Pruning Trees and Tree Reconstruction

Tomás Feder* Carlos Subi†

Abstract

Given a tree T with n vertices, a *k-pruning* of T for $0 \leq k \leq n$ is a sequence $v_1 v_2 \cdots v_k$ of distinct vertices such that $T - \{v_1, \ldots, v_i\}$ is either the empty graph (for $i = k = n$) or a subtree of T, for $1 \leq i \leq k$. Equivalently, v_i is either the last vertex to be removed or a leaf of $T - \{v_1, \ldots, v_{i-1}\}$. If T is instead a rooted tree with root r, then a *rooted k-pruning* of T is a k-pruning such that $v_i \neq r$ for $i < n$.

We denote by $p_k(T)$ the number of k-prunings of T, and by $p_k^r(T)$ the number of rooted k-prunings of T rooted at r. We study the question of whether T can be reconstructed from such pruning numbers or from some associated pruning tree, both in the rooted and unrooted cases.

1 Introduction

Given a tree T with n vertices, a *k-pruning* of T for $0 \leq k \leq n$ is a sequence $v_1 v_2 \cdots v_k$ of distinct vertices such that $T - \{v_1, \ldots, v_i\}$ is either the empty graph (for $i = k = n$) or a subtree of T, for $1 \leq i \leq k$. Equivalently, v_i is either the last vertex to be removed or a leaf of $T - \{v_1, \ldots, v_{i-1}\}$. If T is instead a rooted tree with root r, then a *rooted k-pruning* of T is a k-pruning such that $v_i \neq r$ for $i < n$.

We denote by $p_k(T)$ the number of k-prunings of T, and by $p_k^r(T)$ the number of rooted k-prunings of T rooted at r. We formulate two conjectures.

Conjecture 1 *The tree T can be reconstructed from the numbers $p_k(T)$ for $0 \leq k \leq n$.*

Conjecture 2 *The tree T rooted at r can be reconstructed from the numbers $p_k^r(T)$ for $0 \leq k \leq n$.*

A similar system of properties on tree data was studied before. It was shown that if we count the number of subtrees of a tree with ℓ internal (non-leaf) eges and L leaf edges, then the reconstruction property in this case is true for rooted trees [2] but false even over caterpillars for unrooted trees [1].

We do not prove or disprove our conjectures. However, we show that T cannot be reconstructed from the collection of n numbers $p_n^r(T)$ over all possible choices of root r for T. Therefore an unrooted T cannot be reconstructed from $p_n(T) = \sum_r p_n^r(T)$. These results are based on a simple characterization of $p_n^r(T)$ for a given root r, and a simple property of the numbers $p_n^r(T)$ for two adjacent choices of root r.

We next show that the smallest possible value 1 of $p_n^r(T)$ is achieved uniquely by a path rooted at a leaf, and that the smallest possible value 2^{n-1} of $p_n(T)$ is achieved uniquely by an unrooted path. We also show that the largest possible value $(n-1)!$ of $p_n^r(T)$ is achieved uniquely by a star rooted at the center, and that the largest possible value $2(n-1)!$ of $p_n(T)$ is achieved uniquely by

*268 Waverley Street, Palo Alto, CA 94301. Email: tomas@theory.stanford.edu
†carlos.subi@hotmail.com

an unrooted star. We define an *elongated star* to be a tree T whose only vertex of degree at least 3 is the root r. We show that $p_n^r(T)$ does not always uniquely characterize an elongated star T among elongated stars, and yet the collection of numbers $p_k(T)$ ($p_k^r(T)$) for $0 \le k \le n$ uniquely characterizes each elongated star T among elongated stars (resp. rooted elongated stars).

Finally, we define an associated pruning tree S for T based on prunings (and rooted prunings for root r). The vertices at depth k correspond to the k-prunings (resp. rooted k-prunings) for $0 \le k \le n$, and a k-pruning $v_1 \cdots v_k$ has $k+1$-prunings $v_1 \cdots v_k v_{k+1}$ as children obtained by adding a valid vertex v_{k+1} (similarly for rooted k-prunings). We show that unrooted and rooted trees T can be reconstructed from the associated pruning tree S.

2 Reconstructing T from $p_n^r(T)$ and $p_n(T)$

In this section we study the numbers $p_n^r(T)$ and $p_n(T)$. Let T be a tree rooted at r. We attach a number $m(r')$ to each vertex r' of T, and a number $p(T')$ to the subtree T' rooted at r'. If T' has n' vertices, and the h subtrees T_i' of T' rooted at the children r_i' of r' have n_i' vertices, for $1 \le i \le h$ then we assign multinomials $m(r') = \binom{n'-1}{n_1', n_2', \ldots, n_h'}$ and assign $p(T') = \prod_{v \in T'} m(v)$.

Theorem 1 *We have $p_n^r(T) = p(T)$.*

Proof. Asssume inductively for the subtrees T_i with n_i vertices rooted at the children r_i of r for $1 \le i \le h$ that $p_{n_i}^{r_i}(T_i) = p(T_i)$ is the number of n_i-prunings of the T_i. Then it remains to intermingle these $\prod_{1 \le i \le} p(T_i)$ prunings in $m(r) = \binom{n-1}{n_1, n_2, \ldots, n_h}$ possible ways, and therefore multiplying by $m(r)$ gives the stated result. □

From this we infer the following.

Theorem 2 *Suppose T', T'' are rooted at r', r'' and have n', n'' vertices respectively. Let T be obtained by joining the two roots r', r''. Then $p_n^{r'}(T)/n' = p^{r''_n}(T)/n''$.*

Proof. Denote by $m^r(v)$ and $p^r(T)$ the numbers $m(v)$ and $p(T)$ when T is rooted at r We have $m^{r'}(r') = \binom{n-1}{n''} m^{r''}(r')$ and $m^{r''}(r'') = \binom{n-1}{n'} m^{r'}(r'')$, while $m^{r'}(v) = m^{r''}(v)$ for $v \neq r', r''$. Therefore

$$p^{r'}(T)/\binom{n-1}{n'-1} = p^{r'}(T)/\binom{n-1}{n''} = p^{r''}(T)/\binom{n-1}{n'}$$

and so $p^{r'}(T)/n' = p^{r''}(T)/n''$. □

Given four rooted trees T_1, T_2, T_3, T_4, each with t vertices, we let $T(T_1, T_2, T_3, T_4)$ be the tree rooted at r such that the children of r are r' and r'', the children of r' are the roots of T_1 and T_2, and the children of r'' are the roots of T_3 and T_4. We consider two nonisomorphic trees $T' = T(T_1, T_2, T_3, T_4)$ and $T'' = T(T_1, T_3, T_2, T_4)$.

Theorem 3 *The two nonisomorphic trees T', T'' with $n = 4t + 3$ vertices have $p_n^r(T') = p_n^r(T'')$.*

Proof. The root in both trees satisfies $m(r) = \binom{n-1}{2t+1}$, the children of the root in both trees satisfies $m(r') = m(r'') = \binom{2t}{t}$, and the values of $p(T_1)$, $p(T_2)$, $p(T_3)$, $p(T_4)$, are the same in both trees, so the result follows from Theorem 1. □

Theorem 4 *The two nonisomorphic trees T', T'' with $n = 4t + 3$ vertices have equal multisets $\{p_n^v(T')\} = \{p_n^v(T'')\}$ over all choices of root v.*

Proof. We have $p^r(T') = p^r(T'')$ from Theorem 3. We also have $p^{r'}(T') = p^{r''}(T') = p^{r'}(T'') = p^{r''}(T'')$ from Theorem 2 applied to the two trees joined by an edge from r to a child. We further have, again from Theorem 2, that all values $p^v(T_i)$ where v is the root of T_i are the same in all eight T_i, four in T' and four in T''. Finally, proceeding down the four T_i in T' and T'' gives the same values, again from Theorem 2. This completes the one-to-one correspondence between the two multisets. □

Theorem 5 *The two nonisomorphic trees T', T'' with $n = 4t+3$ vertices have $\{p_n(T')\} = \{p_n(T'')\}$, where the trees are now considered unrooted.*

Proof. The unrooted values $p_n(T')$ are equal to the sum of rooted values $p_n^v(T')$ over all choices of root v, and similarly for T'', so the result follows from the result on multisets in Theorem 4. □

3 Paths, Stars, and Elongated Stars

We first consider paths.

Theorem 6 *The smallest possible value 1 of $p_n^r(T)$ is achieved uniquely by a path rooted at a leaf, and that the smallest possible value 2^{n-1} of $p_n(T)$ is achieved uniquely by an unrooted path.*

Proof. In the rooted case, to achieve the minimum 1, we must always have only 1 leaf other than the root, so T must be a path rooted at a leaf.

In the unrooted case, to achieve the minimum 2^{n-1}, we must always have only 2 leaves, so T must be an unrooted path. □

We next consider stars.

Theorem 7 *The largest possible value $(n-1)!$ of $p_n^r(T)$ is achieved uniquely by a star rooted at the center, and that the largest possible value $2(n-1)!$ of $p_n(T)$ is achieved uniquely by an unrooted star.*

Proof. In both the rooted and unrooted cases, the maximum number of leaves at each iteration is achieved precisely by stars. □

We further consider elongated stars.

Theorem 8 *The number $p_n^r(T)$ does not always uniquely characterize a rooted elongated star T among rooted elongated stars.*

Proof. Let T', T'' be the two elongated stars with root r and paths from r to leaves of lengths $1, 1, 1, 4$ and $2, 2, 3$ respectively. We have $p(T') = m(r)$ and $p(T'') = m(r)$ in both trees, and the two multinomials have the same value, so the result follows. □

We finally show that the conjectures hold over elongated stars.

Theorem 9 *The collection of numbers $p_k(T)$ ($p^r_k(T)$) for $0 \le k \le n$ uniquely characterizes each elongated star T among elongated stars (resp. rooted elongated stars).*

Proof. Consider two rooted elongated stars T', T'' such that the paths from the root r have increasing lengths $\ell'_1, \ell'_2, \ldots, \ell'_{h'}$ and $\ell''_1, \ell''_2, \ldots, \ell''_{h''}$ respectively.

If $h' \neq h''$, then $p^r_1(T') = h' \neq h'' = p^r_2(T'')$ and the result follows, othewise we set $h = h' = h''$. Let j be such that $\ell'_j < \ell''_j$ and $\ell'_i = \ell''_i$ for $1 \le i < j$. For $k \le \ell'_j$, we can isomorphically repeatedly remove leafs from the corresponding paths in this enumaration order, so $p^r_k(T') = p^r_k(T'')$ in these cases for k. However, for $k = \ell'_j + 1$, we can remove the next edge from the jth path in T'' of length $\ell''_j \ge k$ if we have already removed only edges in this path previously, but we do not have such a choice in T', so $p^r_k(T'') > p^r_k(T')$ in this case for k.

This completes the proof for the rooted case. For the unrooted case, we notice that up to $k = \ell'_j + 1$ we may not remove the root r, so the unrooted case is equivalent to the case rooted at r for the values of k that we used in the proof. $\square$

4 Reconstruction from the Pruning Tree

The following is known.

Theorem 10 *[3] An unrooted tree T with h leaves is reconstructible from its h maximal subtrees (each obtained by removing one leaf from T).*

We infer the following.

Corollary 1 *An unrooted tree T is reconstructible from its associated pruning tree S.*

Proof. We proceed from the leaves up by induction. Suppose we have reconstructed the trees corresponding to subtrees of S rooted at the children of a given vertex v of S. Then the tree corresponding to the subtree rooted at v can be reconstructed as well since its children correspond to maximal subtrees. The construction finishes with v being the root of S. $\square$

Define the *weight* of a vertex v in a tree to be the maximum size of a component of $T - v$. The *centroid* of T is the set of vertices v of minimum weight. It is well known that the centroid of T consists of one or two vertices. We call T unicentroidal or bicentroidal accordingly. We prove the equivalent results for rooted T.

Theorem 11 *A rooted tree T with h leaves other than the root is reconstructible from its h maximal rooted subtrees (each obtained by removing one leaf from T if such a leaf is not the root).*

Proof. We proceed by induction on the size of T. The tree T is rooted at a leaf if and only if the h maximal subtrees are rooted at a leaf or T is a path on three vertex rooted at the middle vertex. In this last case T is reconstructible. Otherwise, when all maximal subtrees are rooted at leaves, we may move the root to the vertex adjacent to the corresponding leaf, and remove this leaf. By induction, we reconstruct a rooted T' and adding a leaf at the root of T' and moving the root to this leaf gives T.

Suppose instead that not all maximal subtrees are rooted at a leaf, and thus T is not rooted at a leaf or the path on three vertices with root in the middle. We proceed to reconstruct T without the root from Theorem 10.

If T is bicentroidal, then all maximal subtrees will be unicentroidal, and we may determine the subtree rooted at this centroid of a maximal subtree containing the root r obtained by omitting the subtree of a single other child without r having $(n/2) - 1$ vertices in the subtree. If the root is the centroid in some such maximal subtree, then it will not be the centroid in another maximal subtree, and this and this centroid corresponds to the vertex of the centroid of T that is not the root r, with its corresponding subtree.

If T is unicentroidal, then the root r is this centroid if and only if the root is in the centroid of every maximal subtree. If the root r is not the centroid of T, then we can examine the tree containing r coming out of the centroid of the maximal subtrees, and determine when this subtree has an extra vertex added as a leaf. In that case the subtree is the same as the subtree coming out of the centroid of T containing r. □

The following follows as before for the unrooted case.

Corollary 2 *A rooted tree T is reconstructible from its associated pruning tree S.*

Thus to prove the two conjectures it would suffice to show that the number of vertices at each level of the associated pruning tree S characterizes S (hence T).

References

[1] D. Eisenstat and G. Gordon, Non-isomorphic caterpillars with identical tree data, Discr. Math. 306 (2006), 827–830.

[2] G. Gordon and E. McMahon, A greedoid polynomial which distinguishes rooted arborescences, Proc. Amer. Math. Soc. 107 (1989), 287–298.

[3] F. Harary and E. Palmore, The reconstruction of a tree from its maximal subtrees, Canad. J. Math 18 (1966b), 803–810.