THE NUMBER OF COMPONENTS IN RANDOM LINEAR GRAPHS

By T. L. Austin, R. E. Fagen, W. F. Penney and John Riordan

Silver Spring, Md.; Hughes Aircraft Co., Silver Spring, Md.; Bell Telephone Lab. Inc.

1. Introduction. Given n distinct points, m selections of pairs of points are made independently and at random, each of the $\binom{n}{2}$ possible pairs having the same chance $1/\binom{n}{2}$ of selection at each trial. Once selected, a pair is connected by a line joining its two points, labeled by the order number of its selection; thus, after m selections, a linear graph with n distinct (labeled) points and m distinct (labeled) lines connecting pairs of points is formed. (Note that the rule of formation implies the graph contains no slings but may contain lines in parallel.) In many investigations it is valuable to have the distribution of the number of connected components (each isolated point being counted as a component) in such a random linear graph.

In the following this distribution is found. In addition, simple closed expressions are given for a few special cases of interest, and finally, an approximation for the average number of components.

2. Summary of results. Let $N = \binom{n}{2}$, and let T_{nmp} be the number of graphs (as described above) with n points, m lines and p parts; then, of course, the probability that a graph has p parts is T_{nmp}/N^m. Let $C_{nm} = T_{nm1}$ be the number of the corresponding connected graphs (single component) with n points, m lines, and introduce the following enumerating functions:

$$\begin{aligned} T(x, y, z) &= \sum \sum \sum T_{nmp} \frac{x^n}{n!} \frac{y^m}{m!} z^p = \sum \sum T_{nm}(z) \frac{x^n}{n!} \frac{y^m}{m!} \\ &= \sum T_n(y, z) \frac{x^n}{n!}, \qquad n, m = 0, 1, \cdots, p = 1, 2, \cdots, n \end{aligned} \tag{1}$$

and

$$C(x, y) = \sum \sum C_{nm} \frac{x^n}{n!} \frac{y^m}{m!} = \sum C_n(y) \frac{x^n}{n!}. \tag{2}$$

Then

$$T(x, y, z) = \exp[z\, C(x, y)] \tag{3}$$

is the most concise expression of the relations between enumerators. Since $T_{nm}(1)$ is N^m, $T(x, y, 1)$ is known; hence so is $C(x, y)$, and $T(x, y, z)$ is completely determined by (3).

Received July 14, 1958; revised December 16, 1958.

Indeed, using the abbreviation

$$(4) \qquad \tau_{nk}(m) = \sum \frac{n!\left[k_1\binom{1}{2} + k_2\binom{2}{2} + \cdots + k_n\binom{n}{2}\right]^m}{k_1!\cdots k_n!\,1!^{k_1}2!^{k_2}\cdots n!^{k_n}}$$

with summation over all k-part partitions of n, that is over all non-negative integral solutions of

$$k_1 + 2k_2 + \cdots + nk_n = n$$
$$k_1 + k_2 + \cdots + k_n = k,$$

it turns out that

$$(5) \qquad T_{nmp} = \sum_{k=p}^{n} \tau_{nk}(m)s(k, p),$$

$$(6) \qquad C_{nm} = \sum_{k=1}^{n} (-1)^{k-1}(k-1)!\,\tau_{nk}(m).$$

In (5), $s(k, p)$ is a Stirling number of the first kind defined by

$$(7) \qquad (z)_k = z(z-1)\cdots(z-k+1) = \sum s(k,p)z^p.$$

It is also interesting to notice that

$$(8) \qquad T_{nm}(z) = \sum_{k=0}^{n} (z)_k\, \tau_{nk}(m).$$

The special cases of (5) of most interest are

$$(9) \qquad T_{n,n-1,1} = (n-1)!\,n^{n-2},$$

$$(10) \quad T_{n,n,1} = \tfrac{1}{2}n!(n-1)!\left[1 + n + \frac{n^2}{2!} + \cdots + \frac{n^{n-3}}{(n-3)!} + \frac{n^{n-2}}{(n-2)!}\right].$$

Equation (10) depends on the following auxiliary result which is probably of more interest in graph theory: the number of connected linear graphs with n distinct points and exactly one cycle of length k, for $k > 2$, is $(n)_k\, n^{n-k-1}/2$, while for $k = 2$ it is $(n)_2\, n^{n-3}$. This is a natural extension of the result of Cayley's used in (9) that the number of (free) trees with n distinct points is n^{n-2}, and it is an instance of a more general result appearing in G. W. Ford and G. E. Uhlenbeck [1].

3. Derivation. Consider first the enumerator

$$(11) \qquad T_{nm}(z) = \sum T_{nmp}z^p.$$

As already noticed, $T_{nm}(1) = \sum T_{nmp} = N^m$, since the m lines are chosen independently from the same population of $N = \binom{n}{2}$.

For orientation, the first few evaluations of (11), obtained by easy enumerations, are as follows:

$$
\begin{aligned}
T_{n0}(z) &= z^n \\
T_{n1}(z) &= \binom{n}{2} z^{n-1} \\
T_{n2}(z) &= \binom{n}{2} z^{n-1} + \left[6\binom{n}{3} + 6\binom{n}{4}\right] z^{n-2} \\
T_{n3}(z) &= \binom{n}{2} z^{n-1} + \left[24\binom{n}{3} + 18\binom{n}{4}\right] z^{n-2} \\
&\quad + \left[96\binom{n}{4} + 180\binom{n}{5} + 90\binom{n}{6}\right] z^{n-3}.
\end{aligned}
$$

Notice that

$$
\begin{aligned}
T_{n0}(z) &= zT_{n-1,0}(z) \\
T_{n1}(z) &= z(T_{n-1,1}(z) + (n-1)T_{n-2,0}(z)) \\
T_{n2}(z) &= z(T_{n-1,2}(z) + 2(n-1)T_{n-2,1}(z) + (n-1)T_{n-2,0}(z) \\
&\quad + 6\binom{n-1}{2} T_{n-3,0}(z)\Big).
\end{aligned}
$$

The general form of the recurrence suggested by these may be derived by a slight modification of an argument given by E. N. Gilbert [2]. Thus in the graphs with $n + 1$ labeled vertices, m labeled lines and p parts enumerated by $T_{n+1,m,p}$, the vertex labeled $n + 1$ belongs in a connected part with i other points and j lines, while the remaining $n - i$ points and $m - j$ lines belong to a graph with $p - 1$ parts. Since the labels for the i points and j lines may be chosen in $\binom{n}{i}\binom{m}{j}$ ways, it follows at once that

$$
T_{n+1,m,p} = \sum_{i,j} \binom{n}{i}\binom{m}{j} C_{i+1,j}\, T_{n-i,m-j,p-1}\,. \tag{12}
$$

Multiplying by z^p and summing on p, it is found that

$$
T_{n+1,m}(z) = z \sum \binom{n}{i} C_{i+1,j}\, T_{n-i,m-j}(z)\,. \tag{13}
$$

For boundary conditions note that $T_{1m}(z) = z\delta_{0m}$, with $\delta_{00} = 1$, $\delta_{0m} = 0, m > 0$, and for consistency with equation (13) $T_{0m}(z) = \delta_{0m}$, since $C_{1j} = \delta_{0j}$. Note also that $C_{n0} = 0$, $n > 1$, $C_{n,n-j} = 0$, $j > 1$, and to verify the instances of (13) appearing above, $C_{21} = C_{22} = 1$, $C_{32} = 6$. For concreteness, it may also be noted that

$$T_{20}(z) = z^2, \qquad T_{2m}(z) = z, \qquad m > 0$$
$$T_{30}(z) = z^3, \qquad T_{3m}(z) = 3z^2 + z(3^m - 3), \qquad m > 0.$$

Multiplying (13) by $y^m/m!$ and summing on m leads to

$$T_{n+1}(y, z) = z \sum \binom{n}{i} C_{i+1}(y) T_{n-i}(y, z) \,. \tag{14}$$

Multiplying (14) in its turn by $x^{n+1}/(n+1)!$ leads to

$$\frac{\partial T(x, y, z)}{\partial x} = z \frac{\partial C(x, y)}{\partial x} T(x, y, z) \,. \tag{15}$$

Integrating with respect to x using the boundary conditions $T(0, y, z) = C(0, y) = 0$ gives (3), that is

$$T(x, y, z) = \exp[z\, C(x, y)]. \tag{3}$$

The further results reported above are obtained directly from (3) as follows: first

$$\begin{aligned} T(x, y, z) &= [\exp C(x, y)]^z \\ &= [1 + T(x, y, 1) - 1]^z \\ &= \exp (z)\ (T(x, y, 1) - 1), \quad (z)^k \equiv (z)_k \,. \end{aligned} \tag{16}$$

Next, a basic equation for the Bell multivariate polynomials (see [3], Section 2.8)

$$Y_n(ay_1, \cdots, ay_n) = \sum \frac{n!\, a_k}{k_1! \cdots k_n!} \left(\frac{y_1}{1!}\right)^{k_1} \cdots \left(\frac{y_n}{n!}\right)^{k_n} \tag{17}$$

with summation over all partitions of n, is

$$\sum Y_n(ay_1, \cdots, ay_n) \frac{x^n}{n!} = \sum \frac{a_n}{n!} \left(xy_1 + \frac{x^2 y_2}{2!} + \cdots\right)^n. \tag{18}$$

Hence (16) is equivalent to (equating coefficients of $x^n/n!$),

$$T_n(y, z) = Y_n(aT_1, \cdots, aT_n), \qquad a^k \equiv a_k = (z)_k \tag{19}$$

with $T_n \equiv T_n(y, 1) = \exp(yN)$. Using (17) for the right hand side and equating coefficients of $y^m/m!$ gives (8). Introducing the Stirling numbers of the first kind in (8) by use of (7) gives (5).

Finally the relation (18) along with the instance $z = 1$ of (3) namely $T(x, y, 1) = \exp C(x, y)$ shows that

$$T_n \equiv T_n(y, 1) = Y_n(C_1(y), \cdots, C_n(y)) \tag{20}$$

and the inverse of this (*cf.* [3], equation 2.51) is

$$C_n(y) = Y_n(fT_1, \cdots, fT_n), \qquad f^k \equiv f_k = (-1)^{k-1} (k-1)!. \tag{21}$$

Equating coefficients of $y^m/m!$ again, gives (6).

4. Special cases. While the results above are formally complete, they may become almost impossibly difficult to write out for large n since summation is over all partitions. Special cases obtainable otherwise are a valuable adjunct and as already noted, those given by equations (9) and (10) are independently interesting in the theory of graphs.

The number $T_{n,n-1,1} = C_{n,n-1}$ is the number of graphs with n labeled points, $n - 1$ labeled lines and 1 part, that is the number of free trees with all points and lines labeled. The lines and points are labeled independently. The number of free trees with all points (and no lines) labeled is n^{n-2}, by Cayley's formula, and the number of line labelings is $(n - 1)!$

The number $T_{n,n,1} = C_{n,n}$ is obtained in a similar way, the graphs consisting of a single connected part containing exactly one closed path (cycle) and with all points and lines labeled. The essential enumeration is of such graphs with cycle length k, and with all points (and no lines) labeled.

These graphs may be enumerated by use of a theorem due to Pólya ([3], Chapter 6) since they may be regarded as formed by placing rooted trees at the vertices of the k-sided polygon formed by the cycle. Their enumerator by number of points and number of point labels may be written $d_k(x, y) = \sum_{n,m} d_{nm}(k)x^n y^m/m!$ and by [3], problem 25 of Chapter 6,

$$d_k(x, y) = D_k(r(x, y), r(x^2), \cdots, r(x^k)) \tag{16}$$

with $r(x, y)$ the enumerator of rooted trees by number of points and number of point labels, $r(x) \equiv r(x, 0)$ and $D_k(t_1, t_2, \cdots t_k)$ the cycle index of the dihedral group:

$$\begin{aligned} 2D_k(t_1, t_2, \cdots, t_k) &= Z_k(t_1, t_2, \cdots, t_k) + t_1 t_2^j, \qquad & k = 2j + 1 \\ &= Z_k(t_1, t_2, \cdots, t_k) + S_2 t_2^{j-1}, \qquad & k = 2j \end{aligned}$$

and

$$Z_n(t_1, t_2, \cdots, t_n) = \frac{1}{n} \sum_{d|n} \varphi(d) t_d^{n/d}$$

$$S_2 \equiv S_2(t_1, t_2) = (t_1^2 + t_2)/2 .$$

($\varphi(d)$ is Euler's totient function, the number of integers less than d and relatively prime to d, $\varphi(1) = 1$, and the sum for Z_n, the cycle index of the cyclic group, is over all divisors of n, including 1 and n).

Making the substitution $y = z/x$ in the definition of $d(x, y)$ changes it to the form $d_k(x, z) = d_{k0}(z) + x\, d_{k1}(z) + \cdots$ with $d_{kj}(z) = \sum d_{n+j,n}(k)\, z^n/n!$. Hence the numbers required, $d_{nn}(k)$, are enumerated by $d_{k0}(z)$ which is obtained from (16) as

$$d_{20}(z) = r_0^2(z)/2, \qquad d_{k0}(z) = r_0^k(z)/2k, \qquad k > 2 \tag{17}$$

with

$$r_0(z) = \sum_{n=0} r_{nn}\, z^n/n! = \sum n^{n-1}\, z^n/n!$$

Noting that $r_0(z) = z \exp r_0(z)$ the Lagrange formula

$$f(u) = f(0) + \sum_{n=1} \frac{z^n}{n!}\left[\frac{d^{n-1}}{dx^{n-1}} f'(x)\varphi^n(x)\right]_{x=0}$$

with $u = z\varphi(u)$, gives $d_{k0}(z)$ with $u = r_0(z)$, $\varphi(u) = e^u$ and $f(u) = u^k$ as

$$d_{k0}(z) = \sum_{n=1} [(n)_k\, n^{n-k-1}/2]\, z^n/n!, \qquad k > 2$$

$$d_{20}(z) = \sum (n)_2\, n^{n-3}\, z^n/n!$$

and $T_{n,n,1} = n! \sum_{k=2} d_{nn}(k)$ is obtained as in (10).

At the other extreme, it may be noted that

$$T_{n,m,n-1} = \binom{n}{2}$$

$$T_{n,m,n-2} = (3^m - 3)\binom{n}{3} + 3(2^m - 2)\binom{n}{4}$$

$$\begin{aligned} T_{n,m,n-3} = {} & (4^m - 4.3^m - 3.2^m + 12)\binom{n}{4} \\ & + 10(6^m - 3^m - 3.2^m + 5)\binom{n}{5} \\ & + 15(3^m - 3.2^m + 3)\binom{n}{6} \end{aligned}$$

5. Average number of components. The average number of components can be computed directly by (3); let M_{nm} be the average with n points and m lines, and $A_{nm} = M_{nm}N^m$, $N = \binom{n}{2}$. Then if $A_n(y) = \sum A_{nm}\, y^m/m!$, the relation

$$\frac{\partial}{\partial z} T(x, y, z)\,|_{z=1} = \sum A_n(y)\,\frac{x^n}{n!}$$

follows. Differentiating in (3), leads to $(\partial/\partial z)T(x, y, z)\,|_{z=1} = C(x, y)\; T(x, y, 1)$, whence

(22) $\quad A_n(y) = (C(y) + T(y))^n$, $C^n(y) = C_n(y)$, $T^n(y) = T_n(y, 1)$.

Recalling that $T_n(y, 1) = T_n = \exp(yN)$ and that by (21) $C_n(y)$ is expressible in terms of T_1 to T_n, equation (22) leads to an explicit expression for $A_n(y)$, namely

(23) $$A_n(y) = Y_n(bT_1, \cdots, bT_n),\ b^k \equiv b_k$$

with $b_k = 1$, $b_k = (-1)^k (k - 2)!$, $k > 1$ and T_n as above.

While complete, this has the disadvantage of increasing elaboration with n. The following alternative development is more easily adapted to asymptotic approximation.

Let $S_1, S_2, \cdots, S_n$ denote respectively the number of components which are single points, isolated connected pairs, isolated connected trios, etc. Then $M_{nm} = E(S_1) + E(S_2) + \cdots + E(S_n)$. Now let

$$\begin{aligned}
S_1 &= x_1 + x_2 + \cdots + x_n \\
S_2 &= x_{12} + x_{13} + \cdots + x_{n-1,n} \\
&\vdots \\
S_j &= x_{12\cdots j} + \cdots + x_{n-j+1,n-j+2,\cdots,n-1,n} \\
&\vdots \\
S_n &= x_{12\cdots n}
\end{aligned}$$

where x_1 is 1 or 0 according as point 1 is isolated or not, x_{12} is 1 or 0 according as points 1 and 2 are connected and isolated, or not, etc.; then

$$\begin{aligned}
E(S_1) &= nE(x_1) = np_1 \\
E(S_2) &= \binom{n}{2} E(x_{12}) = \binom{n}{2} p_{12} \\
E(S_j) &= \binom{n}{j} E(x_{12\cdots j}) = \binom{n}{j} p_{12\cdots j}
\end{aligned}$$

where $p_{12\cdots j}$ is the probability that points $1, 2, \cdots, j$ are connected and isolated. Then $M_{nm} = np_1 + \binom{n}{2}p_{12} + \cdots \binom{n}{j}p_{12\cdots j} + \cdots + \binom{n}{n}p_{12\cdots n}$; to estimate the quantity M_{nm} it is necessary only to estimate the probabilities above. To illustrate one approximation which seems quite simple, suppose the approximation is on M_{nn} as a function of n; first, the p's can be estimated as follows:

$$\begin{aligned}
p_1 &= \left[\frac{\binom{n-1}{2}}{\binom{n}{2}}\right]^n = \left[\frac{(n-1)(n-2)}{n(n-1)}\right]^n \\
&= \left[\frac{n-2}{n}\right]^n \sim e^{-2} \quad \text{and} \quad E(S_1) \sim ne^{-2},
\end{aligned}$$

which is exact, except for the asymptotic approximation in the last step. Next, p_{12} can be estimated by the following argument; for points 1 and 2 to be connected and isolated, they must be joined either by a single line (forming a tree with two labeled points), or by two lines (forming a graph with a single cycle), or by three or more lines. Thus

$$\begin{aligned}
p_{12} = \Bigg[nT_{2,1,1}\binom{n-2}{2}^{n-1} &+ \binom{n}{2} T_{2,2,1}\binom{n-2}{2}^{n-2} \\
&+ \binom{n}{3} T_{2,3,1}\binom{n-2}{2}^{n-3} + \cdots\Bigg] \Big/ \binom{n}{2}^n.
\end{aligned}$$

Using (9) and (10), and noting that all except the first term result in terms $0(1)$ and smaller, $E(S_2) \doteq ne^{-4} + 0(1)$. This argument can be continued, and results, in effect, in neglecting all terms which result from counting connections of a $j$tuple by more than $j - 1$ lines; the typical approximation would be

$$E(S_j) \doteq \binom{n}{j-1} \frac{\binom{n}{j} T_{j,j-1,1} \binom{n-j}{2}^{n-j+1}}{\binom{n}{2}^n};$$

for j small in comparison with n this can be further approximated by

$$E(S_j) \doteq \frac{2^{j-1} j^{j-3}}{(j-1)!} ne^{-2j}.$$

Finally, only those $E(S_j)$ need be used that are of significant size; i.e., for $n < 40$, $ne^{-4} < 1$; to simplify, it is sufficient to take only those terms which are greater than 1 and estimate the total contribution of all other terms by 1 (which, in effect, says that the average is most heavily contributed to by one large component and isolated points if $n < e^4$, and by one large component plus isolated points plus isolated pairs if $e^4 < n < e^6$, etc.) Thus, reasonable approximations are

$$M_{nn} \doteq 1 + n/e^2, \qquad n < e^4,$$
$$M_{nn} \doteq 1 + n/e^2 + n/e^4, \qquad e^4 < n < e^6,$$
$$M_{nn} \doteq 1 + n/e^2 + n/e^4 + 2n/e^6, \qquad e^6 < n < e^8, \text{ etc.}$$

The following table indicates these approximations may be satisfactory, at least for n of moderate size (using $M_{nn} \doteq 1 + n/e^2$, values given to 3 places):

M_{nn} = Mean Number of Components

n	Exact	Approx.	Diff.	Rel. Error
3	1.111	1.406	.295	.265
4	1.282	1.541	.259	.202
5	1.462	1.677	.215	.147
6	1.642	1.812	.170	.103
7	1.819	1.947	.128	.071
8	1.993	2.083	.090	.045
9	2.166	2.218	.052	.024
10	2.336	2.353	.017	.007

REFERENCES

[1] G. W. Ford and G. E. Uhlenbeck, "Combinatorial problems in the theory of graphs I," *Proc. Nat. Acad. Sci. USA*, Vol. 42 (1956), pp. 122–128.

[2] E. N. Gilbert, "Enumeration of labelled graphs," *Canadian J. Math.*, Vol. 8 (1957), pp. 405–411.

[3] J. Riordan, *An Introduction to Combinatorial Analysis*, John Wiley and Sons, New York, 1958.