Galled Tree-Child Networks

Yu-Sheng Chang[*]   Michael Fuchs[*]   Guan-Ru Yu[†]

March 5, 2024

Dedicated to Hsien-Kuei Hwang on the occasion of his 60th birthday

Abstract

We propose the class of *galled tree-child networks* which is obtained as intersection of the classes of galled networks and tree-child networks. For the latter two classes, (asymptotic) counting results and stochastic results have been proved with very different methods. We show that a counting result for the class of galled tree-child networks follows with similar tools as used for galled networks, however, the result has a similar pattern as the one for tree-child networks. In addition, we also consider the (suitably scaled) numbers of reticulation nodes of random galled tree-child networks and show that they are asymptotically normal distributed. This is in contrast to the limit laws of the corresponding quantities for galled networks and tree-child networks which have been both shown to be discrete.

1   Introduction

Phylogenetic networks are used to visualize, model, and analyze the ancestor relationship of taxa in reticulate evolution. To make them more relevant for biological applications as well as devise algorithms for them, many subclasses of the class of phylogenetic networks have been proposed; see the comprehensive survey [13]. A lot of recent research work was concerned with fundamental questions such as counting them and understanding the shape of a network drawn uniformly at random from a given class; see, e.g., [2–4,7–12,14,15]. Despite of this, even counting results are still missing for most of the major classes of phylogenetic networks. Two notable exceptions are tree-child networks and galled networks for which such results have been proved in [10,11]. In this work, we consider the intersection of these two network classes. We start with some basic definitions and then explain why we find this class interesting.

First, a phylogenetic network is defined as follows.

Definition 1.1 (Phylogenetic Network)**.** A (rooted) phylogenetic network of size n is a rooted, simple, directed, acyclic graph whose nodes fall into the following three (disjoint) categories:

(a) A unique *root* which has indegree 0 and outdegree 1;

(b) *Leaves* which have indegree 1 and outdegree 0 and are bijectively labeled with labels from the set $\{1, \dots, n\}$;

---

[*]Department of Mathematical Sciences, National Chengchi University, Taipei 116, Taiwan.
[†]Department of Applied Mathematics, National Sun Yat-sen University, Kaohsiung 804, Taiwan.

(c) *Internal nodes* which have indegree and outdegree at least 1 and total degree at least 3.

Moreover, a phylogenetic network is called *binary* if (c) is replaced by

(c') *Internal nodes* which have either indegree 1 and outdegree 2 (*tree nodes*) or indegree 2 and outdegree 1 (*reticulation nodes*).

Remark 1.2. (i) Phylogenetic networks with all internal nodes having indegree equal to 1 are called *phylogentic trees*.

(ii) If not explicitly mentioned, phylogenetic networks will always be binary in the sequel.

We next define galled networks and tree-child networks which are two of the major classes of phylogenetic networks. For the definition, we need the notion of a *tree cycle* which is a pair of edge-disjoint paths in a phylogenetic network that start at a common tree node and end at a common reticulation node with all other nodes being tree nodes.

Definition 1.3. (a) A phylogenetic network is called a *tree-child network* if every non-leaf node has at least one child which is either a reticulation node or a leaf.

(b) A phylogenetic network is called a *galled network* if every reticulation node is in a (necessarily unique) tree cycle.

Remark 1.4. Note that neither the class of tree-child networks is contained in the class of galled networks nor vice versa.

Let $\mathrm{TC}_{n,k}$ and $\mathrm{GN}_{n,k}$ denote the number of tree-child networks and galled networks of size n with k reticulation nodes, respectively. It is not hard to see that $k \leq n-1$ for tree-child networks and $k \leq 2n-2$ for galled networks where both bounds are sharp. Thus, the total numbers are given by:

$$\mathrm{TC}_n := \sum_{k=0}^{n-1} \mathrm{TC}_{n,k} \qquad \text{and} \qquad \mathrm{GN}_n := \sum_{k=0}^{2n-2} \mathrm{GN}_{n,k}. \tag{1}$$

The asymptotic growth of both of these sequences is known. First, in [10], it was proved that for the number of tree-child networks, as $n \to \infty$,

$$\mathrm{TC}_n = \Theta\left(n^{-2/3} e^{a_1(3n)^{1/3}} \left(\frac{12}{e^2}\right)^n n^{2n} \right), \tag{2}$$

where a_1 is the largest root of the Airy function of the first kind. The surprise here was the presence of a *stretched exponential* in the asymptotic growth term. On the other hand, no stretched exponential is contained in the asymptotics of the number of galled networks. More precisely, it was proved in [11] that

$$\mathrm{GN}_n \sim \frac{\sqrt{2e\sqrt[4]{e}}}{4} n^{-1} \left(\frac{8}{e^2}\right)^n n^{2n}. \tag{3}$$

The tools used to establish (2) and (3) were very different: for (2), a bijection to a class of words was proved and a recurrence for these word was found which could be (asymptotically) analyzed with the approach from [6]; for (3), the component graph method introduced in [12] together with the Laplace method and a result from [1] was used.

Another difference was the location in (1) of the terms which dominate the two sums. For tree-child networks, the main contribution comes from networks with k close to $n - 1$ (the maximal-reticulated networks), whereas for galled networks, the main contributions comes from networks with $k \approx n$. In fact, the limit law of the number of reticulation nodes, say R_n, was derived in [5,11] for both network classes if a network of size n is sampled uniformly at random. More precisely, for tree-child networks, it was shown in [5] that, as $n \to \infty$,

$$n - 1 - R_n \xrightarrow{d} \text{Poisson}(1/2),$$

where $\xrightarrow{d}$ denotes convergence in distribution and $\text{Poisson}(\lambda)$ is a Poisson law with parameter λ. A similar discrete limit law was proved in [11] for galled networks, however, the limit law is not Poisson but a mixture of Poisson laws; see Theorem 2 in [11] for details.

Due to the above results and differences, one wonders how the intersection of the class of tree-child networks and galled networks behaves?

Definition 1.5 (Galled Tree-Child Network). A *galled tree-child network* is a network which is both a galled network and a tree-child network.

Let $\text{GTC}_{n,k}$ denote the number of galled tree-child networks of size n with k reticulation nodes. We will show below that $k \leq n - 1$. (See Lemma 3.1 in Section 3.) Set:

$$\text{GTC}_n := \sum_{k=0}^{n-1} \text{GTC}_{n,k}.$$

Then, this sequence has the following first-order asymptotics.

Theorem 1.6. *For the number of galled tree-child networks, we have, as $n \to \infty$,*

$$\text{GTC}_n \sim \frac{1}{2\sqrt[4]{e}} n^{-5/4} e^{2\sqrt{n}} \left(\frac{2}{e^2}\right)^n n^{2n}.$$

Remark 1.7. Note that the asymptotics contains a stretched exponential as does the expansion (2) for tree-child networks, however, the proof will use the tools which were used in [11] to derive (3) for galled networks.

We next consider the number of reticulation nodes R_n of a *random galled tree-child network* which is a galled tree-child network of size n that is sampled uniformly at random from the set of all galled tree-child networks of size n. In contrast to tree-child networks and galled networks, the limit law of R_n (suitably scaled) is continuous.

Theorem 1.8. *The number of reticulation nodes R_n of a random galled tree-child networks satisfies, as $n \to \infty$,*

$$\frac{R_n - \mathbb{E}(R_n)}{\sqrt{\text{Var}(R_n)}} \xrightarrow{d} N(0,1),$$

where $N(0,1)$ denotes the standard normal distribution. Moreover,

$$\mathbb{E}(R_n) \sim n - \sqrt{n} \qquad \text{and} \qquad \text{Var}(R_n) \sim \sqrt{n}/2.$$

The above results show that galled tree-child networks behave quite different from both tree-child networks and galled networks. That is one reason why we find them interesting.
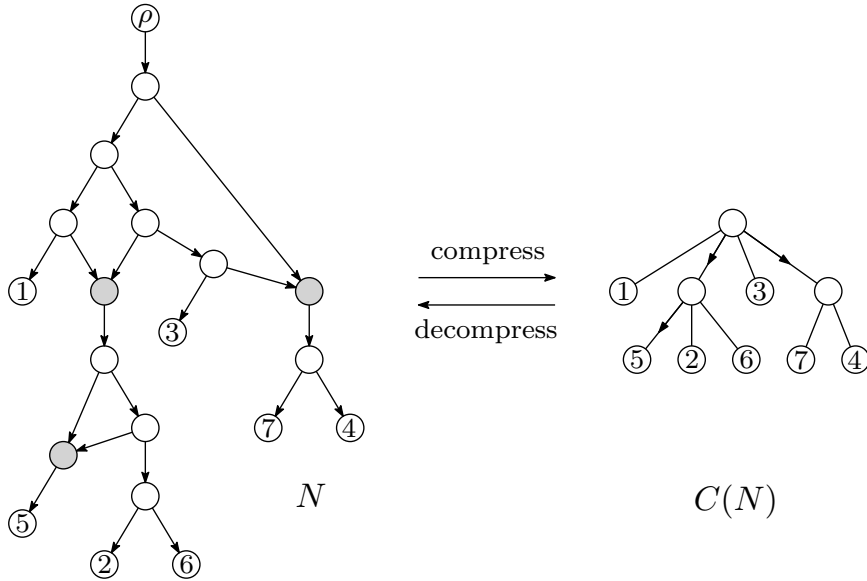
Figure 1: A galled network N and its component graph $C(N)$ which is a phylogenetic tree.

Another reason stems from a recent result which was proved in [4]. In the latter paper, the asymptotics of $\mathrm{GN}_{n,k}$ for fixed k was derived. Let $\mathrm{PN}_{n,k}$ denote the number of phylogenetic networks of size n and k reticulation nodes. (Note that this number is finite, whereas it becomes infinite when summing over k.) Then, one of the main results from [4] implies that for fixed k, as $n \to \infty$,

$$\mathrm{PN}_{n,k} \sim \mathrm{TC}_{n,k} \sim \mathrm{GN}_{n,k} \sim \frac{2^{k-1}\sqrt{2}}{k!}\left(\frac{2}{e}\right)^n n^{n+2k-1}. \tag{4}$$

(The first two asymptotic equivalences were proved in [9,14].) That $\mathrm{TC}_{n,k}$ and $\mathrm{GN}_{n,k}$ have the same first-order asymptotics for fixed k was a surprise since the classes of tree-child networks and galled networks are quite different, e.g., neither contains the other; see Remark 1.4. However, the above result can be explained via the class of galled tree-child networks as will be seen in Section 3 below.

We conclude the introduction with a short sketch of the paper. The proofs of Theorem 1.6 and Theorem 1.8 follow with a similar approach as used for galled networks in [10]. This approach is based on the component graph method from [12] which we will recall in the next section. Then, in Section 3, we will consider $\mathrm{GTC}_{n,k}$ for small and large values of k. Finally, Section 4 will contain the proofs of our main results (Theorem 1.6 and Theorem 1.8). We will conclude the paper with some final remarks in Section 5.

2 The Component Graph Method

The component graph method for galled networks was introduced in [12] and used in [4,11] to prove asymptotic results. It is explained in detail in all these papers. However, to make the current paper self-contained, we will briefly recall it.

Let N be a galled network. Then, by removing all the edges leading to reticulation vertices (these are the so-called *reticulation edges*), we obtain a forest whose trees are called the *tree-*

components of N.

The *component graph* of N, denoted by $C(N)$, is now a rooted, directed, acyclic graph which has a vertex for every tree-component. Moreover, the vertices are connected in the same way as the tree-components have been connected by the reticulation edges. Finally, we attach the leaves in the tree-components to the corresponding vertices in $C(N)$ unless a vertex v of $C(N)$ is a terminal vertex and its corresponding tree-component has exactly one leaf, in which case we use the label of that leaf to label v. Note that every non-root vertex in $C(N)$ has indegree 2 and that $C(N)$ may contain double edges. We will replace such a double edge by a single edge and indicate that it was a double edge by placing an arrow on it; see Figure 1 for a galled network together with its component graph. Also, denote by $\tilde{C}(N)$ the component graph of $C(N)$ with all arrows on edges removed. Then, the authors of [12] made the following important observation.

Proposition 2.1 ([12])**.** *N is a galled network if and only if $\tilde{C}(N)$ is a (not necessarily binary) phylogenetic tree.*

The component graph can be seen as a compression of N. Thus, in order to generate all galled networks of size n, one only needs to list all component graphs (i.e., phylogenetic trees) with n labeled leaves and decompress them.

We will next explain the decompression procedure. For this, we need the notion of *one-component networks*.

Definition 2.2 (One-component Network)**.** A phylogenetic network is called a *one-component network* if every reticulation node has a leaf as its child.

Now, let $C(N)$ be given. We do a breadth-first traversal of $C(N)$ and replace every vertex v by a one-component galled network O_v whose leaves below reticulation vertices are labeled with the first k labels, where k is the number of outgoing edges of v in $C(N)$ that have an arrow on them, and whose size is equal to the outdegree of v. Then, attach the subtrees of v which are connected to v by edges with arrows on them to the leaves of O_v with labels $\{1, \ldots, k\}$, where the subtree with the smallest label is attached to 1, the subtree with the second largest label is attached to 2, etc. Moreover, relabel the remaining leaves of O_v by the remaining subtrees of v (which are all of size 1, i.e., they are leaves) in an order-consistent way. By using all possible one-component galled networks in every step, this gives all possible galled networks with $C(N)$ as component graph. Moreover, if we start from $\tilde{C}(N)$, then we first have to place arrows on all edges whose heads are internal nodes of $C(N)$ and for all remaining edges, we can freely decide if we want to place an arrow on them or not. Overall, this gives the following result which was one of the main results of [12].

Proposition 2.3 ([12])**.** *We have,*

$$\mathrm{GN}_n = \sum_{\mathcal{T}} \prod_v \sum_{j=0}^{c_{\mathrm{lf}}(v)} \binom{c_{\mathrm{lf}}(v)}{j} M_{c(v), c(v) - c_{\mathrm{lf}}(v) + j},$$

where the first sum runs over all (not necessarily binary) phylogenetic trees $\mathcal{T}$ of size n, the product runs over all internal nodes of $\mathcal{T}$, $c(v)$ is the outdegree of v, $c_{\mathrm{lf}}(v)$ is the number of children of v which are leaves, and $M_{n,k}$ denotes the number of one-component galled networks of size n with k reticulation vertices, where the leaves below the reticulation vertices are labeled with labels from the set $\{1, \ldots, k\}$.

For galled tree-child networks, it is now clear that the same formula holds with the only difference that $M_{n,k}$ has to be replaced by the corresponding number of one-component galled tree-child networks. However, this number is the same as the number of one-component tree-child networks.

Lemma 2.4. *Every one-component tree-child network is a one-component galled tree-child network.*

Proof. Let v be a reticulation vertex and consider a pair of edge-disjoint paths from a common tree vertex to v. (Note that such a pair trivially exists.) Then, no internal vertex can be a reticulation vertex because such a reticulation vertex would not be followed by a leaf. Thus, v is in a tree cycle which shows that the network is indeed galled. ∎

Denote by $\mathrm{L}_{n,k}$ the number of one-component tree-child networks of size n and k reticulation vertices, where the labels of the leaves below the reticulation vertices are $\{1,\ldots,k\}$. Then, we have the following analogous result to Proposition 2.3.

Proposition 2.5. *We have,*

$$\mathrm{GTC}_n = \sum_{\mathcal{T}} \prod_v \sum_{j=0}^{c_{\mathrm{lf}}(v)} \binom{c_{\mathrm{lf}}(v)}{j} L_{c(v),c(v)-c_{\mathrm{lf}}(v)+j}, \tag{5}$$

where notation is as in Proposition 2.3 and $L_{n,k}$ was defined above.

Remark 2.6. Using this result, one can obtain the following table for small values of n:

n	GTC_n
1	1
2	3
3	48
4	1,611
5	87,660
6	6,891,615
7	734,112,540
8	101,717,195,895
9	17,813,516,259,420
10	3,857,230,509,496,875

Table 1: The values of GTC_n for $1 \le n \le 10$.

We will deduce all our results from (5). In addition, we will make use of the following results for $L_{n,k}$ which were proved in [3] and [10]. To state them, denote by $\mathrm{OTC}_{n,k}$ the number of one-component tree-child networks of size n with k reticulation vertices and by OTC_n the (total) number of one-component tree-child networks of size n. Then,

$$\mathrm{OTC}_{n,k} = \binom{n}{k} L_{n,k} \tag{6}$$

and

$$\text{OTC}_n = \sum_{k=0}^{n-1} \text{OTC}_{n,k}.$$

(Note that the tree-child property implies the $k \leq n-1$ and this bound is sharp.)

Proposition 2.7 ([3,10]). *(i) We have,*

$$\text{OTC}_{n,k} = \binom{n}{k} \frac{(2n-2)!}{2^{n-1}(n-k-1)!}.$$

(ii) As $n \to \infty$,

$$\text{OTC}_{n,k} = \frac{1}{2\sqrt{e\pi}} n^{-3/2} e^{2\sqrt{n}} \left(\frac{2}{e^2}\right)^n n^{2n} e^{-x^2/\sqrt{n}} \left(1 + \mathcal{O}\left(\frac{1+|x|^3}{n} + \frac{|x|}{\sqrt{n}}\right)\right),$$

where $k = n - \sqrt{n} + x$ and $x = o(n^{1/3})$.

The second result above is a local limit theorem for the (random) number of reticulation vertices of a one-component tree-child network of size n which is picked uniformly at random from all one-component tree-child networks of size n. It implies the following (asymptotic) counting result for OTC_n.

Corollary 2.8 ([10]). *As $n \to \infty$,*

$$\text{OTC}_n \sim \frac{1}{2\sqrt{e}} n^{-5/4} e^{2\sqrt{n}} \left(\frac{2}{e^2}\right)^n n^{2n}.$$

3 Networks with Few and Many Reticulation Nodes

In this section, we consider $\text{GTC}_{n,k}$ for small and large k. We start with large k.

As mentioned in the last section (see the sentence before Proposition 2.7), for tree-child networks, we have that $k \leq n-1$ and this bound is sharp. Clearly, this implies that $k \leq n-1$ also for galled tree-child networks. Again this bound is sharp.

Lemma 3.1. *The number of reticulation vertices of a galled tree-child network of size n is at most $n-1$ where this bound is sharp.*

Proof. Let N be a galled tree-child network of size n. Consider the component graph $\tilde{C}(N)$ of N which is a phylogenetic tree of size n (see Lemma 2.1). The maximal number of reticulation vertices of N is achieved by placing the maximal number of arrows at all outgoing edges of internal vertices v of $\tilde{C}(N)$. Note that this the degree of v, say $c(v)$, minus 1, because placing arrows on all outgoing edges is not possible since $L_{c(v),c(v)} = 0$. Thus, the maximal number of reticulation vertices equals

$$\sum_v (c(v) - 1) = \sum_v c(v) - (\# \text{ internal nodes of } \tilde{C}(N)), \tag{7}$$

where the sums run over all internal vertices of $C(N)$. By the handshake lemma,

$$\sum_v c(v) = (\# \text{ internal nodes of } \tilde{C}(N)) + n$$

which, by plugging into (7), gives the claimed result. ∎

7

The proof of the last lemma also reveals the structure of maximal reticulated galled tree-child networks of size n: They are obtained by decompressing component graphs $\tilde{C}(N)$ that are phylogenetic trees of size n with at least one leaf ℓ attached to every internal vertex v by placing arrows on all outgoing edges of v except the one leading to ℓ. This can be translated into generating functions. Set:

$$\mathrm{M}(z) := \sum_{n \geq 1} \mathrm{GTC}_{n,n-1} \frac{z^n}{n!}, \qquad \mathrm{L}(z) := \sum_{n \geq 1} L_{n,n-1} \frac{z^n}{n!} = \sum_{n \geq 1} \frac{(2n-2)!}{2^{n-1}n!} z^n,$$

where the last line follows from (6) and Proposition 2.7-(i). Then, we have the following result.

Lemma 3.2. *We have,*
$$M(z) = z + zL'(M(z)). \tag{8}$$

Proof. According to the explanation in the paragraph preceding the lemma, a maximal reticulated galled tree-child network is either a leaf or obtained from a maximal reticulated one-component tree-child network with the leafs below the reticulation vertices replaced by maximal reticulation galled tree-child networks. This translates into

$$M(z) = z + \sum_{n \geq 1} L_{n,n-1} \frac{z M(z)^{n-1}}{(n-1)!},$$

where the z inside the sum counts the leaf which is not below the reticulation vertex and the factor $1/(n-1)!$ discards the order of the maximal reticulated galled tree-child networks (counted by $M(z)^{n-1}$) which are attached to the children below the reticulation vertices. The claimed result follows from this. ∎

Note that (8) is of *Lagrangian type*. Thus, we can obtain the asymptotics of $\mathrm{GTC}_{n,n-1}$ by applying Lagrange's inversion formula and the following result from [1].

Theorem 3.3 ([1]). *Let $S(z)$ be a formal power series with $s_0 = 0, s_1 \neq 0$ and $ns_{n-1} \sim \gamma s_n$. Then, for $\alpha \neq 0$ and β real numbers,*

$$[z^n](1 + S(z))^{\alpha n + \beta} \sim \alpha e^{\alpha s_1 \gamma} n s_n.$$

Theorem 3.4. *The number of maximal reticulated galled tree-child networks $\mathrm{GTC}_{n,n-1}$ satisfies, as $n \to \infty$,*

$$\mathrm{GTC}_{n,n-1} \sim \sqrt{e\pi} n^{-1/2} \left(\frac{2}{e^2}\right)^n n^{2n}.$$

Remark 3.5. For tree-child networks, it was proved in [10] that $\mathrm{TC}_n = \Theta(\mathrm{TC}_{n,n-1})$. (This was a main step in the proof of (2).) The above result together with Theorem 1.6 shows that the same is not true for galled tree-child networks.

Proof. Applying the Lagrange inversion formula to (8) gives

$$\mathrm{GTC}_{n,n-1} = n![z^n]M(z) = (n-1)![\omega^{n-1}](1 + L'(\omega))^n. \tag{9}$$

Next, by Stirling's formula, as $n \to \infty$,

$$[z^n]L'(z) = \frac{L_{n+1,n}}{n!} = \frac{(2n)!}{2^n n!} \sim \sqrt{2}\left(\frac{2}{e}\right)^n n^n.$$

Thus, we can apply Theorem 3.3 to (9) with $\gamma = 1/2$ and obtain that, as $n \to \infty$,

$$\mathrm{GTC}_{n,n-1} \sim \sqrt{e}nL_{n,n-1} = \sqrt{e}n\frac{(2n-2)!}{2^{n-1}} \sim \sqrt{e\pi}n^{-1/2}\left(\frac{2}{e^2}\right)^n n^{2n}.$$

This is the claimed result. $\blacksquare$

We next consider $\mathrm{GTC}_{n,k}$ with k small, i.e., the other extreme case of the number of reticulation vertices. Here, we have the following result which explains why the asymptotic expansions of $\mathrm{TC}_{n,k}$ and $\mathrm{GN}_{n,k}$ in (4) are the same.

Theorem 3.6. *For fixed k, as $n \to \infty$,*

$$\mathrm{GTC}_{n,k} \sim \frac{2^{k-1}\sqrt{2}}{k!}\left(\frac{2}{e}\right)^n n^{n+2k-1}. \tag{10}$$

The proof of this result uses ideas from [9].

Proof. First consider galled tree-child networks of size n which are obtained by decompressing phylogenetic trees of size n which have all k arrows on the edges from the root, i.e., the root has at least one leaf and all other children are either internal nodes or leaves (with at most k internal nodes) and all internal nodes have just leaves as children. By Proposition 8 in [9], the number of these galled tree-child network has the same asymptotics as the one on the right-hand side of (10). Moreover, these networks also dominate the asymptotics in the case of tree-child networks. Thus, the remaining galled tree-child networks are asymptotically negligible as their number is bounded above by the number of remaining tree-child networks. $\blacksquare$

4 Proof of the Main Results

In this section, we first prove Theorem 1.6 and then state a result which implies Theorem 1.8.

For the proof of Theorem 1.6, we closely follow the method of proof of (3) from [11]. The main idea is to use (5) to find asymptotic matching upper and lower bounds for GTC_n.

First, for an upper bound, we pick a phylogenetic tree $\mathcal{T}$ of size n (which is considered to be a component graph of a galled tree-child network of size n) and decompress it by picking for internal vertices v of $\mathcal{T}$ *any* one-component tree-child network of size $c(v)$ (where the notation is as in Proposition 2.3). Since, as explained in Section 2, actually only certain one-component tree-child networks are permissible, this modified decompression procedure overcounts the number of galled tree-child networks of size n. More precisely, we consider

$$U_n := \sum_{\mathcal{T}} \prod_v \mathrm{OTC}_{c(v)},$$

where the first sum runs over all phylogenetic trees $\mathcal{T}$ of size n and the product runs over internal vertices of $\mathcal{T}$. Then, we have $\mathrm{GTC}_n \leq U_n$. Next, set

$$U(z) := \sum_{n \geq 1} U_n \frac{z^n}{n!}, \qquad A(z) := \sum_{n \geq 1} \mathrm{OTC}_{n+1}\frac{z^n}{(n+1)!}.$$

Then, the definition of U_n implies the following result.

Lemma 4.1. *We have,*

$$U(z) = z + U(z)A(U(z)).$$

Proof. The networks counted by U_n are either a leaf or a one-component tree-child network with n leaves which are replaced by an unordered sequence of networks of the same type. This gives

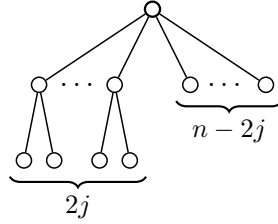$$U(z) = z + \sum_{n \geq 2} \mathrm{OTC}_n \frac{U(z)^n}{n!}$$

from which the claimed result follows. ∎

Now, we can proceed as in the proof of Theorem 3.4 to obtain the following asymptotic result for U_n.

Proposition 4.2. *As $n \to \infty$,*

$$U_n \sim \frac{1}{2\sqrt[4]{e}} n^{-5/4} e^{2\sqrt{n}} \left(\frac{2}{e^2}\right)^n n^{2n}.$$

Next, we need a matching lower bound. Therefore, we consider (5) with the first sum restricted to phylogenetic trees of the shape (where we have removed the leaf labels):

We denote the resulting term by L_n. The decompression procedure from Section 2, then gives the following result.

Lemma 4.3. *We have,*

$$L_n = \sum_{j=0}^{\lfloor n/2 \rfloor} \binom{n}{2j} \frac{(2j)!}{j!2^j} \sum_{\ell=0}^{n-2j} \binom{n-2j}{\ell} L_{n-j,j+\ell}$$

$$= \sum_{j=0}^{\lfloor n/2 \rfloor} \binom{n}{2j} \frac{(2j)!}{j!2^j} \sum_{\ell=0}^{n-2j} \binom{n-2j}{\ell} \frac{(2n-2j-2)!}{2^{n-j-1}(n-2j-\ell-1)!}. \tag{11}$$

Proof. The first equality is explained as in the proof of Lemma 9 in [11] and the second equality follows from (6) and Proposition 2.7-(i). ∎

From this result, we can deduce (matching) first-order asymptotics for L_n which then together with the asymptotics of the upper bound (Proposition 4.2) concludes the proof of Theorem 1.6.

Proposition 4.4. *As $n \to \infty$,*

$$L_n \sim \frac{1}{2\sqrt[4]{e}} n^{-5/4} e^{2\sqrt{n}} \left(\frac{2}{e^2}\right)^n n^{2n}.$$

10

Sketch of the proof. From Stirling's formula (similar to the proof of Proposition 2.7-(ii)),

$$\binom{n-2j}{\ell}\frac{(2n-2j-2)!}{2^{n-j-1}(n-2j-\ell-1)!} \sim \frac{1}{2^{j+1}\sqrt{e\pi}}n^{-3/2}e^{2\sqrt{n}}\left(\frac{2}{e^2}\right)^j n^{2n-2j}e^{-x^2/\sqrt{n}},$$

where $k = n - \sqrt{n} + x$ and this holds uniformly for small x and j (which both may depend on n). Using the Laplace method then gives,

$$\sum_{j=0}^{n-2j}\binom{n-2j}{\ell}\frac{(2n-2j-2)!}{2^{n-j-1}(n-2j-\ell-1)!} \sim \frac{1}{2^{j+1}\sqrt{e}}n^{-5/4}e^{2\sqrt{n}}\left(\frac{2}{e^2}\right)^n n^{2n-2j}$$

uniformly for small j (which again may depend on n). Finally, by plugging the last relation into (11),

$$L_n \sim \frac{1}{2\sqrt{e}}\left(\sum_{j\geq 0}\frac{1}{j!4^j}\right)n^{-5/4}e^{2\sqrt{n}}\left(\frac{2}{e^2}\right)^n n^{2n}$$

which gives the claimed result. ∎

Finally, by refining the above method (see Section 6 of [11] where the same was done for galled networks), we obtain the following result which implies our second main result (Theorem 1.8).

Theorem 4.5. *Let I_n be the number of reticulation vertices of a random galled tree-child network of size n which are not followed by a leaf and R_n be the total number of reticulation vertices. Then, as $n \to \infty$,*

$$\left(I_n, \frac{R_n - n + \sqrt{n}}{\sqrt[4]{n/4}}\right) \xrightarrow{d} (I, R),$$

where I and R are independent with $I \overset{d}{=} \mathrm{Poisson}(1/4)$ and $R \overset{d}{=} N(0,1)$.

5 Conclusion

In this paper, we introduced the class of *galled tree-child network* which is obtained as intersection of the classes of galled networks and tree-child networks. Our reason for doing so was two-fold: (i) Different tools have been used to prove results for galled networks and tree-child networks ([10,11]); consequently, we were curious about which tools apply to the combination of these classes? (ii) It was recently proved that the number of galled networks and tree-child networks have the same first-order asymptotics when the number of reticulation vertices is fixed ([4,9]). Why is that the case?

As for (i), we showed that an asymptotic counting result for galled tree-child networks (Theorem 1.6) can be obtained with the methods for galled networks, however, the result contains a stretched exponential as does the asymptotic result for tree-child networks. In addition, we showed that the number of reticulation vertices for a random galled tree-child networks is asymptotically normal (Theorem 1.8), whereas the limit laws of the same quantities for galled networks and tree-child networks were discrete. As for (ii), we showed that the number of galled tree-child networks also satisfies the same first order asymptotics when the number of reticulation vertices is fixed. This explains the previous results from [4,9].

References

[1] E. A. Bender and L. B. Richmond (1984). An asymptotic expansion for the coefficients of some power series. II. Lagrange inversion, *Discrete Math.*, **50:2-3**, 135–141.

[2] M. Bouvel, P. Gambette, M. Mansouri (2020). Counting phylogenetic networks of level 1 and level 2, *J. Math. Biol.*, **81:6-7**, 1357–1395.

[3] G. Cardona and L. Zhang (2020). Counting and enumerating tree-child networks and their subclasses, *J. Comput. System Sci.*, **114**, 84–104.

[4] Y.-S. Chang and M. Fuchs. Counting phylogenetic networks with few reticulation vertices: galled and reticulation-visible networks, arXiv:2401.08958.

[5] Y.-S. Chang, M. Fuchs, H. Liu, M. Wallner, G.-R. Yu (2022). Enumerative and distributional results for d-combining tree-child networks, arXiv:2209.03850.

[6] A. Elvey Price, W. Fang, M. Wallner. Compacted binary trees admit a stretched exponential, *J. Comb. Theory Ser. A*, **177**, Article 105306.

[7] M. Fuchs, B. Gittenberger, M. Mansouri (2019). Counting phylogenetic networks with few reticulation vertices: tree-child and normal networks, *Australas. J. Combin.*, **73:2**, 385–423.

[8] M. Fuchs, B. Gittenberger, M. Mansouri (2021). Counting phylogenetic networks with few reticulation vertices: exact enumeration and corrections, *Australas. J. Combin.*, **82:2**, 257–282.

[9] M. Fuchs, E.-Y. Huang, G.-R. Yu (2022). Counting phylogenetic networks with few reticulation vertices: a second approach, *Discrete Appl. Math.*, **320**, 140–149.

[10] M. Fuchs, G.-R. Yu, L. Zhang (2021). On the asymptotic growth of the number of tree-child networks, *European J. Combin.*, **93**, 103278, 20pp.

[11] M. Fuchs, G.-R. Yu, L. Zhang (2022). Asymptotic enumeration and distributional properties of galled networks, *J. Comb. Theory Ser. A.*, **189**, 105599, 28 pages.

[12] A. D. M. Gunawan, J. Rathin, L. Zhang (2020). Counting and enumerating galled networks, *Discrete Appl. Math.*, **283**, 644–654.

[13] S. Kong, J. C. Pons, L. Kubatko, K. Wicke (2022). Classes of explicit phylogenetic networks and their biological and mathematical significance, *J. Math. Biol.*, **84**, Paper: 47.

[14] M. Mansouri (2022). Counting general phylogenetic networks, *Australas. J. Combin.*, **83**, 40–86.

[15] M. Pons and J. Batle (2021). Combinatorial characterization of a certain class of words and a conjectured connection with general subclasses of phylogenetic tree-child networks, *Sci. Rep.*, **11**, Article number: 21875.