

A Cube-Root Phase Transition in Tree-Child Networks and the Enumeration Threshold for Galled Networks

Michael Fuchs

Department of Mathematical Sciences
National Chengchi University
Taipei 116
Taiwan

Hao Yu

Department of Mathematics
National University of Singapore
Singapore 119076
Republic of Singapore

August 21, 2026

Abstract

We prove two surprising results about phylogenetic networks. First, we show that the structure of tree-child networks with n leaves and k reticulation nodes undergoes a sharp phase transition at $n^{1/3}$: if $k = o(n^{1/3})$, then a random tree-child network is almost surely a semi-simplex tree-child network, whereas if $k/n^{1/3} \rightarrow \infty$ and $k = o(n^{1/2})$, it is almost surely not. Second, we show that this result implies that the asymptotic counting formula for galled networks with n leaves and a fixed number k of reticulation nodes remains valid in the range $k = o(n^{1/3})$, but not beyond. This is in strong contrast to recently established results for the asymptotic counting formulas for tree-child and normal networks with n leaves and k reticulation nodes, which are valid in the (optimal) range $k = o(n^{1/2})$.

1 Introduction and Results

Over the last few decades, phylogenetic networks have increasingly replaced phylogenetic trees as the standard models for many evolutionary scenarios in biology; see [15, 22]. Consequently, a rich mathematical theory has emerged, with many recent advances. Some of these advances concern the enumeration and stochastic behavior of various network classes; see, e.g., [2, 18, 20–22]. In this paper, we add two new results in this area. The literature most closely related to our results is reviewed below.

We start with the precise definition of a (rooted, binary) phylogenetic network, which is the following graph-theoretic structure.

Definition 1.1 (Phylogenetic Networks). A (rooted, binary) *phylogenetic network* is a simple directed acyclic graph (DAG) whose nodes can be classified as follows.

- A unique *root* of indegree 0 and outdegree 1;
- *Leaves*, which are bijectively labeled nodes of indegree 1 and outdegree 0;
- *Tree nodes*, which are nodes of indegree 1 and outdegree 2;

- *Reticulation nodes*, which are nodes of indegree 2 and outdegree 1.

Reticulation nodes (or *reticulations* for short) are the nodes that distinguish general phylogenetic networks from phylogenetic trees, which form a subclass of the former. The *size* of a network is its number of leaves. Moreover, if a network has n leaves, then these are labeled by labels from the set $\{1, \dots, n\}$.

As the space of phylogenetic networks is huge, many subclasses have been defined to make it more tractable; see [17]. In this paper, we consider two of them. The first is the prominent class of tree-child networks together with two of its subclasses; see [3] where it was introduced.

Definition 1.2 (Tree-Child Networks). A phylogenetic network is called a *tree-child network* if every non-leaf node has at least one child that is a tree node or a leaf. In addition, we call a tree-child network *simplex* if the child of every reticulation node is a leaf, and *semi-simplex* if the child of every reticulation node is the root of a subtree.

Let $\text{TC}_{n,k}$ denote the number of tree-child networks with n leaves and k reticulations. The following asymptotic counting result was recently proved in [26]: as $n \rightarrow \infty$,

$$\text{TC}_{n,k} \sim \frac{2^{k-1}\sqrt{2}}{k!} \left(\frac{2}{e}\right)^n n^{n+2k-1} \quad (1)$$

uniformly in the optimal range $k = o(n^{1/2})$. This result improved earlier results from [9–11], where the above asymptotics was proved only for fixed k . In addition, [26] showed that (1) also holds for *normal networks* (introduced in [24]), which are tree-child networks without *shortcuts*, i.e., edges (u, v) such that v can also be reached from u via a path of length ≥ 2 .

Next, for simplex tree-child networks with n leaves and k reticulations, whose number we denote by $\text{STC}_{n,k}$, the following closed-form expression is known:

$$\text{STC}_{n,k} = \binom{n}{k} \frac{(2n-2)!}{2^{n-1}(n-k-1)!}, \quad (0 \leq k \leq n-1); \quad (2)$$

see [4]. Thus, by Stirling's formula, as $n \rightarrow \infty$,

$$\text{STC}_{n,k} \sim \frac{\sqrt{2}}{2k!} \left(\frac{2}{e}\right)^n n^{n+2k-1}$$

again throughout the optimal range $k = o(n^{1/2})$.

As for the number $\text{SSTC}_{n,k}$ of semi-simplex tree-child networks with n leaves and k reticulations, it was proved in [11] that, for fixed k and as $n \rightarrow \infty$, a tree-child network with n leaves and k reticulations chosen uniformly at random is almost surely semi-simplex, i.e.,

$$\text{SSTC}_{n,k} \sim \frac{2^{k-1}\sqrt{2}}{k!} \left(\frac{2}{e}\right)^n n^{n+2k-1}, \quad (3)$$

for fixed k as $n \rightarrow \infty$. In light of the above results, one would expect that the asymptotics in (3) remains true in the range $k = o(n^{1/2})$. Surprisingly, this turns out to be false.

Proposition 1.3. *As $n \rightarrow \infty$ and $k = o(n^{1/2})$,*

$$\text{SSTC}_{n,k} \sim \frac{2^{k-1}\sqrt{2}}{k!} \left(\frac{2}{e}\right)^n n^{n+2k-1} e^{-\sqrt{2k^3/n}}.$$

Consequently, (3) continues to hold in the optimal range $k = o(n^{1/3})$.

From this result, together with (1) from [26], we obtain the following sharp structural phase transition result for tree-child networks, which is our first main result. To the best of our knowledge, this is the first such result in combinatorial phylogenetics, whereas phase transition results are a central theme in probabilistic combinatorics; see, e.g., [1, 7, 16].

Theorem 1.4. *Let $E_{n,k}$ denote the event that a randomly chosen tree-child network with n leaves and k reticulations is semi-simplex. Then, as $n \rightarrow \infty$ and $k = o(n^{1/2})$,*

$$\mathbb{P}(E_{n,k}) \rightarrow \begin{cases} 1, & \text{if } k = o(n^{1/3}); \\ e^{-\sqrt{2}c^{3/2}}, & \text{if } k \sim cn^{1/3}; \\ 0, & \text{if } k/n^{1/3} \rightarrow \infty. \end{cases}$$

Another consequence of Proposition 1.3 concerns the counting of galled networks, another important network class in phylogenetics; see [14].

Definition 1.5. A phylogenetic network is called a *galled network* if every reticulation is in a *tree cycle*, i.e., a union of two edge-disjoint paths from a common tree node to a common reticulation with all other nodes being tree nodes.

Denote by $\text{GN}_{n,k}$ the number of galled networks with n leaves and k reticulations. It was proved in [5] that for k fixed and as $n \rightarrow \infty$,

$$\text{GN}_{n,k} \sim \text{TC}_{n,k} \sim \frac{2^{k-1}\sqrt{2}}{k!} \left(\frac{2}{e}\right)^n n^{n+2k-1}; \quad (4)$$

see also [6]. One might again guess that this extends to the range $k = o(n^{1/2})$, however, this again turns out to be wrong.

Theorem 1.6. *As $n \rightarrow \infty$ and $k = o(n^{1/2})$,*

$$\text{GN}_{n,k} = \frac{2^{k-1}\sqrt{2}}{k!} \left(\frac{2}{e}\right)^n n^{n+2k-1} \left(e^{-\sqrt{2k^3/n}} + o(1)\right).$$

Thus, the asymptotic result (4) remains valid precisely in the range $k = o(n^{1/3})$. This resolves the open problem posed in the conclusion of [26]. Also, beyond that range (but still within the range $k = o(n^{1/2})$), the ratio of galled networks to tree-child networks tends to 0. This is consistent with the fact that, for a fixed number n of leaves, tree-child networks are asymptotically much more numerous than galled networks; see the main results of [12] and [13].

We conclude the introduction with a brief outline of the paper. In the next section, we prove Proposition 1.3. The proof of Theorem 1.6 is given in Section 3. We end the paper with some final remarks in Section 4.

2 Semi-simplex Tree-Child Networks

In this section, we prove Proposition 1.3. We start with the following exponential generating function, which is motivated by (2):

$$F_k(z) := \sum_{\ell \geq 1} \frac{(2\ell + 2k - 2)!}{2^{\ell+k-1}(\ell - 1)!} \cdot \frac{z^\ell}{\ell!}.$$

Note that this is the exponential generating function of the number of simplex tree-child networks with $\ell + k$ leaves and k reticulations, where the leaves below the reticulations receive labels from the set $\{1, \dots, k\}$. Thus, when $k = 0$, $F_0(z)$ counts phylogenetic trees, whose exponential generating function is given by

$$T(z) := F_0(z) = 1 - \sqrt{1 - 2z};$$

see, e.g., page 129 in [8].

The exponential generating function for the number of semi-simplex tree-child networks is derived from $F_k(z)$ as follows.

Lemma 2.1. *The exponential generating function of the number of semi-simplex tree-child networks with n leaves and k reticulations is given by*

$$F_k(z) \frac{T(z)^k}{k!} = F_k(z) \frac{(1 - \sqrt{1 - 2z})^k}{k!}.$$

Proof. By definition, a semi-simplex tree-child network with n leaves and k reticulations is obtained from a simplex tree-child network with $\ell + k$ leaves and k reticulations, where the leaves below the reticulations are labeled by $1, \dots, k$, by replacing the leaves below the reticulations with an unordered forest of phylogenetic trees whose leaf labels consist of the set $\{1, \dots, n - \ell\}$. The replacement is performed as follows. The leaf with label 1 is replaced by the phylogenetic tree in the forest with the smallest label, the leaf with label 2 is replaced by the phylogenetic tree with the second smallest label, and so on. Finally, the leaves are (suitably) relabeled so that their labels are from the set $\{1, \dots, n\}$. Using exponential generating functions, this construction translates directly into the claimed expression. $\square$

We next need a result for $F_k(z)$.

Lemma 2.2. *For $k \geq 1$,*

$$F_k(z) = \frac{(2k)!}{2^k} (1 - 2z)^{-2k+1/2} p_k(z),$$

where $p_k(z)$ is the polynomial of degree k given by

$$p_k(z) := \sum_{\ell=1}^k \frac{1}{2^{\ell-1} \ell} \binom{2\ell-2}{\ell-1} \binom{2k-2}{2\ell-2} z^\ell. \quad (5)$$

Proof. Using the definition of the Gauss hypergeometric function,

$$F_k(z) = \frac{(2k)!}{2^k} {}_2F_1(k + 1/2, k + 1, 2; 2z). \quad (6)$$

Now recall that

$${}_2F_1(k + 1/2, k + 1, 2; x) = (1 - x)^{-2k+1/2} {}_2F_1(1 - k, 3/2 - k, 2; x).$$

Plugging this into (6) gives the claimed expression with

$$p_k(z) := {}_2F_1(1 - k, 3/2 - k, 2; 2z),$$

where (5) is obtained by straightforward simplification of the coefficient of the Gauss hypergeometric function with the above parameters. (The finite degree of the series comes from the fact that the first parameter is an integer.) $\square$

Remark 2.3. For later use, we remark that $p_k(1/2)$ simplifies to:

$$p_k(1/2) = \sum_{\ell=1}^k \frac{1}{2^{2\ell-1}\ell} \binom{2\ell-2}{\ell-1} \binom{2k-2}{2\ell-2} = \frac{1}{2^{2k}k} \binom{4k-2}{2k-1}. \quad (7)$$

This identity follows either by a direct generating function argument or from standard properties of the Gauss hypergeometric function.

Before we can prove Proposition 1.3 from the exponential generating function in Lemma 2.1, which is completely explicit due to Lemma 2.2, we need a final technical lemma. We formulate this lemma in a more general form, as it will also be used in the proof of Lemma 3.3 in the next section.

We recall that $[z^n]f(z)$ denotes the n -th Maclaurin coefficient of $f(z)$.

Lemma 2.4. *Let $1 \leq \alpha = \Theta(k)$ and $0 \leq m \leq k$. Then, as $n \rightarrow \infty$,*

$$[z^n](1-z)^{-\alpha}(1-\sqrt{1-z})^m = \frac{n^{\alpha-1}}{\Gamma(\alpha)} e^{-m\sqrt{\alpha/n}} \left(1 + \mathcal{O}\left(\sqrt{\frac{k}{n}} + \frac{k^2}{n}\right) \right),$$

uniformly for $1 \leq k = o(n^{1/2})$.

Proof. The coefficient in question can be written as a convolution

$$[z^n](1-z)^{-\alpha}(1-\sqrt{1-z})^m = \sum_{j=0}^n [z^j]\tilde{T}(z)^m \binom{n-j+\alpha-1}{n-j}, \quad (8)$$

where $\tilde{T}(z) = T(z/2) = 1 - \sqrt{1-z}$.

We first consider the binomial coefficient which can be written with the help of the gamma function as follows:

$$\binom{n-j+\alpha-1}{n-j} = \frac{\Gamma(n-j+\alpha)}{\Gamma(\alpha)\Gamma(n-j+1)}.$$

Our goal is to prove for it the following uniform asymptotic expansion: for $0 \leq j \leq n$, as $n \rightarrow \infty$,

$$\frac{\Gamma(n-j+\alpha)}{\Gamma(n-j+1)} = n^{\alpha-1} e^{-\alpha j/n} \left(1 + \mathcal{O}\left(\frac{\alpha j^2}{n^2} + \frac{\alpha^2}{n} + \frac{j}{n}\right) \right). \quad (9)$$

We split the proof into two cases.

First, if $0 \leq j \leq n/2$, by the well-known asymptotics of the ratio of two gamma functions

$$\frac{\Gamma(n-j+\alpha)}{\Gamma(n-j+1)} = (n-j)^{\alpha-1} \left(1 + \mathcal{O}\left(\frac{\alpha^2}{n-j}\right) \right) = (n-j)^{\alpha-1} \left(1 + \mathcal{O}\left(\frac{\alpha^2}{n}\right) \right), \quad (10)$$

where the last step follows since $n-j \geq n/2$. Next, for the leading term, as $n \rightarrow \infty$,

$$\begin{aligned} (n-j)^{\alpha-1} &= n^{\alpha-1} e^{(\alpha-1)\log(1-j/n)} \\ &= n^{\alpha-1} e^{-(\alpha-1)j/n + \mathcal{O}(\alpha j^2/n^2)} \\ &= n^{\alpha-1} e^{-(\alpha-1)j/n} \left(1 + \mathcal{O}\left(\frac{\alpha j^2}{n^2}\right) \right) \end{aligned}$$

$$= n^{\alpha-1} e^{-\alpha j/n} \left(1 + \mathcal{O} \left(\frac{\alpha j^2}{n^2} + \frac{j}{n} \right) \right).$$

Plugging this into (10) yields (9) in this case.

Next, we consider the range $n/2 < j \leq n$. Here, we define

$$R(j; n) = \frac{\Gamma(n-j+\alpha)}{\Gamma(n-j+1)} \cdot \frac{\Gamma(n+1)}{\Gamma(n+\alpha)} \cdot \left(\frac{n+\alpha}{n} \right)^j. \quad (11)$$

Note that the second term in $R(j; n)$ has the expansion

$$\frac{\Gamma(n+1)}{\Gamma(n+\alpha)} = n^{1-\alpha} \left(1 + \mathcal{O} \left(\frac{\alpha^2}{n} \right) \right) \quad (12)$$

and the third term has the expansion

$$\left(\frac{n+\alpha}{n} \right)^j = e^{\alpha j/n} \left(1 + \mathcal{O} \left(\frac{\alpha^2 j}{n^2} \right) \right) = e^{\alpha j/n} \left(1 + \mathcal{O} \left(\frac{\alpha^2}{n} \right) \right)$$

which correspond to the reciprocals of the main terms on the right-hand side of (10). (That is why these terms were added as multiplicative factors.) The sequence $R(j; n)$ has the product representation

$$R(j; n) = \prod_{\ell=0}^{j-1} \left(1 + \frac{n-\alpha\ell}{n(n-\ell+\alpha-1)} \right)$$

whose terms exceed one if and only if $n - \alpha\ell > 0$ which in turn holds if and only if $\ell < n/\alpha$. In that case, since $\alpha \geq 1$, we have

$$\frac{n-\alpha\ell}{n-\ell+\alpha-1} \leq \alpha$$

and thus these terms are at most $1 + \alpha/n$. In addition, there are at most $n/\alpha + 1$ such terms. Consequently,

$$R(j; n) \leq \left(1 + \frac{\alpha}{n} \right)^{n/\alpha+1} = \mathcal{O}(1)$$

for $n/2 < j \leq n$. Combining this with (11) and (12) gives

$$\frac{\Gamma(n-j+\alpha)}{\Gamma(n-j+1)} = \mathcal{O} \left(n^{\alpha-1} e^{-\alpha j/n} \right)$$

for $n/2 < j \leq n$. From this (9) follows also for this range due to the appearance of j/n in the $\mathcal{O}$ -term of (9) which is $j/n > 1/2$.

We now plug (9) into (8) which produces the following sums

$$\sum_{j=0}^n [z^j] \tilde{T}(z)^m \rho^j, \quad \sum_{j=0}^n j [z^j] \tilde{T}(z)^m \rho^j, \quad \sum_{j=0}^n j^2 [z^j] \tilde{T}(z)^m \rho^j, \quad (13)$$

where $\rho = e^{-\alpha/n}$ with $\rho \sim 1$ as $n \rightarrow \infty$. These sums can be estimated by using generating functions because

$$\sum_{j=0}^{\infty} [z^j] \tilde{T}(z)^m \rho^j z^j = \tilde{T}(\rho z)^m$$

and thus

$$\sum_{j=0}^{\infty} j[z^j] \tilde{T}(z)^m \rho^j z^{j-1} = \frac{d}{dz} \tilde{T}(\rho z)^m = m \tilde{T}(\rho z)^{m-1} \tilde{T}'(\rho z) \rho \quad (14)$$

and

$$\begin{aligned} \sum_{j=0}^{\infty} j^2[z^j] \tilde{T}(z)^m \rho^j z^{j-1} &= \frac{d}{dz} \left(z \frac{d}{dz} \tilde{T}(\rho z)^m \right) \\ &= m \tilde{T}(\rho z)^{m-1} \tilde{T}'(\rho z) \rho + m(m-1) z \tilde{T}(\rho z)^{m-2} \tilde{T}'(\rho z)^2 \rho^2 \\ &\quad + m z \tilde{T}(\rho z)^{m-1} \tilde{T}''(\rho z) \rho^2. \end{aligned} \quad (15)$$

We start by deriving an asymptotic expansion of the first sum in (13):

$$\sum_{j=0}^n [z^j] \tilde{T}(z)^m \rho^j = \tilde{T}(\rho)^m - \sum_{j>n} [z^j] \tilde{T}(z)^m \rho^j. \quad (16)$$

Note that

$$\begin{aligned} \tilde{T}(\rho)^m &= (1 - \sqrt{1-\rho})^m = e^{m \log(1-\sqrt{1-\rho})} = e^{-m\sqrt{1-\rho} + \mathcal{O}(m\alpha/n)} \\ &= e^{-m\sqrt{\alpha/n} + \mathcal{O}(m\alpha/n)} = e^{-m\sqrt{\alpha/n}} \left(1 + \mathcal{O}\left(\frac{k^2}{n}\right) \right) \end{aligned} \quad (17)$$

where in the last step, we have used that $\alpha = \Theta(k)$ and $m \leq k$. Moreover,

$$\sum_{j>n} [z^j] \tilde{T}(z)^m \rho^j \leq \frac{1}{n} \sum_{j=0}^{\infty} j[z^j] \tilde{T}(z)^m \rho^j = \frac{m \tilde{T}(\rho)^{m-1} \tilde{T}'(\rho) \rho}{n},$$

where the series is obtained by evaluating (14) at $z = 1$. Note that $\tilde{T}'(\rho) = (1-\rho)^{-1/2}/2 \sim \sqrt{n}/(2\sqrt{\alpha})$ and $\tilde{T}(\rho) \sim \rho \sim 1$ as $n \rightarrow \infty$. Thus,

$$m \tilde{T}(\rho)^{m-1} \tilde{T}'(\rho) \rho = \mathcal{O} \left(\tilde{T}(\rho)^m \frac{m\sqrt{n}}{\sqrt{\alpha}} \right) = \mathcal{O} \left(e^{-m\sqrt{\alpha/n}} \sqrt{kn} \right), \quad (18)$$

where we have again used that $\alpha = \Theta(k)$ and $m \leq k$. Consequently,

$$\sum_{j>n} [z^j] \tilde{T}(z)^m \rho^j = \mathcal{O} \left(e^{-m\sqrt{\alpha/n}} \sqrt{\frac{k}{n}} \right).$$

Combining this with (17) and plugging into (16) gives

$$\sum_{j=0}^n [z^j] \tilde{T}(z)^m \rho^j = e^{-m\sqrt{\alpha/n}} \left(1 + \mathcal{O} \left(\sqrt{\frac{k}{n}} + \frac{k^2}{n} \right) \right). \quad (19)$$

Next, for the second sum in (13):

$$\sum_{j=0}^n j[z^j] \tilde{T}(z)^m \rho^j \leq m \tilde{T}(\rho)^{m-1} \tilde{T}'(\rho) \rho = \mathcal{O} \left(e^{-m\sqrt{\alpha/n}} \sqrt{kn} \right), \quad (20)$$

where the last step follows from (18).

Finally, for the third sum in (13), we upper bound it by (15) evaluated at $z = 1$ which gives

$$\sum_{j=0}^n j^2 [z^j] \tilde{T}(z)^m \rho^j = \mathcal{O} \left(\tilde{T}(\rho)^m \left(\frac{m\sqrt{n}}{\sqrt{\alpha}} + \frac{m^2 n}{\alpha} + \frac{mn^{3/2}}{\alpha^{3/2}} \right) \right),$$

where we have used that $\tilde{T}''(\rho) = (1 - \rho)^{-3/2}/4 \sim n^{3/2}/(4\alpha^{3/2})$ as $n \rightarrow \infty$. Thus, since $\alpha = \Theta(k)$ and $m \leq k$,

$$\sum_{j=0}^n j^2 [z^j] \tilde{T}(z)^m \rho^j = \mathcal{O} \left(e^{-m\sqrt{\alpha/n}} \left(kn + n\sqrt{\frac{n}{k}} \right) \right). \quad (21)$$

Now, we finally can plug (9) into (8) and use (19), (20), and (21) which gives the claimed result. $\square$

Remark 2.5. The last lemma is a uniform version of the transfer theorems from singularity analysis; see Chapter VI in [8]. It could also have been derived using complex analysis; however, we decided to follow the above more elementary approach.

As a corollary, we have the following.

Corollary 2.6. *Let $q_k(z)$ be a polynomial of degree k with non-negative coefficients. Then, with the same assumptions as in Lemma 2.4,*

$$[z^n](1 - z)^{-\alpha}(1 - \sqrt{1 - z})^m q_k(z) = q_k(1) \frac{n^{\alpha-1}}{\Gamma(\alpha)} e^{-m\sqrt{\alpha/n}} \left(1 + \mathcal{O} \left(\sqrt{\frac{k}{n}} + \frac{k^2}{n} \right) \right),$$

uniformly for $1 \leq k = o(n^{1/2})$.

Proof. Set $q_k(z) = \sum_{d=0}^k \omega_d z^d$. Then,

$$[z^n](1 - z)^{-\alpha}(1 - \sqrt{1 - z})^m q_k(z) = \sum_{d=0}^k \omega_d [z^{n-d}](1 - z)^{-\alpha}(1 - \sqrt{1 - z})^m. \quad (22)$$

From the last lemma, uniformly in $0 \leq d \leq k$,

$$[z^{n-d}](1 - z)^{-\alpha}(1 - \sqrt{1 - z})^m = \frac{(n-d)^{\alpha-1}}{\Gamma(\alpha)} e^{-m\sqrt{\alpha/(n-d)}} \left(1 + \mathcal{O} \left(\sqrt{\frac{k}{n}} + \frac{k^2}{n} \right) \right).$$

Note that

$$(n-d)^{\alpha-1} = n^{\alpha-1} e^{(\alpha-1)\log(1-d/n)} = n^{\alpha-1} e^{\mathcal{O}(d/n)} = n^{\alpha-1} \left(1 + \mathcal{O} \left(\frac{k^2}{n} \right) \right)$$

and

$$e^{-m\sqrt{\alpha/(n-d)}} = e^{-m\sqrt{\alpha/n} + \mathcal{O}(md\sqrt{\alpha}/n^{3/2})} = e^{-m\sqrt{\alpha/n}} \left(1 + \mathcal{O} \left(\frac{k^{5/2}}{n^{3/2}} \right) \right).$$

Thus, uniformly in $0 \leq d \leq k$

$$[z^{n-d}](1 - z)^{-\alpha}(1 - \sqrt{1 - z})^m = \frac{n^{\alpha-1}}{\Gamma(\alpha)} e^{-m\sqrt{\alpha/n}} \left(1 + \mathcal{O} \left(\sqrt{\frac{k}{n}} + \frac{k^2}{n} \right) \right).$$

Plugging this into (22) gives the claimed result. $\square$

Now, we can prove Proposition 1.3 in a slightly stronger form.

Proof of Proposition 1.3. From Lemma 2.1,

$$\text{SSTC}_{n,k} = n![z^n]F_k(z) \frac{(1 - \sqrt{1 - 2z})^k}{k!}.$$

Plugging into this the expression for $F_k(z)$ from Lemma 2.2,

$$\begin{aligned} \text{SSTC}_{n,k} &= \frac{(2k)!}{2^k k!} n![z^n](1 - 2z)^{-2k+1/2} (1 - \sqrt{1 - 2z})^k p_k(z) \\ &= \frac{(2k)!}{2^k k!} 2^n n![z^n](1 - z)^{-2k+1/2} (1 - \sqrt{1 - z})^k p_k(z/2). \end{aligned}$$

Applying Corollary 2.6 (with $\alpha = 2k - 1/2$ and $m = k$) and Stirling's formula,

$$\begin{aligned} \text{SSTC}_{n,k} &= \frac{(2k)!}{2^k k!} p_k(1/2) \frac{n^{2k-3/2}}{\Gamma(2k - 1/2)} 2^n n! e^{-\sqrt{2k^3/n}} \left(1 + \mathcal{O}\left(\sqrt{\frac{k}{n}} + \frac{k^2}{n}\right) \right) \\ &= c_k \left(\frac{2}{e}\right)^n n^{n+2k-1} e^{-\sqrt{2k^3/n}} \left(1 + \mathcal{O}\left(\sqrt{\frac{k}{n}} + \frac{k^2}{n}\right) \right), \end{aligned}$$

where

$$c_k := \frac{(2k)! p_k(1/2) \sqrt{2\pi}}{2^k k! \Gamma(2k - 1/2)}. \quad (23)$$

What is left is to simplify the constant. From the duplication formula of the gamma function:

$$\Gamma(2k - 1/2) = \frac{(4k - 3)! \sqrt{\pi}}{2^{4k-3} (2k - 2)!}$$

Plugging this into (23) and using (7) gives

$$c_k = \frac{2^{k-1} \sqrt{2}}{k!}$$

which is the claimed result that in addition also contains an error term. $\square$

3 Galled Networks

In this section, we prove Theorem 1.6, which is our second main result. The main step of the proof is to show that Proposition 1.3 extends to networks that are both tree-child and galled. These networks have been named *galled tree-child networks* in [6]. Let $\text{GTC}_{n,k}$ denote their number with n leaves and k reticulations. A useful feature of this class is that their bivariate exponential generating function, i.e.,

$$G(z, v) := \sum_{n \geq 1} \sum_{k \geq 0} \text{GTC}_{n,k} \frac{z^n}{n!} v^k$$

satisfies an implicit equation involving the function $F_k(z)$ from the last section; see [6].

Lemma 3.1 ([6]). *We have,*

$$G(z, v) = \sum_{j \geq 0} F_j(z) \frac{(vG(z, v))^j}{j!}. \quad (24)$$

Using the Lagrange inversion formula, a closed form expression can be derived from this for the exponential generating function of $\text{GTC}_{n,k}$:

$$E_k(z) := \sum_{n \geq 1} \text{GTC}_{n,k} \frac{z^n}{n!}.$$

Lemma 3.2. *For $k \geq 1$,*

$$E_k(z) = \frac{1}{k+1} \sum_{s=1}^k \binom{k+1}{s} F_0(z)^{k+1-s} \sum_{\substack{j_1 + \dots + j_s = k \\ j_i \geq 1}} \prod_{i=1}^s \frac{F_{j_i}(z)}{j_i!}. \quad (25)$$

Proof. Set $H(z, v) := vG(z, v)$. Then, (24) can be rewritten as:

$$H(z, v) = v\Phi(H(z, v)), \quad \Phi(\omega) = \sum_{j \geq 0} F_j(z) \frac{\omega^j}{j!}.$$

This is exactly the setting to which the Lagrange inversion formula can be applied, which yields

$$E_k(z) = [v^k]G(z, v) = [v^{k+1}]H(z, v) = \frac{1}{k+1} [\omega^k] \left(\sum_{j \geq 0} F_j(z) \frac{\omega^j}{j!} \right)^{k+1}.$$

Expanding the power gives the claimed result. $\square$

We will use this in order to show that $\text{GTC}_{n,k}$ is closely related to $\text{SSTC}_{n,k}$.

Lemma 3.3. *As $n \rightarrow \infty$,*

$$\text{GTC}_{n,k} = \text{SSTC}_{n,k} \left(1 + \mathcal{O} \left(\sqrt{\frac{k}{n}} \right) \right).$$

uniformly for $k = o(n^{1/2})$.

The proof requires a final (technical) lemma. Set $a_k := (2k)!/(2^k k!)$ and

$$b_k := a_k p_k(1/2) = \frac{(4k-2)!}{2^{3k-1} (2k-1)! k!} b_k \sim \frac{8^k k!}{4\sqrt{2\pi} k^2} = \Theta \left(\frac{8^k k!}{k^2} \right), \quad (k \rightarrow \infty), \quad (26)$$

where the polynomials $p_k(z)$ are from Lemma 2.2, the closed-form expression follows from identity (7), and the asymptotic approximations follow from Stirling's formula. Then, we have the following property.

Lemma 3.4. *There exists an absolute constant $c_0 > 0$ such that for all $1 \leq s \leq k$:*

$$\sum_{\substack{j_1 + \dots + j_s = k \\ j_i \geq 1}} b_{j_1} \dots b_{j_s} \leq b_k \left(\frac{c_0}{k} \right)^{s-1} s!.$$

Proof. We use induction on s . First, if $s = 1$, the claim trivially holds. So, assume that the claim is true for $s - 1$. We are going to show it for s . By using the induction hypothesis,

$$\begin{aligned} \sum_{\substack{j_1 + \dots + j_s = k \\ j_i \geq 1}} b_{j_1} \cdots b_{j_s} &= \sum_{r=1}^{k-s+1} b_r \sum_{\substack{j_2 + \dots + j_s = k-r \\ j_i \geq 1}} b_{j_2} \cdots b_{j_s} \\ &\leq c_0^{s-2} (s-1)! \sum_{r=1}^{k-s+1} \frac{b_r b_{k-r}}{(k-r)^{s-2}}. \end{aligned}$$

The claim follows from this, if we can prove that

$$\sum_{r=1}^{k-s+1} \frac{b_r b_{k-r}}{(k-r)^{s-2}} \leq c_0 \frac{b_k}{k^{s-1}} \quad (27)$$

for all $s \leq k$. In order to show this, we use (26) which gives

$$\sum_{r=1}^{k-s+1} \frac{b_r b_{k-r}}{(k-r)^{s-2}} = \mathcal{O} \left(8^k \sum_{r=1}^{k-s+1} \frac{r!(k-r)!}{r^2(k-r)^s} \right), \quad (28)$$

where the implied constant is absolute. We break the sum inside the $\mathcal{O}$ -term into three parts:

$$\sum_{r=1}^{k-s+1} \frac{r!(k-r)!}{r^2(k-r)^s} = \frac{(k-1)!}{(k-1)^s} + \sum_{2 \leq r \leq k/2} \frac{r!(k-r)!}{r^2(k-r)^s} + \sum_{k/2 < r \leq k-s+1} \frac{r!(k-r)!}{r^2(k-r)^s}. \quad (29)$$

For the first part, we have

$$\frac{(k-1)!}{(k-1)^s} = \frac{k!}{k^{s+1}} \cdot \frac{k^s}{(k-1)^s} \leq \frac{k!}{k^{s+1}} e^{s/(k-1)} = \mathcal{O} \left(\frac{k!}{k^{s+1}} \right),$$

where the implied constant is absolute. For the second part, by Stirling's formula,

$$\begin{aligned} \sum_{2 \leq r \leq k/2} \frac{r!(k-r)!}{r^2(k-r)^s} &= \mathcal{O} \left(\frac{1}{e^k} \sum_{2 \leq r \leq k/2} r^{r-3/2} (k-r)^{k-r+1/2-s} \right) \\ &= \mathcal{O} \left(\frac{1}{e^k \sqrt{k}} \sum_{2 \leq r \leq k/2} r^{r-3/2} (k-r)^{k-r+1-s} \right) \\ &= \mathcal{O} \left(\frac{k^{k-1/2-s}}{e^k \sqrt{k}} \sum_{2 \leq r \leq k/2} \left(\frac{r}{k} \right)^{r-3/2} \left(1 - \frac{r}{k} \right)^{k-r+1-s} \right) \\ &= \mathcal{O} \left(\frac{k^k}{e^k k^{s+1}} \sum_{2 \leq r \leq k/2} \left(\frac{r}{k} \right)^{r-3/2} \right) = \mathcal{O} \left(\frac{k^k}{e^k k^{s+1}} \right) = \mathcal{O} \left(\frac{k!}{k^{s+1}} \right), \end{aligned}$$

where from the second-last line to the last line, we used that $r \leq k - s + 1$ (otherwise the sum in (27) is empty), and the implied constant is again absolute. Finally, for the last part in (29):

$$\sum_{k/2 < r \leq k-s+1} \frac{r!(k-r)!}{r^2(k-r)^s} = \mathcal{O} \left(\frac{(k-s+1)!(s-1)!}{k^2} \sum_{r \geq s-1} \frac{1}{r^s} \right) = \mathcal{O} \left(\frac{(k-s+1)!(s-1)!}{k^2(s-1)^{s-1}} \right),$$

where we used that $r!(k-r)!$ is non-decreasing in the range $k/2 < r \leq k-s+1$ and the last step follows by a standard integral approximation. Also, note that the implied constant is again absolute. Now, from the simple bound

$$\binom{k}{s-1} \geq \left(\frac{k}{s-1}\right)^{s-1},$$

we obtain that

$$\frac{(k-s+1)!(s-1)!}{(s-1)^{s-1}} \leq \frac{k!}{k^{s-1}}.$$

Thus,

$$\sum_{k/2 < r \leq k-s+1} \frac{r!(k-r)!}{r^2(k-r)^s} = \mathcal{O}\left(\frac{k!}{k^{s+1}}\right)$$

with the implied constant absolute.

Consequently, all three parts in (29) admit the same upper bound. Applying them to (28) gives

$$\sum_{r=1}^{k-s+1} \frac{b_r b_{k-r}}{(k-r)^{s-2}} = \mathcal{O}\left(\frac{8^k k!}{k^2 k^{s-1}}\right) = \mathcal{O}\left(\frac{b_k}{k^{s-1}}\right),$$

where the last step follows from (26). Since the implied constant is absolute, we have proved (27) which concludes the proof of the lemma. $\square$

With the help of the last lemma, we can now prove Lemma 3.3.

Proof of Lemma 3.3. Note that for $s = 1$, the right hand side of (25) is the exponential generating function of the number of semi-simple tree-child networks with n leaves and k reticulations; see Lemma 2.1. Thus,

$$\text{GTC}_{n,k} = n![z^n]E_k(z) = \text{SSTC}_{n,k} + n![z^n]R_k(z), \quad (30)$$

where

$$R_k(z) := \frac{1}{k+1} \sum_{s=2}^k \binom{k+1}{s} F_0(z)^{k+1-s} \sum_{\substack{j_1+\dots+j_s=k \\ j_i \geq 1}} \prod_{i=1}^s \frac{F_{j_i}(z)}{j_i!}.$$

Plugging the expression for $F_k(z)$ from Lemma 2.2 into this gives

$$R_k(z) = \frac{1}{k+1} \sum_{s=2}^k \binom{k+1}{s} (1-2z)^{-2k+s/2} T(z)^{k+1-s} \underbrace{\sum_{\substack{j_1+\dots+j_s=k \\ j_i \geq 1}} \prod_{i=1}^s a_{j_i} p_{j_i}(z)}_{=: r_{k,s}(z)},$$

where $T(z) = 1 - \sqrt{1-2z}$. Note that $r_{k,s}(z)$ is a polynomial of degree k with non-negative coefficients. Thus, by Corollary 2.6,

$$[z^n]R_k(z) = \frac{1}{k+1} \sum_{s=2}^k \binom{k+1}{s} [z^n](1-2z)^{-2k+s/2} T(z)^{k+1-s} r_{k,s}(z)$$

$$\begin{aligned}
&= \frac{2^n}{k+1} \sum_{s=2}^k \binom{k+1}{s} [z^n] (1-z)^{-2k+s/2} \tilde{T}(z)^{k+1-s} r_{k,s}(z/2) \\
&= \mathcal{O} \left(2^n \sum_{s=2}^k \frac{k! r_{k,s}(1/2)}{(k+1-s)! s!} \frac{n^{2k-s/2-1}}{\Gamma(2k-s/2)} e^{-(k+1-s)\sqrt{(2k-s/2)/n}} \right), \quad (31)
\end{aligned}$$

where $\tilde{T}(z) = T(z/2)$. Now, observe that by Lemma 3.4,

$$r_{k,s}(1/2) = \sum_{\substack{j_1+\dots+j_s=k \\ j_i \geq 1}} \prod_{i=1}^s a_{j_i} p_{j_i}(1/2) = \sum_{\substack{j_1+\dots+j_s=k \\ j_i \geq 1}} b_{j_1} \cdots b_{j_s} \leq b_k \left(\frac{c_0}{k} \right)^{s-1} s!.$$

Moreover,

$$\frac{\Gamma(2k-1/2)}{\Gamma(2k-s/2)} = \mathcal{O}((2k)^{(s-1)/2})$$

and

$$e^{-(k+1-s)\sqrt{(2k-s/2)/n}} = e^{-k\sqrt{(2k-s/2)/n} + \mathcal{O}(s)} = e^{-\sqrt{2k^3/n} + \mathcal{O}(s)}.$$

Plugging this into (31) gives

$$[z^n] R_k(z) = \mathcal{O} \left(\frac{b_k}{\Gamma(2k-1/2)} 2^n n^{2k-3/2} e^{-\sqrt{2k^3/n}} \sum_{s=2}^k \frac{k!}{(k+1-s)! k^{s-1}} \left(\tilde{c}_0 \frac{k}{n} \right)^{(s-1)/2} \right),$$

where $\tilde{c}_0$ is a suitable constant. Note that

$$\frac{b_k}{\Gamma(2k-1/2)} = \mathcal{O} \left(\frac{2^k}{k!} \right)$$

which follows by the same line of arguments as in the end of the proof of Proposition 1.3. (The above constant is up to the term $\sqrt{2\pi}$ equal to c_k .) Also,

$$\sum_{s=2}^k \frac{k!}{(k+1-s)! k^{s-1}} \left(\tilde{c}_0 \frac{k}{n} \right)^{(s-1)/2} \leq \sum_{s=2}^k \left(\tilde{c}_0 \frac{k}{n} \right)^{(s-1)/2} = \mathcal{O} \left(\sqrt{\frac{k}{n}} \right).$$

Thus,

$$[z^n] R_k(z) = \mathcal{O} \left(\frac{2^k}{k!} 2^n n^{2k-3/2} e^{-\sqrt{2k^3/n}} \sqrt{\frac{k}{n}} \right)$$

and consequently, by Stirling's formula,

$$n! [z^n] R_k(z) = \mathcal{O} \left(\frac{2^k}{k!} \left(\frac{2}{e} \right)^n n^{n+2k-1} e^{-\sqrt{2k^3/n}} \sqrt{\frac{k}{n}} \right).$$

Plugging this into (30) completes the proof of the lemma. $\square$

Combining Lemma 3.3 with Proposition 1.3, we obtain the following consequence.

Corollary 3.5. *As $n \rightarrow \infty$ and $k = o(n^{1/2})$,*

$$\text{GTC}_{n,k} \sim \frac{2^{k-1}\sqrt{2}}{k!} \left(\frac{2}{e} \right)^n n^{n+2k-1} e^{-\sqrt{2k^3/n}}.$$

We can now deduce from this our second main result.

Proof of Theorem 1.6. Denote by $\text{PN}_{n,k}$ the number of (general) phylogenetic networks with n leaves and k reticulations. It was proved in [26] that, as $n \rightarrow \infty$

$$\text{PN}_{n,k} \sim \frac{2^{k-1}\sqrt{2}}{k!} \left(\frac{2}{e}\right)^n n^{n+2k-1} \quad (32)$$

uniformly for $k = o(n^{1/2})$. Now observe that a network which is a galled network but not a galled tree-child network is contained in the set of phylogenetic networks which are not tree-child networks. Thus,

$$0 \leq \text{GN}_{n,k} - \text{GTC}_{n,k} \leq \text{PN}_{n,k} - \text{TC}_{n,k}.$$

Using (1) and (32) yields

$$\text{GN}_{n,k} - \text{GTC}_{n,k} = o\left(\frac{2^k}{k!} \left(\frac{2}{e}\right)^n n^{n+2k-1}\right)$$

and consequently, from Corollary 3.5,

$$\text{GN}_{n,k} = \frac{2^{k-1}\sqrt{2}}{k!} \left(\frac{2}{e}\right)^n n^{n+2k-1} \left(e^{-\sqrt{2k^3/n}} + o(1)\right).$$

This proves the theorem. □

4 Conclusion

In a series of papers that started in [9] and was continued in [5,6,10,11,19,25], it was shown that the numbers of general networks ($\text{PN}_{n,k}$), tree-child networks ($\text{TC}_{n,k}$), galled networks ($\text{GN}_{n,k}$), and galled tree-child networks ($\text{GTC}_{n,k}$) with n leaves and k reticulations all admit the same asymptotic formula for fixed k as n tends to infinity, namely,

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

In fact, the results for tree-child and galled networks follow from those for general networks and galled tree-child networks, respectively, as the former class is a superclass of both, whereas the latter class is contained in both. On the other hand, the classes of tree-child networks and galled networks have a non-trivial intersection. Recently, the results for general networks and tree-child networks were shown to remain valid in the range $k = o(n^{1/2})$, but not beyond; see [26]. This raised the question whether the same is true for the other two classes, too. In this paper, we surprisingly proved that the square-root threshold for general and tree-child networks becomes a cube-root threshold for galled networks and galled tree-child networks.

We proved the above results by first deriving the asymptotic formula for the number of semi-simplex tree-child networks and then extending it to galled tree-child networks and finally to galled networks. Our proof is elementary in the sense that no complex-analytic tools are used. However, we could have alternatively used tools from analytic combinatorics (which are based

on complex analysis). We avoided this in order to make the paper also accessible to readers who are not familiar with these tools.

Another result which we proved in this paper, and which in fact is the source of the cube-root phenomenon, is a (sharp) phase transition result for random tree-child networks. More precisely, we proved that for $k = o(n^{1/3})$, a tree-child network with n leaves and k reticulations chosen uniformly at random is almost surely semi-simplex. On the other hand, for $k/n^{1/3} \rightarrow \infty$ and $k = o(n^{1/2})$, the opposite holds, namely, it almost surely is not. Moreover, in the transitional range when $k \sim cn^{1/3}$, we found the limiting probability of the event that a random tree-child network is semi-simplex. To the best of our knowledge, this is the first phase transition result for random phylogenetic networks.

Acknowledgements

We thank Louxin Zhang for encouragement and helpful comments. MF acknowledges partial support by the National Science and Technology Council (NSTC), Taiwan under grants NSTC-113-2115-M-004-004-MY3. HY is financially supported by the Singapore MOE Academic Research Fund Tier 1 [A-8001951-00-00].

References

- [1] B. Bollobás. *Random Graphs*, second edition, Cambridge University Press, Cambridge, 2001.
- [2] M. Bouvel, P. Gambette, M. Mansouri (2020). Counting phylogenetic networks of level 1 and 2, *J. Math. Biol.*, **81:6-7**, 1357–1395.
- [3] G. Cardona, F. Rossello, G. Valiente (2009). Comparison of tree-child phylogenetic networks, *IEEE/ACM Trans. Comput. Biol. Bioinform.*, **6**, 552–569.
- [4] G. Cardona and L. Zhang (2020). Counting and enumerating tree-child networks and their subclasses, *J. Comput. System Sci.*, **114**, 84–104.
- [5] Y.-S. Chang and M. Fuchs (2024). Counting phylogenetic networks with few reticulation vertices: galled and reticulation-visible networks, *Bull. Math. Biol.*, **86**, Paper 76.
- [6] Y.-S. Chang, M. Fuchs, G.-R. Yu (2024). Galled tree-child networks, *LIPICs, Proceedings of the 35th International Meeting on Probabilistic, Combinatorial and Asymptotic Methods for the Analysis of Algorithms*, **302**, Paper 8.
- [7] P. Erdős and A. Rényi (1960). On the evolution of random graphs, *Publ. Math. Inst. Hung. Acad. Sci.*, **5**, 17–61.
- [8] P. Flajolet and R. Sedgewick. *Analytic Combinatorics*, Cambridge University Press, Cambridge, UK, 2009.
- [9] 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.

- [10] 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.
- [11] 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.
- [12] M. Fuchs, G.-R. Yu, L. Zhang (2021). On the asymptotic growth of the number of tree-child networks, *Eur. J. Combin.*, **93**, 103278.
- [13] M. Fuchs, G.-R. Yu, L. Zhang (2022). Asymptotic enumeration and distributional properties of galled networks, *J. Comb. Theory Ser. A*, **189**, 105599.
- [14] D. H. Huson and T. H. Klöpper (2007). Beyond galled trees - decomposition and computation of galled networks. In *Lecture Notes in Computer Science*, pages 211–225.
- [15] D. H. Huson, R. Rupp, C. Scornavacca. *Phylogenetic Networks: Concepts, Algorithms and Applications*, Cambridge University Press, New York, NY, USA, 2011.
- [16] S. Janson, T. Łuczak, A. Rucinski. *Random Graphs*, Wiley-Interscience, New York, NY, USA, 2000.
- [17] 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.
- [18] Z. Lin, F. Liu, J. Liu, J. Liu, G. Xin. Proof of a conjecture on Young tableaux with walls, arXiv:2601.09551.
- [19] M. Mansouri (2022). Counting general phylogenetic networks, *Australas. J. Combin.*, **83**, 40–86.
- [20] C. McDiarmid, C. Semple, D. Welsh (2015). Counting phylogenetic networks, *Ann. Comb.*, **19:1**, 205–224.
- [21] 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.
- [22] M. Steel. *Phylogeny—Discrete and Random Processes in Evolution*, CBMS-NSF Regional Conference Series in Applied Mathematics, 89, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2016.
- [23] B. Stuffer (2022). A branching process approach to level-k phylogenetic networks, *Random Struc. Algor.*, **61:2**, 397–421.
- [24] S. J. Willson (2010). Properties of normal phylogenetic networks, *Bull. Math. Biol.*, **72:2**, 340–358.
- [25] H. Yu, M. Fuchs, G.-R. Yu, L. Zhang (2026). Exact and asymptotic counts of binary phylogenetic networks with a few reticulation events, *J. Comput. Biol.*, **33:4**, 386–400.
- [26] H. Yu and L. Zhang. Asymptotic counting of binary phylogenetic networks, arXiv:2605.23126.