Polynomial thresholds for graceful leaf extensions of arbitrary trees

Akram Younso

Working preprint, 25 November 2026

hexagon:2611.01141v1

Abstract

Let H be any tree on n vertices, and attach new leaves at its vertices in any prescribed numbers. The resulting tree is graceful whenever its final edge count is at least $3{,}955{,}201n^7$. No graceful labeling of H is assumed, and no positive lower bound is required at an individual attachment site. The proof uses the parities of the same prescribed leaf counts to control two competing constructions. For a tree with h internal vertices the resulting sufficient edge bound is $3{,}955{,}201h^7$. The construction lemmas and the uniform estimate are formalized in Lean.

1 The threshold theorem

A tree with N edges is *graceful* if it admits a bijection from its vertices to $\{0, \dots, N\}$ whose edge differences are exactly $1, \dots, N$. For a finite tree H and a vector $\mathbf{a} \in \mathbb{Z}_{\geq 0}^{V(H)}$, let $H(\mathbf{a})$ be obtained by adding a_v new leaves at each $v \in V(H)$. These are the only new vertices; the construction must realize the entire prescribed target.

Theorem 1. *Let H be a tree on $n \geq 1$ vertices. Every prescribed leaf extension $H(\mathbf{a})$ with final edge count*

$$N = n - 1 + \sum_{v \in V(H)} a_v \;\geq\; 3{,}955{,}201n^7$$

is graceful.

The hypothesis uses the number of vertices before attachment and the number of edges after attachment. It allows all new leaves at a single vertex. In particular, a counterexample with N edges must have more than $(N/3{,}955{,}201)^{1/7}$ internal vertices (Corollary 4). The general Graceful Tree Conjecture remains open.

Superdock proved eventual gracefulness with all new leaves at a single prescribed vertex [2, Corollary 5.10]. Iterating such a result lets later thresholds depend on earlier leaf counts; it does not give the uniform bound above. Lapierre and McGuinness [1, Theorem 1.2 and Corollary 1.3] give constructions from bipartite gracefully labeled cores with conditions on the attachment demands. Here no graceful source labeling or bipartition-demand hypothesis is imposed. We use the classical pair-transfer identity, together with a modular construction controlling several attachment ranks and a two-interval allocation for all secondary sites.

1.1 Proof strategy and notation

Call an injective nonnegative rooted labeling *consistent* if its root is labeled zero and its positive vertex labels, as a multiset, are exactly its edge differences. We call labels *ranks*. Scaling preserves consistency. If a consistent tree has ranks in $[0, N]$, an unused positive

rank may be used for a new root leaf: its label and its edge difference coincide. Thus the task is to meet all nonroot demands before filling the complement at zero.

Two elementary operations preserve this equality of multisets. At a parent of rank p, new children of ranks $x, p-x$, with $0 < x < p$, supply differences $p-x, x$. Also leaves of ranks x, y can be moved between parents of ranks p, q when $x + y = p + q$ and the corresponding absolute differences are exchanged. The two constructions below use only vertices of the prescribed extension. We state their quantitative guarantees explicitly. The linked Lean modules give the detailed labeling constructions.

2 Completion from one or two large supplies

Choose a vertex u of maximum demand, let q count vertices of positive demand, and let o count vertices other than u whose demand is odd. Under the edge hypothesis of Theorem 1, some demand is positive, so $0 \le o < q \le n$. Write $S = n + o$. For $q \ge 2$ put

$$T = (q-1)^2\big(1236n + 1200q - 3764 - 1200\max(0, o-1)\big), \qquad L = (q-1)(T-2).$$

For $q = 1$ set $L = 0$; the two-site case is then absent. Define the seed and packing allowances by

$$A(S) = 4096S^6(9 + 3\lceil \log_2 S \rceil), \qquad P(S, L) = 200S^3\big(2(S-1) + 2L + 1\big).$$

Lemma 2 (Completion alternatives)**.** *If $q \ge 2$ and some $v \ne u$ has demand at least T, the extension is graceful. In the absence of such a site, the extension is graceful whenever $N \ge A(S) + P(S, L)$.*

Proof. In the first case both u and v meet the threshold, by maximality of u. For each of the other positive sites reserve one requested leaf if its demand is odd and two if it is even. If r is the number reserved, then

$$r + o = 2(q-2) + (a_v \bmod 2).$$

Thus the enlarged core has at most $j = n + 2q - 5 - \max(0, o-1)$ edges. The two-site construction uses a modular labeling and transfers leaves between the two chosen sites. For $m = q - 1$ its sufficient demand at each is $40m^2(30j + 25 - 30m) + 36m^2(n+1)$, which expands to T. The palette estimates give enough available ranks in the two intervals, and the construction also handles its exceptional double-broom case.

In the other case include one requested leaf at each odd site off u. The prepared core has order S and every remaining demand off u is even. When $q \ge 2$, T is even, so each such residual demand is at most $T - 2$. Their total is at most L (zero also when $q = 1$). The one-site packing bound, applied in the completion lemma, uses $A(S)$ for the prepared rooted labeling and $P(S, L)$ to make room for the residual pairs. At rank p these pairs have the form $\{x, p-x\}$; unused ranks then complete the demand at the root u. This gives a labeling of the entire prescribed extension. □

For example, form a six-vertex tree from the path $0, 1, 2, 3, 4$ by attaching a leaf 5 at vertex 1. Prescribe respectively 24,209, 24,209, and one new leaf at $0, 4, 5$. With $u = 0$ we have $q = 3$ and $o = 2$. The threshold at each of 0 and 4 is 24,208, so the first case already constructs the target. Ignoring the two off-root parity bits would give a threshold of 29,008. This example is included in the checked regression examples.

2.1 How to place two large supplies

The difficulty is to realize many prescribed parent vertices without using a separate interval of labels for each. Give the two large sites ranks 0 and $P = 2Q$. First work modulo Q, identifying a residue with its negative. A sibling pair at input p receives $y, p-y$. On a chain of three edges the successive ranks $y, p-y, p-2y$ give differences, up to sign, $p-y, p-2y, y$. A remaining single edge can halve the input. Each operation thus reproduces its new ranks as edge classes. Free choices of y keep classes from different operations distinct.

The distinguished vertices are the secondary sites and, when needed, one vertex used to restore the path between the two large sites. There are at most $m = q - 1$ marks. A small centered representative of a marked residue gives an even label near 0 or P on lifting by multiplication by two. The reason for imposing divisibility during the modular construction is that halving an even small integer keeps it small; halving an arbitrary residue modulo Q can put it halfway across the interval.

Here is the count behind the threshold T. Propagate powers of two required by the marks backwards through the operations. On the forward pass a request of exponent e restricts a trial to a progression with step at most 2^e. The marked construction associates w witnessing marks to the request, with $2^e \le 4w^2$. If there are a previously used seeds, their indices are disjoint from those witnesses, so $a + w \le j$. Choose Q coprime to 30030. Avoiding old unsigned classes and coincidences in the new operation excludes at most $30a + 23$ trials (residue exclusion count). For $W = 2m^2(30j + 25 - 30m)$ the interval budget is sufficient because

$$2^e(30a + 25) \le 4w^2(30j + 25 - 30w) \le 4m^2(30j + 25 - 30m) = 2W.$$

The second inequality uses $w \le m$ and $45m \le 30j + 25$, under which the cubic is increasing. With margin $B = 4W$ the palette construction chooses $Q > 4W$ and $B \le N - P < B + 60$. Both supplies of size $5B + 36m^2(n + 1) = T$ suffice for the lift and its completion.

To see how two distinct sites can originate at residue zero, consider the path from u to v. For adjacent sites contract their edge, then restore it with length P. For distance at least three write the path as $u, w, z, \ldots, v$. Delete uw, contract wz to a mark and identify u, v as the root. The deletion separates the two banks, so the quotient is a tree. Orient its marked residue to give a small positive lift t. Assign u, v, z, w the ranks $0, P, t, P+t$; the two restored edges have lengths $P+t, P$. At distance two, a spare neighbor of the middle vertex provides the same four ranks. If the middle vertex has no spare neighbor, a two-edge side path from either end provides the five ranks $0, P - t, P, P + t, t$ on the bank path and the side path, respectively. The special lengths are then $P-t, t, P+t, P$. The path restoration checks both orientations.

If neither distance-two repair applies, the tree consists of a path on three vertices with leaves at its ends. Write a, c for the numbers of leaves there. Give the two ends ranks $0, 1$ and the middle rank $c+2$. Use $2, \ldots, c+1$ at the end labeled 1 and $c+3, \ldots, a+c+2$ at the end labeled 0. The first group gives lengths $1, \ldots, c$, the middle edges give $c + 1, c + 2$, and the second group gives the remaining lengths. This settles the exceptional shape directly.

For the ordinary lift, an unused member of a mirror pair $\{x, P - x\}$ can be placed at 0 or P to supply either of the two lengths. After repairing the singly occupied mirror classes, the remaining secondary demands are even. For a high secondary rank $p = P + d$, choose slots $x, P - y$ with $y = x - d$; their sum is p. For a low rank p, begin with slots provisionally at P and take $y = x - p$. Moving $x, P - y$ to p exchanges their lengths $P - x, y$. In each of two disjoint intervals visit the sites in decreasing order of offset, scanning x upwards. Both x and y increase, so an occupied slot spoils at most one trial in each role over the entire scan. This is why the cost of the used labels is paid once across the secondary sites. The two large sites absorb the remaining slots.

2.2 How to complete one large supply

For the other construction root the prepared tree at u. A useful seed needs both a small maximum M and a lower bound on its positive ranks, since a site of rank p has only $\lfloor (p-1)/2 \rfloor$ distinct pairs $\{x, p-x\}$ available below it. The rooted seed construction obtains

$$M \le A(S), \qquad M \le Rp \quad (p > 0), \qquad R = 200S^3.$$

Here is how the two estimates arise. At each vertex follow a child of largest subtree size. Moving into any other child's subtree at least halves the size, so a root path takes at most $h \le \log_2 S$ such steps. Work first with rational ranks, using reflected sibling pairs and triples along the selected paths. Free integer choices avoid earlier ranks and their relevant doubles and halves. The choice scale is $D = 512S(9 + 3\lceil \log_2 S \rceil)$. The construction maintains the range $[D/100, 2DS^3]$ for positive ranks. Their ratio is at most $200S^3$. In the denominator estimate, at most two bits per level are needed, so multiplication by $2^{2(h+1)} \le 4S^2$ gives integer ranks. Multiplying the upper end of the range by $4S^2$ gives exactly $A(S)$ as an upper bound.

For a prepared core of order at least two, dilate the seed by $K = \lfloor N/M \rfloor$. Since $N \ge M + R(2(S-1) + 2L + 1)$, we have $KM > N - M$ and hence $Kp \ge 2(S-1) + 2L + 1$ for each positive rank. At a site of scaled rank Kp, distinct candidate pairs are disjoint. Each occupied positive rank invalidates at most one candidate there. If the site still needs $2b$ leaves, at most $S - 1 + L - 2b$ ranks have been occupied by the seed and earlier secondary allocations. There are at least $S - 1 + L$ candidates, so one can select b intact pairs. Each pair exchanges its labels as differences; assigning all remaining slots to the root completes the labeling. A singleton prepared core gives a star directly.

3 The uniform estimate

Proof of Theorem 1. By Lemma 2 it suffices to show $A(S) + P(S, L) \le 3{,}955{,}201n^7$. The correlation between S and T matters: an additional odd site enlarges the prepared core but also reduces the two-site reserve. For $n \ge 6$ write $B = 2n - 1$ and $C = 3636n - 2564$. Then $S \le B$. Put

$$U = \max(1, q-1)^2(1236n + 1200q - 1200o - 2564).$$

We have $L \le (q-1)U$ by the formula for T when $q \ge 2$, and $2(S-1) + 1 \le U$ since $o < q$ and $S \le 2n - 1$. Consequently

$$2(S-1) + 2L + 1 \le BU, \qquad U \le (n-1)^2(C - 1200S).$$

The function $s^3(C - 1200s)$ is increasing on $[0, B]$, since its derivative is $3s^2(C - 1600s)$ and $1600B \le C$. Hence the packing cost is at most

$$E(n) = 200(2n-1)^4(n-1)^2(1236n - 1364).$$

The leading coefficient here is $200 \cdot 2^4 \cdot 1236 = 3{,}955{,}200$.

For the seed, set $t = \lfloor S/2^{21} \rfloor$. Since $S \le 2^{21}(t+1) \le 2^{21+t}$ and $S \le 2n$, we get

$$A(S) \le 4096(2n)^6(72 + 3t) \le 18{,}874{,}368n^6 + n^7.$$

For $n \ge 32$, substituting $n = x + 32$ into $3{,}955{,}200n^7 - E(n) - 18{,}874{,}368n^6$ gives a polynomial with nonnegative coefficients. In ascending order of degree these are

$$10{,}901{,}718{,}756{,}768, \quad 44{,}127{,}171{,}891{,}744, \quad 6{,}783{,}308{,}678{,}880,$$
$$423{,}670{,}516{,}320, \quad 13{,}283{,}768{,}320, \quad 208{,}588{,}544, \quad 1{,}311{,}232.$$

This absorbs the degree-six seed term and leaves one further copy of n^7 for the seed. For $6 \le n < 32$ direct integer substitution gives $A(2n-1) + E(n) \le 3{,}955{,}201n^7$. For $1 \le n \le 5$ substitute the finitely many $1 \le q \le n$ and $0 \le o < q$ directly into $A(S) + P(S, L)$. The formal uniform estimate includes the polynomial identity and the exact small-case checks. This proves the bound in both completion cases. □

4 Effective consequences

Corollary 3. *For each finite tree H on n vertices, the least nonnegative integer $B(H)$ such that every extension with at least $B(H)$ new leaves is graceful is computable, and*

$$B(H) \le 3{,}955{,}201n^7 - (n-1).$$

The complete finite list of nongraceful attachment vectors is computable.

Proof. Set $U = 3{,}955{,}201n^7 - (n-1)$. Enumerate all nonnegative integer vectors on $V(H)$ of total less than U. Each resulting finite tree can be tested for gracefulness by enumerating its vertex bijections. Theorem 1 certifies every omitted vector. If no vector fails, return $B(H) = 0$; otherwise return one plus the largest failing total. This proves computability, not practical efficiency of this enumeration. □

Corollary 4. *A nongraceful tree with N edges and h internal vertices must satisfy*

$$N < 3{,}955{,}201h^7, \qquad h > (N/3{,}955{,}201)^{1/7}.$$

For each fixed h, all nongraceful trees with at most h internal vertices can therefore be listed by a finite computation.

Proof. Trees of order at most two are graceful. In any larger tree the internal vertices induce a nonempty subtree, and all other vertices are leaves adjacent to it. Apply Theorem 1 to that core. For bounded h only finitely many trees remain below the displayed edge bound. □

Checking the formal proof

Clone the tagged public source and enter its `lean/` directory. With Elan installed, run `lake exe cache get` followed by `lake build`. The manifest pins Mathlib and the toolchain file selects Lean 4.30.0. The default target compiles the graph-level theorems for the prescribed extension and for the internal-vertex bound. It checks their axiom dependencies against propositional extensionality, classical choice, and quotient soundness.

AI contributions

ChatGPT research agents developed the labeling constructions and quantitative bounds. Dot produced the Lean formalizations. Codex helped write and revise the manuscript, checked the correspondence between the stated results and the formal theorems, and ran the clean build and computational checks. The v1.2 release contains the proof sources, build transcript, and declaration audit.

References

[1] L. Lapierre and S. McGuinness, Constructing trees with graceful labelings using caterpillars, *J. Combin. Math. Combin. Comput.* **94** (2015), 313–321.

[2] M. Superdock, The Graceful Tree Conjecture: a class of graceful diameter-6 trees, arXiv:1403.1564 (2014).