ADVANCES IN APPLIED MATHEMATICS **10**, 488–496 (1989)

Applications of Antilexicographic Order. I. An Enumerative Theory of Trees

PÉTER L. ERDŐS

Mathematical Institute of the Hungarian Academy of Sciences, H-1364 Budapest, P.O.B. 127, Hungary

AND

L. A. SZÉKELY*

Eötvös University, H-1088 Budapest, Muzeum krt. 6-8, Hungary and University of New Mexico, Albuquerque, New Mexico 87131

Making a bijection between semilabelled trees and some partitions, we build up a powerful theory for enumeration of trees. Theorems of Cayley, Menon, Clarke, Rényi, Erdélyi—Etherington are among the consequences. The theory of random semilabelled trees turns into the theory of random set partitions. © 1989 Academic Press, Inc.

The motivation for the present research was the perplexing fact that the semifactorial function arises as the solution in both of the following enumeration problems: *What is the number of complete matchings in the complete graph K_{2n}?* and *What is the number of rooted binary trees with n labelled leaves*? These results are folklore, the second author learned the second one from his co-authors of paper [1]. We succeeded in establishing a general bijection between some sets of trees and some sets of partitions. In this way, instead of enumeration of trees, we enumerate partitions. In most cases, this is easier to do. In particular, we solve some new enumeration problems for trees and give new derivation of a number of classical theorems, including Cayley's theorem. The main tool is the antilexicographic order of subsets of an ordered set, which was introduced independently by Kruskal [7] and Katona [5]. We emphasize that our bijection is well computable in both directions. For the theory of enumeration of trees, see [6, 10].

*Research supported in part by OTKA 327 0113.

0196-8858/89 $7.50

DEFINITION. Let X be an ordered set. The *antilexicographic order* $<_{\mathrm{AL}}$ on the power set of X is defined as follows:

$$A <_{\mathrm{AL}} B \Leftrightarrow \max(A \Delta B) = \max\{(A \setminus B) \cup (B \setminus A)\} \in B.$$

DEFINITION. A *semilabelled tree* is a rooted tree on $n + 1$ vertices with k leaves. The leaves are labelled with the numbers $1, 2, \dots, k$. (Although the root may have degree one, it is never to be considered as a leaf.) We refer to the vertices differing from leaves and root as *branching points*. They are unlabelled. The *descendants* of a vertex are the neighbours of it, except the (possible) one in the root direction. The *out-degree* of a vertex is the number of its descendants, i.e., the number of neighbours, except the one in the root direction. The out-degree of the root is the degree of the root, the out-degree of other points is the degree minus one.

THEOREM 1. *The set of semilabelled trees on $n + 1$ vertices with k leaves are in a one-to-one correspondence with the set of partitions of $\{1, 2, \dots, n\}$ into $n - k + 1$ non-empty classes. Hence, the number of semilabelled trees above is the Stirling number of second kind $S(n, n - k + 1)$; and the total number of semilabelled trees on $n + 1$ vertices is the Bell number B_n, the total number of partitions of an n-element set. Furthermore, under this one-to-one correspondence, semilabelled trees with root degree t are in one-to-one correspondence with partitions in which n belongs to a t-element class. The out-degrees of the branching points and root are the cardinalities of the classes.*

Proof. For convenience, we make the bijection between semilabelled trees with labels of an ordered k-element set and partitions of an ordered n-element set. Let the label set of the leaves of the semilabelled trees be $\{x_1, x_2, \dots, x_k\}$ and the set to be partitioned $\{y_1, y_2, \dots, y_n\}$. Both sets are ordered by the subscripts.

We construct a partition for a given tree. We call this map ϕ. Let T denote the tree, $V(T)$ denote its vertex set and r denote the root vertex. We assign to every vertex $v \in V(T)$, $v \neq r$ a set $S(v) \subset \{x_1, x_2, \dots, x_k\}$ as

$$S(v) = \begin{cases} \{v\}, & \text{for } v \in \{x_1, x_2, \dots, x_k\}; \\ \text{the set of leaves separated} & \\ \text{from } r \text{ by } v & \text{otherwise.} \end{cases}$$

It may happen, that some sets $S(v)$ occur with multiplicity, if some branching points have out-degree one. Therefore we use $\leq_{\mathrm{AL}}$ instead of $<_{\mathrm{AL}}$; put $\{S(v)\colon v \in V(T) \setminus \{r\}\}$ into antilexicographic order keeping the following extra property: if $S(v) = S(w)$, and v is closer to the root than w, then $S(v)$ is bigger in the antilexicographic order than $S(w)$. Now we give a partition of $\{y_1, y_2, \dots, y_n\}$ into $n - k + 1$ classes: let rank(v)

denote the place of $S(v)$ in the antilexicographic order $\leq_{AL}$ on $\{S(v): v \in V(T) \setminus \{r\}\}$, so we have $1 \leq \text{rank}(v) \leq n$. The partition classes will be of the following form: $\{y_{\text{rank}(v)}: v \in W\}$, where W is any of the (non-empty) sets of descendants in the tree. We shall separate two steps in the algorithm above: ϕ_1 assigns the ranks as labels to the vertices of the semilabelled tree and ϕ_2 reads the partition classes from the result of ϕ_1. Obviously we have $\phi = \phi_2\phi_1$.

We construct the inverse of our bijection, $\psi = \psi_2\psi_1$, which assigns a semilabelled tree with labels $\{x_1, x_2, \dots, x_k\}$ to a given partition $\mathscr{P}$ on $\{y_1, y_2, \dots, y_n\}$. The first step, ψ_1 makes a rooted tree with n other labelled vertices. Add a root to $\{y_1, y_2, \dots, y_n\}$ in order to get the vertex set, join the elements of the class of y_n to the root and join the elements of the class P to $1 + \max P$. In the second step, ψ_2 deletes the labels of branching points and changes the labels of leaves for $\{x_1, x_2, \dots, x_k\}$ keeping their order.

We are going to prove, that ϕ is bijection and $\phi\psi$ fixes the partitions. We make mathematical induction for $|\mathscr{P}|$ and an inner mathematical induction for $|\cup \mathscr{P}|$ for a given $|\mathscr{P}|$. All the statements are obvious for $|\mathscr{P}| = 1$.

Claim. If $\phi(F) = \mathscr{P}$ and P is the antilexicographically smallest element in $\mathscr{P}$, then $\max\{i: y_i \in P\} \leq k$ and $\{x_i: y_i \in P\}$ is the set of descendants of some vertex in F.

The claim holds, since on a path from a leaf to the root the ranks are increasing and $\text{rank}(x_i) > i$ iff there is a complete set of descendants $\{x_j: j \in J\}$ with $\max J < i$.

The claim implies that ϕ is injection. For the contrary, suppose the trees F and G are mapped to the same partition $\mathscr{P}$. Truncate the trees by deleting the vertices $\{x_i: y_i \in P\}$ and label their ancestor by $\max P$. The new trees are F' and G'. It is easy to see, that $\phi(F') = \mathscr{P} \setminus P = \phi(G')$. By hypothesis, $F' = G'$ and $F = G$ is implied.

We prove that ϕ is onto. Let us be given a partition $\mathscr{P}$ and take P as above. By hypothesis, $\mathscr{P} \setminus P = \phi(T)$ for some semilabelled tree T with label set

$$\{x_i: i \in \{1, 2, \dots, k\} \setminus P \cup \{\max P\}\}.$$

Let T' denote the following tree: delete the label $x_{\max P}$ in T and join the vertices $\{x_i: y_i \in P\}$ to this delabelled vertex. It is easy to check that $\phi(T') = \mathscr{P}$. (This step uses the second induction as well.)

We have to prove $\phi\psi = \text{id}$. It is easy to see, that $\phi_1\psi_2 = \text{id}$ and $\phi_2\psi_1 = \text{id}$. This is enough for us, since $\phi\psi = \phi_2\phi_1\psi_2\psi_1$.

The root degree is equal to the cardinality of the partition class containing n by the description of ψ. $\square$

Under the bijection above, some special trees are in correspondence to special partitions. From now on we make use of the well-known elementary fact that the number of partitions of an n-element set into k_i i-element classes ($i = 1, 2, \ldots, n$) is

$$\frac{n!}{1!^{k_1}k_1!2!^{k_2}k_2! \cdots n!^{k_n}k_n!}.$$

A binary tree with k leaves has $k - 2$ branching points and $2k - 1$ vertices in total, and we have

COROLLARY 2. *The semilabelled binary trees with k leaves are in one-to-one correspondence with the partitions of a $2k - 2$ element set into 2-element subsets, and their number is $(2k - 3)!!$.* □

In a more general setting, we have

COROLLARY 3. *The semilabelled t-ary trees with m branching points are in one-to-one correspondence with the partitions of a $(m + 1)t$- element set into t-element classes, and their number is $((m + 1)t)!/(t!^{m+1}(m + 1)!)$.*

COROLLARY 4. *The number of semilabelled trees on $n + 1$ vertices with maximum degree at most d are in one-to-one correspondence with the partitions of an n-element set into classes not bigger than d.*

COROLLARY 5 (Knuth [6], Poupard [11]). *The number of rooted labelled trees on $n + 1$ vertices with k leaves is $n(n - 1) \cdots (k + 1)S(n, n - k + 1)$.*

Proof. We select k labels out of n in $\binom{n}{k}$ ways for the leaves, build up $S(n, n - k + 1)$ semilabelled trees for each by Theorem 1, and distribute the remaining $(n - k)$ labels for the branching points in $(n - k)!$ ways. □

COROLLARY 6 (Rényi [12]). *The number of labelled trees on $n + 1$ vertices with k leaves is $(n + 1)n(n - 1) \cdots (k + 1)S(n - 1, n - k + 1)$.*

Proof. The number of rooted semilabelled trees on $n + 1$ vertices with k leaves is $S(n, n - k + 1)$, with $S(n - 1, n - k)$ having root degree 1. (We realize them by the class $\{n\}$ in the corresponding partition.) Therefore, the number of rooted semilabelled trees on $n + 1$ vertices with k leaves with root degree at least two is

$$S(n, n - k + 1) - S(n - 1, n - k) = (n - k + 1)S(n - 1, n - k + 1).$$

We label the branching points of these trees in $(n - k)!$ ways, and there are $\binom{n}{k}$ choices for the set of labels of leaves. We have that the number of labelled trees on $n + 1$ vertices with k leaves and with a special label whose degree is at least two is

$$f = \frac{n!}{k!}(n-k+1)S(n-1, n-k+1).$$

Count the ordered pairs $\langle T, v\rangle$, where T is a labelled tree on $n+1$ vertices with k leaves and v is a vertex of it which is not a leaf. On the one hand, the number of pairs is $n-k+1$ times the quantity looked for; on the other hand, the number of pairs is $(n+1)f$. □

We give a new proof of Cayley's theorem. We need the Faá di Bruno formula [6]: Let $D_x^k u$ represent the k^{th} derivative of a function u with respect to x. Then

$$D_x^n w = \sum_{0\le j\le n} \sum_{\sum ik_i = n} D_u^j w \frac{n!}{1!^{k_1}k_1!2!^{k_2}k_2!\dots n!^{k_n}k_n!} \times \left(D_x^1 u\right)^{k_1} \cdots \left(D_x^n u\right)^{k_n},$$

where the inner summation is taken for non-negative integer values of k_i with $\sum k_i = j$.

THEOREM 7 (Cayley [2]). *The number of labelled trees on n vertices is* n^{n-2}.

Proof. We prove the following equivalent of Cayley's theorem: the number of rooted trees with n labelled non-root vertices (i.e., with $n+1$ vertices in total) is $(n+1)^{n-1}$. Different partitions give rise to different semilabelled trees. How many different labelled trees come from a given partition, which contains k_i i-element classes? The corresponding semilabelled tree has $\sum k_i - 1$ branching points. The set of names of leaves can be selected in $\binom{n}{\sum k_i - 1}$ ways, we can name the branching points in $(\sum k_i - 1)!$ ways. Therefore, the number of rooted trees with n labelled non-root vertices is

$$\sum_{n=\sum ik_i} \frac{n!}{1!^{k_1}k_1!2!^{k_2}k_2!\dots n!^{k_n}k_n!}\binom{n}{\sum k_i - 1}\left(\sum k_i - 1\right)!,$$

where the summation is taken by $k_i \ge 0$. We recall the Faá di Bruno formula with the following cast: $u = e^x$, $w = e^{(n+1)x}$ and we evaluate the derivative at $x = 0$. We have

$$(n+1)^n = \sum_{j=\sum k_i} \sum_{n=\sum ik_i} \frac{n!}{1!^{k_1}k_1!2!^{k_2}k_2!\dots n!^{k_n}k_n!} \times (n+1)n(n-1)\cdots(n-j+2)$$

and we divide both sides by $(n+1)$. □

THEOREM 8 (Clarke [3]). *The number of rooted labelled trees on $n+1$ vertices with a root degree k is equal to $\binom{n-1}{k-1}n^{n-k}$.*

Proof. We claim that the number of corresponding semilabelled trees is

$$\binom{n-1}{k-1} \sum_{n-k=\sum ik_i} \frac{(n-k)!}{1!^{k_1}k_1!2!^{k_2}k_2!\dots(n-k)!^{k_{n-k}}k_{n-k}!}.$$

We have $\binom{n}{\sum k_i - 1}(\sum k_i - 1)!$ times more labelled trees for a given sequence k_i, similarly to the proof of Cayley's theorem. Applying the Faá di Bruno formula for $w = e^{nx}$ and $u = e^x$, to compute the $(n-k)^{\text{th}}$ derivative at $x = 0$ we get

$$\binom{n-1}{k-1}n^{n-k} = \binom{n-1}{k-1} \sum_{n-k=\sum ik_i} \frac{(n-k)!\binom{n}{\sum k_i - 1}\left(\sum k_i - 1\right)!}{1!^{k_1}k_1!2!^{k_2}k_2!\dots(n-k)!^{k_{n-k}}k_{n-k}!}.$$

The formula claimed holds, since the neighbours of the root make the antilexicographically greatest partition class, which is the class containing the number n. We have $\binom{n-1}{k-1}$ choices for this class and we find

$$\sum_{n-k=\sum ik_i} \frac{(n-k)!}{1!^{k_1}k_1!2!^{k_2}k_2!\dots(n-k)!^{k_{n-k}}k_{n-k}!}$$

partitions containing the given k-element class. □

THEOREM 9 (Menon [9]). *Suppose $\sum_{i=0}^n d_i = 2n$. The number of labelled trees on $n+1$ points with prescribed degrees $d_0, d_1, \dots, d_n$ is equal to*

$$\frac{(n-1)!}{(d_0-1)!\cdots(d_n-1)!}.$$

Proof. Let k_i denote the number of occurrences of $i + 1$ in the sequence $d_1, \ldots, d_n$. In the first step we enumerate the semilabelled trees on $n + 1$ vertices with root degree d_0 and with k_0 labelled leaves. Among the corresponding partitions, the neighbours of the root make the antilexicographically greatest partition class, which contains the number n. We have $\binom{n-1}{d_0-1}$ choices for this class and we can build up

$$\frac{(n-d_0)!}{1!^{k_1}k_1!2!^{k_2}k_2!\ldots(n-d_0)!^{k_{n-d_0}}k_{n-d_0}!}$$

partitions containing such a given class. Finally, there are $k_1!k_2! \cdots k_{n-d_0}!$ ways to match the branching points and their possible names. Multiplying these numbers together, we have the formula claimed. □

Remark. As it is well known, the Menon theorem implies the Cayley, Clarke, and Rényi theorems by the polynomial theorem. In this way we have an alternative proof for them, which avoids the Faá di Bruno formula.

THEOREM 10 (Erdélyi-Etherington [4]). *Suppose $n_0 = 1 + n_2 + 2n_3 + 3n_4 + \cdots + (m-1)n_m$. The number of rooted, unlabelled, ordered trees (i.e., in which we order the subtrees hanging on the same point) having n_i vertices with out-degree i ($i = 0, 1, \ldots, m$) is equal to*

$$\frac{(n_0 + n_1 + \cdots + n_m - 1)!}{n_0!n_1! \cdots n_m!}.$$

Proof. The number of rooted semilabelled trees with the prescribed out-degrees is

$$\frac{(n_0 + n_1 + \cdots + n_m - 1)!}{n_1!1!^{n_1} \cdots n_m!m!^{n_m}}$$

by Theorem 1. In order to get the number of rooted ordered semilabelled trees we multiply by $1!^{n_1}2!^{n_2} \cdots m!^{n_m}$, since so many ways can we order the subtrees. Finally, we divide by $n_0!$, since any rooted ordered tree has $n_0!$ semilabelling. □

DEFINITION. A rooted binary plane tree is a rooted unlabelled tree, in which each degree is 1 or 3, the root degree is 1, two plane trees are the "same," if there is an isomorphism between them, which keeps the root and the cyclic ordering of edges in the star of each point.

COROLLARY 11 (Lovász [8]). *The number of rooted unlabelled binary plane trees on $2n$ vertices is the Catalan number*

$$\frac{1}{n}\binom{2n-2}{n-1}.$$

Proof. Delete the root and consider its neighbour as a new root. The deleted root defines an order of the two subtrees of the new root.

Being the tree rooted binary, the sequence n_i is determined and the Erdélyi–Etherington theorem may be applied with $n_0 = n$, $n_2 = n - 1$. Alternatively, Corollary 2 can be also applied. Consider the semilabelled binary trees on $2n - 1$ vertices, they are $(2n - 3)!!$ in number. Multiply by 2^{n-1} for the order of subtrees and divide by $n!$ to lose the labels of leaves. □

Choosing a proper definition for random semilabelled trees, under Theorem 1 the theory of random partitions turns into the theory of random semilabelled trees.

DEFINITION. A random semilabelled tree on $n + 1$ vertices is a randomly selected one out of the semilabelled trees on $n + 1$ vertices, whose label set is a beginning segment of the set of natural numbers. Each of them occurs with the same probability.

THEOREM 12. *The distribution of the number of branching points in a random semilabelled tree is asymptotically normal with expected value* $(n/\ln n)(1 + o(1))$ *and with variance* $(n/(\ln n)^2)(1 + o(1))$. *If* $l > 0$ *is a constant and* $\chi(l)$ *counts the number of vertices with out-degree* l *in the random semilabelled tree, then* $\chi(l)$ *is asymptotically normal with expected value* $(r^l/l!)(1 + o(1))$ *and with variance* $(r^l/l!)(1 + o(1))$, *where* r *is the solution of the equation* $re^r = n$; *for a fixed* l, *the random semilabelled tree has a vertex with degree* l *with probability* $1 - o(1)$.

Proof. Apply the theory of random partitions [13] and Theorem 1. Neglect a number one, if necessary. Actually, [13] contains more theorems which apply. □

For rooted, unlabelled trees the next theorem easily follows, but we did not manage to derive a formula for it. Let $[l]$ denote the class of l and $\pi^{\#}$ denote the action of a permutation π on the power set.

THEOREM 13. *The number of rooted trees with* n *other, unlabelled vertices with* k *leaves out of them is equal to* $1/k!$ *times the number of ordered pairs* $\langle \mathcal{P}, \pi \rangle$, *where* $\mathcal{P}$ *is a partition of* $\{1, 2, \ldots, n\}$ *into* $n - k + 1$ *classes and* π

is a permutation of $\{1, 2, \ldots, n\}$ *under the assumptions*:

(i) $\pi^{\#}([n]) = [n]$,

(ii) *for all* $1 \leq l \leq n$ $|[\pi(l)]| = |[l]|$,

(iii) $P \in \mathcal{P}$, $P \neq [n]$ *implies* $\pi(1 + \max P) = 1 + \max \pi^{\#}(P)$.

Furthermore, if we restrict our attention to the enumeration of trees above with a given out-degree sequence, then, on the other side, we have to consider partitions whose class sizes produce the sequence above.

References

1. M. Carter, M. Hendy, D. Penny, L. A. Székely, and N. C. Wormald, On the distribution of lengths of evolutionary trees, *SIAM J. Discrete Math.*, in press.
2. A Cayley, A theorem on trees, *Quart. J. Math.* **23** (1889), 376–379.
3. L. E. Clarke, On Cayley's formula for counting trees, *J. London Math. Soc.* **33** (1958), 471–475.
4. A. Erdélyi and I. M. H. Etherington, Some problems of non-associative combinations II, *Edinburgh Math. Notes* **32** (1941), 7–12.
5. G. O. H. Katona, A theorem of finite sets, *in* "Theory of Graphs, Proceedings, Coll. Tihany, 1966, pp. 187–207, Akad. Kiadó, Budapest, 1966.
6. D. E. Knuth, "The Art of Computer Programming I: Fundamental Algorithms," Addison-Wesley, Reading, MA, 1973.
7. J. B. Kruskal, The number of simplices in a complex, *in* "Mathematical Optimization Techniques," pp. 251–278, Univ. of Calif. Press, Berkeley/Los Angeles, 1963.
8. L. Lovász, "Combinatorial Problems and Exercises," North-Holland, Amsterdam, 1979.
9. V. V. Menon, On the existence of trees with given degrees, *Sankhyá* **26** (1964), 63–68.
10. J. W. Moon, "Counting Labelled Trees," Montreal, Canadian Math. Congress, 1970.
11. Y. Poupard, "Codage et dénombrement de diverses structures apparentées à celle d'arbre," Cahiers du B.U.R.O., Institut de Statisque des Universités Paris, 1971.
12. A. Rényi, Some remarks on the theory of trees, *Publ. Math. Inst. Hungar. Acad. Sci.* **4** (1959), 73–85.
13. V. N. Sachkoff, "Probabilistic Methods of Combinatorial Analysis," Nauka, Moscow, 1978. [Russian]