HOW MANY STEMMATA?

COLIN FLIGHT
University of Birmingham

ABSTRACT

Using the language of graph theory, this paper solves a problem posed by Maas (1958): for some given number of surviving manuscripts, how many different stemmata may exist? With two surviving manuscripts, there are three possible stemmata; with three, there are twenty-two; and the number increases rapidly after that. With ten surviving manuscripts, granted the same assumptions made by Maas, there are more than 100 billion possible stemmata.

Though not of any particular interest for mathematicians, the combinatorial problem discussed in this note is one which arises naturally in the discipline of textual criticism (Greg 1927; Maas 1958).

Suppose that a text is reproduced by means of manuscript copies, each copy becoming in its turn a potential exemplar for others. Provided that every copy derives from a single exemplar (which ceases to be true, for instance, if manuscripts are collated, with readings from one being copied across to another), the textual tradition assumes the form of a tree, as in Figure 1. This is a stemma of the simplest possible kind.

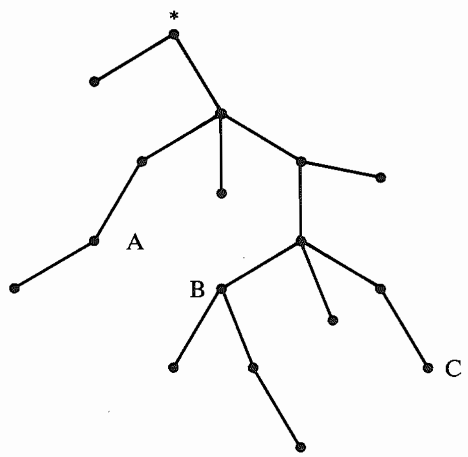


Figure 1. A tree of manuscripts. The vertices (black dots) denote manuscripts; the edges (lines) denote transcription events. A star (*) marks the original. The vertices labelled A, B, C are the manuscripts assumed to survive in Figure 2.

Because transcription is never perfect, some number of textual variants originates with each new copy; and these variants are then transmitted to all descendant copies, except where they become overwritten by subsequent variants. Thus the shape of the tree expresses itself in the distribution of variants among the different manuscripts.

In principle, if all the copies which ever existed survive, it ought to be possible for us, given the textual evidence alone, to reconstruct the tree. Usually, out of all the copies ever made, only some are extant; and in these circumstances the stemma we construct will represent a more or less simplified version of the actual tree, as in Figure 2.

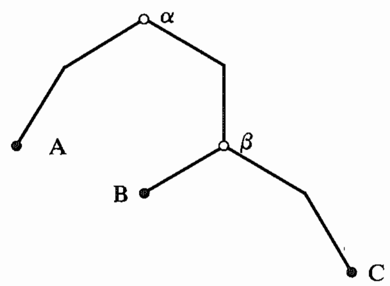


Figure 2. A simple stemma. Some vertices (distinguished as white dots and labelled arbitrarily for reference) denote hypothetical manuscripts; α is the archetype.

Granted some rather special conditions, it is possible to prove the former existence of manuscripts no longer surviving. Thus, in Figure 2, since some but not all of the variants found in **B** will also be found in **C**, and since some but not all of the variants found in **C** will also be found in **B**, we are in a position to prove the existence of β. (It remains to be proved, however, that β is distinct from α, and that α is distinct from **A**.) On the other hand, it will not be possible for us to prove that **C** was copied from a copy of β, rather than from β itself.

When we set about constructing a stemma, we have to rely repeatedly on the principle of exclusion. At each stage we are faced with a number of alternative possibilities; and we then proceed by eliminating as many of these alternatives as we can, until ideally only one is left. Arguments of this type can easily go astray—and often do—through some of the alternatives being overlooked. Hence the question arises: given some number of surviving manuscripts, how many possible stemmata do we have to consider?

This question was asked by Maas (1958); but he provided only a partial answer. (In my notation the values reported by Maas are $G^*(2) = 3$, $G^*(3) = 22$, $G^*(4) = 250$, which is wrong, and $G^*(5) \approx 4000$, which is an underestimate.) The purpose of this note is to present a complete solution for Maas's problem.

Definitions

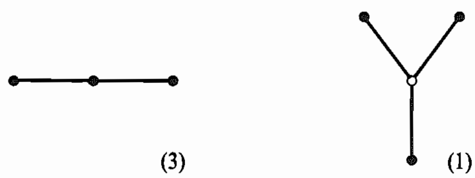
The objects we aim to enumerate are trees, in the sense in which this term is used by graph theorists—for example, by Wilson (1979)—but they are trees of a peculiar kind. I propose to call them rooted Greg trees, in tribute to the textual critic W. W. Greg.

A Greg tree with parameters m, n—informally, an (m, n) tree—is a tree containing m labelled vertices (surviving manuscripts) and n unlabelled vertices (hypothetical manuscripts), the latter required to be of degree at least three.

A rooted Greg tree is a Greg tree with one distinguished vertex, called the root. (This represents the archetype, the point on the tree where we come closest to the original.) An unlabelled vertex forming the root is allowed to be of degree two.

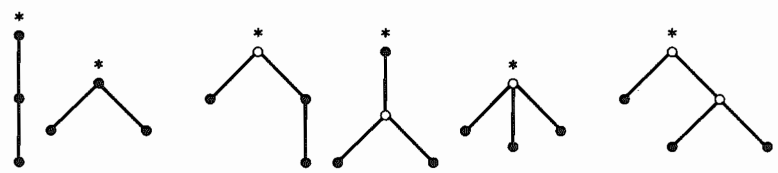
With m = 3, the possible topologies for a Greg tree are as shown in Figure 3, for a rooted Greg tree as in Figure 4. The bracketed numbers denote the number of labellings for each topology.

A stemma of the simplest kind is the same thing as a rooted Greg tree.



(3) (1)

Figure 3. Greg tree topologies for three surviving manuscripts. The bracketed figure denotes the number of labellings for each topology.



(6) (3) (6) (3) (1) (3)

Figure 4. Rooted Greg tree topologies for three surviving manuscripts. Here again, the bracketed figure denotes the number of labellings for each topology. The archetype is marked with a star.

Results

It is easy to see that a Greg tree must have $n \leq m - 2$. In any tree, we can calculate the number of edges in two different ways: by counting the vertices and subtracting one, or by summing the vertex degrees and dividing by two (since every edge is being counted twice). In a Greg tree, by definition, the unlabelled vertices have degree at least three; so the number of edges is at least $(m + 3n)/2$. Hence

$$m + n - 1 = \text{number of edges} \geq (m + 3n)/2,$$

from which it follows that n cannot be greater than $m - 2$.

Let $g(m, n)$ be the number of (m, n) trees, $m \geq 2$, $0 \leq n \leq m - 2$. Obviously, $g(2, 0) = 1$. To find an element in the m-th row, $m > 2$, we take up to three elements from the $(m-1)$-th row, multiply each by an appropriate factor, and then add the answers together. We have

$$g(m, n) = \alpha \cdot g(m-1, n-1) + [\beta + \gamma] \cdot g(m-1, n) + \delta \cdot g(m-1, n+1),$$

where α, β, γ, δ are the factors defined below.

Figure 5. All possible ways of attaching a new vertex to a Greg tree.

We justify this formula by considering the number of ways in which a new labelled vertex (a newly discovered manuscript) might be connected to some existing tree. There are four possible cases, as shown in Figure 5. The new vertex may be

(1) attached to a new unlabelled vertex inserted on any edge in an $(m-1, n-1)$ tree: thus $\alpha = m + n - 3$, the number of existing edges in any such tree;
(2) attached to any vertex (labelled or not) in an $(m-1, n)$ tree: thus $\beta = m + n - 1$;
(3) inserted on any edge in an $(m-1, n)$ tree: thus $\gamma = m + n - 2$;
(4) substituted for any unlabelled vertex in an $(m-1, n+1)$ tree: thus $\delta = n + 1$.

This last case is the surprising one. It corresponds to what would happen if the newly discovered manuscript turned out to be identical with a manuscript whose existence had already been inferred.

By means of this formula, we obtain the triangular array of which the first few rows are shown in Table I. Note that we have

$$g(m, 0) = m^{m-2},$$

which is simply the number of trees with m labelled vertices, and

$$g(m, m-2) = 1 \cdot 3 \cdot 5 \cdot \ldots \cdot (2m - 5),$$

which is the number of trees with vertices all of degree either one or three, the terminal vertices only being labelled. These results are both well known, the first being specially famous (Moon 1967).

The row totals,

$$G(m) = \sum_{n=0}^{m-2} g(m,n),$$

then give the number of (unrooted) Greg trees for m surviving manuscripts.

	n = 0	1	2	3	G(m)
m = 2	1				1
3	3	1			4
4	16	13	3		32
5	125	171	85	15	396

Table I. Values of g(m,n).

In a rooted Greg tree, where one of the unlabelled vertices may be of degree two, the number of edges is at least $(m + 3n - 1)/2$. Hence, by the same sort of argument as before, we must have $n \leq m - 1$.

Let $g^*(m, n)$ be the number of rooted (m, n) trees, $m \geq 1$, $0 \leq n \leq m - 1$. Then $g^*(1, 0) = 1$. For $m > 1$, we find each element in the m-th row by using a formula similar to the one above. We have

$$\begin{aligned} g^*(m, n) = [\alpha + 1]\cdot g^*(m-1, n-1) \\ + [\beta + \gamma + 1]\cdot g^*(m-1, n) \\ + \delta\cdot g^*(m-1, n+1), \end{aligned}$$

where the additional possibilities are as shown in Figure 6.

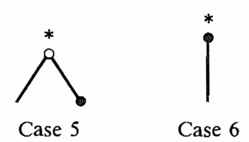


Figure 6. Additional ways of attaching a new vertex to a rooted Greg tree.

This gives us another triangular array, the first few rows of which are shown in Table II. Again, note that we have $g^*(m, 0) = m^{m-1}$, and $g^*(m, m-1) = 1 \cdot 3 \cdot 5 \cdot \ldots \cdot (2m - 3)$.

The row totals,

$$G^*(m) = \sum_{n=0}^{m-1} g^*(m,n),$$

give the number of rooted Greg trees for m surviving manuscripts. That answers Maas's question.

	n = 0	1	2	3	4	G*(m)
m = 1	1					1
2	2	1				3
3	9	10	3			22
4	64	113	70	15		262
5	625	1526	1450	630	105	4336

Table II. Values of g*(m,n).

Given any Greg tree, we can turn it into a rooted Greg tree by placing the root either (1) on any existing vertex or (2) on a new unlabelled vertex inserted for the purpose on any existing edge. Thus there are 2m + 2n − 1 rooted Greg trees corresponding to any (m, n) tree. Hence

$$G^*(m) = \sum_{n=0}^{m-2} [2m + 2n - 1]\cdot g(m,n).$$

This means that we can find both G(m) and G*(m) from Table I alone, ignoring Table II.

Computed values for G(m) and G*(m), $m \leq 12$, are tabulated in Table III. Neither sequence is listed by Sloane (1973); but it seems unlikely that these numbers are significant only in the context of one particular discipline. For any kind of object, and for any kind of relationship, if there is some presumption in favour of a tree-shaped structure, G(m) is the number of hypotheses covering m such objects. Each edge in the tree corresponds to a class of statements which are true for all the objects on one side, false for all those on the other.

m	G(m)	G*(m)
1	1	1
2	1	3
3	4	22
4	32	262
5	396	4 336
6	6 692	91 984
7	143 816	2 381 408
8	3 756 104	72 800 928
9	115 553 024	2 566 606 784
10	4 093 236 352	102 515 201 984
11	164 098 040 448	4 575 271 116 032
12	7 345 463 787 136	225 649 908 491 264

Table III. Values of G(m) and G*(m)

REFERENCES

Greg, Walter W. 1927. *The Calculus of Variants: An Essay on Textual Criticism.* Oxford: Clarendon Press.

Maas, Paul. 1958. *Textual Criticism.* Translated from the German by Barbara Flower. Oxford: Clarendon Press.

Moon, John W. 1967. "Various Proofs of Cayley's Formula for Counting Trees." In *A Seminar on Graph Theory*, ed. by Frank Harary with Lowell Beineke, 70-78. New York [etc]: Holt, Rinehart and Winston.

Sloane, Neil J. A. 1973. *A Handbook of Integer Sequences.* New York: Academic Press.

Wilson, R. J. 1979. *Introduction to Graph Theory.* 2nd ed. New York: Academic Press.