A DICHOTOMY LAW FOR CERTAIN CLASSES OF PHYLOGENETIC NETWORKS

MICHAEL FUCHS AND MIKE STEEL

Abstract. Many classes of phylogenetic networks have been proposed in the literature. A feature of several of these classes is that if one restricts a network in the class to a subset of its leaves, then the resulting network may no longer lie within this class. This has implications for their biological applicability, since some species – which are the leaves of an underlying evolutionary network – may be missing (e.g., they may have become extinct, or there are no data available for them) or we may simply wish to focus attention on a subset of the species. On the other hand, certain classes of networks are 'closed' when we restrict to subsets of leaves, such as (i) the classes of all phylogenetic networks or all phylogenetic trees; (ii) the classes of galled networks, simplicial networks, galled trees; and (iii) the classes of networks that have some parameter that is monotone-under-leaf-subsampling (e.g., the number of reticulations, height, etc.) bounded by some fixed value. It is easily shown that a closed subclass of phylogenetic trees is either all trees or a vanishingly small proportion of them (as the number of leaves grows). In this short paper, we explore whether this dichotomy phenomenon holds for other classes of phylogenetic networks, and their subclasses.

Keywords: Phylogenetic networks, asymptotic enumeration, closure, dichotomy

1. Introduction

Phylogenetic networks provide a way for biologists to represent the evolutionary history of present-day species more accurately than traditional phylogenetic trees allow. Whereas trees can adequately represent past speciation and extinction events, networks can also explicitly display reticulate evolutionary events, such as lateral gene transfer and hybridization [2, 12]. The leaves ('tips') of these networks are typically a set of present-day species, and the root of the network represents their most recent common ancestor. Over the last two decades, the definition and study of various phylogenetic classes, their enumeration and combinatorial properties, and deciphering the relationships between these various classes has been an active area of research [13].

A desirable property of any class of networks is that it satisties a certain 'closure' property. Roughly speaking, this property states that if a network in the class is restricted to a subset of its leaf set, then the induced sub-network remains in that class. For example, the class of phylogenetic trees has this property, as does the class of all phylogenetic networks. However, many other classes between these two extremes (e.g., tree-child networks, tree-based networks, etc.) fail to have this closure property.

If a network class is not closed, it might be hoped that a large closed subclass exists within it, perhaps even one that might even be asymptotically equivalent in

Date: September 26, 2025.

size to the full class, as the number of leaves grow. However, in this paper, we demonstrate that for certain classes of networks this is far from the case: for certain 'tight' classes of networks, every closed subset of the class is either the entire class, or it constitutes a vanishingly small proportion of the entire class (as the number of leaves becomes large). Our methods rely primarily on asymptotic enumeration techniques. We end by posing a general question for further study.

1.1. **Definitions: Phylogenetic networks.** Throughout this paper, all trees and networks are directed graphs, represented by a set of vertices and a set of edges (ordered pairs of distinct vertices). We now recall some standard terminology in the phylogenetic literature. A (binary) *phylogenetic network* on $[n] := \{1, \ldots, n\}$ is a directed acyclic graph with n leaves (vertices of out-degree 0) labeled bijectively by the elements of $[n]$, and with each non-leaf vertex having in-degree 1 and out-degree 2 (tree vertices), or in-degree 2 and out-degree 1 (reticulation vertices or reticulations for short), or in-degree 0 and out-degree 1 (the root of the network at the top of an ancestral root edge). Edges which terminate in a reticulation vertex are called *reticulation edges*; all other edges are called *tree edges*. Two phylogenetic networks are regarded as equivalent if there is a directed graph isomorphism between them that maps i to i for each $i \in [n]$. Three important classes of phylogenetic networks are the following:

- A *tree-child network* is a phylogenetic network for which each non-leaf vertex has at least one of its outgoing edges directed to a tree vertex or a leaf.
- A *normal network* is a tree-child network that has no 'shortcut' edge (i.e., no edge (u, v) for which there is another path from u to v).
- A *phylogenetic tree* is a phylogenetic network with no reticulation vertices.

Thus, tree-child networks include normal networks which, in turn, include phylogenetic trees. For more background and details on phylogenetic networks; see [12]. Fig. 1 illustrates these three classes of networks.

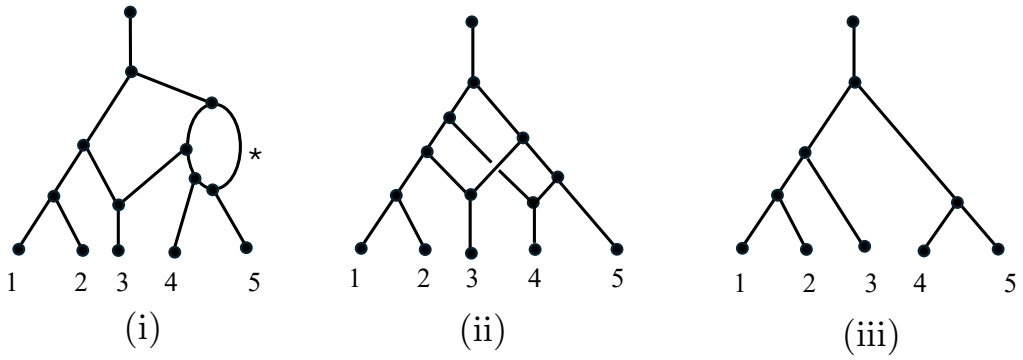

FIGURE 1. Three phylogenetic networks, each having leaf set $\{1, 2, 3, 4, 5\}$. Edges are directed downwards. (i) A tree-child network, (ii) a normal network, and (iii) a phylogenetic tree. Note that although the networks in (i) and (ii) each have two reticulation vertices, the network in (i) is not a normal network due to the presence of a 'shortcut' edge indicated by *.

Apart from the above three classes of phylogenetic networks, two more will play an important role in this paper. For the definition of these two, we need the notion

of a *reticulation cycle* (or *gall*) which is a set of two paths from a common top tree vertex to a common bottom reticulation, with the sets of internal vertices being disjoint. Then, we have the following definitions:

- A *galled network* is a phylogenetic network with each reticulation in exactly one reticulation cycle.
- A *galled tree* is a galled network whose reticulation cycles are edge-disjoint.

Fig. 2 illustrates these two classes of networks.

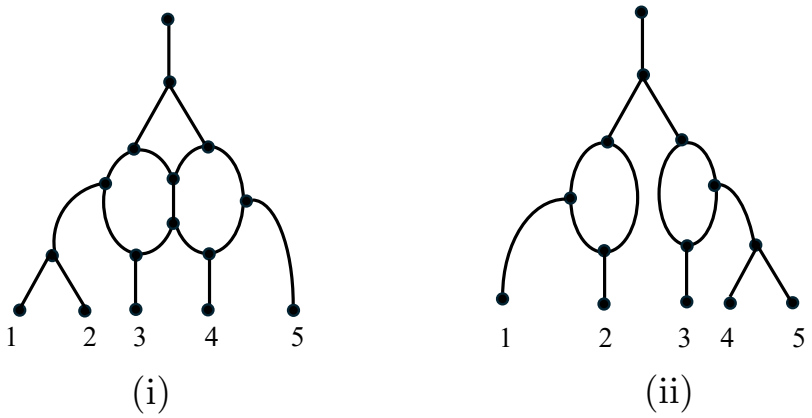

FIGURE 2. (i) A galled network, (ii) a galled tree. Edges are directed downwards.

For more types of networks, we refer the reader to [12].

1.2. **Definitions: Classes of networks.** A *class of networks* $\mathcal{N}$ is the set of all binary networks of a particular type (e.g., tree-child, normal, trees, etc.), and having a leaf set that is a subset of $\mathbb{N} = \{1, 2, 3, \ldots\}$. We will also assume that any class of networks $\mathcal{N}$ satisfies the following two properties:

(P_1) If $N \in \mathcal{N}$, then any relabelling of the leaves of N by distinct elements of $\mathbb{N}$ results in a network that also lies in $\mathcal{N}$[1].

(P_2) For some fixed function f, the number of vertices in any network in $\mathcal{N}$ is bounded above by $f(n)$, where n is the number of leaves of N.

Property (P_1) captures the requirement that classes of networks depend only on the shape of the network and not on the way its leaves are labelled. Property (P_2) is satisfied by many of the well-studied classes of phylogenetic networks (e.g., trees, tree-child networks, galled networks, etc.). In addition, sufficient conditions are known so that a class satisfies (P_2); see [15]. However, (P_2) excludes, for example, the class of *all* phylogenetic networks, or all tree-based networks, since such networks can have an unbounded number of vertices and yet have just two leaves.

[1]Note that because of this property, a class of networks is not necessary a combinatorial class (see, e.g., Definition I.6 in [7]) as the number of networks with a fixed number of leaves is infinite. However, subsequently, we will restrict to sets of networks whose leaf sets are a subset of $[n]$ and these sets will have a finite cardinality.

For a class of networks $\mathcal{N}$ and a subset X of $\mathbb{N}$, let $\mathcal{N}(X)$ be the set of networks in $\mathcal{N}$ that have leaf set X and let $\mathcal{N}_n = \bigcup_{X\subseteq[n]} \mathcal{N}(X)$. Notice that $\mathcal{N}_n \subseteq \mathcal{N}_{n+1}$ and

$$\mathcal{N} = \bigcup_{n\geq 1} \mathcal{N}_n,$$

(i.e., this union equals $\mathcal{N}$ as leaf sets of the networks in $\mathcal{N}$ may be arbitrary subsets of $\mathbb{N} = \{1, 2, 3, \ldots\}$).

For $N \in \mathcal{N}_n$, let $\mathcal{L}_N$ denote the set of leaves of N (a subset of $[n] = \{1, 2, \ldots, n\}$) and, given a subset Y of $\mathcal{L}_N$, let $N|Y$ be the induced phylogenetic network on the restricted leaf set Y. The network $N|Y$ is obtained from N by taking all vertices and edges of N that lie on at least one path from the root of N to at least one leaf in Y, and then suppressing any resulting subdivision vertices (i.e. vertices of in-degree and out-degree equal to 1) and any double edges. Fig. 3 illustrates this notion, and involves both types of suppression.

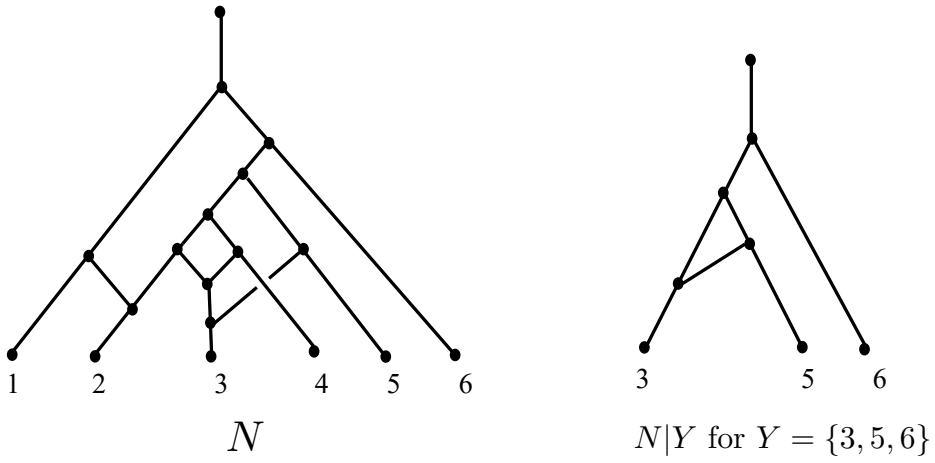

FIGURE 3. Restricting a network to a subset of leaves. Edges are directed downwards.

1.3. **Closed subsets of $\mathcal{N}$ and closed classes.** A subset $\mathcal{C}$ of $\mathcal{N}$ is said to be *closed* if $\mathcal{C}$ satisfies properties $(P_1), (P_2)$ and a further property:

(P_3) If $N \in \mathcal{C}$, and $Y \subseteq \mathcal{L}_N$ then $N|Y \in \mathcal{C}$.

If $\mathcal{C} = \mathcal{N}$, then the class itself is called closed. Examples of closed classes of networks include the following.

- The class of all phylogenetic trees (or any closed subclass, as discussed in the next section);
- Galled trees (i.e., level-1 networks);
- Galled networks;
- The class of networks with at most k reticulation vertices (for any $k \geq 1$);
- The class of networks of height at most k (for any $k \geq 1$);
- An arbitrary union of two or more closed classes;
- The class of *simplicial* networks — i.e., networks for which the child of every reticulation is a leaf;
- The class of *semi-simplicial* networks — i.e., networks for which the child of every reticulation is either a leaf or the root of a tree;

- The class of networks defined by $f(N) \leq k$, where f is a function that satisfies $f(N|Y) \leq f(N)$ for all $Y \subseteq \mathcal{L}_N$ and k is a fixed value (this example generalises some of the entries above).

Examples of classes of networks that are not closed (because they fail to satisfy Properties (P_2) or (P_3) or both) include tree-child networks, normal networks, tree-based networks, and orchard networks. (See [13] for a definition of the latter two network classes.)

2. Closed classes of Phylogenetic trees

Here, 'tree' refers to any finite rooted binary phylogenetic tree. Given any subset X of $[n]$, let $\mathcal{T}(X)$ be the set of all trees on leaf set X. Thus $|\mathcal{T}(X)| = (2|X|-3)!! := (2|X|-3)(2|X|-5)\cdots 1$ (where the empty product is by convention equal to one); see, e.g., Corollary 2.2.4 in [16]. Next, let

$$\mathcal{T}_n = \bigcup_{X\subseteq[n]} \mathcal{T}(X).$$

Notice that $\mathcal{T}_n \subseteq \mathcal{T}_{n+1}$ and

$$|\mathcal{T}_n| = \sum_{j=1}^{n} \binom{n}{j} (2j-3)!! = n! \sum_{j=1}^{n} \frac{C_{j-1}}{(n-j)!2^{j-1}} = \frac{n!}{2^{n-1}} \sum_{j=0}^{n-1} \frac{2^j C_{n-j-1}}{j!},$$

where $C_n = \frac{1}{n+1}\binom{2n}{n}$ denotes the Catalan numbers, which have the asymptotic expansion

$$C_{n-j-1} = \frac{4^{n-j-1}}{\sqrt{\pi n^3}} \left(1 + \mathcal{O}\left(\frac{j+1}{n}\right)\right)$$

uniformly for $j = o(n)$, as $n \to \infty$. In addition, by considering the quotient of consecutive terms of the sequence $2^j C_{n-j-1}/j!, 0 \leq j \leq n-1$, it can be seen that the sequence is decreasing. Thus, by an application of the Laplace method (for a detailed explanation of this method, see Section 4.7 in [7]):

$$|\mathcal{T}_n| \sim \sqrt{\frac{e}{\pi n^3}} 2^{n-1} n! \sim \sqrt{\frac{e}{2}} n^{-1} \left(\frac{2}{e}\right)^n n^n, \qquad (n \to \infty), \tag{1}$$

where we used Stirling's formula in the last step.

We now consider the infinite set

$$\mathcal{T} = \bigcup_{n\geq 1} \mathcal{T}_n,$$

which is the set of *all* finite phylogenetic trees whose leaf set consists of an arbitrary finite subset of $\mathbb{N} = \{1, 2, 3, \ldots\}$.

Examples of closed subsets of $\mathcal{T}$ include:

- $\mathcal{T}$;
- The class of caterpillar trees, i.e., trees with exactly one cherry;
- The class of trees that have height at most k (for any $k \geq 1$);
- The class of trees that have at most k cherries (i.e., pairs of leaves incident with a common vertex) for any $k \geq 1$;
- For a fixed tree, the class of leaf relabellings, and all of the induced subtrees of these trees;

- An arbitrary union of two or more closed subsets of $\mathcal{T}$.

Our first result states that a closed class of trees is either all trees, or an asymptotically negligible proportion of trees.

Proposition 2.1. *Let $\mathcal{C}$ be a closed subset of $\mathcal{T}$. Then either $\mathcal{C} = \mathcal{T}$ or*

$$\lim_{n\to\infty} \frac{|\mathcal{C} \cap \mathcal{T}_n|}{|\mathcal{T}_n|} = 0.$$

Proof. If $\mathcal{C} \neq \mathcal{T}$, then there exists a tree $T \in \mathcal{T}$ that is not present in $\mathcal{C}$. Let τ be the tree shape of T, i.e., T with the labels of the leaves removed. Since $\mathcal{C}$ is closed, none of the trees in $\mathcal{C}$ contains a subtree of shape τ. Let T_n be a tree sampled uniformly at random from $\mathcal{T}_n$. We show that $\mathbb{P}(T_n \in \mathcal{C}) \to 0$ as $n \to \infty$. Let E_n be the event that T_n contains a subtree of shape τ (this subtree shape can be anywhere inside T_n). By the law of total probability, we have

$$\mathbb{P}(E_n) = \mathbb{P}(E_n|T_n \in \mathcal{C})\mathbb{P}(T_n \in \mathcal{C}) + \mathbb{P}(E_n|T_n \notin \mathcal{C})\mathbb{P}(T_n \notin \mathcal{C}). \tag{2}$$

Now, $\mathbb{P}(E_n|T_n \in \mathcal{C}) = 0$ (by definition), and the second (product) term on the right of Eqn. 2 is less than or equal to $\mathbb{P}(T_n \notin \mathcal{C})$. Thus $\mathbb{P}(E_n) \leq \mathbb{P}(T_n \notin \mathcal{C})$. We now apply Corollary 13 of [3], which imples that $\mathbb{P}(E_n)$ tends to 1 as $n \to \infty$. Consequently, as $n \to \infty$, we have $\mathbb{P}(T_n \notin \mathcal{C}) \to 1$ and thus $\mathbb{P}(T_n \in \mathcal{C}) = \frac{|\mathcal{C}\cap\mathcal{T}_n|}{|\mathcal{T}_n|} \to 0$. □

3. Extending Proposition 2.1 beyond trees

For the remainder of this paper we explore the extent to which Proposition 2.1 holds for more general classes of phylogenetic networks. We begin by formalizing this notion.

3.1. **Definitions: Tightness.** We say that a class of networks $\mathcal{N}$ that satisfies (P_1, P_2) is *tight* if the following dichotomy condition holds. For every closed subset $\mathcal{C}$ of $\mathcal{N}$, either $\mathcal{C} = \mathcal{N}$ or

$$\lim_{n\to\infty} \frac{|\mathcal{C} \cap \mathcal{N}_n|}{|\mathcal{N}_n|} = 0. \tag{3}$$

Notice that if $\mathcal{N}$ is **not** a closed class of networks (i.e., it violates condition (P_3)), then the condition that $\mathcal{N}$ is tight is equivalent to the statement that, for any closed subset $\mathcal{C}$ of $\mathcal{N}$, Eqn. 3 holds.

Notice also that every class of networks $\mathcal{N}$ contains a unique maximal closed class with respect to set inclusion, namely, the union of *all* closed classes in $\mathcal{N}$ (which, as noted above, forms a closed class). We denote this maximal closed class for $\mathcal{N}$ by $\mathcal{C}_\mathcal{N}$. Thus, if $\mathcal{N} \neq \mathcal{C}_\mathcal{N}$ (i.e., $\mathcal{N}$ is not closed), then $\mathcal{N}$ is tight if and only if Eqn. 3 holds for the single closed class $\mathcal{C} = \mathcal{C}_\mathcal{N}$.

3.2. **Examples.** We give some examples of classes which are tight and not tight

Example 1: The set $\mathcal{T}$ of all binary phylogenetic trees is tight, by Proposition 2.1.

Example 2: Let $\mathcal{N}$ be the set of **all** galled trees. Then, $\mathcal{N}$ is tight by Corollary 14 of [3] and a similar argument to that used in Proposition 2.1.

Example 3: Let $\mathcal{N}$ be the class of galled networks. Then $\mathcal{N}$ is not tight, since if we take $\mathcal{C}$ to be the closed subclass of simplicial galled networks, we have $\mathcal{C} \subsetneq \mathcal{N}$ but Eqn. 3 does not hold due to results in [11]. More precisely, we have the following.

Proposition 3.1. *If $\mathcal{N}$ is the class of galled networks and $\mathcal{C}$ is the class of simplicial galled networks, then*

$$\lim_{n\to\infty} \frac{|\mathcal{C} \cap \mathcal{N}_n|}{|\mathcal{N}_n|} = e^{-3/8}.$$

Proof. First, note that the class of simplicial galled networks and the class of all simplicial networks coincide (see, for example, Proposition 2 in [5]). Next, it was proved in [11] that the numbers of galled networks (GN_ℓ) and the number of simplicial networks (SN_ℓ) with ℓ leaves that are labelled by the set $X = \{1, \ldots, \ell\}$ admit the following asymptotics, as $\ell \to \infty$,

$$\mathrm{GN}_\ell \sim \frac{\sqrt{2e\sqrt[4]{e}}}{4}\ell^{-1}\left(\frac{8}{e^2}\right)^\ell \ell^{2\ell} \sim \frac{\sqrt{2e\sqrt[4]{e}}}{8\pi}\ell^{-2}8^\ell \ell!^2 \tag{4}$$

and

$$\mathrm{SN}_\ell \sim \frac{\sqrt{2\sqrt{e}}}{4}\ell^{-1}\left(\frac{8}{e^2}\right)^\ell \ell^{2\ell} \sim \frac{\sqrt{2e\sqrt{e}}}{8\pi}\ell^{-2}8^\ell \ell!^2. \tag{5}$$

We now use a similar line of reasoning as for Eqn. 1 in Section 2 to find the asymptotics of $|\mathcal{C} \cap \mathcal{N}_n|$ and $|\mathcal{N}_n|$. First,

$$|\mathcal{N}_n| = \sum_{j=1}^{n} \binom{n}{j} \mathrm{GN}_j = \mathrm{GN}_n + \sum_{j=1}^{n-1} \binom{n}{j} \mathrm{GN}_j$$

and from Eqn. 4, we obtain for the sum-term:

$$\sum_{j=1}^{n-1} \binom{n}{j} \mathrm{GN}_j = \mathcal{O}\left(\sum_{j=1}^{n-1} \binom{n}{j} j^{-2} 8^j j!^2\right) = \mathcal{O}\left(8^n n! \sum_{j=1}^{n-1} \frac{(n-j-1)!}{j!8^j(n-j)}\right),$$

where the summation index was changed from j to $n-j$ in the last step. Next, observe that, as $n \to \infty$,

$$\frac{(n-j-1)!}{j!8^j(n-j)} = \frac{1}{j!8^j} n^{-j-2} n! \left(1 + \mathcal{O}\left(\frac{j^2}{n}\right)\right)$$

uniformly for $j = o(\sqrt{n})$. Moreover, $(n-j-1)!/(j!8^j(n-j)), 1 \le j \le n-1$ is a decreasing sequence. Thus, by applying the Laplace method:

$$\sum_{j=1}^{n-1} \binom{n}{j} \mathrm{GN}_j = \mathcal{O}\left(n^{-3} 8^n n!^2\right).$$

Consequently,

$$|\mathcal{N}_n| \sim \mathrm{GN}_n \sim \frac{\sqrt{2e\sqrt[4]{e}}}{4} n^{-1} \left(\frac{8}{e^2}\right)^n n^{2n}, \qquad (n \to \infty).$$

Likewise, from Eqn. 5,

$$|\mathcal{C} \cap \mathcal{N}_n| \sim \frac{\sqrt{2\sqrt{e}}}{4} n^{-1} \left(\frac{8}{e^2}\right)^n n^{2n}, \qquad (n \to \infty).$$

The claimed result follows from these two asymptotic relations. □

Example 4: Let $\mathcal{N}$ be the class of semi-simplicial networks. Then $\mathcal{N}$ is not tight, since the closed subclass $\mathcal{C}$ of simplicial networks violates Eqn. 3, as this limit cannot tend to zero due to the previous example and the fact that semi-simplicial networks and simplicial networks are both galled networks. In fact, we can even compute the limit precisely.

Proposition 3.2. *If $\mathcal{N}$ is the class of semi-simplicial networks and $\mathcal{C}$ the class of simplicial networks, then*

$$\lim_{n\to\infty} \frac{|\mathcal{C}\cap\mathcal{N}_n|}{|\mathcal{N}_n|} = e^{-1/16}.$$

This is proved in a similar way to Proposition 3.1 by using the following result, which will be established in the appendix.

Theorem 3.3. *The number SSN_ℓ of semi-simplicial networks with ℓ leaves that are labelled by the set $X = \{1,\dots,\ell\}$ admits the asymptotics:*

$$\mathrm{SSN}_\ell \sim \frac{\sqrt{2\sqrt{e\sqrt[4]{e}}}}{4}\ell^{-1}\left(\frac{8}{e^2}\right)^\ell \ell^{2\ell}, \qquad (\ell\to\infty).$$

4. Tree-child and normal networks

In this section, we show that (i) the classes of all tree-child and all normal networks are both tight, and (ii) the class of tree-child networks with at most $k \geq 1$ reticulations is not tight, whereas the class of normal networks with at most $k \geq 1$ reticulations is tight.

4.1. **The classes of all tree-child and all normal networks.** First, let $\mathcal{N}$ be the class of all binary tree-child networks. Observe that the type of argument used to establish the tightness of the class of binary phylogenetic trees (Proposition 2.1) cannot be used to show that $\mathcal{N}$ is tight, since there are tree-child networks (and indeed trees) that have a small probability of being present in a large uniformly sampled tree-child network. To see why, observe that by Corollary 1.10(i) of [6], the expected number of cherries (i.e., pairs of leaves incident with a common vertex) is $\mathcal{O}(1)$. If we now take T to be a complete balanced binary tree with 2^k leaves (and so 2^{k-1} cherries), then the probability that a uniformly sampled network N in $\mathcal{N}$ with n leaves contains T as a subtree (somewhere within N) does not tend to 1 as n grows, provided that k is chosen sufficiently large.

Nevertheless, we now provide a different argument to establish the following result.

Theorem 4.1. *The class of all tree-child networks is tight.*

For the proof of this result, we need the following lemma (which uses the definition of a *reticulation cycle* from Section 1.1).

Lemma 4.2. *The largest closed subclass $\mathcal{C}_\mathcal{N}$ of the class $\mathcal{N}$ of tree-child networks is the class of galled trees.*

Proof. Let $\mathcal{C}$ be a subclass of the class of tree-child networks which contains at least one network N that is not a galled tree. Then, N contains a reticulation v which is the bottom vertex of a reticulation cycle that contains either (i) another reticulation or (ii) a tree vertex u whose child which is not on the cycle also has a reticulation

as descendant. Choose the lowest such vertex on one of the paths from the bottom tree vertex of the reticulation cycle which contains v. Also, in Case (i), pick a leaf $\tilde{u}$ which can be reached from v via a path consisting only of tree vertices (such a path exists due to the tree-child property). Then, the induced subnetwork $N|\{\tilde{u}\}$ is not tree-child (since it contains a reticulation whose child is another reticulation). In Case (ii), let $\tilde{v}$ be a reticulation on the path from $\tilde{u}$ which does not contain v so that all other vertices of this path are tree nodes. Moreover, choose a leaf $\hat{u}$ which again can be reached via a path consisting only of tree vertices from $\tilde{v}$. Then, $N|\{\tilde{u}, \hat{u}\}$ is not tree-child (as it contains a tree vertex with both children reticulations). Thus, in both cases $\mathcal{C}$ is not closed. This proves the desired result. □

Proof of Theorem 4.1. First, recall from [14] that the number of tree-child networks (TC_ℓ) with ℓ leaves that are labelled by the set $X = \{1, \ldots, \ell\}$ is bounded from below by

$$\mathrm{TC}_\ell = \Omega(c^\ell \ell^{2\ell}) \tag{6}$$

for some constant $c > 0$. Consequently,

$$|\mathcal{N}_n| = \sum_{j=1}^{n} \binom{n}{j} \mathrm{TC}_j = \Omega\left(c^n n^{2n}\right). \tag{7}$$

Next, let $\mathcal{C}$ be the class of galled trees. Recall from [4] that the number of galled trees (GT_ℓ) with ℓ leaves that are labelled by the set $X = \{1, \ldots, \ell\}$ is given by:

$$\begin{aligned}\mathrm{GT}_\ell &= \frac{\sqrt{34}(\sqrt{17}-1)}{136}\ell^{-1}\left(\frac{8}{e}\right)^\ell \ell^\ell(1+\mathcal{O}(\ell^{-1})) \\ &= \frac{\sqrt{17}(\sqrt{17}-1)}{136\sqrt{\pi}}\ell^{-3/2}8^\ell\ell!(1+\mathcal{O}(\ell^{-1})), \qquad (\ell\to\infty).\end{aligned}$$

Thus, by the Laplace method:

$$\begin{aligned}|\mathcal{C}\cap\mathcal{N}_n| = \sum_{j=1}^{n}\binom{n}{j}\mathrm{GT}_j &= \frac{\sqrt{17}(\sqrt{17}-1)n!}{136\sqrt{\pi}}\sum_{j=1}^{n}\frac{j^{-3/2}8^j}{(n-j)!}(1+\mathcal{O}(j^{-1})) \\ &\sim \frac{\sqrt{17\sqrt[4]{e}}(\sqrt{17}-1)}{136\sqrt{\pi}}n^{-3/2}8^n n! \\ &\sim \frac{\sqrt{34\sqrt[4]{e}}(\sqrt{17}-1)}{136}n^{-1}\left(\frac{8}{e}\right)^n n^n, \qquad (n\to\infty).\end{aligned}$$

Consequently, $\mathcal{C}$ satisfies Eqn. 3 and from the statement in the third paragraph of Section 3, we obtain that $\mathcal{N}$ is tight as claimed. □

Remark 4.3. In [10], Eqn. 6 was refined to

$$\mathrm{TC}_\ell = \Theta\left(\ell^{-2/3}e^{a_1(3\ell)^{1/3}}\left(\frac{12}{e^2}\right)^\ell \ell^{2\ell}\right), \qquad (\ell\to\infty),$$

where a_1 denotes the largest root of the Airy function of the first kind. Using this result, Eqn. 7 can also be refined as follows

$$|\mathcal{N}_n| = \Theta\left(n^{-2/3}e^{a_1(3n)^{1/3}}\left(\frac{12}{e^2}\right)^n n^{2n}\right), \qquad (n\to\infty).$$

Now, let us turn to the class $\mathcal{N}$ of normal networks. Here, similar to Lemma 4.2, we have the following result.

Lemma 4.4. *The largest closed subclass $\mathcal{C}_{\mathcal{N}}$ of the class $\mathcal{N}$ of normal networks is the class of phylogenetic trees.*

Proof. Since every normal network is also tree-child, Lemma 4.2 implies that $\mathcal{C}_{\mathcal{N}}$ must be contained in the class of galled trees. Assume now that $\mathcal{C}_{\mathcal{N}}$ contains a network N which is not a phylogenetic tree. Then, N contains a reticulation v which is in a reticulation cycle whose vertices apart from v are all tree vertices. Moreover, since all networks are simple, there is at least one tree vertex $\tilde{v}$ which is not the top tree vertex of this reticulation cycle. Let u be a leaf which can be reached via a path that only contains tree vertices from v and let $\tilde{u}$ be a leaf which also can be reached via a path that only contains tree vertices from $\tilde{v}$. Then, the network $N|\{u, \tilde{u}\}$ is not normal. This is a contradiction, since $\mathcal{C}_{\mathcal{N}}$ was assumed to be closed. Thus, $\mathcal{C}_{\mathcal{N}}$ must be the class of phylogenetic trees. □

We now apply this lemma to show that $\mathcal{N}$ is tight as well.

Theorem 4.5. *The class of all normal networks is tight.*

Proof. From [14], we know that the number of normal networks (NN_ℓ) with ℓ leaves that are labelled by the set $X = \{1, \dots, \ell\}$ is bounded from below by

$$\mathrm{NN}_\ell = \Omega(c^\ell \ell^{2\ell})$$

for some constant $c > 0$. Thus,

$$|\mathcal{N}_n| = \sum_{j=1}^{n} \binom{n}{j} \mathrm{NN}_j = \Omega(c^n n^{2n}).$$

Moreover, for the largest closed class $\mathcal{C}_{\mathcal{N}}$ in $\mathcal{N}$, from Lemma 4.4 and Eqn. 1 we have,

$$|\mathcal{C}_{\mathcal{N}} \cap \mathcal{N}_n| = |\mathcal{T}_n| \sim \sqrt{\frac{e}{2}} n^{-1} \left(\frac{2}{e}\right)^n n^n, \qquad (n \to \infty).$$

Consequently, Eqn. 3 holds and thus the claim is proved. □

4.2. **Tree-child and normal networks with a bounded number of reticulations.** Now, suppose that $\mathcal{N}$ is the class of tree-child networks with at most k (fixed) reticulations where $k \geq 1$. We are going to establish the following.

Proposition 4.6. *The class $\mathcal{N}$ of tree-child networks with at most k reticulations is not tight for each $k \geq 1$.*

Proof. We deal with the cases $k \geq 2$ and $k = 1$ separately.

For the case $k \geq 2$, it was shown in [8] that the number of tree-child networks ($\mathrm{TC}_{\ell,k}$) with exactly k reticulations and ℓ leaves that are labelled by the set $X = \{1, \dots, \ell\}$ satisfies, as $\ell \to \infty$,

$$\mathrm{TC}_{\ell,k} = \frac{2^{k-1}\sqrt{2}}{k!} \ell^{2k-1} \left(\frac{2}{e}\right)^\ell \ell^\ell \left(1 + \mathcal{O}(\ell^{-1/2})\right) = \frac{2^{k-1}}{k!\sqrt{\pi}} 2^\ell \ell^{2k-3/2} \ell! \left(1 + \mathcal{O}(\ell^{-1/2})\right).$$

Thus, the number of tree-child networks with at most k reticulations (denoted by $\mathrm{TC}_{\ell,\leq k}$) also satisfies the same asymptotic result, since:

$$\mathrm{TC}_{\ell,\leq k} = \sum_{i=0}^{k} \mathrm{TC}_{\ell,i}.$$

Combining the last two equations gives:

$$|\mathcal{N}_n| = \sum_{j=1}^{n} \binom{n}{j} \mathrm{TC}_{j,\leq k} = \frac{2^{k-1} n!}{k!\sqrt{\pi}} \sum_{j=1}^{n} \frac{2^j j^{2k-3/2}}{(n-j)!} \left(1 + \mathcal{O}(j^{-1/2})\right),$$

and using the Laplace method, we obtain

$$|\mathcal{N}_n| \sim \frac{2^{k-1}}{k!} \sqrt{\frac{e}{\pi}} n^{2k-3/2} 2^n n! \sim \frac{2^{k-1}\sqrt{2e}}{k!} n^{2k-1} \left(\frac{2}{e}\right)^n n^n.$$

Next, by the proof of Lemma 4.2, the largest closed subclass of $\mathcal{N}$ is the class of galled trees with at most k reticulations. Denote this subclass by $\mathcal{C}$. It was shown in [1] that the number of galled trees ($\mathrm{GT}_{\ell,k}$) with exactly k reticulations and ℓ leaves that are labelled by the set $X = \{1, \ldots, \ell\}$ satisfies, as $\ell \to \infty$,

$$\mathrm{GT}_{\ell,k} = \frac{2^{2k-1}\sqrt{2}}{(2k)!} \ell^{2k-1} \left(\frac{2}{e}\right)^{\ell} \ell^{\ell} (1 + \mathcal{O}(\ell^{-1/2})).$$

(Actually, the result in [1] was for normal galled trees; however, the same method of proof shows that the number of all galled trees has the same asymptotics.) Therefore, we obtain

$$|\mathcal{C} \cap \mathcal{N}_n| \sim \frac{2^{2k-1}\sqrt{2e}}{(2k)!} n^{2k-1} \left(\frac{2}{e}\right)^n n^n$$

and thus

$$\lim_{n\to\infty} \frac{|\mathcal{C} \cap \mathcal{N}_n|}{|\mathcal{N}_n|} = \frac{2^k k!}{(2k)!} = \frac{1}{(2k-1)!!} > 0 \quad \text{for } k \geq 2, \tag{8}$$

which proves the claimed result for each $k \geq 2$.

It remains to consider the case $k = 1$. Notice that, in this case, $\mathcal{N} = \mathcal{C}$ and thus, $\mathcal{N}$ is a closed class of networks, in contrast to the cases where $k > 1$. Here, instead of $\mathcal{C}$, we consider the subclass $\mathcal{C}'$ of $\mathcal{N}$ consisting of all simplicial tree-child networks with at most 1 reticulation. Then $\mathcal{C}'$ is a closed subclass of $\mathcal{N}$. Moreover it is clear that $\mathcal{C}' \neq \mathcal{N}$. Next, recall that it was proved in [5] that the number of simplicial tree-child networks ($\mathrm{STC}_{\ell,k}$) with exactly k reticulations and ℓ leaves that are labelled by the set $X = \{1, \ldots, \ell\}$ admits the closed-form expression:

$$\mathrm{STC}_{\ell,k} = \binom{\ell}{k} \frac{(2\ell-2)!}{2^{\ell-1}(\ell-k-1)!}; \tag{9}$$

Thus, by Stirling's formula, as $\ell \to \infty$,

$$\mathrm{STC}_{\ell,k} = \frac{\sqrt{2}}{2k!} \ell^{2k-1} \left(\frac{2}{e}\right)^{\ell} \ell^{\ell} \left(1 + \mathcal{O}(\ell^{-1})\right).$$

and consequently for the number of simplicial tree-child networks ($\mathrm{STC}_{\ell,\leq k}$) with at most k reticulations,

$$\mathrm{STC}_{\ell,\leq k} = \sum_{i=0}^{\ell} \mathrm{STC}_{\ell,i} = \frac{\sqrt{2}}{2k!}\ell^{2k-1}\left(\frac{2}{e}\right)^{\ell}\ell^{\ell}\left(1+\mathcal{O}(\ell^{-1})\right).$$

Then, as above, where we set $k=1$:

$$|\mathcal{C}'\cap\mathcal{N}_n| = \sum_{j=1}^{n}\binom{n}{j}\mathrm{STC}_{j,\leq 1} \sim \frac{\sqrt{2e}}{2}n\left(\frac{2}{e}\right)^{n} n^n$$

which in turn yields

$$\lim_{n\to\infty}\frac{|\mathcal{C}'\cap\mathcal{N}_n|}{|\mathcal{N}_n|} = \frac{1}{2}.$$

This shows that $\mathcal{N}$ is not tight for $k=1$, too. □

Remark 4.7. Note that if $\mathcal{C}'$ denotes the class of simplicial-tree child networks with at most k reticulations, then for all $k\geq 1$, we have

$$\lim_{n\to\infty}\frac{|\mathcal{C}'\cap\mathcal{N}_n|}{|\mathcal{N}_n|} = 2^{-k}.$$

However, this does not show that $\mathcal{N}$ is not tight for $k\geq 2$ as $\mathcal{C}'$ is only a closed class for $k=1$.

Next, let $\mathcal{N}$ be the class of normal networks with at most k (fixed) reticulations where again $k\geq 1$. By Lemma 4.4, the largest closed subclass of $\mathcal{N}$ is the set of phylogenetic trees, which we now denote by $\mathcal{C}$. Then, in contrast to tree-child networks, this class is tight.

Proposition 4.8. *The class $\mathcal{N}$ of normal networks with at most k reticulations is tight for each $k\geq 1$.*

Proof. Normal networks with exactly k reticulations and ℓ leaves that are labelled by the set $X=\{1,\ldots,\ell\}$ have also been counted in [10]; see also [9]. More precisely, denote the number of such networks by $\mathrm{NN}_{\ell,k}$. Then, it was shown in [9, 10] that $\mathrm{NN}_{\ell,k}$ admits the same first-order asymptotics as $\mathrm{TC}_{\ell,k}$:

$$\mathrm{NN}_{\ell,k} = \frac{2^{k-1}\sqrt{2}}{k!}\ell^{2k-1}\left(\frac{2}{e}\right)^{\ell}\ell^{\ell}(1+\mathcal{O}(\ell^{-1/2}), \qquad (\ell\to\infty).$$

Consequently, with the same proof as above, we have

$$|\mathcal{N}_n| \sim \frac{2^{k-1}\sqrt{2e}}{k!}n^{2k-1}\left(\frac{2}{e}\right)^{n} n^n.$$

Applying Eqn. 1 to provide the upper bound on $|\mathcal{C}\cap\mathcal{N}_n|$, where $\mathcal{C}$ denotes the class of phylogenetic trees (see Lemma 4.4), and dividing by $|\mathcal{N}_n|$ shows that Eqn. 3 holds for all $k\geq 1$. □

5. Concluding comments

In this paper, we have introduced the notion of a class of networks being 'tight' and shown that several standard classes of phylogenetic networks satisfy this property. These include the classes of trees, galled trees, tree-child networks, normal networks, and normal networks with at most k reticulations for $k \geq 1$. On the other hand, we have also established that various other network classes fail to have this tightness property, in particular, the classes of galled networks, semi-simplicial networks, and the class of tree-child networks with at most k reticulations for $k \geq 1$. It may therefore be of interest in future work to consider which other classes of phylogenetic networks satisfying property (P_2) are tight.

6. Acknowledgements

We thank the reviewers for many useful comments. The second author thanks Andrew Francis for preliminary discussions regarding the concept of closed classes of networks and the tightness properties of trees. He also thanks the NZ Marsden Fund for research support (23-UOC-003). The first author acknowledges partial support by National Science and Technology Council, Taiwan under research grant NSTC-113-2115-M-004-004-MY3.

References

[1] Agranat-Tamir, L., Fuchs, M., Gittenberger, B. and Rosenberg, N. A. Enumerative combinatorics of unlabelled and labelled time-consistent galled trees. ArXiv:2504.16302.

[2] Bapteste, E., van Iersel, L., Janke, A., Kelchner, S., Kelk, S., McInerney, J., Morrison, D., Nakhleh, L., Steel, M., Stougie, L. and Whitfield, J. (2013). Networks: Expanding evolutionary thinking. Trends Genet. 29(8): 439–441.

[3] Bienvenu, F. and Steel, M. (2024). 0–1 laws for pattern occurrences in phylogenetic trees and networks. Bull. Math. Biol. 86, 94.

[4] Bouvel, M., Gambette, P. and Mansouri, M. (2022). Counting phylogenetic networks of level 1 and 2. J. Math. Biol. 81 (6-7): 1357–1395.

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

[6] Chang, Y.-S., Fuchs, M., Liu, H., Wallner, M. and Yu, G.-R. (2024). Enumerative and distributional results for d-combining tree-child networks. Adv. Appl. Math. 157: 102704. https://doi.org/10.1016/j.aam.2024.102704.

[7] Flajolet, P. and Sedgewick, R. (2013). An Introduction to the Analysis of Algorithms. 2nd ed., Addison-Wesley Professional.

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

[9] Fuchs, M., Steel, M. and Zhang, Q. (2025). Asymptotic enumeration of normal and hybridization networks via tree decoration. Bull. Math. Biol. 87: Paper 69.

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

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

[12] Huson, D.H., Rupp, R. and Scornavacca, C. (2011). Phylogenetic Networks: Concepts, Algorithms and Applications. Cambridge University Press.
[13] Kong, S., Pons, J.C., Kubatko, L. et al. (2022). Classes of explicit phylogenetic networks and their biological and mathematical significance. J. Math. Biol. 84, 47. doi.org/10.1007/s00285-022-01746-y
[14] McDiarmid, C., Semple, C. and Welsh, D. (2015). Counting phylogenetic networks. Ann. Comb. 19, 205–224.
[15] Semple, C. (2017). Size of a phylogenetic network. Discrete Appl. Math. 217, 362–367.
[16] Semple, C. and Steel, M. (2003). Phylogenetics. Oxford University Press.

7. Appendix: Proof of Theorem 3.3

In this appendix, we prove Theorem 3.3. Note that semi-simplicial networks are galled networks which have been investigated in [11]. The proof of Theorem 3.3 relies heavily on tools from this paper and is similar to the proof of Eqn. 4.

One crucial observation in [11] was the following: a galled network is almost surely a simplicial (galled) network N with simplicial networks with two leaves attached to some of the leaves below reticulations of N. Thus, in order to count semi-simplicial networks, we only have to count simplicial networks N with cherries attached to some of the leaves below reticulations of N. For this number $(\widetilde{\mathrm{SSN}}_\ell)$, we have the following closed-form expression, where $N_{\ell+1}^{(k)}$ denotes the number of simplicial networks with ℓ leaves and k reticulations, where the leaves below the reticulations have labels from the set $\{1,\dots,k\}$ (this notation comes from [11]).

Lemma 7.1. *We have*

$$\widetilde{SSN}_\ell = \sum_{j=0}^{\lfloor \ell/2 \rfloor} \binom{\ell}{2j} \frac{(2j)!}{2^j j!} \sum_{i=0}^{\ell-2j} \binom{\ell-2j}{i} N_{\ell-j+1}^{(i+j)}. \tag{10}$$

Proof. For $0 \le j \le \ell/2$ and $0 \le i \le \ell - j$, the number of simplicial networks with $i+j$ reticulations whose leaves are labelled by the set $\{1,\dots,i+j\}$ and with $\ell - j$ leaves in total is given by $N_{\ell-j+1}^{(i+j)}$. For each such network, replace the leaves with labels $\{1,\dots,j\}$ by cherries and then relabel all leaves so that the resulting labels are all different. For the relabelling, there are

$$\binom{\ell}{2j}\binom{\ell-2j}{i}\frac{(2j)!}{2^j j!},$$

ways, as we first have to choose $2j$ labels for the cherries and then have to separate the remaining $\ell - 2j$ leaves into those which are below reticulations (i) and those which are not $(\ell - 2j - i)$. The last factor is because swapping the leaves of a cherry results in the same network and thus the $2j$ labels for the cherries have to be divided into j disjoint sets of size 2. □

Proof of Theorem 3.3. As mentioned at the beginning of this appendix, we know from [11] that

$$\mathrm{SSN}_\ell \sim \widetilde{\mathrm{SSN}}_\ell, \qquad (\ell \to \infty). \tag{11}$$

Thus, we only need to concentrate on the latter sequence for which Eqn. 10 holds. We break the first summation into two parts according to whether $j < \sqrt[4]{\ell}$ or $\sqrt[4]{\ell} \leq j \leq \lfloor \ell/2 \rfloor$. The second part is shown to be an exponentially small function multiplied with $\ell^{2\ell}$ as in the proof of Proposition 4 in [11]. Thus, the main contribution comes from the first part, for which we use the estimate of Lemma 10-(ii) from [11], which reads:

$$\sum_{i=0}^{\ell-2j} \binom{\ell-2j}{\ell} N_{\ell-j+1}^{(i+j)} = \frac{\sqrt{2\sqrt{e}}}{2^{3j+2}} \ell^{-1} \left(\frac{8}{e^2}\right)^{\ell} \ell^{2\ell-2j} \left(1 + \mathcal{O}\left(\frac{j^2}{\ell} + \frac{1}{\sqrt{\ell}}\right)\right)$$

uniformly for $j = o(\sqrt{\ell})$. Plugging this into the first part gives:

$$\widetilde{\mathrm{SSN}}_\ell \sim \left(\sum_{j<\sqrt[4]{\ell}} \frac{1}{j! 16^j}\right) \frac{\sqrt{2\sqrt{e}}}{4} \ell^{-1} \left(\frac{8}{e^2}\right)^{\ell} \ell^{2\ell} \sim \frac{\sqrt{2\sqrt{e\sqrt[4]{e}}}}{4} \ell^{-1} \left(\frac{8}{e^2}\right)^{\ell} \ell^{2\ell}$$

which together with Eqn. 11 proves the claimed result. □