ADVANCES IN APPLIED MATHEMATICS 14, 1–24 (1993)

Decompositions of Leaf-Colored Binary Trees

MICHAEL STEEL

Zentrum für interdisziplinäre Forschung der Universität Bielefeld, Wellenburg 1, Bielefeld 1, D-4800 Germany

Leaf-colored binary trees, with an induced integer "length," arise in biomathematics. We analyse such trees in terms of a natural bipartition of their edge set, and, extending a recent decomposition for binary trees, obtain enumerative formulae.

1. INTRODUCTION

A tree with a colouration of its leaves (degree-one vertices) has an associated non-negative, integer "length" which forms the basis for the widely used "minimum length tree" construction in taxonomy. This procedure estimates the ancestral evolutionary tree of a collection S of species as the tree(s), whose degree-one vertices are S, and which minimizes the sum of the lengths induced by a set of colorations of S. In practice these colorations arise from the characteristics on which S is being compared, for example, nucleotide bases on aligned DNA sequences (see, for example, Felsenstein [7]). The problem of finding a minimum length tree (also called a "maximum parsimony tree"), or even just the minimum possible length, is known to be NP-hard. However, good branch-and-bound and heuristic algorithms are available, and these are necessary when finding an optimal tree for a large number of species since any exhaustive search would have to consider every binary tree with labelled leaves. As is well known, the number of such trees is $(2n-5)!! = (2n-5) \times (2n-7) \times \ldots \times 3 \times 1$, where n is the number of leaves; see [4] for a bijective enumeration of various classes of leaf-labelled trees.

To evaluate the statistical properties of the minimum length tree method it is helpful to enumerate binary trees by their length according to a fixed leaf coloration. This task was taken up by Carter *et al.* [2] who obtained exact formulae in two cases: when the number of colors present is either two (the "bichromatic binary tree theorem"), or larger by one than the

0196-8858/93 $9.00

length of the tree. An exact enumeration formulae was then found for the case when the number of colors present equals the length of the tree (Steel [13]). That paper, and Erdős and Székely [5] also provided a structural proof, based on Menger's theorem, of the bichromatic binary tree theorem. Applications of this theorem to taxonomy were outlined in Steel [13], and implemented by Steel *et al.* [14]. Most recently, Erdős and Székely [6] have obtained an extension for trees of Menger's theorem as a first step toward a structural analysis of trees whose leaves are r-colored, $r > 2$.

In this paper we study the structure of trees which are subject to leaf r-colorations by considering a natural partition of the edge set of a leaf-colored tree T, into two classes: "reducible" and "irreducible" edges (Section 2). This in turn induces a decomposition of T into a set of subtrees over which the length function is additive, but so that these trees cannot be decomposed further while preserving additivity. In one special case this decomposition is shown to be unique. In Section 3 enumerative formulae are derived for bicolored trees (Theorem 3) and 3-colored trees (Theorems 4 and 5), using results from Section 2, the Lagrange inversion formula, and an extension of the methods from Steel [13].

DEFINITIONS. We mostly follow the terminology of Erdős and Székely [5]; for standard graph theoretic terminology see Bondy and Murty [1]. A *semilabelled tree* $T = (V(T), E(T))$ is a finite tree whose leaves (degree-one vertices) only are labelled. Throughout we let L denote the set of labelled leaves, and $n = |L|$. In case all the non-leaf vertices of T have degree three we say that T is *binary*. If T has a distinguished vertex v we say T is *rooted*, and v is the *root* of T. A *leaf r-coloration* is a map $\chi: L \to C$, where C is a set of r colors. If $\{|\chi^{-1}(\alpha)| : \alpha \in C\} = \{a_1 \geq a_2 \geq, \ldots, \geq a_r \geq 1\}$, we say χ is of *type* $(a_1, \ldots, a_r)$. A coloration $\bar{\chi} : V(T) \to C$ is an *extension* of χ if χ and $\bar{\chi}$ agree on L, and the *changing number* of $\bar{\chi}$ is the number of edges whose ends are assigned different colors by $\bar{\chi}$. An extension which has the smallest changing number amongst all extensions is called a *minimal coloration* (according to χ), and its changing number is called the *length* of the tree (according to χ).

We denote the length of T according to χ by $l(T, \chi)$, or more briefly $l(T)$ when χ is clear. If χ is an r-coloration of the leaves of T, and t is a subtree of T, let $\chi|_t$ denote the induced leaf coloration of t, and let $l(t) = l(t, \chi|_t)$.

If a semilabelled tree T consists of two rooted semilabelled trees T_1, T_2 and an edge e, incident with their roots, we write $T = [T_1, T_2]_e$, or just $T = [T_1, T_2]$, as indicated in Fig. 1(a), and say T_1, T_2 are *pendant* in T. For a rooted semilabelled tree T with root r we write $T = [T_1, \ldots, T_s]^r$ if deleting r and its incident edges produces the forest of rooted semi-

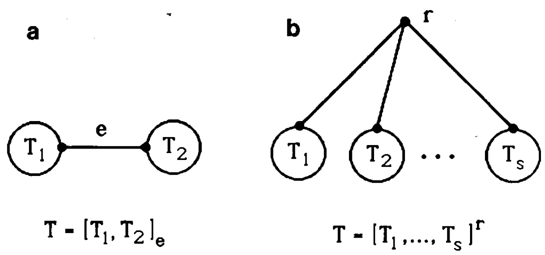


FIG. 1. Elementary edge (a) and root (b) tree decompositions.

labelled trees, $\{T_1, \ldots, T_s\}$, as indicated in Fig. 1(b). Given a rooted semilabelled tree T let $M(T, \chi)$ denote the set of colorations of the root of T which occur in at least one minimal coloration of T that extends χ. When χ is unambiguous we write $M(T) = M(T, \chi)$, and $M(t) = M(t, \chi|_t)$ for a subtree t of T.

Next we summarize some of the basic properties of $l(T)$. As mentioned in [6], the problem of determining the minimum value of $l(T)$ over all semilabelled trees T is a special case of the "multiway cut problem" for graphs, described by Chopra and Rao [3], which is NP-hard even for $r = 3$. However, for trees, $O(r \times n)$ algorithms—in particular, the "Fitch algorithm"—for determining $l(T)$ and a minimal coloration are well known in biomathematics; see, for example, [6, 9, 11]. We first describe a recursive and efficient way to calculate $l(T)$, established by Hartigan [9]. An extension of this result, not required here, also gives an explicit minimal coloration in $O(r \times n)$ time.

THEOREM 1. (1) *If T' is a subdivision of T then $l(T') = l(T)$.*

(2) Suppose $T = [T_1, \ldots, T_s]^r$. For $\alpha \in C$ let $f_\alpha = |\{i\colon \alpha \in M(T_i)\}|$, and let $K = \max\{f_\alpha\colon \alpha \in C\}$. Then,

$$M(T) = \{\alpha \in C\colon f_\alpha = K\}; \qquad l(T) = \sum_{i=1}^{s} l(T_i) + s - K.$$

When $r = 2$, an application of Menger's theorem shows that $l(T)$ equals the maximal number of edge-disjoint paths joining differently colored leaves. For $r > 2$ this result has been generalized by Erdős and Székely [6], as follows.

DEFINITION. By an *Erdős–Székely path system* on T (according to χ) we mean a collection P of oriented paths in T, each connecting a pair of differently colored leaves and having the property that if two paths in P use a common edge of T then both paths have the same direction along that edge and the leaves that the two paths are directed toward have different colors. If P has the maximum cardinality of any Erdős–Székely path system system on T (according to χ) we say that P is *optimal.*

THEOREM 2 (Erdős and Székely [6]). *The size of an optimal Erdős–Székely path system on T equals $l(T, \chi)$.*

A third, basic property of $l(T)$ is an inequality which arises from a recursion on the color set C. For $\alpha \in C$ and T a semilabelled tree having at least one vertex colored α by a leaf coloration χ, let $T^{-\alpha}$ denote the semilabelled tree obtained by deleting from T all leaves colored α, together with their incident edges, let $\chi^{-\alpha} = \chi|_{T^{-\alpha}}$ and let χ_α denote the leaf bicolouration of T by $\{\alpha, \beta\}$ defined by

$$\chi_\alpha(v) = \begin{cases} \alpha, & \text{if } \chi(v) = \alpha \\ \beta, & \text{otherwise.} \end{cases}$$

LEMMA 1. (1) *For any leaf r-coloration, $r - 1 \leq l(T) \leq n - a_1$, and $l(T) = n - a_1$ if and only if T has a minimal coloration which is monochromatic on the non-leaf vertices of T.*

(2) *For all pairs (T, χ), $l(T, \chi) \leq l(T^{-\alpha}, \chi^{-\alpha}) + l(T, \chi_\alpha)$.*

Furthermore, equality holds if $l(T, \chi) = n - a_1$ and $a_\alpha \leq a_2$, where $a_\alpha = |\chi^{-1}(\alpha)|$. In that case $l(T^{-\alpha}, \chi^{-\alpha}) = n - a_\alpha - a_1$ and $l(T, \chi_\alpha) = a_\alpha$.

Proof. (1) Choose r differently colored leaves $v_1, \ldots, v_r$. Then the collection of directed paths from v_1 to v_i, $i = 2, \ldots, r$, is an Erdős–Székely path system of cardinality $r - 1$, and thus $r - 1 \leq l(T)$ by Theorem 2. Coloring all the non-leaf vertices of T by a color which occurs most frequently on the leaves gives an extension $\bar{\chi}$ of χ having changing number $n - a_1$, hence $l(T) \leq n - a_1$, with equality precisely if $\bar{\chi}$ is a minimal coloration according to χ.

(2) Choose the following minimal colorations: $\bar{\chi}_1$ for $T^{-\alpha}$ according to $\chi^{-\alpha}$; and $\bar{\chi}_2$ for T, according to χ_α. Define a coloration $\bar{\chi}$ of $V(T)$ by setting

$$\bar{\chi}(v) = \begin{cases} \alpha, & \text{if } v \in \bar{\chi}_2^{-1}(\alpha) \\ \bar{\chi}_1(v), & \text{otherwise.} \end{cases}$$

Then $\bar{\chi}$ is an extension of χ, so $l(T, \chi)$ is at most the changing number of $\bar{\chi}$, which is the number of edges vv' of T for which $\bar{\chi}(v) \neq \bar{\chi}(v')$. This set is contained in the disjoint union of two sets:

$$C_1 = \{vv' : |\{v, v'\} \cap \bar{\chi}_2^{-1}(\alpha)| = 1\}$$
$$C_2 = \{vv' : \{v, v'\} \cap \bar{\chi}_2^{-1}(\alpha) = \varnothing, \bar{\chi}_1(v) \neq \bar{\chi}_1(v')\}.$$

Now, since $\bar{\chi}_1$ and $\bar{\chi}_2$ are minimal colorations (according to their respective leaf-colorations), we have $|C_1| = l(T, \chi_\alpha)$, and $|C_2| \leq l(T^{-\alpha}, \chi^{-\alpha})$. Thus $l(T) \leq |C_1 \cup C_2| = |C_1| + |C_2| \leq l(T^{-\alpha}, \chi^{-\alpha}) + l(T, \chi_\alpha)$, as required. From part (1) (and since χ_α is a leaf bicoloration) we have

$$l(T, \chi_\alpha) \leq a_\alpha \tag{1.1}$$

and if $a_\alpha \leq a_2$ then part (1) also gives

$$l(T^{-\alpha}, \chi^{-\alpha}) \leq (n - a_\alpha) - a_1. \tag{1.2}$$

Thus if $l(T, \chi) = n - a_1$ then the first part of result (2) gives $n - a_1 = l(T) \leq l(T^{-\alpha}, \chi^{-\alpha}) + l(T, \chi_\alpha) \leq n - a_1$ and so (1.1) and (1.2) must both be equalities, as claimed.

2. Reducible Edges, Irreducible Trees

Definition. If $T = [T_1, T_2]_e$, then by Theorem 1, $l(T, \chi) - l(T_1) - l(T_2) \in \{0, 1\}$. We say e is *reducible* (according to χ) if $l(T, \chi) = l(T_1) + l(T_2)$. If every edge of T is irreducible (i.e, not reducible) according to χ, we say T is *irreducible* (according to χ).

Clearly,

$$l(T) = \max_I \sum_{t \in I} l(t), \tag{2.1}$$

where I ranges over sets of disjoint irreducible semilabelled subtrees of T according to their induced leaf colorations.

Remarks. (1) The notion of reducible edges already exists in biomathematics (see, for example, Rinsma *et al.* [10])—its biological relevance relates to the estimation of the temporal lengths of edges of evolutionary trees constructed by the minimum length tree method—although irreducibility of trees appears to be a new concept.

(2) A set I which realizes equality in (2.1) can be constructed in polynomial time, although in general it is not unique.

(3) In case $r = 2$, the only irreducible binary trees are paths of length 2, with differently colored leaves; however, even for $r = 3$ there is an infinite number of irreducible binary trees.

LEMMA 2. *Given a leaf coloration of* $T = [T_1, T_2]_e$, *the following are equivalent*:

(i) e is reducible

(ii) $M(T_1) \cap M(T_2) \neq \varnothing$

(iii) all minimal colorations of T assign a common color to the endpoints of e.

(iv) at least one optimal Erdős–Székely path system avoids e.

Proof. (i) $\Leftrightarrow$ (ii) Theorem 1 with $s = 2$ implies that for $T' = [T_1, T_2]^r$, $l(T', \chi) - l(T_1) - l(T_2) = 0$ precisely if $M(T_1) \cap M(T_2) \neq \varnothing$, and since $l(T', \chi) = l(T, \chi)$ the result follows.

(ii) $\Leftrightarrow$ (iii) is clear.

(i) $\Rightarrow$ (iv) If e is reducible select, for $i = 1, 2$, an optimal Erdős–Székely path system P_i on T_i (according to $\chi|_{T_i}$). Thus $|P_i| = l(T_i)$. Now $P_1 \cup P_2$ is an Erdős–Székely path system on T and

$$|P_1 \cup P_2| = |P_1| + |P_2| = l(T_1) + l(T_2) = l(T),$$

the last equality holding since e is reducible. Thus $P_1 \cup P_2$ is optimal and avoids e.

(iv) $\Rightarrow$ (i) Suppose an optimal Erdős–Székely path system P avoids e. For $i = 1, 2$, let P_i denote the set of paths in P_i which lie in T_i. Then $|P_1| + |P_2| = |P| = l(T)$. However, for $i = 1, 2$, P_i is an Erdős–Székely path system on T_i and hence $l(T_i) \geq |P_i|$. Thus $l(T_1) + l(T_2) \geq l(T)$, which combined with the reverse inequality (which always holds) gives $l(T_1) + l(T_2) = l(T)$, and so e is reducible.

LEMMA 3. (1) *Suppose* $T = [T_1, T_2]_e$, *where e is an irreducible edge of T according to* χ, *and* e_1 *is an irreducible edge of* T_1 *according to* $\chi|_{T_1}$. *Then* e_1 *is an irreducible edge of T according to* χ.

(2) *Suppose* $T = [T_1, \dots, T_s]^r$, *and let* e_i' *denote the edge of T incident with r and the root of* T_i. *Then T is irreducible according to* χ *if, for* $i = 1, \dots, s$, T_i *is irreducible according to* $\chi|_{T_i}$, *and* e_i' *is irreducible according to* χ.

Proof. (1) Delete e and e_1 from T to partition T into three trees, t_0, t_1, t_2, where we may suppose t_1 lies between t_0 and t_2, and $t_2 = T_2$ (thus t_0, and t_2 are rooted). Then $l(T) = l(T_1) + l(T_2) + 1$, since e is irreducible, and $l(t_0) + l(t_1) + 1 = l(T_1)$ since e_1 is an irreducible edge of T_1. Thus

$$l(T, \chi) = l(t_0) + l(t_1) + l(t_2) + 2. \qquad (2.2)$$

Now writing $T = [t_0, T']_{e_1}$ and noting that $l(T') \leq l(t_1) + l(t_2) + 1$, we deduce from (2.2) that $l(T) > l(t_0) + l(T')$ and so e_1 is an irreducible edge of T.

(2) The "only if" direction is immediate. For the "if" direction, suppose $e'_1, \ldots, e'_s$ are all irreducible and that T is reducible. Then T has a reducible edge of $e_1 \notin \{e'_1, \ldots, e'_s\}$, where we may suppose e_1 lies in T_1. But since $T = [T_1, T']_{e'_1}$, where $T' = [T_2, \ldots, T_s]^r$ we can apply Lemma 3(1) (with $e = e'_1$) to obtain the required contradiction.

The following result is useful in determining whether or not a tree is reducible. Given a set C' of colors, let $S(C')$ denote the "star-shaped" rooted semilabelled tree consisting of leaf set C' together with one other root vertex. Note that $S(C')$ is endowed with a natural leaf $|C'|$-coloration.

Lemma 4. *$T = [T_1, T_2]_e$ is irreducible according to χ if and only if $[T_1, S_2]_e$ and $[S_1, T_2]_e$ are irreducible according to their induced leaf colorations, where $S_i = S(M(T_i))$.*

Proof. Let $M_i = M(T_i)$. Suppose T is irreducible and e' is an edge of $[T_1, S_2]_e$. If $e' = e$ then e' is irreducible since $M_1 \cap M_2 = \varnothing$. If e' is an edge in S_2 incident with a leaf v and the root r of S_2 then let S' denote the tree obtained from S_2 by deleting v and its incident edge, let $T' = [T_1, S']^r$, and let χ' be the induced leaf coloration of T'. By Theorem 1, $M(T', \chi') = M_1 \cup M_2 - \{\chi(v)\}$, and so $M(T', \chi') \cap \chi(v) = \varnothing$, and thus, by Lemma 2, e' is an irreducible edge of $[T_1, S_2]_e$. If e' lies in T_1, delete e' from $[T_1, S_2]_e$ to partition this tree into two subtrees t_1, t_2, where t_1, say, is contained entirely in T_1. Let t'_2 denote the subtree of T for which $T = [t_1, t'_2]_{e'}$. Then by Theorem 1, $M(t_2) = M(t'_2)$ and, since e' is an irreducible edge of T, Theorem 1 implies $M(t_1) \cap M(t_2) = \varnothing$. Thus (again by Theorem 1) e' is an irreducible edge of $[T_1, S_2]_e$. The same applies for $[S_1, T_2]_e$. The converse argument to establish the irreducibility of every edge e' of T is similar, by considering separately the cases $e' = e$; e' lies in T_1; and e' lies in T_2.

As mentioned above, the set of trees I which realizes $l(T)$ in (2.1) is generally not unique, even when $r = 2$. However, there is a special case where this is so.

Definition. We say a semilabelled tree T is *r-tight* (according to χ) if $l(T) = n - a_r$. By Lemma 1 (1), this condition is equivalent to requiring that χ is an r-coloration of type $(k, k, \ldots, k)$ and that $l(T) = n - k$ $(= k(r-1))$, or that, equivalently, every coloration of the non-leaf vertices of T by a single color is a minimal coloration.

LEMMA 5. *Suppose T is r-tight according to χ.*

(i) *If $T = [T_1, T_2]_e$ then T_1 and T_2 are both r-tight precisely when e is reducible, and neither tree is r-tight when e is irreducible.*

(ii) *A unique set I realizes the equality in (2.1) and consists of the irreducible r-tight subtrees of T.*

Proof. (i) Since T is r-tight, either T_1 and T_2 are both r-tight or neither tree is. Thus it suffices to show that T_1, T_2 are r-tight if and only if e is reducible. Suppose first that T_1 and T_2 are r-tight. If T_i has rk_i leaves, then $l(T_i) = k_i(r-1)$ so that

$$l(T_1) + l(T_2) = (k_1 + k_2)(r-1) = l(T),$$

and hence e is reducible.

Conversely, suppose at least one of the pair T_1, T_2 is not r-tight. Let K_i denote the number of leaves of T_i, and for each $\alpha \in C$, let k_α be the number of leaves of T_1 colored α. Thus,

$$\sum_{\alpha \in C} k_\alpha = K_1.$$

There are two possible cases:

(a) $k_\alpha > k_\beta$ for some $\alpha, \beta \in C$.

(b) k_α is a constant value, k' for each $\alpha \in C$, and at least one tree, say T_1 is not r-tight; that is, $l(T_1) \leq k'(r-1) - 1$.

In case (a), since $l(T_1) \leq K_1 - k_\alpha$, and $l(T_2) \leq K_2 - (k - k_\beta)$ (as T_2 has $k - k_\beta$ leaves colored β), we have

$$\begin{aligned} l(T_1) + l(T_2) &\leq k(r-1) - (k_\alpha - k_\beta), \qquad \text{since } K_1 + K_2 = rk, \\ &\leq k(r-1) - 1, \\ &= l(T) - 1, \end{aligned}$$

which implies that e is irreducible. In case (b) let v denote the unique vertex of T_1 which is adjacent to a vertex of T_2. If a minimal coloration of T_1 according to χ_1 assigns color $\gamma \in C$ to v, extend this coloration to T by assigning color γ to all the non-leaf vertices of T_2. Since $l(T)$ is at most the changing number of this coloration,

$$\begin{aligned} l(T) &\leq (k'(r-1) - 1) + (k - k')(r-1) \\ &= k(r-1) - 1, \qquad \text{a contradiction.} \end{aligned}$$

Thus case (b) cannot occur.

(ii) First note that the trees in any set I which realizes the equality in (2.1) must collectively cover every leaf of T, for let n_t denote the number of leaves of each tree t in I and let $f(t)$ be the largest number of monochromatic leaves in t. Then,

$$\sum_{t\in I} n_t \leq rk, \qquad l(t) \leq n_t - f(t), \qquad \sum_{t\in I} f(t) \geq k,$$

and these inequalities (in particular, the first) are compatible with an equality in (2.1) for I, only if they are all equalities.

Suppose I_1, I_2 are two sets which realize equality in (2.1). We show $I_1 = I_2$ by induction on $m = |I_1| + |I_2|$. This holds trivially for $m = 2$, so suppose $m > 2$. Since the trees in each of I_1 and I_2 collectively cover all the leaves of T there exists, for $j = 1, 2$, a tree $T_j \in I_j$ which is pendant in T, and for which $T_1 \cap T_2 \neq \varnothing$. If $T_1 \neq T_2$, it follows that one of these trees is pendant in the other, say T_2 is pendant in T_1. But by part (i), any pendant tree of T which lies in I_1 or I_2 (in particular, T_1 and T_2) is r-tight, and so, writing $T_1 = [T_2, T']_e$, for some pair e, T', a second application of part (i) shows that T_1 is reducible, a contradiction. Thus there exists an irreducible r-tight tree T_0 common to I_1, I_2, and for which $T = [T_0, T_0']_e$ for some T_0', e. Then by part (i) T_0' is r-tight, and since $I_1 - \{T_0\}$ and $I_2 - \{T_0\}$ both realize the equality in (2.1) for T_0', it follows by induction that $I_1 - \{T_0\} = I_2 - \{T_0\}$, which completes the proof.

3. Application to Binary Trees

Definition. By a *rooted semilabelled binary tree* T we mean either an isolated labelled leaf or a semilabelled binary tree on two or more vertices, for which one edge has been subdivided—the new vertex being the *root* of T.

Let $F_k(a_1, \ldots, a_r)$ denote the set of semilabelled binary trees of length k according to a *fixed* coloration of type $(a_1, \ldots, a_r)$, and let $f_k(a_1, \ldots, a_r)$ be the cardinality of $F_k(a_1, \ldots, a_r)$, as in Carter *et al.* [2]. We will suppose $a_r > 1$, since if $a_r = 1$ one has

$$f_k(a_1, \ldots, a_{r-1}, 1) = (2n - 5) f_{k-1}(a_1, \ldots, a_{r-1}).$$

Exact expressions for $f_k(a_1, \ldots, a_r)$ have so far been found in three cases: $r = 2$ (Carter *et al.* [2]), $k = r - 1$ (Carter *et al.* [2]) and $k = r$ (Steel [13]). In the case $r = 2$ the bichromatic binary tree theorem, proved in [2, 5, 13], states

$$f_k(a,b) = \frac{(2n-3k)\psi_k(a,b)b(n)}{b(n-k+2)}, \qquad n = a+b,$$

where $b(n) = (2n-5)!! = (2n-5) \times (2n-7) \times \cdots \times 3 \times 1$, which is the number of binary trees on n leaves, and

$$\psi_k(a,b) = (k-1)!N(a,k)N(b,k),$$

where

$$N(n,k) = \begin{cases} \binom{2n-k-1}{k-1} \times b(n-k+2), & \text{if } k \leq n \\ 0, & \text{if } k > n \end{cases}$$

which is the number of forests consisting of k rooted semilabelled binary trees on a total of n labelled leaves (see [2]).

We first extend this theorem by adding as an additional parameter the number of irreducible edges and the colors a root vertex can take over all minimal colorations.

DEFINITIONS. Let $f_k(a_1,\ldots,a_r|s)$ denote the number of trees in $F_k(a_1,\ldots,a_r)$ which have exactly s irreducible edges, and let $f_k'(a_1,\ldots,a_r)$ denote the number of pairs, (T,e), where $T \in F_k(a_1,\ldots,a_r)$ and e is an irreducible edge of T. Thus,

$$f_k(a_1,\ldots,a_r) = \sum_{s\geq 0} f_k(a_1,\ldots,a_r|s),$$

$$f_k'(a_1,\ldots,a_r) = \sum_{s\geq 0} s \times f_k(a_1,\ldots,a_r|s).$$

A link between the latter quantity and $f_k(a_1,\ldots,a_r)$ is provided by the following result.

LEMMA 6. *If* $a_r = 2$,

$$\begin{aligned} f_k(a_1,\ldots,a_r) - (2n-7)\big(f_{k-1}(a_1,\ldots,a_{r-1}) \\ +(2n-6)f_{k-2}(a_1,\ldots,a_{r-1})\big) \\ = 2\big(f_{k-1}'(a_1,\ldots,a_{r-1}) - f_{k-2}'(a_1,\ldots,a_{r-1})\big). \end{aligned}$$

Proof. First note that a binary tree with p leaves has exactly $2p-3$ edges. For $T \in F_k(a_1,\ldots,a_r)$, $a_r = 2$, let v, v' denote the two leaves of T

assigned a least frequently occurring color. Then T is obtained from a unique tree $T' \in F_j(a_1, \ldots, a_{r-1})$, $j < k$, in precisely one of the following ways:

(1) Make v and v' ends of a new edge e, then subdivide both e and an edge e' of T' and make these two new vertices ends of a new edge. There are $(2(n-2)-3)$ ways to do this and in this case $j = k - 1$.

(2) Subdivide two different edges of T' and make each of v and v' adjacent to exactly one of the new vertices by a new edge. There are $(2(n-2)-3)(2(n-2)-4)$ ways to do this, and in this case $j = k - 2$.

(3) Twice subdivide an edge, e, of T' and make each of v and v' adjacent to exactly one of the new vertices by a new edge. For each edge there are two ways to do this. Writing $T' = [T_1, T_2]_e$ and $M_i = M(T_i; \chi|_{T_i})$, an application of Theorem 1 shows that

$$j = \begin{cases} k-1 & \text{if } M_1 \cap M_2 = \emptyset \\ k-2 & \text{if } M_1 \cap M_2 \neq \emptyset. \end{cases}$$

In view of Lemma 2, this implies that $j = k - 1$ precisely if e is an irreducible edge of T', and $j = k - 2$ otherwise. Thus the number of trees in $T \in F_k(a_1, \ldots, a_r)$ arising from case (3) is precisely $2f'_{k-1}(a_1, \ldots, a_{r-1}) + 2((2(n-2)-3)f_{k-2}(a_1, \ldots, a_{r-1}) - f'_{k-2}(a_1, \ldots, a_{r-1}))$. The lemma now follows.

For a nonempty subset M of C let $f_{k,M}(a_1, \ldots, a_r)$ denote the number of rooted semilabelled binary trees of length k according to a fixed coloration χ of type $(a_1, \ldots, a_r)$, and for which $M(T, \chi) = M$. Clearly,

$$\sum_M f_{k,M}(a_1, \ldots, a_r) = (2n-3)f_k(a_1, \ldots, a_r).$$

In the following theorem, $n = a + b$, and in part (2a), $n!!$ is $n \times (n-2) \times (n-4), \ldots,$ where the last term is 1 or 2 depending on whether n is odd or even, respectively.

THEOREM 3. (1) *The number of irreducible edges in a semilabelled binary tree T of length k according to some leaf bicoloration, lies between k and $3k-2$.*

(2a) *The semilabelled binary trees which attain the lower bound in part (1) are precisely those which have a unique minimal coloration. The number* *of such trees is*:

$$f_k(a,b|k) = \frac{(2n-4k)\psi_k(a,b)(2n-k-5)!!}{(2n-3k-1)!!}$$

(2b) $f_k(a,b|k+1) = 0$

(3) $f_k(a,b|3k-2) = \psi_k(a,b)(3k-3)!/(2k-1)!$

(4) $f_k'(a,b) = k2^{k-1}\psi_k(a,b)(n-2)!/(n-k-1)!$

(5) *writing* $C = \{\alpha,\beta\}$, *where* $|\chi^{-1}(\alpha)| = a$, *we have*

$$f_{k,M}(a,b) = \begin{cases} \dfrac{2(a-k)\psi_k(a,b)b(n+1)}{b(n-k+2)} & \text{if } M = \{\alpha\}, \\ \dfrac{2(b-k)\psi_k(a,b)b(n+1)}{b(n-k+2)} & \text{if } M = \{\beta\} \\ \dfrac{k\psi_k(a,b)b(n+1)}{b(n-k+2)} & \text{if } M = \{\alpha,\beta\}. \end{cases}$$

Proof. We first summarize some recent results relating to the enumeration of semilabelled binary trees and establish a mild generalization of one of these formulae. Suppose $F = \{t_1,\ldots,t_{r-1}\}$ is a forest of rooted semilabelled binary trees, t_0 is a semilabelled binary tree, and the leaf sets of t_0 and the trees in F are disjoint. As in [13] let $\mathrm{Ext}(t_0;F)$ denote the set of semilabelled binary trees which contain disjoint subtrees $T_0,\ldots,T_{r-1}$ such that for $i \geq 0$, each T_i is a subdivision of t_i, and if $i > 0$ the root of t_i, when regarded as a vertex of T_i, lies on the path from any vertex in T_i to T_0. A complementary, recursive description of $\mathrm{Ext}(t_0;F)$, given in [5], is as follows. For $|F| = 1$, $\mathrm{Ext}(t_0;\{t_1\})$ is the set of trees which can be obtained by subdividing by a new vertex an edge of t_0, and making this new vertex adjacent to the root of t_1 by a new edge. For $|F| > 1$, a tree T lies in $\mathrm{Ext}(t_0;F)$ precisely if $T \in \mathrm{Ext}(T';\{t\})$ for some $t \in F$ and $T' \in \mathrm{Ext}(t_0;F-\{t\})$. We call such a description of T a *recursive Ext construction* (*of* T). From Steel [13], or Erdős and Székely [5], we have

$$|\mathrm{Ext}(t_0;F)| = \frac{e_0 b(n)}{b(n-r+2)}, \tag{3.1}$$

where $e_0 = |E(t_0)|$ and $r-1 = |F|$. Let

$$\mathrm{Ext}_k(t_0;F) = \left\{T \in \mathrm{Ext}(t_0;F) \colon |V(T) - \bigcup_{i\geq 0} V(T_i)| = k\right\}.$$

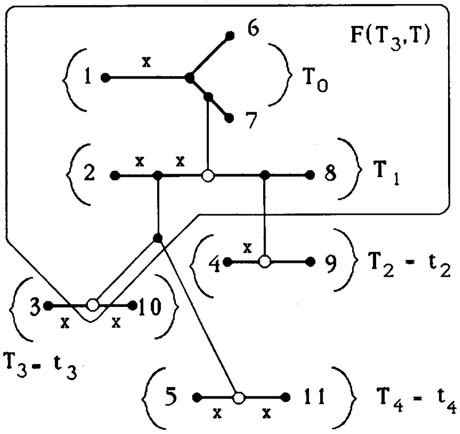


FIG. 2. A tree in $\mathrm{Ext}_1(t_0, F)$, $F = \{t_1, \ldots, t_4\}$ regarded as a tree in $F_5(5, 6)$ with $s = 1$, $d = 2$. The $i(T) = 8$ irreducible edges are marked by x, the fibre above T_3, $F(T_3, T)$, is circled.

From [13],

$$|\mathrm{Ext}_k(t_0; F)| = \frac{e_0(k + r - 1)!}{k!2^k} \times \binom{2n - r - k - 4}{r - k - 1}, \tag{3.2}$$

where $r - 1 = |F|$. A representation of a tree in $\mathrm{Ext}_1(t_0; F)$, $|F| = 4$, is shown in Fig. 2.

Next we generalize (3.1) by allowing conditional restrictions on the possible edges which may be subdivided at each step of any recursive Ext construction of a tree in $\mathrm{Ext}(t_0; F)$. The *fibre* above T_i $(i \geq 0)$ in T, denoted $F(T_i, T)$, is defined as follows: For $i > 0$, $F(T_i, T)$ is the minimal subtree of T which contains the root of T_i and all the trees in $\{T_0, T_1, \ldots, T_{r-1}\} - \{T_i\}$ encountered on the path from T_i to T_0; for $i = 0$ we let $F(T_0, T) = \varnothing$. For example, in Fig. 2, the fibre above the tree $T_3 = t_3$ is depicted as an enclosed region and consists of T_0, T_1, one other vertex, and three other edges. Suppose the set of edges of t_i which may be subdivided to form T_i in $T \in \mathrm{Ext}(t_0; F)$ is restricted to lie in a subset (of the edge set of t_i), which may depend on $F(T_i, T)$ by some pre-given rule, but always has the same cardinality, denoted e_i', and let $\mathrm{Ext}'(t_0; F)$ denote the set of trees in $\mathrm{Ext}(t_0; F)$ which satisfy this requirement for $i = 0, \ldots, r - 1$. Then we claim there is the following strengthening of (3.1),

$$|\mathrm{Ext}'(t_0; F)| = \frac{e_0'(E' + 2r - 4)!!}{E'!!}, \tag{3.3}$$

where $E' = \sum_{i=0}^{r-1} e_i'$ and $r - 1 = |F|$.

(Proof of the claim: Apply induction on r. The result clearly holds when $r = 2$, so suppose $r > 2$. For $T \in \mathrm{Ext}'(t_0; F)$, consider the set $S(T)$ of the subtrees T_i (subdivisions of the trees in F) of T for which the fibre above T_i consists only of T_0 and possibly vertices of degree 2. By induction, the number of trees in $\mathrm{Ext}'(t_0; F)$ for which $S(T)$ is a given set S is

$$e_0'(e_0' + 2) \cdots (e_0' + 2|S| - 2) \times \sum_{t \in S} e_t' \times \frac{(E' - e_0' + 2(r - |S|) - 4)!!}{(E' - e_0)!!}.$$

Summing over all sets S of size p this expression depends only on E', e_0', p, and r. Thus, summing over p, $|\mathrm{Ext}'(t_0; F)|$ depends only on E', e_0', and r, and so we may take $e_1' = E' - e_0'$, and $e_i' = 0$ for $i > 2$; in which case it is easily shown that (3.3) holds.)

Next we describe an inverse, constructive aspect of a decomposition from [13] which allows the realization of all bicolored semilabelled binary trees of a given length in terms of $\mathrm{Ext}(t_0; F)$ for simply colored choices of t_0, F. By a *bichromatic tree* we mean a rooted semilabelled binary tree $T = [T_1, T_2]^r$ and a leaf bicoloration χ of T under which T_1 and T_2 are each monochromatic, but oppositely colored. Let $H(a, b, k)$ denote the set of forests consisting of exactly k bichromatic trees and for which a total of exactly $a \geq 0$ leaves are assigned one color and $b \geq 0$ leaves are assigned the other. Clearly,

$$|H(a, b, k)| = k! \times N(a, k) N(b, k) = k\psi_k(a, b), \qquad (3.4)$$

where $N(m, k)$ and $\psi_k(a, b)$ are as in the definitions preceding Theorem 3. Given a rooted semilabelled binary tree t, let $t^\wedge$ denote the semilabelled tree obtained by deleting from t the root, and identifying its two incident edges. Set

$$G_k(a, b) = \{(F, t, T) \colon F \in H(a, b, k), t \in F, T \in \mathrm{Ext}(t^\wedge; F - \{t\})\},$$

and let $\mathbf{f}$ denote the projection $\mathbf{f}((F, t, T)) = T$. By the decomposition from [13],

$$\mathbf{f} \text{ is } k\text{-to-1 from } G_k(a, b) \text{ onto } F_k(a, b). \qquad (3.5)$$

An example of the construction of a tree in $F_5(5, 6)$ in this way is illustrated in Fig. 2, where $\chi(\{1, \ldots, 5\}) = \{\alpha\}$, $\chi(\{6, \ldots, 11\}) = \{\beta\}$.

We now describe recursively the set of irreducible edges in T for $(F, t, T) \in G_k(a, b)$. Suppose $T \in \mathrm{Ext}(t^\wedge; F - \{t\})$, where $t \in F$ and F is a set of k bichromatic trees. For convenience we write t_0 for $t^\wedge$, and F' for $F - \{t\}$. By the recursive description of Ext we have $T \in \mathrm{Ext}(T'; \{t'\})$, where $T' \in \text{Ext}(t_0; F' - \{t'\})$ for some $t' \in F'$. Let e' denote the edge of T' which was subdivided by the new vertex v' in forming T.

By Theorem 1 it is easily seen that

(i) an edge $\neq e'$ of T' is irreducible in T precisely if it is irreducible in T'.

(ii) Subdividing e' creates two new edges which are both irreducible (resp. both reducible) in T precisely in e' is irreducible (resp. reducible) in T'.

(iii) The edge of T incident with v' and the root of t is reducible.

Regarding the edges of t', write $t' = [t_\alpha, t_\beta]^r$, where the leaves of t_α, t_β are colored α, β, respectively. Then,

(iv) edges in t_α or t_β are reducible in T.

Regarding the two remaining edges of t' which are incident with r, let T_j denote the first tree in $F' - \{T_i\}$ encountered in the directed path from T_i to T_0, and let v denote the degree-two vertex of T_j where this path first meets T_j.

By case (ii) above (applied inductively), the two edges of T_j which are incident with v are either both reducible or both irreducible in T. If they are both reducible then all minimal colorations of T assign the same value $\gamma(\in \{\alpha, \beta\})$ to v. Then it is easily seen that

(v) In this case the edge of t' incident with r and the root of t_γ is reducible in T, while the other edge of t' incident with r is irreducible in T.

Similarly,

(vi) if the edges of T_j that are incident with v are both irreducible in T then both edges of t' incident with r are irreducible in T.

In this last case we say that t' is *descended from an irreducible edge*. Cases (v) and (vi) are illustrated by taking $t = t_2, t_3$ (respectively) in Fig. 2.

Since t_0 has exactly one irreducible edge, (i) through (vi) gives

$$i(T) = k + d + s, \tag{3.6}$$

where

$i(T)$ is the number of irreducible edges of T, and $k = l(T)$.

d is the number of trees in F' which are descended from an irreducible edge in any recursive Ext construction of T.

s is the number of edges of times an irreducible edge is subdivided.

For example, for the construction illustrated by Fig. 2 we have $s = 1$, $d = 2$. We now apply these results to establish the claims in the theorem.

(1) By Lemma 2, $i(T) \geq l(T)$. Also, since $s \leq d \leq k - 1$ we have $i(T) \leq 3k - 2$, establishing (1).

(2a) By the equivalence of (1) and (3) in Lemma 2, $i(T) = k$ precisely if T has a unique minimal coloration. If $i(T) = k$, then $d = s = 0$. In this case in any recursive Ext construction of T no tree from F' can be descended from an irreducible edge, and no subdivisions of irreducible edge are allowed. Thus the set of edges of $t \in F'$ which can be subdivided in the construction of T is always one edge less than the full edge set of t (this prohibited edge being dependent on the fibre of t in T). Applying (3.4) (with $E' = 2n - 3k - 1$) and (3.5) gives the equation in (2a).

(2b) If $i(T) = k + 1$ then $s + d = 1$. But this is impossible since $s \leq d$, and yet $d = 1$ clearly implies $s = 1$.

(3) Let $\nu = |V(T) - \bigcup_{i \geq 0} V(T_i)|$, where T_i is, as usual, the subdivision of t_i in T. Then

$$s + \nu \leq k - 1, \tag{3.7}$$

since the vertices counted by s and ν are a subset of vertices created by edge subdivisions in the construction of T. Suppose $i(T) = 3k - 2$. Then from $s \leq d \leq k - 1$ we have $s = d = k - 1$, and thus $\nu = 0$, by (3.7). Hence $i(T) = 3k - 2$ precisely if T is obtained from a recursive Ext construction in such a way that subdivisions are confined to the irreducible edge of t_0, and, for each tree t in F', the two edges of t which are incident with the root. Thus in a recursive Ext construction of T we may as well regard $t_0, t_1, \ldots, t_{k-1}$ as having just two leaves, and at the end of the construction replace each pair of leaves by the corresponding rooted subtrees of t_i, for $i = 0, \ldots, k - 1$. Thus the number of trees in $\mathrm{Ext}(t_0; F')$ with $i(T) = 3k - 2$ is just $|\mathrm{Ext}_0(t_0'; F^{\#})|$, where $|F^{\#}| = k - 1$ and where t_0' and the trees in $F^{\#}$ all have just two leaves. The result now follows from (3.2) (taking $k = 0$) and (3.4).

(4), (5) Let

$$T_M = T_M(x_1, \ldots, x_r, z) = \sum_{a_1, \ldots, a_r, k \geq 0} \frac{f_{k,M}(a_1, \ldots, a_r)}{a_1! \cdots a_r!} x_1^{a_1} \cdots x_r^{a_r} z^k.$$

Order the elements of C as $\alpha_1, \ldots, \alpha_r$ so that $|\chi^{-1}(\alpha_i)| = a_i$. Then by Theorem 1, the collection $\{T_M : M \neq \varnothing\}$ satisfies the system of simultaneous quadratic equations:

$$T_M = \frac{1}{2}T_M^2 + \sum_{\{\{A,B\}:\, A\cap B = M\}} T_A T_B + \sum_{\{\{A,B\}:\, A\cap B=\varnothing,\, A\cup B = M\}} zT_A T_B + \sum_{i=1}^{r} x_i \delta_{i,M},$$

where, for $i = 1, \ldots, r$,

$$\delta_{i,M} = \begin{cases} 1 & \text{if } M = \{\alpha_i\} \\ 0 & \text{otherwise.} \end{cases}$$

In case $r = 2$, writing $T_i = T_{\{\alpha_i\}}$ for $i = 1, 2$, and $T_3 = T_{\{\alpha_1, \alpha_2\}}$ these above equations can be written $T_i = x_i \Phi_i(T_1, T_2, T_3)$, where $x_3 = z$, and

$$\Phi_i(w_1, w_2, w_3) = \begin{cases} \left(1 - \frac{1}{2}w_1 - w_3\right)^{-1}, & i = 1 \\ \left(1 - \frac{1}{2}w_2 - w_3\right)^{-1}, & i = 2, \\ w_1 w_2\left(1 - \frac{1}{2}w_3\right)^{-1}, & i = 3. \end{cases}$$

As in [12] we can apply the multivariate Lagrange inversion formula for monomials (see Goulden and Jackson [8]) to evaluate T_i and T_1T_2 (this is considerably simpler than applying the full multivariate Lagrange inversion formula, as in the calculation by Carter *et al.* [2] of $(T_1 + zT_2 + T_3)$). Omitting the details of the calculation, the results now follow from the identities:

$$f_k'(a, b) = a!b!\left[x_1^a x_2^b z^{k-1}\right] T_1 T_2;$$
$$f_{k,M}(a, b) = a!b!\left[x_1^a x_2^b z^k\right] T_M(x_1, x_2, z),$$

where "[ ]" denotes coefficient extraction, as in [8]. This completes the proof of Theorem 3.

Remarks. (i) The upper bound $31(T) - 2$ in Theorem 3(1) is realized by the semilabelled binary caterpillar tree and bicoloration described in an example below.

(ii) The combination of result (4) with Lemma 6 gives an exact expression for $f_k(a, b, 2)$.

(iii) Result (5) is a refinement of the bichromatic binary tree theorem, as this follows from (5) immediately. Result (5), while not stated explicitly as such by Carter *et al.* [2], is implicit in the formulae given in that paper.

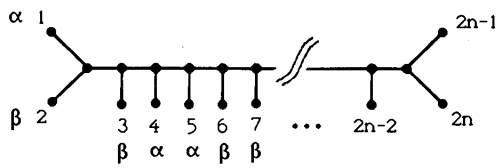


FIG. 3. A bicolored binary tree with a Fibonacci number of minimal colorations.

(iv) The average number of irreducible edges in trees chosen from $F_k(a, b)$ is $f_k'(a, b)/f_k(a, b)$ which, by the previous theorem, is independent of a and b for $n = a + b$ fixed.

(v) A recursive Ext construction of T also allows, for example, an immediate identification of those edges e of T for which every minimal coloration of T according to χ always colors the ends of e differently. More significantly, one can also recursively count the number of minimal colorations of T.

EXAMPLE. In general, a semilabelled tree can have several minimal colorations according to a leaf coloration χ, as an example by Rinsma *et al.* [10] shows. Indeed the number of minimal colorations can grow exponentially with $1(T)$, even when $r = 2$ as the following example illustrates. Consider the semilabelled binary "caterpillar" tree J_{2n} on $2n$ leaves, labelled $1, \ldots, 2n$, as in Fig. 3, and the leaf bicoloration χ, where

$$\chi(i) = \begin{cases} \alpha, & \text{if } i = 0, 1 \pmod 4 \\ \beta, & \text{otherwise.} \end{cases}$$

By Theorem 2 (and Lemma 1(1)) it is easily seen that $1(J_{2n}, \chi) = n$. Using a recursive Ext construction it can be shown that the number F_n of minimal colorations of J_{2n} with respect to χ satisfies the recurrence $F_n = F_{n-1} + F_{n-2}$, $n > 1$, and so, since $F_0 = F_1 = 1$, it follows that F_n is the nth Fibonacci number.

Finally, we show how the results developed in Section 2 allow the classification of irreducible semilabelled binary trees when $r = 3$, leading to their exact enumeration, and thereby providing an expression for $f_{2k}(k, k, k)$.

DEFINITION. A rooted semilabelled binary tree T, subject to a leaf 3-coloration χ, is said to be *4-spread* if T has four leaves and the shape of the tree in Fig. 4(a), and χ assigns a color, denoted $f(T, \chi)$, to a pair of leaves of T which are separated by four edges, and χ assigns each of the other two colors to each of the remaining leaves of T'.

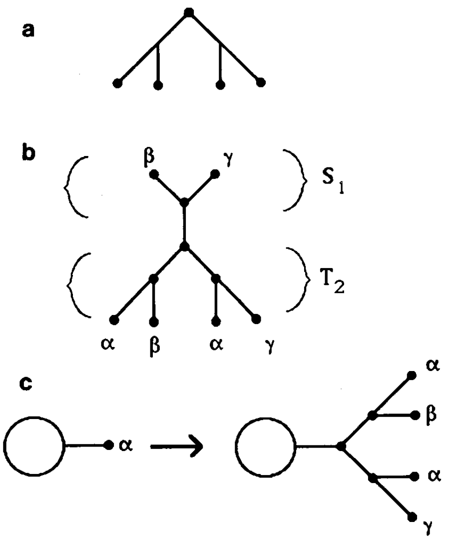


FIG. 4. Tree representations for the proof of Lemma 7.

The following lemma shows that up to a permutation of the leaf set, all semilabelled binary trees which are irreducible according to a leaf 3-coloration are obtained from a semilabelled tree on three differently colored leaves by successively applying the type of transformation illustrated in Fig. 4(c) (where the three colors α, β, γ, may also be permuted in this transformation).

LEMMA 7. *A semilabelled binary tree T is irreducible according to a leaf 3-coloration χ if and only if either T has just three differently colored leaves or $T = [T_1, T_2]$, where T_2 is 4-spread, and $[T_1, \{v\}]$ is irreducible according to χ', where $\chi'|_{T_1} = \chi|_{T_1}$, $\chi'(v) = f(T_2, \chi|_{T_2})$.*

Proof. (only if) Since T is binary, T can be represented as $T = [T_1, T_2]_e$, where T_2 either has three leaves or T_2 has four leaves and the shape shown in Fig. 4(a). Suppose T is irreducible according to a leaf 3-coloration χ. As in Lemma 4, let $M_i = M(T_i, \chi|_{T_i})$, $S_i = S(M_i)$. Then by Lemma 4, $[S_1, T_2]$ is irreducible. However, by considering cases it can readily be checked that this implies that $|M_1| = 2$, $M_1 = \{\alpha, \beta\}$, say, and that $[S_1, T_2]$ is the semilabelled tree with leaf 3-coloration as in Fig. 4(b), where α, β, γ are the three colours in C. Thus T_2 is 4-spread. This in turn implies $M_2 = \{\alpha\}$ and so $[T_1, S_2]$ is a binary tree and irreducible according to the induced leaf 3-coloration, χ'.

(if) From the "only if" direction, it follows by induction that if a tree T is irreducible according to a leaf 3-coloration χ then χ must be 3-tight, and hence if any non-leaf vertex v of T is distinguished as a root of T then $M(T, \chi) = \{\alpha, \beta, \gamma\}$. Thus if $T' = [T_1, \{v\}]$ is irreducible according to χ', where $\chi'(v) = \alpha$, say, we must have $M(T_1, \chi_1) = \{\beta, \gamma\}$ by Theorem 1. A second application of Lemma 4 now shows that if T_2 is a 4-spread pendant subtree of T and with $f(T_2, \chi|_{T_2}) = \alpha$, then $T = [T_1, T_2]$ is irreducible according to χ.

THEOREM 4. *Let $I_k(a_1, a_2, a_3)$ be the number of irreducible trees in $F_k(a_1, a_2, a_3)$. Then,*

$$I_k(a_1, a_2, a_3) = \begin{cases} \dfrac{2(p!)^2(4p-3)!}{(3p-1)!}, & \text{if } a_1 = a_2 = a_3 = p,\ k = 2p \\ 0, & \text{otherwise.} \end{cases}$$

Proof. The restriction for which $I_k(a_1, a_2, a_3)$ is nonzero follows from the previous lemma. Thus let $I(n) = I_{2k}(k, k, k)$, $n = 3k$. Let $d_T(v, v')$ denote the number of edges in the path in T between vertices v and v'. We say that a rooted semilabelled binary tree T having root r, is *even* (resp. *odd*) if $d_T(r, j) \equiv 0 \pmod 2$ (resp. $d_T(r, j) \equiv 1 \pmod 2$) for all leaves j. For $n \geq 1$, let E_n (resp. O_n) denote the number of even (resp. odd) rooted semilabelled binary trees on n leaves. Let

$$E(x) = \sum_{n \geq 1} \frac{E_n x^n}{n!}; \qquad O(x) = \sum_{n \geq 1} \frac{O_n x^n}{n!}.$$

Then by the standard rooted binary tree decomposition $T = [T_1, T_2]^r \to \{T_1, T_2\}$ we have

$$E(x) = \tfrac{1}{2} O(x)^2 + x; \qquad O(x) = \tfrac{1}{2} E(x)^2.$$

Thus $E(x) = x(1 - \frac{1}{8}E(x)^3)^{-1}$ and so, by the Lagrange inversion formula (see [8]), we have

$$\frac{O_{n-1}}{(n-1)!} = [x^{n-1}]\frac{1}{2}E(x)^2 = \frac{1}{(n-1)} \times [\lambda^{n-2}]\lambda\left(1 - \frac{1}{8}\lambda^3\right)^{-(n-1)},$$

where $[\lambda^p]f(\lambda)$ denotes the coefficient of λ^p in $f(\lambda)$. Thus,

$$O_{n-1} = \begin{cases} (n-2)!\dbinom{4k-3}{k-1}2^{3-n}, & \text{if } n = 3k, \\ 0, & \text{otherwise.} \end{cases} \tag{3.8}$$

Next, we say that a semilabelled binary tree T is *even* if $d_T(i, j) \equiv 0 \pmod 2$ for all leaves i, j. Equivalently, $T = [T_1, \{v\}]$ is even if and only if the rooted tree T_1 is odd (as defined earlier). Thus, if $E(n)$ denotes the number of even semilabelled binary trees on n leaves, then,

$$E(n) = O_{n-1}. \tag{3.9}$$

Let Ω_k denote the set of pairs (T, χ), where χ is an 3-coloration of type (k, k, k) of $3k$ labelled leaves and T is an irreducible binary tree according to χ. Counting Ω_k by first enumerating trees for each χ, and then summing over all choices of χ gives

$$|\Omega_k| = I(3k) \times \frac{(3k)!}{(k!)^3}.$$

Conversely, we may first fix a semilabelled binary tree T, and count the r-colorations of the leaves for which T is 3-tight, and then sum over all choices for T. Now, the even semilabelled binary trees have precisely the same topology as the irreducible 3-colored trees, as can readily be shown by the same type of inductive argument used to establish Lemma 7. In this way, using Lemma 7,

$$|\Omega_k| = E(3k) \times 3 \times 2^{n-2}.$$

Combining the two expressions for $|\Omega_k|$, gives $I(3k)$ in terms of $E(n)$ and, thus, by (3.8) and (3.9) the theorem follows.

THEOREM 5.

$$f_{2k}(k,k,k) = (k!)^3 \sum_{s=1}^{k} [x^k] \frac{Q(x)^s}{s!} \times \frac{b(n)}{b(n-s+2)},$$

where $[x^i]Q(x) = 2(4i-3)!(6i-3)/(3i-1)!i!$ *and* $n = 3k$.

Proof. Suppose $S = \{t_0, \dots, t_{k-1}\}$ is a set of semilabelled binary trees having disjoint leaf sets. Let $\mathrm{Ext}(S)$ denote the set of semilabelled binary trees which contain disjoint subtrees that are homeomorphically equivalent to S (i.e., modulo vertices of degree 2). Clearly, $\mathrm{Ext}(S)$ is the disjoint union of $\mathrm{Ext}(t_0; F)$ over all forests F which can be constructed from $t_1, \dots, t_{k-1}$ by subdividing by a new (root) vertex an edge of each tree. Thus by Result (3.1) in the proof of Theorem 3,

$$|\mathrm{Ext}(S)| = \frac{b(n)}{b(n-k+2)} \times \prod_{i=0}^{k-1} e_i, \qquad \text{where } e_i = |E(t_i)|.$$

For $t_1 \geq 0, \ldots, t_k \geq 0$, let $S(\mathbf{t})$ denote the set of all collections S of irreducible 3-colored semilabelled binary trees on disjoint leaf sets chosen from $3k$ leaves and for which t_i of the trees in S have $3i$ leaves. Thus if $s = |S|$ we have

$$s = \sum_i t_i, \qquad k = \sum_i i t_i. \tag{3.10}$$

Let $T(k)$ denote the set of 3-tight semilabelled binary trees on $3k$ leaves, and let $G(k) = \{(\mathbf{t}, T, S): S \in S(\mathbf{t}), T \in \text{Ext}(S)\}$. Then the projection

$$\phi : G(k) \to T(k)$$
$$\phi(\mathbf{t}, S, T) = T$$

is a bijection, by Lemma 5. Thus, by the above expression for Ext(S),

$$f_{2k}(k,k,k) = |F(k)| = |G(k)|$$
$$= \sum_{s=1}^{k} \sum_{\mathbf{t}} \nu(\mathbf{t}) \times \prod_{i=1}^{s} I(3i)^{t_i} \times \frac{b(n)}{b(n-s+2)} \times \prod_{i=1}^{s} (6i-3)^{t_i}, \tag{3.11}$$

where $\mathbf{t}$ ranges over all vectors, having non-negative integer components, and satisfying (3.10), and $\nu(\mathbf{t})$ is the number of ways of partitioning $k = \sum_{i=1}^{s} i t_i$ labelled leaves of each of three colors into s blocks, where t_i of the blocks have size $3i$, and with each block receiving the same number of each color. Thus,

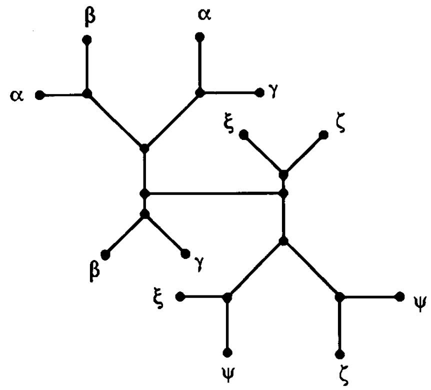
$$\nu(\mathbf{t}) = \frac{1}{\prod_{i=1}^{s} t_i!} \times \left(\frac{k!}{\prod_{i=1}^{s} (i!)^{t_i}} \right)^3.$$

Substituting this into (3.11), together with the expressions for $I(3i)$ given by Theorem 4, and then collecting terms, completes the proof.

EXAMPLE. To find $f_6(3,3,3)$ note that $[x^3](Q(x)^s/s!)$ consists of just one product term for each $s \in \{1, 2, 3\}$. Applying the theorem gives

$$f_6(3,3,3) = 19{,}116.$$

An interesting remaining question is whether there is a simple and constructive characterization of semilabelled binary trees which are irreducible according to a leaf r-coloration for $r = 4$ or for other values of $r > 3$ (analogous, perhaps, to Lemma 7, although respecting the additional



$l(T) = 9 \neq n - a_1$

FIG. 5. An irreducible binary tree, which has no minimal coloration under which the non-leaf vertices are monochromatic.

possibilities allowed by Lemma 3(2)). Note that, unlike the case $r = 3$, irreducible r-colored binary trees can fail to be r-tight, as the example in Fig. 5 illustrates.

ACKNOWLEDGMENTS

This work was supported by the Alexander von Humboldt-Stiftung.

REFERENCES

1. J. A. BONDY AND U. S. R. MURTY, "Graph Theory with Applications," Macmillan Press, London, 1976.
2. M. CARTER, M. D. HENDY, D. PENNY, L. A. SZÉKELY, AND N. C. WORMALD, On the distribution of lengths of evolutionary trees, *SIAM J. Discrete Math.* **3** (1990), 38–47.
3. S. CHOPRA AND M. R. RAO, On the multiway cut polyhedron, *Networks* **21** (1991), 51–89.
4. P. L. ERDŐS AND L. A. SZÉKELY, Applications of antilexicographic order 1: an enumerative theory of trees, *Adv. Appl. Math.* **10** (1989), 488–496.
5. P. L. ERDŐS AND L. A. SZÉKELY, Counting bichromatic evolutionary trees, *Discrete Appl. Math.*, in press.
6. P. L. ERDŐS AND L. A. SZÉKELY, Evolutionary trees: An integer multicommodity max-flow-min-cut theorem, *Adv. Appl. Math.*, in press.

7. J. Felsenstein, Phylogenies from molecular sequences: Inference and reliability, *Annu. Rev. Genet.* **22** (1988), 521–565.
8. I. P. Goulden and D. M. Jackson, "Combinatorial Enumeration," Wiley, New York, 1983.
9. J. A. Hartigan, Minimum mutation fits to a given tree, *Biometrics* **29** (1973), 53–65.
10. I. Rinsma, M. D. Hendy, and D. Penney, Minimally coloured trees, *Math. Biosci.* **98** (1990), 201–210.
11. D. Sankoff, Minimal mutation trees of sequences, *SIAM J. Appl. Math.* **28**, No. 1 (1975), 35–42.
12. M. A. Steel, Distributions on bicoloured evolutionary trees, *Bull. Austral. Math. Soc.* **41** (1990), 159–160.
13. M. A. Steel, Distributions on bicoloured binary trees arising from the Principle of Parsimony, *Discrete Appl. Math.*, in press.
14. M. A. Steel, M. D. Hendy, and D. Penny, Significance of the length of the shortest tree, *J. Classification* **9** (1992), 71–90.