→ 798
1035

Scientific Applications

M. STUART LYNN, Editor

On the Computer Enumeration of Finite Topologies

J. W. EVANS
IBM Scientific Center, Los Angeles, California*
F. HARARY
University of Michigan,† Ann Arbor, Michigan
AND
M. S. LYNN
IBM Scientific Center, Houston, Texas

The problem of enumerating the number of topologies which can be formed from a finite point set is considered both theoretically and computationally. Certain fundamental results are established, leading to an algorithm for enumerating finite topologies, and computed results are given for $n \leqq 7$. An interesting side result of the computational work was the unearthing of a theoretical error which had been induced into the literature; the use of the computer in combinatorics represents, chronologically, an early application, and this side result underscores its continuing usefulness in this area.

It seems to have become an almost classic remark that there are no interesting problems concerning topologies on a finite number of points. To a topologist this may be true; however, from a combinatorial point of view, it is interesting to determine how many different topologies there are on n points.

A word of explanation is in order. There are really two distinct, although related, enumeration problems: either we may consider the points as distinguished (the labeled case), or we may only count the number of homeomorphism classes of topological spaces (the unlabeled case).

Our object is to enumerate the labeled topologies with n points. A finite topology is characterized axiomatically by taking a prescribed collection of the subsets of a set V with n points as open, such that the union and intersection of two open sets are open, as are the empty set and V itself. A "labeled topology" has its points labeled with the integers $1, 2, \cdots, n$. Two labeled topologies are called *homeomorphic* if there is a 1-1 correspondence between their point sets which preserves open sets. By an "unlabeled topology" or just *a topology* is meant a homeomorphism class of labeled topologies.

In this paper, we establish certain fundamental results leading to an algorithm for enumerating finite topologies and give computed results for $n \leq 7$. A side result of this computational work was to unearth an error which had previously appeared in the literature (see section on T_0-Topologies), perhaps underscoring the continuous usefulness of the computer in combinatorics.

* Present address: National Institutes of Health, Bethesda, Md.
† Department of Mathematics. This author's work was supported in part by the U. S. Air Force Office of Scientific Research under grant AF-AFOSR-74-65.

Topologies and Transitive Digraphs

The enumeration of labeled topologies will be formulated with the help of a lemma, anticipated by Krishnamurthy [6], who expressed the observation in terms of matrices. We use the terminology of directed graphs given in [4]. A *labeled digraph* D has its set V of n points labeled with the integers $1, 2, \cdots, n$. As usual, D is *transitive* if whenever the (directed) lines uv and vw are in D, so is uw. Let $Q(v)$ denote the set of all points of D which can reach v along a directed path. Thus when D is transitive, $Q(v)$ is the set of points adjacent to v, and in particular $v \in Q(v)$ since the point v itself may be regarded as a path of length 0 from v to v.

LEMMA 1. *There is a 1-1 correspondence between the labeled topologies with n points and the labeled transitive digraphs with n points.*

PROOF. With any topology T, one can associate a digraph $D(T)$ as follows. The point set of $D(T)$ is that of T. For two distinct points u and v of T, u will be adjacent to v in $D(T)$ provided u is in every neighborhood of v (open set containing v). Clearly $D(T)$ is transitive, and is uniquely determined by T.

We next show that to each labeled transitive digraph D with n points, there corresponds a unique labeled topology. Define $T(D)$ as that topology with the same point set as D, in which the basic open sets are all sets $Q(v)$, $v \in V$. Since D is transitive, $Q(v)$ consists of v and all points adjacent to it. By definition then, every open set in $T(D)$ is of the form $Q(W) = \bigcup Q(w)$, $w \in W$, for some set W of points. We now show that $T(D)$ is indeed a topology. By definition, it is immediate that the union of two open sets is open. It is sufficient to show that the intersection of two open sets is open if we prove it for two basic open sets $Q(v_1)$ and $Q(v_2)$. Consider the set of points v adjacent to both v_1 and v_2. By transitivity, this set is the union of all the sets $Q(v)$ for $v \in Q(v_1) \cap Q(v_2)$, and hence is open.

To establish a 1-1 correspondence, it only remains to observe that $D(T(D)) = D$, which is a direct consequence of the defining constructions.

It is easy to see that this lemma is still valid when we replace the word "labeled" by "unlabeled," or equivalently omit the word "labeled" from the lemma.

A *loop* is a directed line joining a point to itself. By definition [4, p. 9], a digraph has no loops. A relation is *reflexive* if every point has a loop. Obviously it makes no difference to the number of transitive relations whether every point has a loop or no point does, and this holds both for the labeled and unlabeled cases.

COROLLARY. *For any positive integer n, there are equally many topologies, transitive digraphs, and reflexive transitive relations on n points.*

We might remark that it has been pointed out by both Davis [1] and Harary [2] that the enumeration of transitive digraphs is a particularly intractable problem.

T_0-Topologies

We note that, from the defining relations, a labeled topology T is T_0 (i.e., given two distinct points, there exists an open set containing one but not the other) if and only if $D(T)$ is acyclic (has no directed cycles); the same also holds true for the unlabeled case. In this section, we prove our main theorem that expresses the number of labeled topologies on n points in terms of the numbers of T_0-topologies on m points ($m = 1, \cdots, n$).

Let us digress momentarily, however, to note that a transitive digraph is acyclic if and only if it is oriented (asymmetric). It appears to have been indicated in the literature (see [3] for example) that the number of acyclic, oriented, unlabeled digraphs is

$$\frac{1}{n+1}\binom{2n}{n}, \tag{1}$$

this result having been inferred from a result suggested by Wine and Freund [8] and proved by L. Moser. In fact, although the authors of [8] would (in other terminology) appear to have wished to enumerate acyclic, oriented, unlabeled digraphs, they actually enumerated only a subset of these. Thus, for example, the digraph consisting of two disjoint oriented lines would not be included in their enumeration.

This error was first brought to light by the fact that if γ_n is the number of acyclic, oriented, labeled digraphs with n points, then in particular (see Table I), $\gamma_6 > \frac{6!}{7}\binom{12}{6}$, whereas the right-hand side should have been an *upper* bound for γ_6. This emphasizes, perhaps, a familiar and useful aspect of the computer in combinatorial research.

Returning to the main theme of this paper, we first state two lemmas which are given in [4]. For brevity, we will use *transgraph* to mean a transitive digraph.

LEMMA 2. *Every strong component of a transgraph is complete and symmetric.*

TABLE I

n	1	2	3	4	5	6	7
τ_n	1	4	29	355	6942	209 527	9 535 241
γ_n	1	3	19	219	4231	130 023	6 129 859

LEMMA 3. *The condensation of a transgraph with m strong components is itself an acyclic transgraph on m points.*

Now let $C_n^{(m)}$ denote the set of all labeled transgraphs on n points with m strong components, and let us write C_n for $C_n^{(n)}$, the set of all labeled, acyclic transgraphs on n points. Correspondingly, we let C_n' denote the set of all unlabeled, acyclic transgraphs on n points. Let $C' \in C_n'$ be an isomorphism class, containing $|C'|$ labeled transgraphs; we may clearly represent C' by removing the integer labels from any labeled transgraph $C \in C'$. We shall call the process of assigning distinct integer labels to the "points" of C' one of labeling C'; thus there are $|C'|$ ways of doing this. As mentioned in Harary and Read [5], the number $s(C')$ of symmetries (or automorphisms) of C' is in fact given by

$$|C'| = \frac{n!}{s(C')}. \tag{2}$$

Now consider any partition of $V_n = \{1, 2, \cdots, n\}$ into m parts, $m \leqq n$. It is shown in Riordan [7, p. 99] that there are $S(n, m)$ such partitions, where $S(n, m)$ is a Stirling number of the second kind. An *admissible* (m, n) *labeling* of an unlabeled transgraph $C' \in C_m'$ is determined by assigning to each of the m points of C' one of the parts of a given partition of V_n into m parts. Thus, there are $S(n, m)\,|C'|$, different admissible (m, n) labelings of C'. Furthermore, each such admissible (m, n)-labeled C' can be mapped into a unique labeled transgraph $C \in C_n^{(m)}$, C being that labeled transgraph on n points whose condensation is some member of C', and which is labeled so that the labels in the rth strong component (in any order, by virtue of Lemmas 2 and 3) are precisely those integers which appear in that part of the partition of V_n which is assigned to the rth point of C'.

Such a mapping is clearly 1-1 and onto. Hence, if there are $\gamma_n^{(m)}$ transgraphs in the set $C_n^{(m)}$, and $\gamma_m = \gamma_m^{(m)}$ the number of (acyclic) transgraphs in C_m, then

$$\gamma_n^{(m)} = \sum_{C' \in C_m'} S(n, m)\,|C'| = S(n, m) \sum_{C' \in C_m'} |C'| = S(n, m)\gamma_m .$$

We have thus proved the main result.

THEOREM 1. *If τ_n is the number of labeled topologies on n points, and γ_m is the number of labeled acyclic transgraphs on m points, then*

$$\tau_n = \sum_{m=1}^{n} S(n, m)\gamma_m . \tag{3}$$

In the next section we shall indicate how (3) has been used as the basis for an algorithm to compute τ_n for $n \leqq 7$

Algorithm

The algorithm basically consists of inductively computing all the adjacency matrices of the acyclic transgraphs in C_n from those in C_{n-1}, and then using (3) to obtain τ_n. This makes use of the fact (see [4]) that, if $C \in C_n$ and v is any point of C, then the digraph $C - v$ is also transitive and acyclic; that is, $C - v \in C_{n-1}$. We also note that if $C \in C_n$ and $A = A(C)$ denotes the adjacency matrix of

C, then, from the transitivity of C,

$$[(I + A)^2]^{\#} = [I + A]^{\#} \qquad (4)$$

where, as in [3], the notation $[M]^{\#}$ denotes the operation of replacing all nonzero entries of a matrix M by 1. Conversely, it may be seen that if the associated digraph, C, of a null-diagonal, binary matrix, A, of order n satisfying (4) is acyclic, then $C \in C_n$.

The above remarks then form the basis for the following lemma in which β^T denotes the transpose of column vector β.

LEMMA 4. *If a null-diagonal, binary matrix $A = (a_{ij})$ of order n is partitioned in the form*

$$A = \begin{bmatrix} A_1 & \alpha \\ \beta^T & 0 \end{bmatrix}, \qquad (5)$$

where A_1 is a square matrix and α, β are column vectors of order $n-1$, then the associated digraph C of A belongs to C_n if and only if:

(i) *the associated digraph of A_1 belongs to C_{n-1};*

(ii) $a_{ij} = 0 \quad (1 \leqq i, \quad j \leqq n-1)$ *implies* $\alpha_i\beta_j = 0 (1 \leqq i, j \leqq n-1)$;

(iii) $\alpha_i = 0 (1 \leqq i \leqq n-1)$ *implies* $a_{ij}\alpha_j = 0 (1 \leqq j \leqq n-1)$;

(iv) $\beta_j = 0 (1 \leqq j \leqq n-1)$ *implies* $a_{ij}\beta_i = 0 (1 \leqq j \leqq n-1)$;

(v) *D is acyclic.*

Lemma 4 has been directly used to derive an algorithm to compute the adjacency matrices of all transgraphs in C_n for $n \leqq 7$, noting that in bordering matrices such as A_1 in (5), we do not have to consider all possibilities for α and β, but only those with inner product $\alpha^T \cdot \beta = 0$, else we should contradict (v) by producing directed cycles of length 2. We also note that in testing for (v) by the usual process of successively deleting null rows and corresponding columns of A as far as possible, we always attempt to delete the last row at any stage before any other row, since if this ever becomes null, the remaining submatrix is clearly acyclic since its associated digraph is a subgraph of an acyclic digraph by (i).

Essentially using these ideas, the values for τ_n and γ_n shown in Table I have been computed[1] on the IBM 7094 computer of the UCLA Western Data Processing Center. It is to be regretted that γ_n appears to grow at essentially the same rate as τ_n since this prevents further enumeration for reasons of both time and space. In fact, since it can easily be shown that there are

$$\delta(n) = \sum_{s=1}^{n} \binom{n}{s} (2^s - 1)^{n-s} \qquad (6)$$

labeled digraphs on n points containing no path of length greater than 1, and since these are clearly transitive and acyclic, then merely by considering the largest terms in (6),

$$n!\,2^{\binom{n}{2}} \geqq \gamma_n \geqq \delta(n) \geqq \bar{\delta}(n) \qquad (7)$$

[1] The authors would like to thank J. H. Beeman in this connection for his invaluable assistance in programming this algorithm as well as for contributing several ideas to its success.

where

$$\bar{\delta}(n) = \begin{cases} \binom{2k}{k} (2^k - 1)^k + \binom{2k}{k-1} [(2^{k-1} - 1)^{k+1} + (2^{k+1} - 1)^{k-1}], & \text{if } n = 2k, \\ \binom{2k+1}{k} [(2^k - 1)^{k+1} + (2^{k+1} - 1)^k], & \text{if } n = 2k+1. \end{cases}$$

The left-hand inequality in (7) is a direct consequence of the fact that the adjacency matrix of an acyclic digraph may be permuted into upper triangular form; see [4].

Since $\bar{\delta}(n)$ is asymptotically dominated by terms of order $2^{n^2/4}$, it might serve as a caution to those who would calculate further numbers τ_n, using algorithms of the kind discussed in this section. We note that the values of τ_n for $n = 3$ were obtained by Krishnamurthy [6] using a different algorithm.

To verify directly that $\gamma_3 = 19$ and $\tau_3 = 29$, we may apply eq. (2) to the five unlabeled acyclic transgraphs D_i on 3 points shown in Figure 1 and also to the four nonacyclic transgraphs D_6, D_7, D_8, D_9 on 3 points in Figure 2. This information is summarized in Table II, in which the sum of the first 5 entries in the last row is $\gamma_3 = 19$ and the total row sum is $\tau_3 = 29$.

In actual fact, (6) and (7) are, with $s = 2$, a special case of an easily derivable result which we state without proof; namely,

$$\gamma_n \geqq \sum \frac{n!}{a_1!\,a_2!\cdots a_s!} \prod_{i=1}^{s} (2^{a_i} - 1)^{a_{i+1}}, \qquad (8)$$

where the sum is taken over all compositions (ordered partitions) of n.

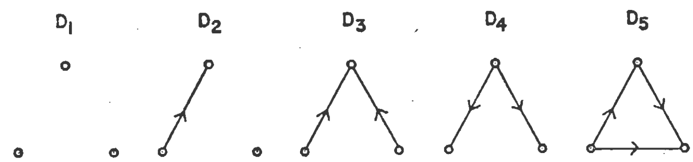

FIG. 1.

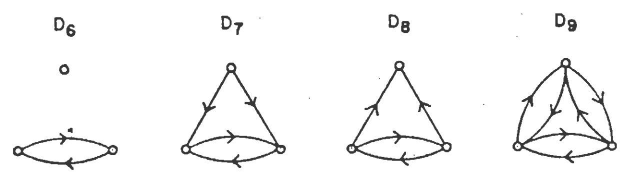

FIG. 2

TABLE II

Digraph D	D_1	D_2	D_3	D_4	D_5	D_6	D_7	D_8	D_9
Symmetry number $s(d)$	6	1	2	2	1	2	2	2	6
Number of labeling $6/s(D)$	1	6	3	3	6	3	3	3	1

RECEIVED SEPTEMBER, 1966

REFERENCES

1. DAVIS, R. L. The number of structures of finite relations. *Proc. Am. Math. Soc.* *4* (1953), 486–495.

(*Continued on page 313*)

TABLE XII. Examples of Summed Coincidence Matrices

7 × 7 (1917 cases)

1860	0	56	69	79	47	178
1	1482	252	257	30	291	6
32	231	1543	81	115	158	44
68	50	90	1262	70	85	30
281	36	152	100	1172	34	63
94	73	101	51	77	997	383
101	135	31	58	54	317	1402

9 × 8 (2676 cases)

2562	21	72	226	429	344	323	192	214
1	2359	509	343	151	132	411	141	86
78	203	1943	615	309	58	201	175	58
236	213	152	1510	775	144	162	133	84
386	124	205	58	1280	1279	487	222	126
91	519	206	105	141	1072	1666	115	26
217	110	166	82	313	102	417	2129	138
128	134	10	231	51	9	33	298	2271

10 × 8 (737 cases)

569	81	36	43	56	34	46	39	33	57
1	506	155	89	28	39	49	72	36	92
41	54	442	203	100	53	16	55	15	88
29	50	25	270	144	262	40	167	55	42
48	19	60	48	218	128	309	27	23	52
44	59	63	47	55	277	87	323	51	17
22	46	17	57	83	21	164	85	457	20
45	27	181	54	13	70	38	75	55	552

In an attempt to derive a weighting function which would reflect the true probability of a one (1) occurring in the (i, j)-position of an $m \times n$ coincidence matrix, a third set of tabulations was made. The coincidence matrices were formed for all word-misspelling pairs. The elements were multiplied by the frequency of occurrence of the pair, and sums were formed of all matrices having equal dimensions. Examples of these summed matrices are presented in Table XII. The columns correspond to letters in the correctly spelled word, the rows to letters in the misspelling. With the exception of the regularly large values along the axis of the matrix, there is little regularity. This can probably be accounted for by the relatively small number of different correctly spelled words used in the study. If the elements of the matrix are taken to be values of a function over the x,y-plane, a contour map of the function may be drawn. Such maps have been constructed for a number of the matrices. They reveal a rather vague tendency for "ridges" and "valleys" to follow diagonals. This phenomenon has been tentatively ascribed to the distribution of vowels in the words. However, no unequivocal evidence is available at this time.

The irregularity noted above precluded the use of these tabulations in determining a probabilistic weighting function.

Received November, 1966

REFERENCES

1. Baskin, H. IBM Research Center, private communications.
2. Savin, H. B. U. of Pennsylvania, private communications.
3. Selfridge, L. D. IBM Research Center, private communications.
4. Bethel, J. P. (Ed.) A guide to pronunciation. In *Webster's New Collegiate Dictionary*, G. &. C. Merriam Co., Springfield, Mass., 1960.
5. Blair, C. R. A program for correcting spelling errors. *Inform. Contr. 3*, (1960), 60–67.
6. Cochran, W. G. *Sampling Techniques*. John Wiley, Inc., New York, 1953, 160–187.
7. Conlin, D. A. *Grammar for Written English*. Houghton Mifflin Co., Boston, 1961.
8. Damerau, F. A technique for computer detection and correction of spelling errors. *Comm. ACM 7*, 3 (Mar. 1964), 171–176.
9. Davidson, L. Retrieval of misspelled names in an airlines passenger record system. *Comm. ACM 5*, 3 (Mar. 1962), 169–171.
10. Faulk, R. D. An inductive approach to language translation. *Comm. ACM 7*, 11 (Nov. 1964), 647–653.
11. Glantz, H. T. On the recognition of information with a digital computer. *J. ACM 4*, (1957), 178–188.
12. Kennedy, A. G. *Current English*. Ginn and Co., Boston, 1935, pp. 166–262.
13. Leutenegger, R. R. *The Sounds of American English*. Scott, Foresman and Co., Chicago, 1963.
14. Masters, H. V. A study of spelling errors. Ph.D. Thesis. U. of Iowa, 1927, unpublished.
15. Mood, A. McF. *Introduction to the Theory of Statistics*. McGraw Hill, New York, 1950, 388–389.
16. Morrison, H. W. Computer processing of responses in verbal training. IBM Research Rep. RC 1133.
17. Neilson, W. A. (Ed.) *A Guide to Pronunciation*. G. & C. Merriam Co., Springfield, Mass., 1935.
18. Pearson, E. S., and Hartley, H. O. *Biometrika Tables for Statisticians*. Cambridge U. Press, Cambridge, 1958, pp. 61–62.
19. Thorelli, L. E. Automatic correction of errors in text. *BIT 2*, (1962), 45–62.
20. Yum, K. S. An experimental test of the law of assimilation. *J. Exp. Psychol. 14*, (1931), 73–74.

EVANS, Et Al.——cont'd from p. 297

2. Harary, F. Unsolved problems in the enumeration of graphs. *Publ. Math. Inst. Hung. Acad. Sci. 5* (1960), 63–95.
3. Harary, F. Combinatorial problems in graphical enumeration. In E. F. Beckenbach (Ed.), *Applied Combinatorial Mathematics*, Wiley, New York, 1964.
4. Harary, F., Norman, R. Z., and Cartwright, D. *Structural Models: An Introduction to the Theory of Directed Graphs*. Wiley, New York, 1965.
5. Harary, F., and Read, R. C. The probability of a given 1-choice structure. *Psychometrika* (1966).
6. Krishnamurthy, V. On the number of topologies on a finite set. *Am. Math. Monthly 73* (1966), 154–157.
7. Riordan, J. *An Introduction to Combinatorial Analysis*. Wiley, New York, 1958.
8. Wine, R. L., and Freund, J. E. On the enumeration of decision patterns involving n means. *Ann. Math. Stat. 28* (1957), 256–259.