ON THE EVOLUTION OF RANDOM GRAPHS

by

P. ERDÖS and A. RÉNYI

Institute of Mathematics
Hungarian Academy of Sciences, Hungary

1. Definition of a random graph

Let $E_{n,N}$ denote the set of all graphs having n given labelled vertices $V_1, V_2, \cdots$, V_n and N edges. The graphs considered are supposed to be not oriented, without parallel edges and without slings (such graphs are sometimes called linear graphs). Thus a graph belonging to the set $E_{n,N}$ is obtained by choosing N out of the possible $\binom{n}{2}$ edges between the points $V_1, V_2, \cdots, V_n$, and therefore the number of elements of $E_{n,N}$ is equal to $\binom{\binom{n}{2}}{N}$. A random graph $\Gamma_{n,N}$ can be defined as an element of $E_{n,N}$ chosen at random, so that each of the elements of $E_{n,N}$ have the same probability to be chosen, namely $1\Big/\binom{\binom{n}{2}}{N}$. There is however an other slightly different point of view, which has some advantages. *We may consider the formation of a random graph as a stochastic process* defined as follows: At time $t=1$ we choose one out of the $\binom{n}{2}$ possible edges connecting the points $V_1, V_2, \cdots, V_n$, each of these edges having the same probability to be chosen; let this edge be denoted by e_1. At time $t=2$ we choose one of the possible $\binom{n}{2}-1$ edges, different from e_1, all these being equiprobable. Continuing this process at time $t=k+1$ we choose one of the $\binom{n}{2}-k$ possible edges different from the edges $e_1, e_2, \cdots, e_k$ already chosen, each of the remaining edges being equiprobable, i.e. having the probability $1\big/\left\{\binom{n}{2}-k\right\}$. We denote by $\Gamma_{n,N}$ the graph consisting of the vertices $V_1, V_2, \cdots$, V_n and the edges $e_1, e_2, \cdots, e_N$.

1) Other not equivalent but closely connected notions of random graphs are as follows: 1) We may define a random graph $\Gamma^*_{n,N}$ by dropping the restriction that there should be no parallel edges; thus we may suppose that e_{k+1} may be equal with probability $1\big/\binom{n}{2}$ with each of the $\binom{n}{2}$ edges, independently of whether they are contained in the sequence of edges $e_1, e_2, \cdots, e_k$ or not. These random graphs are considered in the paper [3]. 2) We may decide with respect to each of the $\binom{n}{2}$ edges, whether they should form part of the random graph considered or not, the probability of including a given edge being $p=N\big/\binom{n}{2}$ for each edge and the decisions concerning different edges being independent. We denote the random graph thus obtained by $\Gamma^{**}_{n,N}$. These random graphs have been considered in the paper [4].

The two definitions are clearly equivalent[1]. According to the second definition the number of edges of a random graph is interpreted as *time*, and according to this interpretation we may investigate the *evolution* of a random graph, i. e. the step-by-step unravelling of the structure of $\Gamma_{n,N}$ when N increases.

The evolution of random graphs may be considered as a (rather simplified) model of the evolution of certain real communication-nets, e. g. the railway-, road- or electric network system of a country or some other unit, or of the growth of structures of anorganic or organic matter, or even of the development of social relations. Of course, if one aims at describing such a real situation, our model of a random graph should be replaced by a more complicated but more realistic model. The following possible lines of generalization of the considered stochastic process of the formation of a random graph should be mentioned here:

a) One may distinguish different sorts of vertices, and/or edges—by a usual terminology one may consider *coloured* vertices resp. edges.

b) One may attribute different probabilities to the different edges; this can be done e.g. by attributing a weight, $W_e \geqq 0$ to each of the $\binom{n}{2}$ possible edges e so that $\sum W_e=1$ and to suppose that e_1 is equal to the edge e with probability W_e and that after $e_1, e_2, \cdots, e_k$ have been chosen, e_{k+1} is equal to any edge e not occurring among the edges $e_1, e_2, \cdots, e_k$ with probability $\frac{W_e}{S_k}$ where $S_k=\sum_{e \neq e_j\ (j=1,2,\cdots,k)} W_e$. An other alternative is to admit that the probability of choosing an edge $e=(V_i, V_j)$ after k other edges have already been chosen, should depend on the number of edges starting from the points V_i resp. V_j which have already been chosen.

In what follows we consider only the simple random graph-formation process, described above, i.e. we consider only the random graphs $\Gamma_{n,N}$.

Our main aim is to show through this special case that the evolution of a random graph shows very clear-cut features. The theorems we have proved belong to two classes. The theorems of the first class deal with the appearance of certain subgraphs (e.g. trees, cycles of a given order etc.) or components, or other local structural properties, and show that for many types of local structural properties A a definite "threshold" $A(n)$ can be given, so that if $\frac{N(n)}{A(n)} \to 0$ for $n \to +\infty$ then the probability that the random graph $\Gamma_{n,N(n)}$ has the structural property A tends to 0 for $n \to +\infty$, while for $\frac{N(n)}{A(n)} \to +\infty$ for $n \to +\infty$ the probability that $\Gamma_{n,N(n)}$ has the structural property A tends to 1 for $n \to +\infty$. In many cases still more can be said: there exists a "threshold function" for the property A, i. e. a probability distribution function $F_A(x)$ so that if $\lim_{n\to+\infty} \frac{N(n)}{A(n)}=x$ the probability that $\Gamma_{n,N(n)}$ has the property A tends to $F_A(x)$ for $n \to +\infty$.

The theorems of the second class are of similar type, only the properties A considered are not of a local character, but global properties of the graph $\Gamma_{n,N}$ (e. g. connectivity, total number of components, etc.).

In the next § we briefly describe the process of evolution of the random graph $\Gamma_{n,N}$. The proofs, which are completely elementary, and are based on the asymptotic evaluation of combinatorial formulae and on some well-known general methods of probability theory, are published in the papers [1] and [2].

2. *The evolution of* $\Gamma_{n,N}$

If n is a fixed large positive integer and N is increasing from 1 to $\binom{n}{2}$, the evolution of $\Gamma_{n,N}$ passes through five clearly distinguishable phases. These phases correspond to ranges of growth of the number N of edges, these ranges being defined in terms of the number n of vertices.

Phase 1. corresponds to the range $N(n) = o(n)$. For this phase it is characteristic that $\Gamma_{n,N(n)}$ consists almost surely (i. e. with probability tending to 1 for $n \to +\infty$) exclusively of components which are trees. Trees of order k appear only when $N(n)$ reaches the order of magnitude $n^{\frac{k-2}{k-1}}$ $(k=3, 4, \cdots)$. If $N(n) \sim \rho n^{\frac{k-2}{k-1}}$ with $\rho > 0$, then the probability distribution of the number of components of $\Gamma_{n,N(n)}$ which are trees of order k tends for $n \to +\infty$ to the Poisson distribution with mean value $\lambda = \frac{(2\rho)^{k-1} k^{k-2}}{k!}$. If $\frac{N(n)}{n^{\frac{k-2}{k-1}}} \to +\infty$ then the distribution of the number of components which are trees of order k is approximately normal with mean $M_n = n\frac{k^{k-2}}{k!}\left(\frac{2N(n)}{n}\right)^{k-1} e^{-\frac{2kN(n)}{n}}$ and with variance also equal to M_n. This result holds also in the next two ranges, in fact it holds under the single condition that $M_n \to +\infty$ for $n \to +\infty$.

Phase 2. corresponds to the range $N(n) \sim cn$ with $0 < c < 1/2$.

In this case $\Gamma_{n,N(n)}$ already contains cycles of any fixed order with probability tending to a positive limit: the distribution of the number of cycles of order k in $\Gamma_{n,N(n)}$ is approximately a Poisson-distribution with mean value $\frac{(2c)^k}{2k}$. In this range almost surely all components of $\Gamma_{n,N(n)}$ are either trees or components consisting of an equal number of edges and vertices, i. e. components containing exactly one cycle. The distribution of the number of components consisting of k vertices and k edges tends for $n \to +\infty$ to the Poisson distribution with mean value $\frac{(2ce^{-2c})^k}{k!}\left(1+k+\frac{k^2}{2!}+\cdots+\frac{k^{k-3}}{(k-3)!}\right)$. In this phase though not all, but still almost all (i. e. $n-o(n)$) vertices belong to components which are trees. The mean number of components is $n - N(n) + O(1)$, i. e. in this range by adding a new edge the number of components decreases by 1, except for a finite number of steps.

Phase 3. corresponds to the range $N(n) \sim cn$ with $c \geqq 1/2$. When $N(n)$ passes the threshold $n/2$, the structure of $\Gamma_{n,N(n)}$ changes abruptly. As a matter of fact this sudden change of the structure of $\Gamma_{n,N(n)}$ is the most surprising fact discovered by the investigation of the evolution of random graphs. While for $N(n) \sim cn$ with $c < 1/2$ the greatest component of $\Gamma_{n,N(n)}$ is a tree and has (with probability tending to 1 for $n \to +\infty$) approximately $\frac{1}{\alpha}\left(\log n - \frac{5}{2}\log\log n\right)$ vertices, where $\alpha = 2c - 1 - \log 2c$, for $N(n) \sim n/2$ the greatest component has (with probability tending to 1 for $n \to +\infty$) approximately $n^{2/3}$ vertices and has a rather complex structure. Moreover for $N(n) \sim cn$ with $c > 1/2$ the greatest component of $\Gamma_{n,N(n)}$ has (with probability tending to 1 for $n \to +\infty$) approximately $G(c)n$ vertices, where

(1) $$G(c)=1-\frac{1}{2c}\sum_{k=1}^{+\infty}\frac{k^{k-1}}{k!}\left(2ce^{-2c}\right)^k$$

(clearly $G(1/2)=0$ and $\lim_{c\to+\infty} G(c)=1$).

Except this "giant" component, the other components are all relatively small, most of them being trees, the total number of vertices belonging to components, which are trees being almost surely $n(1-G(c))+o(n)$ for $c \geqq 1/2$.

As regards the mean number of components[2], this is for $N(n) \sim cn$ with $c>1/2$ asymptotically equal to $\frac{n}{2c}\left(X(c)-\frac{X^2(c)}{2}\right)$, where

(2) $$X(c)=\sum_{k=1}^{+\infty}\frac{k^{k-1}}{k!}\left(2ce^{-2c}\right)^k=2c\left(1-G(c)\right)$$

The evolution of $\Gamma_{n,N(n)}$ in Phase 3. may be characterized by that the small components (most of which are trees) melt, each after another, into the giant component, the smaller components having the larger chance of "survival"; the survival time of a tree of order k which is present in $\Gamma_{n,N(n)}$ with $N(n) \sim cn$, $c>1/2$ is approximately exponentially distributed with mean value $n/2k$.

Phase 4. corresponds to the range $N(n) \sim c\, n \log n$ with $c \leqq 1/2$. In this phase the graph almost surely becomes connected. If

(3) $$N(n)=\frac{n}{2k}\log n+\frac{k-1}{2k}n\log\log n+yn+o(n)$$

then there are with probability tending to 1 for $n\to+\infty$ only trees of order $\leqq k$ outside the giant component, the distribution of the number of trees of order k having in the limit again a Poisson distribution with mean value $\frac{e^{-2ky}}{k\cdot k!}$.
Thus for $k=1$, i. e. for

(4) $$N(n)=\frac{n}{2}\log n+yn+o(n)$$

$\Gamma_{n,N(n)}$ consists, with probability tending to 1 for $n\to+\infty$, only of a connected component containing $n-O(1)$ points and a few isolated points, the distribution of the number of these being approximately a Poisson distribution with mean value e^{-2y}. Thus in case (4) the probability that the whole graph $\Gamma_{n,N(n)}$ is connected tends to $e^{-e^{-2y}}$ for $n\to+\infty$ and thus this probability approaches 1 as y increases.

This last result has been obtained by us already in 1958 (see [2]). The probability of $\Gamma^{**}_{n,N}$ being connected has been investigated by E. N. *Gilbert* (see [4]).

It should be mentioned that the investigation of $\Gamma^{**}_{n,N}$ can be reduced to that of $\Gamma_{n,N}$ as follows[3]: $\Gamma^{**}_{n,N}$ can be obtained by first choosing the value k of a random variable $\mathfrak{H}$ having the binomial distribution $P(\mathfrak{H}=k)=\binom{\binom{n}{2}}{k}p^k(1-p)^{\binom{n}{2}-k}$ where

2) The mean number of components of $\Gamma^{*}_{n,N}$ has been investigated in [3]. Our results for $\Gamma_{n,N}$ are however more far reaching.

3) This idea has been used by J. *Hájek* [5] in the theory of sampling from a finite population who has shown in this way that the Lindeberg-type conditions given by us [6] for the validity of the central limit theorem for samples from a finite population are not only sufficient but also necessary.

$p=N/\binom{n}{2}$ and then choosing $\Gamma_{n,k}$. In this way one can show that the threshold and the threshold function for connectivity of $\Gamma^{**}_{n,N}$ are the same as that of $\Gamma_{n,N}$. It should be mentioned that this does not follow from the inequalities given by *Gilbert* [4].

Phase 5. consists of the range $N(n)\sim(n\log n)\,w(n)$ where $w(n)\to+\infty$. In this range the whole graph is not only almost surely connected, but the orders of all points are almost surely asymptotically equal. Thus the graph becomes in this phase "asymptotically regular".

References

[1] P. Erdös-A.Rényi: On the evolution of random graphs, *Publications of the Mathematical Institute of the Hungarian Academy of Sciences* 5 (1960) (in print).

[2] P. Erdös-A.Rényi: On random graphs, *Publicationes Mathematicae*, 6 (1959) 290–297.

[3] T. L. Austin, R. E. Fagen, W. F. Penney and J. Riordan: The number of components in random linear graphs, *Annals of Mathematical Statistics*, 30 (1959) 747–754.

[4] E. N. Gilbert: Random graphs, *Annals of Mathematical Statistics*, 30 (1959), 1141–1144.

[5] J. Hájek: Limiting distributions in simple random sampling from a finite population, *Publications of the Mathematical Institute of the Hungarian Academy of Sciences* 5 (1960) (in print).

[6] P. Erdös and A. Rényi: On the central limit theorem for samples from a finite population, *Publications of the Mathematical Institute of the Hungarian Academy of Sciences*, 4 (1959) 49–61.

RÉSUMÉ

Soit $E_{n,N}$ l'ensemble de tous les graphes possédants n sommets donnés et ayant N arcs. Nous considérons seulement des graphes non-orientés et sans boucles. Un graph aléatoir $\Gamma_{n,N}$ est défini comme un élément de l'ensemble $E_{n,N}$ choisi au hasard tel que tous les éléments de $E_{n,N}$ ont la même probabilité d'être choisis.

Les auteurs considerent les propriétés probables de $\Gamma_{n,N}$ quand n et N tends vers l'infini d'un tel façon que $N=N(n)$ est une fonction donnée de n.