Asymptotic enumeration of rooted binary unlabeled galled trees with a fixed number of galls

Lily Agranat-Tamir [1]
Department of Biology, Stanford University, Stanford, CA 94305, USA

Michael Fuchs
Department of Mathematical Sciences, National Chengchi University, Taipei 116, Taiwan

Bernhard Gittenberger
Department of Discrete Mathematics and Geometry, Technische Universität Wien, Wiedner Hauptstraße 8-10/104, A-1040 Wien Austria

Noah A. Rosenberg
Department of Biology, Stanford University, Stanford, CA 94305, USA

Abstract

Galled trees appear in problems concerning admixture, horizontal gene transfer, hybridization, and recombination. Building on a recursive enumerative construction, we study the asymptotic behavior of the number of rooted binary unlabeled (normal) galled trees as the number of leaves n increases, maintaining a fixed number of galls g. We find that the exponential growth with n of the number of rooted binary unlabeled normal galled trees with g galls has the same value irrespective of the value of $g \geq 0$. The subexponential growth, however, depends on g; it follows $c_g n^{2g-3/2}$, where c_g is a constant dependent on g. Although for each g, the exponential growth is approximately 2.4833^n, summing across *all* g, the exponential growth is instead approximated by the much larger 4.8230^n.

2012 ACM Subject Classification Mathematics of computing, enumeration; Mathematics of computing, generating functions.

Keywords and phrases galled trees, generating functions, phylogenetics, unlabeled trees.

Digital Object Identifier 10.4230/LIPIcs.AofA.2024.19

Funding *Michael Fuchs*: National Science and Technology Council grant NSTC-111-2115-M-004-002-MY2
Noah A. Rosenberg: National Science Foundation grant BCS-2116322.

1 Introduction

Rooted binary trees are a staple of mathematical phylogenetic analysis, as they are used to represent diverse biological processes taking place in time—including the evolution of species, the evolution of genes among those species, and the divergence of populations [9, 21, 24]. The root represents a common ancestor, and the leaves represent subsequent biological entities, often in the present day. Viewed as objects evolving in time, by extension of "vertical" inheritance that occurs in genetic transmission from parents to offspring, biological divergences are viewed as taking place vertically on the tree. Mathematical phylogenetic analyses of trees have produced rich contributions to algorithmic and combinatorial studies.

Certain evolutionary events, however, involve *merging* rather than *divergence* of biological lineages. Such events include the recombination that occurs during gamete formation, population admixture, horizontal gene transfer, and hybridization. To describe processes that include these events, we must look beyond trees to *phylogenetic networks* [14, 17, 18].

[1] corresponding author

35th International Conference on Probabilistic, Combinatorial and Asymptotic Methods for the Analysis of Algorithms.
Editors: C. Mailler and S. Wild; Article No. 19; pp. 19:1–19:14
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

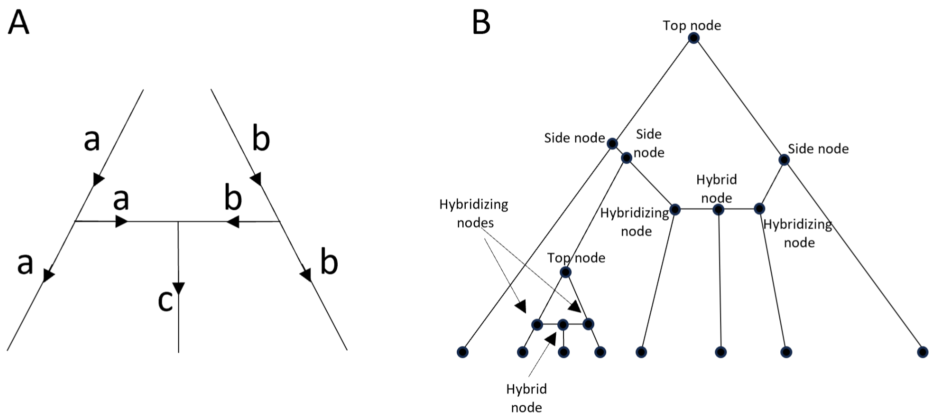

Figure 1 Features of a gall in a galled tree. (A) A gall as a representation of a biological merging event. Biological lineages a and b each bifurcate, with one branch of each bifurcation merging to form lineage c. (B) Nomenclature for the various nodes in a gall.

Among the phylogenetic networks, galled trees are some of the simplest. As their name suggests, they are tree-like, but they can contain certain internal nodes with in-degree 2 and out-degree 1, representing permitted classes of mergings. Galled trees are named for the growths, termed *galls*, which appear in plants but which do not disrupt their tree-like structure. They were first introduced in the study of recombination [15, 16, 23].

Mathematically, a galled tree allows each vertex or edge in a graph to be contained in at most one cycle. An additional requirement is needed for galled trees to be meaningful for biological processes such as hybridization. In a hybridization event, two biological lineages, a and b, each bifurcate; a merging event occurs between two branches, one from each bifurcation, producing a third lineage, c (Figure 1A). The structure of the event requires that when viewed graphically, a gall—a cycle in the graph—contains at least four nodes. These include a *top node*, two *hybridizing nodes*, and one *hybrid node*. Additional *side nodes* are permitted, and we regard the hybridizing nodes as special side nodes (Figure 1B). The requirement that galls have at least these four nodes (i.e. the top node must not be a hybridizing node) is equivalent to a requirement that galled trees be *normal*.

Many enumerative problems on galled trees have been investigated [3, 4, 5, 22]; this study concerns rooted binary unlabeled normal galled (non-plane) trees. Given number of galls g, as the number of leaves $n \to \infty$, what is the growth of the size of this class? The case of $g = 0$ is the enumeration of rooted binary unlabeled trees, and we previously studied $g = 1$ [1]. Building on a recurrence for rooted binary unlabeled normal galled trees with n leaves and g galls, we obtain a generating function for $g = 2$. We find the asymptotic behavior of the number of trees with n leaves and $g = 2$ galls, and we obtain asymptotics for each $g > 2$. In our main result, Theorem 10, we report that the number of galled trees with n leaves and g galls has the form $\beta_g n^{2g-\frac{3}{2}}\rho^{-n}$, where ρ is the radius of convergence of the generating function for the $g = 0$ case, and β_g is a constant that depends solely on g.

2 Definitions

We define our concepts formally. We assume that all networks and trees are binary; we usually drop the term *binary*. A *rooted phylogenetic network* is a directed acyclic graph in which four properties hold. (i) There exists a unique node with in-degree 0 and out-degree

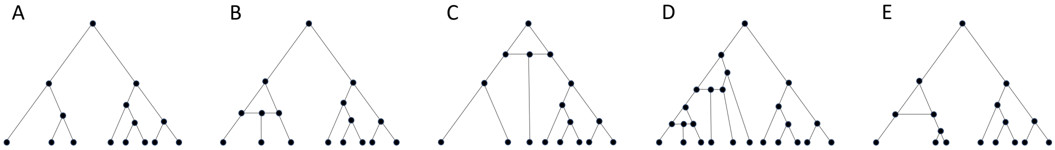

Figure 2 Rooted binary unlabeled galled trees. (A) A tree with no galls. (B) A galled tree with one gall. (C) A galled tree with a root gall. (D) A galled tree with two galls. (E) A galled tree that is not a normal galled tree and that is not included in the class of galled trees that we enumerate.

2. This node is the *root node*. (ii) *Leaf nodes* possess in-degree 1 and out-degree 0. (iii) Non-leaf, non-root nodes possess in-degree 2 and out-degree 1 or in-degree 1 and out-degree 2. (iv) Edges are directed away from the root. Nodes with in-degree 2 and out-degree 1 are *reticulation nodes* (or *hybrid nodes*). Nodes with in-degree 1 and out-degree 2 are *tree nodes*.

A *rooted galled tree* is a rooted phylogenetic network with three additional properties. (v) Each reticulation node a_r has a unique ancestor node r so that exactly two non-overlapping paths of edges connect r to a_r. Ignoring the direction of the edges, the two paths from r to a_r produce a cycle C_r. The cycle is termed a *gall*. (vi) Consider galls C_r and C_s, associated with reticulation nodes a_r and a_s, $a_r \neq a_s$. The sets of nodes in the galls C_r and C_s are disjoint. (vii) Ancestor node r and reticulation node a_r are separated by two or more edges. Condition (vii) encodes the requirement that we consider only *normal* galled trees (Figure 2).

We generally drop the terms *rooted* and *normal*, and refer only to *galled trees*, and where a distinction is necessary, *labeled* and *unlabeled* galled trees. Although a galled tree is not technically a tree due to the presence of cycles, we continue to refer to galled trees as trees. We similarly refer to the galled trees rooted at internal nodes of a galled tree as *subtrees*. Our view of galls as representations of biological merging events leads us to depict hybridizing nodes and their associated hybrid node on a horizontal line, representing the simultaneity of these nodes when a galled tree is taken to represent a structure evolving in time [2, 20].

A basic result describes the maximal number of galls possible in a galled tree with n leaves. A gall contains three or more descendant subtrees: one from the reticulation node, two from the hybridizing nodes, and one for each additional side node. Hence, the smallest galled tree possesses $n = 3$ leaves. Adding a gall to a galled tree involves replacing one subtree with at least three subtrees, so that each gall adds at least two leaves. For a tree with g galls, the number of leaves satisfies $n \geq 2g + 1$, or $g \leq \lfloor \frac{n-1}{2} \rfloor$ [20].

We will need to consider *compositions*, ordered lists of positive integers that sum to a specified value. We denote by $C(a, b)$ the compositions of a natural number a into b parts. $C(a, b)$ is the set of ordered lists of positive integers of length b, $(i_1, i_2, \ldots, i_b)$, with sum equal to a. We denote by $C_p(a, b)$ the subset of $C(a, b)$ containing the *palindromic* compositions of a, that is, the compositions $(i_1, i_2, \ldots, i_b)$ for which $i_j = i_{b-j+1}$ for each j from 1 to b.

3 Previous work

We review a number of results. The rooted binary unlabeled galled trees generalize the rooted binary unlabeled trees without galls. Letting U_n denote the number of rooted binary unlabeled trees with no galls and letting $\mathcal{U}(t)$ denote the generating function $\sum_{n \geq 0} U_n t^n$,

$$\mathcal{U}(t) = t + \frac{1}{2}\mathcal{U}^2(t) + \frac{1}{2}\mathcal{U}(t^2). \tag{1}$$

Denoting the radius of convergence by ρ, as $t \to \rho^-$, we have $\mathcal{U}(t) \sim 1 - \gamma\sqrt{1 - t/\rho}$, where $\gamma \approx 1.1300$ and $\rho \approx 0.4027$ [8, p. 55] [10, pp. 476-477]. The asymptotic approximation for the number of rooted binary unlabeled trees (with no galls) is,

$$U_n = [t^n]\mathcal{U}(t) \sim \frac{\gamma}{2\Gamma(\frac{1}{2})} n^{-\frac{3}{2}} \rho^{-n}. \tag{2}$$

In our previous work on rooted binary unlabeled normal galled trees [1] (henceforth "unlabeled galled trees"), we obtained a recursion enumerating the A_n unlabeled galled trees with n leaves and another recursion enumerating the $E_{n,g}$ unlabeled galled trees with a specified number of galls g. We specifically considered the case of $g = 1$. We also studied the asymptotics of A_n and $E_{n,1}$ through their generating functions. The generating function for unlabeled galled trees, considering all possible numbers of galls, was found to be [1, eq. 36]

$$\mathcal{A}(t) = t + \frac{1}{2}\mathcal{A}^2(t) + \frac{1}{2}\mathcal{A}(t^2) + 1 - \frac{1}{1 - \mathcal{A}(t)} + \frac{\mathcal{A}(t)}{2[1 - \mathcal{A}(t)]^2} + \frac{\mathcal{A}(t)}{2[1 - \mathcal{A}(t^2)]}. \tag{3}$$

The three leftmost terms, identical to the generating function $\mathcal{U}(t)$ (eq. (1)), arise from the galled trees in which two subtrees descend immediately from the root. The other terms arise from galled trees with a gall that contains the root, a *root gall*.

Using the *asymptotics of implicit tree-like classes* theorem [10, pp. 467-468], we obtained the asymptotics of the number of galled trees with n leaves, A_n [1, eq. 42]: $A_n = [t^n]\mathcal{A}(t) \sim \left[\delta/\left(2\Gamma(\frac{1}{2})\right)\right] n^{-\frac{3}{2}} \alpha^{-n}$, where $\delta \approx 0.2793$ and $\alpha \approx 0.2073$. $\mathcal{A}(t)$ has convergence radius about half that of $\mathcal{U}(t)$, so that galled trees are much more numerous than the trees without galls.

We also derived the generating function $\mathcal{E}_1(t)$ and asymptotic growth of the number of unlabeled galled trees with exactly one gall. We state these results as propositions.

▶ **Proposition 1.** [1, eq. 48] *The generating function $\mathcal{E}_1(t)$ for the number of unlabeled galled trees with 1 gall satisfies*

$$\mathcal{E}_1(t) = \frac{1}{1 - \mathcal{U}(t)} - \frac{1}{[1 - \mathcal{U}(t)]^2} + \frac{\mathcal{U}(t)}{2[1 - \mathcal{U}(t)]^3} + \frac{\mathcal{U}(t)}{2[1 - \mathcal{U}(t)][1 - \mathcal{U}(t^2)]}. \tag{4}$$

▶ **Proposition 2.** [1, eq. 50] *The asymptotic growth of the number $E_{n,1}$ of unlabeled galled trees with n leaves and 1 gall satisfies*

$$E_{n,1} \sim \frac{1}{2\gamma^3\Gamma(\frac{3}{2})} n^{\frac{1}{2}} \rho^{-n} = \frac{1}{\gamma^3\sqrt{\pi}} n^{\frac{1}{2}} \rho^{-n}. \tag{5}$$

Proposition 2 follows from the fact that as $t \to \rho^-$, $\mathcal{E}_1(t) \sim 1/[2\gamma^3(1 - t/\rho)^{\frac{3}{2}}]$. $\mathcal{E}_1(t)$ in eq. (4) depends on $\mathcal{U}(t)$. Eq. (5) clarifies that the exponential growth of the number of unlabeled galled trees with one gall is the same as that of the number of unlabeled galled trees with no galls; only the subexponential growth differs. We will generalize this result.

4 Recursion

4.1 Recursion for g galls, $E_{n,g}$

In [1, eq. 27], we obtained a recursion for $E_{n,g}$, the number of unlabeled galled trees with n leaves and exactly g galls; Table 3 reported the numerical values $E_{n,g}$ up to $n = 18$. The base cases are $E_{1,0} = 1$ and $E_{1,g} = 0$ for $g \geq 1$. We also write $E_{m,\ell} = 0$ when m is not a positive integer, ℓ is not a positive integer, or both.

▶ **Proposition 3.** *For* (n, g) *with* $n \geq 2$ *and* $0 \leq g \leq \lfloor \frac{n-1}{2} \rfloor$*, the number of unlabeled galled trees with* n *leaves and* g *galls is*

$$E_{n,g} = \frac{1}{2}\Bigg[\Bigg(\sum_{\mathbf{c}\in C(n,2)}\sum_{\mathbf{d}\in C(g+2,2)}\prod_{i=1}^{2} E_{c_i,d_i-1}\Bigg) + E_{\frac{n}{2},\frac{g}{2}} \tag{6}$$

$$+\Bigg(\sum_{k=3}^{n}(k-2)\sum_{\mathbf{c}\in C(n,k)}\sum_{\mathbf{d}\in C(g-1+k,k)}\prod_{i=1}^{k} E_{c_i,d_i-1}\Bigg) \tag{7}$$

$$+\Bigg(\sum_{a=1}^{\lfloor\frac{n-1}{2}\rfloor}\sum_{\mathbf{c}\in C_p(n,2a+1)}\sum_{\mathbf{d}\in C_p\left(g-1+(2a+1),2a+1\right)}\prod_{i=1}^{a+1} E_{c_i,d_i-1}\Bigg)\Bigg]. \tag{8}$$

The approach is to use a recursion at the root node. We sum over all products of possible counts of subtrees, each with fewer than n leaves. Pairs of galled trees that are reflections of one another over the root—or the axis connecting the top node to the reticulation node of the root gall—are the same unlabeled galled tree, explaining the leading $\frac{1}{2}$. We add back terms for galled trees that are symmetric over the root, which are not double-counted.

Line (6) in Proposition 3 enumerates galled trees with n leaves and g galls that do not have a root gall. The first term traverses combinations of numbers of leaves in the two subtrees summing to n by traversing compositions $\mathbf{c}$ of n into 2 parts $\big(\mathbf{c} \in C(n, 2)\big)$. It also traverses combinations of placements of the g galls in the two subtrees. Because subtrees can possess 0 galls, these combinations are identified from compositions of $g + 2$ into 2 parts, subtracting 1 gall in each part $\big(\mathbf{d} \in C(g + 2, 2)\big)$. The second term adds back the galled trees with identical subtrees; this term is nonzero only if both n and g are even.

Line (7) counts galled trees with n leaves and g galls that do have a root gall. It traverses the possible number k of subtrees descending from side nodes, hybridizing nodes, and the hybrid node of the root gall (3 to n, the number of leaves). It then traverses the $k-2$ possible nodes in the root gall where the hybrid node can be placed: all k nodes except immediate descendants of the root. We then traverse the possible combinations of the n leaves and $g - 1$ remaining (non-root) galls into the k subtrees, again allowing subtrees with no galls.

Line (8) adds back half the galled trees with n leaves and g galls that have a root gall and that are symmetric over the reticulation node. Here, a is the possible number of subtrees of the root gall on each side of the reticulation node, so that the root gall has $2a + 1$ subtrees in total. The composition of the n leaves into $2a + 1$ subtrees and the composition of the $g - 1$ galls into those subtrees are both palindromic. Given these compositions, a tree is specified by its subtrees of one side of the reticulation node and the subtree of the reticulation node.

4.2 Recursion for two galls, $E_{n,2}$

For $g = 2$, for $n \geq 2$, the recursion for $E_{n,g}$ becomes

$$E_{n,2} = \frac{1}{2}\Bigg[\Bigg(\sum_{c=1}^{n-1}\sum_{d=0}^{2} E_{c,d}E_{n-c,2-d}\Bigg) + E_{\frac{n}{2},1}$$

$$+\sum_{k=3}^{n}(k-2)\sum_{\mathbf{c}\in C(n,k)}\sum_{\mathbf{d}\in C(k+1,k)}\prod_{i=1}^{k} E_{c_i,d_i-1}$$

$$+\sum_{a=1}^{\lfloor\frac{n-1}{2}\rfloor}\sum_{\mathbf{c}\in C_p(n,2a+1)}\sum_{\mathbf{d}\in C_p(2a+2,2a+1)}\prod_{i=1}^{a+1} E_{c_i,d_i-1}\Bigg]$$

$$= \frac{1}{2}\Bigg[\Bigg(2\sum_{m=1}^{n-1} U_m E_{n-m,2} + \sum_{m=1}^{n-1} E_{m,1}E_{n-m,1}\Bigg) + E_{\frac{n}{2},1}$$
$$+ \sum_{k=3}^{n}(k-2) \sum_{m=k-1}^{n-1} \sum_{\mathbf{c}\in C(m,k-1)} \Bigg(\prod_{i=1}^{k-1} U_{c_i}\Bigg) k E_{n-m,1}$$
$$+ \sum_{a=1}^{\lfloor \frac{n-1}{2} \rfloor} \sum_{\mathbf{c}\in C_p(n,2a+1)} \Bigg(\prod_{i=1}^{a} U_{c_i}\Bigg) E_{c_{a+1},1}\Bigg]. \tag{9}$$

Recall here that $E_{m,1} = 0$ if $m \notin \mathbb{N}$. In the first line, m gives the number of leaves in the "left" subtree of the root and $n-m$ is the number in the "right" subtree (the left–right distinction is solely for convenience, as we consider non-plane trees, in which the particular embedding of a tree in the plane is disregarded). In the second line, k is the number of subtrees of the root gall, m is the number of leaves across the $k-1$ subtrees of the root gall that *do not* contain a gall, and $n-m$ is the number of leaves in the subtree with the second gall.

5 Analysis of $E_{n,2}$

5.1 Generating function

Using the recursion in eq. (9), we now find the generating function of $E_{n,2}$, which we define by $\mathcal{E}_2(t) = \sum_{n\geq 0} E_{n,2}t^n$. Eq. (9) holds for all $n \geq 0$ because $E_{n,2} = 0$ for $n \leq 4$ and $E_{n,1} = 0$ for $n \leq 2$. We can add terms involving U_0, $E_{0,1}$, and $E_{0,2}$, all of which equal zero. Then

$$\mathcal{E}_2(t) = \sum_{n\geq 0} E_{n,2}t^n = \frac{1}{2}\Bigg[\underbrace{\sum_{n\geq 0}\Bigg(\Bigg(2\sum_{m=0}^{n} U_m E_{n-m,2}\Bigg) + \Bigg(\sum_{m=0}^{n} E_{m,1}E_{n-m,1}\Bigg) + E_{\frac{n}{2},1}\Bigg)t^n}_{\mathcal{E}_{2_{\mathrm{i}}}(t)}$$
$$+ \underbrace{\sum_{n\geq 0}\Bigg(\sum_{k=3}^{n}(k-2)k \sum_{m=k-1}^{n-1} \sum_{\mathbf{c}\in C(m,k-1)} \Bigg(\prod_{i=1}^{k-1} U_{c_i}\Bigg) E_{n-m,1}\Bigg)t^n}_{\mathcal{E}_{2_{\mathrm{ii}}}(t)}$$
$$+ \underbrace{\sum_{n\geq 0}\Bigg(\sum_{a=1}^{\lfloor \frac{n-1}{2} \rfloor} \sum_{\mathbf{c}\in C_p(n,2a+1)} \Bigg(\prod_{i=1}^{a} U_{c_i}\Bigg) E_{c_{a+1},1}\Bigg)t^n}_{\mathcal{E}_{2_{\mathrm{iii}}}(t)}\Bigg]. \tag{10}$$

We now simplify the three terms of $\mathcal{E}_2(t)$:

$$\mathcal{E}_{2_{\mathrm{i}}}(t) = 2\sum_{m\geq 0}\sum_{n\geq m}(U_m t^m)(E_{n-m,2}t^{n-m}) + \sum_{m\geq 0}\sum_{n\geq m}(E_{m,1}t^m)(E_{n-m,1}t^{n-m}) + \sum_{n\geq 0} E_{\frac{n}{2},1}t^n$$
$$= 2\sum_{m\geq 0}(U_m t^m)\sum_{\ell\geq 0}(E_{\ell,2}t^\ell) + \sum_{m\geq 0}(E_{m,1}t^m)\sum_{\ell\geq 0}(E_{\ell,1}t^\ell) + \sum_{n\geq 0} E_{n,1}t^{2n}$$
$$= 2\mathcal{U}(t)\,\mathcal{E}_2(t) + \mathcal{E}_1^2(t) + \mathcal{E}_1(t^2). \tag{11}$$

For $\mathcal{E}_{2_{\mathrm{ii}}}(t)$, we obtain

$$\begin{aligned}
\mathcal{E}_{2_{\mathrm{ii}}}(t) &= \sum_{k\geq 3}(k-2)k \sum_{m\geq k-1} \sum_{\mathbf{c}\in C(m,k-1)} \prod_{i=1}^{k-1} U_{c_i} t^{c_i} \sum_{n\geq m} E_{n-m,1} t^{n-m} \\
&= \sum_{k\geq 3}(k-2)k \sum_{i_1\geq 0} \sum_{i_2\geq 0} \cdots \sum_{i_{k-1}\geq 0} U_{i_1} U_{i_2} \cdots U_{i_{k-1}} t^{i_1+i_2+\ldots+i_{k-1}} \sum_{\ell\geq 0} E_{\ell,1} t^{\ell} \\
&= \sum_{k\geq 3}(k-2)k\,\mathcal{U}^{k-1}(t)\,\mathcal{E}_1(t) = \mathcal{E}_1(t)\Big[\sum_{k\geq 2}(k^2-1)\,\mathcal{U}^k(t)\Big] \\
&= \mathcal{E}_1(t)\Big[\Big(\sum_{k\geq 0} k^2 \mathcal{U}^k(t)\Big) - \mathcal{U}(t) - \Big(\sum_{k\geq 0} \mathcal{U}^k(t)\Big) + 1 + \mathcal{U}(t)\Big] \\
&= \mathcal{E}_1(t)\left[\frac{\mathcal{U}(t)+\mathcal{U}^2(t)}{[1-\mathcal{U}(t)]^3} - \frac{1}{1-\mathcal{U}(t)} + 1\right].
\end{aligned} \tag{12}$$

Finally, $\mathcal{E}_{2_{\mathrm{iii}}}(t)$ becomes

$$\begin{aligned}
\mathcal{E}_{2_{\mathrm{iii}}}(t) &= \sum_{a\geq 1} \sum_{m\geq a} \sum_{\mathbf{c}\in C(m,a)} \prod_{i=1}^{a} U_{c_i} t^{2c_i} \sum_{n\geq 2m} E_{n-2m,1} t^{n-2m} \\
&= \sum_{a\geq 1} \sum_{i_1\geq 0} \sum_{i_1\geq 0} \cdots \sum_{i_a\geq 0} U_{i_1} U_{i_2} \cdots U_{i_a} t^{2i_1+2i_2+\ldots+2i_a} \sum_{\ell\geq 0} E_{\ell,1} t^{\ell} \\
&= \sum_{a\geq 1} \mathcal{U}^a(t^2)\,\mathcal{E}_1(t) = \frac{\mathcal{E}_1(t)}{1-\mathcal{U}(t^2)} - \mathcal{E}_1(t).
\end{aligned} \tag{13}$$

Summing the three parts, we obtain the following proposition.

▶ **Proposition 4.** *The generating function $\mathcal{E}_2(t)$ for the number of unlabeled galled trees with 2 galls satisfies*

$$\mathcal{E}_2(t) = \frac{\mathcal{E}_1(t)}{2[1-\mathcal{U}(t)]}\left[\mathcal{E}_1(t) + \frac{\mathcal{U}(t)+\mathcal{U}^2(t)}{[1-\mathcal{U}(t)]^3} - \frac{1}{1-\mathcal{U}(t)} + \frac{1}{1-\mathcal{U}(t^2)}\right] + \frac{\mathcal{E}_1(t^2)}{2[1-\mathcal{U}(t)]}. \tag{14}$$

5.2 Asymptotic analysis

To analyze the asymptotics of $\mathcal{E}_2(t)$ as $t \to \rho^-$, we take the highest-order terms in Proposition 4, that is, the terms with the highest power of $1 - t/\rho$ in the denominator. We recall $\mathcal{U}(t) \sim 1 - \gamma\sqrt{1-t/\rho}$. From Proposition 1, $\mathcal{E}_1(t) \sim 1/[2\gamma^3(1-t/\rho)^{\frac{3}{2}}]$. We have:

$$\mathcal{E}_2(t) \sim \frac{\mathcal{E}_1^2(t)}{2[1-\mathcal{U}(t)]} + \frac{2\mathcal{E}_1(t)}{2[1-\mathcal{U}(t)]^4} = \frac{5}{8\gamma^7(1-t/\rho)^{7/2}}. \tag{15}$$

To obtain a result for the coefficients $E_{n,2}$, we use the transfer formula (Corollary VI.1, page 392 and Theorem VI.4, page 393 in [10])—according to which, if $f(t)$ is Δ-analytic with a singularity at b, and $f(t) \sim (1-\frac{t}{b})^{-a}$ as $\frac{t}{b} \to 1$ with t in Δ, and $a \notin \{0,-1,-2,\ldots\}$, then $[t^n]f(t) \sim n^{a-1}b^{-n}/\Gamma(a)$. Here, ρ fulfills the role of b and $\frac{7}{2}$ that of a.

▶ **Proposition 5.** *The asymptotic growth of the number $E_{n,2}$ of unlabeled galled trees with n leaves and 2 galls satisfies*

$$E_{n,2} \sim \frac{5}{8\gamma^7\Gamma(\frac{7}{2})} n^{\frac{5}{2}} \rho^{-n} = \frac{1}{3\gamma^7\sqrt{\pi}} n^{\frac{5}{2}} \rho^{-n}. \tag{16}$$

We note the appearance of ρ^{-n} and $n^{5/2}$ to obtain the following corollary.

▶ **Corollary 6.** *The exponential growth of $\mathcal{E}_2(t)$ is the same as that of $\mathcal{U}(t)$ and $\mathcal{E}_1(t)$; however, its subexponential growth is greater.*

6 Analysis of $E_{n,g}$

6.1 Generating function

We denote the generating function of the number of galled trees with exactly g galls by $\mathcal{E}_g(t) = \sum_{n\geq 0} E_{n,g}t^n$. Similarly to the case of $g = 2$, we use the recursion we had calculated for $E_{n,g}$ in Proposition 3 to derive the generating function. From Proposition 3, we can decompose the generating function by

$$\mathcal{E}_g(t) = \frac{1}{2}\Bigg[\underbrace{\sum_{n\geq 0}\Bigg(\Bigg(\sum_{\mathbf{c}\in C(n,2)}\sum_{\mathbf{d}\in C(g+2,2)}\prod_{i=1}^{2}E_{c_i,d_i-1}\Bigg) + E_{\frac{n}{2},\frac{g}{2}}\Bigg)t^n}_{\mathcal{E}_{g_{\text{i}}}(t)}$$
$$+ \underbrace{\sum_{n\geq 0}\Bigg(\sum_{k=3}^{n}(k-2)\sum_{\mathbf{c}\in C(n,k)}\sum_{\mathbf{d}\in C(g-1+k,k)}\prod_{i=1}^{k}E_{c_i,d_i-1}\Bigg)t^n}_{\mathcal{E}_{g_{\text{ii}}}(t)}$$
$$+ \underbrace{\sum_{n\geq 0}\Bigg(\sum_{a=1}^{\lfloor\frac{n-1}{2}\rfloor}\sum_{\mathbf{c}\in C_p(n,2a+1)}\sum_{\mathbf{d}\in C_p(g-1+2a+1,2a+1)}\prod_{i=1}^{a+1}E_{c_i,d_i-1}\Bigg)t^n}_{\mathcal{E}_{g_{\text{iii}}}(t)}\Bigg]. \quad (17)$$

where $E_{n,g} = 0$ for pairs with $n = 0$ or $n = 1$ and $g \geq 1$. The terms in the decomposition are

$$\mathcal{E}_{g_{\text{i}}}(t) = 2\sum_{m\geq 0}\sum_{n\geq m}(U_m t^m)(E_{n-m,g}t^{n-m}) + \sum_{j=1}^{g-1}\sum_{m\geq 0}\sum_{n\geq m}(E_{m,j}t^m)(E_{n-m,g-j}t^{n-m})$$
$$+ \sum_{n\geq 0}E_{\frac{n}{2},\frac{g}{2}}t^n$$
$$\mathcal{E}_{g_{\text{ii}}}(t) = \sum_{\ell=1}^{g-1}\sum_{k\geq 3}(k-2)\binom{k}{\ell}\sum_{m\geq k-\ell}\sum_{\mathbf{c}\in C(m,k-\ell)}\prod_{i=1}^{k-\ell}U_{c_i}t^{c_i}$$
$$\times\sum_{n\geq m}\sum_{\tilde{\mathbf{c}}\in C(n-m,\ell)}\sum_{\mathbf{d}\in C(g-1,\ell)}\prod_{j=1}^{\ell}E_{\tilde{c}_j,d_j}t^{\tilde{c}_j} \quad (18)$$
$$\mathcal{E}_{g_{\text{iii}}}(t) = \sum_{\ell=0}^{\lfloor\frac{g-1}{2}\rfloor}\sum_{a\geq 1}\binom{a}{\ell}\sum_{m_1\geq a-\ell}\sum_{\mathbf{c}\in C(m_1,a-\ell)}\prod_{i=1}^{a-\ell}U_{c_i}t^{2c_i}$$
$$\times\sum_{m\geq m_1+\ell}\sum_{\tilde{\mathbf{c}}\in C(m-m_1,\ell)}\sum_{b=\ell}^{\lfloor\frac{g-1}{2}\rfloor}\sum_{\mathbf{d}\in C(b,\ell)}\prod_{j=1}^{\ell}E_{\tilde{c}_j,d_j}t^{2c_j}\sum_{n\geq 2m}E_{n-2m,g-1-2b}t^{n-2m}, \quad (19)$$

where it is convenient to denote U_n by $E_{n,0}$ for terms with $g - 1 - 2b = 0$ in $\mathcal{E}_{g_{\text{iii}}}(t)$.

In $\mathcal{E}_{g_{\text{i}}}(t)$, j is the number of galls in the left subtree of the root, supposing both subtrees possess at least one gall. In $\mathcal{E}_{g_{\text{ii}}}(t)$, ℓ is the number of subtrees of the root gall that possess at least one gall; k is the number of subtrees of the root gall, so that $\binom{k}{\ell}$ counts ways to select which ℓ subtrees possess galls; and m is the number of leaves in the $k - \ell$ remaining subtrees.

Similarly, in $\mathcal{E}_{g_{\text{iii}}}(t)$, for symmetric root galls, ℓ is the number of subtrees of the left side of the root gall that contain galls; a is the number of subtrees of the left side of the root gall;

m_1 is the sample size in the $a-\ell$ subtrees that do not possess galls; $m-m_1$ is the sample size in the ℓ subtrees that do possess galls; and b is the number of galls in those ℓ subtrees.

We now solve each part of the decomposition:

$$\begin{aligned}\mathcal{E}_{g_{\mathrm{i}}}(t) &= 2\sum_{m\ge 0}(U_m t^m)\sum_{\ell\ge 0}(E_{\ell,g}t^\ell) + \sum_{j=1}^{g-1}\sum_{m\ge 0}(E_{m,j}t^m)\sum_{\ell\ge 0}(E_{\ell,g-j}t^\ell) + \sum_{n\ge 0}E_{n,\frac{g}{2}}t^{2n}\\ &= 2\mathcal{U}(t)\,\mathcal{E}_g(t) + \left(\sum_{j=1}^{g-1}\mathcal{E}_j(t)\,\mathcal{E}_{g-j}(t)\right) + \mathcal{E}_{\frac{g}{2}}(t^2). \qquad (20)\end{aligned}$$

where $\mathcal{E}_\ell(t)=0$ for $\ell\notin\mathbb{N}$. The second part produces

$$\begin{aligned}\mathcal{E}_{g_{\mathrm{ii}}}(t) &= \sum_{\ell=1}^{g-1}\sum_{k\ge\max(\ell,3)}(k-2)\binom{k}{\ell}\sum_{i_1\ge 0}\sum_{i_2\ge 0}\cdots\sum_{i_{k-\ell}\ge 0}U_{i_1}U_{i_2}\cdots U_{i_{k-\ell}}t^{i_1+i_2+\ldots+i_{k-\ell}}\\ &\quad\times\sum_{\mathbf{d}\in C(g-1,\ell)}\sum_{j_1\ge 0}\sum_{j_2\ge 0}\cdots\sum_{j_\ell\ge 0}E_{j_1,d_1}E_{j_2,d_2}\cdots E_{j_\ell,d_\ell}t^{j_1+j_2+\ldots+j_\ell}\\ &= \sum_{\ell=1}^{g-1}\left(\sum_{k\ge\max(\ell,3)}(k-2)\binom{k}{\ell}\mathcal{U}^{k-\ell}(t)\right)\sum_{\mathbf{d}\in C(g-1,\ell)}\prod_{j=1}^{\ell}\mathcal{E}_{d_j}(t)\\ &= \sum_{\ell=1}^{g-1}\left(\frac{3\mathcal{U}(t)-2+\ell}{[1-\mathcal{U}(t)]^{\ell+2}} + [\![\ell=1]\!]\right)\sum_{\mathbf{d}\in C(g-1,\ell)}\prod_{j=1}^{\ell}\mathcal{E}_{d_j}(t). \qquad (21)\end{aligned}$$

Here, $[\![\cdot]\!]$ denotes the Iverson bracket. Finally, for the third part,

$$\begin{aligned}\mathcal{E}_{g_{\mathrm{iii}}}(t) &= \sum_{\ell=0}^{\lfloor\frac{g-1}{2}\rfloor}\sum_{a\ge 1}\binom{a}{\ell}\sum_{i_1\ge 0}\sum_{i_2\ge 0}\cdots\sum_{i_{a-\ell}\ge 0}U_{i_1}U_{i_2}\cdots U_{i_{a-\ell}}t^{2i_1+2i_2+\ldots+2i_{a-\ell}}\\ &\quad\times\sum_{b=\ell}^{\lfloor\frac{g-1}{2}\rfloor}\sum_{\mathbf{d}\in C(b,\ell)}\sum_{j_1\ge 0}\sum_{j_2\ge 0}\cdots\sum_{j_\ell\ge 0}E_{j_1,d_1}E_{j_2,d_2}\cdots E_{j_\ell,d_\ell}t^{2j_1+2j_2+\ldots 2j_\ell}\\ &\quad\times\sum_{j\ge 0}E_{j,g-1-2b}t^j\\ &= \sum_{\ell=0}^{\lfloor\frac{g-1}{2}\rfloor}\left(\sum_{a\ge 1}\binom{a}{\ell}\mathcal{U}^{a-\ell}(t^2)\right)\sum_{b=\ell}^{\lfloor\frac{g-1}{2}\rfloor}\sum_{\mathbf{d}\in C(b,\ell)}\left(\prod_{j=1}^{\ell}\mathcal{E}_{d_j}(t^2)\right)\mathcal{E}_{g-1-2b}(t)\\ &= \sum_{\ell=0}^{\lfloor\frac{g-1}{2}\rfloor}\left(\frac{1}{[1-\mathcal{U}(t^2)]^{\ell+1}} - [\![\ell=0]\!]\right)\sum_{b=\ell}^{\lfloor\frac{g-1}{2}\rfloor}\sum_{\mathbf{d}\in C(b,\ell)}\left(\prod_{j=1}^{\ell}\mathcal{E}_{d_j}(t^2)\right)\mathcal{E}_{g-1-2b}(t). \qquad (22)\end{aligned}$$

6.2 Asymptotic analysis

$\mathcal{E}_g(t)$ is the sum $\frac{1}{2}[\mathcal{E}_{g_{\mathrm{i}}}(t)+\mathcal{E}_{g_{\mathrm{ii}}}(t)+\mathcal{E}_{g_{\mathrm{iii}}}(t)]$ (eq. (17)). We denote $\mathcal{E}'_{g_{\mathrm{i}}}(t) = \left(\sum_{j=1}^{g-1}\mathcal{E}_j(t)\,\mathcal{E}_{g-j}(t)\right) + \mathcal{E}_{\frac{g}{2}}(t^2)$ and have $\mathcal{E}_g(t) = \frac{1}{2[1-\mathcal{U}(t)]}[\mathcal{E}'_{g_{\mathrm{i}}}(t)+\mathcal{E}_{g_{\mathrm{ii}}}(t)+\mathcal{E}_{g_{\mathrm{iii}}}(t)]$. From eqs. (20)-(22), $\mathcal{E}_g(t)$ is a rational function in $\mathcal{U}(t)$ and $\mathcal{E}_\ell(t)$ for $1\le\ell\le g-1$, as well as in $\mathcal{U}(t^2)$ and $\mathcal{E}_\ell(t^2)$ for $1\le\ell\le g-1$.

▶ **Proposition 7.** *The generating function $\mathcal{E}_g(t)$ for the number of unlabeled galled trees with g galls satisfies as $t\to\rho^-$*

$$\mathcal{E}_g(t) \sim \frac{\delta_g}{\gamma^{4g-1}(1-t/\rho)^{2g-1/2}}, \qquad (23)$$

where δ_g is a constant dependent on g satisfying $\delta_1 = \frac{1}{2}$, and for $g \geq 2$,

$$\delta_g = \frac{1}{2}\sum_{\ell=1}^{g-1}\left[\delta_\ell\delta_{g-\ell} + (\ell+1)\sum_{\mathbf{d}\in C(g-1,\ell)}\prod_{j=1}^{\ell}\delta_{d_j}\right]. \tag{24}$$

Proof. We proceed by induction. The claim holds for $g = 1$ (Proposition 1) and $g = 2$ (eq. (15)), with $\delta_2 = \frac{1}{2}[\frac{1}{2}\frac{1}{2}+2\frac{1}{2}] = \frac{5}{8}$. We assume inductively that for $\ell = 1, 2, \ldots, g-1$, $\mathcal{E}_\ell(t) \sim \delta_\ell/[\gamma^{4\ell-1}(1-t/\rho)^{2\ell-1/2}]$, with constants δ_ℓ as in eq. (24). By the inductive hypothesis, the convergence radius of $\mathcal{E}_\ell(t)$ for each ℓ, $1 \leq \ell \leq g-1$, is ρ. Because $t^2 < t$ for $t < \rho$, $\mathcal{U}(t^2)$ and $\mathcal{E}_\ell(t^2)$ can be treated as constants when finding the asymptotic behavior of $\mathcal{E}_g(t)$. As a result, using the inductive hypothesis, all terms in $\mathcal{E}_g(t)$ take the form $c/[\gamma^m(1-t/\rho)^{m/2}]$, and we must find the terms with the maximal power of $1/\sqrt{1-t/\rho}$.

We examine $\mathcal{E}'_{g_{\text{i}}}(t)$, $\mathcal{E}_{g_{\text{iii}}}(t)$, and then $\mathcal{E}_{g_{\text{ii}}}(t)$. By the inductive hypothesis,

$$\begin{aligned}\mathcal{E}'_{g_{\text{i}}}(t) &\sim \sum_{j=1}^{g-1}\left[\frac{\delta_j}{\gamma^{4j-1}(1-t/\rho)^{2j-1/2}}\cdot\frac{\delta_{g-j}}{\gamma^{4(g-j)-1}(1-t/\rho)^{2(g-j)-1/2}}\right]\\ &\sim \sum_{j=1}^{g-1}\frac{\delta_j\delta_{g-j}}{\gamma^{4g-2}(1-t/\rho)^{2g-1}}\end{aligned} \tag{25}$$

$$\mathcal{E}_{g_{\text{iii}}}(t) \sim \sum_{\ell=0}^{\lfloor\frac{g-1}{2}\rfloor}\sum_{b=\ell}^{\lfloor\frac{g-1}{2}\rfloor}\left(\frac{1}{[1-\mathcal{U}(\rho^2)]^{\ell+1}}\sum_{\mathbf{d}\in C(b,\ell)}\prod_{j=1}^{\ell}\mathcal{E}_{d_j}(\rho^2)\right)\frac{\delta_{g-1-2b}}{\gamma^{4g-8b-5}(1-t/\rho)^{2g-4b-5/2}}. \tag{26}$$

Because the largest power of $1/(1-t/\rho)$ in $\mathcal{E}_{g_{\text{iii}}}(t)$ is less than $2g-1$, its largest power in $\mathcal{E}'_{g_{\text{i}}}(t)$, $\mathcal{E}_{g_{\text{iii}}}(t)$ does not affect the asymptotics of $\mathcal{E}_g(t)$.

For $\mathcal{E}_{g_{\text{ii}}}(t)$, for any $\ell = 1, 2, \ldots, g-1$, two quantities determine the power of $1/\sqrt{1-t/\rho}$: both $\sum_{\mathbf{d}\in C(g-1,\ell)}\prod_{j=1}^{\ell}\mathcal{E}_{d_j}(t)$ and $[3\mathcal{U}(t)-2+\ell]/[1-\mathcal{U}(t)]^{\ell+2} + [\![\ell=1]\!]$. First, according to the inductive hypothesis, for each ℓ, $1 \leq \ell \leq g-1$, noting $\sum_{j=1}^{\ell} d_j = g-1$,

$$\begin{aligned}\sum_{\mathbf{d}\in C(g-1,\ell)}\prod_{j=1}^{\ell}\mathcal{E}_{d_j}(t) &\sim \sum_{\mathbf{d}\in C(g-1,\ell)}\prod_{j=1}^{\ell}\frac{\delta_{d_j}}{\gamma^{4d_j-1}(1-t/\rho)^{2d_j-1/2}}\\ &\sim \sum_{\mathbf{d}\in C(g-1,\ell)}\frac{\prod_{j=1}^{\ell}\delta_{d_j}}{\gamma^{4g-4-\ell}(1-t/\rho)^{2g-2-\ell/2}}.\end{aligned} \tag{27}$$

Second, for ℓ, $1 \leq \ell \leq g-1$, from $\mathcal{U}(t) \sim 1-\gamma\sqrt{1-t/\rho}$,

$$\left(\frac{3\mathcal{U}(t)-2+\ell}{[1-\mathcal{U}(t)]^{\ell+2}} + [\![\ell=1]\!]\right) \sim \frac{\ell+1}{\gamma^{\ell+2}(1-t/\rho)^{(\ell+2)/2}}. \tag{28}$$

Combining eqs. (27) and (28), we obtain

$$\begin{aligned}\mathcal{E}_{g_{\text{ii}}}(t) &\sim \sum_{\ell=1}^{g-1}\sum_{\mathbf{d}\in C(g-1,\ell)}\frac{\prod_{j=1}^{\ell}\delta_{d_j}}{\gamma^{4g-4-\ell}(1-t/\rho)^{2g-2-\ell/2}}\cdot\frac{\ell+1}{\gamma^{\ell+2}(1-t/\rho)^{(\ell+2)/2}}\\ &\sim \sum_{\ell=1}^{g-1}\frac{(\ell+1)\sum_{\mathbf{d}\in C(g-1,\ell)}\prod_{j=1}^{\ell}\delta_{d_j}}{\gamma^{4g-2}(1-t/\rho)^{2g-1}}.\end{aligned} \tag{29}$$

The proof is concluded by noting

$$\mathcal{E}_g(t) \sim \left[\sum_{j=1}^{g-1} \frac{\delta_j \delta_{g-j}}{\gamma^{4g-2}(1-t/\rho)^{2g-1}} + \sum_{\ell=1}^{g-1} \frac{(\ell+1)\sum_{\mathbf{d}\in C(g-1,\ell)} \prod_{j=1}^{\ell} \delta_{d_j}}{\gamma^{4g-2}(1-t/\rho)^{2g-1}}\right] \frac{1}{2\gamma(1-t/\rho)^{1/2}}$$
$$\sim \frac{\sum_{\ell=1}^{g-1} \left[\delta_\ell \delta_{g-\ell} + (\ell+1)\sum_{\mathbf{d}\in C(g-1,\ell)} \prod_{j=1}^{\ell} \delta_{d_j}\right]}{2\gamma^{4g-1}(1-t/\rho)^{2g-1/2}}$$
$$\sim \frac{\delta_g}{\gamma^{4g-1}(1-t/\rho)^{2g-1/2}}. \tag{30}$$

◀

▶ **Theorem 8.** *The asymptotic growth of the number $E_{n,g}$ of unlabeled galled trees with n leaves and a fixed number of galls $g \geq 1$ satisfies*

$$E_{n,g} \sim \frac{\delta_g}{\gamma^{4g-1}\Gamma(2g-\frac{1}{2})} n^{2g-\frac{3}{2}} \rho^{-n} \sim \frac{2^{2g-1}\delta_g}{\gamma^{4g-1}(4g-3)!!\sqrt{\pi}} n^{2g-\frac{3}{2}} \rho^{-n}. \tag{31}$$

Proof. The first step follows from the transfer formula. For the second step of eq. (31), we recall $\Gamma(n+\frac{1}{2}) = [(2n-1)!!/2^n]\sqrt{\pi}$ with and $2g-\frac{1}{2} = (2g-1)+\frac{1}{2}$. ◀

The δ_g have a relationship with the Catalan numbers, $C_m = \binom{2m}{m}/(m+1)$.

▶ **Proposition 9.** *The numbers $\{\delta_g\}_{g\geq 1}$ satisfy $2^{2g-1}\delta_g = C_{2g-1}$.*

Proof. We prove the result by showing that the generating function $\mathcal{D}(t) = \sum_{g\geq 1} 2^{2g-1}\delta_g t^{2g-1}$ is the odd part of the generating function of the Catalan numbers, $\mathcal{C}_O(t) = \sum_{g\geq 1} C_{2g-1} t^{2g-1}$.

$\mathcal{C}_O(t)$ satisfies $\mathcal{C}_O(t) = \frac{1}{2}\sum_{n\geq 0}[C_n t^n - C_n(-t)^n] = \sum_{n\geq 1} C_{2n-1} t^{2n-1}$, where $\mathcal{C}(t) = (1-\sqrt{1-4t})/(2t)$ is the generating function of the Catalan numbers. Hence, $\mathcal{C}_O(t) = \left[1-\frac{1}{2}(\sqrt{1-4t}+\sqrt{1+4t})\right]/(2t)$. From the recursion for δ_g (Proposition 7),

$$\mathcal{D}(t) = t + \sum_{g\geq 2}\left(\sum_{\ell=1}^{g-1} 2^{2g-2}\delta_\ell\delta_{g-\ell}\right)t^{2g-1} + \sum_{g\geq 2}\left[\sum_{\ell=1}^{g-1}(\ell+1)2^{2g-2}\sum_{\mathbf{d}\in C(g-1,\ell)}\prod_{j=1}^{\ell}\delta_{d_j}\right]t^{2g-1}$$
$$= t + \left[\sum_{\ell\geq 1} 2^{2\ell-1}\delta_\ell t^{2\ell-1} \sum_{g\geq \ell+1} 2^{2(g-\ell)-1}\delta_{g-\ell} t^{2(g-\ell)-1}\right]t$$
$$+ \left[\sum_{\ell\geq 1}(\ell+1)(2t)^\ell \sum_{g\geq \ell+1}\sum_{\mathbf{d}\in C(g-1,\ell)}\prod_{j=1}^{\ell} 2^{2d_j-1}\delta_{d_j} t^{2d_j-1}\right]t$$
$$= t + t\mathcal{D}^2(t) + t\sum_{\ell\geq 1}(\ell+1)[2t\mathcal{D}(t)]^\ell$$
$$= t + t\mathcal{D}^2(t) + \frac{2t^2\mathcal{D}(t)}{[1-2t\mathcal{D}(t)]^2} + \frac{2t^2\mathcal{D}(t)}{1-2t\mathcal{D}(t)}. \tag{32}$$

Solving for $\mathcal{D}(t)$, we obtain four solutions, only one of which has the correct limit of 0 as $t \to 0$; this root is equal to $\mathcal{C}_O(t)$.

◀

▶ **Theorem 10.** *The number of unlabeled galled trees with n leaves and any fixed number of galls $g \geq 0$ has asymptotic approximation*

$$E_{n,g} \sim \frac{2^{2g-1}}{(2g)!\,\gamma^{4g-1}\sqrt{\pi}} n^{2g-\frac{3}{2}} \rho^{-n}. \tag{33}$$

Table 1 The subexponential portion $c_g n^{2g-\frac{3}{2}}$ of the growth $c_g n^{2g-\frac{3}{2}}\rho^{-n}$ with the number of leaves n of $E_{n,g}$, the number of galled trees with exactly g galls. Quantities are computed according to eq. (2) for $g=0$ and Theorems 8 and 10 for $g \geq 1$.

Number of galls g	Exact constant c_g	Approximate value of c_g	$n^{2g-\frac{3}{2}}$
0	$\frac{\gamma}{2\sqrt{\pi}}$	0.3188	$n^{-\frac{3}{2}}$
1	$\frac{1}{\gamma^3\sqrt{\pi}}$	0.3910	$n^{\frac{1}{2}}$
2	$\frac{5}{15\gamma^7\sqrt{\pi}} = \frac{8}{24\gamma^7\sqrt{\pi}} = \frac{1}{3\gamma^7\sqrt{\pi}}$	0.0799	$n^{\frac{5}{2}}$
3	$\frac{42}{945\gamma^{11}\sqrt{\pi}} = \frac{32}{720\gamma^{11}\sqrt{\pi}} = \frac{2}{45\gamma^{11}\sqrt{\pi}}$	0.0065	$n^{\frac{9}{2}}$
4	$\frac{429}{135135\gamma^{15}\sqrt{\pi}} = \frac{128}{40320\gamma^{15}\sqrt{\pi}} = \frac{1}{315\gamma^{15}\sqrt{\pi}}$	2.8638×10^{-4}	$n^{\frac{13}{2}}$
5	$\frac{4862}{344459425\gamma^{19}\sqrt{\pi}} = \frac{512}{3628800\gamma^{19}\sqrt{\pi}} = \frac{2}{14175\gamma^{19}\sqrt{\pi}}$	7.8062×10^{-6}	$n^{\frac{17}{2}}$

Proof. The Catalan numbers satisfy $C_n = 2^n(2n-1)!!/(n+1)!$, so that

$$\frac{2^{2g-1}\delta_g}{(4g-3)!!} = \frac{C_{2g-1}}{(4g-3)!!} = \frac{2^{2g-1}[2(2g-1)-1]!!}{(4g-3)!!\,(2g-1+1)!} = \frac{2^{2g-1}}{(2g)!}.$$

The case of $g=0$ is included, as $E_{n,0} \sim [2^{-1}/(\gamma^{-1}\sqrt{\pi})]n^{-\frac{3}{2}}\rho^{-n} = [\gamma/2\sqrt{\pi}]n^{-\frac{3}{2}}\rho^{-n} \sim U_n$. ◀

Table 1 depicts the subexponential growth of $E_{n,g}$ for each g from 1 to 5. For $g=1$ and $g=2$, the theorem recovers the values obtained in Propositions 2 and 5.

▶ **Corollary 11.** *The exponential growth of the number $E_{n,g}$ of unlabeled trees with n leaves and a fixed number of galls $g \geq 1$ is the same as that of U_n, the number of unlabeled trees with no galls; however, the subexponential growth is greater by a factor of $4n^2/[\gamma^4(2g+1)(2g+2)]$.*

7 Discussion

We have studied the number of rooted binary unlabeled galled trees with a fixed number of galls, analyzing the exponential growth of this quantity as the number of leaves increases. We have found that the exponential growth, with the increase in the number of leaves n, of the number of galled trees with a fixed number of galls is independent of the number of galls g (Corollary 11). This independence includes the case of $g=0$ galls, the classic case of rooted binary unlabeled trees. It also implies that the number of galled trees whose number of galls is in some finite set G also has this same exponential growth.

The exponential growth with n of the number of galled trees with fixed g or with g in a finite set of values contrasts with the much greater increase in A_n, the number of galled trees with no restriction on the number of galls. This much larger growth for A_n is explained by the increase in the subexponential component with increasing g of the number of galled trees with n leaves and g galls, and the fact that with no maximum number of galls, as n increases, the number of terms in $A_n = \sum_{g\geq 0}^{\lfloor (n-1)/2 \rfloor} E_{n,g}$ grows without bound.

Our analysis produced a recursion for the Catalan numbers with odd indices: $C_{2n-1} = \sum_{m=1}^{n-1} C_{2m-1}C_{2(n-m)-1} + \sum_{m=1}^{n-1}(m+1)2^m \sum_{\mathbf{d}\in C(n-1,m)} C_{2d_j-1}$. The first part comes from terms of $C_n = \sum_{m=0}^{n-1} C_m C_{(n-1)-m}$ with odd m and $(n-1)-m$; the second substitutes a sum involving Catalan numbers with odd index for terms with even m and $(n-1)-m$.

The difference across values of g in the growth of the number of trees with exactly $g \geq 0$ galls lies in the subexponential component, $c_g n^{2g-\frac{3}{2}}$. Related problems involving labeled phylogenetic networks show this same pattern, in which incrementing a constant associated with network complexity does change the subexponential growth but not the exponential growth. In particular, this pattern is seen with increasingly many reticulation nodes in various network classes [6, 7, 11, 12, 13, 19]; the subexponential growth often includes a factor of n^2, as in our case. Note additionally that beginning from $g = 1$, the constant c_g in the asymptotic approximation for $E_{n,g}$ decreases with g (eq. (31), Table 1). This property also holds for the labeled normal networks of Fuchs et al. [11, 12, 13].

The study here deals with the asymptotic enumeration of galled trees when the number of *galls* is fixed. Using the bivariate function $\mathcal{A}(t, u) = \sum_{n \geq 0} \sum_{g \geq 0} E_{n,g} t^n u^g$, Section 5.6 of our previous study of galled trees showed that for a fixed number of *leaves*, the number of galls follows an asymptotic normal distribution [1, eq. 56]. The marginal analysis fixing the number of galls contributes a perspective on the bivariate distribution different from that of the previous analysis.

We comment that we could potentially have derived our generating functions by the symbolic method [10]. Our approach instead began with constructive enumeration of possible cases, continuing the analysis based on a recursion derived in our previous study of galled trees [1] in order to find the generating functions. The symbolic method, which we defer to a subsequent article, potentially leads to simpler derivations that enable quick comparisons of relationships among enumerations for different types of galled trees.

By analyzing the asymptotics of $E_{n,g}$ for arbitrary g, this work solves unsolved problems from [1], who only analyzed $E_{n,1}$ and $A_n = \sum_{g \geq 0}^{\lfloor (n-1)/2 \rfloor} E_{n,g}$. The analysis has potential to assist in other scenarios with unlabeled phylogenetic networks indexed by a fixed quantity.

References

1. L. Agranat-Tamir, S. Mathur, and N. A. Rosenberg. Enumeration of rooted binary unlabeled galled trees. *Bulletin of Mathematical Biology*, 86:45, 2024. doi:10.1007/s11538-024-01270-8.
2. F. Bienvenu, A. Lambert, and M. Steel. Combinatorial and stochastic properties of ranked tree-child networks. *Random Structures and Algorithms*, 60:653–689, 2022. doi:10.1002/rsa.21048.
3. M. Bouvel, P. Gambette, and M. Mansouri. Counting phylogenetic networks of level 1 and 2. *Journal of Mathematical Biology*, 81:1357–1395, 2020. doi:10.1007/s00285-020-01543-5.
4. G. Cardona and L. Zhang. Counting and enumerating tree-child networks and their subclasses. *Journal of Computer and System Sciences*, 114:84–104, 2020. doi:10.1016/j.jcss.2020.06.001.
5. K.-Y. Chang, W.-K. Hon, and S. V. Thankachan. Compact encoding for galled-trees and its applications. In *2018 Data Compression Conference*, pages 297–306, Snowbird, UT, 2018. doi:10.1109/DCC.2018.00038.
6. Y. S. Chang and M. Fuchs. Counting Phylogenetic Networks with Few Reticulation Vertices: Galled and Reticulation-Visible Networks. *Bull. Math. Biol.*, 86(7):76, 2024. doi:10.1007/s11538-024-01309-w.
7. Y.-S. Chang, M. Fuchs, and G.-R. Yu. Galled tree-child networks. In *35th International Conference on Probabilistic, Combinatorial and Asymptotic Methods for the Analysis of Algorithms*, volume 302, Article 2 of *Leibniz International Proceedings in Informatics, LIPIcs*. Schloss Dagstuhl—Leibniz-Zentrum für Informatik, Wadern, 2024.
8. L. Comtet. *Advanced Combinatorics*. Reidel, Boston, 1974. doi:10.1007/978-94-010-2196-8.
9. J. Felsenstein. *Inferring Phylogenies*. Sinauer Associates Inc., Sunderland MA, 2004.

10 P. Flajolet and R. Sedgewick. *Analytic Combinatorics*. Cambridge University Press, Cambridge, 2009. doi:10.1017/CBO9780511801655.

11 M. Fuchs, B. Gittenberger, and M. Mansouri. Counting phylogenetic networks with few reticulation vertices: tree-child and normal networks. *Australasian Journal of Combinatorics*, 73:385–423, 2019.

12 M. Fuchs, B. Gittenberger, and M. Mansouri. Counting phylogenetic networks with few reticulation vertices: exact enumeration and corrections. *Australasian Journal of Combinatorics*, 81:257–282, 2021.

13 M. Fuchs, E.-Y. Huang, and G.-R. Yu. Counting phylogenetic networks with few reticulation vertices: a second approach. *Discrete Applied Mathematics*, 320:140–149, 2022. doi:10.1016/j.dam.2022.03.026.

14 D. Gusfield. *ReCombinatorics*. MIT Press, Cambridge MA, 2014.

15 D. Gusfield, S. Eddhu, and C. Langley. Efficient reconstruction of phylogenetic networks with constrained recombination. In *Proceedings of the IEEE Computer Society Conference on Bioinformatics*, pages 363–374, 2003. doi:10.1109/CSB.2003.1227337.

16 D. Gusfield, S. Eddhu, and C. H. Langley. The fine structure of galls in phylogenetic networks. *INFORMS Journal on Computing*, 16:459–469, 2004. doi:10.1287/ijoc.1040.0099.

17 D. H. Huson, R. Rupp, and C. Scornavacca. *Phylogenetic Networks: Concepts, Algorithms and Applications*. Cambridge University Press, Cambridge, 2010. doi:10.1017/CBO9780511974076.

18 S. Kong, J. C. Pons, L. Kubatko, and K. Wicke. Classes of explicit phylogenetic networks and their biological and mathematical significance. *Journal of Mathematical Biology*, 84:47, 2022. doi:10.1007/s00285-022-01746-y.

19 M. Mansouri. Counting general phylogenetic networks. *Australasian Journal of Combinatorics*, 83:40–86, 2022.

20 S. Mathur and N. A. Rosenberg. All galls are divided into three or more parts: recursive enumeration of labeled histories for galled trees. *Algorithms in Molecular Biology*, 18:1, 2023. doi:10.1186/s13015-023-00224-4.

21 C. Semple and M. Steel. *Phylogentics*. Oxford University Press, Oxford, 2003.

22 C. Semple and M. Steel. Unicyclic networks: compatibility and enumeration. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 3:84–91, 2006. doi:10.1109/TCBB.2006.14.

23 Y. S. Song. A concise necessary and sufficient condition for the existence of a galled-tree. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 3:186–191, 2006. doi:10.1109/TCBB.2006.15.

24 T. Warnow. *Computational Phylogenetics*. Cambridge University Press, Cambridge, 2018. doi:10.1017/9781316882313.