On random graphs I.

Dedicated to O. Varga, at the occasion of his 50th birthday.

By P. ERDŐS and A. RÉNYI (Budapest).

Let us consider a "random graph" $\Gamma_{n,N}$ having n possible (labelled) vertices and N edges; in other words, let us choose at random (with equal probabilities) one of the $\binom{\binom{n}{2}}{N}$ possible graphs which can be formed from the n (labelled) vertices $P_1, P_2, \ldots, P_n$ by selecting N edges from the $\binom{n}{2}$ possible edges $\overline{P_iP_j}$ $(1 \leqq i < j \leqq n)$. Thus the effective number of vertices of $\Gamma_{n,N}$ may be less than n, as some points P_i may be not connected in $\Gamma_{n,N}$ with any other point P_j; we shall call such points P_i *isolated* points. We consider the isolated points also as belonging to $\Gamma_{n,N}$. $\Gamma_{n,N}$ is called completely connected if it effectively contains all points $P_1, P_2, \ldots, P_n$ (i, e. if it has no isolated points) and is connected in the ordinary sense. In the present paper we consider asymptotic statistical properties of random graphs for $n \to +\infty$. We shall deal with the following questions:

1. What is the probability of $\Gamma_{n,N}$ being completely connected?

2. What is the probability that the greatest connected component (subgraph) of $\Gamma_{n,N}$ should have effectively $n-k$ points? $(k=0, 1, \ldots)$.

3. What is the probability that $\Gamma_{n,N}$ should consist of exactly $k+1$ connected components? $(k=0, 1, \ldots)$.

4. If the edges of a graph with n vertices are chosen successively so that after each step every edge which has not yet been chosen has the same probability to be chosen as the next, and if we continue this process until the graph becomes completely connected, what is the probability that the number of necessary steps ν will be equal to a given number l?

As (partial) answers to the above questions we prove the following four theorems. In Theorems 1, 2, and 3 we use the notation

$$N_c = \left[\frac{1}{2} n \log n + c n\right] \tag{1}$$

where c is an arbitrary fixed real number ($[x]$ denotes the integer part of x).

Theorem 1. *Let $P_0(n, N_e)$ denote the probability of Γ_{n,N_e} being completely connected. Then we have*

$$\lim_{n\to+\infty} P_0(n, N_e) = e^{-e^{-2c}}. \tag{2}$$

Theorem 2. *Let $P_k(n, N_e)$ $(k = 0, 1, \ldots)$ denote the probability of the greatest connected component of $\Gamma_{n,N}$ consisting effectively of $n-k$ points. (Clearly $P_0(n, N_e)$ has the same meaning as before.) Then we have*

$$\lim_{n\to+\infty} P_k(n, N_e) = \frac{(e^{-2c})^k e^{-e^{-2c}}}{k!}, \tag{3}$$

i. e. the number of points outside the greatest connected component of $\Gamma_{n,N}$ is distributed in the limit according to Poisson's law with mean value e^{-2c}.

Theorem 3. *Let $H_k(n, N_e)$ denote the probability of Γ_{n,N_e} consisting exactly of $k+1$ disjoint connected components. (Clearly $H_0(n, N_e) = P_0(n, N_e)$.) Then we have*

$$\lim_{n\to+\infty} H_k(n, N_e) = \frac{(e^{-2c})^k e^{-e^{-2c}}}{k!}, \tag{4}$$

i. e. the number of connected components of $\Gamma_{n,N}$ diminished by one is in the limit distributed according to Poisson's law with mean value e^{-2c}.

Theorem 4. *Let the edges of a random graph with possible vertices $P_1, P_2, \ldots, P_n$ be chosen successively among the possible edges $\overline{P_iP_j}$ in such a manner, that at each stage every edge which has not yet been chosen has the same probability to be chosen as the next, and let us continue this process until the graph becomes (completely) connected. Let ν_n denote the number of edges of the resulting connected random graph Γ. Then we have*

$$P\left(\nu_n = \left[\frac{1}{2} n \log n\right] + l\right) \sim \frac{2}{n} e^{-\frac{2l}{n} - e^{-\frac{2l}{n}}} \tag{5}$$

for $|l| = O(n)$ and

$$\lim_{n\to+\infty} P\left(\frac{\nu_n - \frac{1}{2} n \log n}{n} < x\right) = e^{-e^{-2x}}. \tag{6}$$

As regards the first question, previously P. Erdős and H. Whitney have obtained some less complete results. They proved that if $N > \left(\frac{1}{2} + \varepsilon\right) n \log n$ where $\varepsilon > 0$ then the probability of $\Gamma_{n,N}$ being connected tends to 1 if

$n \to +\infty$, but if $N < \left(\frac{1}{2} - \varepsilon\right) n \log n$ with $\varepsilon > 0$ then the probability of $\Gamma_{n,N}$ being connected, tends to 0 if $n \to +\infty$. They did not publish their result. This result is contained in Theorem 1.

Let $G_{n,N}$ denote a graph having n vertices and N edges. Let $C(n, N)$ denote the number of (completely) connected graphs $G_{n,N}$. Of course the solution of the above problems would be easy if a simple formula were known for $C(n, N)$. As a matter of fact, e. g. the probability $P_0(n, N)$ of $\Gamma_{n,N}$ being completely connected is given simply by

$$P_0(n, N) = \frac{C(n, N)}{\binom{\binom{n}{2}}{N}}. \tag{7}$$

Unfortunately such a simple formula is not known. The special case $N = n-1$ has been considered already by A. CAYLEY [1] who proved that $C(n, n-1) = n^{n-2}$ (for other proofs see O. DZIOBEK [2], A. PRÜFER [3], G. PÓLYA [4]). For the general case, it has been shown by R. I. RIDDELL and G. E. UHLENBECK [5] (see also E. N. GILBERT [6]), that

$$\sum_{n=1}^{\infty} \sum_{N=1}^{\infty} \frac{C(n, N) x^n y^N}{n!} = \log\left(1 + \sum_{k=1}^{\infty} \frac{(1+y)^{\binom{k}{2}} x^k}{k!}\right). \tag{8}$$

Relation (8) is a consequence of the following recursive formula for $C(n, N)$:

$$\binom{\binom{n+1}{2}}{N} = \sum_{m=0}^{n} \binom{n}{m} \sum_{M=m}^{\binom{m+1}{2}} C(m+1, M) \binom{\binom{n-m}{2}}{N-M}. \tag{9}$$

Unfortunately neither (8) nor (9) helps much to deduce the asymptotic properties of $C(n, N)$. In the present paper we follow a more direct approach.

All four theorems are based on the rather surprising Lemma given below.

Let us call the graph Γ_{n,N_c} with the n possible vertices $P_1, P_2, \ldots, P_n$ and N_c edges, of type A, if it consists of a connected graph having $n-k$ effective vertices and of k isolated points ($k = 0, 1, \ldots$). Any graph Γ_{n,N_c} which is not of type A shall be called to be of type $\bar{A}$. Then the following lemma holds.

Lemma. *Let $P(\bar{A}, n, N_c)$ denote the probability of Γ_{n,N_c} being of type $\bar{A}$. Then we have*

$$\lim_{n \to +\infty} P(\bar{A}, n, N_c) = 0. \tag{10}$$

Thus for a large n almost all graphs Γ_{n,N_c} are of type A.

PROOF OF THE LEMMA. Let M be a large positive number, which will be chosen later. All graphs G_{n,N_e} having the possible vertices $P_1, P_2, \ldots, P_n$ and N_e edges, where N_e is defined by (1) shall be divided into two classes: Those in which the greatest connected component consists of not less than $n-M$ points shall constitute the class E_M, all others the class $\bar{E}_M$. Let $\mathcal{N}(\bar{E}_M, n, N_e)$ denote the number of graphs G_{n,N_e} belonging to the class $\bar{E}_M$, i. e. the number of graphs in which the greatest connected component consists of less then $n-M$ points.

If the graph consists of r connected components having l_i points $(i=1, 2, \ldots, r)$ then

$$\sum_{i=1}^{r} l_i = n \quad \text{and} \quad \sum_{i=1}^{r} \binom{l_i}{2} \geqq N_e$$

and therefore if $L = \max_{(i)} l_i$ we have

$$\frac{L-1}{2} \geqq \frac{N_e}{n}$$

and thus $L > \frac{2N_e}{n}$; this implies that

$$M \leqq n - \frac{2N_e}{n}. \tag{11}$$

Therefore we have

$$\mathcal{N}(\bar{E}_M, n, N_e) \leqq \sum_{M<s<n-\frac{2N_e}{n}} \binom{n}{s} \binom{\binom{n}{2} - s(n-s)}{N_e} \tag{12}$$

because if the $n-s$ points belonging to the greatest connected component of the graph are fixed, the $s(n-s)$ edges connecting one of the points of this component with a point outside this component can not belong to the graph. Thus, denoting by $P(\bar{E}_M, n, N_e)$ the probability of Γ_{n,N_e} belonging to the class $\bar{E}_M$, we have

$$P(\bar{E}_M, n, N_e) = \frac{\mathcal{N}(\bar{E}_M, n, N_e)}{\binom{\binom{n}{2}}{N_e}} \leqq \sum_{M<s<n-\frac{2N_e}{n}} \binom{n}{s} \frac{\binom{\binom{n}{2} - s(n-s)}{N_e}}{\binom{\binom{n}{2}}{N_e}}. \tag{13}$$

We obtain for $n \geqq n_0$, by using some elementary estimations,

$$\binom{n}{s} \frac{\binom{\binom{n}{2}-s(n-s)}{N_c}}{\binom{\binom{n}{2}}{N_c}} \leqq \frac{e^{(3-2c)s}}{s!} \quad \text{for} \quad s \leqq \frac{n}{2} \tag{14}$$

and

$$\binom{n}{s} \frac{\binom{\binom{n}{2}-s(n-s)}{N_c}}{\binom{\binom{n}{2}}{N_c}} \leqq \frac{e^{(3-2c)(n-s)}}{(n-s)!} \quad \text{for} \quad s \geqq \frac{n}{2} \tag{15}$$

We obtain from (13), (14) and (15)

$$P(\bar{E}_M, n, N_c) \leqq \sum_{s > \frac{2N_c}{n}} \frac{e^{(3-2c)s}}{s!} + \sum_{s>M} \frac{e^{(3-2c)s}}{s!} \quad \text{for} \quad n \geqq n_0. \tag{16}$$

Thus we have

$$\lim_{n \to +\infty} P(\bar{E}_{\log\log n}, n, N_c) = 0. \tag{17}$$

Thus to prove our Lemma it is sufficient to show that denoting by $\bar{A}E_{\log\log n}$ the class of those graphs which belong to both $\bar{A}$ and $E_{\log\log n}$ and by $P(\bar{A}E_{\log\log n})$ the probability that Γ_{n, N_c} belongs to the class $\bar{A}E_{\log\log n}$ we have

$$\lim_{n \to +\infty} P(\bar{A}E_{\log\log n}, n, N_c) = 0. \tag{18}$$

Now we have

$$P(\bar{A}E_{\log\log n}, n, N_c) \leqq \sum_{s=2}^{\log\log n} \binom{n}{s} \left(\sum_{r=1}^{\binom{s}{2}} \binom{\binom{s}{2}}{r} \frac{\binom{\binom{n-s}{2}}{N_c-r}}{\binom{\binom{n}{2}}{N_c}} \right) \tag{19}$$

because if the $n-s$ points forming the greatest connected component of a graph belonging to the class $\bar{A}E_{\log\log n}$ are fixed, then if r is the number of edges connecting some of the s points outside this connected component we must have $r \geqq 1$, and the r edges in question can be chosen in $\binom{\binom{s}{2}}{r}$ ways,

and the remaining $N_e - r$ edges which connect points of the connected component having $n-s$ points can be chosen in less than $\binom{\binom{n-s}{2}}{N_e-r}$ ways. Thus we obtain

$$P(\bar{A}E_{\log\log n}, n, N_e) \leqq \frac{\log n}{n} \sum_{s=2}^{\log\log n} \frac{2^{\binom{s}{2}} e^{-2se}}{s!} \leqq \frac{e^{e^{-2e}} \cdot e^{\frac{1}{2}(\log\log n)^2}}{n} \tag{20}$$

which proves (18). Thus our lemma is proved.

Now we turn to the proof of our theorems.

PROOF OF THEOREM 1. Let $\mathcal{C}_0(n, N_e)$ denote the number of completely connected graphs G_{n,N_e}. Then $\mathcal{C}_0(n, N_e)$ is equal to the number of graphs G_{n,N_e} of type A having no isolated points. Now let $\mathcal{C}_0'(n, N_e)$ denote the number of *all* graphs G_{n,N_e} (including graphs of type $\bar{A}$) having no isolated point. Then clearly

$$\mathcal{C}_0'(n, N_e) = \sum_{k=0}^{n} (-1)^k \binom{n}{k} \binom{\binom{n-k}{2}}{N_e}. \tag{21}$$

As the left hand side of (21) is contained between any two consecutive partial sums of the sum on the right of (21), and for any fixed value of k we have

$$\lim_{n\to+\infty} \frac{\binom{n}{k}\binom{\binom{n-k}{2}}{N_e}}{\binom{\binom{n}{2}}{N_e}} = \frac{e^{-2ke}}{k!}. \tag{22}$$

Thus we obtain

$$\lim_{n\to\infty} \frac{\mathcal{C}_0'(n, N_e)}{\binom{\binom{n}{2}}{N_e}} = \sum_{k=0}^{\infty} \frac{(-1)^k e^{-2ke}}{k!} = e^{-e^{-2e}}. \tag{23}$$

But clearly

$$0 \leqq \frac{\mathcal{C}_0'(N, N_e) - \mathcal{C}_0(n, N_e)}{\binom{\binom{n}{2}}{N_e}} \leqq P(\bar{A}, n, N_e). \tag{24}$$

Thus, applying our Lemma, Theorem 1 follows.

PROOF OF THEOREM 2. According to our Lemma we have to consider only graphs of type A. Now the number of graphs G_{n,N_c} of the type A having a connected component consisting of $n-k$ points is clearly equal to $\binom{n}{k}$ multiplied by the number of connected graphs G_{n-k,N_c}. Thus it follows that

$$P_k(n,N_c)\sim\binom{n}{k}\frac{\binom{\binom{n-k}{2}}{N_c}}{\binom{\binom{n}{2}}{N_c}}P_0(n-k,N_c). \tag{25}$$

Taking into account that for a fixed value of k

$$\lim_{n\to\infty}\frac{N_c-\frac{1}{2}(n-k)\log(n-k)}{n-k}=c$$

with respect to (22) and (2) we obtain the assertion of Theorem 2.[1])

PROOF OF THEOREM 3. As we may restrict ourselves to graphs of type A, Theorem 3 follows from Theorem 2 immediately.

PROOF OF THEOREM 4. Clearly if $\nu_n=\left[\frac{1}{2}n\log n\right]+l=N+1$ then before choosing the last edge, we had a disconnected graph $G_{n,N}$ which can be made completely connected by adding one edge. With respect to our Lemma we may suppose that $G_{n,N}$ consists of a connected graph having $n-1$ vertices and an isolated point. As the last edge can be chosen in $n-1$ ways among the remaining $\binom{n}{2}-N$ edges it follows from Theorem 2

$$P\left(\nu_n=\left[\frac{1}{2}n\log n\right]+l\right)\sim\frac{2}{n}P_1(n,N)\sim\frac{2}{n}e^{-\frac{2l}{n}-e^{-\frac{2l}{n}}} \tag{26}$$

provided that $\frac{|l|}{n}$ is bounded.

---

[1]) Compare Theorem 2 with the following known result (see [7], Exercise No. 7 of Chapter IV, p. 134): If N balls are distributed at random in n boxes and $R_k(n,N)$ denotes the probability that exactly k boxes remain empty, then we have for any fixed real x and for $k=0,1,\ldots$

$$\lim_{n\to\infty}R_k(n,n\log n+xn)=\frac{(e^{-x})^k e^{-e^{-x}}}{k!}.$$

Thus (5) is proved. To prove (6) we have only to sum the probabilities (5) for $l < nx$ and obtain

$$(27) \qquad P\left(\frac{\nu_n - \frac{1}{2} n \log n}{n} < x\right) \sim \sum_{l<nx} \frac{2}{n} e^{-\frac{2l}{n} - e^{-\frac{2l}{n}}} \sim \int_{-\infty}^{2x} e^{-t-e^{-t}} dt.$$

As however

$$(28) \qquad \int_{-\infty}^{2x} e^{-t-e^{-t}} dt = \int_{e^{-2x}}^{\infty} e^{-u} du = e^{-e^{-2x}}$$

Theorem 4 is proved.

The following more general questions can be asked: Consider the random graph $\Gamma_{n, N(n)}$ with n possible vertices and $N(n)$ edges. What is the distribution of the number of vertices of the greatest connected component of $\Gamma_{n, N(n)}$ and the distribution of the number of its components? What is the typical structure of $\Gamma_{n, N(n)}$ (in the sense in which, according to our Lemma, the typical structure of Γ_{n, N_c} is that it belongs to type A)? We have solved these problems in the present paper only in the case $N(n) = \frac{1}{2} n \log n + cn$. We schall return to the general case in an other paper [8].

Bibliography.

[1] A. Cayley, A theorem on trees, *Quart. J. Pure Appl. Math.* **23** (1889), 376—378. (See also *The Collected Papers of A. Cayley*, Cambridge 1897, Vol. 13, pp. 26—28.)

[2] O. Dziobek, Eine Formel der Substitionstheorie, *S.—B. Berlin. Math. Ges.* **17** (1917) 64—67.

[3] A. Prüfer, Neuer Beweis eines Satzes über Permutationen, *Arch. Math. Phys.* **27** (1918), 142—144.

[4] G. Pólya, Kombinatorische Anzahlbestimmungen für Gruppen, Graphen und chemische Verbindungen, *Acta Math.* **68** (1937), 145—255.

[5] R. I. Riddell, Jr. and G. E. Uhlenbeck, On the theory of the virial development of the equation of state of monoatomic gases. *J. Chem. Phys.* **21** (1953), 2056—2064.

[6] E. N. Gilbert, Enumeration of labelled graphs, *Canad. J. Math.* **8** (1956), 405—411.

[7] A. Rényi, Valószínűségszámítás, (Textbook of probability theory) Budapest, 1954 (Hungarian).

[8] P. Erdős and A. Rényi, On the evolution of random graphs, *Publications of the Mathematical Institute of the Hungarian Academy of Sciences* **5** (1960) (in print).

(Received November 19, 1958.)