The Annals of Probability
1973, Vol. 1, No. 5, 876-881

THE EXPECTED NUMBER OF COMPONENTS IN RANDOM LINEAR GRAPHS

BY ROBERT F. LING

University of Chicago

Exact, approximate, asymptotic, and computational formulas are derived for the expected number of components of any given size in a random linear graph. A theorem generalizes some asymptotic results of Austin, Fagen, Penney, and Riordan.

1. Introduction. Let $T_{n,r}$ denote the set of all linear graphs having n given labeled vertices and r edges; that is, the graphs are nondirected, without slings and without parallel edges. Throughout this paper, expressions of the form $\binom{\binom{p}{2}}{q}$ will be denoted by $N(p, q)$. The number of elements of $T_{n,r}$ is $N(n, r)$. In this paper, a random graph $\Gamma_{n,r}$ is defined as an element of $T_{n,r}$ chosen at random, so that each of the elements of $T_{n,r}$ has the same probability of being chosen, namely, $1/N(n, r)$. Such a model has been considered by Erdös and Rényi (see, e.g., [2], [3], [4], [5]). Austin, Fagen, Penney, and Riordan [1] considered a similar definition of random graphs, where both vertices and edges are labeled and two vertices can be connected by more than one edge; namely, the edges of their random graphs are considered as samples with replacement from $\binom{n}{2}$ edges. Gilbert [6] considered random graphs with labeled vertices but with a fixed probability p for each edge, independent of the number of vertices in the graph. This definition is different from the previous two in that the number of edges in the graph is random. According to Erdös and Rényi ([4] page 20), the probabilistic properties of $\Gamma_{n,r}$ under the first two formulations of random graphs are in general (if the number of edges r is not too large) asymptotically equal. In Section 2, exact, approximate, and asymptotic formulas for the expected number of components of size j (the number of connected subgraphs of $\Gamma_{n,r}$ with exactly j vertices) are presented. More asymptotic formulas and two computational formulas for the expected number of isolated vertices of a random graph are given in Section 3. The results in this paper can be interpreted as the probable structure of certain cluster analysis problems (see [8]).

2. Expected Number of Components of $\Gamma_{n,r}$. Let $C_{n,r}$ denote the number of connected graphs of $T_{n,r}$. The range of nonvanishing $C_{n,r}$ can trivially be seen to be $r = \overline{n-1}(1)\binom{n}{2}$. This range will be implicitly assumed in subsequent references to $C_{n,r}$. It follows from a result of Riddell and Uhlenbeck ([9] page 2060)

Received June 5, 1972; revised November 30, 1972.

AMS 1970 *subject classifications*. Primary 60C05; Secondary 05C30, 62E20.

Key words and phrases. Random linear graphs, connected subgraphs, expected number of components, trees, asymptotic approximations.

that

$$(1)\qquad C_{n,r} = \sum_{m=1}^{n} \frac{(-1)^{m+1}}{m} \sum' \frac{n!}{n_1!\, n_2! \cdots n_m!} \binom{\sum_{k=1}^{m} \binom{n_k}{2}}{r},$$

where $\sum'$ sums over all m-part partitions (or compositions) of n; that is, over all nonnegative integral solutions of $\sum_{k=1}^{m} n_k = n$, $n_1 \leqq n_2 \leqq \cdots \leqq n_m$.

THEOREM 1. *Let $\Gamma_{n,r,j}$ denote the number of components of $\Gamma_{n,r}$ of size j, then*

$$(2)\qquad E(\Gamma_{n,r,j}) = \frac{\binom{n}{j}}{N(n, r)} \sum_l C_{j,l} N(n-j, r-l),$$

where $C_{j,l}$ is defined by (1). *Alternatively,*

$$(3)\qquad E(\Gamma_{n,r,j}) = \binom{n}{j} \sum_l \frac{C_{j,l} H(l \mid \binom{j}{2} + j(n-j), r, \binom{n}{2})}{\binom{\binom{j}{2} + j(n-j)}{l}},$$

where H is the hypergeometric distribution defined by

$$H(l \mid s, r, t) = \binom{s}{l}\binom{t-s}{r-l}/\binom{t}{r}.$$

PROOF OF THEOREM 1. Consider random graphs of n labelled vertices V_1, $V_2, \cdots, V_n$ and r edges. Without loss of generality, let r_1 be the number of edges associated with the vertices $(V_1, V_2, \cdots, V_j)$; r_2 be the number of edges associated with $(V_{j+1}, \cdots, V_n)$; and r_3 be the edges between $(V_1, \cdots, V_j)$ and $(V_{j+1}, \cdots, V_n)$, $r = r_1 + r_2 + r_3$. Then (r_1, r_2, r_3) has a multivariate hypergeometric distribution (see Johnson and Kotz ([7] page 301), namely,

$$P(r_1 = n_1, r_2 = n_2, r_3 = n_3) = N(j, n_1)N(n-j, n_2)\binom{j(n-j)}{n_3}/N(n, r),$$

with $n_1 + n_2 + n_3 = r$. The probability that $(V_1, V_2, \cdots, V_j)$ is connected by l edges and isolated from other vertices is easily seen to be

$$P(r_1 = l, r_2 = r - l, r_3 = 0) \cdot C_{j,l}/N(j, l) = C_{j,l} \cdot N(n-j, r-l)/N(n, r).$$

$\therefore\ P((V_1, \cdots, V_j)$ is an isolated component$)$

$$= \sum_l C_{j,l} \cdot N(n-j, r-l)/N(n, r),$$

from which (2) follows. (3) is obtained by the substitution of the following identify into (2):

$$N(n-j, r-l)/N(n, r) = \binom{\binom{n}{2} - \binom{j}{2} - j(n-j)}{r-l} \Big/ N(n, r)$$

$$= H(l \mid \tbinom{j}{2} + j(n-j), r, \tbinom{n}{2}) \Big/ \binom{\binom{j}{2} + j(n-j)}{l}.$$

The proof of Theorem 1 is complete.

If we approximate $E(\Gamma_{n,r,j})$ by enumerating only those components of size j that are connected by $(j-1)$ or j edges (that is, trees or connected subgraphs containing exactly one cycle), we have

COROLLARY 1.1. *For $j \geqq 3$,*

$$E(\Gamma_{n,r,j}) \doteq \binom{n}{j}/N(n, r)\Big\{N(n-j, r-j+1)\cdot j^{j-2} + N(n-j, r-j)\frac{(j-1)!}{2}\Big(1 + j + \frac{j^2}{2!} + \cdots + \frac{j^{j-3}}{(j-3)!}\Big)\Big\}.$$

THEOREM 2. *Let $r = \alpha n$, $\alpha > 0$. Then for large n,*

$$E(\Gamma_{n,r,j}) \sim j^{j-2}\binom{n}{j}\binom{\alpha n}{j-1}\binom{n-j}{2}^{\alpha n-j+1}\binom{n}{2}^{-\alpha n}(j-1)!\,. \tag{4}$$

For j small in comparison to n and n large,

$$E(\Gamma_{n,r,j}) \sim n2^{j-1}j^{j-3}\alpha^{j-1}e^{-2\alpha j}/(j-1)!\,. \tag{5}$$

PROOF OF THEOREM 2. Approximating $E(\Gamma_{n,r,j})$ by the first nonvanishing term given in (3), we have

$$E(\Gamma_{n,r,j}) \doteq \binom{n}{j}j^{j-2}H\Big(j-1 \,\Big|\, \binom{j}{2} + j(n-j), \alpha n, \binom{n}{2}\Big)\Big/\binom{\binom{j}{2} + j(n-j)}{j-1}.$$

For n large or αn small compared to $\binom{n}{2}$, $H(j-1 \mid \binom{j}{2} + j(n-j), \alpha n, \binom{n}{2})$ admits the binomial approximation

$$\binom{\alpha n}{j-1}\left(\frac{j(n-j)+\binom{j}{2}}{\binom{n}{2}}\right)^{j-1}\left(1-\frac{j(n-j)+\binom{j}{2}}{\binom{n}{2}}\right)^{\alpha n-j+1}$$
$$= \binom{\alpha n}{j-1}\Big(j(n-j)+\binom{j}{2}\Big)^{j-1}\binom{n-j}{2}^{\alpha n-j+1}\Big/\binom{n}{2}^{\alpha n}.$$

The above, together with

$$\binom{\binom{j}{2} + j(n-j)}{j-1} \sim \Big(\binom{j}{2} + j(n-j)\Big)^{j-1}/(j-1)!$$

yield (4) of the theorem. For n large and j small compared to n, we have the Poisson approximation

$$H\Big(j-1 \,\Big|\, \binom{j}{2} + j(n-j), \alpha n, \binom{n}{2}\Big)$$
$$\sim \left(\alpha n\cdot\frac{\binom{j}{2}+j(n-j)}{\binom{n}{2}}\right)^{j-1}\Big/\left((j-1)!\exp\left(-\alpha n\frac{\binom{j}{2}+j(n-j)}{\binom{n}{2}}\right)\right)$$
$$\sim (2\alpha j)^{j-1}\exp(-2\alpha j)/(j-1)!\,.$$

Furthermore,

$$\binom{n}{j}\Big/\binom{\binom{j}{2} + j(n-j)}{j-1} \sim \frac{n^j}{j!}\frac{(j-1)!}{(jn)^{j-1}} = \frac{n}{j^j}.$$

Thus,

$$E(\Gamma_{n,r,j}) \sim j^{j-2}\frac{(2\alpha j)^{j-1}\exp(-2\alpha j)}{(j-1)!}\cdot\frac{n}{j^j} = n2^{j-1}j^{j-3}\alpha^{j-1}e^{-2\alpha j}/(j-1)!$$

The proof is complete.

For the special case $r = n$ (that is, $\alpha = 1$), (4) and (5) become

$$E(\Gamma_{n,n,j}) \sim j^{j-2}\binom{n}{j}\binom{n}{j-1}\binom{n-j}{2}^{n-j+1}\binom{n}{2}^{-n},$$ and

$$E(\Gamma_{n,n,j}) \sim n\cdot 2^{j-1}j^{j-3}e^{-2j}/(j-1)!$$

TABLE 1

Exact $E(\Gamma_{n,r,j})$ and approximations

A. $E(\Gamma_{n,r,j}) = \left[\binom{n}{j}\Big/\binom{\binom{n}{2}}{r}\right] \sum_{l \geq j-1} C_{j,l} \binom{\binom{n-j}{2}}{r-l}$, $\quad C_{j,l}$ *defined by* (1)

B. $E(\Gamma_{n,r,j}) \doteq \left[\binom{n}{j}\Big/\binom{\binom{n}{2}}{r}\right]\left[\binom{\binom{n-j}{2}}{r-j+1} j^{j-2} + \binom{\binom{n-j}{2}}{r-j} \frac{(j-1)!}{2}\left(1 + j + \frac{j^2}{2!} + \cdots + \frac{j^{j-3}}{(j-3)!}\right)\right]$

C. $E(\Gamma_{n,r,j}) \sim j^{j-2}\binom{n}{j}\binom{\alpha n}{j-1}\binom{n-j}{2}^{\alpha n-j+1}\binom{n}{2}^{-\alpha n}(j-1)!\,, \quad r = \alpha n$

D. $E(\Gamma_{n,r,j}) \sim n2^{j-1}j^{j-3}\alpha^{j-1}e^{-2\alpha j}/(j-1)!\,, \quad r = \alpha n$

$\alpha = 0.5$ ($r = 0.5n$)	$n = 20$				$n = 40$			
j	A	B	C	D	A	B	C	D
3	.6186	.6186	.5875	.4979	1.1024	1.1024	1.0792	.9957
4	.3677	.3673	.3262	.2442	.5893	.5891	.5584	.4884
5	.2611	.2599	.2135	.1404	.3721	.3716	.3396	.2807
6	.2066	.2045	.1546	.8924(−1)	.2614	.2606	.2285	.1785
7	.1732	.1703	.1189	.6082(−1)	.1977	.1965	.1647	.1216
8	.1459	.1430	.9365(−1)	.4362(−1)	.1576	.1560	.1247	.8724(−1)
9	.1151	.1133	.7192(−1)	.3253(−1)	.1306	.1285	.9770(−1)	.6506(−1)
10	.7486(−1)	.7486(−1)	.4921(−1)	.2502(−1)	.1110	.1084	.7847(−1)	.5004(−1)

$\alpha = 1$ ($r = n$)	$n = 20$				$n = 40$			
j	A	B	C	D	A	B	C	D
3	.7994(−1)	.7994(−1)	.8759(−1)	.9915(−1)	.1805	.1805	.1876	.1983
4	.3122(−1)	.3097(−1)	.3130(−1)	.3578(−1)	.6738(−1)	.6729(−1)	.6757(−1)	.7157(−1)
5	.1514(−1)	.1456(−1)	.1309(−1)	.1513(−1)	.3001(−1)	.2982(−1)	.2852(−1)	.3027(−1)
6	.8716(−2)	.7788(−2)	.6042(−2)	.7078(−2)	.1501(−1)	.1480(−1)	.1331(−1)	.1416(−1)
7	.5876(−2)	.4562(−2)	.2984(−2)	.3549(−2)	.8247(−2)	.7952(−2)	.6657(−2)	.7099(−2)
8	.4661(−2)	.2857(−2)	.1548(−2)	.1873(−2)	.4855(−2)	.4535(−2)	.3504(−2)	.3746(−2)
9	.4417(−2)	.1881(−2)	.8317(−3)	.1028(−2)	.3039(−2)	.2707(−2)	.1918(−2)	.2056(−2)
10	.5086(−2)	.1288(−2)	.4590(−3)	.5816(−3)	.2010(−2)	.1676(−2)	.1082(−2)	.1163(−2)

These special cases were derived in ([1] page 754), using an approach very different from that used in this paper.

The following corollary, which gives better insight into the magnitude of $E(\Gamma_{n,r,j})$, is easily derived from (5) through the use of Stirling's approximation:

COROLLARY 2.1. *For j small in comparison to n and n large,*

$$E(\Gamma_{n,r,j}) \sim \frac{n}{2\alpha(2\pi)^{\frac{1}{2}}} j^{-\frac{3}{2}} \exp\{(1 - 2\alpha + \log 2\alpha)j\},$$

where $r = \alpha n$, $\alpha > 0$.

Table 1 gives some comparisons of the exact $E(\Gamma_{n,r,j})$ given by Theorem 1 to the approximations given by Corollary 1.1 and Theorem 2. The values in the Table are rounded to four significant figures. The notation $a(b)$ means $a \times 10^b$.

3. Special case: Expected number of isolated vertices. Using the same notations as in the previous sections, we have

THEOREM 3.

$$E(\Gamma_{n,r,1}) = nH(0 \mid n-1, r, \tbinom{n}{2}) \qquad \textit{for all } n, r. \tag{6}$$

$$E(\Gamma_{n,r,1}) \sim n\left(1 - \frac{2}{n}\right)^r, \qquad r = o(n). \tag{7}$$

$$E(\Gamma_{n,r,1}) \sim ne^{-2\alpha}, \qquad r = \alpha n,\ \alpha > 0. \tag{8}$$

$$E(\Gamma_{n,r,1}) \sim n^{1-2\alpha}, \qquad r = \alpha n \log n,\ \alpha > 0. \tag{9}$$

$$E(\Gamma_{n,r,1}) \sim n \exp(-2\alpha n^\beta), \qquad r = \alpha n^{1+\beta},\ \alpha > 0,\ 0 < \beta < 1. \tag{10}$$

PROOF OF THEOREM 3. (6) is a special case of (3), since $C_{1,l} = 0$ for $l > 0$. For r small compared to $\binom{n}{2}$, $H(0 \mid n-1, r, \binom{n}{2})$ is adequately approximated by the binomial probability

$$\binom{r}{0}\left(\frac{n-1}{\binom{n}{2}}\right)^0\left(1 - \frac{n-1}{\binom{n}{2}}\right)^{r-0} = \left(1 - \frac{2}{n}\right)^r,$$

from which (7) follows. (8), (9), and (10) are results of applying the following fact to $(1 - 2n^{-1})^r$:

$$\left(1 + \frac{a\phi(n)}{n}\right)^{bn} \to_{n\to\infty} \exp[ab\phi(n)] \qquad \text{if} \quad \phi(n) = o(n^{\frac{1}{2}}).$$

Following are two easy corollaries of (6) stated without proof. These give computational formulas for $E(\Gamma_{n,r,1})$. Each formula avoids the use of binomial coefficients which could easily overflow in machine computation if n is not small.

COROLLARY 3.1.

$$E(\Gamma_{n,r+1,1}) = \left(1 - \frac{n \dot{-} 1}{\binom{n}{2} - r}\right) E(\Gamma_{n,r,1}), \qquad r \geqq 1.$$

COROLLARY 3.2.

$$E(\Gamma_{n,r,1}) = n \prod_{k=0}^{r-1}\left(1 - \frac{n-1}{\binom{n}{2} - k}\right), \qquad r \geqq 1.$$

REFERENCES

[1] AUSTIN, T. L., FAGEN, R. E., PENNY, W. F., and RIORDAN, J. (1959). The number of components in random linear graphs. *Ann. Math. Statist.* **30** 747–754.

[2] ERDÖS, P. (1959). Graph theory and probability. *Canad. J. Math.* **11** 34–38.

[3] ERDÖS, P. and RÉNYI, A. (1959). On random graphs I. *Publ. Math. Debrecen.* **6** 290–297.

[4] ERDÖS, P. and RÉNYI, A. (1960). On the evolution of random graphs. *Publ. Math. Inst. Hungar. Acad. Sci.* **5** 17–61.

[5] ERDÖS, P. and RÉNYI, A. (1961). On the evolution of random graphs. *Bull. Inst. Internat. Statist. Tokyo.* **0** 343–347.

[6] GILBERT, E. N. (1959). Random graphs. *Ann. Math. Statist.* **30** 1141–1144.

[7] JOHNSON, N. L. and KOTZ, S. (1969). *Discrete Distributions.* Houghton Mifflin, Boston.

[8] LING, R. F. (1971). Cluster analysis. Ph. D. dissertation, Yale University.

[9] RIDDELL, R. J., JR. and UHLENBECK, G. E. (1953). On the virial development of the equation of state of monoatomic gases. *J. Chem. Phys.* **21** 2056–2064.

UNIVERSITY OF CHICAGO
GRADUATE SCHOOL OF BUSINESS
5836 GREENWOOD AVENUE
CHICAGO, ILLINOIS 60637