arbitrarily r points of G ($r=0, 1, \ldots$) among the connected components of the remaining graph G^* the number of odd components is less than $r+2$.

In the present paper we deal with the question: what is the probability that the random graph $\Gamma_{n,N}$ has a factor of degree one?

Clearly a graph having at least one isolated point cannot possess a factor of degree one. As according to our mentioned result in case $N=\frac{1}{2}n\log n+cn+o(n)$ the random graph $\Gamma_{n,N}$ contains with probability not tending to 0 isolated points, it seems natural to deal only with the case when $\dfrac{N-{}^1/_2\, n\log n}{n}\to+\infty$; of course one has to suppose also that n is even, $n=2m$. For this case we shall prove the following

THEOREM 1. *Let us suppose that n is even, $n=2m$ further that*

$$(0.5)\qquad n\to\infty \quad \text{and} \quad N=\tfrac{1}{2}n\log n+\omega(n)n \quad \text{where} \quad \lim_{n\to+\infty}\omega(n)=+\infty.$$

Let F denote the class of graphs containing a factor of degree one; then we have under condition (0.5)

$$(0.6)\qquad \lim P_{n,N}(F)=1.$$

If $N=\frac{1}{2}n\log n+O(n)$, as mentioned above, with probability near to 1 $\Gamma_{n,N}$ consists of a connected component and a certain number of isolated points. With the same method as used to prove Theorem 1 one can prove that if the connected component of $\Gamma_{n,N}$ consists of an even number of points, it has with probability near to 1 a factor of degree one. As the proof of this result is almost the same as that of Theorem 1, we do not go into the details.

It should be mentioned that the results of the present paper are closely related to a previous result of ours (see [5]) concerning random zero-one matrices. As a matter of fact to every n by n zero-one matrix M there corresponds an even graph G, namely the graph consisting of n „red" and n „blue" points in which there is no edge connecting two points having the same colour and the j-th red point is connected with the k-th blue point if and only if the k-th entry of the j-th row of M is 1. Now clearly the graph G has a factor of degree one if and only if the permanent of M is positive; more exactly the permanent of M is equal to the number of different factors of degree one of G.

Now in [5] we have proved that if the n by n random zero-one matrix $M_{n,N}$ is obtained by choosing at random with uniform distribution one of the $\binom{n^2}{N}$ possible n by n zero-one matrices which contain N ones and n^2-N zeros, then the probability that $M_{n,N}$ has a positive permanent tends to 1 if $n\to+\infty$ and $\frac{N-n\log n}{n}\to+\infty$. According to what has been said above, this result can be interpreted as a result concerning the existence of a factor of degree one of the even graph corresponding to the matrix $M_{n,N}$. It should be added that the problem investigated in the present paper is much more difficult than the corresponding problem for even graphs solved in [5]. Thus for instance in [5] we made use of the well known theorem of D. KÖNIG; the corresponding tool in the present paper is the much deeper theorem of TUTTE mentioned above.

In § 1 we collected some simple inequalities needed in the sequel. § 2 contains the proof of Theorem 1.

§ 1. Some inequalities

The following (well known or obvious) inequalities will be needed.

(1. 1) $$\binom{n}{r}\leqq\frac{n^r}{r!}\qquad (r=1,2,\dots;\quad n=r,r+1,\dots),$$

(1. 2) $$\frac{x-a}{y-a}\leqq\frac{x}{y}\quad\text{for}\quad 0<a\leqq x\leqq y,$$

(1. 3) $$\frac{\binom{B-A}{b-a}}{\binom{B}{b}}\leqq\left(1-\frac{A-a}{B-a}\right)^{b-a}\left(\frac{b}{B}\right)^a$$

if $0\leqq b-a\leqq B-A,\ 0\leqq a\leqq b\leqq B,\ 0\leqq A.$

(1. 4) $$1-x\leqq e^{-x}\quad\text{if}\quad x\geqq 0.$$

(1. 5) $$n!\geqq\left(\frac{n}{e}\right)^n\quad\text{for}\quad n\geqq 1.$$

(1. 6) If $\lambda>1,\quad 0<\delta<\frac{1}{\lambda e}$

we have

$$\sum_{n\delta\lambda e\leqq r\leqq n}\binom{n}{r}\delta^r=O\left(\frac{1}{\lambda^{\delta\lambda en}}\right).$$

(1. 7) For $0<p<\frac{1}{2}$ one has

$$\sum_{k\leqq np}\binom{n}{k}=O\left(\sqrt{n}\cdot e^{nh(p)}\right)$$

where $h(p)=p\log\frac{1}{p}+(1-p)\log\frac{1}{1-p}$.

(1. 8) If $0<\alpha<1$, $c>0$

$$\sum_{\frac{nc}{\log n}\leqq k}\binom{n}{k}\frac{1}{n^{\alpha k}}=e^{-\alpha c n+o(n)}.$$

(1. 9)
$$1-x+\frac{x^2}{2}\leqq e^{-x+\frac{x^3}{3}}\quad\text{if}\quad 0\leqq x<1.$$

(1. 10)
$$\binom{n}{k}\leqq\frac{n^k}{k!}e^{-\frac{k(k-1)}{2n}}\quad\text{if}\quad 0<k\leqq n.$$

§ 2. Proof of Theorem 1

The proof consists of 8 steps. A characteristic feature of the proof is that we prove certain assertions first in a weak form and use this weak result to improve itself.

1st step. Let A_r denote the class of all graphs from which one can omit r points so that among the components of the remaining graph there are $\geqq r+2$ odd components. In what follows $o(1)$ will denote a quantity which tends to 0 if n and N tend to $+\infty$ so that condition (0. 5) is satisfied.

LEMMA 1.
$$\sum_{r\geqq n\sqrt{\frac{2}{\log n}}}P_{n,N}(A_r)=o(1).$$

PROOF OF LEMMA 1. Let us consider a graph G belonging to the class A_r. Let us choose in each of $r+2$ odd components one point. Clearly these points cannot be connected with each other in G. Thus we get by using (1. 3) and (1. 4) for sufficiently large n

$$P_{n,N}(A_r)\leqq\binom{n}{r+2}\frac{\binom{\binom{n}{2}-\binom{r+2}{2}}{N}}{\binom{\binom{n}{2}}{N}}\leqq\binom{n}{r+2}\left(1-\frac{(r+2)(r+1)}{n(n-1)}\right)^N\leqq\binom{n}{r}e^{-n}$$

and thus

$$\sum_{r\geqq n\sqrt{\frac{2}{\log n}}}P_{n,N}(A_r)\leqq\left(\frac{2}{e}\right)^n.$$

2nd step. We prove now

LEMMA 2.
$$\sum_{\frac{4n\log\log n}{\log n}<r<n\sqrt{\frac{2}{\log n}}}P_{n,N}(A_r)=o(1).$$

PROOF OF LEMMA 2. According to [1] we may suppose that G is connected because under condition (0. 5) the probability of $\Gamma_{n,N}$ being disconnected is $o(1)$. Suppose now that $\Gamma_{n,N}$ belongs to the class A_r. Then by selecting r points from $\Gamma_{n,N}$ (we shall call these r points "*separating points*") the remaining graph contains $\geqq r+2$ odd components.

From each of these components we can select at least one point which is connected with a separating point. We shall call these "*contact-points*". These points are certainly not connected with each other in $\Gamma_{n,N}$. The r separating points can be selected in $\binom{n}{r}$ ways; the mentioned contact points in the components can be selected in $\binom{n-r}{r+2}$ ways; each of these points may be connected with any one of the r separating points. As the contact-points can not be connected with each other, for the choice of the remaining $N-(r+2)$ edges of the graph there are only $\binom{n}{2}-\binom{r+2}{2}$ possibilities. Thus we get

$$P_{n,N}(A_r) \leqq \binom{n}{r}\binom{n-r}{r+2} r^{r+2} \frac{\binom{\binom{n}{2}-\binom{r+2}{2}}{N-(r+2)}}{\binom{\binom{n}{2}}{N}}$$

and thus, using the inequalities (1. 1)—(1. 6) we obtain

$$\sum_{\frac{4n\log\log n}{\log n} \leqq r < n\sqrt{\frac{2}{\log n}}} P_{n,N}(A_r) \leqq \log^2 n \sum_{\frac{4n\log\log n}{\log n} < r} \binom{n}{r}\left(\frac{3}{\log n}\right)^r = O\left(e^{-\frac{n}{\sqrt{\log n}}}\right).$$

3*rd step*. Let $D(d)$ denote the set of those graphs G from the points of which one can select a subset S_1 of $k \geqq d$ points and another subset S_2 disjoint from S_1 of $l \geqq d$ points so that no point belonging to the set S_1 is connected with a point belonging to the set S_2 in G.

LEMMA 3.

$$P_{n,N}\left(D\left(n\sqrt{\frac{2}{\log n}}\right)\right) = o(1).$$

PROOF OF LEMMA 3. Clearly for sufficiently large values of n

$$P_{n,N}\left(D\left(n\sqrt{\frac{2}{\log n}}\right)\right) \leqq \sum_{k,\, l \geqq n\sqrt{\frac{2}{\log n}}} \sum \binom{n}{k}\binom{n}{l} \frac{\binom{\binom{n}{2}-kl}{N}}{\binom{\binom{n}{2}}{N}} \leqq \left(\frac{2}{e}\right)^{2n}.$$

4*th step*. We need the following very simple

LEMMA 4. *If $a_0 \geqq a_1 \geqq \ldots \geqq a_s > 0$ is a sequence of positive numbers with sum $A = \sum_{i=0}^{s} a_i$ such that there does not exist a number j $(0 \leqq j \leqq s)$ for which both $a_0 + \ldots + a_j \geqq B$ and $a_{j+1} + \ldots + a_s \geqq B$ where $0 < B \leqq A/3$, then we have $a_0 > A - B$.*

PROOF OF LEMMA 4. If $B \leqq a_0 \leqq A - B$ then $a_1 + \ldots + a_s = A - a_0 \geqq A - (A-B) = B$ which contradicts our hypothesis. Thus either $a_0 < B$ or $a_0 > A - B$; in the first case let i be defined by $a_0 + \ldots + a_{i-1} < B \leqq a_0 + \ldots + a_i$; it follows $a_0 + \ldots + a_i < B + a_0 < 2B$ and thus $a_{i+1} + \ldots + a_s = A - (a_0 + \ldots + a_i) > A - 2B \geqq B$ which is again a contradiction. Thus we have $a_0 > A - B$.

Let $\Gamma_{n,N}$ belong to the class A_r. According to Lemmas 1 and 2 we may suppose that

$$r < \frac{4n \log\log n}{\log n}.$$

Let $a_0 \geqq a_1 \geqq \ldots \geqq a_s$ $(s \geqq r+2)$ denote the numbers of points of the connected components of the graph $\Gamma_{n,N}$ after the r separating points have been removed. In view of Lemma 3 one can suppose that the sequence a_i satisfies the conditions of Lemma 4 with $A = n - r$, $B = \dfrac{\sqrt{2}\, n}{\sqrt{\log n}}$; thus by Lemma 4

$$a_0 \geqq n - \frac{4n \log\log n}{\log n} - \frac{\sqrt{2}\, n}{\sqrt{\log n}}.$$

Thus

$$a_1 + \ldots + a_s < \frac{4n \log\log n}{\log n} + \frac{\sqrt{2} \cdot n}{\sqrt{\log n}}.$$

5th step. We prove first the following

LEMMA 5. *Let $H_n(k, r)$ denote the set of those graphs G having n points, among the points of which one can select a subset S_1 having k elements and another subset S_2 disjoint from S_1 having $n-k-r$ elements such that no point in S_1 is connected with a point in S_2 by an edge of G. Then if $0 < \delta < \frac{1}{2}$, $0 < \varepsilon < \frac{1}{2}$ and $c > \dfrac{h(\delta)}{1 - \varepsilon - \delta}$, we have*

$$\sum_{\substack{\frac{cn}{\log n} < k < \varepsilon n \\ 0 \leqq r < \delta n}} P_{n,N}(H_n(k, r)) = o(1).$$

PROOF OF LEMMA 5.

$$P_{n,N}(H_n(k,r)) \leqq \binom{n}{k}\binom{n-k}{r} \frac{\binom{\binom{n}{2} - k(n-k-r)}{N}}{\binom{\binom{n}{2}}{N}} \leqq$$

$$\leqq \binom{n}{k}\binom{n}{r} e^{-\frac{k(n-k-r)}{n}\log n}.$$

Thus, using (1. 7) we get

$$\sum_{\substack{0 \leqq r < \delta n \\ \frac{cn}{\log n} < k < \varepsilon n}} \sum P_{n,N}\big(H_n(k,r)\big) = O\left(\sqrt{n}\, e^{nh(\delta)} \sum_{\frac{cn}{\log n} < k} \binom{n}{k} \frac{1}{n^{k(1-\varepsilon-\delta)}}\right)$$

and therefore by (1. 8) it follows

$$\sum_{\substack{0 \leqq r < \delta n \\ \frac{cn}{\log n} < k < \varepsilon n}} \sum P_{n,N}\big(H_n(k,r)\big) = O\big(\sqrt{n}\; e^{n(h(\delta) - c(1-\varepsilon-\delta)) + o(n)}\big).$$

Thus our lemma is proved.

Applying Lemma 5 with

$$\delta = \frac{4 \log\log n}{\log n}, \quad \varepsilon = \frac{4 \log\log n}{\log n} + \sqrt{\frac{2}{\log n}}$$

we obtain that we may suppose

$$a_1 + \ldots + a_s < \frac{8n \log\log^2 n}{\log^2 n}$$

because the probability of the opposite inequality tends to 0. Now we may suppose without restricting the generality that $s = r+1$ and that the numbers $a_1, \ldots, a_{r+1}$ are all odd. Thus we may suppose

$$r+1 \leqq a_1 + \ldots + a_{r+1} = k < \frac{8n \log\log^2 n}{\log^2 n}.$$

6th step. We have reduced the problem to the investigation of graphs having $r < \dfrac{8n \operatorname{loglog}^2 n}{\log^2 n}$ separating points such that after removing these points the remaining graph contains $r+1$ odd components with $a_1, \ldots, a_{r+1}$ points, such that $k = a_1 + \ldots + a_{r+1} < \dfrac{8n \operatorname{loglog}^2 n}{\log^2 n}$. Let us denote by s the number of edges connecting one of the separating points with one of the k points belonging to the $r+1$ odd components. We distinguish two cases. Either $s \geqq r+8$ or $s < r+8$. We deal first with the case $s \geqq r+8$. The probability of such a configuration clearly does not exceed

$$\Delta = \sum_{\substack{r < \frac{8n \log\log^2 n}{\log^2 n} \\ r+1 \leqq k < \frac{8n \log\log^2 n}{\log^2 n}}} \binom{n}{r} \binom{n-r}{k} \sum_{r+8 \leqq s \leqq kr} \binom{kr}{s} \frac{\dbinom{\binom{n}{2} - k(n-k)}{N-s}}{\dbinom{\binom{n}{2}}{N}}.$$

Using the inequalities (1. 1)—(1. 10) we obtain

$$\Delta \leqq \sum_{\substack{r<\frac{8\log\log^2 n}{\log^2 n} \\ r+1\leqq k<\frac{8n\log\log^2 n}{\log^2 n}}} \frac{n^{k+r}}{k!\,r!} \sum_{r+8\leqq s\leqq kr} \binom{kr}{s}\left(\frac{\log n}{n}\right)^s \cdot \frac{1}{n^k} = O(n^{e^2-8}) = o(1).$$

7th step. It remains to deal with the case when $s \leqq r+7$. We shall prove that one can suppose that $k \leqq s$ because the probability of $k \geqq s+1$ is $o(1)$. As a matter of fact the following lemma is valid.

LEMMA 6. *Let E_k denote the set of those graphs which contain a subset S of k points which are connected by $\leqq k-1$ edges with points outside S. Then we have*

$$\sum_{1\leqq k\leqq \frac{n}{2\log n}} P_{n,N}(E_k) = o(1).$$

PROOF OF LEMMA 6. We shall call a subset S of a graph G a *loosely joined set* if S has k points and the number of edges connecting points of S with points outside S is less than k. We shall call S a *primitive loosely joined set* if it is loosely joined and no one of its proper subsets is loosely joined. It is easy to see that a primitive loosely joined set is a connected subgraph. As a matter of fact if S were a primitive loosely joined set, which can be splitted into two sets S_1 and S_2 such that there is no edge in G between points of S_1 and points of S_2, let k_1 and k_2 denote the number of points in S_1 and S_2 respectively $(k_1+k_2=k)$ and j_1 resp. j_2 denote the number of edges going out from the set S_1 resp. S_2. According to our supposition S is primitive loosely connected; thus $j_1 \geqq k_1$ and $j_2 \geqq k_2$ which implies $j_1+j_2 \geqq k$; thus S is not loosely connected, which contradicts our supposition.

Let E_k^* denote the class of graphs containing a primitive loosely connected subset of order k. As $\sum_{k\leqq K} E_k = \sum_{k\leqq K} E_k^*$ it is sufficient to prove that

$$\sum_{1\leqq k\leqq \frac{n}{2\log n}} P_{n,N}(E_k^*) = o(1).$$

Now evidently

$$P_{n,N}(E_1^*) = o(1)$$

because a primitive loosely connected set of order 1 is nothing else than an isolated point, and we have proved that the probability of $\Gamma_{n,N}$ containing an isolated point tends to 0 for $n\to\infty$, $N=\frac{1}{2}n\log n + n\omega(n)$ with $\omega(n)\to\infty$.

Thus we have to prove only that

$$\sum_{2\leqq k\leqq \frac{n}{2\log n}} P(E_k^*) = o(1).$$

Clearly a primitive loosely connected set of order 2 is simple a pair of points P_1, P_2 such that P_1 and P_2 are connected by an edge, further P_1 is connected with a single other point P_3 by an edge and P_2 is not connected with any other point except P_1. Thus

$$P_{n,N}(E_2^*) \leqq n(n-1)(n-2)\frac{\binom{\binom{n}{2}-(2n-3)}{N-2}}{\binom{\binom{n}{2}}{N}} = O\left(\frac{\log^2 n}{n}\right).$$

As any connected graphs of order k contains at least one spanning tree (having $k-1$ edges) and from k points a tree can be formed in k^{k-2} different ways, we have for $3 \leqq k \leqq \dfrac{n}{2\log n}$

$$P_{n,N}(E_k^*) \leqq \binom{n}{k} k^{k-2} \sum_{j=0}^{k-1} \binom{k(n-k)}{j} \frac{\binom{\binom{n}{2}-k(n-k)}{N-j-k+1}}{\binom{\binom{n}{2}}{N}} = O\left(\frac{\left(\frac{e^2 \log^2 n}{n}\right)^{k-1}}{k^2}\right).$$

Thus

$$\sum_{1 \leqq k \leqq \frac{n}{2\log n}} P_{n,N}(E_k^*) = o(1)$$

which proves Lemma 6.

8th step. We have reduced the problem to the study of graphs G which are of the following type:

a) There can be selected in G r separating points with $r \leqq \dfrac{8n \log\log^2 n}{\log^2 n}$ such that removing these the graph G falls into components among which there are $r+1$ odd components $C_1, \ldots, C_{r+1}$ of orders $a_1, \ldots, a_{r+1}$.

b) There are in G s edges connecting a separating point with a point in one of the components $C_1, \ldots, C_{r+1}$ where $r+1 \leqq s \leqq r+7$.

c) Putting $k = a_1 + \ldots + a_{r+1}$ we have $k \leqq s$.

As each of the components C_j $(j=1, \ldots, r+1)$ is connected with at least one separating point and the number of such points is less than the number of the C_j, there exist two components C_{i_1} and C_{i_2} which are connected with the same separating point P. Clearly the sum of orders a_{i_1} and a_{i_2} of C_{i_1} and C_{i_2} can not exceed 8, and the sum of the numbers b_1 and b_2 of edges going out from C_{i_1} and C_{i_2} does not exceed 8 either. Thus if K denotes the class of graphs with the above mentioned properties a), b), c), we have

$$P_{n,N}(K) \leqq \sum_{\substack{1 \leqq a_{i_1} \leqq 7 \\ 1 \leqq a_{i_2} \leqq 8-a_{i_1} \\ 1 \leqq b_1 \leqq 7 \\ 1 \leqq b_2 \leqq 8-b_1}} n \binom{n}{a_{i_1}} \binom{n}{a_{i_2}} \binom{na_{i_1}}{b_1-1} \binom{na_{i_2}}{b_2-1} \frac{\binom{\binom{n-a_{i_1}-a_{i_2}}{2}}{N-b_1-b_2}}{\binom{\binom{n}{2}}{N}}$$

from which we get

$$P_{n,N}(K) = O\left(\frac{\log^8 n}{n}\right).$$

Thus Theorem 1 is proved.

(Received 7 September 1965)

References

[1] P. ERDŐS—A. RÉNYI, On random graphs. I, *Publicationes Mathematicae (Debrecen)*, **6** (1959), pp. 290—297.
[2] P. ERDŐS—A. RÉNYI, On the evolution of random graphs, *Publ. Math. Inst. Hung. Acad. Sci.*, **5A** (1960), pp. 17—61.
[3] P. ERDŐS—A. RÉNYI, On the strength of connectedness of a random graph, *Acta Math. Acad. Sci. Hung.*, **12** (1961), pp. 261—267.
[4] W. T. TUTTE, The factorization of linear graphs, *Journal London Math. Soc.*. **22** (1947), pp. 107—111.
[5] P. ERDŐS—A. RÉNYI, On random matrices, *Publ. Math. Inst. Hung. Acad. Sci.*, **8A** (1963), pp. 455—461.