At most Cubic Caterpillars not Reconstrible from Subtree Data

Tomás Feder* Carlos Subi†

1 Introduction

Given a tree T with n vertices and m edges, we define the quantities ℓ_i for integer $i \geq 0$ to be the number of paths of length i contained in T. Notice that $\ell_0 = n$ and $\ell_1 = m$.

We consider the following question: When can the tree T be reconstructed, up to isomorphism, from the quantities ℓ_i for all $i \geq 0$?

The answer is known to be negative even for caterpillars [2]. We thus restrict our attention to trees of degrees at most three.

We first to show two nonisomorphic trees T_1 and T_2 of degrees at most three with $n = 11$ and $m = n - 1 = 10$ that have the same sequence of ℓ_i with $i \geq 0$. The trees T_1 and T_2 thus show that T cannot always be reconstructed from the sequence of ℓ_i.

The trees T_1 and T_2, however, are not caterpillars, they are caterpillars plus one vertex and one edge. We show that if T is assumed to be a caterpillar, then T of degrees at most three cannot in general be reconstructed from the sequence of ℓ_i with $i \geq 0$. In fact there are diameters d for which 2^{d-3} nonisomorphic caterpillars exist, all with the same ℓ_i (without assuming degree at most three).

Finally we ask if trees or specifically caterpillars of degrees at most three are reconstructible from subtree data (not just path data), namely the number of subtrees with ℓ internal vertices and L leaves. The answer is known to be negative for caterpillars without assuming degree at most three [1], and David Eisentat finally found a counterexample heree as well for at most cubic caterpillars by computer.

2 Non-reconstructibility with $n = 11$ and $m = n - 1 = 10$

We show the following.

Theorem 1 *There exist two non-isomorphic trees T_1 and T_2 of degrees at most three with $n = 11$ and $m = n - 1 = 10$ that have the same sequence of ℓ_i with $i \geq 0$.*

Proof. Let A be the tree rooted at a with a single child a'', and with two paths of length 2 coming out of a'', and let B be the tree rooted at b with a single child b'', and with b'' having two children, one of which is a leaf, and the other one which has two leaf children. Let C be the tree obtained by joining A and B with the edge ab. Finally, let T_1 be obtained from C by adding a new edge aa' to a new vertex a', and let T_2 be obtained from C by adding a new edge bb' to a new vertex b'.

*268 Waverley Street, Palo Alto, CA 94301. Email: tomas@theory.stanford.edu

†carlos.subi@hotmail.com

The trees T_1 and T_2 are non-isomorphic, since any path of maximum diameter 5 has the edge ab as the middle edge, so any isomorphism between T_1 and T_2 would have to flip the edge ab, considering the degrees of a and b differ in T_1 and in T_2. Since there is no root-preserving isomorphism from A to B, non-isomorphism follows.

To see that all ℓ_i with $i \geq 0$ are the same in T_1 and T_2, noticing that two trees are the same if we remove a', b' and get back to C, so the number of paths of length i not ending in a' or b' for the two trees are the same as the number of such paths in C. As for the number of paths of length i ending in a' or b' for the two trees, we notice that for the two trees A and B, the number of paths of length i starting in a or b respectively are the same, namely 1 forlength 1, 2 for length 2, 2 for length 3, and 0 for any greater length. Therefore the paths starting at a' or b' can be put in one-to-one correspondence, by flipping A and B, since the counts for A and B are the same. □

3 Negative and Positive Examples for Caterpillars

We show a preliminary result that will allow us to show non-reconstructibility of caterpillars.

Theorem 2 *Consider the polynomials*

$$g(x) = \sum_{i=0}^{n} a_i x^i,$$

$$h(x) = x^n g(1/x) = \sum_{i=0}^{n} a_i x^{n-i},$$

with integer coefficients $a_i \geq 1$. and

$$\begin{aligned} f(x) &= g(x)h(x) \\ &= a_0 a_n + (a_0 a_{n-1} + a_1 a_n)x + (a_0 a_{n-2} + a_1 a_{n-1} + a_2 a_n)x^2 \\ &\quad + \cdots + (a_0^2 + \cdots + a_n^2)x^n + \cdots + a_n a_0 x^{2n}. \end{aligned}$$

Notice that the coefficients of x^i and $x^{2n} - i$ in $f(x)$ are the same for all $0 \leq i \leq 2n$, and that $a_0 + \cdots + a_n = g(1) = \sqrt{f(1)}$ is reconstructible from $f(x)$. Then we cannot reconstruct the pair $g(x), h(x)$ from the polynomial $f(x)$. In fact a single $f(x)$ can lead to 2^{n-1} decompositions $g(x), h(x)$ Also, a single $f(x)$ leads to two decompositions will coefficients a_i equal to 1 or 2.

Proof. Let $f(x) = (x+2)(2x+1)(x+3)(3x+1)\cdots(x+(n+1))((n+1)x+1)$. The factors in $f(x)$ come in n consecutive symmetric pairs. We may choose in 2^{n-1} different ways which of the two factors of each pair to include in one or the other of $g(x), h(x)$.

For example, for $n = 2$, we may choose $g(x) = (x+2)(x+3)$, $h(x) = (2x+1)(3x+1)$ or $g(x) = (x+2)(3x+1)$, $h(x) = (2x+1)(x+3)$. This gives the two solutions $g(x) = (x^2+5x+6)$, $h(x) = (6x^2+5x+1)$ and $g(x) + (3x^2+7x+2)$, $h(x) = (2x^2+7x+3)$.

Similarly, letting $g(x)$ be the product of either x^3+x+1 or x^3+x^2+1 with either $x^6+x^5+x^2+x+1$ or $x^6+x^5+x^4+x+1$ (each pair having reverse coefficients) gives

$$g(x) = (x^3+x+1)(x^6+x^5+x^2+x+1) = x^9+x^8+x^7+2x^6+2x^5+x^4+2x^3+2x^2+2x+1$$

$$g(x) = (x^3+x^2+1)(x^6+x^5+x^2+x+1) = x^9+2x^8+x^7+x^6+2x^5+2x^4+2x^3+2x^2+x+1$$

with coefficients 1 and 2. $\square$

The polynomial decomposition naturally corresponds to a caterpillar for $g(x)$ with path

$$v_0, v_1, v_2, \dots, v_{n-1}, v_n$$

and the number of leafs attached to v_i respectively

$$a_0, a_1 - 1, a_2 - 1, \dots, a_{n-1} - 1, a_n.$$

Then if we let the a_i be coefficients of $g(x)$, then the coefficients of x^i in $f(x)$ for $0 \leq i < n$ are the ℓ_{n+2-i}, while the coefficient of x^2 equal to $a_0^2 + \cdots + a_n^2$ and $g(1) = a_0 + \cdots + a_n$ are equivalently replaced by $\ell_0 - 1 = \ell_1 = g(1) + 1$ and $\ell_2 = a_0(a_0+1)/2 + \cdots + a_n(a_n+1)/2 + 1$.

The diameter is then $n + 2$ and the number of different caterpillars of this diameter with the same ℓ_i is $2^{n-1} = 2^{d-3}$.

For example, as in the theorem, we may let $a_0, a_1 - 1, a_2$ be $1, 4, 6$ or $3, 6, 2$ and the polynomial with coefficients 1 and 2 gives two caterpillars of degreeas at most three.

Corollary 1 *There can be 2^{d-3} non-isomorphic caterpillars of diameter $d \geq 4$ that all have the same sequence of ℓ_i with $i \geq 0$, and two non-isomorphic caterpillars with degrees at most three that give the same sequence.*

If the caterpillar is required to be symmetric, that is, $a_i = a_{n-i}$ for all $0 \leq i \leq n$, then the caterpillar can indeed be reconstructed from the ℓ_i. To see this, first $a_0 = a_n$ can be obtained from $a_0 a_n$, then $a_1 = a_{n-1}$ can be obtained from $a_0 a_{n-1} + a_1 a_n$, then $a_2 = a_{n-2}$ can be obtained from $a_0 a_{n-2} + a_1 a_{n-1} + a_2 a_n$, and so on and so forth.

An alternative proof of the symmetric case is that in this case $h(x) = g(x)$, so $g(x)h(x) = g'(x)h'(x)$ gives $g^2(x) = {g'}^2(x)$, so $g'(x) = g(x)$ or $g'(x) = -g(x)$, and since the coefficents are positive, we have $g'(x) = g(x)$.

If the symmetry $a_i = a_{n-i}$ is only required in a middle range $k \leq i \leq n-k$ for some k, and outside this range the values a_j are given (not necessarily symmetrically), then the remaining middle range a_i can similarly be reconstructed.

Finally, if the symmetry $a_i = a_{n-i}$ is required in the range $1 \leq i \leq n-1$, with a_0, a_n possibly different, and the value $a_1 = a_{n-1}$ is given, then the reconstruction of the remaining a_i. First, the values $a_0 a_n$ and $a_0 a_{n-1} + a_1 a_n$ give us the sum and product of a_0, a_n, which give us a_0, a_n up to exchange, then the previous case completes the reconstruction of the middle range a_i with $2 \leq i \leq n-2$.

Suppose we assume in addition for two caterpillars that the paths counts starting at either end are the same. This means that if $C = a_0, a_1, ..., a_n$ and $C' = a'_0, a'_1, ..., a'_n$ are the two degree sequences, then $a_0 + a_n = a'_0 + a'_n, a_1 + a_{n-1} = a'_1 + a'_{n-1}$, etcetera. Then we have $g(x)h(x) = g'(x)h'(x)$, and in this special case $g'(x) = g(x) + p(x), h'(x) = h(x) - p(x)$, so $g(x)h(x) = g(x)h(x) + p(x)(h(x) - g(x) - p(x))$, so p(x)=0 or g'(x)=h(x) and both representations are the same.

We now consider the case of subtree data, and see that the two caterpillars from the polynomial in the theorem do not give the same subtree data. In fact, the blocks of coefficent 2 must be of the same number of same lengths in both caterpillars, as seen by considering subtrees with $L = \ell + 2$ leaves and ℓ internal nodes in decreasing order.

The counterexample of David Eisenstat has degree sequences $(2, 2, 3, 2, 2, 2, 3, 3, 2, 3, 2)$ and $(2, 2, 3, 3, 2, 3, 2, 2, 2, 3, 2)$.

References

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

[2] G. Gordon and E. McDonnell, Trees with the same degree sequence and path numbers, Discrete Mathematics 147 (1995), 297–300.