ENUMERATION OF LABELLED GRAPHS

E. N. GILBERT

1. Introduction. The number of connected linear graphs having V vertices labelled $1, \ldots, V$ and λ (unlabelled) lines is found below. Similar formulas are found for graphs in which slings, lines "in parallel," or both are allowed and for directed graphs with or without slings or parallel lines. Some of these graphs are also counted when the lines are labelled and the vertices are unlabelled. Another type of graph which is counted is connected, has no cycles of odd length (even graph), and has L labelled lines and λ unlabelled vertices.

Two graphs with labelled vertices are counted as the same if and only if for all i and j the same number of lines go from the vertex labelled i to the vertex labelled j in both graphs. Consequently two topologically equivalent graphs may be counted as distinct labelled graphs if they are labelled differently. An enumeration of some unlabelled connected graphs has been given by Riddell and Uhlenbeck (**9**) and by Harary (**4**). Riddell and Uhlenbeck also count connected graphs with labelled vertices in which slings, lines in parallel, and directed lines are ruled out.

2. Graphs with labelled vertices. The kinds of labelled graphs mentioned above are easy to count if one removes the condition that the graph be connected. For example the number of labelled graphs (including disconnected graphs) which have V vertices labelled $1, \ldots, V$ and λ unlabelled lines and no pair of vertices joined by more than one line (no lines in *parallel*) is clearly the binomial coefficient $\binom{\frac{1}{2}V(V-1)}{\lambda}$. For, every such graph is just a collection of λ of the $\frac{1}{2}V(V-1)$ lines which can be drawn between pairs of distinct vertices. More generally let P be any property of connected graphs. Let $T_{V,\lambda}$ be the total number of graphs having V labelled vertices, λ unlabelled lines and such that every connected component of the graph has the property P. Suppose the numbers $T_{V,\lambda}$ are known and consider the number $C_{V,\lambda}$ of connected graphs of V labelled vertices and λ unlabelled lines which have the property P. To get a recurrence equation for $C_{V,\lambda}$ note that in a graph J with $V + 1$ vertices and λ lines the vertex labelled $V + 1$ belongs to a connected component K having some number v of other vertices and some number μ of lines. The remaining part $J - K$ of J has $V - v$ vertices and $\lambda - \mu$ lines. There are $\binom{V}{v}$ ways in which v of the labels $1, \ldots, V$ can be chosen to be

assigned to the graph K, then $C_{v+1,\mu}$ ways of picking K, and $T_{V-v,\lambda-\mu}$ ways of picking $J - K$. Hence we conclude

$$(1) \qquad T_{V+1,\lambda} = \sum_{v,\mu} \binom{V}{v} C_{v+1,\mu} T_{V-v,\lambda-\mu}.$$

In (1) we must make the convention that $T_{0,\lambda} = 1$ if $\lambda = 0$ and $T_{0,\lambda} = 0$ otherwise. Introducing the generating functions

$$C_V(y) = \sum_{\lambda} C_{V,\lambda} y^{\lambda}$$

and

$$T_V(y) = \sum_{\lambda} T_{V,\lambda} y^{\lambda}$$

(1) assumes a simple form

$$(2) \qquad T_{V+1}(y) = \sum_{v} \binom{V}{v} C_{v+1}(y) T_{V-v}(y)$$

which relates $C_{V+1}(y)$ (the term $v = V$) to $C_1(y), \ldots, C_V(y)$ and the known $T_n(y)$.

For computing purposes (2) is quite convenient. However, we will also solve (2) for $C_{V+1}(y)$ explicitly in terms of $T_V(y)$. This solution is obtained in the form of the generating function

$$(3) \qquad C(x, y) = \sum_{V} C_V(y) \, x^V / V!.$$

A compact derivation is achieved using the symbolic method **(2; 6)**. We use a special pair of parentheses $\{\ldots\}_s$ to enclose expressions which are to be interpreted symbolically. The expressions inside such parentheses will be analytic functions depending on x and two letters C and T. The entire expression, including parentheses, stands for the formula which is obtained when the analytic function is expanded into a power series and the term $T^m C^n x^i$ is replaced by $T_m(y) \, C_n(y) x^i$ for all m, n, i. For example (3) becomes

$$C(x, y) = \{\exp Cx\}_s$$

and (2) becomes
$$(4) \qquad T_{V+1}(y) = \{C(T + C)^V\}_s.$$

Multiplying both sides of (4) by $x^V / V!$ and summing on V one derives

$$\{T \exp Tx\}_s = \{C \exp (T + C)x\}_s$$
$$= \{C \exp Cx\}_s \{\exp Tx\}_s$$
$$(5) \qquad \{C \exp Cx\}_s = \{T \exp Tx\}_s / \{\exp Tx\}_s.$$

Integrating both sides of (5) with respect to x from 0 to x one also has

$$\{\exp Cx\}_s - \{C^0\}_s = \log\{\exp Tx\}_s.$$

If by convention we put $\{C^0\}_s = C_0(y) = 0$, then

$$(6) \qquad C(x, y) = \{\exp Cx\}_s = \log\{\exp Tx\}_s.$$

Either of (5) or (6) provides an explicit solution for $C_V(y)$ which, in the usual notation, is given by the following theorem.

THEOREM 1. *Let $T_V(y)$ be the generating function for the number of graphs with V labelled vertices, λ lines, and such that each component has a given property P. Let $C_V(y)$ be the generating function for the number of connected graphs having V labelled vertices, λ lines, and having property P. Then $C_V(y)$ is $(V-1)!$ times the coefficient of x^{V+1} in the power series for the quotient*

$$\sum_{V=0}^{\infty} T_{V+1}(y) \frac{x^V}{V!} \Big/ \sum_{V=0}^{\infty} T_V(y) \frac{x^V}{V!} .$$

$C_V(y)$ is also $V!$ times the coefficient of x^V in the power series for

$$\log \sum_{V=0}^{\infty} T_V(y) \, x^V / V!.$$

In these series $T_0(y) = 1$ by convention.

All the steps leading up to Theorem I can be justified using results in a paper on the symbolic method by Bell **(2)**. In all our applications the power series $\{\exp Cx\}_s$ has a zero radius of convergence; nevertheless $C_V(y)$ is obtained from either form of the generating series by formally expanding the generating series just as though it converged. One can also prove Theorem I using traditional methods but the proof is laborious. For example, expanding the generating function

$$\log\left(1 + \sum_{V=1}^{\infty} T_V(y) \, x^V / V!\right)$$

into a formal power series in x, Theorem I states that

(7) $$C_V(y) = - V! \sum_{b_1,\ldots,b_V} \frac{(\Sigma b_i - 1)!}{b_1! \ldots b_V!} \left(- \frac{T_1(y)}{1!}\right)^{b_1} \ldots \left(- \frac{T_V(y)}{V!}\right)^{b_V}$$

where the sum is taken over all partitions $b_1 + 2b_2 + \ldots + Vb_V = V$. It is then possible to retrace the steps of our symbolic proof backward to show that the expression (7) is indeed the solution of the recurrence equation (2).

It was noted above that there are

$$\binom{\frac{1}{2}V(V-1)}{\lambda}$$

graphs with λ lines and V labelled vertices and no lines in parallel. Hence putting

(8) $$T_V(y) = (1 + y)^{\frac{1}{2}V(V-1)}$$

in Theorem I or in (2) we count connected graphs with V labelled vertices and λ lines, none in parallel. Setting $y = 1$ we count these graphs by vertices, allowing any number of lines. For $V = 1, 2, \ldots, 5$ we find 1, 1, 4, 38, and 728 connected graphs. The 38 connected graphs with four labelled vertices fall into the six topologically distinct types shown in Figure 1.
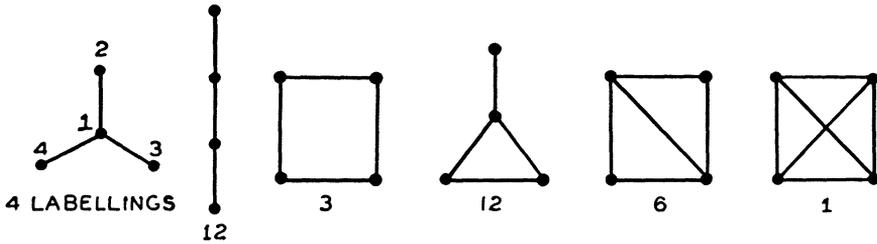
Figure 1

Similarly to count graphs in which any number of lines in parallel are allowed $T_{V,\lambda}$ is the number of combinations of λ lines drawn from $\frac{1}{2}V(V-1)$ different kinds of lines with repetitions allowed, i.e.

$$T_{V,\lambda} = \begin{pmatrix} \frac{1}{2}V(V-1) + \lambda - 1 \\ \lambda \end{pmatrix}$$

and

(9) $$T_V(y) = (1-y)^{-\frac{1}{2}V(V-1)}.$$

A *sling* is a line of a graph which has both its end-points the same. Two slings which share the same vertex will be considered to be in parallel. Slings were ruled out in the above enumerations. If slings are to be allowed the same sort of argument applies. Now there are V additional kinds of lines to choose from so that the terms $\frac{1}{2}V(V-1)$ in (8) and (9) are to be replaced by $\frac{1}{2}V(V+1)$.

Similarly various kinds of directed graphs may be counted. For instance if the graphs are to be composed entirely of lines directed between different vertices but slings are excluded there are $V(V-1)$ kinds of lines and the exponents of (8) and (9) are to be multiplied by 2. Note that this modification of (8) counts two lines joining the same pair of vertices as not in parallel if they have opposite direction. To count directed graphs with slings the exponents in (8) and (9) are changed to V^2, which is the number of kinds of directed lines including the V slings (which can have only one direction). A diadic relation may be interpreted as a directed graph in which there is a line directed from i to j if i has the given relation to j. In a paper enumerating structures of relations, Davis **(3)** counts certain kinds of topologically different (i.e. unlabelled) directed graphs.

Our results are summarized most compactly by the generating functions of the following theorem.

THEOREM II. *The number of connected graphs having V labelled vertices and λ unlabelled lines is $V!$ times the coefficient of $x^V y^\lambda$ in a generating series of the form*

$$\log\left(1 + \sum_{i=1}^{\infty} \frac{(1 + \alpha y)^{\alpha\beta(i)} x^i}{i!}\right),$$

where

$$\alpha = \begin{cases} -1 \ \textit{if lines in parallel are allowed,} \\ \ \ \ 1 \ \textit{otherwise,} \end{cases}$$

and

$$\beta(i) = \begin{cases} \dbinom{i}{2} & \textit{for (undirected) graphs without slings,} \\ \dbinom{i+1}{2} & \textit{for (undirected) graphs with slings,} \\ i(i-1) & \textit{for directed graphs without slings,} \\ i^2 & \textit{for directed graphs with slings.} \end{cases}$$

The formula with $\alpha = 1$, $\beta(i) = \dbinom{i}{2}$ was derived by Riddell and Uhlenbeck **(8)** using quite different techniques.

3. Graphs with labelled lines. Graphs having labelled lines and un-labelled vertices may be counted by exactly the same technique as used in §2. Equation (1) remains true if we interpret $C_{V,\lambda}$ as the number of connected graphs with property P and having V lines labelled $1, 2, \ldots, V$ and λ unlabel-led vertices and $T_{V,\lambda}$ as the number of (perhaps disconnected) graphs having V labelled lines, λ unlabelled vertices and such that every component has property P. With this change in viewpoint Theorem I counts connected graphs with labelled lines if the corresponding graphs (not necessarily con-nected) can be counted.

As an example of such an enumeration, let P be the property that every cycle of the graph must be of even length. König **(7**, Ch. XI, §4**)** calls such a graph an *even graph*. We now use L for the number of labelled lines. Lines in parallel will be permitted but slings are cycles of odd length and so are ex-cluded automatically.

The vertices of even graphs separate into two classes A and B with the property that every line has one end-point in A and one in B. The connections among the L end-points in A and the L end-points in B form two groupings of the L lines into clusters; all lines belonging to the same cluster are connected together at a single vertex. It will be convenient first to count even graphs in which each vertex carries a label A or B to show to which of the two classes it belongs. Let a denote the number of vertices in A and b the number in B. The number of graphs with these values of a and b is just $S(L, a) \, S(L, b)$ where $S(L, k)$ is the number of ways of putting L different objects (the end-points) into k groups (the vertices) so that no group is empty. $S(L, k)$ is a Stirling number of second kind **(5**, p. 179**)** and is given by the generating function

(10) $$\sum_{L, k} S(L, k) \, y^k \frac{x^L}{L!} = \exp[y(\exp x - 1)].$$

The generating function $T_L(y)$ for the number of graphs with λ vertices is then

$$T_L(y) = \sum_{a,\lambda} S(L, a)\, S(L, \lambda - a)\, y^\lambda$$

(11)
$$= \left(\sum_k S(L, k)\, y^k \right)^2.$$

Theorem I together with (11) solves the problem for even graphs with A, B labels on the vertices. When the A, B labels are removed one finds that most of the even graphs were counted twice (once for each choice of the A, B labelling). However, those graphs for which the A and B partitions were alike were counted just once. Of course there is only one connected even graph of this sort (the graph with two vertices and all L lines in parallel). Hence the generating series $c(x, y)$ for even graphs is one half the sum of the series for even graphs with A, B labels plus the series

$$\sum_{L=1}^{\infty} 1 \cdot y^2 \frac{x^L}{L!} = y^2 (\exp x - 1).$$

The final result is

THEOREM III. *The number of even graphs with L lines labelled $1, 2, \ldots, L$ and λ unlabelled vertices is $L!$ times the coefficient of $x^v y^\lambda$ in the generating series*

$$\frac{1}{2} \left\{ y^2 (\exp x - 1) + \log \left(1 + \sum_{k=1}^{\infty} T_k(y)\, x^k / k! \right) \right\}$$

where $T_k(y)$ is given by (11).

To count even graphs by lines allowing any number of vertices, set $y = 1$. Then $T_k(1) = G_k^2$ where G_k is the number of ways of grouping k distinct objects into any number of clusters. A table of G_k is given by Bell **(1)**. For $L = 1, 2, \ldots, 5$ there are $1, 2, 8, 60, 672$ even graphs. The 60 even graphs with 4 labelled lines fall into 10 topologically distinct types shown in Figure 2.

A simple enumeration may be given for connected directed graphs allowing lines in parallel and slings. If we drop the connectedness requirement the number of such graphs which have L labelled lines and λ unlabelled vertices
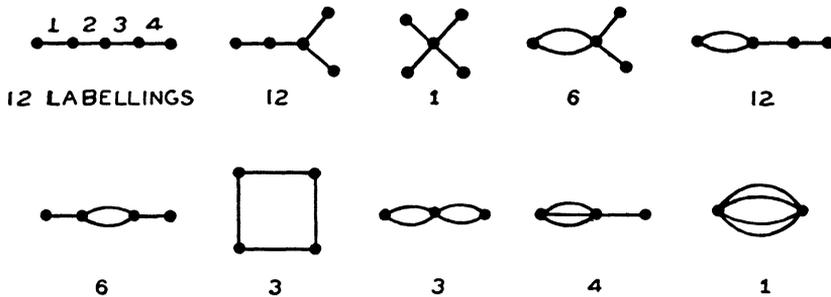

Figure 2

is $S(2L, \lambda)$; for, the vertices represent clusters of end-points taken from the $2L$ distinct end-points of the L directed lines. Theorem I now applies with

$$T_L(y) = \sum_\lambda S(2L, \lambda) y^\lambda.$$

This enumeration can be modified to rule out slings. The number $S(2L, \lambda)$ must now be replaced by the number of groupings of $2L$ distinct end-points into λ clusters such that no cluster contains both end-points of a line. This number is easily found by the principle of inclusion and exclusion (8)

$$T_{L,\lambda} = \sum_k (-1)^k \binom{L}{k} S(2L - k, \lambda).$$

Undirected connected graphs with lines in parallel allowed but with no slings may now be counted by noting that, with one exception, each such graph contributes 2^L directed graphs to the preceding enumeration when the lines are directed in all possible ways. The exception is the graph with two vertices (L lines in parallel), which contributes only 2^{L-1} directed graphs because a simultaneous reversal of all L directions on this graph does not produce a new directed graph.

REFERENCES

1. E. T. Bell, *Exponential polynomials*, Ann. Math., *35* (1934), 258–277.
2. ———, *Postulational bases for the umbral calculus*, Amer. J. Math., *62* (1940), 717–724.
3. R. L. Davis, *The number of structures of finite relations*, Proc. Amer. Math. Soc., *4* (1953), 486–495.
4. F. Harary, *The number of linear, directed, rooted and connected graphs*, Trans. Amer. Math. Soc., *78* (1955), 445–463.
5. C. Jordan, *Calculus of Finite Differences* (New York, 1950).
6. I. Kaplansky and J. Riordan, *Multiple matching and runs by the symbolic method*, Ann. Math. Stat. (3), *16* (1945), 272–277.
7. D. König, *Theorie der endlichen und unendlichen Graphen* (New York, 1950).
8. E. Netto, *Lehrbuch der Combinatorik* (Leipzig, 1901).
9. R. J. Riddell, Jr. and G. E. Uhlenbeck, *On the theory of the virial development of the equation of state of monoatomic gases*, J. Chem. Phys., *21* (1953), 2056–2064.

Bell Telephone Laboratories,
Murray Hill, N.J.