ON THE EVOLUTION OF RANDOM GRAPHS

by

P. ERDŐS and A. RÉNYI

Dedicated to Professor P. Turán at his 50th birthday.

Introduction

Our aim is to study the probable structure of a random graph $\Gamma_{n,N}$ which has n given labelled vertices $P_1, P_2, \ldots, P_n$ and N edges; we suppose that these N edges are chosen at random among the $\binom{n}{2}$ possible edges, so that all $\binom{\binom{n}{2}}{N} = C_{n,N}$ possible choices are supposed to be equiprobable. Thus if $G_{n,N}$ denotes any one of the $C_{n,N}$ graphs formed from n given labelled points and having N edges, the probability that the random graph $\Gamma_{n,N}$ is identical with $G_{n,N}$ is $\frac{1}{C_{n,N}}$. If A is a property which a graph may or may not possess, we denote by $\mathbf{P}_{n,N}(A)$ the probability that the random graph $\Gamma_{n,N}$ possesses the property A, i. e. we put $\mathbf{P}_{n,N}(A) = \frac{A_{n,N}}{C_{n,N}}$ where $A_{n,N}$ denotes the number of those $G_{n,N}$ which have the property A.

An other equivalent formulation is the following: Let us suppose that n labelled vertices $P_1, P_2, \ldots, P_n$ are given. Let us choose at random an edge among the $\binom{n}{2}$ possible edges, so that all these edges are equiprobable. After this let us choose an other edge among the remaining $\binom{n}{2} - 1$ edges, and continue this process so that if already k edges are fixed, any of the remaining $\binom{n}{2} - k$ edges have equal probabilities to be chosen as the next one. We shall study the "evolution" of such a random graph if N is increased. In this investigation we endeavour to find what is the "typical" structure at a given stage of evolution (i. e. if N is equal, or asymptotically equal, to a given function $N(n)$ of n). By a "typical" structure we mean such a structure the probability of which tends to 1 if $n \to +\infty$ when $N = N(n)$. If A is such a property that $\lim_{n\to+\infty} \mathbf{P}_{n,N(n)}(A) = 1$, we shall say that „almost all" graphs $G_{n,N(n)}$ possess this property.

The study of the evolution of graphs leads to rather surprising results. For a number of fundamental structural properties A there exists a function $A(n)$ tending monotonically to $+\infty$ for $n \to +\infty$ such that

$$\lim_{n\to+\infty} \mathbf{P}_{n,N(n)}(A) = \begin{cases} 0 & \text{if} \quad \lim\limits_{n\to+\infty} \dfrac{N(n)}{A(n)} = 0 \\ 1 & \text{if} \quad \lim\limits_{n\to+\infty} \dfrac{N(n)}{A(n)} = +\infty . \end{cases} \tag{1}$$

If such a function $A(n)$ exists we shall call it a *"threshold function"* of the property A.

In many cases besides (1) it is also true that there exists a probability distribution function $F(x)$ so that if $0 < x < +\infty$ and x is a point of continuity of $F(x)$ then

$$\lim_{n\to+\infty} \mathbf{P}_{n,N(n)}(A) = F(x) \quad \text{if} \quad \lim_{n\to+\infty} \frac{N(n)}{A(n)} = x . \tag{2}$$

If (2) holds we shall say that $A(n)$ is a *„regular threshold function"* for the property A and call the function $F(x)$ the *threshold distribution function* of the property A.

For certain properties A there exist two functions $A_1(n)$ and $A_2(n)$ both tending monotonically to $+\infty$ for $n\to+\infty$, and satisfying $\lim\limits_{n\to+\infty} \dfrac{A_2(n)}{A_1(n)} = 0$, such that

$$\lim_{n\to+\infty} \mathbf{P}_{n,N(n)}(A) = \begin{cases} 0 & \text{if} \quad \lim\limits_{n\to+\infty} \dfrac{N(n)-A_1(n)}{A_2(n)} = -\infty \\ 1 & \text{if} \quad \lim\limits_{n\to+\infty} \dfrac{N(n)-A_1(n)}{A_2(n)} = +\infty. \end{cases} \tag{3}$$

Clearly (3) implies that

$$\lim_{n\to+\infty} \mathbf{P}_{n,N(n)}(A) = \begin{cases} 0 & \text{if} \quad \limsup\limits_{n\to+\infty} \dfrac{N(n)}{A_1(n)} < 1 \\ 1 & \text{if} \quad \liminf\limits_{n\to+\infty} \dfrac{N(n)}{A_1(n)} > 1 . \end{cases} \tag{4}$$

If (3) holds we call the pair $\big(A_1(n), A_2(n)\big)$ a pair of *"sharp threshold"*-functions of the property A. It follows from (4) that if $\big(A_1(n), A_2(n)\big)$ is a pair of sharp threshold functions for the property A then $A_1(n)$ is an (ordinary) threshold function for the property A and the threshold distribution function figuring in (2) is the degenerated distribution function

$$F_1(x) = \begin{cases} 0 & \text{for} \quad x \leq 1 \\ 1 & \text{for} \quad x > 1 \end{cases}$$

and convergence in (2) takes place for every $x \neq 1$. In some cases besides (3) it is also true that there exists a probability distribution function $G(y)$ defined for $-\infty < y < +\infty$ such that if y is a point of continuity of $G(y)$ then

$$\text{(5)} \qquad \lim_{n \to +\infty} \mathbf{P}_{n,N(n)}(A) = G(y) \quad \text{if} \quad \lim_{n \to +\infty} \frac{N(n) - A_1(n)}{A_2(n)} = y .$$

If (5) holds we shall say that we have a *regular sharp threshold* and shall call $G(y)$ the *sharp-threshold distribution function of the property* A.

One of our chief aims will be to determine the threshold respectively sharp threshold functions, and the corresponding distribution functions for the most obvious structural properties, e. g. the presence in $\Gamma_{n,N}$ of subgraphs of a given type (trees, cycles of given order, complete subgraphs etc.) further for certain global properties of the graph (connectedness, total number of connected components, etc.).

In a previous paper [7] we have considered a special problem of this type; we have shown that denoting by C the property that the graph is connected, the pair $C_1(n) = \frac{1}{2} n \log n$, $C_2(n) = n$ is a pair of strong threshold functions for the property C, and the corresponding sharp-threshold distribution function is $e^{-e^{-2y}}$; thus we have proved[1] that putting $N(n) = \frac{1}{2} n \log n + yn + o(n)$ we have

$$\text{(6)} \qquad \lim_{n \to +\infty} \mathbf{P}_{n,N(n)}(C) = e^{-e^{-2y}} \qquad (-\infty < y < +\infty) .$$

In the present paper we consider the evolution of a random graph in a more systematic manner and try to describe the gradual development and step-by-step unravelling of the complex structure of the graph $\Gamma_{n,N}$ when N increases while n is a given large number.

We succeeded in revealing the emergence of certain structural properties of $\Gamma_{n,N}$. However a great deal remains to be done in this field. We shall call in § 10. the attention of the reader to certain unsolved problems. It seems to us further that it would be worth while to consider besides graphs also more complex structures from the same point of view, i. e. to investigate the laws governing their evolution in a similar spirit. This may be interesting not only from a purely mathematical point of view. In fact, the evolution of graphs may be considered as a rather simplified model of the evolution of certain communication nets (railway, road or electric network systems, etc.) of a country or some other unit. (Of course, if one aims at describing such a real situation, one should replace the hypothesis of equiprobability of all connections by some more realistic hypothesis.) It seems plausible that by considering the random growth of more complicated structures (e. g. structures consisting of different sorts of "points" and connections of different types) one could obtain fairly reasonable models of more complex real growth processes (e. g.

[1] Partial result on this problem has been obtained already in 1939 by P. ERDŐS and H. WHITNEY but their results have not been published.

the growth of a complex communication net consisting of different types of connections, and even of organic structures of living matter, etc.).

§§ 1—3. contain the discussion of the presence of certain components in a random graph, while §§ 4—9. investigate certain global properties of a random graph. Most of our investigations deal with the case when $N(n) \sim cn$ with $c > 0$. In fact our results give a clear picture of the evolution of $\Gamma_{n,N(n)}$ when $c = \frac{N(n)}{n}$ (which plays in a certain sense the role of time) increases. In § 10. we make some further remarks and mention some unsolved problems.

Our investigation belongs to the combinatorical theory of graphs, which has a fairly large literature. The first who enumerated the number of possible graphs with a given structure was A. CAYLEY [1]. Next the important paper [2] of G. PÓLYA has to be mentioned, the starting point of which were some chemical problems. Among more recent results we mention the papers of G. E. UHLENBECK and G. W. FORD [5] and E. N. GILBERT [6]. A fairly complete bibliography will be given in a paper of F. HARARY [8]. In these papers the probabilistic point of view was not explicitly emphasized. This has been done in the paper [9] of one of the authors, but the aim of the probabilistic treatment was there different: the existence of certain types of graphs has been shown by proving that their probability is positive. Random trees have been considered in [14].

In a recent paper [10] T. L. AUSTIN, R. E. FAGEN, W. F. PENNEY and J. RIORDAN deal with random graphs from a point of view similar to ours. The difference between the definition of a random graph in [10] and in the present paper consists in that in [10] it is admitted that two points should be connected by more than one edge ("parallel" edges). Thus in [10] it is supposed that after a certain number of edges have already been selected, the next edge to be selected may be any of the possible $\binom{n}{2}$ edges between the n given points (including the edges already selected). Let us denote such a random graph by $\Gamma^*_{n,N}$. The difference between the probable properties of $\Gamma_{n,N}$ resp. $\Gamma^*_{n,N}$ are in most (but not in all) cases negligible. The corresponding probabilities are in general (if the number N of edges is not too large) asymptotically equal. There is a third possible point of view which is in most cases almost equivalent with these two; we may suppose that for each pair of n given points it is determined by a chance process whether the edge connecting the two points should be selected or not, the probability for selecting any given edge being equal to the same number $p > 0$, and the decisions concerning the different edges being completely independent. In this case of course the number of edges is a random variable, having the expectation $\binom{n}{2}p$; thus if we want to obtain by this method a random graph having in the mean N edges we have to choose the value of p equal to $\frac{N}{\binom{n}{2}}$. We shall denote such a random graph by $\Gamma^{**}_{n,N}$. In many (though not all) of the problems treated in the present paper it does not cause any essential difference if we consider instead of $\Gamma_{n,N}$ the random graph $\Gamma^{**}_{n,N}$.

Comparing the method of the present paper with that of [10] it should be pointed out that our aim is to obtain threshold functions resp. distributions, and thus we are interested in asymptotic formulae for the probabilities considered. Exact formulae are of interest to us only so far as they help in determining the asymptotic behaviour of the probabilities considered (which is rarely the case in this field, as the exact formulae are in most cases too complicated). On the other hand in [10] the emphasis is on exact formulae resp. on generating functions. The only exception is the average number of connected components, for the asymptotic evaluation of which a way is indicated in § 5. of [10]; this question is however more fully discussed in the present paper and our results go beyond that of [10]. Moreover, we consider not only the number but also the character of the components. Thus for instance we point out the remarkable change occuring at $N \sim \frac{n}{2}$. If $N \sim nc$ with $c < 1/2$ then with probability tending to 1 for $n \to +\infty$ all points except a bounded number of points of $\Gamma_{n,N}$ belong to components which are trees, while for $N \sim nc$ with $c > \frac{1}{2}$ this is no longer the case. Further for a fixed value of n the average number of components of $\Gamma_{n,N}$ decreases asymptotically in a linear manner with N, when $N \leqq \frac{n}{2}$, while for $N > \frac{n}{2}$ the formula giving the average number of components is not linear in N.

In what follows we shall make use of the sysmbols O and o. As usually $a(n) = o\big(b(n)\big)$ (where $b(n) > 0$ for $n = 1, 2, \ldots$) means that $\lim\limits_{n \to +\infty} \frac{|a(n)|}{b(n)} = 0$, while $a(n) = O\big(b(n)\big)$ means that $\frac{|a(n)|}{b(n)}$ is bounded. The parameters on which the bound of $\frac{|a(n)|}{b(n)}$ may depend will be indicated if it is necessary; sometimes we will indicate it by an index. Thus $a(n) = O_\varepsilon\big(b(n)\big)$ means that $\frac{|a(n)|}{b(n)} \leqq K(\varepsilon)$ where $K(\varepsilon)$ is a positive constant depending on ε. We write $a(n) \sim b(n)$ to denote that $\lim\limits_{n \to +\infty} \frac{a(n)}{b(n)} = 1$.

We shall use the following definitions from the theory of graphs. (For the general theory see [3] and [4].)

A finite non-empty set V of labelled points $P_1, P_2, \ldots, P_n$ and a set E of different unordered pairs (P_i, P_j) with $P_i \in V$, $P_j \in V$, $i \neq j$ is called a *graph*; we denote it sometimes by $G = \{V, E\}$; the number n is called the *order* (or *size*) of the graph; the points $P_1, P_2, \ldots, P_n$ are called the *vertices* and the pairs (P_i, P_j) the *edges* of the graph. Thus we consider *non-oriented finite graphs without parallel edges and without slings.* The set E may be empty, thus a collection of points (especially a single point) is also a graph.

A graph $G_2 = \{V_2, E_2\}$ is called a *subgraph* of a graph $G_1 = \{V_1, E_1\}$ if the set of vertices V_2 of G_2 is a subset of the set of vertices V_1 of G_1 and the set E_2 of edges of G_2 is a subset of the set E_1 of edges of G_1.

A sequence of k edges of a graph such that every two consecutive edges and only these have a vertex in common is called a *path* of order k.

A cyclic sequence of k edges of a graph such that every two consecutive edges and only these have a common vertex is called a *cycle* of order k.

A graph G is called *connected* if any two of its points belong to a path which is a subgraph of G.

A graph is called a *tree* of order (or size) k if it has k vertices, is connected and if none of its subgraphs is a cycle. A tree of order k has evidently $k - 1$ edges.

A graph is called a *complete graph* of order $\binom{k}{2}$ if it has k vertices and $\binom{k}{2}$ edges. Thus in a complete graph of order k any two points are connected by an edge.

A subgraph G' of a graph G will be called an *isolated subgraph* if all edges of G one or both endpoints of which belong to G', belong to G'. A connected isolated subgraph G' of a graph G is called a *component* of G. The number of points belonging to a component G' of a graph G will be called the *size* of G'.

Two graphs shall be called *isomorphic*, if there exists a one-to-one mapping of the vertices carrying over these graphs into another.

The graph $\overline{G}$ shall be called *complementary graph* of G if $\overline{G}$ consists of the same vertices $P_1, P_2, \ldots, P_n$ as G and of those and only those edges (P_i, P_j) which do not occur in G.

The number of edges starting from the point P of a graph G will be called the *degree* of P in G.

A graph G is called a *saturated even graph of type* (a, b) if it consists of $a + b$ points and its points can be split in two subsets V_1 and V_2 consisting of a resp. b points, such that G contains any edge (P, Q) with $P \in V_1$ and $Q \in V_2$ and no other edge.

A graph is called *planar*, if it can be drawn on the plane so that no two of its edges intersect.

We introduce further the following definitions: If a graph G has n vertices and N edges, we call the number $\frac{2N}{n}$ the "*degree*" of the graph. (As a matter of fact $\frac{2N}{n}$ is the average degree of the vertices of G.) If a graph G has the property that G has no subgraph having a larger degree than G itself, we call G a *balanced* graph.

We denote by $\mathbf{P}(\ldots)$ the probability of the event in the brackets, by $\mathbf{M}(\xi)$ resp. $\mathbf{D}^2(\xi)$ the mean value resp. variance of the random variable ξ. In cases when it is not clear from the context in which probability space the probabilities or respectively the mean values and variances are to be understood, this will be explicitly indicated. Especially $\mathbf{M}_{n,N}$ resp. $\mathbf{D}^2_{n,N}$ will denote the mean value resp. variance calculated with respect to the probabilities $\mathbf{P}_{n,N}$.

We shall often use the following elementary asymptotic formula:

$$(7) \qquad \binom{n}{k} \sim \frac{n^k e^{-\frac{k^2}{2n} - \frac{k^3}{6n^2}}}{k\,!} \text{ valid for } k = o(n^{3/4})\,.$$

Our thanks are due to T. GALLAI for his valuable remarks.

§ 1. Thresholds for subgraphs of given type

If N is very small compared with n, namely if $N = o(\sqrt{n})$ then it is very probable that $\Gamma_{n,N}$ is a collection of isolated points and isolated edges, i. e. that no two edges of $\Gamma_{n,N}$ have a point in common. As a matter of fact the probability that at least two edges of $\Gamma_{n,N}$ shall have a point in common is by (7) clearly

$$1 - \frac{\binom{n}{2N}(2N)\,!}{2^N N!\binom{\binom{n}{2}}{N}} = O\left(\frac{N^2}{n}\right).$$

If however $N \sim c\sqrt{n}$ where $c > 0$ is a constant not depending on n, then the appearance of trees of order 3 will have a probability which tends to a positive limit for $n \to +\infty$, but the appearance of a connected component consisting of more than 3 points will be still very improbable. If N is increased while n is fixed, the situation will change only if N reaches the order of magnitude of $n^{2/3}$. Then trees of order 4 (but not of higher order) will appear with a probability not tending to 0. In general, the threshold function for the presence of trees of order k is $n^{\frac{k-2}{k-1}}$ $(k = 3, 4, \ldots)$. This result is contained in the following

Theorem 1. *Let $k \geqq 2$ and l $\left(k - 1 \leqq l \leqq \binom{k}{2}\right)$ be positive integers. Let $\mathcal{B}_{k,l}$ denote an arbitrary not empty class of connected balanced graphs consisting of k points and l edges. The threshold function for the property that the random graph considered should contain at least one subgraph isomorphic with some element of $\mathcal{B}_{k,l}$ is $n^{2-\frac{k}{l}}$.*

The following special cases are worth mentioning

Corollary 1. *The threshold function for the property that the random graph contains a subgraph which is a tree of order k is $n^{\frac{k-2}{k-1}}$ $(k = 3, 4, \ldots)$.*

Corollary 2. *The threshold function for the property that a graph contains a connected subgraph consisting of $k \geqq 3$ points and k edges (i. e. containing exactly one cycle) is n, for each value of k.*

Corollary 3. *The threshold function for the property that a graph contains a cycle of order k is n, for each value of $k \geqq 3$.*

Corollary 4. *The threshold function for the property that a graph contains a complete subgraph of order* $k \geqq 3$ *is* $n^{2\left(1-\frac{1}{k-1}\right)}$.

Corollary 5. *The threshold function for the property that a graph contains a saturated even subgraph of type* (a, b) (*i. e. a subgraph* consisting of $a+b$ points $P_1, \ldots, P_a, Q_1, \ldots Q_b$ and of the ab edges (P_i, Q_j) *is* $n^{2-\frac{a+b}{ab}}$.

To deduce these Corollaries one has only to verify that all 5 types of graphs figuring in Corollaries 1—5. are balanced, which is easily seen.

Proof of Theorem 1. Let $B_{k,l} \geqq 1$ denote the number of graphs belonging to the class $\mathscr{B}_{k,l}$ which can be formed from k given labelled points. Clearly if $P_{n,N}(\mathscr{B}_{k,l})$ denotes the probability that the random graph $\Gamma_{n,N}$ contains at least one subgraph isomorphic with some element of the class $\mathscr{B}_{k,l}$, then

$$\mathbf{P}_{n,N}(\mathscr{B}_{k,l}) \leqq \binom{n}{k} B_{k,l} \frac{\binom{\binom{n}{2}-l}{N-l}}{\binom{\binom{n}{2}}{N}} = O\left(\frac{N^l}{n^{2l-k}}\right). \tag{1.1}$$

As a matter of fact if we select k points (which can be done in $\binom{n}{2}$ different ways) and form from them a graph isomorphic with some element of the class $\mathscr{B}_{k,l}$ (which can be done in $B_{k,l}$ different ways) then the number of graphs $G_{n,N}$ which contain the selected graph as a subgraph is equal to the number of ways the remaining $N-l$ edges can be selected from the $\binom{n}{2}-l$ other possible edges. (Of course those graphs, which contain more subgraphs isomorphic with some element of $\mathscr{B}_{k,l}$ are counted more than once.)

Now clearly if $N = o(n^{2-\frac{k}{l}})$ then by

$$\mathbf{P}_{n,N}(\mathscr{B}_{k,l}) = o(1)$$

which proves the first part of the assertion of Theorem 1. To prove the second part of the theorem let $\mathscr{B}_{k,l}^{(n)}$ denote the set of all subgraphs of the complete graph consisting of n points, isomorphic with some element of $\mathscr{B}_{k,l}$. To any $S \in \mathscr{B}_{k,l}^{(n)}$ let us associate a random variable $\varepsilon(S)$ such that $\varepsilon(S) = 1$ or $\varepsilon(S) = 0$ according to whether S is a subgraph of $\Gamma_{n,N}$ or not. Then clearly (we write in what follows for the sake of brevity $\mathbf{M}$ instead of $\mathbf{M}_{n,N}$)

$$\mathbf{M}\left(\sum_{S \in \mathscr{B}_{k,l}^{(n)}} \varepsilon(S)\right) = \sum_{S \in \mathscr{B}_{k,l}^{(n)}} \mathbf{M}(\varepsilon(S)) = \binom{n}{k} B_{k,l} \frac{\binom{\binom{n}{2}-l}{N-l}}{\binom{\binom{n}{2}}{N}} \sim \frac{B_{k,l}}{k!} \frac{(2N)^l}{n^{2l-k}}. \tag{1.2}$$

On the other hand if S_1 and S_2 are two elements of $\mathcal{B}_{k,l}^{(n)}$ and if S_1 and S_2 do not contain a common edge then

$$\mathbf{M}(\varepsilon(S_1)\,\varepsilon(S_2)) = \frac{\binom{\binom{n}{2} - 2l}{N - 2l}}{\binom{\binom{n}{2}}{N}}.$$

If S_1 and S_2 contain exactly s common points and r common edges $(1 \leq r \leq l-1)$ we have

$$\mathbf{M}(\varepsilon(S_1)\,\varepsilon(S_2)) = \frac{\binom{\binom{n}{2} - 2l + r}{N - 2l + r}}{\binom{\binom{n}{2}}{N}} = O\left(\frac{N^{2l-r}}{n^{4l-2r}}\right).$$

On the other hand the intersection of S_1 and S_2 being a subgraph of S_1 (and S_2) by our supposition that each S is balanced, we obtain $\frac{r}{s} \leq \frac{l}{k}$ i. e. $s \geq \frac{rk}{l}$ and thus the number of such pairs of subgraphs S_1 and S_2 does not exceed

$$B_{k,l}^2 \sum_{j \geq \frac{rk}{l}}^{k} \binom{n}{k}\binom{k}{j}\binom{n-k}{k-j} = O\left(n^{2k - \frac{rk}{l}}\right).$$

Thus we obtain

$$\mathbf{M}\left(\Big(\sum_{S \in \mathcal{B}_{k,l}^{(n)}} \varepsilon(S)\Big)^2\right) =$$

(1.3)

$$= \sum_{S \in \mathcal{B}_{k,l}^{(n)}} \mathbf{M}(\varepsilon(S)) + \frac{n!\,B_{k,l}^2}{k!^2(n-2k)!} \frac{\binom{\binom{n}{2} - 2l}{N - 2l}}{\binom{\binom{n}{2}}{N}} + O\left(\left(\frac{N^l}{n^{2l-k}}\right)^2 \sum_{r=1}^{l} \left(\frac{n^{2-\frac{k}{l}}}{N}\right)^r\right).$$

Now clearly

$$\frac{n!}{k!^2(n-2k)!} \frac{\binom{\binom{n}{2} - 2l}{N - 2l}}{\binom{\binom{n}{2}}{N}} \leq \binom{n}{k}^2 \frac{\binom{\binom{n}{2} - l}{N - l}^2}{\binom{\binom{n}{2}}{N}^2}.$$

If we suppose that

$$\frac{N}{n^{2-\frac{k}{l}}} = \omega \to +\infty ,$$

it follows that we have

$$\text{(1.4)} \qquad \mathbf{D}^2\Big(\sum_{S\in\mathcal{B}_{k,l}^{(n)}} \varepsilon(S)\Big) = O\left(\frac{\Big(\sum_{S\in\mathcal{B}_{k,l}^{(n)}} \mathbf{M}(\varepsilon(S))\Big)^2}{\omega}\right).$$

It follows by the inequality of *Chebysheff* that

$$\mathbf{P}_{n,N}\left(\Big|\sum_{S\in\mathcal{B}_{k,l}^{(n)}} \varepsilon(S) - \sum_{S\in\mathcal{B}_{k,l}^{(n)}} \mathbf{M}(\varepsilon(S))\Big| > \frac{1}{2}\sum_{S\in\mathcal{B}_{k,l}^{(n)}} \mathbf{M}(\varepsilon(S))\right) = O\left(\frac{1}{\omega}\right)$$

and thus

$$\text{(1.5)} \qquad \mathbf{P}_{n,N}\left(\sum_{S\in\mathcal{B}_{k,l}^{(n)}} \varepsilon(S) \leq \frac{1}{2}\sum_{S\in\mathcal{B}_{k,l}^{(n)}} \mathbf{M}(\varepsilon(S))\right) = O\left(\frac{1}{\omega}\right).$$

As clearly by (1.2) if $\omega \to +\infty$ then $\sum_{S\in\mathcal{B}_{k,l}^{(n)}} \mathbf{M}(\varepsilon(S)) \to +\infty$ it follows not only that the probability that $\Gamma_{n,N}$ contains at least one subgraph isomorphic with an element of $\mathcal{B}_{k,l}$ tends to 1, but also that with probability tending to 1 the number of subgraphs of $\Gamma_{n,N}$ isomorphic to some element of $\mathcal{B}_{k,l}$ will tend to $+\infty$ with the same order of magnitude as ω^l.

Thus Theorem 1 is proved.

It is interesting to compare the thresholds for the appearance of a subgraph of a certain type in the above sense with probability near to 1, with the number of edges which is needed in order that the graph should have *necessarily* a subgraph of the given type. Such "compulsory" thresholds have been considered by P. TURÁN [11] (see also [12]) and later by P. ERDŐS and A. H. STONE [17]). For instance for a tree of order k clearly the compulsory threshold is $\left[\frac{n(k-2)}{2}\right] + 1$; for the presence of at least one cycle the compulsory threshold is n while according to a theorem of P. TURÁN [11] for complete subgraphs of order k the compulsory threshold is $\frac{(k-2)}{2(k-1)}(n^2 - r^2) + \binom{r}{2}$ where $r = n - (k-1)\left[\frac{n}{k-1}\right]$. In the paper [13] of T. KŐVÁRI, V. T. SÓS and P. TURÁN it has been shown that the compulsory threshold for the presence of a saturated even subgraph of type (a, a) is of order of magnitude not greater than $n^{2-\frac{1}{a}}$. In all cases the "compulsory" thresholds in TURÁN's sense are of greater order of magnitude as our "probable" thresholds.

§ 2. Trees

Now let us turn to the determination of threshold distribution functions for trees of a given order. We shall prove somewhat more, namely that if $N \sim \varrho\, n^{\frac{k-2}{k-1}}$ where $\varrho > 0$, then the number of trees of order k contained in $\Gamma_{n,N}$ has in the limit for $n \to +\infty$ a Poisson distribution with mean value $\lambda = \frac{(2\varrho)^{k-1} k^{k-2}}{k!}$. This implies that the threshold distribution function for trees of order k is $1 - e^{-\lambda}$.

In proving this we shall count only *isolated trees* of order k in $\Gamma_{n,N}$, i. e. trees of order k which are isolated subgraphs of $\Gamma_{n,N}$. According to Theorem 1. this makes no essential difference, because if there would be a tree of order k which is a subgraph but not an isolated subgraph of $\Gamma_{n,N}$, then $\Gamma_{n,N}$ would have a connected subgraph consisting of $k+1$ points and the probability of this is tending to 0 if $N = o\left(n^{\frac{k-1}{k}}\right)$ which condition is fulfilled in our case as we suppose $N \sim \varrho n^{\frac{k-2}{k-1}}$.

Thus we prove

Theorem 2a. *If* $\lim\limits_{n\to+\infty} \frac{N(n)}{n^{\frac{k-2}{k-1}}} = \varrho > 0$ *and* τ_k *denotes the number of isolated trees of order* k *in* $\Gamma_{n,N(n)}$ *then*

$$\lim_{n\to+\infty} \mathbf{P}_{n,N(n)}(\tau_k = j) = \frac{\lambda^j e^{-\lambda}}{j!} \tag{2.1}$$

or $j = 0, 1, \ldots,$ *where*

$$\lambda = \frac{(2\varrho)^{k-1} k^{k-2}}{k!}. \tag{2.2}$$

For the proof we need the following

Lemma 1. *Let* $\varepsilon_{n1}, \varepsilon_{n2}, \ldots, \varepsilon_{nl_n}$ *be sets of random variables on some probability space; suppose that* $\varepsilon_{ni}(1 \leq i \leq l_n)$ *takes on only the values* 1 *and* 0. *If*

$$\lim_{n\to+\infty} \sum_{1\leq i_1<i_2<\ldots<i_r\leq l_n} \mathbf{M}(\varepsilon_{ni_1}\varepsilon_{ni_2}\ldots\varepsilon_{ni_r}) = \frac{\lambda^r}{r!} \tag{2.3}$$

uniformly in r *for* $r = 1, 2, \ldots,$ *where* $\lambda > 0$ *and the summation is extended over all combinations* $(i_1, i_2, \ldots, i_r)$ *of order* r *of the integers* $1, 2, \ldots, l_n$, *then*

$$\lim_{n\to+\infty} \mathbf{P}\left(\sum_{i=1}^{l_n} \varepsilon_{n_i} = j\right) = \frac{\lambda^j e^{-\lambda}}{j!} \qquad (j = 0, 1, \ldots) \tag{2.4}$$

i. e. the distribution of the sum $\sum_{i=1}^{l_n} \varepsilon_{ni}$ *tends for* $n \to +\infty$ *to the Poisson-distribution with mean value* λ.

Proof of Lemma 1. Let us put

$$P_n(j) = \mathbf{P}\left(\sum_{i=1}^{l_n} \varepsilon_{n_i} = j\right). \tag{2.5}$$

Clearly

$$\sum_{1\leq i_1<i_2<\ldots<i_r\leq l_n} \mathbf{M}(\varepsilon_{ni_1}\varepsilon_{ni_2}\ldots\varepsilon_{ni_r}) = \sum_{j=r}^{+\infty}\binom{j}{r}P_n(j) \tag{2.6}$$

thus it follows from (2.3) that

$$\lim_{n\to+\infty}\sum_{j=r}^{+\infty}P_n(j)\binom{j}{r} = \frac{\lambda^r}{r!} \qquad (r=1,2,\ldots) \tag{2.7}$$

uniformly in r.

It follows that for any z with $|z|<1$

$$\lim_{n\to+\infty}\sum_{r=1}^{\infty}\left(\sum_{j=r}^{+\infty}P_n(j)\binom{j}{r}\right)z^r = \sum_{r=1}^{\infty}\frac{(\lambda z)^r}{r!} = e^{\lambda z}-1. \tag{2.8}$$

But

$$\sum_{r=1}^{\infty}\left(\sum_{j=r}^{+\infty}P_n(j)\binom{j}{r}\right)z^r = \sum_{j=0}^{+\infty}P_n(j)\,(1+z)^j-1. \tag{2.9}$$

Thus choosing $z=x-1$ with $0<x\leq 1$ it follows that

$$\lim_{n\to+\infty}\sum_{j=0}^{+\infty}P_n(j)\,x^j = e^{\lambda(x-1)} \qquad \text{for } 0<x\leq 1. \tag{2.10}$$

It follows easily that (2.10) holds for $x=0$ too. As a matter of fact putting $G_n(x)=\sum_{j=0}^{+\infty}P_n(j)\,x^j$, we have for $0<x\leq 1$

$$|P_n(0)-e^{-\lambda}| \leq |G_n(x)-e^{\lambda(x-1)}| + |G_n(x)-P_n(0)| + |e^{\lambda(x-1)}-e^{-\lambda}|.$$

As however

$$|G_n(x)-P_n(0)| \leq x\sum_{j=1}^{+\infty}P_n(j) \leq x$$

and similarly

$$|e^{\lambda(x-1)}-e^{-\lambda}| \leq x$$

it follows that

$$|P_n(0)-e^{-\lambda}| \leq |G_n(x)-e^{\lambda(x-1)}| + 2x.$$

Thus we have

$$\limsup_{n\to+\infty}|P_n(0)-e^{-\lambda}| \leq 2x;$$

as however $x>0$ may be chosen arbitrarily small it follows that

$$\lim_{n\to+\infty}P_n(0) = e^{-\lambda}$$

i. e. that (2.10) holds for $x = 0$ too. It follows by a well-known argument that

$$\lim_{n\to+\infty} P_n(j) = \frac{\lambda^j e^{-\lambda}}{j!} \qquad (j = 0, 1, \ldots). \tag{2.11}$$

As a matter of fact, as (2.10) is valid for $x = 0$, (2.11) holds for $j = 0$. If (2.11) is already proved for $j \leqq s - 1$ then it follows from (2.10) that

$$\lim_{n\to+\infty} \sum_{j=s}^{+\infty} P_n(j)\, x^{j-s} = \sum_{j=s}^{+\infty} \frac{\lambda^j e^{-\lambda}}{j!} x^{j-s} \qquad \text{for } 0 < x \leqq 1. \tag{2.12}$$

By the same argument as used in connection with (2.10) we obtain that (2.12) holds for $x = 0$ too. Substituting $x = 0$ into (2.12) we obtain that (2.11) holds for $j = s$ too. Thus (2.11) is proved by induction and the assertion of Lemma 1 follows.

Proof of Theorem 2a. Let $T_k^{(n)}$ denote the set of all trees of order k which are subgraphs of the complete graph having the vertices $P_1, P_2, \ldots, P_n$. If $S \in T_k^{(n)}$ let the random variable $\varepsilon(S)$ be equal to 1 if S is an *isolated* subgraph of $\Gamma_{n,N}$; otherwise $\varepsilon(S)$ shall be equal to 0. We shall show that the conditions of Lemma 1 are satisfied for the sum $\sum_{S\in T_k^{(n)}} \varepsilon(S)$ provided that $N = N(n) \sim \sim \varrho n^{\frac{k-2}{k-1}}$ and λ is defined by (2.2). As a matter of fact we have for any $S \in T_k^{(n)}$

$$\mathbf{M}(\varepsilon(S)) = \frac{\binom{\binom{n-k}{2}}{N-k+1}}{\binom{\binom{n}{2}}{N}} = \left(\frac{2N}{n^2}\right)^{k-1} e^{-\frac{2Nk}{n}} \left(1 + O\left(\frac{N}{n^2}\right)\right). \tag{2.13}$$

More generally if $S_1, S_2, \ldots, S_r$ ($S_j \in T_k^{(n)}$) have pairwise no point in common then clearly we have for each fixed $k \geqq 1$ and $r \geqq 1$ provided that $n \to +\infty$, $N \to +\infty$

$$\mathbf{M}(\varepsilon(S_1)\,\varepsilon(S_2)\ldots\varepsilon(S_r)) = \frac{\binom{\binom{n-rk}{2}}{N-r(k-1)}}{\binom{\binom{n}{2}}{N}} = \left(\frac{2N}{n^2}\right)^{(k-1)r} e^{-\frac{2Nrk}{n}} \left(1 + O\left(\frac{r^2N}{n^2}\right)\right) \tag{2.14}$$

where the bound of the O term depends only on k. If however the S_j ($j = = 1, 2, \ldots, r$) are not pairwise disjoint, we have

$$\mathbf{M}(\varepsilon(S_1)\,\varepsilon(S_2)\ldots\varepsilon(S_r)) = 0. \tag{2.15}$$

Taking into account that according to a classical formula of CAYLEY [1] the number of different trees which can be formed from k labelled points is equal to k^{k-2}, it follows that

$$\sum \mathbf{M}(\varepsilon(S_1)\,\varepsilon(S_2)\,\ldots\,\varepsilon(S_r)) = \left(\frac{k^{k-2}}{k!}\right)^r \frac{n^{kr}}{r!}\left(\frac{2N}{n^2}\right)^{rk-1} e^{-\frac{2Nrk}{n}}\left(1+O\left(\frac{r^2N}{n^2}\right)\right) \tag{2.16}$$

where the summation on the left hand side is extended over all r-tuples of trees belonging to the set $T_k^{(n)}$ and the bound of the O-term depends only on k. Note that (2.16) is valid independently of how N is tending to $+\infty$. This will be needed in the proof of Theorem 3.

Thus we have, uniformly in r

$$\lim_{\frac{N(n)}{n^{\frac{k-2}{k-1}}}\to\varrho} \sum \mathbf{M}(\varepsilon(S_1)\,\varepsilon(S_2)\,\ldots\,\varepsilon(S_r)) = \frac{\lambda^r}{r!} \qquad \text{for } r=1,2,\ldots \tag{2.17}$$

where λ is defined by (2.2).

Thus our Lemma 1 can be applied; as $\tau_k = \sum_{S\in T_k^{(n)}} \varepsilon(S)$ Theorem 2 is proved.

We add some remarks on the formula, resulting from (2.16) for $r=1$

$$\mathbf{M}(\tau_k) = \frac{n^2}{2N}\,\frac{\left(\frac{2N}{n}e^{-\frac{2N}{n}}\right)^k k^{k-2}}{k!}\left(1+O\left(\frac{N}{n^2}\right)\right). \tag{2.18}$$

Let us investigate the functions $m_k(t) = \frac{k^{k-2}\,t^{k-1}\,e^{-kt}}{k!}$ ($k=1,2,\ldots$). According to (2.18) $nm_k\left(\frac{2N}{n}\right)$ is asymptotically equal to the average number of trees of order k in $\Gamma_{n,N}$. For a fixed value of k, considered as a function of t, the value of $m_k(t)$ increases for $t<\frac{k-1}{k}$ and decreases for $t>\frac{k-1}{k}$; thus for a fixed value of n the average number of trees of order k reaches its maximum for $N\sim\frac{n}{2}\left(1-\frac{1}{k}\right)$; the value of this maximum is

$$M_k^* \sim n\,\frac{\left(1-\frac{1}{k}\right)^{k-1} e^{-(k-1)}\,k^{k-2}}{k!}.$$

For large values of k we have evidently

$$M_k^* \sim \frac{n}{\sqrt{2\pi}\,k^{5/2}}.$$

It is easy to see that for any $t > 0$ we have

$$m_k(t) \geqq m_{k+1}(t) \qquad (k = 1, 2, \ldots).$$

The functions $y = m_k(t)$ are shown on Fig. 1.

It is natural to ask what will happen with the number τ_k of isolated trees of order k contained in $\Gamma_{n,N}$ if $\frac{N(n)}{n^{\frac{k-2}{k-1}}} \to +\infty$. As the Poisson distribution $\left\{\frac{\lambda^j e^{-\lambda}}{j!}\right\}$ is approaching the normal distribution if $\lambda \to +\infty$, one can guess that τ_k will be approximately normally distributed. This is in fact true, and is expressed by

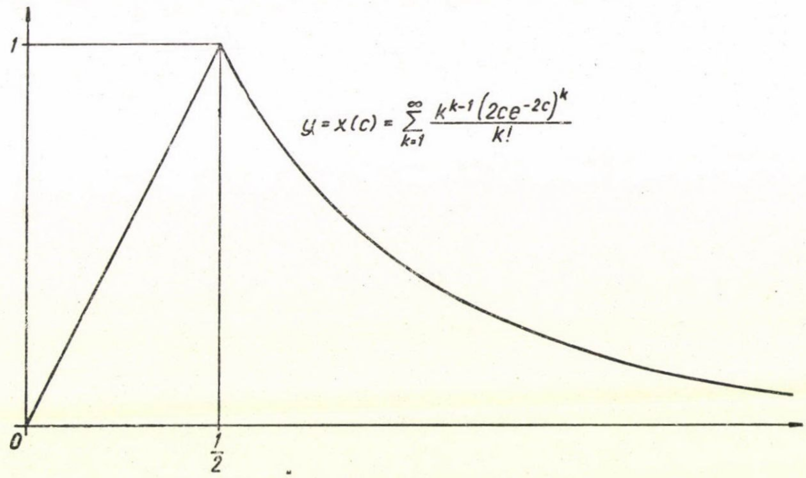

Figure 1a.

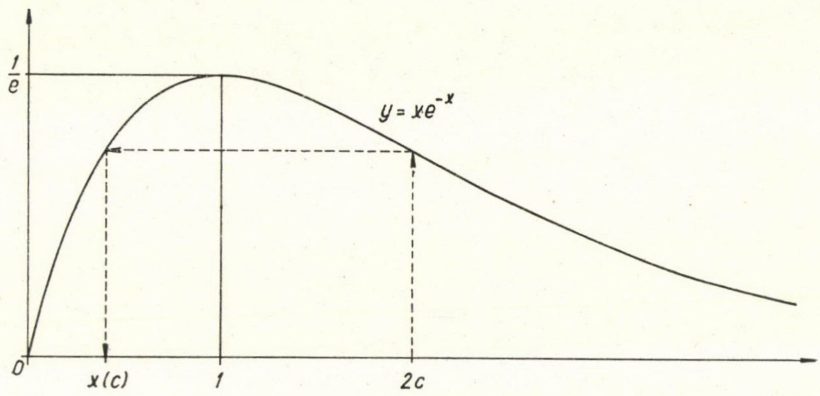

Figure 1b.

Theorem 2b. *If*

$$(2.19)\qquad \frac{N(n)}{n^{\frac{k-2}{k-1}}} \to +\infty$$

but at the same time

$$(2.20)\qquad \lim_{n\to+\infty} \frac{N(n) - \frac{1}{2k} n \log n - \frac{k-1}{2k} n \operatorname{loglog} n}{n} = -\infty ,$$

then denoting by τ_k the number of disjoint trees of order k contained as subgraphs in $\Gamma_{n,N(n)}$ ($k = 1, 2, \ldots$), we have for $-\infty < x < +\infty$

$$(2.21)\qquad \lim_{n\to+\infty} \mathbf{P}_{n,N(n)}\left(\frac{\tau_k - M_{n,N(n)}}{\sqrt{M_{n,N(n)}}} < x\right) = \Phi(x)$$

where

$$(2.22)\qquad M_{n,N} = n\frac{k^{k-2}}{k!}\left(\frac{2N}{n}\right)^{k-1} e^{-\frac{2kN}{n}}$$

and

$$(2.23)\qquad \Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{u^2}{2}}\,du .$$

Proof of Theorem 2b. Note first that the two conditions (2.19) and (2.20) are equivalent to the single condition $\lim_{n\to+\infty} M_{n,N(n)} = +\infty$, and as $\mathbf{M}(\tau_k) \sim M_{n,N}$ this means that the assertion of Theorem 2b can be expressed by saying that the number of isolated trees of order k is asymptotically normally distributed always if n and N tend to $+\infty$ so, that the average number of such trees is also tending to $+\infty$. Let us consider

$$\mathbf{M}(\tau_k^r) = \mathbf{M}\Big(\big(\sum_{S\in T_k^{(n)}} \varepsilon(S)\big)^r\Big) .$$

Now we have evidently, using (2.16)

$$\mathbf{M}(\tau_k^r) = \left(1 + O\left(\frac{r^2 N}{n^2}\right)\right)\sum_{j=1}^{r}\left(\sum_{\sum_{i=1}^{j} h_i = r,\, h_i\geq 1} \frac{r!}{h_1!\, h_2! \ldots h_j!}\right)\frac{M_{n,N}^j}{j!}$$

where $M_{n,N}$ is defined by (2.22). Now as well known (see [16], p. 176)

$$(2.24)\qquad \frac{1}{j!}\sum_{\sum_{i=1}^{j} h_i = r,\ h_i\geq 1} \frac{r!}{h_1!\, h_2! \ldots h_j!} = \sigma_r^{(j)}$$

where $\sigma_r^{(j)}$ are the Stirling numbers of the second kind (see e. g. [16], p. 168) defined by

$$x^r = \sum_{j=1}^{r} \sigma_r^{(j)} x(x-1)\ldots(x-j+1). \tag{2.25}$$

Thus we obtain

$$\mathbf{M}(\tau_k^r) = \left(1 + O\left(\frac{r^2 N}{n^2}\right)\right) \cdot \sum_{j=1}^{r} \sigma_r^{(j)} M_{n,N}^j . \tag{2.26}$$

Now as well known (see e. g. [16], p. 202)

$$e^{\lambda(e^x-1)} - 1 = \sum_{j=1}^{+\infty} \sum_{r=j}^{+\infty} \sigma_r^{(j)} \frac{x^r \lambda^j}{r!} = \sum_{r=1}^{\infty} \frac{x^r}{r!} \left(\sum_{j=1}^{r} \sigma_r^{(j)} \lambda^j \right). \tag{2.27}$$

Thus it follows that

$$\sum_{j=1}^{r} \sigma_r^{(j)} \lambda^j = \left[\frac{d^r}{dx^r} e^{\lambda(e^x-1)} \right]_{x=0} = \sum_{k=0}^{+\infty} \frac{\lambda^k}{k!} e^{-\lambda} k^r . \tag{2.28}$$

We obtain therefrom

$$\mathbf{M}\left(\left(\frac{\tau_k - M_{n,N}}{\sqrt{M_{n,N}}} \right)^r \right) = \left(\frac{1}{M_{n,N}^{r/2}} \sum_{k=0}^{+\infty} \frac{M_{n,N}^k}{k!} e^{-M_{n,N}} (k - M_{n,N})^r \right) \left(1 + O\left(\frac{r^2 N}{n^2} \right) \right). \tag{2.29}$$

Now evidently $\sum_{k=0}^{+\infty} \frac{\lambda^k}{k!} e^{-\lambda} (k-\lambda)^r$ is the r-th central moment of the Poisson distribution with mean value λ. It can be however easily verified that the moments of the Poisson distribution appropriately normalized tend to the corresponding moments of the normal distribution, i. e. we have for $r = 1, 2, \ldots$

$$\lim_{\lambda \to +\infty} \frac{1}{\lambda^{\frac{r}{2}}} \left(\sum_{k=1}^{+\infty} \frac{\lambda^k e^{-\lambda}}{k!} (k-\lambda)^r \right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^r e^{-\frac{x^2}{2}} dx . \tag{2.30}$$

In view of (2.29) this implies the assertion of Theorem 2b.

In the case $N(n) = \frac{1}{2k} n \log n + \frac{k-1}{2k} n \operatorname{loglog} n + yn + o(n)$ when the average number of isolated trees of order k in $\Gamma_{n,N(n)}$ is again finite, the following theorem is valid.

Theorem 2c. *Let τ_k denote the number of isolated trees of order k in $\Gamma_{n,N}$ $(k = 1, 2, \ldots)$. Then if*

$$N(n) = \frac{1}{2k} n \log n + \frac{k-1}{2k} n \operatorname{loglog} n + yn + o(n) \tag{2.31}$$

where $-\infty < y < +\infty$, we have

$$\lim_{n \to +\infty} \mathbf{P}_{n,N(n)}(\tau_k = j) = \frac{\lambda^j e^{-\lambda}}{j!} \qquad (j = 0, 1, \ldots) \tag{2.32}$$

where

$$\lambda = \frac{e^{-2ky}}{k \cdot k!} . \tag{2.33}$$

Proof of Theorem 2c. It is easily seen that under the conditions of Theorem 2c

$$\lim_{n \to +\infty} \mathbf{M}_{n,N(n)}(\tau_k) = \lambda .$$

Similarly from (2.16) it follows that for $r = 1, 2, \ldots$

$$\lim_{n \to +\infty} \sum_{S_j \in T_k^{(n)}} \mathbf{M}_{n,N(n)}\big(\varepsilon(S_1)\, \varepsilon(S_2) \ldots \varepsilon(S_r)\big) = \frac{\lambda^r}{r!}$$

and the proof of Theorem 2c is completed by the use of our Lemma 1 exactly as in the proof of Theorem 2a.

Note that Theorem 2c generalizes the results of the paper [7], where only the case $k = 1$ is considered.

§ 3. Cycles

Let us consider now the threshold function of cycles of a given order. The situation is described by the following

Theorem 3a. *Suppose that*

$$N(n) \sim c\, n \text{ where } c > 0 . \tag{3.1}$$

Let γ_k denote the number of cycles of order k contained in $\Gamma_{n,N}$ ($k = 3, 4, \ldots$). Then we have

$$\lim_{n \to +\infty} \mathbf{P}_{n,N(n)}(\gamma_k = j) = \frac{\lambda^j e^{-\lambda}}{j!} \qquad (j = 0, 1, \ldots) \tag{3.2}$$

where

$$\lambda = \frac{(2\, c)^k}{2\, k} . \tag{3.3}$$

Thus the threshold distribution corresponding to the threshold function $A(n) = n$ for the property that the graph contains a cycle of order k is $1 - e^{-\frac{1}{2k}(2c)^k}$.

It is interesting to compare Theorem 3a with the following two theorems:

Theorem 3b. *Suppose again that* (3.1) *holds. Let γ_k^* denote the number of isolated cycles of order k contained in $\Gamma_{n,N}$ ($k = 3, 4, \ldots$). Then we have*

$$\lim_{n \to +\infty} \mathbf{P}_{n,N(n)}(\gamma_k^* = j) = \frac{\mu^j e^{-\mu}}{j!} \qquad (j = 0, 1, \ldots) \tag{3.4}$$

where

$$\mu = \frac{(2\, c\, e^{-2c})^k}{2\, k} . \tag{3.5}$$

Remark. Note that according to Theorem 3b for isolated cycles there does not exist a threshold in the ordinary sense, as $1 - e^{-\mu}$ reaches its maximum $1 - e^{-\frac{1}{2ke^k}}$ for $c = \frac{1}{2}$ $\left(\text{i. e. for } N(n) \sim \frac{n}{2}\right)$ and then again decreases ;

thus the probability that $\Gamma_{n,N}$ contains an *isolated* cycle of order k never approaches 1.

Theorem 3c. *Let δ_k denote the number of components of $\Gamma_{n,N}$ consisting of $k \geqq 3$ points and k edges. If* (3.1) *holds then we have*

$$\lim_{n\to+\infty} \mathbf{P}_{n,N(n)}(\delta_k = j) = \frac{\omega^j e^{-\omega}}{j!} \qquad (j = 0, 1, \dots) \tag{3.6}$$

where

$$\omega = \frac{(2\,c\,e^{-2c})^k}{2\,k}\left(1 + k + \frac{k^2}{2!} + \dots + \frac{k^{k-3}}{(k-3)!}\right). \tag{3.7}$$

Proof of Theorems 3a., 3b. and 3c. As from k given points one can form $\frac{1}{2}(k-1)!$ cycles of order k we have evidently for fixed k and for $N = O(n)$

$$\mathbf{M}(\gamma_k) = \frac{1}{2}\binom{n}{k}(k-1)!\,\frac{\binom{\binom{n}{2}-k}{N-k}}{\binom{\binom{n}{2}}{N}} \sim \frac{\left(\frac{2N}{n}\right)^k}{2\,k} \tag{3.8}$$

while

$$\mathbf{M}(\gamma_k^*) = \frac{1}{2}\binom{n}{k}(k-1)!\,\frac{\binom{\binom{n-k}{2}}{N-k}}{\binom{\binom{n}{2}}{N}} \sim \frac{\left(\frac{2N}{n}e^{-\frac{2N}{n}}\right)^k}{2\,k}. \tag{3.9}$$

As regards Theorem 3c it is known (see [10] and [15]) that the number of connected graphs $G_{k,k}$ (i. e. the number of connected graphs consisting of k labelled vertices and k edges) is exactly

$$\Theta_k = \frac{1}{2}(k-1)!\left(1 + k + \frac{k^2}{2} + \dots + \frac{k^{k-3}}{(k-3)!}\right). \tag{3.10}$$

Now we have clearly

$$\mathbf{M}(\delta_k) = \binom{n}{k}\Theta_k\,\frac{\binom{\binom{n-k}{2}}{N-k}}{\binom{\binom{n}{k}}{N}} \sim \frac{\left(\frac{2N}{n}e^{-\frac{2N}{n}}\right)^k}{2\,k}\left(1 + k + \frac{k^2}{2!} + \dots + \frac{k^{k-3}}{(k-3)!}\right). \tag{3.11}$$

For large values of k we have (see [15])

$$\Theta_k \sim \sqrt{\frac{\pi}{8}}\, k^{k-1/2} \tag{3.12}$$

and thus

$$\mathbf{M}(\delta_k) \sim \frac{\left(\frac{2N}{n} e^{1-\frac{2N}{n}}\right)^k}{4k}. \tag{3.13}$$

For $N \sim \frac{n}{2}$ we obtain by some elementary computation using (7) that for large values of k (such that $k = o(n^{3/4})$).

$$\mathbf{M}(\delta_k) \sim \frac{e^{-\frac{k^3}{n^2}}}{4k}. \tag{3.14}$$

Using (3.8), (3.9) and (3.11) the proofs of Theorems 3a, 3b and 3c follow the same lines as that of Theorem 2a, using Lemma 1. The details may be left to the reader.

Similar results can be proved for other types of subgraphs, e. g. complete subgraphs of a given order. As however these results and their proofs have the same pattern as those given above we do not dwell on the subject any longer and pass to investigate *global properties of the random graph* $\Gamma_{n,N}$.

§ 4. The total number of points belonging to trees

We begin by proving

Theorem 4a. *If* $N = o(n)$ *the graph* $\Gamma_{n,N}$ *is, with probability tending to* 1 *for* $n \to +\infty$, *the union of disjoint trees.*

Proof of Theorem 4a. A graph consists of disjoint trees if and only if there are no cycles in the graph. The number of graphs $G_{n,N}$ which contain at least one cycle can be enumerated as was shown in § 1 for each value k of the length of this cycle. In this way, denoting by T the property that the graph is a union of disjoint trees, and by $\overline{T}$ the opposite of this property, i. e. that the graph contains at least one cycle, we have

$$\mathbf{P}_{n,N}(\overline{T}) \leq \sum_{k=3}^{n} \binom{n}{k} (k-1)!\, \frac{\binom{\binom{n}{2}-k}{N-k}}{\binom{\binom{n}{2}}{N}} = O\left(\frac{N}{n}\right). \tag{4.1}$$

It follows that if $N = o(n)$ we have $\lim\limits_{n\to+\infty} \mathbf{P}_{n,N}(T) = 1$ which proves Theorem 4a.

If N is of the same order of magnitude as n i. e. $N \sim cn$ with $c > 0$, then the assertion of Theorem 4a is no longer true. Nevertheless if $c < 1/2$, still almost all points (in fact $n - O(1)$ points) of $\Gamma_{n,N}$ belong to isolated trees. There is however a surprisingly abrupt change in the structure of $\Gamma_{n,N}$ with $N \sim cn$ when c surpasses the value $\frac{1}{2}$. If $c > 1/2$ in the average only a positive fraction of all points of $\Gamma_{n,N}$ belong to isolated trees, and the value of this fraction tends to 0 for $c \to +\infty$.

Thus we shall prove

Theorem 4b. *Let $V_{n,N}$ denote the number of those points of $\Gamma_{n,N}$ which belong to an isolated tree contained in $\Gamma_{n,N}$. Let us suppose that*

$$\lim_{n\to+\infty} \frac{N(n)}{n} = c > 0 . \tag{4.2}$$

Then we have

$$\lim_{n\to+\infty} \frac{\mathbf{M}(V_{n,N(n)})}{n} = \begin{cases} 1 & \text{for } c \leq 1/2 \\ \dfrac{x(c)}{2c} & \text{for } c > \dfrac{1}{2} \end{cases} \tag{4.3}$$

where $x = x(c)$ is the only root satisfying $0 < x < 1$ of the equation

$$x e^{-x} = 2 c e^{-2c} , \tag{4.4}$$

which can also be obtained as the sum of a series as follows:

$$x(c) = \sum_{k=1}^{\infty} \frac{k^{k-1}}{k!} (2 c e^{-2c})^k . \tag{4.5}$$

Proof of Theorem 4b. We shall need the well known fact that the inverse function of the function

$$y = x e^{-x} \qquad (0 \leq x \leq 1) \tag{4.6}$$

has the power series expansion, convergent for $0 \leq y \leq \frac{1}{e}$

$$x = \sum_{k=1}^{+\infty} \frac{k^{k-1} y^k}{k!} . \tag{4.7}$$

Let τ_k denote the number of isolated trees of order k contained in $\Gamma_{n,N}$. Then clearly

$$V_{n,N} = \sum_{k=1}^{n} k \tau_k \tag{4.8}$$

and thus

$$\mathbf{M}(V_{n,N}) = \sum_{k=1}^{n} k \mathbf{M}(\tau_k) . \tag{4.9}$$

By (2.18), if (4.2) holds, we have

$$\lim_{n\to+\infty} \frac{1}{n} \mathbf{M}(\tau_k) = \frac{1}{2c} \frac{k^{k-2}}{k!} (2 c e^{-2c})^k . \tag{4.10}$$

Thus we obtain from (4.10) that for $c \leq 1/2$

$$\liminf_{n\to+\infty} \frac{\mathbf{M}(V_{n,N(n)})}{n} \geq \frac{1}{2c} \sum_{k=1}^{s} \frac{k^{k-1}(2ce^{-2c})^k}{k!} \quad \text{for any } s \geq 1. \tag{4.11}$$

As (4.11) holds for any $s \geq 1$ we obtain

$$\liminf_{n\to+\infty} \frac{\mathbf{M}(V_{n,N(n)})}{n} \geq \frac{1}{2c} \sum_{k=1}^{\infty} \frac{k^{k-1}(2ce^{-2c})^k}{k!}. \tag{4.12}$$

But according to (4.7) for $c \leq 1/2$ we have

$$\sum_{k=1}^{\infty} \frac{k^{k-1}(2ce^{-2c})^k}{k!} = 2c.$$

Thus it follows from (4.12) that for $c \leq 1/2$

$$\liminf_{n\to+\infty} \frac{\mathbf{M}(V_{n,N(n)})}{n} \geq 1. \tag{4.13}$$

As however $V_{n,N(n)} \leq n$ and thus $\limsup\limits_{n\to+\infty} \dfrac{M(V_{n,N(n)})}{n} \leq 1$ it follows that if (4.2) holds and $c \leq 1/2$ we have

$$\lim_{n\to+\infty} \frac{\mathbf{M}(V_{n,N(n)})}{n} = 1. \tag{4.14}$$

Now let us consider the case $c > \dfrac{1}{2}$. It follows from (2.18) that if (4.2) holds with $c > 1/2$ we obtain

$$\mathbf{M}(V_{n,N(n)}) = \frac{n^2}{2N} \sum_{k=1}^{n} \frac{k^{k-1}}{k!} \left(\frac{2N(n)}{n} e^{-\frac{2N(n)}{n}} \right)^k + O(1) \tag{4.15}$$

where the bound of the term $O(1)$ depends only on c. As however for $N(n) \sim \sim nc$ with $c > 1/2$

$$\sum_{k=n+1}^{\infty} \frac{k^{k-1}}{k!} \left(\frac{2N(n)}{n} e^{-\frac{2N(n)}{n}} \right)^k = O\left(\frac{1}{n^{3/2}} \right)$$

it follows that

$$\mathbf{M}(V_{n,N(n)}) = \frac{n^2}{2N(n)} x\left(\frac{N(n)}{n} \right) + O(1) \tag{4.16}$$

where $x = x\left(\dfrac{N(n)}{n}\right)$ is the only solution with $o < x < 1$ of the equation $xe^{-x} = \dfrac{2N(n)}{n} e^{-\frac{2N(n)}{n}}$. Thus it follows that if (4.2) holds with $c > 1/2$ we have

$$\lim_{n\to+\infty} \frac{\mathbf{M}(V_{n,N(n)})}{n} = \frac{x(c)}{2c} \tag{4.17}$$

where $x(c)$ is defined by (4.5).

The graph of the function $x(c)$ is shown on Fig. 1a; its meaning is shown by Fig. 1b. The function

$$y = \begin{cases} 1 & \text{for } c \leq 1/2 \\ \dfrac{x(c)}{2c} & \text{for } c > 1/2 \end{cases}$$

is shown on Fig. 2a.

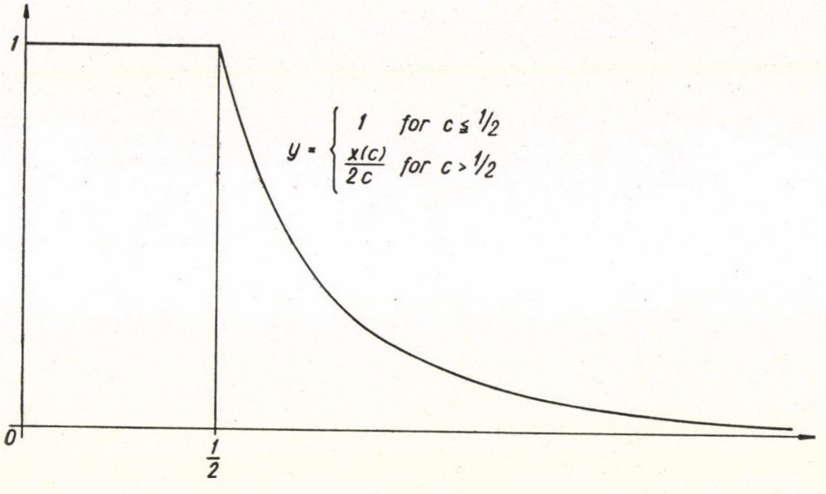

Figure 2a.

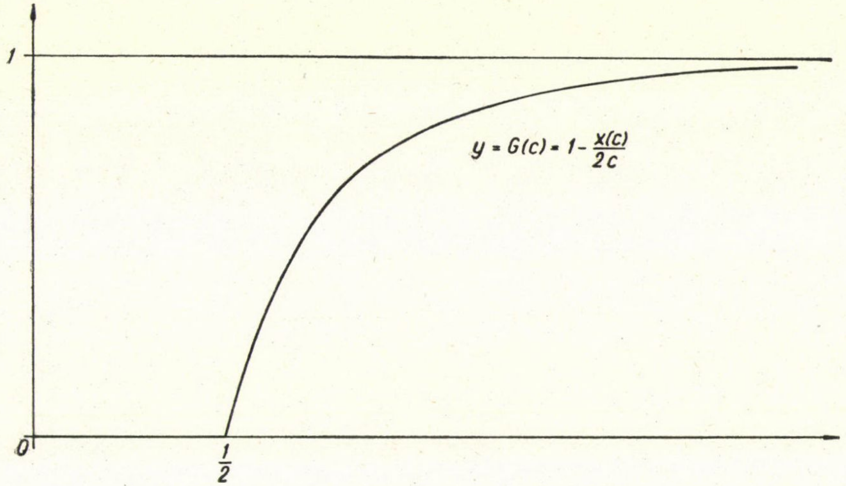

Figure 2b.

Thus the proof of Theorem 4b is complete. Let us remark that in the same way as we obtained (4.16) we get that if (4.2) holds with $c < 1/2$ we have

$$\mathbf{M}(V_{n,N(n)}) = n - O(1) \tag{4.18}$$

where the bound of the $O(1)$ term depends only on c. (However (4.18) is not true for $c = \frac{1}{2}$ as will be shown below.)

It follows by the well known inequality of Markov

$$\mathbf{P}(\xi > a) \leq \frac{1}{a}\mathbf{M}(\xi) \tag{4.19}$$

valid for any nonnegative random variable ξ and any $a > \mathbf{M}(\xi)$, that the following theorem holds:

Theorem 4c. *Let $V_{n,N}$ denote the number of those points of $\Gamma_{n,N}$ which belong to isolated trees contained in $\Gamma_{n,N}$. Then if ω_n tends arbitrarily slowly to $+\infty$ for $n \to +\infty$ and if* (4.2) *holds with $c < 1/2$ we have*

$$\lim_{n\to+\infty} \mathbf{P}(V_{n,N(n)} \geq n - \omega_n) = 1\,. \tag{4.20}$$

The case $c > 1/2$ is somewhat more involved. We prove

Theorem 4d. *Let $V_{n,N}$ denote the number of those points of $\Gamma_{n,N}$ which belong to an isolated tree contained in $\Gamma_{n,N}$. Let us suppose that* (4.2) *holds with $c > 1/2$. It follows that if ω_n tends arbitrarily slowly to $+\infty$, we have*

$$\lim_{n\to+\infty} \mathbf{P}\left(\left|V_{n,N(n)} - \frac{n^2}{2\,N(n)}x\left(\frac{N(n)}{n}\right)\right| > \sqrt{n}\,\omega_n\right) = 0 \tag{4.21}$$

where $x = x\left(\frac{N(n)}{n}\right)$ is the only solution with $0 < x < 1$ of the equation

$$xe^{-x} = \frac{2\,N(n)}{n}e^{-\frac{2N(n)}{n}}\,.$$

Proof. We have clearly, as the series $\sum_{k=1}^{\infty}\frac{k^k}{k!}(2\,ce^{-2c})^k$ is convergent, $\mathbf{D}^2\,(V_{n,N(n)}) = O(n)$. Thus (4.21) follows by the inequality of *Chebyshev.*

Remark. It follows from (4.21) that we have for any $c > 1/2$ and any $\varepsilon > 0$

$$\lim_{n\to+\infty} \mathbf{P}\left(\left|\frac{V_{n,N(n)}}{n} - \frac{x(c)}{2\,c}\right| < \varepsilon\right) = 1 \tag{4.22}$$

where $x(c)$ is defined by (4.5).

As regards the case $c = {}^1/_2$ we formulate the theorem which will be needed latter.

Theorem 4e. *Let $V_{n,N}(r)$ denote the number of those points of $\Gamma_{n,N}$ which belong to isolated trees of order $\geq r$ and $\tau_{n,N}(r)$ the number of isolated trees of order $\geq r$ contained in $\Gamma_{n,N}$. If $N(n) \sim \frac{n}{2}$ we have for any $\delta > o$*

$$\lim_{n \to +\infty} \mathbf{P}\left(\left|\frac{V_{n,N(n)}(r)}{n} - \sum_{k=r}^{+\infty} \frac{k^{k-1}}{k!} e^{-k}\right| < \delta\right) = 1 \tag{4.23}$$

and

$$\lim_{n \to +\infty} \mathbf{P}\left(\left|\frac{\tau_{n,N(n)}(r)}{n} - \sum_{k=r}^{+\infty} \frac{k^{k-2}}{k!} e^{-k}\right| < \delta\right) = 1\,. \tag{4.24}$$

The proof follows the same lines as those of the preceding theorems.

§ 5. The total number of points belonging to cycles

Let us determine first the average number of all cycles in $\Gamma_{n,N}$. We prove that this number remains bounded if $N(n) \sim cn$ and $c < {}^1/_2$ but not if $c = {}^1/_2$.

Theorem 5a. *Let $H_{n,N}$ denote the number of all cycles contained in $\Gamma_{n,N}$. Then we have if $N(n) \sim cn$ holds with $c < \frac{1}{2}$*

$$\lim_{n \to +\infty} \mathbf{M}(H_{n,N(n)}) = \frac{1}{2} \log \frac{1}{1-2c} - c - c^2 \tag{5.1}$$

while we have for $c = \frac{1}{2}$

$$\mathbf{M}(H_{n,N(n)}) \sim \frac{1}{4} \log n. \tag{5.2}$$

Proof. Clearly if γ_k is the number of all cycles of order k contained in $\Gamma_{n,N}$ we have

$$H_{n,N} = \sum_{k=1}^{n} \gamma_k.$$

Now (5.1) follows easily, taking into account that (see (3.8))

$$\mathbf{M}(\gamma_k) = \frac{1}{2}\binom{n}{k}(k-1)!\,\frac{\binom{\binom{n}{2}-k}{N-k}}{\binom{\binom{n}{2}}{N}} = \frac{\left(\frac{2N}{n}\right)^k}{2k}\left(1 + O\left(\frac{k^2}{n}\right)\right). \tag{5.3}$$

If $c = {}^1/_2$ we have by (3.8)

$$\mathbf{M}(\gamma_k) \sim \frac{1}{2\,k} e^{-\frac{3k^2}{2n}}. \tag{5.4}$$

As $\sum_{k=3}^{n} \frac{1}{2\,k} e^{-\frac{3k^2}{2n}} \sim \frac{1}{4} \log n$, it follows that (5.2) holds. Thus Theorem 5a is proved.

Let us remark that it follows from (5.2) that (4.18) is not true for $c = {}^1/_2$.

Similarly as before we can prove corresponding results concerning the random variable $H_{n,N}$ itself.

We have for instance in the case $c = {}^1/_2$ for any $\varepsilon > o$

$$\lim_{n\to+\infty} \mathbf{P}\left(\left| \frac{H_{n,N(n)}}{\log n} - \frac{1}{4} \right| < \varepsilon \right) = 1. \tag{5.5}$$

This can be proved by the same method as used above: estimating the variance and using the inequality of Chebyshev.

An other related result, throwing more light on the appearance of cycles in $\Gamma_{n,N}$ runs as follows.

Theorem 5b. *Let K denote the property that a graph contains at least one cycle. Then we have if $N(n) \sim nc$ holds with $c \leq {}^1/_2$*

$$\lim_{n\to+\infty} \mathbf{P}_{n,N(n)}(K) = 1 - \sqrt{1-2\,c}\, e^{c+c^2}. \tag{5.6}$$

Thus for $c = \frac{1}{2}$ it is „almost sure" that $\Gamma_{n,N(n)}$ contains at least one cycle, while for $c < \frac{1}{2}$ the limit for $n \to +\infty$ of the probability of this is less than 1.

Proof. Let us suppose first $c < \frac{1}{2}$. By an obvious sieve (taking into account that according to Theorem 1 the probability that there will be in $\Gamma_{n,N(n)}$ with $N(n) \sim nc$ $(c < {}^1/_2)$ two circles having a point in common is negligibly small) we obtain

$$\lim_{n\to+\infty} \mathbf{P}_{n,N(n)}(\overline{K}) = e^{-\lim\limits_{n\to+\infty} \mathbf{M}(H_{n,N(n)})} = \sqrt{1-2\,c}\, e^{c+c^2}. \tag{5.7}$$

Thus (5.6) follows for $c < {}^1/_2$. As for $c \to {}^1/_2$ the function on the right of (5.6) tends to 1, it follows that (5.6) holds for $c = {}^1/_2$ too. The function $y = 1 - \sqrt{1-2c}\, e^{c+c^2}$ is shown on Fig. 3.

We prove now the following

Theorem 5c. *Let $H^*_{n,N}$ denote the total number of points of $\Gamma_{n,N}$ which belong to some cycle. Then we have for $N = N(n) \sim cn$ with $0 < c < {}^1/_2$*

$$\lim_{n\to+\infty} \mathbf{M}(H^*_{n,N(n)}) = \frac{4\,c^3}{1-2\,c}. \tag{5.8}$$

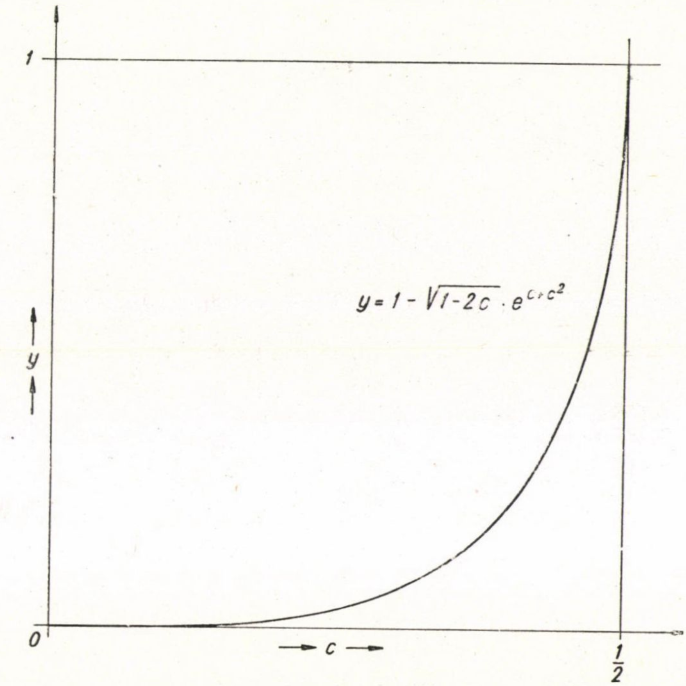

Figure 3.

Proof of Theorem 5c. As according to Theorem 1 the probability that two cycles should have a point in common is negligibly small, we have by (5.3)

$$\mathbf{M}(H^*_{n,N(n)}) \sim \sum_{k=1}^{n} k\,\gamma_k \sim \frac{(2\,c)^3}{2(1-2\,c)} = \frac{4\,c^3}{1-2\,c}.$$

The size of that part of $\Gamma_{n,N}$ which does not consist of trees is still more clearly shown by the following

Theorem 5d. *Let $\vartheta_{n,N}$ denote the number of those points of $\Gamma_{n,N}$ which belong to components containing exactly one cycle. Then we have for $N = N(n) \sim \sim cn$ in case $c \neq {}^1/_2$*

$$(5.9)\qquad \lim_{n\to+\infty} \mathbf{M}(\vartheta_{n,N(n)}) = \frac{1}{2}\sum_{k=3}^{+\infty}(2\,ce^{-2c})^k\left(1+\frac{k}{1!}+\frac{k^2}{2!}+\ldots+\frac{k^{k-3}}{(k-3)!}\right)$$

while for $c = {}^1/_2$ we have

$$(5.10)\qquad \mathbf{M}(\vartheta_{n,N(n)}) \sim \frac{\Gamma\left(\frac{1}{3}\right)}{12}\,n^{2/3}$$

where $\Gamma(x)$ denotes the gamma-function $\Gamma(x) = \int_0^\infty t^{x-1}\,e^{-t}\,dt$ for $x > 0$.

Proof of Theorem 5d. (5.9) follows immediately from (3.11); for $c = 1/2$ we have by (3.14)

$$\mathbf{M}(\vartheta_{n,N(n)}) \sim \frac{1}{4} \sum_{k=3}^{n} e^{-\frac{k^3}{n^2}} \sim \frac{\Gamma\left(\frac{1}{3}\right)}{12} n^{2/3}.$$

Remark. Note that for $c \to {}^1/_2$

$$\frac{1}{2} \sum_{k=3}^{\infty} (2\,ce^{-2c})^k \left(1 + \frac{k}{1!} + \ldots + \frac{k^{k-3}}{(k-3)!}\right) \sim \frac{1}{4(1-2\,c)^2}.$$

Thus the average number of points belonging to components containing exactly one cycle tends to $+\infty$ as $\frac{1}{4(1-2\,c)^2}$ for $c \to {}^1/_2$.

We now prove

Theorem 5e. *For $N(n) \sim cn$ with $0 < c < {}^1/_2$ all components of $\Gamma_{n,N(n)}$ are with probability tending to 1 for $n \to +\infty$, either trees or components containing exactly one cycle.*

Proof. Let $\psi_{n,N}$ denote the number of points of $\Gamma_{n,N}$ belonging to components which contain more edges than vertices and the number of vertices of which is less than $\sqrt{\log n}$. We have clearly for $N(n) \sim cn$ with $c < {}^1/_2$

$$\mathbf{M}(\psi_{n,N(n)}) \leq \sum_{k=4}^{[\sqrt{\log n}]} k \binom{n}{k} 2^{\binom{k}{2}} \frac{\binom{\binom{n-k}{2}}{N-k-1}}{\binom{\binom{n}{2}}{N}} = O\left(n^{\frac{\log 2}{2} - 1}\right).$$

Thus

$$\mathbf{P}(\psi_{n,N(n)} \geqq 1) = O\left(\frac{1}{n^{1-\frac{\log 2}{2}}}\right).$$

On the other hand by Theorem 4c the probability that a component consisting of more than $\sqrt{\log n}$ points should not be a tree tends to 0. Thus the assertion of Theorem 5e follows.

§ 6. The number of components

Let us turn now to the investigation of the average number of components of $\Gamma_{n,N}$. It will be seen that the above discussion contains a fairly complete solution of this question. We prove the following

Theorem 6. *If $\zeta_{n,N}$ denotes the number of components of $\Gamma_{n,N}$ then we have if $N(n) \sim cn$ holds with $0 < c < \frac{1}{2}$*

$$\mathbf{M}(\zeta_{n,N(n)}) = n - N(n) + O(1) \tag{6.1}$$

where the bound of the O-term depends only on c. If $N(n) \sim \frac{n}{2}$ we have

$$\mathbf{M}(\zeta_{n,N(n)}) = n - N(n) + O(\log n). \tag{6.2}$$

If $N(n) \sim cn$ holds with $c > \frac{1}{2}$ we have

$$\lim_{n\to+\infty} \frac{\mathbf{M}(\zeta_{n,N(n)})}{n} = \frac{1}{2c}\left(x(c) - \frac{x^2(c)}{2}\right) \tag{6.3}$$

where $x = x(c)$ is the only solution satisfying $0 < x < 1$ of the equation $xe^{-x} = 2ce^{-2c}$, i. e.

$$x(c) = \sum_{k=1}^{\infty} \frac{k^{k-1}}{k!}(2ce^{-2c})^k. \tag{6.4}$$

Proof of Theorem 6. Let us consider first the case $c < \frac{1}{2}$. Clearly if we add a new edge to a graph, then either this edge connects two points belonging to different components, in which case the number of components is decreased by 1, or it connects two points belonging to the same component in which case the number of components does not change but at least one new cycle is created. Thus[2]

$$\zeta_{n,N} - (n - N) \leq H_{n,N} \tag{6.5}$$

where $H_{n,N}$ is the total number of cycles in $\Gamma_{n,N}$. Thus by Theorem 5a it follows that (6.1) holds.

Similarly (6.2) follows also from Theorem 5a. Now we consider the case $c > \frac{1}{2}$.

It is easy to see that for $0 \leq y \leq \frac{1}{e}$ we have (see e. g. [14])

$$\sum_{k=1}^{+\infty} \frac{k^{k-2} y^k}{k!} = x - \frac{x^2}{2} \tag{6.6}$$

where

$$x = \sum_{k=1}^{+\infty} \frac{k^{k-1} y^k}{k!} \tag{6.7}$$

[2] In fact according to a well known theorem of the theory of graphs (see [4], p. 29) being a generalization of Euler's theorem on polyhedra we have $N - n + \zeta_{n,N} = \varkappa^*_{n,N}$, where $\varkappa^*_{n,N}$ — the „cyclomatic number" of the graph $\Gamma_{n,N}$ — is equal to the maximal number of independent cycles, in $\Gamma_{n,N}$ (For a definition of independent cycles see [4] p. 28).

x can be characterized also as the only solution satisfying $0 < x \leq 1$ of the equation $xe^{-x} = y$.

It follows that if $N(n) \sim nc$ holds with $c < {}^1\!/_2$ we have

(6.8) $$\mathbf{M}(\zeta_{n,N(n)}) = \frac{n^2}{2\,N(n)}\left(\frac{2\,N(n)}{n} - \frac{4\,N^2(n)}{2\,n^2}\right) + O(1) = n - N(n) + O(1)$$

which leads to a second proof of the first part of Theorem 6.

To prove the second part, let us remark first that the number of components of order greater than A is clearly $\leq \frac{n}{A}$. Thus if $\zeta_{n,N}(A)$ denotes the number of components of order $\leq A$ of $\Gamma_{n,N}$ we have clearly

(6.9) $$\mathbf{M}(\zeta_{n,N}) = \mathbf{M}(\zeta_{n,N}(A)) + O\left(\frac{n}{A}\right).$$

The average number of components of fixed order k which contain at least k edges will be clearly according to Theorem 1 of order $\left(\frac{N}{n}\right)^k$, i. e. bounded for each fixed value of k. As A can be chosen arbitrarily large we obtain from (6.9) that

(6.10) $$\mathbf{M}(\zeta_{n,N}) \sim \sum_{k=1}^{n} \mathbf{M}(\tau_k).$$

According to (2.18) it follows that

(6.11) $$\mathbf{M}(\zeta_{n,N}) \sim \frac{n^2}{2\,N} \sum_{k=1}^{+\infty} \frac{k^{k-2}}{k!}\left(\frac{2\,N}{n}\, e^{-\frac{2N}{n}}\right)^k$$

and thus, according to (6.6) if $N(n) \sim cn$ holds with $c > {}^1\!/_2$ we have

(6.12) $$\lim_{n \to +\infty} \frac{\mathbf{M}(\zeta_{n,N(n)})}{n} = \frac{1}{2\,c}\left(x(c) - \frac{x^2(c)}{2}\right)$$

where $x(c)$ is defined by (6.4). Thus Theorem 6 is completely proved.

Let us add some remarks. Theorem 6 illustrates also the fundamental change in the structure of $\Gamma_{n,N}$ which takes place if N passes $\frac{n}{2}$. While the average number of components of $\Gamma_{n,N}$ (as a function of N with n fixed) decreases linearly if $N \leq \frac{n}{2}$ this is no longer true for $N > \frac{n}{2}$; the average number of components decreases from this point onward more and more slowly. The graph of

(6.13) $$z(c) = \lim_{\frac{N(n)}{n} \to c} \frac{\mathbf{M}(\zeta_{n,N(n)})}{n} = \begin{cases} 1 - c & \text{for} \quad 0 \leq c \leq \frac{1}{2} \\ \frac{1}{2\,c}\left(x(c) - \frac{x^2(c)}{2}\right) & \text{for} \quad c > {}^1\!/_2 \end{cases}$$

as a function of c is shown by Fig. 4.

From Theorem 6 one can deduce easily that in case $N(n) \sim cn$ with $c < {}^1/_2$ we have for any sequence ω_n tending arbitrarily slowly to infinity

$$\lim_{n \to +\infty} \mathbf{P}(|\zeta_{n,N(n)} - n + N(n)| < \omega_n) = 1 \tag{6.14}$$

(6.14) follows easily by remarking that clearly $\zeta_{n,N} \geq n - N$.

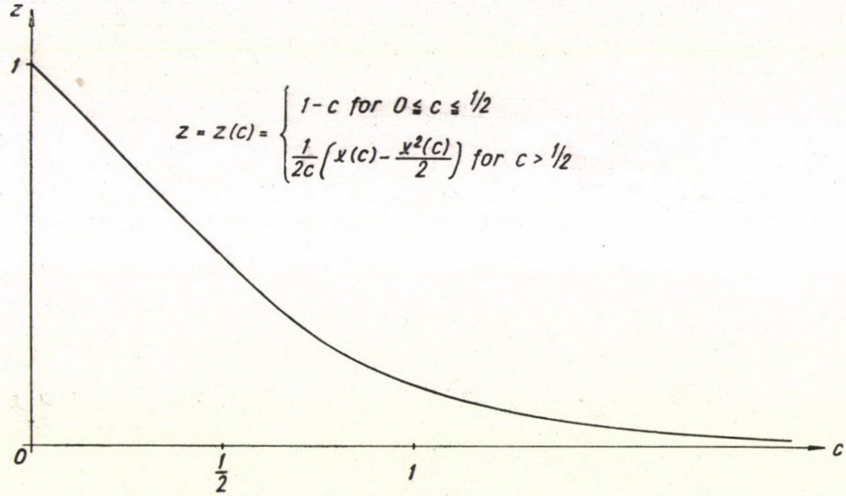

Figure 4.

For the case $N(n) \sim cn$ with $c \geq {}^1/_2$ one obtains by estimating the variance of $\zeta_{n,N(n)}$ and using the inequality of Chebyshev that for any $\varepsilon > 0$

$$\lim_{n \to +\infty} \mathbf{P}\left(\left| \frac{\zeta_{n,N(n)}}{n} - \frac{1}{2\,c}\left(x(c) - \frac{x^2(c)}{2} \right) \right| < \varepsilon \right) = 1\,. \tag{6.15}$$

The proof is similar to that of (4.21) and therefore we do not go into details.

§ 7. The size of the greatest tree

If $N \sim cn$ with $c < {}^1/_2$ then as we have seen in § 6 all but a finite number of points of $\Gamma_{n,N}$ belong to components which are trees. Thus in this case the problem of determining the size of the largest component of $\Gamma_{n,N}$ reduces to the easier question of determining the greatest tree in $\Gamma_{n,N}$. This question is answered by the following.

Theorem 7a. *Let $\Delta_{n,N}$ denote the number of points of the greatest tree which is a component of $\Gamma_{n,N}$. Suppose $N = N(n) \sim cn$ with $c \neq {}^1/_2$. Let ω_n be a sequence* *tending arbitrarily slowly to* $+\infty$. *Then we have*

$$\lim_{n\to+\infty} \mathbf{P}\left(\Delta_{n,N(n)} \geqq \frac{1}{\alpha}\left(\log n - \frac{5}{2}\log\log n\right) + \omega_n\right) = 0 \tag{7.1}$$

and

$$\lim_{n\to+\infty} \mathbf{P}\left(\Delta_{n,N(n)} \geqq \frac{1}{\alpha}\left(\log n - \frac{5}{2}\log\log n\right) - \omega_n\right) = 1 \tag{7.2}$$

where

$$e^{-\alpha} = 2\,ce^{1-2c} \qquad \text{(i. e. } \alpha = 2\,c - 1 - \log 2\,c \tag{7.3}$$

and thus $\alpha > 0$.)

Proof of Theorem 7a. We have clearly

$$\mathbf{P}(\Delta_{n,N(n)} \geqq z) = \mathbf{P}\left(\sum_{k\geqq z} \tau_k \geqq 1\right) \leqq \sum_{k\geqq z} \mathbf{M}(\tau_k) \tag{7.4}$$

and thus by (2.18)

$$\mathbf{P}(\Delta_{n,N(n)} \geqq z) = O\left(\frac{ne^{-\alpha z}}{z^{5/2}}\right). \tag{7.5}$$

It follows that if $z_1 = \frac{1}{\alpha}\left(\log n - \frac{5}{2}\log\log n\right) + \omega_n$
we have

$$\mathbf{P}(\Delta_{n,N(n)} \geqq z_1) = O(e^{-\alpha\omega_n}). \tag{7.6}$$

This proves (7.1). To prove (7.2) we have to estimate the mean and variance of τ_{z_2} where $z_2 = \frac{1}{\alpha}\left(\log n - \frac{5}{2}\log\log n\right) - \omega_n$. We have by (2.18)

$$\mathbf{M}(\tau_{z_2}) \sim \frac{\alpha^{5/2}}{2\,c\sqrt{2\,\pi}}\, e^{\alpha\omega_n} \tag{7.7}$$

and

$$\mathbf{D}^2(\tau_{z_2}) = O\big(\mathbf{M}(\tau_{z_2})\big). \tag{7.8}$$

Clearly

$$\mathbf{P}(\Delta_{n,N(n)} \geqq z_2) \geqq \mathbf{P}(\tau_{z_2} \geqq 1) = 1 - \mathbf{P}(\tau_{z_2} = 0)$$

and it follows from (7.7) and (7.8) by the inequality of Chebyshev that

$$\mathbf{P}(\tau_{z_2} = 0) = O(e^{-\alpha\omega_n}). \tag{7.9}$$

Thus we obtain

$$\mathbf{P}(\Delta_{n,N(n)} \geqq z_2) \geqq 1 - O(e^{-\alpha\omega_n}). \tag{7.10}$$

Thus (7.2) is also proved.

Remark. If $c < \frac{1}{2}$ the greatest tree which is a component of $\Gamma_{n,N}$ with $N \sim cn$ is — as mentioned above — at the same time the greatest component of $\Gamma_{n,N}$, as $\Gamma_{n,N}$ contains with probability tending to 1 besides trees only components containing a single circle and being of moderate size. This follows evidently from Theorem 4c. As will be seen in what follows (see § 9) for $c > \frac{1}{2}$ the situation is completely different, as in this case $\Gamma_{n,N}$ contains a very large component (in fact of size $G(c)n$ with $G(c) > 0$) which is not a tree. Note that if we put $c = \frac{1}{2k} \log n$ we have $\alpha = \frac{1}{k} \log n$ and $\frac{1}{\alpha} \log n \sim k$ in conformity with Theorem 2c.

We can prove also the following

Theorem 7b. *If $N \sim cn$, where $c \neq \frac{1}{2}$ and $e^{-\alpha} = 2ce^{1-2c}$ then the number of isolated trees of order $h = \frac{1}{\alpha}\left(\log n - \frac{5}{2} \log\log n\right) + l$ resp. of order $\geq h$ (where l is an arbitrary real number such that h is a positive integer) contained in $\Gamma_{n,N}$ has for large n approximately a Poisson distribution with the mean value $\lambda = \frac{\alpha^{5/2} e^{-\alpha l}}{2c\sqrt{2\pi}}$ resp. $\mu = \frac{\alpha^{5/2} e^{-\alpha l}}{2c\sqrt{2\pi}(1-e^{-\alpha})}$.*

Corollary. The probability that $\Gamma_{n,N(n)}$ with $N(n) \sim nc$ where $c \neq \frac{1}{2}$ does not contain a tree of order $\geq \frac{1}{\alpha}\left(\log n - \frac{5}{2} \log\log n\right) + l$ tends to $\exp\left(-\frac{\alpha^{5/2} e^{-\alpha l}}{2c\sqrt{2\pi}(1-e^{-\alpha})}\right)$ for $n \to +\infty$, where $\alpha = 2c - 1 - \log 2c$.

The size of the greatest tree which is a component of $\Gamma_{n,N}$ is fairly large if $N \sim \frac{n}{2}$. This could be guessed from the fact that the constant factor in the expression $\frac{1}{\alpha}\left(\log n - \frac{5}{2} \log\log n\right)$ of the ,,probable size'' of the greatest component of $\Gamma_{n,N}$ figuring in Theorem 7a becomes infinitely large if $c = \frac{1}{2}$.

For the size of the greatest tree in $\Gamma_{n,N}$ with $N \sim \frac{n}{2}$ the following result is valid:

Theorem 7c. *If $N \sim \frac{n}{2}$ and $\Delta_{n,N}$ denotes again the number of points of the greatest tree contained in $\Gamma_{n,N}$, we have for any sequence ω_n tending to $+\infty$ for $n \to +\infty$*

(7.11) $$\lim_{n \to +\infty} \mathbf{P}(\Delta_{n,N} \geq n^{2/3} \omega_n) = 0$$

and

(7.12) $$\lim_{n \to +\infty} \mathbf{P}\left(\Delta_{n,N} \geq \frac{n^{2/3}}{\omega_n}\right) = 1 .$$

Proof of Theorem 7c. We have by some simple computation using (7)

$$\mathbf{M}(\tau_k) = \frac{\binom{n}{k} k^{k-2} \binom{\binom{n-k}{2}}{N-k+1}}{\binom{\binom{n}{2}}{N}} \sim \frac{n k^{k-2} e^{-k}}{k!} e^{-\frac{k^3}{6n^2}}. \tag{7.13}$$

Thus it follows that

$$\mathbf{P}(\Delta_{n,N} \geq n^{2/3}\omega_n) \leq \sum_{k \geq n^{2/3}\omega_n} \mathbf{M}(\tau_k) = O\left(\frac{1}{\sqrt{\omega_n}}\right) \tag{7.14}$$

which proves (7.11).

On the other hand, considering the mean and variance of $\tau^* = \sum_{k \geq \frac{n^{2/3}}{\omega_n}} \tau_k$, it follows that

$$\mathbf{M}(\tau^*) \geq A\,\omega_n^{3/2} \text{ where } A > 0 \text{ and } \mathbf{D}^2(\tau^*) = O(\omega_n^{3/2})$$

and (7.12) follows by using again the inequality of *Chebyshev*. Thus Theorem 7c is proved.

The following theorem can be proved by developing further the above argument and using Lemma 1.

Theorem 7d. *Let $\tau(\mu)$ denote the number of trees of order $\geq \mu n^{2/3}$ contained in $\Gamma_{n,N(n)}$ where $0 < \mu < +\infty$ and $N(n) \sim \frac{n}{2}$. Then we have*

$$\lim_{n \to +\infty} \mathbf{P}_{n,N(n)}(\tau(\mu) = j) = \frac{\lambda^j e^{-\lambda}}{j!} \qquad (j = 0, 1, \ldots) \tag{7.15}$$

where

$$\lambda = \frac{1}{\sqrt{12\pi}} \int_{\frac{1}{6}\mu^3}^{+\infty} \frac{e^{-x}\,dx}{x^{3/2}}. \tag{7.16}$$

§ 8. When is $\Gamma_{n,N}$ a planar graph?

We have seen that the threshold for a subgraph containing k points and $k + d$ edges is $n^{2-\frac{k}{k+d}}$; thus if $N \sim cn$ the probability of the presence of a subgraph having k points and $k + d$ edges in $\Gamma_{n,N}$ tends to 0 for $n \to +\infty$, for each particular pair of numbers $k \geq 4$, $d \geq 1$. This however does not imply that the probability of the presence of a graph of arbitrary order having more edges than vertices in $\Gamma_{n,N}$ with $N \sim nc$ tends also to 0 for $n \to +\infty$. In fact this is not true for $c \geq {}^1/_2$ as is shown by the following

Theorem 8a. *Let $\chi_{n,N}(d)$ denote the number of cycles of $G_{n,N}$ of arbitrary order which are such that exactly d diagonals of the cycle belong also to $\Gamma_{n,N}$. Then if* $N(n) = \frac{n + \lambda\sqrt{n}}{2} + o(\sqrt{n})$ where $-\infty < \lambda < +\infty$, *we have*

$$\lim_{n\to+\infty} \mathbf{P}(\chi_{n,N(n)}(d) = j) = \frac{\varrho^j e^{-\varrho}}{j!} \qquad (j = 0, 1, \ldots) \tag{8.1}$$

where

$$\varrho = \frac{1}{2\cdot 6^d \cdot d!} \int_0^{+\infty} y^{2d-1} e^{\frac{\lambda y}{\sqrt{3}}} \cdot e^{-\frac{y^2}{2}} dy . \tag{8.2}$$

Proof of Theorem 8a. We have clearly as the number of diagonals of a k — gon is equal to $\frac{k(k-3)}{2}$

$$\mathbf{M}(\chi_{n,N}(d)) = \sum_{k=4}^{n} \frac{1}{2}\binom{n}{k}(k-1)!\binom{\frac{k(k-3)}{2}}{d}\frac{\binom{\binom{n}{2}-\binom{k}{2}}{N-k-d}}{\binom{\binom{n}{2}}{N}} \tag{8.3}$$

and thus if $N(n) = \frac{n + \lambda\sqrt{n}}{2} + o(\sqrt{n})$

$$\mathbf{M}(\chi_{n,N(n)}(d)) \sim \frac{1}{2^{d+1}\cdot d!\, n^d} \sum_{k=4}^{n} k^{2d-1}\left(1 + \frac{\lambda}{\sqrt{n}}\right)^k e^{-\frac{3k^2}{2n}} . \tag{8.4}$$

It follows from (8.4) that

$$\lim_{n\to+\infty} \mathbf{M}(\chi_{n,N(n)}(d)) = \frac{1}{2\cdot 6^d\, d!} \int_0^{\infty} y^{2d-1} e^{\frac{\lambda y}{\sqrt{3}} - \frac{y^2}{2}} dy . \tag{8.5}$$

The proof can be finished by the same method as used in proving Theorem 2a.

Remark. Note that Theorem 8a implies that if $N(n) = \frac{n}{2} + \omega_n\sqrt{n}$ with $\omega_n \to +\infty$ then the probability that $\Gamma_{n,N(n)}$ contains cycles with any prescribed number of diagonals tends to 1, while if $N(n) = \frac{n}{2} - \omega_n\sqrt{n}$ the same probability tends to 0. This shows again the fundamental difference in the structure of $\Gamma_{n,N}$ between the cases $N < \frac{n}{2}$ and $N > \frac{n}{2}$. This difference can be expressed also in the form of the following

Theorem 8b. *Let us suppose that $N(n) \sim nc$. If $c < \frac{1}{2}$ the probability* *that the graph $\Gamma_{n,N(n)}$ is planar is tending to* 1 *while for* $c > \frac{1}{2}$ *this probability tends to* 0.

Proof of Theorem 8b. As well known trees and connected graphs containing exactly one cycle are planar. Thus the first part of Theorem 8b follows from Theorem 5e. On the other hand if a graph contains a cycle with 3 diagonals such that if these diagonals connect the pairs of points (P_i, P'_i) $(i = 1, 2, 3)$ the cyclic order of these points in the cycle is such that each pair (P_i, P'_i) dissects the cycle into two paths which both contain two of the other points then the graph is not planar. Now it is easy to see that among the $\binom{\frac{k(k-3)}{2}}{3}$ triples of 3 diameters of a given cycle of order k there are at least $\binom{k}{6}$ triples which have the mentioned property and thus for large values of k approximately one out of 15 choices of the 3 diagonals will have the mentioned property. It follows that if $N(n) = \frac{n}{2} + \omega_n \sqrt{n}$ with $\omega_n \to +\infty$, the probability that $\Gamma_{n,N(n)}$ is not planar tends to 1 for $n \to +\infty$. This proves Theorem 8b. We can show that for $N(n) = \frac{n}{2} + \lambda \sqrt{n}$ with any real λ the probability of $\Gamma_{n,N(n)}$ not being planar has a positive lower limit, but we cannot calculate ts value. It may even be 1, though this seems unlikely.

§ 9. On the growth of the greatest component

We prove in this § (see Theorem 9b) that the size of the greatest component of $\Gamma_{n,N(n)}$ is for $N(n) \sim cn$ with $c > 1/2$ with probability tending to 1 approximately $G(c)n$ where

$$G(c) = 1 - \frac{x(c)}{2c} \tag{9.1}$$

and $x(c)$ is defined by (6.4). (The curve $y = G(c)$ is shown on Fig. 2b).

Thus by Theorem 6 for $N(n) \sim cn$ with $c > 1/2$ almost all points of $\Gamma_{n,N(n)}$ (i. e. all but $o(n)$ points) belong either to some small component which is a tree (of size at most $1/\alpha\,(\log n - \frac{5}{2}\log\log n) + O(1)$ where $\alpha = 2c - 1 - \log 2c$ by Theorem 7a) or to the single "giant" component of the size $\sim G(c)n$.

Thus the situation can be summarized as follows: the largest component of $\Gamma_{n,N(n)}$ is of order $\log n$ for $\frac{N(n)}{n} \sim c < 1/2$, of order $n^{2/3}$ for $\frac{N(n)}{n} \sim \frac{1}{2}$ and of order n for $\frac{N(n)}{n} \sim c > 1/2$. This double "jump" of the size of the largest component when $\frac{N(n)}{n}$ passes the value $1/2$ is one of the most striking facts concerning random graphs. We prove first the following

Theorem 9a. *Let $\mathcal{H}_{n,N}(A)$ denote the set of those points of $\Gamma_{n,N}$ which belong to components of size $> A$, and let $H_{n,N}(A)$ denote the number of elements of the set $\mathcal{H}_{n,N}(A)$. If $N_1(n) \sim (c - \varepsilon)\, n$ where $\varepsilon > 0$, $c - \varepsilon \geqq {}^1/_2$ and $N_2(n) \sim cn$ then with probability tending to 1 for $n \to +\infty$ from the $H_{n,N_1(n)}(A)$ points belonging to $\mathcal{H}_{n,N_1(n)}(A)$ more than $(1 - \delta)\, H_{n,N_1(n)}(A)$ points will be contained in the same component of $\Gamma_{n,N_2(n)}$ for any δ with $0 < \delta < 1$ provided that*

$$A \geqq \frac{50}{\varepsilon^2 \delta^2}. \tag{9.2}$$

Proof of Theorem 9a. According to Theorem 2b the number of points belonging to trees of order $\leq A$ is with probability tending to 1 for $n \to +\infty$ equal to

$$n \left(\sum_{k=1}^{A} \frac{k^{k-1}}{k!} [2(c - \varepsilon)]^{k-1} e^{-2(c-\varepsilon)} \right) + o(n)\,.$$

On the other hand, the number of points of $\Gamma_{n,N_1(n)}$ belonging to components of size $\leq A$ and containing exactly one cycle is according to Theorem 3c $o(n)$ for $c - \varepsilon \geqq {}^1/_2$ (with probability tending to 1), while it is easy to see, that the number of points of $\Gamma_{n,N_1(n)}$ belonging to components of size $\leq A$ and containing more than one cycle is also bounded with probability tending to 1.)

Our last statement follows by using the inequality (4.19) from the fact that the average number of components of the mentioned type is, as a simple calculation similar to those carried out in previous §§, shows, of order $O\left(\frac{1}{n}\right)$.

Let $E_n^{(1)}$ denote the event that

$$|H_{n,N_1(n)}(A) - nf(A, c - \varepsilon)| < \tau\, nf(A, c - \varepsilon) \tag{9.3}$$

where $\tau > 0$ is an arbitrary small positive number which will be chosen later and

$$f(A, c) = 1 - \frac{1}{2c} \sum_{k=1}^{A} \frac{k^{k-1}}{k!} (2\, ce^{-2c}) > 0 \tag{9.4}$$

and let $\overline{E_n^{(1)}}$ denote the contrary event. It follows from what has been said that

$$\lim_{n \to +\infty} \mathbf{P}(\overline{E_n^{(1)}}) = 0\,. \tag{9.5}$$

We consider only such $\Gamma_{n,N_1(n)}$ for which (9.3) holds.

Now clearly $\Gamma_{n,N_2(n)}$ is obtained from $\Gamma_{n,N_1(n)}$ by adding $N_2(n) - N_1(n) \sim n\varepsilon$ new edges at random to $\Gamma_{n,N_1(n)}$. The probability that such a new edge should connect two points belonging to $\mathcal{H}_{n,N_1(n)}(A)$, is at least $\dfrac{\binom{H_{n,N_1(n)}(A)}{2} - N_2(n)}{\binom{n}{2}}$,

and thus by (9.3) is not less than $(1 - 2\tau)\, f^2\,(A, c - \varepsilon)$, if n is sufficiently large and τ sufficiently small.

As these edges are chosen independently from each other, it follows by the law of large numbers that denoting by ν_n the number of those of the $N_2(n) - N_1(n)$ new edges which connect two points of $\mathcal{H}_{n,N_1(n)}$ and by $E_n^{(2)}$ the event that

$$\nu_n \geqq \varepsilon(1 - 3\tau) f^2(A, c - \varepsilon)\, n \tag{9.6}$$

and by $\overline{E}_n^{(2)}$ the contrary event, we have

$$\lim_{n \to +\infty} \mathbf{P}(\overline{E}_n^{(2)}) = 0\,. \tag{9.7}$$

We consider now only such $\Gamma_{n,N_2(n)}$ for which $E_n^{(2)}$ takes place. Now let us consider the subgraph $\Gamma^*_{n,N_2(n)}$ of $\Gamma_{n,N_2(n)}$ formed by the points of the set $\mathcal{H}_{n,N_1(n)}(A)$ and only of those edges of $\Gamma_{n,N_2(n)}$ which connect two such points.

We shall need now the following elementary

Lemma 2. *Let $a_1, a_2, \ldots, a_r$ be positive numbers, $\sum_{j=1}^{r} a_j = 1$. If $\max_{1 \leq j \leq r} a_j \leqq \alpha$ then there can be found a value k $(1 \leq k \leqq r - 1)$ such that*

$$\frac{1-\alpha}{2} \leqq \sum_{j=1}^{k} a_j \leqq \frac{1+\alpha}{2}$$

(9.8) *and*

$$\frac{1-\alpha}{2} \leqq \sum_{j=k+1}^{n} a_j \leqq \frac{1+\alpha}{2}\,.$$

Proof of Lemma 2. Put $S_j = \sum_{i=1}^{j} a_i$ $(j = 1, 2, \ldots, r)$. Let j_0 denote the least integer, for which $S_j > {}^1/_2$. In case $S_{j_0} - {}^1/_2 > {}^1/_2 - S_{j_0-1}$ choose $k = j_0 - 1$, while in case $S_{j_0} - {}^1/_2 \leqq {}^1/_2 - S_{j_0-1}$ choose $k = j_0$. In both cases we have $| S_k - {}^1/_2 | \leqq \frac{a_{j_0}}{2} \leqq \frac{\alpha}{2}$ which proves our Lemma.

Let the sizes of the components of $\Gamma^*_{n,N_2(n)}$ be denoted by $b_1, b_2, \ldots, b_r$. Let $E_n^{(3)}$ denote the event

$$\max b_j > H_{n,N_1(n)}(A)\,(1 - \delta) \tag{9.9}$$

and $\overline{E}_n^{(3)}$ the contrary event. Applying our Lemma with $\alpha = 1 - \delta$ to the numbers $a_j = \frac{b_j}{H_{n,N_1(n)}(A)}$ it follows that if the event $\overline{E}_n^{(3)}$ takes place, the set $\mathcal{H}_{n,N_1(n)}(A)$ can be split in two subsets $\mathcal{H}'_n$ and $\mathcal{H}''_n$ containing H'_n and H''_n points such that $H'_n + H''_n = H_{n,N_1(n)}(A)$ and

$$H_{n,N_1(n)}(A)\,\frac{\delta}{2} \leqq \min(H'_n, H''_n) \leqq \max(H'_n, H''_n) \leqq H_{n,N_1(n)}(A)\left(1 - \frac{\delta}{2}\right) \tag{9.10}$$

further no point of $\mathcal{H}'_n$ is connected with a point of $\mathcal{H}''_n$ in $\Gamma^*_{n,N_2(n)}$.

It follows that if a point P of the set $\mathcal{H}_{n,N_1(n)}(A)$ belongs to $\mathcal{H}'_n$ (resp. $\mathcal{H}''_n$) then all other points of the component of $\Gamma_{n,N_1(n)}$ to which P belongs are also contained in $\mathscr{H}'_n$ (resp. $\mathscr{H}''_n$). As the number of components of size $> A$ of $\Gamma_{n,N_1(n)}$ is clearly $< \frac{H_{n,N_1(n)}(A)}{A}$ the number of such divisions of the set $\mathscr{H}_{n,N_1(n)}(A)$ does not exceed $2^{\frac{1}{A} H_{n,N_1(n)}(A)}$.

If further $\bar{E}_n^{(3)}$ takes place then every one of the ν_n new edges connecting points of $\mathscr{H}_{n,N_1(n)}(A)$ connects either two points of $\mathscr{H}'_n$ or two points of $\mathscr{H}''_n$. The possible number of such choices of these edges is clearly

$$\binom{\binom{H'_n}{2}+\binom{H''_n}{2}}{\nu_n}.$$

As by (9.10)

$$(9.11)\qquad \frac{\binom{H'_n}{2}+\binom{H''_n}{2}}{\binom{H_n}{2}} \leq \frac{\delta^2}{4}+\left(1-\frac{\delta}{2}\right)^2 = 1-\delta+\frac{\delta^2}{2} \leq 1-\frac{\delta}{2}$$

it follows that

$$(9.12)\qquad \mathbf{P}(\bar{E}_n^{(3)}) \leq 2^{\frac{1}{A} H_{n,N_1(n)}(A)}\left(1-\frac{\delta}{2}\right)^{\varepsilon(1-3\tau)f^2(A,c-\varepsilon)n}$$

and thus by (9.3) and (9.6)

$$(9.13)\qquad \mathbf{P}(\bar{E}_n^{(3)}) \leq \exp\left[nf(A,c-\varepsilon)\left(\frac{(1+\tau)\log 2}{A}-\frac{\varepsilon(1-3\tau)f(A,c-\varepsilon)\delta}{2}\right)\right].$$

Thus if

$$(9.14)\qquad A\,\varepsilon\delta(1-3\tau)\,f(A,c-\varepsilon) > (1+\tau)\log 4$$

then

$$(9.15)\qquad \lim_{n\to+\infty} \mathbf{P}(\bar{E}_n^{(3)}) = 0\,.$$

As however in case $c-\varepsilon > {}^1/_2$ we have $f(A,c-\varepsilon) \geq G(c-\varepsilon) > 0$ for any A, while in case $c-\varepsilon = {}^1/_2$

$$(9.15a)\qquad f\left(A,\frac{1}{2}\right) = 1-\sum_{k=1}^{A}\frac{k^{k-1}}{k!\,e^k} = \sum_{k=A+1}^{\infty}\frac{k^{k-1}}{k!\,e^k} \geq \frac{1}{2\sqrt{A}} \quad \text{if } A \geq A_0$$

the inequality (9.13) will be satisfied provided that $\tau < \frac{1}{10}$ and $A > \frac{50}{\varepsilon^2\delta^2}$. Thus Theorem 9a is proved.

Clearly the "giant" component of $\Gamma_{n,N_2(n)}$ the existence of which (with probability tending to 1) has been now proved, contains more than

$$(1-\tau)(1-\delta)\,nf(A,c-\varepsilon)$$

points. By choosing ε, τ and δ sufficiently small and A sufficiently large, $(1 - \tau)(1 - \delta) f(A, c - \varepsilon)$ can be brought as near to $G(c)$ as we want. Thus we have incidentally proved also the following

Theorem 9b. *Let $\varrho_{n,N}$ denote the size of the greatest component of $\Gamma_{n,N}$. If $N(n) \sim cn$ where $c > 1/2$ we have for any $\eta > 0$*

$$\lim_{n\to+\infty} \mathbf{P}\left(\left|\frac{\varrho_{n,N(n)}}{n} - G(c)\right| < \eta\right) = 1 \tag{9.16}$$

where $G(c) = 1 - \dfrac{x(c)}{2c}$ and $x(c) = \sum_{k=1}^{\infty} \dfrac{k^{k-1}}{k!}(2c\, e^{-2c})^k$ *is the solution satisfying* $0 < x(c) < 1$ *of the equation* $x(c)\, e^{-x(c)} = 2ce^{-2c}$.

Remark. As $G(c) \to 1$ for $c \to +\infty$ it follows as a corollary from Theorem 9b that the size of the largest component will exceed $(1 - \alpha)n$ if c is sufficiently large where $\alpha > 0$ is arbitrarily small. This of course could be proved directly. As a matter of fact, if the greatest component of $\Gamma_{n,N(n)}$ with $N(n) \sim nc$ would not exceed $(1 - \alpha)n$ (we denote this event by $B_n(\alpha, c)$) one could by Lemma 2 divide the set V of the n points $P_1, \ldots, P_n$ in two subsets V' resp. V'' consisting of n' resp. n'' points so that no two points belonging to different subsets are connected and

$$\frac{\alpha n}{2} \leq \min(n', n'') \leq \max(n', n'') \leq \left(1 - \frac{\alpha}{2}\right) n. \tag{9.17}$$

But the number of such divisions does not exceed 2^n, and if the n points are divided in this way, the number of ways N edges can be chosen so that only points belonging to the same subset V' resp. V'' are connected, is

$$\binom{\binom{n'}{2} + \binom{n''}{2}}{N}.$$

As $\binom{n'}{2} + \binom{n''}{2} \leq \dfrac{n^2}{2}\left(1 - \dfrac{\alpha}{2}\right)$, it follows

$$\mathbf{P}(B_n(\alpha, c)) \leq 2^n \left(1 - \frac{\alpha}{2}\right)^{N(n)} \leq 2^n e^{-\frac{N(n)\alpha}{2}}. \tag{9.18}$$

Thus if $\alpha c > \log 4$, then

$$\lim_{n\to+\infty} \mathbf{P}(B_n(\alpha, c)) = 0 \tag{9.19}$$

which implies that for $c > \dfrac{\log 4}{\alpha}$ and $N(n) \sim cn$ we have

$$\lim_{n\to+\infty} \mathbf{P}(\varrho_{n,N(n)} \geqq (1 - \alpha)\, n) = 1. \tag{9.20}$$

We have seen that for $N(n) \sim cn$ with $c > {}^1/_2$ the random graph $\Gamma_{n,N(n)}$ consists with probability tending to 1, neglecting $o(n)$ points, only of isolated trees (there being approximately $\frac{n}{2c} \frac{k^{k-2}}{k!} (2c\, e^{-2c})^k$ trees of order k) and of a single giant component of size $\sim G(c)n$.

Clearly the isolated trees melt one after another into the giant component, the "danger" of being absorbed by the "giant" being greater for larger components. As shown by Theorem 2c for $N(n) \sim \frac{1}{2k} n \log n$ only isolated trees of order $\leq k$ survive, while for $\frac{N(n) - 1/2\, n \log n}{n} \to +\infty$ the whole graph will with probability tending to 1 be connected.

An interesting question is: what is the "life-time" distribution of an isolated tree of order k which is present for $N(n) \sim cn$? This question is answered by the following

Theorem 9c. *The probability that an isolated tree of order k which is present in $\Gamma_{n,N_1(n)}$ where $N_1(n) \sim cn$ and $c > {}^1/_2$ should still remain an isolated tree in $\Gamma_{n,N_2(n)}$ where $N_2(n) \sim (c + t)\, n$ $(t > 0)$ is approximately e^{-2kt}; thus the „life-time" of a tree of order k has approximately an exponential distribution with mean value $\frac{n}{2k}$ and is independent of the "age" of the tree.*

Proof. The probability that no point of the tree in question will be connected with any other point is

$$\prod_{j=N_1(n)+1}^{N_2(n)} \left(\frac{\binom{n-k}{2} - j + k}{\binom{n}{2} - j} \right) \sim e^{-2kt}.$$

This proves Theorem 9c.

§ 10. Remarks and some unsolved problems

We studied in detail the evolution of $\Gamma_{n,N}$ only till N reaches the order of magnitude $n \log n$. (Only Theorem 1 embraces some problems concerning the range $N(n) \sim n^\alpha$ with $1 < \alpha < 2$.) We want to deal with the structure of $\Gamma_{n,N(n)}$ for $N(n) \sim cn^\alpha$ with $\alpha > 1$ in greater detail in a fortcoming paper; here we make in this direction only a few remarks.

First it is easy to see that $\Gamma_{n,\binom{n}{2} - N(n)}$ is really nothing else, than the complementary graph of $\Gamma_{n,N(n)}$. Thus each of our results can be reformulated to give a result on the probable structure of $\Gamma_{n,N}$ with N being not much less than $\binom{n}{2}$. For instance, the structure of $\Gamma_{n.N}$ will have a second abrupt change when N passes the value $\binom{n}{2} - \frac{n}{2}$; if $N < \binom{n}{2} - cn$ with $c > {}^1/_2$ then the complementary graph of $\Gamma_{n,N}$ will contain a connected graph of order $f(c)n$, while for $c < {}^1/_2$ this (missing) "giant" will disappear.

To show a less obvious example of this principle of getting result for N near to $\binom{n}{2}$, let us consider the maximal number of pairwise independent points in $\Gamma_{n,N}$. (The vertices P and Q of the graph Γ are called *independent* if they are not connected by an edge).

Evidently if a set of k points is independent in $\Gamma_{n,N(n)}$ then the same points form a complete subgraph in the complementary graph $\overline{\Gamma}_{n,N(n)}$. As however $\overline{\Gamma}_{n,N(n)}$ has the same structure as $\Gamma_{n,\binom{n}{2}-N(n)}$ it follows by Theorem 1, that there will be in $\Gamma_{n,N(n)}$ almost surely no k independent points if $\binom{n}{2}-N(n)=o\left(n^{2\left(1-\frac{1}{k-1}\right)}\right)$ i. e. if $N(n)=\binom{n}{2}-o\left(n^{2\left(1-\frac{1}{k-1}\right)}\right)$ but there will be in $\Gamma_{n,N(n)}$ almost surely k independent points if $N(n)=\binom{n}{2}-\omega_n n^{2\left(1-\frac{1}{k-1}\right)}$ where ω_n tends arbitrarily slowly to $+\infty$. An other interesting question is: what can be said about the degrees of the vertices of $\Gamma_{n,N}$. We prove in this direction the following

Theorem 10. *Let $D_{n,N(n)}(P_k)$ denote the degree of the point P_k in $\Gamma_{n,N(n)}$ (i.e. the number of points of $\Gamma_{n,N(n)}$ which are connected with P_k by an edge). Put*

$$\underline{D}_n=\min_{1\leq k\leq n} D_{n,N(n)}(P_k) \quad \text{and} \quad \overline{D}_n=\max_{1\leq k\leq n} D_{n,N(n)}(P_k).$$

Suppose that

$$\lim_{n\to+\infty}\frac{N(n)}{n\log n}=+\infty. \tag{10.1}$$

Then we have for any $\varepsilon>o$

$$\lim_{n\to+\infty}\mathbf{P}\left(\left|\frac{\overline{D}_n}{\underline{D}_n}-1\right|<\varepsilon\right)=1. \tag{10.2}$$

We have further for $N(n)\sim cn$ for any k

$$\lim_{n\to+\infty}\mathbf{P}\left(D_{n,N(n)}(P_k)=j\right)=\frac{(2c)^j e^{-2c}}{j!} \qquad (j=0,1,\ldots). \tag{10.3}$$

Proof. The probability that a given vertex P_k shall be connected by exactly r others in $\Gamma_{n,N}$ is

$$\frac{\binom{n-1}{r}\binom{\binom{n-1}{2}}{N-r}}{\binom{\binom{n}{2}}{N}}\sim\frac{\left(\frac{2N}{n}\right)^r e^{-\frac{2N}{n}}}{r!}$$

thus if $N(n) \sim cn$ the degree of a given point has approximately a Poisson distribution with mean value $2c$. The number of points having the degree r is thus in this case approximately

$$n \frac{(2c)^r e^{-2c}}{r!} \qquad (r = 0, 1, \ldots).$$

If $N(n) = (n \log n)\,\omega_n$ with $\omega_n \to +\infty$ then the probability that the degree of a point will be outside the interval $\frac{2N(n)}{n}(1-\varepsilon)$ and $\frac{2N(n)}{n}(1+\varepsilon)$ is approximately

$$\sum_{|k - 2\log n \cdot \omega_n| > \varepsilon \cdot 2 \log n \cdot \omega_n} \frac{(2\omega_n \cdot \log n)^k e^{-2\omega_n \log n}}{k!} = O\left(\frac{1}{n^{\varepsilon^2 \omega_n}}\right)$$

and thus this probability is $o\left(\frac{1}{n}\right)$, for any $\varepsilon > 0$.

Thus the probability that the degrees of not all n points will be between the limit $(1 \pm \varepsilon)\, 2\omega_n \log n$ will be tending to 0. Thus the assertion of Theorem 10 follows.

An interesting question is: what will be the chromatic number of $\Gamma_{n,N}$? (The *chromatic number* $Ch(\Gamma)$ of a graph Γ is the least positive integer h such that the vertices of the graph can be coloured by h colours so that no two vertices which are connected by an edge should have the same colour.)

Clearly every tree can be coloured by 2 colours, and thus by Theorem 4a almost surely $Ch(\Gamma_{n,N}) = 2$ if $N = o(n)$. As however the chromatic number of a graph having an equal number of vertices and edges is equal to 2 or 3 according to whether the only cycle contained in such a graph is of even or odd order, it follows from Theorem 5e that almost surely $Ch(\Gamma_{n,N}) \leq 3$ for $N(n) \sim nc$ with $c < 1/2$.

For $N(n) \sim \frac{n}{2}$ we have almost surely $Ch(\Gamma_{n,N(n)}) \geq 3$.

As a matter of fact, in the same way, as we proved Theorem 5b, one can prove that $\Gamma_{n,N(n)}$ contains for $N(n) \sim \frac{n}{2}$ almost surely a cycle of odd order. It is an open problem how large $Ch(\Gamma_{n,N(n)})$ is for $N(n) \sim cn$ with $c > 1/2$?

A further result on the chromatic number can be deduced from our above remark on independent vertices. If a graph Γ has the chromatic number h, then its points can be divided into h classes, so that no two points of the same class are connected by an edge; as the largest class has at least $\frac{n}{h}$ points, it follows that if f is the maximal number of independent vertices of Γ we have $f \geq \frac{n}{h}$. Now we have seen that for $N(n) = \binom{n}{2} - o\left(n^{2\left(1-\frac{1}{k}\right)}\right)$ almost surely $f \leq k$; it follows that for $N(n) = \binom{n}{2} - o\left(n^{2\left(1-\frac{1}{k}\right)}\right)$ almost surely $Ch(\Gamma_{n,N(n)}) > \frac{n}{k}$.

Other open problems are the following: for what order of magnitude of $N(n)$ has $\Gamma_{n,N(n)}$ with probability tending to 1 a *Hamilton-line* (i.e. a path which passes through all vertices) resp. in case n is even *a factor of degree* 1 (i.e. a set of disjoint edges which contain all vertices).

An other interesting question is: what is the threshold for the appearance of a "topological complete graph of order k" i.e. of k points such that any two of them can be connected by a path and these paths do not intersect. For $k > 4$ we do not know the solution of this question. For $k = 4$ it follows from Theorem 8a that the threshold is $\frac{n}{2}$. It is interesting to compare this with an (unpublished) result of G. DIRAC according to which if $N \geqq 2n - 2$ then $G_{n,N}$ contains certainly a topological complete graph of order 4.

We hope to return to the above mentioned unsolved questions in an other paper.

Remark added on May 16, 1960. It should be mentioned that N. V. SMIRNOV (see e. g. *Математический Сборник* **6** (1939) p. 6) has proved a lemma which is similar to our Lemma 1.

(Received December 28, 1959.)

REFERENCES

[1] CAYLEY, A.: *Collected Mathematical Papers.* Cambridge, 1889—1897.
[2] PÓLYA, G.: "Kombinatorische Anzahlbestimmungen für Gruppen, Graphen und chemische Verbindungen". *Acta Mathematica* **68** (1937) 145—254.
[3] KÖNIG, D.: *Theorie der endlichen und unendlichen Graphen.* Leipzig, 1936.
[4] BERGE, C.: *Théorie des graphes et ses applications.* Paris, Dunod, 1958.
[5] FORD, G. W.—UHLENBECK, G. E.: "Combinatorial problems in the theory of graphs, I." *Proc. Nat. Acad. Sci.* **42** (1956) USA 122—128.
[6] GILBERT, E. N.: "Enumeration of labelled graphs". *Canadian Journal of Math.* **8** (1957) 405—411.
[7] ERDŐS, P.—RÉNYI, A.: "On random graphs, I". *Publicationes Mathematicae (Debrecen)* **6** (1959) 290—297.
[8] HARARY, F.: "Unsolved problems in the enumeration of graphs" In this issue, p. 63.
[9] ERDŐS, P.: „Graph theory and probability." *Canadian Journal of Math.* **11** (1959) 34—38.
[10] AUSTIN, T. L.—FAGEN, R. E.—PENNEY, W. F.—RIORDAN, J.: "The number of components in random linear graphs". *Annals of Math. Statistics* **30** (1959) 747—754.
[11] TURÁN P. : "Egy gráfelméleti szélsőértékfeladatról". *Matematikai és Fizikai Lapok* **48** (1941) 436—452.
[12] TURÁN, P.: "On the theory of graphs", *Colloquium Mathematicum* **3** (1954) 19—30.
[13] KŐVÁRY, T.—SÓS, V. T.—TURÁN, P.: "On a problem of K. Zarenkiewicz". *Colloquium Mathematicum* **3** (1954) 50—57.
[14] RÉNYI, A.: "Some remarks on the theory of trees". *Publications of the Math. Inst. of the Hung. Acad. of Sci.* **4** (1959) 73—85.
[15] RÉNYI, A.: "On connected graphs, I.". *Publications of the Math. Inst. of the Hung. Acad. of Sci.* **4** (1959) 385—387.
[16] JORDAN, CH.: *Calculus of finite differences.* Budapest, 1939.
[17] ERDŐS, P.—STONE, A. H.: "On the structure of linear graphs". *Bull. Amer. Math. Soc.* **52** (1946) 1087—1091.

О РАЗВЁРТЫВАНИЕ СЛУЧАЙНЫХ ГРАФОВ

P. ERDŐS и A. RÉNYI

Резюме

Пусть даны n точки $P_1, P_2, \ldots, P_n$, и выбираем случайно друг за другом N из возможных $\binom{n}{2}$ ребер (P_i, P_j) так что после того что выбраны k ребра каждый из других $\binom{n}{2} - k$ ребер имеет одинаковую вероятность быть выбранным как следующии. Работа занимается вероятной структурой так получаемого случайного графа $\Gamma_{n,N}$ при условии, что $N = N(n)$ известная функция от n и n очень большое число. Особенно исследуется изменение этой структуры если N нарастает при данном очень большом n. Случайно развёртывающий граф может быть рассмотрен как упрощенный модель роста реальных сетей (например сетей связы).