RANDOM GRAPHS

By E. N. Gilbert

Bell Telephone Laboratories, Inc., Murray Hill, New Jersey

1. Introduction. Let N points, numbered $1, 2, \cdots, N$, be given. There are $N(N-1)/2$ lines which can be drawn joining pairs of these points. Choosing a subset of these lines to draw, one obtains a graph; there are $2^{N(N-1)/2}$ possible graphs in total. Pick one of these graphs by the following random process. For all pairs of points make random choices, independent of each other, whether or not to join the points of the pair by a line. Let the common probability of joining be p. Equivalently, one may erase lines, with common probability $q = 1 - p$ from the complete graph.

In the random graph so constructed one says that *point i is connected to point j* if some of the lines of the graph form a path from i to j. If i is connected to j for every pair i, j, then the graph is said to be *connected*. The probability P_N that the graph is connected, and also the probability R_N that two specific points, say 1 and 2, are connected, will both be found.

As an application, imagine the N points to be N telephone central offices and suppose that each pair of offices has the same probability p that there is an idle direct line between them. Suppose further that a new call between two offices can be routed via other offices if necessary. Then R_N is the probability that there is some way of routing a new call from office 1 to office 2 and P_N is the probability that each office can call every other office.

Exact expressions for P_N and R_N are given in Section 2. These results are unwieldy for large N. Bounds on P_N and R_N derived in Section 3 show that

$$P_N \sim 1 - Nq^{N-1} \tag{1}$$

and

$$R_N \sim 1 - 2q^{N-1} \tag{2}$$

asymptotically as $N \to \infty$.

Other related results appear in a paper by Austin, Fagen, Penney, and Riordan [1]. These authors use a different random process to pick a graph and they find a generating function for the distribution of the number of connected pieces in the random graph.

2. Exact results. P_N may be expressed in terms of the number $C_{N,L}$ of connected graphs having N labeled points and L lines. Since each such graph has probability $p^L q^{-L+N(N-1)/2}$ of being the chosen graph, it follows that

$$P_N = \sum_L C_{N,L} p^L q^{-L+N(N-1)/2}.$$

Received January 26, 1959; revised July 29, 1959.

In this formula the range of summation is $N - 1 \leqq L \leqq N(N-1)/2$. In [3] and [4] a generating series for $C_{N,L}$ was given in the form

$$\sum_{N,L}^{\infty} C_{N,L} \frac{x^N y^L}{N!} = \log\left(1 + \sum_{i=1}^{\infty} \frac{x^i(1+y)^{i(i-1)/2}}{i!}\right).$$

This result is easily converted into a generating series for P_N, viz.,

$$\sum_{N=1}^{\infty} P_N \frac{x^N q^{-N(N-1)/2}}{N!} = \log\left(1 + \sum_{i=1}^{\infty} \frac{x^i q^{-i(i-1)/2}}{i!}\right). \tag{3}$$

It may be noted that, when $0 \leqq q < 1$ and $x \neq 0$, neither series in (3) converges. The equality in (3) merely signifies that P_N may be found by formally expanding the logarithm into a power series and collecting coefficients of x^N. One can perform the expansion analytically to obtain an explicit formula

$$P_N = \sum_{r_1,\cdots,r_N} \frac{(-1)^{n-1}(n-1)!N!q^{(N^2-1^2r_1-\cdots-N^2r_N)/2}}{r_1!\cdots r_N!(1!)^{r_1}\cdots(N!)^{r_N}}.$$

The sum is extended over all non-negative integer solutions of $r_1 + 2r_2 + \cdots + Nr_N = N$ (i.e. over all partitions of N). The letter n in the sum is $n = r_1 + \cdots + r_N$.

The first few instances of this formula are

$$P_1 = 1$$
$$P_2 = 1 - q$$
$$P_3 = 1 - 3q^2 + 2q^3$$
$$P_4 = 1 - 4q^3 - 3q^4 + 12q^5 - 6q^6$$
$$P_5 = 1 - 5q^4 - 10q^6 + 20q^7 + 30q^8 - 60q^9 + 24q^{10}$$
$$P_6 = 1 - 6q^5 - 15q^8 + 20q^9 + 120q^{11} - 90q^{12} - 270q^{13} + 360q^{14} - 120q^{15}.$$

For larger values of N the number of terms in the formula for P_N increases rapidly. P_N may then be computed more easily by means of the recurrence relation

$$1 - P_N = \sum_{K=1}^{N-1} \binom{N-1}{k-1} P_k q^{k(N-k)}. \tag{4}$$

The kth term of (4) is the probability that point 1 is connected to exactly $k - 1$ of the $N - 1$ other points. Then (4) follows by noting that point 1 is connected to 0, 1, $\cdots$, or $N - 1$ other points with probability 1.

The argument which was used to derive (4) may be modified to give the following formula for R_N:

$$1 - R_N = \sum_{k=1}^{N-1} \binom{N-2}{k-1} P_k q^{k(N-k)}. \tag{5}$$

TABLE 1

$q =$	.1	.3	.5	.7	.9
P_2	.90000	.70000	.50000	.30000	.10000
P_3	.97200	.78400	.50000	.21600	.02800
P_4	.99581	.89249	.59375	.21865	.01293
P_5	.99949	.95751	.71094	.25626	.00810
P_6	.99994	.98497	.81569	.31690	.00624
R_2	.90000	.70000	.50000	.30000	.10000
R_3	.98100	.84700	.75000	.36300	.10900
R_5	.99980	.98143	.85353	.52528	.13134
R_7	.9999980	.99850	.96302	.70634	.16118

The kth term of (5) is the probability that point 1 is connected to exactly k of the $N - 2$ points $3, \cdots, N$. Then the sum is the probability that points 1 and 2 are not connected.

Using these results, R. W. Hamming and the author computed numerical values of P_N and R_N which appear in Table 1.

3. Bounds. The formulas of Section 2 solve the problem for small N only. In this section we estimate P_N and R_N for large N. As N increases, the number of paths by which points 1 and 2 may be joined increases. Then it is not surprising that $R_N \to 1$ as $N \to \infty$ for every fixed $p > 0$. That $P_N \to 1$ too is less obvious since increasing N also increases the number of pairs of points to be connected. Indeed, Table 1 shows P_N decreasing for $N \leqq 6$ when $q = .9$. The more precise results (1) and (2) follow from the bounds which we now derive.

THEOREM 1:

$$\left\{1 - \frac{N-1}{2} q^{N-1}\right\} N q^{N-1} \leqq 1 - P_N$$

and

$$1 - P_N \leqq q^{N-1}\{(1 + q^{(N-2)/2})^{N-1} - q^{(N-2)(N-1)/2}\} + q^{N/2}\{(1 + q^{(N-2)/2})^{N-1} - 1\}.$$

THEOREM 2:

$$(2 - q^{N-2})q^{N-1} \leqq 1 - R_N \leqq 2\, q^{N-1}(1 + q^{(N-2)/2})^{N-2}.$$

The lower bound in Theorem 2 is just the probability that at least one of the two points 1, 2 is connected to no other point.

A similar idea is used in Theorem 1. A lower bound on $1 - P_N$ is the probability T that at least one of the points $1, 2, \cdots, N$ is connected to no other point. Let E_i denote the event that point i is connected to no other point; then T is the union of the events $E_1, \cdots, E_N$. A lower bound on T (and hence on

$1 - P_N$) is provided by an inequality of Bonferroni (see Feller [2], p. 100):

$$\sum_i P(E_i) - \sum_{i<j} P(E_i E_j) \leqq T.$$

Since $P(E_i) = q^{N-1}$ and Pr $(E_i E_j) - q^{2N-3}$, we obtain the lower bound stated.

The upper bounds are obtained using (4) and (5). In both cases we bound P_k by 1. To bound $q^{k(N-k)}$ we use the fact that $x(N - x)$ is a convex function of x. Then

$$k(N - k) \geqq \frac{(N-2)k + N}{2} \qquad \text{if } 1 \leqq k \leqq \frac{N}{2},$$

$$k(N - k) \geqq \frac{(N-2)(N-k) + N}{2} \qquad \text{if } \frac{N}{2} \leqq k \leqq N - 1,$$

and

$$q^{k(N-k)} \leqq q^{N/2} \{q^{(N-2)k/2} + q^{(N-2)(N-k)/2}\}$$

for $1 \leqq k \leqq N - 1$. When these bounds are inserted into (4) and (5), the sums reduce to the expression shown in Theorems 1 and 2.

When N becomes large the bounds are in close agreement. It follows from Theorems 1 and 2 that

$$P_N = 1 - Nq^{N-1} + O(N^2 q^{3N/2}),$$

and

$$R_N = 1 - 2q^{N-1} + O(Nq^{3N/2}).$$

Checking these approximate formulas against P_N and R_N in Table 1, it appears likely that Nq^{N-1} and $2q^{N-1}$ will represent $1 - P_N$ and $1 - R_N$ to within 3% when $q \leqq .3$ and $N \geqq 6$. For the same degree of approximation, larger values of q will require larger values of N.

REFERENCES

[1] T. L. Austin, R. E. Fagen, W. F. Penny, and John Riordan, "The number of components in random linear graphs," *Ann. Math. Stat.*, Vol. 30 (1959) pp. 747–754.
[2] William Feller, *An Introduction to Probability Theory and its Applications*, Vol. I, Second Edition, Wiley, New York, 1957.
[3] E. N. Gilbert, "Enumeration of labelled graphs," *Canadian J. Math.*, Vol. 8 (1956), pp. 405–411.
[4] R. J. Riddell, Jr. and G. E. Uhlenbeck, "On the virial development of the equation of state of monoatomic gases," *J. Chem. Phys.*, Vol. 21 (1953), pp. 2056–2064.