J. Soc. Indust. Appl. Math.
Vol. 9, No. 4, December, 1961
Printed in U.S.A.

RANDOM PLANE NETWORKS*

E. N. GILBERT†

1. Introduction. Recently random graphs have been studied as models of communications networks. Points (vertices) of a graph represent stations; lines of a graph represent two-way channels. In the literature [1, 3, 5] each pair of stations has some probability (the same for all pairs of stations regardless of their separation) of being joined by a channel. Such a model cannot represent accurately a network of shortrange stations spread over a wide area. The random plane networks of this paper provide a simple model in which the range of the stations is a parameter.

To construct a *random plane network*, first pick points from the infinite plane by a Poisson process with density D points per unit area. Next join each pair of points by a line if the pair is separated by distance less than R.

The random plane network represents an infinite communication network of stations with range R. Another application is to the study of spread of a contagious disease. Points represent individuals susceptible to the disease. If sick individuals infect all others within distance R, the disease spreads along lines of the network.

The random plane network falls into disjoint connected components. In the first application a station can relay messages to any other station in the same component. In the second application an entire component becomes infected when one of its members acquires the disease. The function of main interest here is $P(N)$, the probability that a point belongs to a component containing at least $N - 1$ additional points.

Fig. 1 shows two networks, obtained from the same random point pattern but with different values of R. The parameter E in Fig. 1 is the expected number of points in a circle of radius R; $E = \pi R^2 D$. Not all lines of the network are drawn, just enough to show the connected components. Naturally, the components for $E = 4.7$ are bigger than those for $E = 1.8$ but there is even a qualitative difference. The small components of the $E = 1.8$ network are so merged together in the $E = 4.7$ network that most of the points belong to a single large component. This phenomenon suggests that there is a component with infinitely many points in the infinite plane. Then the limit

$$\lim_{N\to\infty} P(N) = P(\infty)$$

need not be zero. It represents the probability of belonging to an infinite component.

* Received by the editor December 14, 1960 and in revised form March 6, 1961.
† Bell Telephone Laboratories, Inc., Murray Hill, New Jersey.

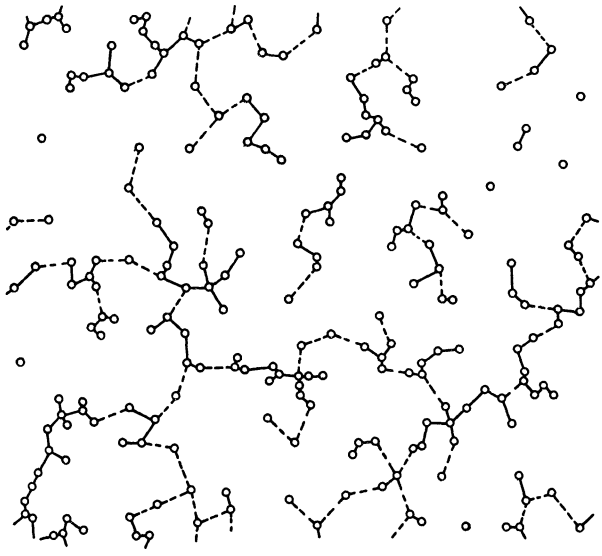

FIG. 1. *Random plane network.*

Section 2 shows that $P(\infty) = 0$ when $E < 1.75$. Section 3 shows $P(\infty) > 0$ when E is sufficiently large. Thus there exists a critical value E_c which is the largest value of E for which $P(\infty) = 0$. In the communication model, the network provides only local communication if $E \leqq E_c$ and provides some long-distance communication if $E > E_c$. In the disease model, a widespread epidemic can occur only if $E > E_c$.

The bounds on E_c, which §2 and §3 provide, are far apart. Section 5 describes a computer simulation of random plane networks. The simulation gives an estimate $E_c = 3.2$ and also gives curves of $P(N)$ and $P(\infty)$.

2. A branching process. The component containing a given point P might have been constructed step-by-step as follows. First pick the points which are connected directly to P. These are found by a Poisson process of density D in the circle $A(1)$ of radius R centered at P. These points, if any, will be said to belong to the *first generation* of descendants of P. Continuing, the *second generation* points are to be chosen within distance R of first generation points but more than distance R away from P. In general, in choosing generation $n + 1$, draw the set $A(n + 1)$ of all points lying within distance R of a point of generation n and not in $A(1), A(2), \cdots, A(n)$. Then generation $n + 1$ is obtained by a Poisson process of density D in the set

$A(n+1)$. If at some generation the Poisson process produces no points, then the construction stops and the component is finite.

A related but simpler construction, which will be called construction B, proceeds as follows. Construct the first generation as before. Suppose the nth generation has been found and call its points $q_1, q_2, \cdots, q_k$. Construct circles $C_1, C_2, \cdots, C_k$ of radius R and centered at $q_1, q_2, \cdots, q_k$. Pick points independently from $C_1, C_2, \cdots, C_k$ by Poisson processes of density D. These points form generation $n+1$.

Clearly the probability of obtaining N or more points by process B is greater than $P(N)$. In fact, the component in the random network might have been constructed by process B, deleting at each step any new point which belongs to a circle C_i already considered (in the same or earlier generations).

Construction B is a simple branching process (see [4], Ch. 12) of the sort which occurs in the study of survival of family names. Here the probability p_k that a point produces k descendants in the next generation is

$$p_k = \frac{E^k e^{-E}}{k!} \qquad (k = 0, 1, 2, \cdots).$$

The probability s that B produces infinitely many points may be computed as an upper bound on $P(\infty)$. Since $s = 0$ for $E \leqq 1$, we have $E_c \geq 1$. If $E > 1$, s is the smallest real root of

$$s = 1 - e^{-Es}. \tag{1}$$

(see [4], p. 275).

A simple refinement B^* of construction B replaces the circles C_i of radius R by certain lune-shaped figures. Note that circles C_i, C_i' surrounding a point q_i and its descendant q_i' in the next generation always intersect in an area at least $(2\pi/3 - \sqrt{3}/2)R^2$. This area would be forbidden to descendants of q_i' in the random plane network. The remaining part of C_i' is a lune of area $(\pi/3 + \sqrt{3}/2)R^2$. In B^*, the descendants of q_i' are picked by a Poisson process of density D in this lune. Using B^* the bound on E_c will be improved to

$$E_c \geqq \frac{1}{\frac{1}{3} + \frac{\sqrt{3}}{2\pi}} = 1.75. \tag{2}$$

In construction B^*, the first generation again has a number K of points distributed according to the Poisson Law with mean E. Each of these K points initiates a simple branching process in which numbers of offspring

have the Poisson distribution, now with mean

$$E^* = \left(\frac{\pi}{3} + \frac{\sqrt{3}}{2}\right) R^2 D = \left(\frac{1}{3} + \frac{\sqrt{3}}{2\pi}\right) E.$$

As in (1), each of the K branching processes has a probability s^* of continuing forever, where now

$$s^* = 0 \qquad \text{(if } E^* \leqq 1\text{)},$$

$$s^* = 1 - e^{-E^* s^*} \qquad \text{(if } E^* > 1\text{)}.$$

The probability S that at least one of the branching processes continues forever is

$$S = 1 - \sum_{K=0}^{\infty} \frac{(E^*)^K e^{-E^*}}{K!} (1 - s^*)^K = 1 - e^{-E^* s^*}. \tag{3}$$

Since $P(\infty) \leqq S$, (3) shows that $P(\infty) = 0$ when $s^* = 0$, i.e., when $E^* \leqq 1$. Then (2) follows.

The results derived from construction B apply to a somewhat more general kind of random plane network. Instead of connecting points (x, y) and (x', y') if and only if they are closer than R apart, a line may be drawn by a random choice. The probability of drawing a line may be an arbitrary function $f(|x - x'|, |y - y'|)$ of the absolute coordinate differences of the end points (x, y), (x', y'). However, the mean number E of lines to a point,

$$E = D \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(|x|, |y|)\, dx\, dy = 4D \int_0^{\infty} \int_0^{\infty} f(x, y)\, dx\, dy,$$

is assumed to exist. Again $E_c \geqq 1$ follows. Moreover, the number of lines at a point still has a Poisson distribution and s, given by (1), remains an upper bound on $P(\infty)$.

3. A percolation process. Percolation processes [2, 6, 7, 8, 9] produce random networks as models of porous media through which a fluid seeps. The percolation process produces a network G by deleting lines at random from a fixed underlying graph G'. All points of G' remain as points of G. The remaining lines (those of G) represent tunnels through which fluid may seep. In a typical special case the underlying graph G' might be one of the regular arrangements of points and lines which occur in crystallography. In any case, lines of G' are deleted independently and all with the same probability q.

The authors cited find, for certain graphs G', that the fluid can percolate infinitely far if q is below a critical value. Thus the random plane network problem strongly resembles a percolation problem. Unfortunately the two

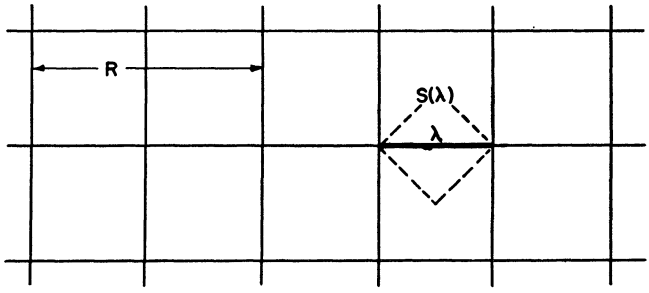

FIG. 2.

are enough different that the theorems of percolation theory do not apply directly here. Suppose one seeks a percolation process which will produce random plane networks. Since any point of the plane may become a station in the network, G' should contain all points of the plane and all lines of length R or less. The present theorems of percolation theory do not allow this; however, one might consider the limiting case as $M \to \infty$ of a percolation on a graph G' with M points per unit area. In order to keep the number of lines per unit area of G fixed at the correct value $\rho E/2$, one must set $q = 1 - O(M^{-2})$. In the limit as $M \to \infty$, not only do almost all points of G have no line but, further, almost all points which have a line have no other line.

A correspondence between random networks and a certain percolation process can, however, supply bounds on E_c. The percolation process in question has a square lattice as its underlying graph. This square lattice, shown in Fig. 2, has all lines of length $L = R/2$. Imagine this lattice to be drawn on the same plane as the random network. Surround each line λ of the square lattice by a square $S(\lambda)$ (shown in dotted lines) having λ as diagonal. In the percolation process, λ is deleted if and only if the square $S(\lambda)$ contains no point of the random network. Thus each line of the lattice has probability

$$p = 1 - e^{-DL^2/2}$$

of remaining. If p is larger than some critical probability p_c, this percolation process has probability 1 of producing an infinite component. Hammersley [8] shows $.35 < p_c < .65$. Harris [9] improves the lower bound to $.50 \leqq p_c$. Let P_1 be a point of the random network. P_1 belongs to some dotted square, say $S(\lambda_1)$. Let $F(p)$ be the probability that λ_1 belongs to an infinite component of the percolation graph. If λ_1 belongs to an infinite component, one can find an infinite path $\lambda_1, \lambda_2, \cdots$, in which (for $i = 1, 2, \cdots$) λ_i and λ_{i+1} have a common end point. Then our random network contains points $P_2, P_3, \cdots$ in $S(\lambda_2), S(\lambda_3), \cdots$, and P_i and P_{i+1} are closer to-

gether than $2L = R$. Thus we conclude

$$P(\infty) \geqq F(p) = F(1 - e^{-E/8\pi}). \tag{4}$$

Now (4) provides $E_c \leqq 8\pi \log_e(1/p_c)$ but this is only a very weak upper bound on E_c. For example, even using the conjectured exact value $p_c = 0.50$ of the critical probability in the percolation process, (4) only provides

$$E_c \leqq 8\pi \log_e 2 = 17.4.$$

4. Hex. The following argument suggests a better upper bound for E_c. The argument depends on well-known properties of the game of hex. Hex is played on a honeycomb pattern of hexagonal cells. A typical cell $H(u, v)$ has its center at the point with Cartesian coordinates

$$x = \left(u + \frac{v}{2}\right) b, \qquad y = vb \frac{\sqrt{3}}{2},$$

where b is the separation between cell centers and u and v assume integer values. A hexboard consists of all cells $H(u, v)$ such that $|u| \leqq K, |v| \leqq K$ for some integer K (see Fig. 3). One player, say White, tries to acquire enough cells to have a connected path from $u = -K$ to $u = K$. His opponent, Black, needs a connected path from $v = -K$ to $v = K$. No draw is possible in hex; indeed every assignment of cells to White and Black finds

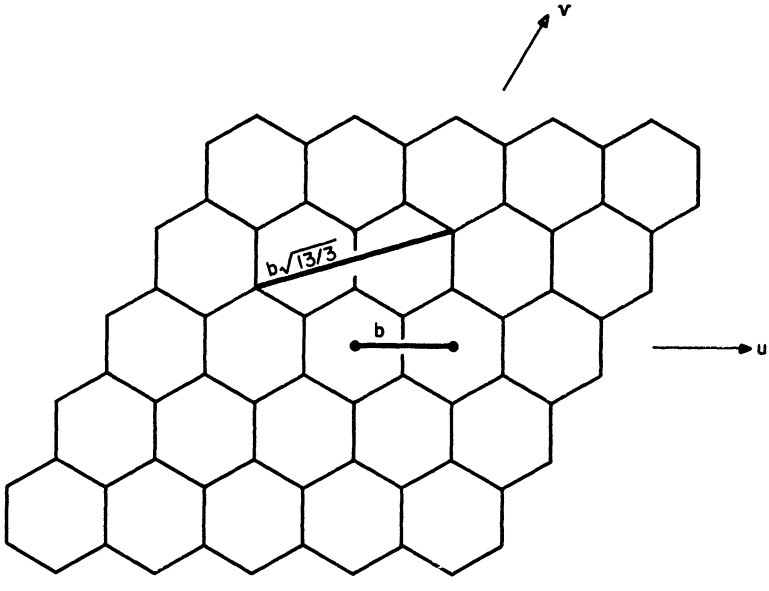

FIG. 3. *Hexboard* ($K = 2$).

exactly one of the players with a winning path. A proof of this property of hex will appear in a forthcoming book on communications by J. R. Pierce.

Imagine a hexboard drawn on the plane of the random plane network. Give White cell $H(u, v)$ if and only if it contains a point of the network. The probability that Black receives a particular cell is

$$Q = e^{-Db^2\sqrt{3}/2}.$$

Suppose Db^2 is chosen so as to make $Q = .50$, i.e., $Db^2 = (2 \log_e 2)/\sqrt{3}$. Then the game is symmetrical and Black and White have each probability $\frac{1}{2}$ of winning; there is probability $\frac{1}{2}$ of a connected path of hexagons from $u = -K$ to $u = K$. If in addition, one picks $b = R\sqrt{3/13}$, then points in neighboring hexagons are within distance R. A path by which White wins now provides a component of the random network containing at least $2K + 1$ points. The final result is:

THEOREM. *Let* $E = (26\pi/3\sqrt{3}) \log_e 2 = 10.9$, *let* K *be an integer arbitrarily large, and let* H_K *be a hexboard of* $(2K + 1)^2$ *cells with separation* $b = R\sqrt{3/13}$ *between cell centers. There is probability at least* $\frac{1}{2}$ *that a component of the random network joins a pair of opposite sides of the hexboard and contains at least* $2K + 1$ *points.*

Since K can be arbitrarily large, the theorem shows that arbitrarily

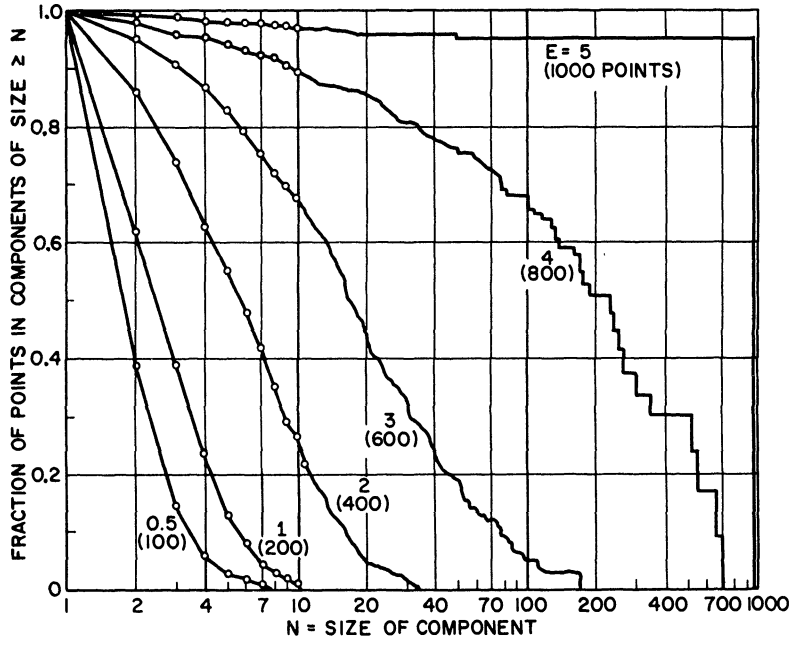

FIG. 4. *Estimates of* $P(N)$ *from* 11 *networks.*

large components are easy to find when $E \geqq 10.9$. Unfortunately $E_c \leqq 10.9$ does not follow. One can imagine that each hexboard H_K contains a $(2K + 1)$-point component and yet no component of the random network is infinite. For example, the components of the network might be arranged as a family of concentric polygons, becoming arbitrarily large but each finite. Such arrangements seem very unlikely, and it would be interesting to have a way of deriving $E_c \leqq 10.9$ from the theorem.

5. An experiment. The theoretical results of §2 and §3 are too weak to estimate E_c, $P(N)$, or $P(\infty)$ accurately. A computer experiment provided better numerical results. An IBM 7090 computer simulated the construction of random plane networks in a large square. For each network the computer printed a table of the fraction of points of the network which belonged to components of size N or more. This fraction is an estimate of $P(N)$.

Each curve in Fig. 4 uses results obtained from 11 independent networks with the stated value of E. The fractions of points belonging to com-

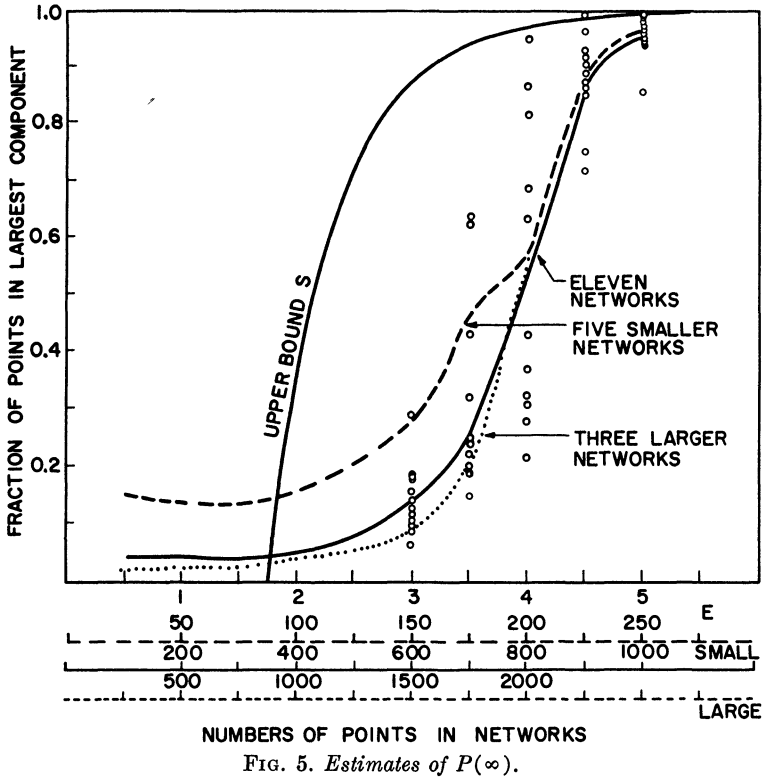

FIG. 5. *Estimates of* $P(\infty)$.

ponents of size N or more were averaged over the 11 networks and plotted as an estimate of $P(N)$. Since the networks had 1000 points each at $E = 5$ and were proportionally smaller for $E < 5$, Fig. 4 does not estimate $P(N)$ reliably when N is large.

In a random plane network, a point has probability $P(\infty)$ of belonging to a component of maximum size. This follows immediately when $E_c < E$ since ∞ is certainly a maximum size; also, when $E \leqq E_c$, an infinite component has probability zero and yet arbitrarily large finite component sizes have nonzero probabilities. Thus, in the computer simulation the fraction of points belonging to the component of largest size was used as an estimate of $P(\infty)$. For each value of E, these estimates from 11 networks were averaged together to give a point on the solid curve shown in Fig. 5. Since the computer only produced finite square approximations to random plane networks, the values plotted in Fig. 5 do not become zero for small E. In an attempt to obtain improved results, some networks were constructed in a square of 2.5 times larger area. There were three of the larger networks for each value of E; they produced the dotted curve in Fig. 5. Similarly a dashed curve shows estimates of $P(\infty)$ from networks (five for each value of E) of area smaller by a factor .25. In Fig. 5 networks were available at $E = .5, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5$, and 5.0. The corresponding numbers of points in the networks of the three sizes are shown on three additional scales on Fig. 5. The upper bound S on $P(\infty)$ was computed from (3) and appears in Fig. 5 for comparison.

In Fig. 5 the estimates of $P(\infty)$ are so large for small E that it is difficult to estimate E_c accurately. If the true $P(\infty)$ vs. E curve hits the E axis with nonzero slope (as for the S vs. E curve of the branching process), then Fig. 5 suggests a value of E_c somewhere near 3.2. However, the author has no proof that the true process resembles a branching process to that extent.

In the range $3 \leqq E \leqq 5$, the estimates of $P(\infty)$ from the 11 individual networks are widely scattered. For each E in this range the 11 estimates appear on Fig. 5 as isolated points. Points for $E \leqq 3$ are not shown because they clustered well about the mean line. Increasing the size of the square would decrease the scatter. However, for many communication network applications, the actual number of stations is less than 1000 and such scatter may be of interest.

The following calculations give rough indications of the sizes of the experimental errors for the curves in Figs. 4 and 5.

In Fig. 5 the 11 points for $E = 4$ spread out with a standard deviation of .26. Take .26 as an estimate of the true standard deviation of the fraction f of points in the largest component when 800 points are placed in a square and when R is chosen to make $E = 4$. The curve plotted shows the average of 11 independent measurements of f. This average is a random variable

with standard deviation $.26/\sqrt{11} = .077$. In a similar way standard deviations for other points on the $P(\infty)$ curve and for the $P(N)$ curves of Fig. 4 were found. These are tabulated below. Of course, these figures only represent errors in measurement for finite squares and tell nothing about how well these squares approximate the infinite plane.

E	$P(3)$	$P(10)$	$P(30)$	$P(100)$	$P(\infty)$
1	.008				
2	.014	.021			.004
3	.007	.010	.041		.018
4	.003	.003	.022	.043	.077
5					.011

Since networks for a range of values of E were needed, much computing time was saved by using old networks as parts of new networks with higher E. In a typical run of the computer the range R and the area A of the square were fixed parameters. Starting with a network of just one point the computer added points one at a time to obtain networks with increasing values of E. Each new point was selected by generating a pair of pseudo-random numbers for coordinates. At suitable intervals the computer printed out the required data. Thus the 110 networks used in plotting the solid curve of Fig. 5 came from 11 runs, each of which constructed networks at $E = .5$, $1.0, \cdots$, and 5.0. This construction gave each network exactly $EA/\pi R^2$ points instead of a true Poisson distribution of points.

Points near the edges of a square tend to belong to small components because connection paths which go outside the square are missing. To offset these edge effects the computation identified opposite sides of the square, making the square topologically equivalent to a torus. Thus in Fig. 1 with $E = 4.7$ small clusters of points at the top of the square are connected to the large component at the bottom of the square.

In programming this experiment it was important to avoid using a storage unit for each pair of points to remember whether or not the pair was joined by a line. It was also desired to compute the distance between each pair of points at most once. Both objectives were achieved by storing the components in the computer as lists. To do this the computer stored a single function $I_{\text{next}}(J)$, the name of the point (if any) which followed point J in its list. Other functions which were used were the coordinates $X(J)$, $Y(J)$ of point J, the name $K(J)$ of the list (component) containing point J, the number $N(K)$ of points in list K, the first and last points in list K, and the number $M(N)$ of lists of length N points.

In order to add point I to the network the computer began by putting I temporarily in a list by itself. It then began computing distances from I to $1, 2, \cdots, I - 1$. Each time a point J was found, within distance R of

point I but not already in list $K(I)$, the lists $K(I)$ and $K(J)$ were combined into a single list. To combine lists, the computer had only to change some of the values of the eight functions mentioned above.

In simulating a network of M points the main storage requirement was for the $8M$ values of the eight functions. The IBM 7090 has adequate storage to handle 3000 points. Each of the eleven 1000 point cases took two minutes on the IBM 7090. The 2000 point cases took seven minutes each.

REFERENCES

1. T. L. Austin, R. E. Fagen, W. F. Penney and John Riordan, *The number of components in random linear graphs*, Ann. Math. Statist., 30 (1959), pp. 747–754.
2. S. R. Broadbent and J. M. Hammersley, *Percolation processes, crystals and mazes*, Proc. Cambridge Philos. Soc., 53 (1957), pp. 629–641.
3. P. Erdös and A. Rényi, *On the evolution of random graphs*, A. Magyar Tudományos Akadémia Matematikai Kutato Intézétenek Közlemenyei (Publ. of the Math. Inst. of Hungarian Acad. of Sci.), Vol. V, Series A, Fasc. 1–2 (1960), pp. 17–61.
4. W. Feller, *An Introduction to Probability Theory and Its Applications*, Wiley, New York, 1957.
5. E. N. Gilbert, *Random graphs*, Ann. Math. Statist., 30 (1959), pp. 1141–1144.
6. J. M. Hammersley, *Percolation processes, the connective constant*, Proc. Cambridge Philos. Soc., 53 (1957), pp. 642–645.
7. ———, *Percolation processes, lower bounds for the critical probability*, Ann. Math. Statist., 28 (1957), pp. 790–795.
8. ———, *Bornes supérieures de la probabilité critique dans un processus de filtration*, Le Calcul des Probabilités et ses Applications. Paris, 15–20 juillet 1958, pp. 17–37. Colloques Internationaux du Centre National de la Recherche Scientifique, LXXXVII. Centre National de la Recherche Scientifique, Paris, 1959.
9. T. E. Harris, *A lower bound for the critical probability in a certain percolation process*, Proc. Cambridge Philos. Soc., 56 (1960), pp. 13–20.