An Enumerative and Probabilistic Study of Various Phylogenetic Network Classes

多種系統發生網路類別的計數與機率研究

Yu-Sheng Chang 張裕昇

Supervised by: Michael Fuchs 符麥克

Compile date: December 31, 2025

Preface

This thesis, entitled ”An Enumerative and Probabilistic Study of Various Phylogenetic Network Classes”, is dedicated to advancing the mathematical understanding of phylogenetic networks. Phylogenetic networks are graphical models used in biology to represent complex evolutionary relationships among species or genes. Unlike traditional phylogenetic trees, networks are able to capture events such as hybridization, horizontal gene transfer, genome fusion, and recombination, making them useful in studying organisms like viruses, bacteria, and plants where such processes occur. This work abstracts these types of problems into the realm of graph theory.

The study primarily focuses on four main classes of phylogenetic networks: Tree-child (TC) networks, Galled (GN) networks, Reticulation-visible (RV) networks, and Galled Tree-child (GTC) networks, along with a generalization of tree-child networks to the d-combining case.

The methodologies employed rely on fundamental concepts from Graph Theory, the crucial graph-theoretic tool for enumeration, the Component Graph Method (CGM), and several Asymptotic Methods, including Generating Functions, Singularity Analysis, the Laplace Method, and the Lagrange Inversion Formula.

The research provides in-depth analyses across several key topics: determining sharp bounds for the maximal number of reticulations, deriving closed-form expressions for networks with a small number of reticulations k, counting maximally reticulated networks, conducting asymptotic counting for fixed k and for the total number of networks, and investigating probabilistic properties of these structures, such as the limit distribution of the number of reticulation nodes and the analysis of the Sackin Index. These results enhance the enumerative and probabilistic understanding of phylogenetic network classes.

前言

本論文題爲「An Enumerative and Probabilistic Study of Various Phylogenetic Network Classes」（各種系統發生網絡類別的枚舉與機率研究）。系統發生網絡 (Phylogenetic networks) 是生物學中使用的圖形模型，用於表示物種或基因之間複雜的演化關係。與傳統的系統發生樹 (Phylogenetic trees) 不同，系統發生網絡能夠捕捉雜交、水平基因轉移、基因組融合和重組等事件，因此在研究病毒、細菌和植物等複雜演化過程的生物體中非常有用。本研究將這些問題抽象化爲圖論領域的數學問題。

論文主要關注四個核心網絡類別：Tree-child (TC) 網絡、Galled (GN) 網絡、Reticulation-visible (RV) 網絡以及 Galled Tree-child (GTC) 網絡，同時也探討了TC網絡的 d-結合（d-combining）推廣。

研究所採用的主要方法學包括圖論的基礎概念、用於枚舉分析的組件圖方法 (Component Graph Method, CGM），以及多種漸近方法，例如生成函數、奇異點分析、拉普拉斯方法和拉格朗日反演公式。

本論文提供了對這些網絡類別的枚舉和性質的深入分析，內容涵蓋以下關鍵主題：確定網點 (reticulations) 數量 k 的最緊緻上界（sharp bound）、推導少量 k 的封閉形式 (closed form)、計算最大 k 的網絡數量、針對固定 k 的漸近計數，以及探討這些網絡的機率特性，包括隨機網絡中 k 的極限分佈 及 Sackin Index 等形狀參數。本研究期望透過這些結果，能加深對系統發生網絡結構和計數行爲的理解。

Contents

Chapter 1

Introduction

1.1 Graph Theory

This section establishes the foundational concepts and structural tools needed for the combinatorial analysis carried out in later chapters. We begin in Section 1.1.1 with the formal definition of the main object of study, phylogenetic networks, together with their basic components such as nodes, edges, and paths, and the related properties that will be used throughout the thesis. In Section 1.1.2, we introduce the particular classes of networks that form the focus of this work. We also review several related network families and describe the inclusion relationships among them, providing a structural map that situates our objects within the broader landscape of phylogenetic network models. Next, in Section 1.1.3, we present the central graph-theoretic tool used in the enumeration aspects of this thesis: the component-graph method. We explain its general principles and then discuss in detail how this method applies specifically to the network classes considered here. Finally, we turn to dup-trees, which are combinatorially equivalent to one-component galled networks and they arise naturally in many counting problems. Although their definition is simple, the corresponding counting question is significantly more complicated. Further refinements of dup-trees, such as trip-trees, are discussed in the Appendix A.

1.1.1 Basic Knowledge

Phylogenetic networks are graphical models used in biology to represent complex evolutionary relationships among species or genes. Unlike traditional phylogenetic trees, which depict only branching evolutionary paths, phylogenetic networks can capture events such as *hybridization*, *horizontal gene transfer*, *genome fusion*, and *recombination*. Thus, networks provide a more accurate representation of evolutionary history by allowing for reticulate patterns of inheritance. They are especially useful in studying organisms where such complex evolutionary processes occur, such as viruses, bacteria, and plants.

Abstracting these types of problems into mathematics falls within the realm of graph theory. Below, we define a phylogenetic network.

Definition 1.1.1. A (binary) phylogenetic network N with n leaves is a rooted, simple, directed acyclic graph (DAG) which is composed of 4 types of nodes:

Type	Number of in-coming edges	Number of out-going edges
root	0	1
leaves	1	0
tree nodes	1	2
reticulation nodes	2	1

where N contains one unique root and n leaves which are bijectively by $\{1, 2, \ldots, n\}$.

Figure 1.1: **From left to right**: root, leaf, reticulation node, and tree node.

Remark 1 (Potential ambiguities).

1. A phylogenetic network is *planted* if the out-degree of the root is one. In some articles, the root has out-degree two. In this thesis, we assume by default that phylogenetic networks are planted. Nevertheless, as there exists a (trivial) bijection between the two types of graphs, they can be interchanged freely.
2. A single labeled node can be regarded as an exceptional but valid network depending on the context.

Networks without any reticulation nodes are called phylogenetic trees.

Definition 1.1.2. A *phylogenetic tree* is a phylogenetic network without reticulation nodes.

A simple relationship between numbers of nodes is proved by counting the total in-degrees and out-degrees.

Proposition 1 (Fundamental equality)**.** For a phylogenetic network with n leaves, k reticulations, and t tree nodes, the relation between n, k and t is

$$n + k = t + 1. \tag{1.1}$$

Thus, the total number of nodes is

$$n + k + t + 1 = 2(n + k) = 2(t + 1).$$

For convenience of discussion, we provide the definitions of commonly used objects.

We start with two basic concepts:

Definition 1.1.3 (Edge and Path)**.** Let $e = (u, v)$ be a pair of nodes and $P = v_1v_2 \ldots v_n$ be a sequence of nodes in a DAG $G = (V, E)$. We say ...

1. e is an *edge* if (u, v) belongs to E.
2. P is a *(directed) path* if the (v_j, v_{j+1})'s are edges in E.

Next, we gives definitions of special types of nodes, edges, and paths, and other objects/relations.

Definition 1.1.4 (Nodes)**.** Let u and v be two nodes in a DAG G. We say/denote ...

1. v is *unary* if v has exactly one in-coming edge and one out-going edge.
2. v is *external* if v has no out-going edges.
3. v is *internal* if v has at least one out-going edge.
4. v is *visible* if there's a leaf ℓ such that every path from ρ to ℓ contains v.
5. v is *stable* if v is visible.
6. v is an *inner* reticulation if it is a reticulation and is not followed by a leaf.
7. v is an *outer* reticulation if it is a reticulation and is followed by a leaf.
8. v and u are *siblings* if they have a common parent.
9. v is *free* if it is a tree node and both its children are not reticulations.
10. u can reach v, $u \prec_G v$, if there is a path in G from u to v and $u \neq v$.
11. $u \preceq_N v$ if $u = v$ or $u \prec_N v$.
12. $\deg_{\text{in}}(v)$ counts the number of in-coming edges of v.

13. $\deg_{\text{out}}(v)$ counts the number of out-going edges of v.
14. $\deg(v)$ counts the number of edges incident to v.

Definition 1.1.5 (Edges)**.** Let $e = (u, v)$ be an edge in a network N. We say …

1. e is a *tree-edge* if v is a tree node.
2. e is a *reticulation-edge* if v is a reticulation node.
3. e is a *leaf-edge* if v is of a leaf.
4. e is a *shortcut* if there is another path from u to v.
5. e is *free* if it is an out-going edge of a free tree node.
6. e is a *bridge* if the deletion of e disconnect N.
7. A reticulation-edge e is *valid* if the sub-network N' obtained by deleting e from N and then appling the following 3 operations until none is applicable,
    - delete an unlabeled vertex of out-degree zero;
    - suppress an unary node;
    - replace a pair of parallel edges by a single edge.

    results in a network with exactly two nodes and three edges fewer than N.

Definition 1.1.6 (Paths)**.** Let $p = v_1v_2 \dots v_n$ and $q = u_1u_2 \dots u_m$ be two paths.

1. p is a *tree-path* if each edge v_iv_{i+1} is a tree-edge or leaf-edge.
2. p and q are *internally disjoint* if $v_i \neq u_j$ for $1 < i < n$ and $1 < j < m$.
3. v *has a tree-path to a leaf* if there's a tree-path p with $v_1 = v$ and v_n is a leaf.
4. A *tree cycle* is a pair of internally disjoint paths p and q from a common tree node to a common reticulation with all remaining nodes being tree nodes.

Definition 1.1.7 (Others)**.** Let v be a node, and $p = v_1v_2 \dots v_n$ be a path in a network $N = (V, E)$.

1. A *cherry* is the induced subgraph of N on two leaves and their common parent.
2. A *twin-cherry* is a cherry but the two leaves have repeated labels.

3. A *reticulated-cherry* is an induced subgraph of N on two leaves and their parents p_1, p_2 such that p_1 is a parent of p_2 and p_2 is a reticulation in N.
4. A *dup-tree* is a multi-labeled phylogenetic tree with each label used at most twice.
5. A *trip-tree* is a multi-labeled phy. trees with each label used once or three times.
6. A phylogenetic tree is *twin-cherry-free* if is contains no twin-cherry.
7. A *bridgeless component* is a maximal induced subgraph of N without bridges.
8. A *cluster* $c_N(v)$ of v is the vertex set $\{\ell \text{ is a leaf in } N : v \preceq_G \ell\}$.
9. A *leaf numbering function* f is a bijection from the leaves set to $[n]$.
10. A *characteristic vector* $C_f[v]$ for v under f is a bit vector of length n where the i-th bit equals 1 if and only if the cluster $c_N(v)$ contains $f^{-1}(i)$.
11. An *interval* is a maximal consecutive sequence of ones in a bit vector.
12. The *spread* I_f of f is defined as $\max_{v \in V} I_f(v)$, where $I_f(v)$ denotes the number of intervals in $C_f[v]$.
13. The *minimum spread* of N is the minimum value of I_f over all leaf numbering functions f.

Remark 2 (Potential ambiguities).

1. The root node is generally considered an internal node, unless it is the only node in the graph.
2. In many articles, the definition of tree-edge includes leaf-edge. In this thesis, they are different. We speak of *non-reticulation-edge* instead.

Many subsequent results simplify when using the notation of double factorials.

Definition 1.1.8.

$$n!! := \prod_{k=0}^{\lfloor n/2 \rfloor - 1} (n - 2k), \tag{1.2}$$

That is,

$$(2n)!! = (2n) \cdot (2n-2) \cdot (2n-4) \cdots 2 = 2^n\, n!,$$

and

$$(2n+1)!! = (2n+1) \cdot (2n-1) \cdot (2n-3) \cdots 1 = \frac{(2n+1)!}{2^n\, n!} = \frac{(2n+2)!}{2^{n+1}\,(n+1)!}.$$

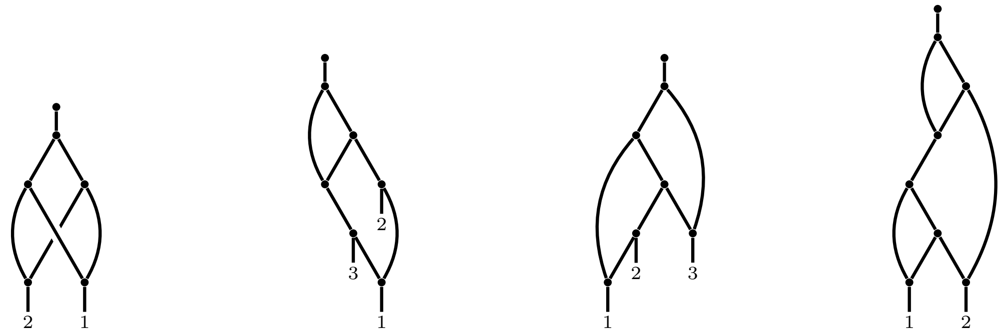
1.1.2 Classes of Networks

We introduce the four classes of phylogenetic networks that we mainly consider in this thesis and show the inclusion relation between them. From a biological perspective, reticulation events are often uncommon in evolutionary history. Consequently, many definitions of classes of phylogenetic networks are formulated with the aim of reducing the number of reticulation vertices.

Definition 1.1.9. A phylogenetic network is . . .

1. *tree-child* if each internal vertex has at least one non-reticulation child.
2. *galled* if each reticulation vertex is in a tree cycle.
3. *galled tree-child* if it is both galled and tree-child.
4. *reticulation-visible* if each reticulation vertex is visible.

For short, we call them **main classes** of phylogenetic networks. The four small networks shown in Figure 1.2 provide a representative overview of the defining features of each class.



$\in \mathcal{GN} \setminus \mathcal{TC}$ $\in \mathcal{TC} \setminus \mathcal{GN}$ $\in \mathcal{GTC}$ $\in \mathcal{RV} \setminus (\mathcal{TC} \cup \mathcal{GN})$

Figure 1.2: **Left 1**: A galled network which is not tree-child; **Left 2**: A tree-child network which is not galled; **Right 2**: A galled tree-child network; **Right 1**: A reticulation-visible network which is neither galled nor tree-child.

We directly observe the subset relationships between them.

Proposition 2. We have,

$$\mathcal{GTC} \subseteq \mathcal{TC}, \mathcal{GN} \qquad \text{and} \qquad \mathcal{TC}, \mathcal{GN} \subseteq \mathcal{RV}.$$

Proof. The first two relations are trivial by definition.

Let N be a tree-child network of size n. Let r be a reticulation in N. By the tree-child property, we recursively obtain a tree-path P to a leaf ℓ, say $P = rt_1t_2 \ldots t_m\ell$, where the t_j's are tree nodes.

If there is no intermediate tree node in P, then r is directly followed by the leaf ℓ, thus r is visible. Assume P contains at least one tree node. Now suppose there is a path P' from the root to the leaf ℓ without containing r. Then P' must go to ℓ via a vertex t_j in P which leads to a contradiction since t_j is a tree node. Hence, $\mathcal{TC} \subseteq \mathcal{RV}$.

Next, we prove $\mathcal{GN} \subseteq \mathcal{RV}$. Let r be a reticulation node in a galled network N. Denote by S the set of all nodes reachable from r. Because N is acyclic, this set S must contain at least one leaf ℓ. Next, suppose, for a contradiction, that there exists a path P from the root to ℓ that does not go through r. Then P must enter S through a reticulation $r' \in S$ without passing through r. However, this means that r' has no tree-cycle since one of the path back in direction of the root must pass through r and this violates the galled property. Thus, r is visible. ■

See the Venn diagram in Figure 1.3 for a visualization of the set inclusions from Proposition 2.

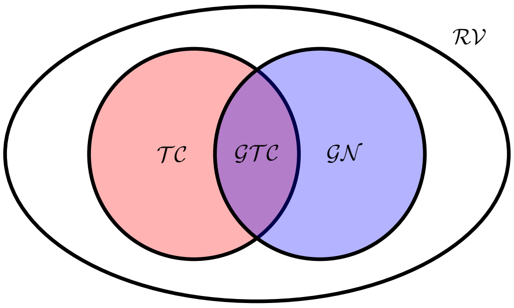


Figure 1.3: The simple Venn diagram for the four main classes

Abbreviations. The symbols, notations, and abbreviations used throughout the thesis are defined as follows. The calligraphic capitals $\mathcal{X}$ denote families of objects; $\mathcal{X}_n$ (resp. $\mathcal{X}_{n,k}$) denotes those with n leaves (resp. n leaves and k reticulations). The upright roman letters X denote the cardinality of such objects if it is finite. Table 1.1 shows the class names, abbreviations (mostly two capital letters), set notations, and cardinality notations. For a class of networks $\mathcal{X}$, the total number of networks denoted by X_n satisfies

$$\mathrm{X}_n = \sum_k \mathrm{X}_{n,k}.$$

An important type of network for the next sections is a one-component network which has all reticulations at the fringe, making the network more tree-like.

Definition 1.1.10 (One-component network)**.** A phylogenetic network is called a *one-component* network if every reticulation node has a leaf as its child.

For a class of networks $\mathcal{X}$, we use $\mathcal{OX}$ for its one-component version and $\mathrm{OX}_n, \mathrm{OX}_{n,k}$ for the cardinalities of $\mathcal{OX}_n, \mathcal{OX}_{n,k}$, respectively.

Abbr.	**Class Name**	**Set**	**Card.**
PN	Phylogenetic network	$\mathcal{PN}, \mathcal{PN}_{n,k}$	$\mathrm{PN}_{n,k}$
TC	Tree-child	$\mathcal{TC}, \mathcal{TC}_n, \mathcal{TC}_{n,k}$	$\mathrm{TC}_n, \mathrm{TC}_{n,k}$
GN	Galled	$\mathcal{GN}, \mathcal{GN}_n, \mathcal{GN}_{n,k}$	$\mathrm{GN}_n, \mathrm{GN}_{n,k}$
RV	Reticulation-visible	$\mathcal{RV}, \mathcal{RV}_n, \mathcal{RV}_{n,k}$	$\mathrm{RV}_n, \mathrm{RV}_{n,k}$
GTC	Galled Tree-child	$\mathcal{GTC}, \mathcal{GTC}_n, \mathcal{GTC}_{n,k}$	$\mathrm{GTC}_n, \mathrm{GTC}_{n,k}$

Table 1.1: The symbols, notations, and abbreviations for the main classes.

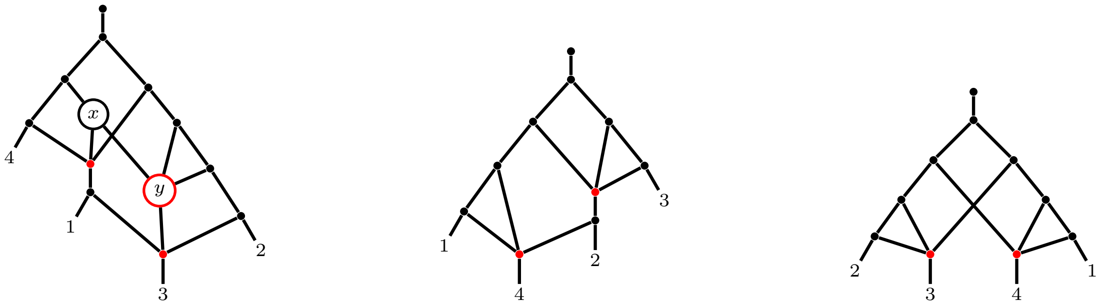


Figure 1.4: **Left**: A 3-combining phylogenetic network which is not a tree-child network (because both children of the tree node x are reticulation nodes and the only child of the reticulation node y is also a reticulation node); **Middle**: a 3-combining tree-child network; **Right**: a 3-combining one-component tree-child network.

Next, we introduce some generalizations of phylogenetic networks.

Definition 1.1.11 (d-combining and q-furcating)**.** A *d-combining* phylogenetic network is a network with reticulation nodes redefined as having in-degree d. Note that each reticulation has exactly d in-coming edges, not at most d ones.

Similarly, a *q-furcating* phylogenetic network is a network with tree nodes redefined as having out-degree q.

A d-combining phylogenetic network is *tree-child* if each internal vertex has at least one non-reticulation vertex in its set of children.

See Figure 1.4 for examples with $d = 3$.

Remark 3 (Possible differences of equivalent definitions in d-combining networks). For a class of networks, it may have different but equivalent definitions in the binary case. They may not hold if extending to d-combining or q-furcating networks, necessitating careful examination of possible discrepancies.

Other classes of networks. To make the content of the thesis self-contained, we list the definitions of other classes of networks which will be used below.

Definition 1.1.12 (Other classes of networks)**.** Let k be a natural number. A phylogenetic network N is . . .

1. *normal* if it is tree-child and has no shortcut edge.
2. *tree-sibling* if each reticulation has at least a tree node or a leaf as sibling.
3. *stack-free* if each reticulation has a non-reticulation node as its child.
4. *nearly stable* if for each node v, either it or its parents is visible.
5. *stable-child* if each node has a visible child.
6. *nearly tree-child* if each reticulation has at least one parent having a tree-path to a leaf.
7. *genetically stable* if each reticulation is visible and has at least one visible parent.
8. *level-k* if each bridgeless component of N contains at most k reticulations,
9. *galled-tree* if it is level-1.
10. *regular* if (1) $c_N(u) \neq c_N(v)$ for $u \neq v$, (2) $u \preceq_N v \iff c_N(u) \subseteq c_N(v)$, and (3) it has no shortcuts.
11. *orchard* or *cherry-picking* if there is a sequence of cherry reductions that reduces it into a single node, where a cherry reduction includes two kinds of operations on a network:
    (1) Given a cherry, delete one of its leaves and suppress the resulting unary node;
    (2) Given a reticulated cherry, delete the reti.-edge and suppress unary nodes.
12. *valid* if all reticulation-edges are valid.
13. *FU-stable* if it is (1) stack-free and (2) has the property that any two distinct tree vertices have distinct sets of children.
14. *spread-k* if its minimum spread is at most k.
15. *k-reticulated* if for any tree node or the root v, there are at most k reticulations that can be reached from v by at least two internally disjoint paths from v.

16. *tree-based* if there is a phylogenetic tree $\mathcal{T}$ of size n such that N can be obtained from $\mathcal{T}$ via the following steps:
    (1) Subdivide the edges of $\mathcal{T}$, i.e., attach any number of unary nodes on any edges;
    (2) Add edges between pairs of unary nodes by keeping it a DAG;
    (3) Suppress the remaining unary nodes.

For grammatical reasons, a hyphen is used to join two or more words into a compound adjective that modifies a noun (i.e. network). Also we give the abbreviations for these classes in Table 1.2. These abbreviations are not necessarily widely used, however, they can provide a reference for comparison or selection in future research. Note that Level-1 and galled-tree are equivalent. In addition, we prefer using cherry-picking instead of “O”rchard for avoiding the ambiguity with “O”ne-component networks.

Abbr.	**Class Name**
NN	Normal
TS	Tree-sibling
SF	Stack-free
NS	Nearly stable
SC	Stable-child
NTC	Nearly tree-child
GS	Genetically stable
L1	Level-1
Lk	Level-k

Abbr.	**Class Name**
GT	Galled-tree
RN	Regular
CP	Cherry-picking
VN	Valid
FU	FU-stable
Sk	Spread-k
kR	k-reticulated
TB	Tree-based

Table 1.2: The abbreviations for other classes.

We next give equivalent definitions for tree-child networks.

Proposition 3 (Equivalent definitions for tree-child networks)**.** Let N be a phylogenetic network. The following statements are equivalent:

1. N is tree-child (i.e. each internal node has at least one non-reticulation child);
2. Every node has a tree-path to a leaf;
3. N is stack-free and no reticulations share a common parent;
4. Each node is visible.

Proof. (1. → 2.) We construct the required tree-path for any node step-by-step by choosing a child which is a tree node or leave due to the hypothesis.

(2. → 3.) The existence of the tree-paths from the reticulations lead to the stack-free property and the existence of the tree-paths from the tree nodes give the second property.

(3. → 1.) Let v be an internal node in a stack-free network with no reticulations sharing a common parent. If v is a reticulation node, then by the stack-free property, its child must be a tree node or leaf; if v is a tree node, then it has at least one tree node or leaf as child, otherwise, the second property is violated; finally, it is clear that the child of the root cannot be a reticulation.

(1. → 4.) According to the Proposition 2, it suffices to show that every tree node is visible. Let t_1 be a tree node in a tree-child network and $P = t_1 t_2 t_3 \ldots t_{n-1} \ell$ be a corresponding tree-path to a leaf ℓ where the t_j's are tree nodes. Suppose, by contradiction, there's a path P' (from the root) to ℓ without passing through t_1. Then P' meets P at some node t_j. This leads to a contradiction since t_j is a tree node and its in-coming edge cannot belongs to P and P' at the same time. Hence, every path to ℓ contains t.

(4. → 3.) Let N be a network satisfying the condition 4. Suppose, by contradiction, that there is a reticulation r_1 followed by another reticulation r_2. By the hypothesis, there is a leaf $\ell = \ell(r_1)$ such that every path P to ℓ contains r_1. P must pass r_2 through the reticulation-edge $r_1 r_2$. Choose a path Q to r_2 passing through the other reticulation-edge of r_2. Then Q does not contains r_1, otherwise, the network is cyclic. Now, we attach the sub-path of P from r_2 to ℓ directly below Q. The resulting path is a path to ℓ that does not contain r_1, which leads to a contradiction. Therefore, N is stack-free.

On the other hand, suppose, by contradiction, that there is a tree node t followed by two reticulations r_1 and r_2. Let $\ell = \ell(t)$ be a leaf, by the hypothesis, such that every path to ℓ contains t. Let P be a path to ℓ containing t. P passes through either tr_1 or tr_2 and we assume it passes tr_1. Choose a path Q to r_1 passing through the other reticulation-edge of r_1. If Q does not pass through tr_2, then the path composed of Q and the sub-path of P from r_1 to ℓ leads to a contradiction; if Q does pass through tr_2, then choose a path R to r_2 passing though the other reticulation-edge of r_2. Concatenate R, the sub-path of Q from r_2 to r_1, and the sub-path of P from r_1 to ℓ. The resulting path also leads to a contradiction. Consequently, the second property holds. ■

We also call the second statement in Proposition 3 the tree-path property.

Subset-relation graph. Figure 1.5 in [33] shows the subset-relationship between all the above classes of phylogenetic networks. Consider the classes as points and the subset-relation as partial order, we may read the graph as the Hasse diagram of a poset (partially ordered set). The smallest element is the set of phylogenetic trees at the bottom and the largest element is the set of general phylogenetic networks at the top.

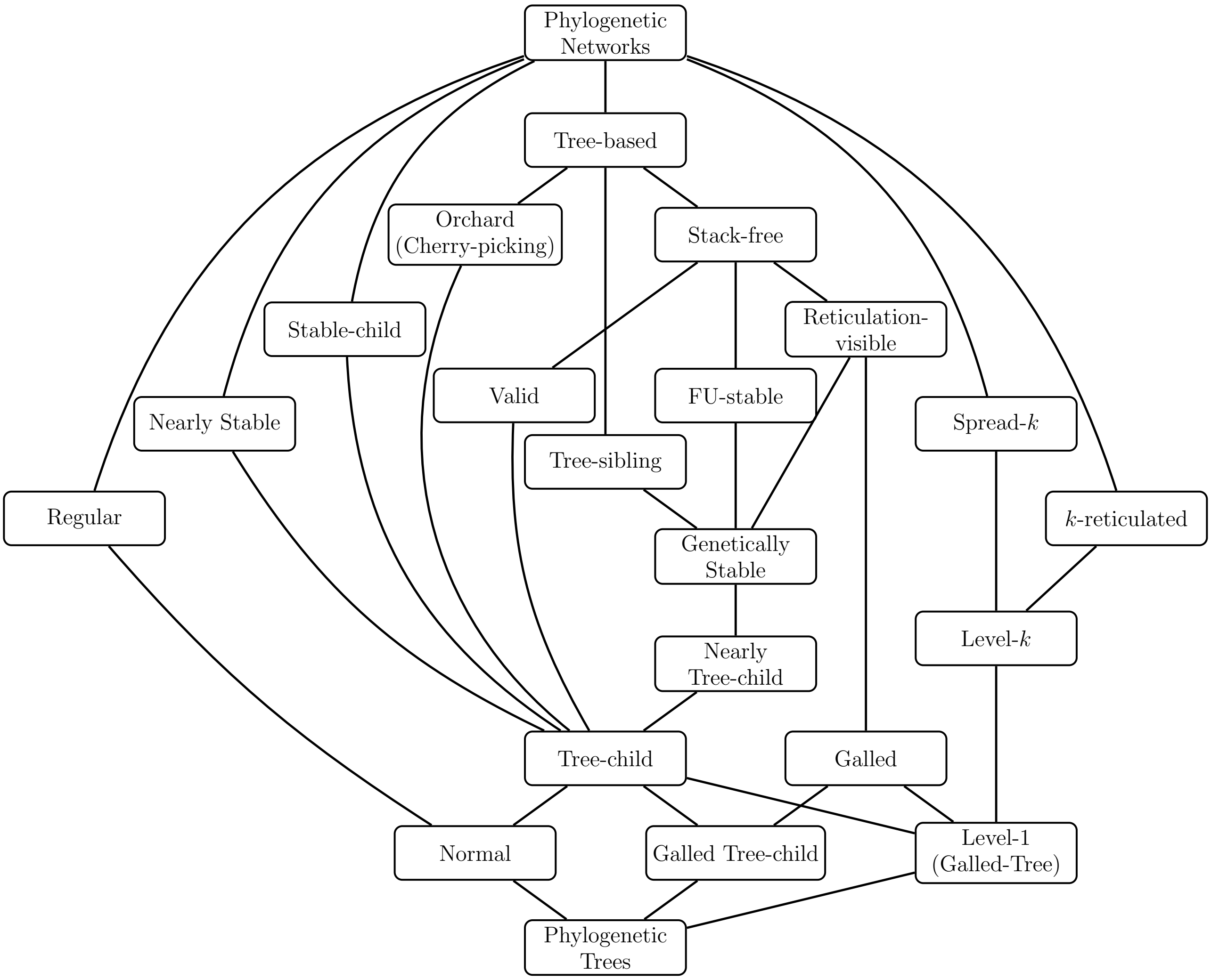


Figure 1.5: The subset-relationship among classes of phylogenetic networks.

Venn diagram for some classes. Apart from the simple Venn diagram for the main classes in Figure 1.3, Steel provided a Venn diagram for describing the relationships between more networks in 2016; see Figure 1.6 and [39].

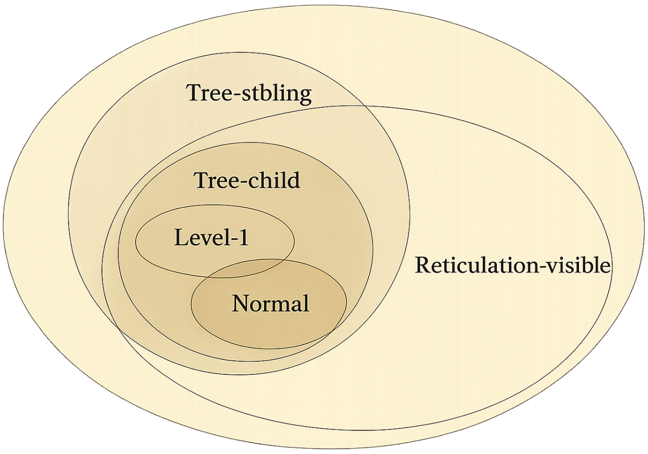


Figure 1.6: A Venn diagram for some classes

1.1.3 The Component Graph Method (CGM)

The component graph method (CGM) was introduced by Zhang in 2017. It was used by him and his co-authors to solve algorithmic problems for phylogenetic networks; see [28, 30, 9, 29]. For example, it was used to solve the algorithmic counting problem for tree-child networks in [12], galled networks in [24, 11], and reticulation-visible networks in [11].

The component graph method consists of the two following processes. First, the compression process is a de-information process, it maps the phylogenetic networks into simpler or smaller objects by loosing some information. The resulting graphs are called component graphs. And second, the de-compression process is a re-information process, it generates a bunch of phylogenetic networks from a component graph by plugging so-called components into its vertices and also requiring some appropriate combinatorial adjustments. Eventually, the component graph method gives not only an exact counting formulae, but also the construction process.

Definition 1.1.13 (A formal component method)**.** Let $\mathcal{X}$ collect the phylogenetic networks we intend to count and N be a phylogenetic network in $\mathcal{X}$. Let r_j be a reticulation node in N, r_0 be the root ρ and C be a function defined on $\mathcal{X}$. Define the *components* of N and the *component graph* of N as follows.

1. A *component* $\{c_j\}$ of N is a maximal connected subgraph of N, rooted at r_j, after doing *a series of combinatorial steps*, such as removing nodes, deleting/cutting edges, doing bijection or else. They are normally disjoint;
2. The *component graph* $C(N)$ of N is a DAG (V, E) where

$$V = \{c_j\}$$

is the set of components of N and E collects edges retaining *sufficient information* to reconstruct N.

We say a function C defined on $\mathcal{X}$ is a *component method* on $\mathcal{X}$ if:

1. The image graphs $C(\mathcal{X})$ are constructable;
2. The collection $C^{-1}(C(\mathcal{X}))$ is a partition of $\mathcal{X}$;
3. For each $Y \in C(\mathcal{X})$, the cardinality of $C^{-1}(Y)$ is determined.

Consequently, a component method gives the formulae,

$$|\mathcal{X}| = \sum_{Y \in C(\mathcal{X})} \left|C^{-1}(Y)\right|.$$

The method still varies due to the flexibility of defining the components and the component graph, which involves various minor adjustments depending on the different classes of phylogenetic networks.

We illustrate the differences in the component method by applying it to galled networks and tree-child networks. Here are some operations in graph theory we use in this paragraph.

- Delete a vertex: remove it and its adjacent edges;
- Suppress an unary node u: say the adjacent edges are $\gamma \to u \to \omega$. Delete u and add a new edge $\gamma \to \omega$;
- De-stem a reticulation node r_j: say the reticulation edges are $u_{j1} \to r_j$ and $u_{j2} \to r_j$. Remove (not delete) them, add two leaves n_{j1} and n_{j2} ***temporarily*** labeled by r_j and two edges $u_{j1} \to n_{j1}$ and $u_{j2} \to n_{j2}$.

We call the original labels *essential labels*, as opposed to the temporary ones. Accordingly, leaves with essential and temporary labels are referred to as *essential* and *temporary* leaves, respectively.

1.1.3.1 d-combining tree-child networks

We define the component graph method on d-combining phylogenetic networks by extending the definition for bi-combing phylogenetic networks. The idea comes from Zhang's work in [9] on (bi-combing) tree-child networks. Fortunately, we can directly apply the component graph method of bi-combining case without any adjustments to d-combining tree-child networks.

Example 1.1.1 (Component graph method for d-combining tree-child networks)**.** For a tree-child network N, the component graph $C(N)$ of N is obtained by the following steps:

1. Delete all reticulation edges and suppress all unary nodes.

   The resulting graph is a forest consisting of $k + 1$ (directed) trees (called *tree-components*) which are either rooted at the network root or at a reticulation node; see Figure 1.8, left and middle for $d = 2$ and Figure 1.7, left for $d = 3$;

2. Create vertices c_j labeled by j for the tree-component which contains the j-th smallest leaf-label. This defines the vertex set $\{c_j\}$ of $C(N)$;

   The set of leaf-labels of each tree-component partitions $\{1, \ldots, n\}$. The blocks of this partition can be indexed by the rank of the smallest element of the block; these indices are then given as label to the corresponding nodes of the component graph

3. Add m edges from c_i to c_j if there are m reticulation edges of r_j originating from c_i in N for $m = 1, 2, \ldots, d$. This defines the edge set of $C(N)$.

Figure 1.7: Left: The 3-combining network (with two reticulation nodes) from Figure 1.4, (a). The three tree-components when incoming edges of the reticulation nodes are removed are encircled in red, blue, and yellow. Right: The corresponding component graph.

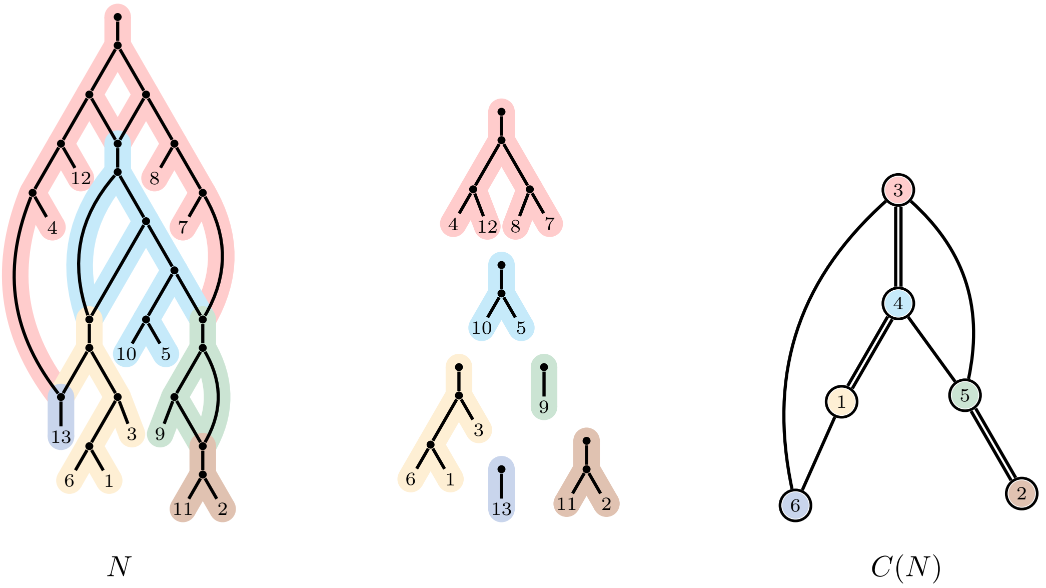


Figure 1.8: **Left**: A tree-child network with 13 leaves and 5 reticulation vertices having the components highlighted. **Middle**: The tree-components of the network. **Right**: The component graph of the network.

After performing the above combination steps on all d-combining tree-child networks, we obtain all component graphs. We can then describe the graphical structure of the component graph more generally.

Definition 1.1.14 (Component graphs for d-combining tree-child networks)**.** A component graph for a d-combining tree-child network is a vertex-labeled DAG such that the following properties hold:

1. Every non-root node has in-degree d;
2. The root has 0 in-degree;
3. Multi-edges between two nodes are allowed.

We denote by $\mathcal{K}_m^{(d)}$ the set of component graphs with m nodes and by $\mathcal{K}_{m,s}^{(d)}$ the set of component graphs with m nodes of which s are leaves. Set $\mathrm{K}_m^{(d)} = |\mathcal{K}_m^{(d)}|$ and $\mathrm{K}_{m,s}^{(d)} = |\mathcal{K}_{m,s}^{(d)}|$. Then,

$$\mathrm{K}_m^{(d)} = \sum_{s=1}^{m-1} \mathrm{K}_{m,s}^{(d)},$$

since the number of leaves of a component graph is at least 1 and at most $m-1$. (For the star-component graph; see Figure 2.10).

This, together with the following recursive formula for $\mathrm{K}_{m,s}^{(d)}$, makes it possible to compute $\mathrm{K}_m^{(d)}$. (See [9, Theorem 15] for the bi-combining case.)

Theorem 1 (Counting the component graphs of d-combining tree-child networks)**.** For $m \geq 2$,

$$\mathrm{K}_{m,s}^{(d)} = \sum_{t=1}^{m-s-1} \binom{m}{s} \beta^{(d)}(m,s,t) \mathrm{K}_{m-s,t}^{(d)}, \quad (1 \leq s \leq m-1),$$

with initial $\mathrm{K}_{1,1}^{(d)} = 1$ and

$$\beta^{(d)}(m,s,t) = \sum_{n=0}^{t} (-1)^n \binom{t}{n} \binom{(m-s-n)+d-1}{d}^s.$$

Proof. The recurrence can be obtained by the following way of constructing all component graphs in $\mathcal{K}_{m,s}^{(d)}$ from those in $\mathcal{K}_{m-s,t}^{(d)}$:

(i) Choose t with $1 \leq t \leq m-s-1$ and a graph G in $\mathcal{K}_{m-s,t}^{(d)}$;

(ii) Add s new nodes, labelled by $\{1', \cdots, s'\}$, to G such that (a) these nodes become the new leaves and (b) all old leaves have at least one out-going edge (i.e., all of them become internal nodes). By the inclusion–exclusion principle, there are

$$\beta^{(d)}(m,s,t) = \sum_{0 \leq n \leq t} (-1)^n \binom{t}{n} \binom{m-s-n+d-1}{d}^s$$

ways of choosing the d incoming edges for the new leaves; here, the counting is done such that n of the old leaves are not used as parents of the new leaves;

(iii) Choose s leaves from the set of labels $\{1, \ldots, m\}$ and use them to re-label the new leaves; use the remaining labels to re-label the remaining nodes in an order-consistent way. ■

For $d = 2$, the result is the same as [9, Theorem 15].

On the other hand, the components of d-combining tree-child networks can be defined as well.

Definition 1.1.15 (Components for d-combining tree-child networks)**.** A component for a d-combining tree-child network is a tree-component, which is simply a phylogenetic tree.

The following exact formulae for d-combining cases was first derived by Cardona and Zhang in 2020, who discussed bi-combining case, with some ambiguities fixed.

Theorem 2 ([9])**.** Let $\prod_{n,k+1}$ be the set of partitions of $[n]$ into $k+1$ blocks $\{B_i\}_{i=1}^{k+1}$ such that $\langle \min B_i \rangle_i$ is increasing. Then

$$\mathrm{TC}_{n,k}^{(d)} = \frac{1}{2^{n-k-1}} \sum_{\{B_i\}_{i=1}^{k+1} \in \prod_{n,k+1}} \sum_{G \in \mathcal{K}_{k+1}^{(d)}} \prod_{j=1}^{k+1} \frac{(2b_j + g_j - 2)!}{(b_j - 1)! \prod_{\ell=1}^{k+1} g_{j,\ell}} \tag{1.3}$$

where $b_j = |B_j|$ for $1 \leq j \leq k+1$, $g_{j,\ell}$ is the number of edges in G which are directed from node j to node ℓ, and $g_j = \sum_\ell g_{j,\ell}$ is the out-degree of node j for $1 \leq j \leq k+1$.

Proof. By the description preceding the theorem, the formula is explained as follows:

(i) G is the chosen graph in $\mathcal{K}_{k+1}^{(d)}$ and $\{B_j\}_{j=1}^{k+1}$ is the chosen partition in $\prod_{n,k+1}$.

(ii) The number of possible phylogenetic trees assigned to the nodes of G is:

$$\prod_{j=1}^{k+1} (2b_j - 3)!! = \prod_{j=1}^{k+1} \frac{(2b_j - 2)!}{2^{b_j - 1}(b_j - 1)!}, \tag{1.4}$$

where we assume that b_j is the size of the block belonging to node j in G.

(iii) The number of ways of adding nodes to the phylogenetic tree of node j is:

$$(2b_j - 1) \cdots (2b_j - 1 + g_j - 1)$$

(iv) Connecting the nodes which have been added to the phylogenetic tree of node j to the root of the phylogenetic tree of node ℓ gives every tree-child networks $(g_{j,\ell})!$ times. Thus, the number of tree-child networks arising from a fixed choice of G, $\{B_j\}_{j=1}^{k+1}$ and a set of phylogenetic trees (whose number of leaves equals to b_j for $1 \leq j \leq k+1$) is:

$$\prod_{j=1}^{k+1} \frac{(2b_j - 1) \cdots (2b_j - 1 + g_j - 1)}{\prod_{\ell=1}^{k+1} (g_{j,\ell})!} \tag{1.5}$$

(v) Finally, multiplying (1.4) and (1.5) gives

$$\prod_{j=1}^{k+1}\frac{(2b_j-2)!}{2^{b_j-1}(b_j-1)!}\frac{(2b_j-1)\cdots(2b_j-1+g_j-1)}{\prod_{\ell=1}^{k+1}(g_{j,\ell})!}=\frac{1}{2^{(\sum_j b_j)-k-1}}\prod_{j=1}^{k+1}\frac{(2b_j+g_j-2)!}{(b_j-1)!\prod_{\ell=1}^{k+1}(g_{j,\ell})!}$$
$$=\frac{1}{2^{n-k-1}}\prod_{j=1}^{k+1}\frac{(2b_j+g_j-2)!}{(b_j-1)!\prod_{\ell=1}^{k+1}(g_{j,\ell})!},$$

which is the claimed factor (in front of and inside) of the double sum in (1.3). ■

1.1.3.2 Galled networks

In this subsection, we discuss the component graph method for galled networks. The components and the component graph are very different from the ones of d-combining tree-child networks.

Example 1.1.2. For a galled network N, the component graph $C(N)$ of N is obtained by the following steps:

1. De-stem each reticulation node which gives the components c_j of N;
2. For each component c_j with essential leaves ℓ_{jk}, create a node v_j, map c_j into v_j, duplicate ℓ_{jk} and connect them to v_j;
3. Add a black-arrowed edge from v_i to v_j if both reticulation edges of c_j come from c_i in N; see the middle graph in Figure 1.9;

   An edge without attaching a black arrow from v_i to v_j means that there is only one of the reticulation edges of c_j coming from c_i. Due to the galled property, the reticulation edges of c_j cannot originate from different c_i's, hence such a situation will not happen.
4. Suppress unary nodes $v_{j'}$ by retaining the black arrow and the leaf label.

Denote by $\tilde{C}(N)$ the component graph of $C(N)$ with all arrows on edges removed. We observe that $\tilde{C}(N)$ is a (not necessarily binary) phylogenetic tree.

Theorem 3 ([30]). Let N be a phylogenetic network. Then if N is galled then $\tilde{C}(N)$ is a (not necessarily binary) phylogenetic tree. Conversely, all trees occur as $\tilde{C}(N)$ for some galled network N.

See Figure 1.9 for a galled network and its component graph. Note that in the transitional graph, an unary node happens only when a reticulation is directly followed by a leaf. This explains why Step 4 in Example 1.1.2 is necessary and why the component graph is a general phylogenetic tree.

The following definitions directly illustrate the component graphs and components.

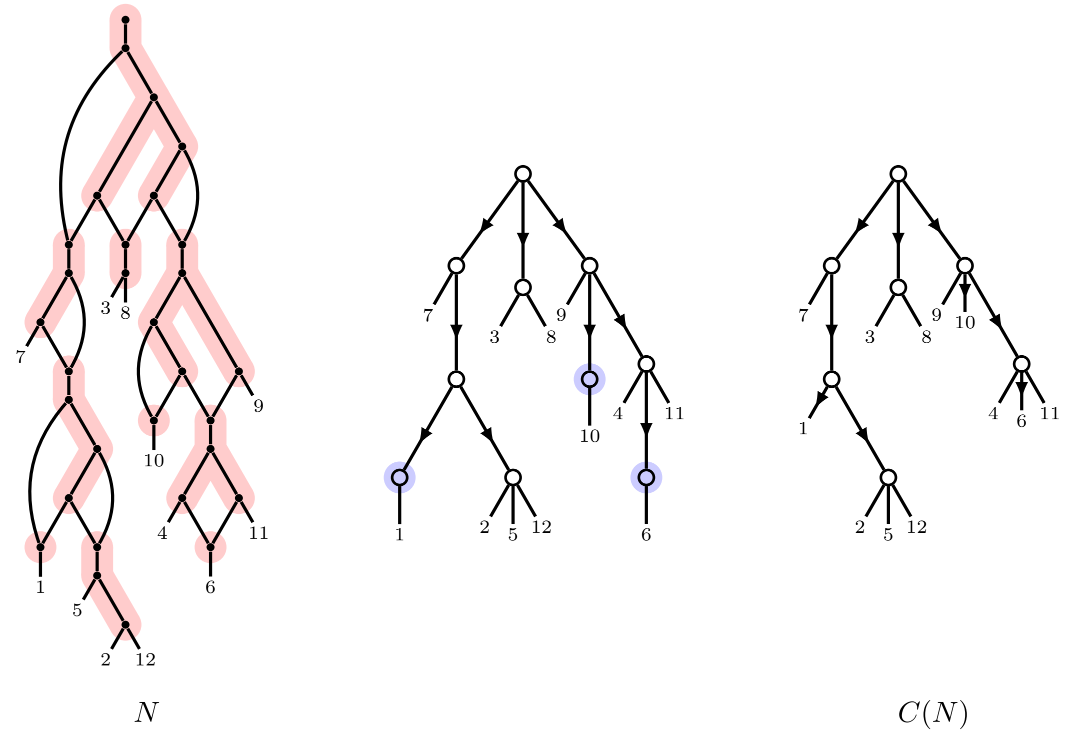


Figure 1.9: **Left**: A galled network with 12 leaves and 9 reticulation vertices having the components highlighted with red background. **Middle**: The transitionary graph having the specific substructures highlighted with blue background. **Right**: The component graph of the network. Note that it is a (not necessarily binary) phylogenetic tree with some black arrows on its edges.

Definition 1.1.16 (Component graphs for galled networks)**.** A component graph for a galled network is a phylogenetic tree such that the following properties hold:

1. Each non-leaf-edge must have a black arrow on it;
2. Each leaf-edge may or may not have a black arrow on it.

We denote the set of such phylogenetic trees by $\mathcal{T}$ and $\mathcal{T}_n$ for ones of size n.

Definition 1.1.17 (Components for galled networks)**.** Let N be a galled network. A component c of N is an one-component galled network.

One may notice that the components in Example 1.1.2 and one-component galled networks are not equal mathematical objects, but they are bijectively equivalent.

Remark 4. The components in Example 1.1.2 are bijectively equivalent to one-component galled networks.

Proof. Let c be a component in Example 1.1.2 with n essential leaves and k pairs of temporary leaves ($n+2k$ leaves in total). For each pair of temporary leaves ℓ_1 and ℓ_2 with label x, delete them, connect their parents to a new node which is followed by a new leaf labeled by x. We have a one-component galled network with $n+k$ leaves and k reticulations. ■

One-component galled networks were counted in [29]. Denote by $\mathrm{OGN}_{n,k}$ the number of one-component galled networks with n leaves and k reticulation vertices, by $M_{n,k}$ the number of one-component galled networks whose leaves below the reticulation vertices are labeled by $\{1,\ldots,k\}$, and by $\mathcal{M}_{n,k}$ the set which collects such networks with n leaves and k reticulations. Then, the following result was proved in [29]. (Note that part (i) is trivial.)

Proposition 4 ([29]).

1. We have,

$$\mathrm{OGN}_{n,k} = \binom{n}{k} M_{n,k}.$$

2. For $2 \le k \le n$,

$$\begin{aligned} M_{n,k} &= (n+k-2)M_{n,k-1} + (k-1)M_{n,k-2} \\ &\quad + \frac{1}{2}\sum_{1\le d\le k-1}\binom{k-1}{d}(2d-1)!!\left(M_{n-d,k-1-d} - M_{n+1-d,k-1-d}\right), \end{aligned} \tag{1.6}$$

initial values $M_{n,0} = (2n-3)!!$ and $M_{n,1} = (n-1)(2n-3)!!$.

The proof of 2. will be shown later in Section 1.1.4.

By performing the above combinatorial steps in reverse in Example 1.1.2, we get an exact formula for counting galled networks with n leaves.

Theorem 4 ([29]). We have,

$$\mathrm{GN}_n = \sum_{\mathcal{T}}\prod_{v}\sum_{j=0}^{\deg_{\mathrm{lf}}(v)}\binom{\deg_{\mathrm{lf}}(v)}{j} M_{\deg_{\mathrm{out}}(v),\deg_{\mathrm{out}}(v)-\deg_{\mathrm{lf}}(v)+j},$$

where

1. The first sum runs over all (not necessarily binary) phylogenetic trees $\mathcal{T}$ of size n (see Definition 1.1.16);
2. The product runs over all internal nodes of $\mathcal{T}$;
3. $\deg_{\mathrm{lf}}(v)$ is the number of children of v which are leaves;
4. $M_{n,k}$ 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 $\{1,\ldots,k\}$.

In order to better understand the decompression steps, we provide a step-by-step example in Figure 1.10. Therefore, fix the component graph of a galled network. In order to decompress it, first pick a one-component galled network N^o for the root r of the component graph (the network over the first arrow in the example), where the number of leaves, say ℓ_r, of N^o is the out-degree of r and the number of reticulation vertices, say k_r, is the number of arrows on the outgoing edges of r. ($\ell_r = k_r = 3$ in the example.) Moreover, let the leaves below the reticulation vertices in N^o have labels $\{1, \ldots, k_r\}$. Next, remove r from the component graph which gives a forest of ℓ_r trees. (The subtrees rooted at B, C, and D in the example.) The trees which have been attached by edges with arrows to r (all trees in the example) go to the leaves below reticulation vertices in N^o and the remaining trees (which consist of just one labeled vertex) are used the relabel the remaining leaves of N^o in an order consistent way (i.e., the smallest goes to $k_r + 1$, the second smallest to $k_r + 2$, etc.). Moreover, for the trees which go to the leaves below reticulation vertices, the order in which they are attached is the increasing order of their smallest leaf label (i.e., the one with the smallest leaf label - the subtree rooted at B in the example as the smallest label of this subtree is 1 - goes to 1, the one with the second smallest leaf label - the subtree rooted at C in the example as the smallest label of this subtree is 3 - goes to 2, etc.). Then, continue with the non-leave vertices of the trees in a recursive way. This gives, by picking all possible one-component networks in each step, all possible galled networks whose component graph is the given component graph.

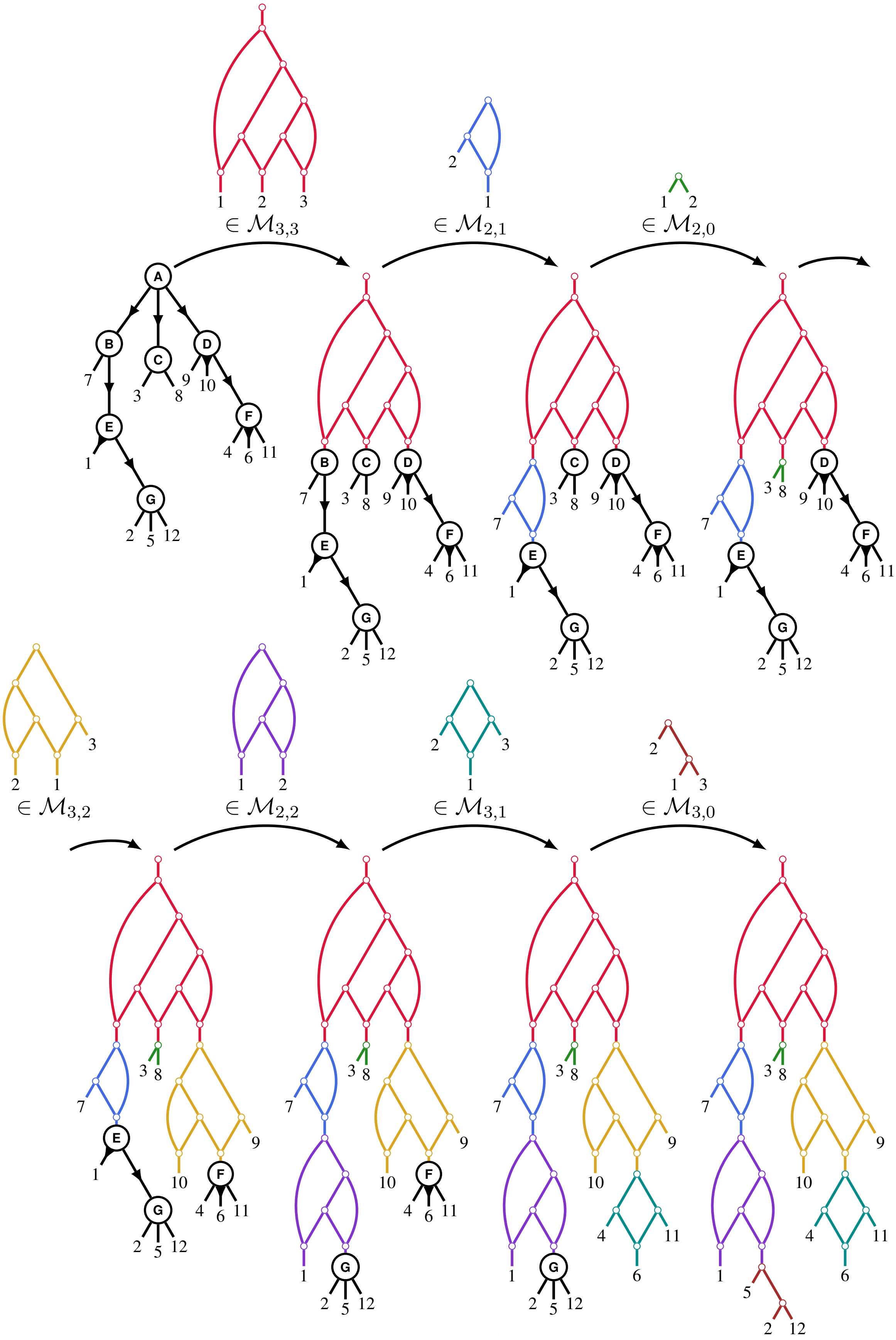


Figure 1.10: Step-by-step decompression of the component graph from Figure 1.9; the nodes are processed in the order indicated and the one-component networks replacing nodes are above the arrows. Note that except for the first of these networks, the root edge has to be removed in all others.

1.1.3.3 Reticulation-visible networks

We use the same combinatorial process which was used for galled networks also for reticulation-visible networks; see Figures 1.11 for a reticulation-visible network and its component graph.

Theorem 5 ([30])**.** Let N be a phylogenetic network. If N is reticulation-visible then $\tilde{C}(N)$ is a (not necessarily binary) tree-child network with all vertices of in-degree at most 2 and no reticulation vertex is just followed by one tree vertex. Conversely, all such tree-child networks occur as $\tilde{C}(N)$ of a reticulation-visible network N

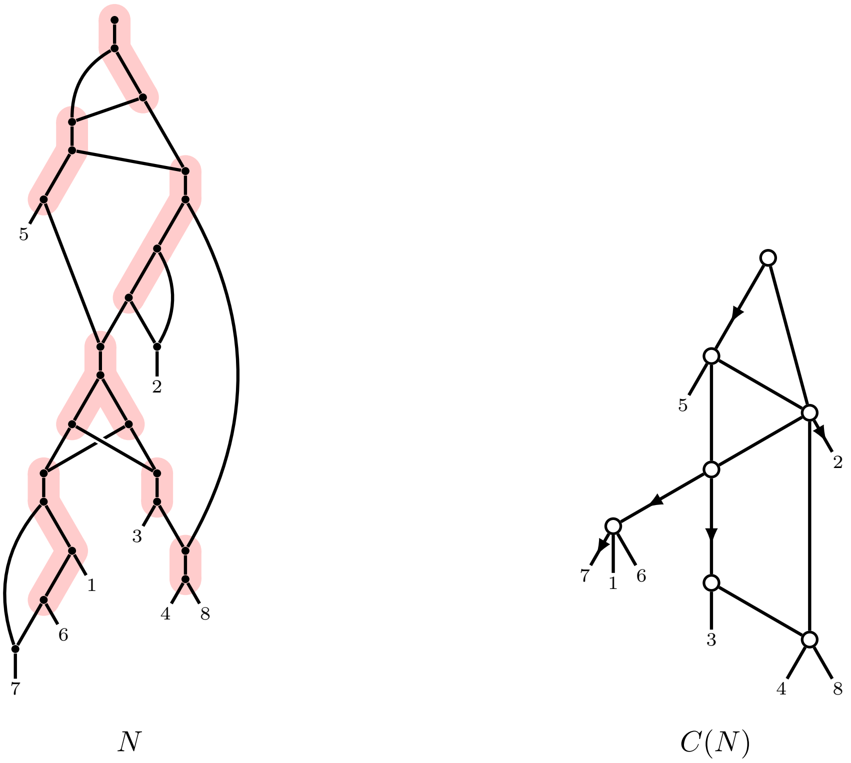


Figure 1.11: **Left**: A reticulation-visible network with 8 leaves and 8 reticulation vertices (with the components highlighted); **Right**: The component graph of the network. Note that it is tree-child network.

We next again define the component graphs and components.

Definition 1.1.18 (Component graphs for reticulation-visible networks)**.** A component graph of a reticulation-visible network is a general tree-child network such that:

1. Every non-root internal node has either in-degree 2 or in-degree 1 with a black arrow on its in-coming edge(s);
2. No internal node has only one child.

Let $\hat{\mathcal{TC}}$ collect all such tree-child networks.

Definition 1.1.19 (Components for reticulation-visible networks)**.** A component of a reticulation-visible network is one-component reticulation-visible network.

Note that every reticulation in a one-component network must be visible by the leaf it is followed by, hence they are one-component reticulation-visible networks. Thus, the set of one-component networks and the set of one-component reticulation-visible networks coincides. Likewise, the set of one-component reticulation visible networks and the set of one-component galled networks also coincide.

Lemma 6. The set of one-component reticulation-visible networks is the same as the set of one-component galled networks.

The following proof shows that more classes of phylogenetic networks have the property that their sets of one-component networks equal to one-component galled networks.

Proof. Let N be an one-component phylogenetic network. Since every reticulation is directed followed by a leaf, it is clearly contained in a tree cycle. That means N is also galled which shows that $\mathcal{OPN} \subseteq \mathcal{OGN}$.

On the other hand, there is a subset relation chain in Figure 1.5 from galled networks to phylogenetic networks:

$$\mathcal{GN} \subseteq \mathcal{RV} \subseteq \mathcal{SF} \subseteq \mathcal{TB} \subseteq \mathcal{PN}$$

and thus the same chain also holds for the corresponding classes of one-component networks. Consequently, we can concludes that

$$\mathcal{OGN} = \mathcal{ORV} = \mathcal{OSF} = \mathcal{OTB} = \mathcal{OPN}. \quad \blacksquare$$

As a consequent, after doing decompression on the components graph of reticulation-visible networks, we have the exact counting formula.

Theorem 7. We have,

$$\mathrm{RV}_n = \sum_{\hat{\mathcal{TC}}} \prod_{v} \sum_{j=0}^{\deg_{\mathrm{lf}}(v)} \binom{\deg_{\mathrm{lf}}(v)}{j} M_{\deg_{\mathrm{out}}(v), c_1(v)+j},$$

where

1. the first sum runs over all (not necessarily binary) tree-child networks $\hat{\mathcal{TC}}$,
2. the product runs over all internal nodes of the tree-child network, and
3. $c_1(v)$ is the number of children of v which are note leaves and have in-degree 1.

We again explain the decompression process. Therefore, consider the component graph of a reticulation-visible network (which by the above result is a particular tree-child network; also, assume that there are arrows on the edges); see Figure 1.12 for an example. The decompressing works as for a galled networks, with the only difference of how the children of a node are attached to the leaves of the one-component network which replaces the node. More precisely, in which order should this be done? (Again, the children attached with edges having arrows go to the leaves below reticulation vertices, the others not.) Here, the tree-child property comes into play. It implies that that for every vertex, there exists a set of leaves which can be only reached from this vertex. All children of a node have such a set and we can order these sets (and consequently the children) according to their smallest elements. This is then used to attach the children of a node to the leaves of the one-component network which replaces the node.

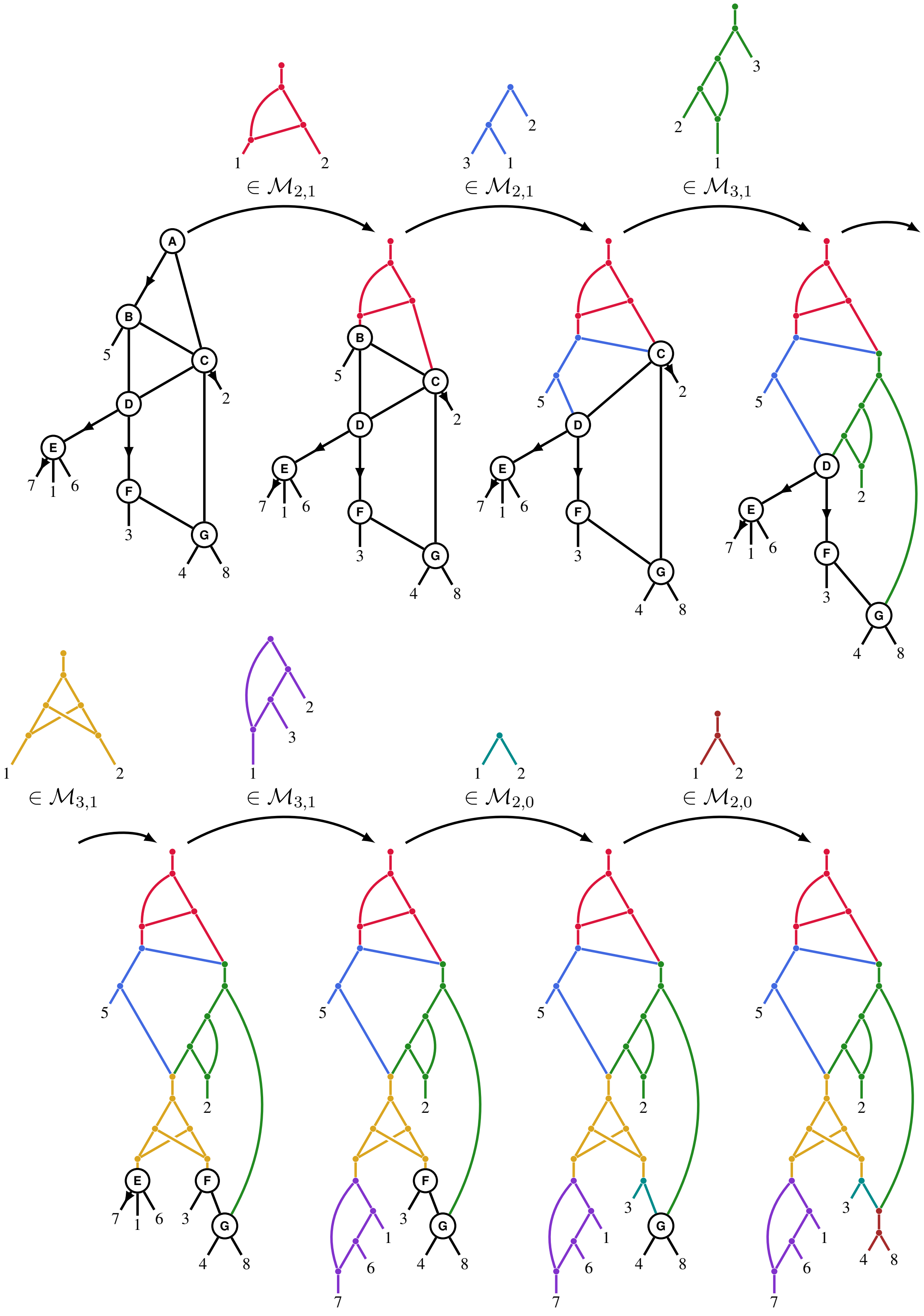


Figure 1.12: Step-by-step decompression of the component graph from Figure 1.11; again the nodes are processed in the indicated order and the one-component networks are above the arrows. Here, the root edges of the latter networks have to be removed if and only if the replaced nodes in the component graph have indegree 1 (and thus the incoming edge has an arrow on it).

1.1.3.4 Galled tree-child networks

For galled tree-child networks, it is now clear that the same formula holds as in Theorem 4 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.

Definition 1.1.20 (Component graphs for galled tree-child)**.** A component graph of a galled tree-child network is a general phylogenetic tree such that:

1. Each non-leaf-edge must have a black arrow on it;
2. Each leaf-edge may or may not have a black arrow on it;
3. Each internal node has at least one out-going edge without a black arrow.

Note that to avoid that all component graphs appear as $C(N)$ of galled tree-child networks N, we have additionally added the third statement compared to the Definition 1.1.16 for galled networks; however, this property is not needed below; see Remark 5.

Lemma 8. 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. ■

Definition 1.1.21 (Components for galled tree-child networks)**.** A component of a galled tree-child network is one-component tree-child network.

Denote by $\mathcal{L}_{n,k}$ the set of one-component tree-child networks of size n and k reticulation vertices, where the labels of the leaves below the reticulation vertices are $\{1, \dots, k\}$, and $\mathrm{L}_{n,k}$ the cardinality of $\mathcal{L}_{n,k}$.

Theorem 9. 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 $\{1, \dots, k\}$ for $0 \leq k \leq n-1$ is given by

$$L_{n,k} = \frac{(2n-2)!}{2^{n-1}(n-k-1)!}.$$

In later chapters, we will extend $L_{n,k}$ to d-combining ones.

Then, we have the following analogous result to Theorem 4.

Theorem 10. We have,

$$\mathrm{GTC}_n = \sum_{\mathcal{T}} \prod_{v} \sum_{j=0}^{c_{\mathrm{lf}}(v)} \binom{c_{\mathrm{lf}}(v)}{j} L_{c(v),c(v)-c_{\mathrm{lf}}(v)+j}, \tag{1.7}$$

where notation is as in Theorem 4 and $L_{n,k}$ was defined above.

Remark 5. The third statement in Definition 1.1.20 can be ignored because for an internal node v with all out-going edges having arrows, there is no corresponding one-component tree-child networks since $L_{\deg_{\mathrm{out}} v, \deg_{\mathrm{out}} v} = 0$.

Remark 6. Using the above 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 1.3: The values of GTC_n for $1 \leq n \leq 10$.

1.1.4 Dup-trees

In this subsection, we discuss the counting of one-component galled networks. An important and related counting problem introduced in [29] is to count twin-cherry-free dup-trees with multi-labels $[n] \uplus [i]$ for $1 \leq i \leq n$, where twin-cherry-free means that the dup-tree has no cherry whose leaves have identical labels. According to Remark 4, counting one-component galled networks is equivalent to counting the components of galled networks, which are exactly the twin-cherry-free dup-trees.

The counting of dup-trees will be done by unrooting them and then exploring a multi-to-multi relation. The method can be generalized to count *trip-trees* which directly correspond to one-component 3-combining networks; see Appendix A.

The following proposition is to count the number of un-rooted dup-trees labeled by $[n] \uplus [i+1]$; the corresponding result for rooted dup-trees will follow from it; see part (ii) of Proposition 4.

Proposition 5 ([29])**.**

$$\begin{aligned}\mathrm{UD}_{n,i+1} = (n+i-2)\,\mathrm{UD}_{n,i} + i\,\mathrm{UD}_{n,i-1} + \\ \frac{1}{2}\sum_{1\leq d\leq i}\binom{i}{d}(2d-1)!!\,\left(\mathrm{UD}_{n-d,i-d} - \mathrm{UD}_{n-d+1,i-d}\right), \quad (1.8)\end{aligned}$$

for $0 \leq i < n$ with initials $\mathrm{UD}_{n,0} = (2n-5)!!$.

Counting dup-trees. We need the following sets:

1. $\mathcal{T}_n$ ($\mathcal{UT}_n$) collects the (un-rooted) phylogenetic trees with n leaves;
2. $\mathcal{D}_{n,i}$ ($\mathcal{UD}_{n,i}$) collects the (un-rooted) dup-trees with labels $[n] \uplus [i]$ which are twin-cherry-free;
3. $\mathcal{UD}^{(1)}_{n,i}$ collects the un-rooted dup-trees with labels $[n] \uplus [i]$ with exactly one twin-cherry,

for $0 \leq i \leq n$ where $\uplus$ is the disjoint union. When $i = 0$, there are no duplicated labels. Thus,

$$\mathrm{D}_{n,0} = \mathrm{T}_n \quad \text{and} \quad \mathrm{UD}_{n,0} = \mathrm{UT}_n.$$

Moreover, by rooting at a leaf labeled by $n+1$, we obtain

$$\mathrm{T}_n = \mathrm{UT}_{n+1} = (2n-3)!! \quad \text{and} \quad \mathrm{D}_{n,i} = \mathrm{UD}_{n+1,i}.$$

Our main idea is to construct the set $\mathcal{UD}_{n,i+1}$ by attaching a leaf labeled by $i+1$ to the edges of an un-rooted dup-tree in $\mathcal{UD}_{n,i}$, for $0 \leq i \leq n-1$. More precisely, we will perform the following steps:

1. Compress the dup-trees in $\mathcal{UD}_{n,i+1}$. (Define b-edge and red dot.)

2. Classify them into a partition $\{\mathcal{O}_{m,j}\}$.

3. Remove a red dot from $\mathcal{O}_{m,j}$ (call $\mathcal{P}^{(k)}_{m,j}$) and then add a red dot.

4. Establish the formula between $\mathcal{UD}_{n,i}$ and $\mathcal{UD}_{n,i+1}$.

5. Compute the remaining terms.

First, compress the dup-trees in $\mathcal{UD}_{n,i+1}$ as follows. Let T be a dup-tree in $\mathcal{UD}_{n,i+1}$. Condense the leaves ℓ labeled by $i+1$ into single points by deleting ℓ, coloring its parents p red and p's incident edges blue; see Figure 1.13. For short, we call blue edges, which carry one or two *red dots* that represent leaves labeled by $i+1$, a *b-edge*. The result tree will still be called T.

Figure 1.13: Left: a b-edge containing two red dots. Right: a b-edge containing one red dot.

Second, classify T as follows:

1. The number j of b-edges where $j = 1, 2$.

2. The number n_0 of the symmetric structures involving b-edges where $n_0 = 0, 1, \ldots, j$. Draw a cherry followed by two triangles to represent a symmetric structure; a circle represents the remaining part.

3. The distribution of the j b-edges among one circle and $2\,n_0$ triangles such that each cherry contains at least one b-edge and the emptiness of the circle is allowed. (Note that the b-edges are identical if they contains the same number of red dots. And every pair of triangles is considered unordered.)

4. In one cherry, if both triangles contain one b-edge, then there are two cases, depending on whether the two b-edges are located at the ***same positions*** in each triangle or not. It yes, draw a curved double arrow between the two blue edges; if no, the b-edges are not plotted exactly at the same position in the triangle.

Overall, this gives 8 different sets $\mathcal{O}_{m,j}$ to which T belongs. $\{\mathcal{O}_{m,j}\}$ forms a partition of $\mathcal{UD}_{n,i+1}$ and thus,

$$\mathrm{UD}_{n,i+1} = \sum_{m,j} \mathrm{O}_{m,j}.$$

Representative graphs for each $\mathcal{O}_{m,j}$ are shown in Figure 1.14.

Third, remove a red dot from $\mathcal{O}_{m,j}$ and then find its multiplicity number m by adding back a red dot. More precisely, for a representative graph of $\mathcal{O}_{m,j}$, choose a red dot (there are two choices) and remove it. The resulting graphs $P^{(k)}_{m,j}$ may or may not be identical; see Figure 1.14. (They are identical for $O_{2,1}$ and $O_{1,1}$, respectively.) Next, for each $\mathcal{P}^{(k)}_{m,j}$, count the number of edges in $\mathcal{P}^{(k)}_{m,j}$ which yield $\mathcal{O}_{m,j}$ by attaching a red dot on it. We call the number of such edges the *multiplicity* of $\mathcal{P}^{(k)}_{m,j}$ which is shown right next to $\mathcal{P}^{(k)}_{m,j}$ in Figure 1.14. The index m of $\mathcal{O}_{m,j}$ is the sum of the multiplicities of $\mathcal{P}^{(k)}_{m,j}$.

Fourth, establish the formula between $\mathcal{UD}_{n,i}$ and $\mathcal{UD}_{n,i+1}$. Let T be a dup-tree in $\mathcal{UD}_{n,i+1}$. Removing one of the leaves labeled by $i+1$ in T might or might not create a twin-cherry, that is, the resulting tree falls into $\mathcal{UD}_{n,i}$ or $\mathcal{UD}^{(1)}_{n,i}$. For a dup-tree in $\mathcal{UD}_{n,i}$, there are $2(n+i)-4$ edges that do not create a twin-cherry by attaching a leaf labeled by $i+1$; for a dup-tree in $\mathcal{UD}^{(1)}_{n,i}$, there are exactly two edges which eliminate the twin-cherry. This gives the formula and with the previous equation, we have

$$\begin{aligned}(2n+2i-4)\ \mathrm{UD}_{n,i} + 2\,\mathrm{UD}^{(1)}_{n,i} &= \sum_{m,j} m\,\mathrm{O}_{m,j}\\ &= 2\sum_{m,j}\mathrm{O}_{m,j} + \sum_{m,j}(m-2)\,\mathrm{O}_{m,j}\\ &= 2\,\mathrm{UD}_{n,i+1} - \mathrm{O}_{1,1} + \mathrm{O}_{3,1} + 2\,\mathrm{O}_{4,1}.\end{aligned}$$

Note that we choose 2 as the coefficient for $\mathrm{UD}_{n,i+1}$ because, in addition to reducing the number of $\mathcal{O}_{m,j}$ to consider, it almost completes the counting.

Finally, it suffices to count $\mathrm{UD}^{(1)}_{n,i}$, $\mathrm{O}_{1,1}$, $\mathrm{O}_{3,1}$ and $\mathrm{O}_{4,1}$. One thing we need to consider is how to generate the labels $[n] \uplus [i]$.

1. $\mathrm{UD}^{(1)}_{n,i} = i\,\mathrm{UD}_{n,i-1}$.

   Proof. Without loss of generality, we assume that the twin-cherry uses label i; replacing it by a leaf labeled i gives a dup-tree in $\mathcal{UD}_{n,i-1}$. ■

2. $\mathrm{O}_{1,1} = \sum_{1\le d\le i} \binom{i}{d} (2d-1)!!\,\mathrm{UD}_{n-d,i-d}$.

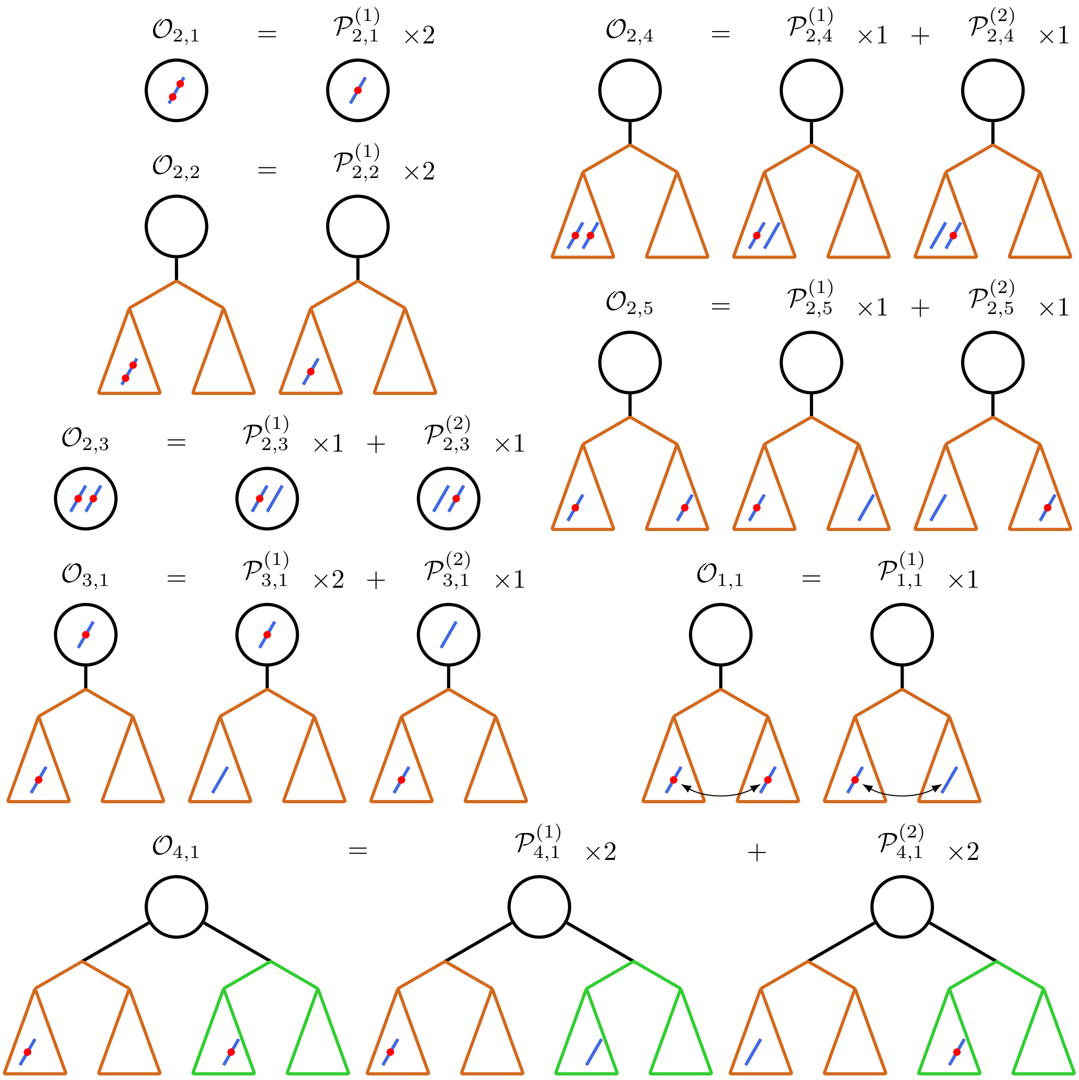


Figure 1.14: **Left of "="**: the representative graph of $\mathcal{O}_{m,j}$; **Right of "="**: the representative graph of $P^{(k)}_{m,j}$ with its multiplicity numbers.

Proof. First, recall that there are $(2n-3)!!$ rooted phylogenetic trees with n leaves and $2n-1$ edges. We directly construct the graphs in $\mathcal{O}_{1,1}$. Therefore, choose a phylogenetic tree with d leaves and attach a leaf labeled by $i+1$ to one of its edges. Next, duplicate the resulting tree and join their roots which represents the chocolate cherry. Next, choose a dup-tree in $\mathcal{UD}_{n-d,i-d}$ labeled by $\{d+1,\dots,n\} \uplus \{d+1,\dots,i\}$. Join the node labeled by $i+1$ to the root of the cherry part. Finally, relabel the original d leaves of the phylogenetic tree. ■

3. $\mathrm{O}_{3,1} + 2\,\mathrm{O}_{4,1} = \sum\limits_{1\le d\le i} \binom{i}{d} (2d-1)!!\,\mathrm{UD}_{n-d+1,i-d}.$

Proof. Similarly, consider the graphs $\mathcal{O}_{3,1}$ and $\mathcal{O}_{4,1}$. Except the common chocolate cherry with one b-edge, the situations described by the two sets show that the other b-edge can occur *anywhere*. Choose a phylogenetic tree with d leaves. Duplicate it, join their roots and attach a leaf labeled by $i+1$ to one chosen edge from the first tree. Second, choose a dup-tree in $\mathcal{UD}_{n-d+1,i-d}$ labeled by $\{d+1,\dots,n+1\} \uplus \{d+1,i\}$. Join the node labeled by $n+1$ to the root of the cherry part. Note that the label $i+1$ is already used, so there is no need to attach one more red dot. Finally, we do the same relabeling process as in the previous case. This shows the claimed formula. ■

Overall, we obtain the result

$$\begin{aligned}\mathrm{UD}_{n,i+1} = (n+i-2)\,\mathrm{UD}_{n,i} + i\,\mathrm{UD}_{n,i-1} + \\ \frac{1}{2}\sum_{1\le d\le i}\binom{i}{d}(2d-1)!!\,\left(\mathrm{UD}_{n-d,i-d} - \mathrm{UD}_{n-d+1,i-d}\right), \quad (1.9)\end{aligned}$$

for $0 \le i < n$ with initials $\mathrm{UD}_{n,0} = (2n-5)!!$. ■

1.2 Asymptotic Methods

In this section, we review several analytic tools that will be used throughout the paper. We begin with the framework of *generating functions*, which provides a bridge between combinatorial structures and analytic methods. We then present the principles of *singularity analysis*, enabling us to extract precise asymptotic estimates of coefficients from the local behavior of generating functions near their dominant singularities. Following this, we briefly recall the *Laplace method*, a classical technique for approximating integrals that arise in asymptotic enumeration. Finally, we summarize the *Lagrange inversion formula*, which offers an explicit way to determine coefficients of implicitly defined generating functions. These tools together form the analytic foundation for the results developed in the subsequent sections.

1.2.1 Generating Functions

Generating functions are a powerful tool in combinatorics and discrete mathematics that encode sequences as algebraic expressions, typically formal power series. Their main uses include: encoding sequences, finding closed-form expressions, and solving recurrence relations and counting problems. Most importantly, they build a bridge between combinatorics and analysis, allowing us to use analytical tools to understand and manipulate them.

Given a sequence, we define the most useful types of generating functions.

Definition 1.2.1 (Generating functions)**.** Let $\langle a_n \rangle_{n \geq 0}$ be a sequence of numbers.

The *ordinary generating function* (OGF) of a_n is defined as

$$A(z) = \sum_{n \geq 0} a_n z^n. \tag{1.10}$$

The *exponential generating function* (EGF) of a_n is defined as

$$A(z) = \sum_{n \geq 0} a_n \frac{z^n}{n!}. \tag{1.11}$$

Some basic operators on sequences or generating functions will directly affect the other. We list the fundamental ones below.

Proposition 6 (Basic operations)**.** Let $\langle a_n \rangle_{n \geq 0}$ and $\langle b_n \rangle_{n \geq 0}$ be two sequences.

Let $A(z)$ and $B(z)$ be the OGFs of a_n and b_n, respectively. Then the function is the OGF of the sequence, where

function	**sequence**
$A(z)+B(z)$	a_n+b_n
$zA(z)$	a_{n-1}
$A'(z)$	$(n+1)a_{n+1}$
$A(z)B(z)$	$\sum_{k=0}^{n} a_n b_{n-k}$

Let $\alpha(z)$ and $\beta(z)$ be the EGFs of a_n and b_n, respectively. Then the function is the EGF of the sequence, where

function	**sequence**
$\alpha(z)+\beta(z)$	a_n+b_n
$z\alpha(z)$	na_{n-1}
$\alpha'(z)$	a_{n+1}
$\alpha(z)\beta(z)$	$\sum_{k=0}^{n} \binom{n}{k} a_n b_{n-k}$

Notice that $a_n = b_n = 0$ for $n < 0$ if it is used.

Depending on how the sequence is defined, not all corresponding generating functions can have a simple or closed form. Here we give several examples of generating functions.

Example 1.2.1 (From sequences to generating functions)**.**

1. (Fibonacci numbers) Let f_n be the n-th Fibonacci number which is defined by $f_n = f_{n-1} + f_{n-2}$ for $n \geq 2$ with initials $f_0 = 0$ and $f_1 = 1$. Then the OGF $F(z)$ is

$$F(z) = \frac{z}{1 - z - z^2}.$$

2. (Unlabeled rooted binary plane trees) Let c_n count the number of unlabeled rooted binary plane trees with n internal nodes (the size). (A tree is *plane* if the subtrees of any node are ordered; it is *binary* if every node has either outdegree 0 or 2.) By definition, a unlabeled rooted binary plane tree is either a single node or a root with its left and right subtrees unlabeled rooted binary plane trees of total size $n-1$. Thus, for $n \geq 1$,

$$c_n = \sum_{k=0}^{n-1} c_k c_{(n-1)-k}.$$

With the recurrence, we obtain the functional equation for its OGF $C(z)$:

$$C(z) = 1 + zC(z)^2 = \frac{1 - \sqrt{1-4z}}{2z}.$$

The term c_n is called Catalan numbers.

3. (Derangements) Let d_n be the number of permutations of size n that have no fixed points, which can be counted by the inclusion-exclusion principle

$$d_n = n! - \sum_{i=1}^{n} (-1)^{i+1} \binom{n}{i} (n-i)! \text{ for } n \geq 1 \text{ with } d_0 = 1.$$

Then, using the rules from Proposition 6 the EGF $D(z)$ is

$$D(z) = \frac{\mathrm{e}^{-z}}{1-z}.$$

4. (Merge sort) Let a_n be the cost of the merge sort of size n which is defined by $a_n = a_{\lceil n/2 \rceil} + a_{\lfloor n/2 \rfloor} + n - 1$ for $n \geq 2$ with $a_1 = 0$. Then its OGF $A(z)$ is

$$A(z) = \frac{(1+z)^2}{z} A(z^2) + \frac{z^2}{(1-z)^2} = \frac{z}{(1-z)^2} \sum_{j \geq 0} z^{2^j}.$$

Conversely, given a function $A(z)$, we define a notation for its n-coefficient.

Definition 1.2.2 (Coefficients)**.** Let $A(z)$ be an analytic function around $z = 0$.

Let a_n be the n-th coefficient of the formal power series of $A(z)$, which is denoted by

$$a_n = [z^n]\, A(z). \tag{1.12}$$

If $A(z)$ is an EGF, then its n-th coefficient is denoted by

$$a_n = n!\, [z^n]\, A(z). \tag{1.13}$$

There are various methods to extract the n-th coefficient depending on the given generating function. Below, we extract the coefficients directly; the resulting sequence may or may not admit a closed form.

Example 1.2.2 (From generating functions to sequences)**.** Direct method.

1. (Fibonacci numbers) Let $F(z) = z/(1 - z - z^2)$ be the OGF of the Fibonacci numbers, $\phi_1 = (1 + \sqrt{5})/2$ be the golden ratio, and $\phi_2 := (1 - \sqrt{5})/2$. Then

$$[z^n]\, F(z) = [z^n]\, \frac{z}{(1 - \phi_1 z)(1 - \phi_2 z)} = [z^n]\, \frac{1/\sqrt{5}}{1 - \phi_1 z} - \frac{1/\sqrt{5}}{1 - \phi_2 z} = \frac{\phi_1^n - \phi_2^n}{\sqrt{5}}.$$

2. (Catalan numbers) Let $C(z) = (1 - \sqrt{1 - 4z})/(2z)$ be the OGF of the Catalan numbers. Then

$$[z^n]\, \frac{1 - \sqrt{1-4z}}{2z} = \frac{-1}{2}\, [z^{n+1}]\, \sqrt{1 - 4z} = \frac{-1}{2}\, (-4)^{n+1} \binom{1/2}{n+1} = \frac{1}{n+1} \binom{2n}{n}.$$

3. (Merge sort) Let $A(z)$ be the OGF of the cost of the merge sort. Then

$$[z^n]\, \frac{z}{(1-z)^2} \sum_{j \geq 0} z^{2^j} = [z^n]\, \frac{z}{1-z} \sum_{n \geq 1} \lfloor \log_2 n \rfloor\, z^n = n \lfloor \log_2 n \rfloor + n - 2^{\lfloor \log_2 n \rfloor + 1} + 1.$$

4. (Bernoulli numbers) Let $B(z) = z/(\mathrm{e}^z - 1)$ be the EGF of the Bernoulli numbers. Then the n-th Bernoulli number is given by the double sum,

$$n!\, [z^n]\, \frac{z}{\mathrm{e}^z - 1} = \text{(skip)} = \sum_{k=0}^{n} \frac{1}{k+1} \sum_{\ell=0}^{k} (-1)^\ell \binom{k}{\ell} \ell^n.$$

In general, the coefficients can be computed by using power series expansion or Cauchy's Coefficient Formula.

Remark 7. We retain only its symbolic form:

$$a_n = \frac{A^{(n)}(0)}{n!} \quad \text{or} \quad a_n = \frac{1}{2\pi \mathrm{i}} \oint_{\gamma} \frac{A(z)}{z^{n+1}} \, \mathrm{d}z, \tag{1.14}$$

where γ is an oriented counterclockwise, simple, and closed contour enclosing the origin.

1.2.2 Singularity Analysis

Singular analysis studies how the singularities (i.e., non-analytic points such as poles or branch points) of a generating function influence the asymptotic behavior of its coefficients. The method starts with a Δ-domain, which is a specially shaped region in the complex plane. The idea is that when a function has a singularity on the boundary of its disk of convergence, one needs to understand how the function behaves near that singularity in order to extract asymptotics of its coefficients.

Definition 1.2.3 (Δ-domain and Δ-analytic)**.** Define the open domain $\Delta(\phi, R)$ by

$$\Delta(\phi, R) = \{z : |z| < R, z \neq 1 \text{ and } |\arg(z-1)| > \phi\}$$

for $0 < \phi < \pi/2$ and $R > 1$. Let $\zeta \neq 0$ be a complex number. We say …

1. A domain is a $\underline{\Delta\text{-}\textit{domain at } 1}$ if it is $\Delta(\phi, R)$ for some ϕ and R;
2. A domain is a $\underline{\Delta\text{-}\textit{domain at } \zeta}$ if it is $\zeta \cdot \Delta_0$ for a Δ-domain Δ_0 (see Figure 1.15);
3. A function is $\underline{\Delta\text{-}\textit{analytic}}$ if it is analytic in some Δ-domain.

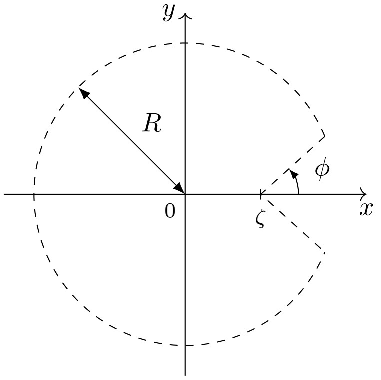


Figure 1.15: A Δ-domain at ζ.

This means it is the disk of radius R centered at the origin, but with small angular sectors (of width 2ϕ) around ζ removed.

Theorem 11 (Closure property)**.** Let $\alpha \notin \mathbb{Z}_{\leq 0}$ and $\beta \in \mathbb{R}$. We have,

1. Entire functions are Δ-analytic, so are

$$\frac{1}{(1-z)^{\alpha}} \quad \text{and} \quad \left(\frac{1}{z}\log\left(\frac{1}{1-z}\right)\right)^{\beta}$$

and products of these functions.

2. (Closure properties) Let $f(z)$ and $g(z)$ be two Δ-analytic functions, so are

$$f(z)+g(z),\ f(z)-g(z),\ f(z)g(z),\ f(g(z)),\ f'(z)\ \text{ and }\ \int^z f(z).$$

Standard function scale. Consider Δ-analytic functions whose expansion at a singularity ζ involves the form

$$f(z) \sim \frac{1}{\left(1-\frac{z}{\zeta}\right)^{\alpha}} \left(\frac{1}{z}\log\frac{1}{1-\frac{z}{\zeta}}\right)^{\beta}.$$

The following theorem gives a corresponding asymptotic results for the right-hand side of this asymptotic equivalence with singularity at $\zeta = 1$.

Theorem 12 (Standard function scale and variations)**.** Let

$$f(z) = \frac{1}{(1-z)^{\alpha}} \left(\frac{1}{z}\log\frac{1}{1-z}\right)^{\beta}.$$

Then,

1. If $\alpha \notin \mathbb{Z}_{\leq 0}$ and $\beta \notin \mathbb{Z}_{\geq 0}$, then

$$[z^n]\, f(z) \sim \frac{n^{\alpha-1}}{\Gamma(\alpha)} (\log n)^{\beta} \sum_{j\geq 0} \frac{C_j}{(\log n)^j}, \tag{1.15}$$

   where $C_j = \binom{\beta}{j}\Gamma(\alpha)\frac{\mathrm{d}^j}{\mathrm{d}s^j}\Gamma(s)^{-1}\Big|_{s=\alpha}$.

2. If $\alpha \notin \mathbb{Z}_{\leq 0}$ and $\beta \in \mathbb{Z}_{\geq 0}$, then

$$[z^n]\, f(z) \sim \frac{n^{\alpha-1}}{\Gamma(\alpha)} \sum_{j\geq 0} \frac{E_j(\log n)}{n^j}, \tag{1.16}$$

   where the E_j are polynomials of degree β.

3. If $\alpha \in \mathbb{Z}_{\leq 0}$ and $\beta \notin \mathbb{Z}_{\geq 0}$, then

$$[z^n]\, f(z) \sim n^{\alpha-1} (\log n)^{\beta-1} \sum_{j\geq 0} \frac{D_{j+1}}{(\log n)^j}, \tag{1.17}$$

   where $D_j = \binom{\beta}{j}\frac{\mathrm{d}^j}{\mathrm{d}s^j}\Gamma(s)^{-1}\Big|_{s=\alpha}$.

4. If $\alpha \in \mathbb{Z}_{\leq 0}$ and $\beta \in \mathbb{Z}_{\geq 0}$, then

$$[z^n]\, f(z) \sim n^{\alpha-1} \sum_{j\geq 0} \frac{F_j(\log n)}{n^j}, \tag{1.18}$$

   where $F_j(z)$ is a polynomial of degree $\beta - 1$.

We only prove the case with $\alpha \notin \mathbb{Z}_{\leq 0}$ and $\beta = 0$, the four statements in Theorem 12 can be derived similarly; for the full proof see [17, Section VI.2].

Proof of Theorem 12 for $\alpha \notin \mathbb{Z}_{\leq 0}$ and $\beta = 0$. Consider $f(z) = (1-z)^{-\alpha}$ with one singularity at $\zeta = 1$. Start with the Cauchy's coefficient formula,

$$f_n = [z^n]\frac{1}{(1-z)^\alpha} = \frac{1}{2\pi i}\oint_\gamma \frac{(1-z)^{-\alpha}}{z^{n+1}}\,\mathrm{d}z, \tag{1.19}$$

where γ is an counterclockwise oriented, simple, and closed contour enclosing the origin.

Let $R \geq 1$ be a positive real number. We choose $\gamma \equiv \gamma_1$, where the contour γ_1 is positively oriented and consists of the four parts:

1. $\gamma_{1,1} = \left\{z = 1 + \frac{1}{n}\mathrm{e}^{i\phi} : |\phi - \pi| \leq \frac{\pi}{2}\right\}$,
2. $\gamma_{1,2} = \left\{z = x + \frac{i}{n} : 1 \leq x \text{ and } |z| \leq R\right\}$,
3. $\gamma_{1,3} = \left\{z = R\mathrm{e}^{i\phi} : -\pi < \phi \leq \pi\right\} \setminus \left\{z = R\mathrm{e}^{i\phi} : |\phi| < \frac{\pi}{2} \text{ and } |\Im(z)| < \frac{1}{n}\right\}$,
4. $\gamma_{1,4} = \left\{z = x - \frac{i}{n} : 1 \leq x \text{ and } |z| \leq R\right\}$,

Next, since the integrand along $\gamma_{1,4}$ decreases as $\mathcal{O}(R^{-n})$, we can let R tend to infinity and choose $\gamma \equiv \mathcal{H}(n)$ which is a Hankel contour consisting of,

$$\mathcal{H}^-(n) = \left\{x - \frac{i}{n} : x \geq 1\right\}, \quad \mathcal{H}^0(n) = \gamma_{1,1}, \quad \text{and} \quad \mathcal{H}^-(n) = \left\{x + \frac{i}{n} : x \geq 1\right\}.$$

Now, a change of variable,

$$z = 1 + \frac{t}{n}$$

in (1.19) gives the form

$$f_n = \frac{n^{\alpha-1}}{2\pi i}\int_{\mathcal{H}} (-t)^{-\alpha}\left(1 + \frac{t}{n}\right)^{-n-1}\mathrm{d}t, \tag{1.20}$$

where $\mathcal{H}$ is the Hankel contour winding around 0, being at distance 1 from the positive real axis. For fixed t, consider the asymptotic expansion,

$$\left(1 + \frac{t}{n}\right)^{-n-1} = \mathrm{e}^{-(n+1)\log(1+t/n)} = \mathrm{e}^{-t}\left(1 + \frac{t^2 - 2t}{2n} + \frac{3t^4 - 20t^3 + 24t^2}{24n^2} + \cdots\right), \tag{1.21}$$

which shows that the integrand in (1.20) converges pointwise to $(-t)^{-\alpha}\mathrm{e}^{-t}$. Substitution of the asymptotic form

$$\left(1 + \frac{t}{n}\right)^{-n-1} = \mathrm{e}^{-t}\left(1 + \mathcal{O}\left(\frac{1}{n}\right)\right),$$

as $n \to \infty$, inside the integral (1.20) suggests that

$$f_n = \frac{n^{\alpha-1}}{2\pi i} \int_{\mathcal{H}} (-t)^{-\alpha} \mathrm{e}^{-t} \, \mathrm{d}t \left(1 + \mathcal{O}\left(\frac{1}{n}\right)\right) = \frac{n^{\alpha-1}}{\Gamma(\alpha)} \left(1 + \mathcal{O}\left(\frac{1}{n}\right)\right),$$

where we used Hankel's formula for the Gamma function:

$$\frac{1}{2\pi \mathrm{i}} \int_{\mathcal{H}} (-t)^{-\alpha} \mathrm{e}^{-t} \, \mathrm{d}t = \frac{1}{\Gamma(\alpha)}.$$

To justify this formal argument, we proceed as follows:

1. Split the contour $\mathcal{H}$ according to $\Re(t) \le \log^2 n$ and $\Re(t) > \log^2 n$.
2. Verify that the part corresponding to $\Re(t) > \log^2 n$ is negligible in the scale of the problem; for instance,
$$\left(1 + \frac{t}{n}\right)^{-n} = \mathcal{O}\left(\exp\left(-\log^2 n\right)\right) \text{ for } \Re(t) \ge \log^2 n.$$
3. Use a terminating form of (1.21) to develop an expansion to any predetermined order, with uniform error terms, for the part corresponding to $\Re(t) \le \log^2 n$. (This is possible because $t/n = \mathcal{O}\left(\log^2 n/n\right)$ is small.)

These considerations validate term-by-term integration of expansion (1.21) within the integral of (1.20), so that the full expansion of f_n is determined as follows: a term of the form t^r/n^s in the expansion (1.21) induces, by Hankel's formula, a term of the form $n^{-s}/\Gamma(\alpha - r)$. (The expansion so obtained is non-degenerate provided α differs from a negative integer or zero.) Since

$$\frac{1}{\Gamma(\alpha - k)} = \frac{1}{\Gamma(\alpha)} (\alpha - 1)(\alpha - 2) \cdots (\alpha - k),$$

the expansion (1.16) for $\beta = 0$ follows. ■

Denote by $\mathcal{S}$ the standard scale of functions:

$$\mathcal{S} := \left\{ \frac{1}{(1-z)^{\alpha}} \left(\frac{1}{z} \log \frac{1}{1-z}\right)^{\beta} : \alpha, \beta \in \mathbb{C} \right\}. \tag{1.22}$$

Transfer. To translate an approximation of a function near a singularity into an asymptotic approximation of its coefficients, we require a way to extract coefficients of *error terms* (known usually in $\mathcal{O}(\cdot)$ or $o(\cdot)$ form) in the expansion of a function near a singularity. Analyticity in a Δ-domain is the basic condition for *transfer* to coefficients of error terms in asymptotic expansions.

Theorem 13 (Transfer, Big-Oh and little-oh)**.** Let α, β be arbitrary real numbers and let $f(z)$ be a function that is Δ–analytic.

1. If $f(z)$ satisfies the condition,

$$f(z) = \mathcal{O}\left(\frac{1}{(1-z)^\alpha}\left(\frac{1}{z}\log\frac{1}{1-z}\right)^\beta\right).$$

in the intersection of a neighborhood of 1 and its Δ-domain, then its coefficient satisfies

$$[z^n]f(z) = \mathcal{O}\left(n^{\alpha-1}\log^\beta n\right).$$

2. If $f(z)$ satisfies the condition,

$$f(z) = o\left(\frac{1}{(1-z)^\alpha}\left(\frac{1}{z}\log\frac{1}{1-z}\right)^\beta\right).$$

in the intersection of a neighborhood of 1 and its Δ-domain, then its coefficient satisfies

$$[z^n]f(z) = o\left(n^{\alpha-1}\log^\beta n\right).$$

This result is proved by extending the method used for Theorem 12; see [17, Section VI.3].

An immediate corollary of Theorem 13 is the possibility of transferring asymptotic equivalences to coefficients:

Corollary 1.2.1. Let $f(z)$ be a Δ-analytic and

$$f(z) \sim \frac{1}{(1-z)^\alpha}, \quad \text{as } z \to 1 \text{ in } \Delta,$$

with $\alpha \notin \{0, -1, -2, \ldots\}$. Then, the coefficients of f satisfy

$$[z^n]f(z) \sim \frac{n^{\alpha-1}}{\Gamma(\alpha)}.$$

The process of singularity analysis. In Theorem 12 and 13, we have developed a collection of statements granting us the existence of correspondences between properties of a function $f(z)$ singular at an isolated point ($z = 1$) and the asymptotic behavior of its coefficients $f_n = [z^n]f(z)$. At this stage, we now possess the necessary tools to justify the term-by-term transfer from a singular expansion of a function, namely, its local expansion at a dominant singularity, to an asymptotic estimation of its coefficients. We therefore state the following theorem which can again be found in [17, Section VI.4].

Theorem 14 (Singularity analysis, single singularity)**.** Let $f(z)$ be a function analytic at 0 with a singularity at ζ, such that $f(z)$ can be continued to the Δ-domain $\zeta \cdot \Delta_0$. Assume that there are two functions σ, τ where σ is a (finite) linear combination of functions in $\mathcal{S}$ and $\tau \in \mathcal{S}$, so that

$$f(z) = \sigma\left(\frac{z}{\zeta}\right) + \mathcal{O}\left(\tau\left(\frac{z}{\zeta}\right)\right), \quad \text{as } z \to \zeta \text{ in } \zeta \cdot \Delta_0.$$

Then the coefficients of $f(z)$ satisfy the asymptotic estimation,

$$f_n = \zeta^{-n}\sigma_n + \mathcal{O}\left(\zeta^{-n}\tau_n^*\right),$$

where $\sigma_n = [z^n]\sigma(z)$ has its coefficients determined by Theorem 12 and $\tau_n^* = n^{\alpha-1}\log^\beta n$ if $\tau(z) = (1-z)^{-\alpha}(1/z\ \log(1-z)^{-1})^\beta$.

The singularity analysis process consists of a series of steps as follows:

Proposition 7 (The singularity analysis process)**.** Let $f(z)$ be a function analytic at 0 whose coefficients are to be asymptotically analyzed.

1. **Preparation.** This consists in locating dominant singularities and checking analytic continuation.
    (a) **Locate singularities.** Determine the dominant singularities of $f(z)$ (assumed not to be entire). Check that $f(z)$ has a single singularity ζ on its circle of convergence.
    (b) **Check continuation.** Establish that $f(z)$ is analytic in some Δ-domain at ζ.
2. **Singular expansion.** Analyze the $f(z)$ as z tends to ζ in the domain $\zeta \cdot \Delta_0$. More precisely, as $z \to 1$, determine in that domain an expansion of the form,

$$f(z) = \sigma\left(\frac{z}{\zeta}\right) + \mathcal{O}\left(\tau\left(\frac{z}{\zeta}\right)\right) \quad \text{with } \tau(z) = o(\sigma(z)).$$

   For the method to succeed, the functions σ and τ should belong to the standard scale of functions $\mathcal{S}$; see Theorem 12.

3. **Transfer.** Translate the main term $\sigma(z)$ by applying Theorem 12; transfer the error term with Theorem 13 and conclude that

$$[z^n]f(z) = \zeta^{-n}\sigma_n + \mathcal{O}\left(\zeta^{-n}\tau_n^*\right),$$

   as $n \to \infty$, where $\sigma_n = [z^n]\sigma(z)$ and $\tau_n^* = [z^n]\tau(z)$ provided the corresponding exponent $\alpha \notin \mathbb{Z}_{\leq 0}$, otherwise, the factor $1/\Gamma(\alpha) = 0$ should be dropped.

Remark 8 (Singularity analysis with multiple singularities). Assume $f(z)$ has finitely many dominant singularities on their circle of convergence. The process can be applied independently to the dominant singularities. One must analyze the local expansion at each singularity and combine their contributions to the coefficients; see Example 1.2.4-4.

We apply the singularity analysis process in the following example.

Example 1.2.3 (Labelled 2-regular graphs)**.** Let $R(z)$ be the EGF of the number of labelled 2-regular graphs. They are composed of a set of undirected cycles with size ≥ 3, hence,

$$R(z) = \exp\left(\frac{1}{2}\left(\log\left(\frac{1}{1-z}\right) - z - \frac{z^2}{2}\right)\right) = \frac{\mathrm{e}^{-z/2-z^2/4}}{\sqrt{1-z}}.$$

Note that this function has a single singularity at $z = 1$. We follow step by step the singularity analysis process mentioned above.

1. **Preparation.** The function $R(z)$ is the product of an entire function $\mathrm{e}^{-z/2-z^2/4}$ and a Δ-analytic function $(1-z)^{-1/2}$, so $R(z)$ is also Δ-analytic with a single singularity at $z = 1$.

2. **Singular expansion.** The asymptotic expansion of $R(z)$ near $z = 1$ is obtained from the standard expansion of $\mathrm{e}^{-z/2-z^2/4}$ at $z = 1$,

$$\mathrm{e}^{-z/2-z^2/4} = \mathrm{e}^{-3/4} + \mathrm{e}^{-3/4}(1-z) + \frac{\mathrm{e}^{-3/4}}{4}(1-z)^2 - \frac{\mathrm{e}^{-3/4}}{12}(1-z)^3 + \cdots,$$

   and the factor $(1-z)^{-1/2}$ itself is already an asymptotic expansion and it is valid in any Δ-domain. The multiplication gives us a complete expansion,

$$R(z) \sim \frac{\mathrm{e}^{-3/4}}{\sqrt{1-z}} + \mathrm{e}^{-3/4}\sqrt{1-z} + \frac{\mathrm{e}^{-3/4}}{4}(1-z)^{3/2} - \frac{\mathrm{e}^{-3/4}}{12}(1-z)^{5/2} + \cdots, \quad (1.23)$$

   from which terminating forms, with an $\mathcal{O}$–error term, can be extracted.

3. **Transfer.** We take for instance the expansion in (1.23) limited to two terms plus an error term:

$$R(z) = \frac{\mathrm{e}^{-3/4}}{\sqrt{1-z}} + \mathrm{e}^{-3/4}\sqrt{1-z} + \mathcal{O}\left((1-z)^{3/2}\right).$$

   The singularity analysis process allows the transfer of (1.23) to coefficients, which we can present as follows,

$$\begin{aligned}
[z^n]\frac{\mathrm{e}^{-3/4}}{\sqrt{1-z}} &= \mathrm{e}^{-3/4}\binom{n-1/2}{-1/2} &&\sim \frac{\mathrm{e}^{-3/4}}{\sqrt{\pi n}}\left(1 - \frac{1}{8n} + \frac{1}{128n^2} + \cdots\right),\\
[z^n]\mathrm{e}^{-3/4}\sqrt{1-z} &= \mathrm{e}^{-3/4}\binom{n-3/2}{-3/2} &&\sim \frac{\mathrm{e}^{-3/4}}{2\sqrt{\pi n^3}}\left(1 + \frac{3}{8n} + \cdots\right)\\
[z^n]\mathcal{O}\left((1-z)^{3/2}\right) &= \mathcal{O}\left(n^{-5/2}\right).
\end{aligned}$$

The terms are then collected, with each expansion truncated at the level of the dominant error term, resulting in a three-term expansion in this case. In what follows, we will no longer provide the detail of such computations; instead, we will simply present the expansion of the function alongside the corresponding expansion of its coefficients, as illustrated by the following correspondence ($\longrightarrow$).

$$\begin{aligned} R(z) &= \frac{\mathrm{e}^{-3/4}}{\sqrt{1-z}} + \mathrm{e}^{-3/4}\sqrt{1-z} + \mathcal{O}\left((1-z)^{3/2}\right) \\ \longrightarrow \quad r_n &= \frac{\mathrm{e}^{-3/4}}{\sqrt{\pi n}} - \frac{5\,\mathrm{e}^{-3/4}}{8\sqrt{\pi n^3}} + \mathcal{O}\left(n^{-5/2}\right). \end{aligned}$$

Some more examples for obtaining the main asymptotic term only are given in the next example.

Example 1.2.4.

1. (Fibonacci numbers) Let $F(z) = z/(1-z-z^2)$ be the OGF of the Fibonacci numbers f_n, and $\phi_1 := (1+\sqrt{5})/2, \phi_2 := (1-\sqrt{5})/2$ such that $1/\phi_1, 1/\phi_2$ are the two roots of $1-z-z^2$. Since $|1/\phi_1| < 1 < |1/\phi_2|$, we expand the singular expansion of $F(z)$ at the dominant singularity $z = 1/\phi_1$.

$$F(z) = \frac{\sqrt{5}}{5(1-\phi_1 z)} - \frac{1+\sqrt{5}}{10} + \mathcal{O}\left(z - \frac{1}{\phi_1}\right).$$

By applying Theorem 12, we have the main asymptotic of the coefficient,

$$f_n \sim [z^n]\frac{\sqrt{5}}{5(1-\phi_1 z)} = \frac{\phi_1^n}{\sqrt{5}}.$$

2. (Catalan numbers) Let $C(z)$ be the OGF of the Catalan numbers c_n, Expand the singular expansion of $C(z)$ at $z = 1/4$ (with branch cut $\mathbb{R}_{\geq 1/4}$),

$$C(z) = 4 - 2\sqrt{1-4z} - 8z - \mathcal{O}\left((1-4z)^{3/2}\right).$$

By applying Theorem 12, we have the main asymptotic of the coefficient,

$$c_n \sim [z^n] - 2\sqrt{1-4z} = -2\,4^n[z^n]\frac{1}{(1-z)^{-1/2}} \sim \frac{4^n}{\sqrt{\pi}\,n^{3/2}},$$

which also matches the asymptotic of its exact solution.

3. (Derangements) Let $D(z) = \mathrm{e}^{-z}/(1-z)$ be the EGF of the number of derangements d_n. Expand it at $z = 1$,

$$D(z) = \frac{\mathrm{e}^{-1}}{1-z} - \mathrm{e}^{-1} + \mathcal{O}\left(1-z\right) \quad \Longrightarrow \quad D(z) \sim \frac{\mathrm{e}^{-1}}{1-z}.$$

Then the first asymptotic of the n-th coefficient is

$$[z^n]D(z) = \frac{d_n}{n!} \sim [z^n]\,\frac{\mathrm{e}^{-1}}{1-z} = \mathrm{e}^{-1} \approx 0.368.$$

We have about 36.8% chance to obtain a derangement from a random permutation of size n.

4. (Bernoulli numbers) Let $B(z) = z/(\mathrm{e}^z - 1)$ be the EGF of the Bernoulli numbers B_n with singularities $\chi_n = 2n\pi i$ for $n \in \mathbb{Z} \setminus \{0\}$. χ_1 and χ_{-1} are the two singularities with smallest modulus $|\chi_1| = |\chi_{-1}| = 2\pi$. Expand $B(z)$ at $z = \chi_1$ and $z = \chi_{-1}$, respectively:

$$\begin{aligned} B(z) &\sim \frac{2\pi i}{z-\chi_1} + (1-\pi i) + \mathcal{O}\left(z-\chi_1\right) \quad \text{as } z \to \chi_1, \text{ and} \\ B(z) &\sim \frac{-2\pi i}{z-\chi_{-1}} + (1+\pi i) + \mathcal{O}\left(z-\chi_{-1}\right) \quad \text{as } z \to \chi_{-1}. \end{aligned}$$

The asymptotics is:

$$\begin{aligned} \frac{B_n}{n!} \sim [z^n]\frac{2\pi i}{z-\chi_1} + \frac{-2\pi i}{z-\chi_{-1}} &= -\frac{2\pi i}{(2\pi i)^{n+1}} + \frac{2\pi i}{(-2\pi i)^{n+1}} \\ &= \begin{cases} 0, & \text{if } n \text{ is odd;} \\ -2(2\pi i)^{-n} & \text{if } n \text{ is even.} \end{cases} \end{aligned}$$

Setting $n = 2m$ gives

$$B_{2m} \sim (-1)^{m+1}\frac{2\,(2m)!}{(2\pi)^{2m}}.$$

5. (Stirling number of first kind) Let $s_{n,k}$ be the Stirling number of first kind, which counts the number of permutations of size n with k cycles. The bivariate generating function $S(z,u)$ is

$$S(z,u) = \sum_{n\geq 0}\sum_{k=0}^{n} s(n,k)\,u^k\,\frac{z^n}{n!} = \mathrm{e}^{u\,\log\frac{1}{1-z}} = (1-z)^{-u}.$$

Then $S_u(z,1)$ is the OGF of average number of cycles in a random permutation. Using singularity analysis gives:

$$[z^n]S_u(z,1) = [z^n]\frac{1}{1-z}\,\log\frac{1}{1-z} \sim \log n.$$

This shows that there is about $\log n$ cycles in a uniformly picked permutation of size n.

1.2.3 Laplace Method

The Laplace method is used to estimate integrals (or sums) where the dominant contribution comes from around the maximum value of the function (or terms). In combinatorics, it is often used to approximate coefficients of generating functions or estimate growth rates of sequences.

Theorem 15 (Laplace method for a real integral)**.** Let f and g be infinitely differentiable real-valued functions defined over some compact interval I of the real line. Assume that $|g(x)|$ attains its maximum at a unique point x_0 interior to I and that $f(x_0), g(x_0), g''(x_0) \neq 0$. Then the integral

$$I_n = \int_I f(x) g(x)^n \,\mathrm{d}x$$

admits a complete asymptotic expansion

$$I_n \sim \sqrt{\frac{2\pi}{\lambda n}}\, f(x_0) g(x_0)^n \left(1 + \sum_{j\geq 1} \frac{\delta_j}{n^j}\right), \qquad \text{where } \lambda = -\frac{g''(x_0)}{g(x_0)}.$$

The above cited version can be found in Appendix B in [17].

A direct application is to obtain the main asymptotics of the Stirling formula.

Example 1.2.5 (Stirling formula by Laplace method for real integrals)**.** We start from the known representation of $n!$:

$$I(n) = \int_0^\infty \mathrm{e}^{-nx} x^n \,\mathrm{d}x = \frac{n!}{n^{n+1}}.$$

The maximum occurs at $x_0 = 1$ and we split the integral into two parts,

$$I(n) = \underbrace{\int_0^2 \mathrm{e}^{-nx} x^n \,\mathrm{d}x}_{=:I_1} + \underbrace{\int_2^\infty \mathrm{e}^{-nx} x^n \,\mathrm{d}x}_{=:I_2}.$$

Applying Theorem 15 ($f(x) = 1, g(x) = x\mathrm{e}^{-x}$) on I_1 gives

$$I_1 = \sqrt{\frac{2\pi}{n}}\mathrm{e}^{-n}\left(1 + \mathcal{O}\left(\frac{1}{n}\right)\right).$$

Estimating the tail I_2:

$$\begin{aligned} I_2 &= \left(2\mathrm{e}^{-2}\right)^n \int_0^\infty \mathrm{e}^{-nx}\left(1 + \frac{x}{2}\right)^n \mathrm{d}x, && \text{by mapping } x \mapsto x+2, \\ &< \left(2\mathrm{e}^{-2}\right)^n \int_0^\infty \mathrm{e}^{-nx}\mathrm{e}^{nx/2} \,\mathrm{d}x = \frac{2}{n}\left(2\mathrm{e}^{-2}\right)^n \downarrow 0, && \text{since } \log(1 + x/2) < x/2. \end{aligned}$$

Finally, we have the main asymptotic term of $n!$,

$$n! = n^n \mathrm{e}^{-n}\sqrt{2\pi n}\left(1 + \mathcal{O}\left(\frac{1}{n}\right)\right).$$

Before discussing the Laplace method for sums, we consider the Euler–Maclaurin formula. It provides an approximation of a finite sum by an integral, with correction terms involving derivatives and Bernoulli numbers. It is extremely useful in analysis, number theory, and asymptotics because it connects discrete and continuous viewpoints. A proof can be found in many textbooks on discrete mathematics, e.g., [27, Section 9.5, Equation (9.78)].

Theorem 16 (Euler–Maclaurin formula)**.** Let $a < b$ be two integers and $f(x)$ be a function which is continuously differentiable $2m$ times. Then

$$\sum_{k=a}^{b} f(k) = \int_a^b f(x)\,\mathrm{d}x + \frac{f(a)+f(b)}{2} + \sum_{j=1}^{m} \frac{B_{2j}}{(2j)!} \cdot f^{(2j-1)}(x)\Big|_a^b + R_m,$$

where B_i is the i-th Bernoulli number and the remainder term R_m is defined by

$$R_m = -\int_a^b f^{(2m)}(x) \frac{B_{2m}(\{x\})}{(2m)!}\,\mathrm{d}x,$$

where $B_i(x)$ is the i-th Bernoulli polynomial and $\{x\} = x - \lfloor x \rfloor$.

With the same example, the Euler–Maclaurin formula gives a weaker version of Stirling's formula.

Example 1.2.6 (Stirling formula via Euler–Maclaurin formula)**.** Define $A(n)$ by

$$A(n) = \log n! = \sum_{k=1}^{n} \log k.$$

Applying the Euler–Maclaurin formula with $a = 1$, $b = n$, $f(x) = \log x$ and $m = 1$ gives

$$A(n) = \int_1^n \log(x)\,\mathrm{d}x + \frac{f(1)+f(n)}{2} + \frac{B_2}{2!}\frac{1}{x}\Big|_1^n + R_1, \tag{1.24}$$

where

$$R_1 = \int_1^n \frac{B_2(\{x\})}{2\,x^2}\,\mathrm{d}x.$$

The error term can be estimated as follows:

$$R_1 = \mathcal{O}\left(\int_1^n \frac{\mathrm{d}x}{x^2}\right) = \mathcal{O}\left(1\right).$$

Define a constant c that collects the constants of all terms in the above equation (including the remainder term), that is,

$$c = 1 + \frac{f(1)}{2} - \frac{B_2}{2!} + c',$$

where c' is the constant part in R_1. We have an asymptotic expansion of $A(z)$,

$$A(n) = n\log n - n + \frac{1}{2}\log n + c + o(1).$$

Thus,

$$n! = e^c\sqrt{n}\left(\frac{n}{\mathrm{e}}\right)^n(1+o(1)),$$

where e^c is some constant (which equals $\sqrt{2\pi}$ as we know).

Although the terms provided by the Euler–Maclaurin formula yield an exact identity, they do not always give an asymptotic expansion. One still needs to show that the error term is of a smaller asymptotic order.

Laplace Method for sums. The main idea of the Laplace method for sums is very similar to that for integrals, namely, the major term is only obtained by a certain range of values. Compared with the above method, one difference is that the summand here is usually not continuous or differentiable. Thus, we rely on Gaussian approximation to convert it to a continuous and integrable form.

Take a finite or infinite sum $A(n)$ defined by

$$A(n) = \sum_{k \in \mathcal{K}} a_{n,k}.$$

1. Find the major range $\mathcal{J}$ of the sum such that the sum over the tails $\mathcal{J}^c$ of $a_{n,k}$ is negligible:

$$A(n) = \sum_{k \in \mathcal{J}} a_{n,k} + \sum_{k \in \mathcal{J}^c} a_{n,k} \quad \text{and} \quad \sum_{k \in \mathcal{J}^c} a_{n,k} = o\left(\sum_{k \in \mathcal{J}} a_{n,k} \right).$$

2. Centrally approximate $a_{n,k}$ by a Gaussian within the major range.

3. Apply the Euler–Maclaurin formula to obtain the asymptotics.

4. The tails of Gaussian integral are negligible.

A classical case is that the summand has only one maximum and nearby values decay rapidly (e.g. exponentially).

Example 1.2.7 (The Laplace method for sums)**.**

$$P(n) = \sum_{k=0}^{n} \frac{(n-k)^k (n-k)!}{n!} = \sqrt{\frac{\pi n}{2}} + \mathcal{O}\left(1\right).$$

Proof. Let $p(k) = p_{n,k} = (n-k)^k (n-k)!/n!$. Then, $p(0) = 1$ is the maximum and consider the values $p(x)$ near 0. The major range of the sum is for $x \le n^{1/2+\epsilon}$ since

$$\sum_{k=n^{1/2+\epsilon}}^{n} p(k) = o\left(\sum_{k=0}^{n^{1/2+\epsilon}} p(k) \right). \tag{1.25}$$

The valid range for a Gaussian-like approximation of $p(x)$ is for $x = o(n^{2/3})$,

$$\log \frac{p(x)}{p(0)} = -\frac{x(x+1)}{2n} + \mathcal{O}\left(\frac{x^3}{n^2} \right).$$

In summary, we choose $x = n^{3/5}$ and for $0 \le k \le n^{3/5}$,

$$p(k) = \mathrm{e}^{-k^2/2n} \left(1 - \frac{k}{2n} + \mathcal{O}\left(\frac{k^3}{n^2} \right) \right). \tag{1.26}$$

We now split the sum $P(n)$ into two parts by the ranges ,

$$P(n) = \sum_{0}^{n^{3/5}} p(k) + \sum_{n^{3/5}}^{n} p(k).$$

The second sum is approximated by (1.25) and in the valid range, we plug the asymptotic expansion of $p(k)$ in (1.26) into the summand.

$$P(n) = \left(\sum_{0}^{n^{3/5}} \mathrm{e}^{-k^2/2n} \right) (1 + o(1)) \, .$$

Apply the Euler–Maclaurin formula ($m = 1$) on the summation part,

$$S(n) := \sum_{0}^{n^{3/5}} \mathrm{e}^{-k^2/2n} = \left(\int_{0}^{n^{3/5}} \mathrm{e}^{-x^2/2n} \mathrm{d}x \right) + \mathcal{O}(1),$$

where non-main terms are contained in the Big $\mathcal{O}$ of 1. Since the tail integral of $\mathrm{e}^{-x^2/2n}$ is negligible, we can extend the range to infinity:

$$\begin{aligned} S(n) &= \int_{0}^{\infty} \mathrm{e}^{-x^2/2n} \mathrm{d}x - \int_{n^{3/5}}^{\infty} \mathrm{e}^{-x^2/2n} \mathrm{d}x + \mathcal{O}(1) \\ &= \sqrt{\frac{\pi n}{2}} + \mathcal{O}\left(1\right), \end{aligned}$$

where the second integral is bounded by a constant. Overall, we obtain

$$P(n) = S(n)(1 + o(1)) = \sqrt{\frac{\pi n}{2}} + \mathcal{O}\left(1\right). \quad \blacksquare$$

1.2.4 Lagrange Inversion Formula

Lagrange inversion formula, a special case of *Lagrange inversion theorem*, is one of the most useful and powerful tools in combinatorics. It can help us to find the coefficients of a generating function that is defined implicitly.

We show the common two versions of this formula together as follow.

Theorem 17 (Lagrange Inversion Formula)**.** Define three formal power series below:

$$A(z)=\sum_{n\geq 0}a_n z^n, \quad F(u)=\sum_{n\geq 0}f_n u^n, \quad \text{and} \quad \Phi(u)=\sum_{n\geq 0}\phi_n u^n$$

such that $F(0)=0$, $F'(0)\neq 0$, $\Phi(0)\neq 0$ and they satisfy the *Lagrange type equations*,

$$z=F(A(z)) \quad \text{and} \quad A(z)=z\Phi(A(z)), \tag{1.27}$$

that is, $\Phi(u)=u/F(u)$. Then, for each integer $n\geq 0$,

$$[z^n]\,A(z)=\frac{1}{n}\left[u^{n-1}\right]\left(\frac{u}{F(u)}\right)^n=\frac{1}{n}\left[u^{n-1}\right]\Phi(u)^n. \tag{1.28}$$

More general, let $G(u)$ be a formal power series, we have

$$[z^n]\,G(A(z))=\frac{1}{n}\left[u^{n-1}\right]G'(u)\left(\frac{u}{F(u)}\right)^n=\frac{1}{n}\left[u^{n-1}\right]G'(u)\Phi(u)^n. \tag{1.29}$$

The Lagrange inversion formula can be proved either by analytic or elementary way; see, e.g. [17, Appendix A] or [40].

Example 1.2.8 (Unlabeled rooted m-ary plane trees)**.** Let t_n be the number of unlabeled rooted m-ary plane trees with n vertices (including both internal and external vertices) and $T(z)$ be the OGF of t_n. Such a tree might be a single node or consist of a root which is followed by a sequence of m such trees. This gives the functional equation,

$$T(z)=z+z\,T(z)^m.$$

Let $F(\omega)=\omega/(1+\omega^m)$ with $F(0)=0$ and $F'(0)=1\neq 0$. Applying the Lagrange inversion formula gives

$$\begin{aligned}[z^n]\,T(z)&=\frac{1}{n}\left[\omega^{n-1}\right](1+\omega^m)^n\\&=\begin{cases}\dfrac{1}{n}\dbinom{n}{(n-1)/m}, & \text{if } m\mid n-1,\\[2ex] 0, & \text{if } m\nmid n-1.\end{cases}\end{aligned}$$

While the classical Lagrange Inversion Formula provides exact coefficient expressions for implicitly defined generating functions of the Lagrange type equation

$$A(z) = z\,\Phi(A(z)),$$

it is often insufficient for asymptotic purposes. In many combinatorial applications, what is essentially required is not a closed-form expression for the coefficients, but rather a precise description of their asymptotic growth. Extracting such asymptotics directly from the Lagrange formula can become technically involved, especially when singularity analysis must later be applied to evaluate the resulting integral representations.

Singular Inversion addresses this issue by shifting the point of view from exact coefficients to the analytic structure of the solution $A(z)$. Under mild smoothness assumptions, implicit equations of the form above fall into the smooth implicit-function schema, which ensures that the dominant singularity of $A(z)$ is of universal square-root type. More precisely, near the dominant singularity ρ, the solution admits a local expansion of the form

$$A(z) = \tau - \alpha\sqrt{1 - z/\rho} + O(1 - z/\rho),$$

from which coefficient asymptotics of the universal form

$$[z^n]\;A(z) \sim C\,\rho^{-n} n^{-3/2}.$$

follow immediately.

Theorem 18 (Singular Inversion; Theorem VI.6 in [17])**.** Let Φ be an analytic function at 0 satisfying the conditions:

$$\Phi(0) \neq 0, \quad [u^n]\,\Phi(u) \geq 0, \quad \text{and} \quad \Phi(u) \neq \phi_0 + \phi_1 u,$$

and within the open disc of convergence R of Φ at 0, there exists a positive solution $0 < \tau < R$ to the characteristic equation:

$$\Phi(\tau) - \tau\Phi'(\tau) = 0.$$

Let $A(z)$ be the solution of the Lagrange type equation (1.27), $A = z\Phi(A)$, satisfying $A(0) = 0$. Then, the quantity $\rho = \dfrac{\tau}{\Phi(\tau)}$ is the radius of convergence of $A(z)$ at 0, and the singular expansion of $A(z)$ near ρ is of the form

$$A(z) = \tau - \alpha_1\sqrt{1 - \frac{z}{\rho}} + \sum_{j \geq 2}(-1)^j \alpha_j \left(1 - \frac{z}{\rho}\right)^{j/2}, \quad \alpha_1 = \sqrt{\frac{2\Phi(\tau)}{\Phi''(\tau)}},$$

with α_j's are computable constants.

Additionally assume that Φ is aperiodic. Then one has that

$$[z^n]\,A(z) \sim \sqrt{\frac{\Phi(\tau)}{2\Phi''(\tau)}}\,\frac{\rho^{-n}}{\sqrt{\pi n^3}}\left(1+\sum_{k=1}^{\infty}\frac{\beta_k}{n^k}\right),$$

for a family β_k of computable constants.

Example 1.2.9 (Catalan number)**.** Let $C(z)$ be the OGF of the Catalan numbers and set $\tilde{C}(z) := C(z)-1$ so that $\tilde{C}(0)=0$. Then,

$$\tilde{C}(z) = z\,\Phi(\tilde{C}(z)),$$

with $\Phi(u)=(1+u)^2$ which satisfies the conditions in Theorem 18: $\Phi(0)=1$, Φ has non-negative coefficients and is not a polynomial of degree ≤ 1. The positive solution $\tau=1$ satisfies the character equation $\Phi(\tau)=\tau\Phi'(\tau)$. Then, by Theorem 18, $\rho=1/4$ is the radius convergence of $\tilde{C}(z)$ at 0, the singular expansion of $\tilde{C}(z)$ near $1/4$ is

$$\tilde{C}(z) = 1-2\sqrt{1-4z}+\cdots,$$

The asymptotic of coefficients is

$$[z^n]\tilde{C}(z) \sim \frac{4^n}{\sqrt{\pi n^3}}.$$

In 1984, Bender and Richmond [6] gave an asymptotic result related to the Lagrange inversion formula of a different type. More precisely, they derived asymptotic expansions for the coefficients of $(1+A(z))^{\alpha n+\beta}$ where $A(z)$ is a formal power series with rapidly growing coefficients.

Theorem 19 ([6])**.** Let $A(z)$ be a formal power series with coefficient a_n where $a_0=0$ and $a_1\neq 0$. Let p_n be the coefficient of z^n in $(1+A(z))^{\alpha n+\beta}$ where $\alpha\neq 0$ and β are fixed complex numbers. Let $R>0$ be a fixed integer.

1. If $na_{n-1}\sim\gamma a_n$, then
$$p_n = \alpha \mathrm{e}^{\alpha a_1\gamma} n a_n + \mathcal{O}\left(a_n\right). \tag{1.30}$$
2. If $na_{n-1}=o(a_n)$, then
$$p_n = \sum_{k=0}^{R-1} d_{n,k}a_{n-k} + \mathcal{O}\left(n^{R+1}a_{n-R}\right), \tag{1.31}$$
where $D_n(z)=\sum_{n,k} d_{n,k}z^k = (\alpha n+\beta)(1+A(z))^{\alpha n+\beta-1}$.

3. In either case, then $p_{n-1} = o(p_n)$ and

$$\sum_{k=R}^{n-R} |p_k p_{n-k}| = \mathcal{O}\left(p_{n-R}\right). \tag{1.32}$$

Chapter 2

Old and New Results for the Main Classes

In this chapter, we present the results from the following papers:

1. Y.-S. Chang and Michael Fuchs. *Counting phylogenetic networks with few reticulation vertices: galled and reticulation-visible networks.* Bulletin of Mathematical Biology 86.7 (2024): 76.
2. Y.-S. Chang, Michael Fuchs, Hexuan Liu, Michael Wallner, and Guan-Ru Yu. *Enumerative and distributional results for d-combining tree-child networks.* Advances in Applied Mathematics 157 (2024): 102704.
3. Y.-S. Chang, Michael Fuchs, and Guan-Ru Yu. *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 (2024).

These papers provide an in-depth analysis of the enumeration and properties of the main classes of phylogenetic networks as described in Section 1.1.9. The studies primarily focus on counting these networks and understanding their typical shapes by studying distribution of parameters, such as the number of reticulation nodes; see [11], [12] and [13].

We organize the presentation of results for the (d-combining) tree-child, galled, reticulation-visible, and galled-tree-child phylogenetic networks by topic sections. In each topic, we will cover a brief historical background of known results. And we show our main results, their proofs, and relevant references in each subsection for the main classes. Possible future work will be placed in Chapter 3.

2.1 The maximal number of reticulations

The topic we are discussing in this section is the natural question of determining the range of the number k of reticulations in a network N with fixed number n of leaves. First, it is obvious that the minimal k is 0 for any class of networks and that N in this case is a phylogenetic tree. The maximal k will vary for different classes. In this section, we call the maximal number of reticulations *the sharp bound* defined as follow.

Definition 2.1.1. (The bound for a class of networks) Let $\mathcal{X}$ be a class of networks and $\mathrm{X}_{n,k}$ be the cardinality of $\mathcal{X}_{n,k}$. We say ...

1. $\mathcal{X}$ has an infinite bound if for some n,

$$\mathrm{X}_{n,k} \neq 0 \quad \text{for infinitely many } k.$$

   We also say the bound of $\mathcal{X}$ is infinite or $\mathcal{X}$ has no bound.

2. $\mathcal{X}$ has a (finite) *bound* k_0 if for all n there exists $k_0 = k_0(n)$

$$\mathrm{X}_{n,k} = 0 \quad \text{for all } k > k_0(n).$$

   We also say k_0 is a bound for $\mathcal{X}$.

3. a bound k_0 is *sharp* if for all n,

$$\mathrm{X}_{n,k} \neq 0 \quad \text{for } 0 \leq k \leq k_0(n).$$

Related works. There are many related articles to this topic. We present them ordered according to the publication date.

1. In 2007, Cardona, Rosselló and Valiente [8] showed the bound $n-1$ for tree-child networks; Huson and Klöpper [32] applied a decomposition theorem to show the bound $2(n-1)$ for galled networks.

2. In 2009, Willson [42] found the bound $n-2$ for normal networks by using Cardona et al's methodology.

3. In 2015, Gambette, Gunawan, Labarre, Vialette, and Zhang [25] showed the upper bound $4(n-1)$ for reticulation-visible networks by applying the concept of dummy leaves and removal of reticulation branches. Note that this bound is not sharp; in the same year, Bordewich and Semple [7] and Gunawan and Zhang [31] showed that the upper bound can be reduced to the

optimal one $3(n-1)$. Meanwhile, Gunawan and Zhang re-proved the bound $2(n-1)$ for galled networks by their method and thus reproved the result from Huson and Klöpper in 2009. And they also provided the bounds $3(n-1)$ and $7(n-1)$ for nearly stable and stable-child networks, respectively.

4. In 2018, Gambette, Gunawan, Labarre, Vialette, and Zhang [26] proposed a new methods offering further insight into reticulation-visible networks, genetically stable networks, nearly stable networks, nearly tree-child networks, and galled networks. These findings revealed the new upper bounds for genetically stable networks and nearly tree-child networks which have at most $2(n-1)$ and $n-1$ reticulation nodes.

5. In 2024, Hao-Jun Li proved that the sharp bounds for level-1 and level-k networks are $n-1$ and $k(n-1)$. In addition, he summarized the bounds and proved sharpness for all classes of networks we mentioned above with comparison of three different method; see Hao-Jun Li. *Finite and Infinite Classes of Phylogenetic Networks*, Master Thesis, National Chengchi University, 2024.

We summarize the sharp bounds for the above classes of networks in Table 2.1. One interesting thing is that the sharp bound of normal networks is the only one which is not divisible by $n-1$ in Table 2.1.

$\mathcal{X}$	Abbr.	the sharp bound	proved in [...]
Tree-child	TC	$n-1$	[8]
Galled	GN	$2(n-1)$	[32, 31, 26]
Normal	NN	$n-2$	[42]
Nearly stable	NS	$3(n-1)$	[31, 26]
Stable-child	SC	$7(n-1)$	[31]
Reticulation-visible	RV	$3(n-1)$	[25, 31, 7, 26]
Nearly tree-child	NTC	$n-1$	[26]
Genetically stable	GS	$2(n-1)$	[26]
Level-1	L1	$n-1$	Li
Level-k	Lk	$k(n-1)$	Li

Table 2.1: A summary of results on the maximal number of reticulations with n leaves for many classes of networks.

Definition 2.1.2 (Maximally reticulated networks)**.** A phylogenetic network N is called *maximally reticulated* (in a class of phylogenetic networks $\mathcal{X}$) if $\mathcal{X}$ has sharp bound $k(n)$ and N belongs to $\mathcal{X}_{n,k(n)}$.

For example, the tree-child networks with n leaves and $n-1$ reticulation nodes are maximally reticulated.

A common property seen in the above table is that the maximal k for all considered classes is of linear type. A simple question arises, namely, whether there is a class with has a bound which grows faster than n? In 2017, Semple [38] defined a class of networks whose bound has a quadratic form and used a chain method in the proof.

Definition 2.1.3 (Reticulation chains)**.** Let N be a network. The sequence of nodes $C = t_1r_1t_2r_2\ldots t_kr_kt_{k+1}$ is a *reticulation chain* if the t_j's are tree nodes, the r_j's are reticulations, and (t_j, r_j) and (t_{j+1}, r_j) are edges in N. A reticulation chain is *closed* if $t_1 = t_{k+1}$. A reticulation chain is *overlapping* if there exists $i \neq j$ such that r_i is an ancestor of r_j.

Theorem 20 ([38])**.** Let N be a phylogenetic network with n leaves and k reticulations. If N is stack-free and has no closed or overlapping reticulation chains, then the sharp bound is

$$k \leq n^2 + 3n - 3. \tag{2.1}$$

To be finite or not to be, that is the question. For each class, the bound falls into two categories: infinite or finite. In the above examples, it was always finite. Actually, the bound for some classes may be infinite. The simplest example is the class of general networks.

Example 2.1.1. (The bound for general networks is infinite) Consider two initial networks N_0 and N_1. N_0 is a phylogenetic tree with 2 leaves and N_1 is made from N_0 by adding a reticulation node whose in-coming edges are starting from the two leaf-edges of N_0 and are followed by a leaf with label 3. Next, define a structure that is made of two tree nodes and two reticulations where the tree nodes share the common reticulations as children. Then it has two in-coming edges and two out-going edges. Finally, any number of replicas of the structure can be connected to N_0 and N_1 without altering the number of leaves. Therefore, $\mathrm{PN}_{2,k}, \mathrm{PN}_{3,k} \neq 0$ for all k. By Definition 2.1.1, $\mathcal{PN}$ has no bound.

To prove the infiniteness of a class, it suffices to find a repeatable structure.

A simple observation is that the (in-)finiteness can be inherited via the subset relation.

Remark 9 (Finiteness and subset relationship). According to the po-set in the Figure 1.5, if a class has a finite bound, so do its smaller classes. Conversely, if a class has infinite bound, so do its greater classes.

Here we list all classes with infinite bounds: orchad, valid, FU-stable, tree-sibling and thus by the remark stack-free and tree-based networks. On the other hand, for finite bounds, with the remark, it suffices to show the finitness for reticulation-visible, stable-child, nearly stable and level-k networks.

Bounds and sharpness for main classes. We start with the bound for d-combining tree-child networks and galled networks whose proofs follow the methods from previous papers.

Theorem 21. The number of reticulations of a d-combining tree-child network of size n is at most $n-1$ and the bound is sharp.

Proof. An equivalent definition of tree-child networks is that every vertex v has a tree-path to a leaf; see Proposition 3. Then the root ρ and the k reticulations posses $k+1$ tree-paths to leaves and they are node-wise disjoint, otherwise, there is a common node of two paths and it must be a reticulation node which contradicts the definition of tree-path. Also, the labels of the $k+1$ leaves are distinct. Thus, the bound is given by $k+1 \leq n$.

On the other hand, we prove the sharpness by construction. We start with the bi-combining case. Create n directed paths of length 1 with labels from 1 to n, respectively. Choose one as the root-path, say P_0 and call the others P_j where $1 \leq j \leq n-1$. Let $N = P_0$. For each j, connect the root of P_j with two edges to two non-reticulation-edges of N and again name the resulting graph N. Repeat this until all paths are connected. We have a tree-child networks N with n leaves and $n-1$ reticulations (the original root-path does not contribute to the number of reticulations). A similar argument shows the result for d-combining tree-child networks. ■

Theorem 22. The number of reticulations of a galled network of size n is at most $2(n-1)$ and the bound is sharp.

Proof. Let N be a galled network. The component graph $C(N)$ of N is a (not necessarily binary) phylogenetic trees of size n. Let I collect the internal nodes of $C(N)$. There are $|I|-1+n$ non-root nodes and they can also be counted by the out-going edges of internal nodes. Since each internal node has at least two out-going edges, we have

$$|I|-1+n = \sum_{v\in I} \deg_{out}(v) \geq 2\cdot|I|\,.$$

Then $|I| \leq n-1$. Each internal non-root node in $C(N)$ corresponds to a inner reticulations of N. Therefore, N as at most $n-2$ inner reticulations. On the other hand, each black arrow on a leaf-edge in T corresponds to an outer reticulation of N. Since T has n leaves, N has at most n outer reticulations. To sum up, N in total has at most $(n-2)+n = 2(n-1)$ reticulations.

In order to show that this bound is sharp, let N be a galled network with $\tilde{C}(N)$ a binary phylogenetic tree of size n. There are $n-1$ internal nodes which gives $n-2$ inner reticulations. Attaching black arrows to each leaf-edge will give n outer reticulations. By decompression process, the resulting network is galled with $2n-2$ reticulations. ■

The proof in addition shows that the maximally reticulated galled networks are decompressed from binary phylogenetic trees with a maximal number of arrows.

2.1.1 Reticulation-visible networks

The sharp bound $3(n-1)$ was proved in many papers and we describe more details on their methods. In 2015, Gunawan and Zhang [31] classified the vertices by a weighted function $w(\cdot)$. In 2016, Bordewich and Semple [7, Section 5] used mathematical induction. In 2018, Gambette et al. [26] classified the reticulations into inner or outer ones, the proof is shown below. We provide a simplified proof.

Theorem 23. The number of reticulations of a reticulation-visible network of size n is at most $3(n-1)$ and the bound is sharp.

Proof in [26]. Let N be a reticulation-visible network, $C = C(N)$ be the component graph of N, and c_j be a component. Due to the visibility property, each component c_j contains either a leaf of N or the parents of an inner reticulation node r_0. In the first case, let k_1 be the number of such components. N has n leaves, then $k_1 \leq n$. In the second case, let k_2 be the number of such components. For each component, the parents must be tree nodes, otherwise, say one of them is a reticulation r_1, but then we have a stack of two reticulations and r_1 is not visible. Hence, k_2 is bounded by half of the number of tree nodes, $k_2 \leq t/2$. Since the root component contains no reticulation, we have

$$k+1 = k_1 + k_2 \leq n + \frac{t}{2} = n + \frac{n+k-1}{2} \quad \Longrightarrow \quad k \leq 3(n-1). \quad ■$$

In [11], we gave a new proof which in addition also gives the number of maximally reticulated reticulation-visible networks; see the asymptotic result in Section 2.4.

Alternative Proof. We use the component graph method from Section 1.1.3.3. Since the component graph of a reticulation-visible network is a tree-child network with all vertices of in-degree at most

2 and no reticulation vertex has just one child that is, moreover, a tree vertex, we first fix such a tree-child network $\tilde{C}$ (without arrows on the edges).

The maximal number of reticulation vertices a network decompressed from $\tilde{C}$ can have is obtained by placing arrows on all pendant edges except the ones directly below reticulation vertices. Let $r(\tilde{C})$ be the number of these edges plus the number of reticulation vertices of $\tilde{C}$ plus the number of internal vertices with exactly one incoming edge. Then, our goal is to find those $\tilde{C}$ which maximize $r(\tilde{C})$. Note that $r(\tilde{C})$ remains invariant if we replace vertices with in-degree 2 and out-degree at least 2 by a reticulation vertex followed by a tree vertex and do not count the additional created edge. This is the set of $\tilde{C}$, we will consider in the sequel. (Thus, for the decompressing procedure, we first have to merge reticulation vertices followed by just one tree vertices, if there are any such vertices.)

We start with the following claims.

Claim 1: $r(\tilde{C})$ is maximized only for binary tree-child networks $\tilde{C}$.

Sub-proof. Assume that $\tilde{C}$ has at least one vertex, say v, with out-degree ≥ 3. Replacing v with a vertex of out-degree 2 and attaching to it one of the children of v and a vertex whose children are the remaining children of v clearly gives a new tree-child network with $r(\tilde{C})$ increased by 1 (since we have created a new internal vertex with indegree 1). By iterating this procedure, we end up with a tree-child network which is binary. This proves our claim. ■

Claim 2: The maximum of $r(\tilde{C})$ over the set of all binary tree-child networks is $3n-3$ and this bound is achieved if and only if $\tilde{C}$ is a maximally reticulated binary tree-child network.

Sub-proof. Note that in any binary tree-child network, we have $n + k = t + 2$, where t is the number of tree vertices; see in [23, Section 1]. Thus,

$$r(\tilde{C}) = n + k + t - k = 2n + k - 2.$$

Since $k \leq n-1$ (see Theorem 21) with this bound achieved exactly by the maximally reticulated tree-child networks, the claim follows. ■

Overall, we have proved so far that the maximal number of reticulation vertices of a reticulation-visible network is $3\ell-3$ and this bound is achieved if and only if the component graph of the network is a maximally reticulated binary tree-child network. Finally, the tree vertices of these networks have one child which is a reticulation vertex and one child which is not; see [23, Lemma 1]. Thus, they are replaced by a one-component network which has 2 leaves exactly one of which is below a reticulation vertex. However, the number of these one-component networks is $M_{2,1} = 1$. Consequently, the decompression of every maximally reticulated binary tree-child network gives exactly one reticulation-visible network; see Figure 2.1 for the smallest example (see also [7]) and Figure 2.2 for a larger example. ■

Figure 2.1: **Left**: The smallest example of a maximally reticulated reticulation-visible network with 2 leaves (with the tree-components highlighted). **Right**: The component graph of the network; note that it is a maximally reticulated binary tree-child network.

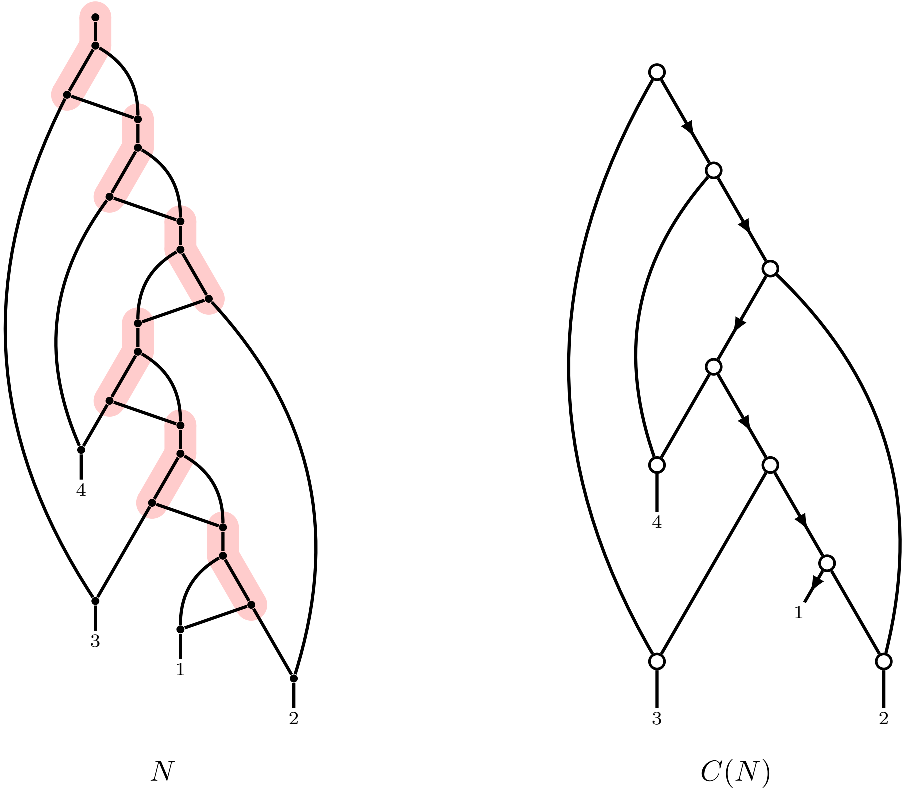


Figure 2.2: A maximally reticulated reticulation-visible network with 4 leaves and 9 reticulation vertices. The component graph is a maximally reticulated binary tree-child network. Note that N is the only network obtained by decompressing $C(N)$.

The maximally reticulated reticulation-visible networks are those networks which are derived from the component graph that in each component contains either one leaf or the parents of exactly one inner reticulation.

2.1.2 Galled tree-child networks

For tree-child networks, we have that the sharp bound equals $n-1$; see Theorem 21. Clearly, this implies that $k \leq n-1$ also for galled tree-child networks. Again this bound is sharp.

Theorem 24. The number of reticulations of a galled tree-child network of size n is at most $n-1$ and the bound is sharp.

Proof. The component graph $\tilde{C}(N)$ of N is a (not necessarily binary) phylogenetic tree of size n; see Example 1.1.2. 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 $\tilde{C}(N)$. For each internal node v, the maximal number of arrows is $\mathrm{deg}_{\text{out}}(v)-1$, because placing arrows on all outgoing edges is not possible since $L_{\mathrm{deg}_{\text{out}}(v),\mathrm{deg}_{\text{out}}(v)}=0$ ($L_{n,k}$ is the number of one-component tree-child networks with n leaves and k reticulations). Thus, the maximal number of reticulation vertices equals

$$\sum_{v}(\mathrm{deg}_{\text{out}}(v)-1)=\sum_{v}\mathrm{deg}_{\text{out}}(v)-(\#\text{ internal nodes of }\tilde{C}(N)), \tag{2.2}$$

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

$$\sum_{v}\mathrm{deg}_{\text{out}}(v)=(\#\text{ internal nodes of }\tilde{C}(N))-1+n$$

which, by plugging into (2.2), gives $n-1$. ■

The proof also reveals the structure of maximally reticulated galled tree-child networks of size n: They are obtained by decompressing component graphs, phylogenetic trees of size n where for each internal node v has one leaf ℓ as child such that all outgoing edges of v except the one leading to ℓ are placed arrows.

2.2 Closed-form expressions for small k and arbitrary n

This topic aims to provide an initial understanding of the classes by examining how their structures differ from that of phylogenetic trees, particularly in terms of how sensitive it is to the number of reticulations. By focusing on networks with only a few reticulations, we seek to understand the extent to which these reticulations influence the overall behavior of the class. Even with the aid of techniques such as the component graph method (Section 1.1.3) or generating functions techniques (Section 1.2.1), the computational complexity of exact counting increases sharply starting from $k \geq 3$. In this section, we provide expressions for $\mathrm{X}_{n,k}$ when the number of reticulations k is small. We first recall what has been proved for the main classes.

For $k = 0$, every class of phylogenetic networks consists only of phylogenetic trees. Thus,

$$\mathrm{PN}_{n,0} = \mathrm{RV}_{n,0} = \mathrm{GN}_{n,0} = \mathrm{TC}_{n,0} = \mathrm{TC}^{(d)}_{n,0} = \mathrm{GTC}_{n,0} = (2n-3)!!.$$

For $k = 1$, the results are given by Cardona and Zhang [9] and Fuchs, Gittenberger and Mansouri [20].

$$\mathrm{PN}_{n,1} = \mathrm{RV}_{n,1} = \mathrm{GN}_{n,1} = \mathrm{TC}_{n,1} = \mathrm{GTC}_{n,1} = n\,(2n-1)!! - 2^{n-1}n!.$$

In addition both of these papers also give $\mathrm{TC}_{n,2}$ for $k = 2$:

$$\mathrm{TC}_{n,2} = \frac{n(n-1)(3n+2)}{3}(2n-1)!! - n(n-1)2^n n!.$$

For $k = 3$, Fuchs et al. [20] give a result for $\mathrm{TC}_{n,3}$:

$$\mathrm{TC}_{n,3} = \frac{(n+2)n(n-1)(n-2)}{3}\,(2n+1)!! - \frac{n(n-1)(n-2)(48n+31)}{3}\,2^{n-4}n!.$$

We will add results for d-combining networks, galled networks, and reticulation-visible networks in the three paragraphs below.

2.2.1 d-combining tree-child networks

In Section 1.1.3.1, the Formula 1.3 makes it possible to derive formulas for $\mathrm{TC}^{(d)}_{n,k}$ for small values of k, d. With small k, e.g. $k = 1$ or 2, we may run through all component graphs in $\mathcal{K}^{(d)}_{k+1}$. In the course of the computation, we see that the results can be presented by a series of generating functions $f_d(z)$ defined below.

Proposition 8. Set

$$f_d(z) := \sum_{m\geq 1} \frac{(2m+d-2)!}{(m-1)!m!} z^m.$$

(i) For $k = 1$,

$$\mathrm{TC}_{n,1}^{(d)} = \frac{n!}{d!2^{n-2}}[z^n]f_d(z)f_0(z). \tag{2.3}$$

(ii) For $k = 2$,

$$\mathrm{TC}_{n,2}^{(d)} = \frac{n!}{d!\,2^{n-3}}\sum_{\ell=0}^{d}\frac{1}{(d-\ell)!\,\ell!}[z^n]f_{2d-\ell}(z)\,f_\ell(z)\,f_0(z) + \frac{n!}{(d!)^2 2^{n-2}}[z^n]\,f_{2d}(z)\,f_0^2(z). \tag{2.4}$$

Proof. We start with $k = 1$. By using (1.3), we have

$$\mathrm{TC}_{n,1}^{(d)} = \frac{1}{2^{n-2}}\sum_{\{B_j\}_{j=1}^2\in\prod_{n,2}}\ \sum_{G\in\mathcal{K}_2^{(d)}}\ \prod_{j=1}^{2}\frac{(2b_j+g_j-2)!}{(b_j-1)!\,\prod_{\ell=1}^{2}(g_{j,\ell})!}.$$

Observe that $\mathcal{K}_2^{(d)}$ contains only two graphs, namely, the graph consisting of a root to which a child is attached by d edges and either the root has label 1 or 2; see the left graph in Figure 2.3. Consequently,

$$\mathrm{TC}_{n,1}^{(d)} = \frac{1}{d!\,2^{n-2}}\sum_{\{B_j\}_{j=1}^2\in\prod_{n,2}}\left(\frac{(2b_1+d-2)!}{(b_1-1)!}\cdot\frac{(2b_2-2)!}{(b_2-1)!}+\frac{(2b_1-2)!}{(b_1-1)!}\cdot\frac{(2b_2+d-2)!}{(b_2-1)!}\right).$$

Summing according to the size of the blocks in the partition $\{B_j\}_{j=1}^2$, we have

$$\begin{aligned}\mathrm{TC}_{n,1}^{(d)} &= \frac{1}{d!\,2^{n-2}}\sum_{\substack{b_1+b_2=n,\\ b_1,b_2\geq 1}}\binom{n-1}{b_1-1}\frac{(2b_1+d-2)!\,(2b_2-2)!+(2b_1-2)!\,(2b_2+d-2)!}{(b_1-1)!\,(b_2-1)!}\\ &= \frac{1}{d!\,2^{n-2}}\sum_{b=1}^{n-1}\binom{n}{b}\frac{(2b+d-2)!\,(2n-2b-2)!}{(b-1)!\,(n-b-1)!}\end{aligned}$$

which translates into (2.3) by using the generating function $f_d(z)$.

Next, we consider $k = 2$. Here, again by (1.3),

$$\mathrm{TC}_{n,2}^{(d)} = \frac{1}{2^{n-3}}\sum_{\{B_j\}_{j=1}^3\in\prod_{n,3}}\ \sum_{G\in\mathcal{K}_3^{(d)}}\ \prod_{j=1}^{3}\frac{(2b_j+g_j-2)!}{(b_j-1)!\,\prod_{\ell=1}^{3}(g_{j,\ell})!}.$$

In contrast to $k = 1$, there are now more possibilities for G: G has three vertices v_1, v_2, v_3 one of which is the root (say v_1); v_1 is connected to v_2 by d edges and to v_3 by ℓ edges with $0 \leq \ell \leq d$; moreover, v_2 is connected to v_3 by $d-\ell$ edges; finally, there are 3 possible labelings if $\ell = d$ and $3! = 6$ possible labelings if $\ell < d$. See the left and right graphs in Figure 2.3, respectively. Thus,

$$\begin{aligned}\mathrm{TC}_{n,2}^{(d)} = \frac{1}{2^{n-3}}\sum_{b_1+b_2+b_3=n}\binom{n}{b_1,b_2,b_3}\frac{1}{3!}\\ \times\left(\sum_{\ell=0}^{d-1}\frac{3!}{d!\ell!(d-\ell)!}\cdot\frac{(2b_1+d+\ell-2)!}{(b_1-1)!}\cdot\frac{(2b_2+d-\ell-2)!}{(b_2-1)!}\cdot\frac{(2b_3-2)!}{(b_3-1)!}\right.\end{aligned}$$

$$+\frac{3}{(d!)^2}\cdot\frac{(2b_1+2d-2)!}{(b_1-1)!}\cdot\frac{(2b_2-2)!}{(b_2-1)!}\cdot\frac{(2b_3-2)!}{(b_3-1)!}\Bigg).$$

From this, by using the generating $f_d(z)$, we obtain

$$\mathrm{TC}_{n,2}^{(d)}=\frac{n!}{d!2^{n-3}}\sum_{\ell=1}^{d}\left(\frac{1}{\ell!(d-\ell)!}[z^n]f_{2d-\ell}(z)f_\ell(z)f_0(z)\right)+\frac{n!}{(d!)^2 2^{n-2}}[z^n]f_{2d}(z)f_0^2(z)$$

which is equivalent to (2.4). ■

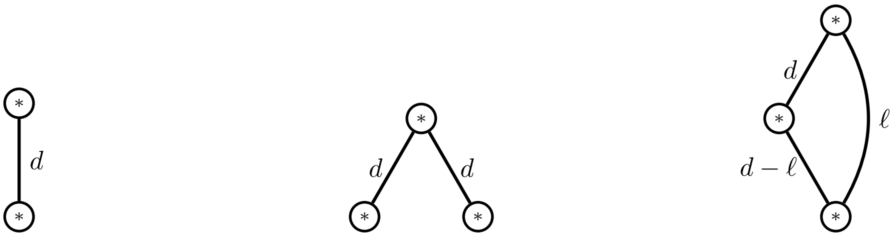


Figure 2.3: **Left**: The component graph in $\mathcal{K}_2^{(d)}$ without the labelings where two labelings can be chosen for the $*$'s; **Middle**: A component graph in $\mathcal{K}_3^{(3)}$ without the labelings. It has 3 different labelings; **Right**: The d component graphs in $\mathcal{K}_3^{(d)}$ without the labelings. Here, $0 \le \ell < d$ and each component graph has $3! = 6$ labelings.

In order to simplify these expressions, we need two technical lemmas. The first gives a recursive way of computing $f_d(z)$.

Lemma 25. Set $X := \sqrt{1-4z}$. Then, $f_d(X)$ can be recursively computed as

$$f_d(X) = (-X^{-1}+X)f'_{d-1}(X) + (d-2)f_{d-1}(X), \qquad (d \ge 1)$$

with initial condition $f_0(X) = 1/2 - X/2$.

Proof. First, for the initial condition, observe that

$$\frac{(2m-2)!}{(m-1)!m!} = \frac{1}{m}\binom{2m-2}{m-1}$$

and thus $f_0(z)$ is the generating function of the (shifted) Catalan numbers. Consequently,

$$f_0(z) = \frac{1-\sqrt{1-4z}}{2} = \frac{1-X}{2}.$$

Next, in order to prove the recurrence, we have

$$\begin{aligned} f_d(z) &= 2\sum_{m\ge1}\frac{m(2m+d-3)!}{(m-1)!m!}z^m + (d-2)\sum_{m\ge1}\frac{(2m+d-3)!}{(m-1)!m!}z^m \\ &= 2zf'_{d-1}(z) + (d-2)f_{d-1}(z). \end{aligned}$$

Writing this in terms of X gives the claimed result. ■

From this lemma, we see that the first few values of $f_d(X)$ are:

$$f_1(X)=\frac{1}{2X}-\frac{1}{2},\quad f_2(X)=\frac{1}{2X^3}-\frac{1}{2X},\quad f_3(X)=\frac{3}{2X^5}-\frac{3}{2X^3}.$$

In particular, it is not hard to see that $f_d(X)$ for $d\geq 2$ is a (finite) linear combinations of terms of the form X^{-m} with m odd.

We need a second technical lemma, which will help us with the extraction of coefficients.

Lemma 26. Let $X := \sqrt{1-4z}$. Then, for odd m

$$[z^n]\,X^{-m}=\frac{1}{\binom{m-1}{(m-1)/2}}\binom{n+(m-1)/2}{(m-1)/2}\binom{2n+m-1}{n+(m-1)/2}$$

and for even m

$$[z^n]\,X^{-m}=4^n\binom{n+(m-2)/2}{(m-2)/2}.$$

Proof. Note that

$$[z^n]\,X^{-m}=\binom{-m/2}{n}(-4)^n=\frac{m(m+2)\cdots(m+2n-2)}{n!}\,2^n.$$

From this both claims follow by standard manipulations. ■

Now, we can simplify (2.3) for small values of d; recall the notation (1.2) of double factorials.

Corollary 2.2.1. The number of tree-child networks with n leaves and 1 reticulation node is

(i) For $d=2$,

$$\mathrm{TC}_{n,1}^{(2)}=n(2n-1)!!-2^{n-1}\,n!.$$

(ii) For $d=3$,

$$\mathrm{TC}_{n,1}^{(3)}=\frac{n(2n+1)}{3}(2n-1)!!-n^2(2n-2)!!.$$

Proof. First, from (2.3) and the initial condition in Lemma 25:

$$\begin{aligned}\mathrm{TC}_{n,1}^{(d)}&=\frac{n!}{d!2^{n-1}}[z^n]f_d(z)-\frac{n!}{d!2^{n-1}}[z^n]f_d(z)X\\&=\frac{(2n+d-2)!}{d!2^{n-1}(n-1)!}-\frac{n!}{d!2^{n-1}}[z^n]f_d(z)X.\end{aligned}\tag{2.5}$$

The second term becomes for $d=2$,

$$[z^n]f_2(z)X=[z^n]\left(\frac{1}{2\,X}-\frac{1}{2\,X^3}\right)X=-\frac{1}{2}[z^n]X^{-2}=-\frac{4^n}{2}$$

and for $d=3$,

$$[z^n]f_3(z)X=[z^n]\left(\frac{3}{2X^5}-\frac{3}{2X^3}\right)X=\frac{3}{2}[z^n]X^{-4}-\frac{3}{2}[z^n]X^{-2}=\frac{3\cdot 4^n}{2}n.$$

Plugging this into (2.5) and standard manipulations give the claimed result. ■

Remark 10. The formula for $d = 2$ is known; see the first paragraph in Section 2.2.

Remark 11. The method of proof gives the following structural result for general d and $n \geq 2$:

$$\mathrm{TC}_{n,1}^{(d)} = \binom{2n+d-2}{d}(2n-3)!! - p_d(n)(2n-2)!!,$$

where $p_d(n)$ is a polynomial of degree $d-1$. Note that $\mathrm{TC}_{1,1}^{(d)} = 0$.

Likewise, we can also simplify (2.4) for small values of d.

Corollary 2.2.2. The number of tree-child networks with n leaves and 2 reticulation node is

(i) for $d = 2$,

$$\mathrm{TC}_{n,2}^{(2)} = n(n-1)\left(\frac{3n+2}{3}(2n-1)!! - (2n)!!\right);$$

(ii) for $d = 3$,

$$\mathrm{TC}_{n,2}^{(3)} = n(n-1)\left(\frac{70n^2+244n+177}{315}(2n+1)!! - \frac{16n+13}{48}(2n+2)!!\right).$$

Remark 12. The formula for $d = 2$ is again known; see again the first paragraph of Section 2.2.

2.2.2 Galled networks

In this section, we derive formulas for galled networks with a small number of reticulation vertices. Due to the tree-like structure of the component graphs of galled networks in Definition 1.1.16, we use generating function techniques and the method of singularity analysis which is described in Section 1.2.2.

Theorem 27. We have

(i) For $k = 2$,

$$\mathrm{GN}_{n,2} = \frac{6n^4+31n^3+30n^2-7n-9}{3}(2n-3)!! - (7n+10)\,2^{n-2}(n+1)!. \tag{2.6}$$

(ii) For $k = 3$,

$$\begin{aligned}\mathrm{GN}_{n,3} = {} & \frac{140n^6+3184n^5+17195n^4+34125n^3+19475n^2-8599n-6090}{105}(2n-3)!! \\ & - \frac{225n^3+2045n^2+5878n+5448}{3}2^{n-5}(n+1)!, \end{aligned} \tag{2.7}$$

Remark 13. The formula for $k = 2$ already appeared in [9] (with typos which we have corrected here).

Since the component graph method is based on one-component galled networks, we start by analyzing their number. First, from the initial values and (1.6), $M_{n,k}$ for small values of k equals

$$\begin{aligned}
M_{n,0} &= (2n-3)!!,\\
M_{n,1} &= (n-1)(2n-3)!!,\\
M_{n,2} &= (2n-1)(n-1)^2(2n-5)!!.
\end{aligned}$$

From this, we observe the following pattern.

Lemma 28. For fixed $k \geq 1$,

$$M_{n,k} = p_k(n)(2(n-k)-3)!!, \qquad (n \geq k),$$

where $p_k(n)$ is a polynomial of degree $2k$ with leading coefficient 2^k.

Proof. This follows by induction on k. First, the claim holds for $k = 1, 2, 3$ and also for $k = 0$ when $n \geq 1$. (This is why we used $(2(n-k)-3)!!$ instead of $(2(n-k+1)-3)!!$.)

Assume now that it holds for $2 \leq k' < k$. Then, by (1.6), we see that it also holds for k where the highest order leading term comes from the first term on the right-hand side of (1.6), namely, it is the highest order leading term of $p_{k-1}(n)$ multiplied by $2n^2$. This proves the claim. ■

Next, we consider the exponential generating function of $M_{n,k}$.

$$M_k(z) := \sum_{n \geq k} M_{n,k} \frac{z^n}{n!}.$$

From the expressions for $M_{n,k}$ for $k = 0, 1, 2$, we obtain that

$$M_0(z) = 1 - \sqrt{1-2z}, \quad M_1(z) = \frac{(1-\sqrt{1-2z})^2}{2\sqrt{1-2z}}$$

and

$$M_2(z) = \frac{(4z^2 - 7z + 4 + \sqrt{1-2z})(1-\sqrt{1-2z})^2}{6(1-2z)^{3/2}}.$$

Also, we have the following asymptotic result for $M_k(z)$ for all $k \geq 1$, where we use the notion of Δ-analyticity of a function $f(z)$ at z_0, i.e., $f(z)$ is an analytic function in a domain of the form

$$\Delta := \{z \,:\, |z| < r,\ |\arg(z - z_0)| > \phi\},$$

for some $r > |z_0|$ and $0 < \phi < \pi/2$.

Lemma 29. For $k \geq 1$, $M_k(z)$ is Δ-analytic for some Δ-domain at $1/2$ and satisfies, as $z \to 1/2$, in the Δ-domain:

$$M_k(z) \sim \frac{(2k-3)!!}{2^k(1-2z)^{k-1/2}}.$$

Proof. Note that

$$P_k(z) := \sum_{\ell \geq k} (2(\ell-k)-3)!! \frac{z^\ell}{\ell!} = \sum_{\ell \geq 0} (2\ell-3)!! \frac{z^\ell}{\ell!} = 2 - \sqrt{1-2z}.$$

From Lemma 28, we see that the function $M_k(z)$ is built from $P_k(z)$ as a linear combination of

$$D^j P_k(z),$$

where $0 \leq j \leq 2k$ and D^j is the j-th iteration of $D := z\frac{d}{dz}$. This shows that $M_k(z)$ is Δ-analytic with singularity expansion, as $z \to 1/2$,

$$M_k(z) \sim 2^k D^{2k} P_k(z) = \frac{(2k-3)!!}{2^k(1-2z)^{k-1/2}}$$

which is the claim. ■

We also need the following closely related generating function

$$F_k(z) := \sum_{n \geq 0} M_{n+k,k} \frac{z^n}{n!},$$

which satisfies the following result.

Corollary 2.2.3. For $k \geq 1$, $F_k(z)$ is Δ-analytic for some Δ-domain at $1/2$ and satisfies, as $z \to 1/2$ in the Δ-domain:

$$F_k(z) \sim \frac{(4k-3)!!}{2^k(1-2z)^{2k-1/2}}. \tag{2.8}$$

Proof. Note that

$$F_k(z) = \frac{d^k}{dz^k} M_k(z). \tag{2.9}$$

Thus, the result follows from the closure properties of singularity analysis; see Section 1.2.2. ■

Remark 14. Note that $k = 0$ has a different singularity expansion since $F_0(z) = 1 - \sqrt{1-2z}$. Also, from the above results for $M_k(z)$ and (2.9).

$$F_1(z) = \frac{z}{(1-2z)^{3/2}} \quad \text{and} \quad F_2(z) = \frac{3 - z + 7z^2 - 4z^3}{(1-2z)^{7/2}}.$$

We now consider general galled networks which are built from one-component galled networks and component graphs. This was described in detail in Section 1.1.3.2 and the recursive method there

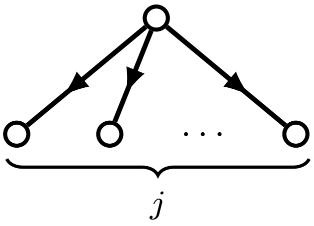


Figure 2.4: The component graph of a galled network with its root having j out-going edges. Each out-going edge provides one counting on the number of reticulaiton, i.e. v^1. Each leaf is replaced by other galled network, i.e. $G(z, v)$.

can be translated into generating functions (because the component graphs are trees). Since we are interested in reticulation vertices, we need to keep track of them. We therefore consider the generating function

$$G(z, v) := \sum_{n\geq 0}\sum_{k\geq 0} \mathrm{GN}_{n,k}\, v^k \,\frac{z^n}{n!}.$$

Then, we have the following result.

Proposition 9. We have,

$$G(z, v) = \sum_{j\geq 0} F_j(z) \frac{(vG(z, v))^j}{j!}. \tag{2.10}$$

Proof. According to the decompressing procedure of component graphs of galled networks described in Section 1.1.3.2, every galled network is obtained from a one-component galled network (this is the term $F_j(z)$) with the leaves below the reticulation vertices replaced by an (unordered) sequence of galled networks (this is the term $G(z, v)^j/j!$; v^j keeps track of the reticulation vertices); see Figure 2.4. (Note that the relabeling of the leaves of the one-component galled network which replaces the root of the component graph described in the previous section is "automatically" obtained from the product of the exponential generating functions which corresponds in combinatorics to the product of labeled combinatorial classes; see, e.g., Chapter II in [17] were this is explained in detail.) ■

The exponential generating function for the number of galled networks with k reticulation vertices, i.e.,

$$E_k(z) := \sum_{n\geq k} \mathrm{GN}_{n,k} \frac{z^n}{n!} \tag{2.11}$$

is obtained from $G(z, v)$ by partial differentiation

$$E_k(z) = \frac{1}{k!}\frac{\partial^k}{\partial v^k} G(z, v)\Big|_{v=0}.$$

From Proposition 9, we obtain a recurrence.

Lemma 30. For $k \geq 1$,

$$E_k(z) = \sum_{j=1}^{k} \frac{F_j(z)}{j!} \sum_{\ell_1+\cdots+\ell_j=k-j} E_{\ell_1}(z)\cdots E_{\ell_j}(z). \tag{2.12}$$

Proof. Differentiating (2.10) k-times gives

$$\begin{aligned} E_k(z) &= \frac{1}{k!}\sum_{j=1}^{k} \frac{F_j(z)}{j!} \frac{\mathrm{d}^k}{\mathrm{d}v^k} (vG(z,v))^j \Big|_{v=0} \\ &= \frac{1}{k!}\sum_{j=1}^{k} \frac{F_j(z)}{j!} \binom{k}{j} j! \frac{\mathrm{d}^{k-j}}{\mathrm{d}z^{k-j}} G(z,v)^j \\ &= \sum_{j=1}^{k} \frac{F_j(z)}{j!(k-j)!} \sum_{\ell_1+\cdots+\ell_j=k-j} \binom{k-j}{\ell_1!\cdots\ell_j!} \ell_1! E_{\ell_1}(z)\cdots \ell_j! E_{\ell_j}(z). \qquad \blacksquare \end{aligned}$$

The claim is obtained by simplifying this.

Proof of Theorem 27. Note that $E_0(z) = 1 - \sqrt{1-2z}$. Then, from (2.12) and Remark 14,

$$E_1(z) = F_1(z)E_0(z) = \frac{z(1-\sqrt{1-2z})}{(1-2z)^{3/2}}.$$

Moreover, by using (2.12) and Remark 14 once again,

$$\begin{aligned} E_2(z) &= F_1(z)E_1(z) + \frac{F_2(z)E_0(z)^2}{2} \\ &= \frac{12z^4 - 18z^3 + 17z^2 - 36z + 21 + (12z^3 - 10z^2 + 15z - 21)\sqrt{1-2z}}{3(1-2z)^{7/2}}. \end{aligned}$$

Extracting coefficients gives the claimed result (2.6) for $\mathrm{GN}_{n,2}$.

As for the result (2.7) of $\mathrm{GN}_{n,3}$, the same method can be used, only the resulting computation is more tedious (and therefore best done with mathematical software, e.g., Maple). $\blacksquare$

2.2.3 Reticulation-visible networks

According to the Section 1.1.3.3, we may derive the formula of $\mathrm{RV}_{n,k}$ for small k. For $k = 2, 3$, we run though all component graphs, which are (general) tree-child networks defined in Definition 1.1.18, and insert the dup-trees $\mathrm{M}_{n,k}$ into each node of it. The combinatorial step corresponds to the generating functions $F_k(z)$ defined in the last Section which help us obtain the results below.

Theorem 31. We have

(i) For $k = 2$,

$$\mathrm{RV}_{n,2} = \frac{6n^4 + 7n^3 + 6n^2 - n - 3}{3}(2n-3)!! - \left(2n^2 + 2n + 1\right) 2^{n-1} n!. \tag{2.13}$$

(ii) For $k = 3$,

$$\mathrm{RV}_{n,3} = \frac{4n^6 + 20n^5 + 33n^4 - 32n^3 - 76n^2 + 12n + 12}{3}(2n-3)!! \\ - \frac{48n^4 + 175n^3 + 99n^2 - 262n - 264}{3} 2^{n-4} n!. \quad (2.14)$$

Let N be a reticulation-visible network. We remove the leaves and their pendant edges from $C(N)$ except those whose pendant edge has an arrow on it. Then, the resulting component graph is a (unlabeled) rooted simple DAG with every non-root vertex of in-degree 2. (Note that an edge with an arrow on it is actually a double edge and thus counts as two edges.) Moreover, this DAG has exactly $k+1$ vertices.

Using these DAGs and the decompression procedure explained in Section 1.1.3.3, we can derive the exponential generating function of $\mathrm{RV}_{n,k}$.

Proposition 10. Let $\mathcal{D}_m$ be the set of (unlabelled) rooted DAGs with m vertices in which non-root vertices have in-degree 2 and double edges between two vertices are allowed. Then, for given k,

$$\sum_{n\geq 1} \mathrm{RV}_{n,k} \frac{z^n}{n!} = \sum_{G\in\mathcal{D}_{k+1}} \frac{1}{\mathrm{m}(G)} \prod_{v\in G} \sum_{n\geq n_0} M_{n+\deg_{\mathrm{out}}(v), c_1(v)} \frac{z^n}{n!}, \quad (2.15)$$

where

1. $\mathrm{m}(G)$ counts symmetries in G,
2. the product runs over all vertices in G, and
3. the final sum on the right-hand side is the generating function with respect to the number of labeled leaves which are attached to v, where
4. $c_1(v)$ is the number of children of v with an arrow on their edges, and
5. n_0 is 0 or 1 according to whether $c_1(v) > 0$ or $c_1(v) = 0$, respectively.

Now, to find $\mathrm{RV}_{n,2}$, we start from the DAGs in the set $\mathcal{D}_3$ which are listed in Figure 2.5. Note that $m(A) = 2$ (due to the symmetry about the root), $m(B) = m(C) = 1$, and for each vertex v, the exponential generating function $f_v(z)$ from Proposition 10 equals (here, $X = \sqrt{1-2z}$):

$$f_{a_1}(z) = F_2(z) = \sum_{n\geq 0} M_{n+2,2} \frac{z^n}{n!} = \frac{15}{4} X^{-7} - \frac{3}{2} X^{-5} + \frac{1}{4} X^{-3} + \frac{1}{2} X^{-1};$$

$$f_{a_2}(z) = f_{a_3}(z) = f_{b_3}(z) = f_{c_3}(z) = F_0(z) = \sum_{n\geq 1} M_{n,0} \frac{z^n}{n!} = 1 - X;$$

$$f_{b_1}(z) = f_{b_2}(z) = F_1(z) = \sum_{n\geq 0} M_{n+1,1} \frac{z^n}{n!} = \frac{1}{2} X^{-3} - \frac{1}{2} X^{-1};$$

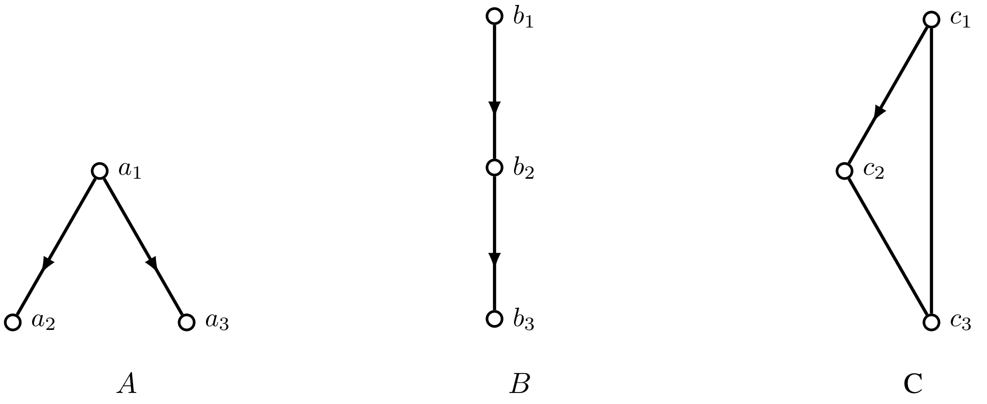


Figure 2.5: The 3 DAGs from the set $\mathcal{D}_3$ where for convenience, we have labeled the vertices. Call the reticulation vertices of the decompressed reticulation-visible networks r_1 and r_2. Then, A gives all the networks where r_1 and r_2 are not in an ancestor-descendant relationship; B gives all networks where r_1 is above r_2 but they are not in a tree-cycle; and C gives all networks where r_1 is in the tree cycle of r_2.

$$\begin{aligned} f_{c_1}(z) &= \sum_{n\geq 0} M_{n+2,1}\,\frac{z^n}{n!} = \frac{3}{2}X^{-5} - \frac{1}{2}X^{-3}; \\ f_{c_2}(z) &= \sum_{n\geq 1} M_{n+1,0}\,\frac{z^n}{n!} = X^{-1} - 1, \end{aligned}$$

Thus, from Proposition 10, we have for the exponential generating function of $\mathrm{RV}_{n,2}$,

$$\begin{aligned} \frac{1}{2}\prod_{j=1}^{3} f_{a_j}(z) + \prod_{j=1}^{3} f_{b_j}(z) + \prod_{j=1}^{3} f_{c_j}(z) &= \frac{(3 - z + 7z^2 - 4z^3)(1 - z - \sqrt{1-2z})}{(1-2z)^{7/2}} \\ &= \frac{(1-X)^2(15 - 6X^2 + X^4 + 2X^6)}{8X^7}. \end{aligned}$$

What is left is to extract coefficients which can be done with the following lemma.

Lemma 32. We have, for n large enough,

$$[z^n]X^d = \begin{cases} 0, & \text{if } d \geq 0 \text{ and } d \text{ is even, } k := \frac{d}{2}; \\ (-1)^{k+1}(2k+1)!!\,\dfrac{(2n-2k-3)!!}{n!}, & \text{if } d \geq 0 \text{ and } d \text{ is odd, } k := \frac{d-1}{2}; \\ 2^n \dbinom{n+k-1}{k-1}, & \text{if } d < 0 \text{ and } d \text{ is even, } k := -\frac{d}{2}; \\ \dfrac{1}{(2k-3)!!}\,\dfrac{(2n+2k-3)!!}{n!}, & \text{if } d < 0 \text{ and } d \text{ is odd, } k := -\frac{d-1}{2}. \end{cases}$$

Applying the lemma gives the following result.

$$\begin{aligned} \mathrm{RV}_{n,2} &= [z^n]\frac{(1-X)^2(15 - 6X^2 + X^4 + 2X^6)}{8X^7} \\ &= \frac{6n^4 + 7n^3 + 6n^2 - n - 3}{3}(2n-3)!! - \left(2n^2 + 2n + 1\right)\, 2^{n-1}n!. \end{aligned}$$

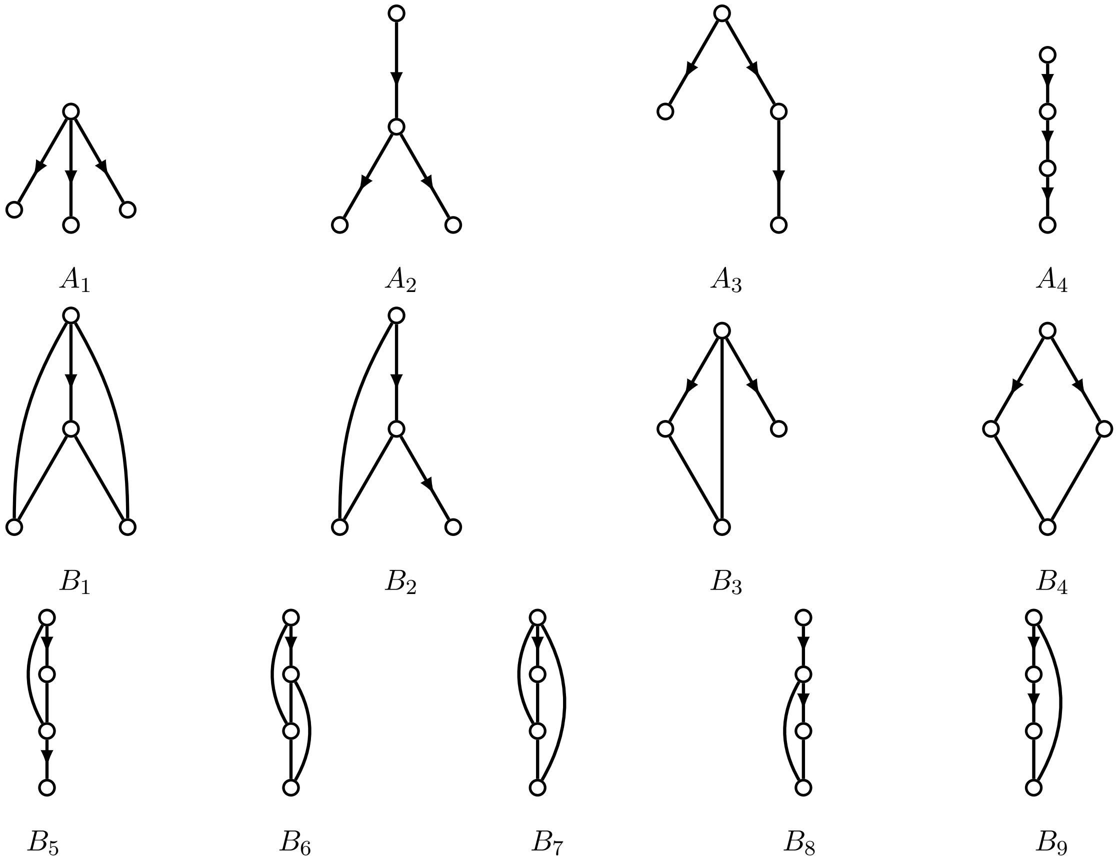


Figure 2.6: The 13 DAGs from the set $\mathcal{D}_4$.

We next consider $k = 3$. Here the set $\mathcal{D}_4$ has 13 elements; see Figure 2.6. Note that these DAGs fall into two types: the A_j's are tree structures and the B_j's are not. (The latter generate the reticulation-visible networks which are not galled networks.) Next, we note that $m(A_1) = 6, m(A_2) = m(B_1) = m(B_4) = 2$ and the values are 1 in all other cases. Thus, by Proposition 10, we obtain for the exponential generating function $f_A(z)$ arising by the A_j's,

$$f_A(z) = \frac{4z^3(29 + 12z + 29z^2 - 37z^3 + 36z^4 - 14z^5)}{(1-2z)^{11/2}\,(1+\sqrt{1-2z})^3} + \frac{6z^3(3 - z + 7z^2 - 4z^3)}{(1-2z)^5\,(1+\sqrt{1-2z})^2} + \frac{2z^4}{(1-2z)^{9/2}\,(1+\sqrt{1-2z})},$$

and for the exponential generating function $f_B(z)$ arising from the B_j's,

$$f_B(z) = \frac{(1-X)^2(258 - 105X - 153X^2 - 16X^3 + 26X^4 + 7X^5 + 3X^6 - 2X^7 - 2X^8)}{8\,X^{10}}.$$

Then, by extracting coefficients of $f_A(z) + f_B(z)$ with Lemma 32,

$$\mathrm{RV}_{n,3} = [z^n]f_A(z) + f_B(z) = \frac{4n^6 + 20n^5 + 33n^4 - 32n^3 - 76n^2 + 12n + 12}{3}(2n-3)!! - \frac{48n^4 + 175n^3 + 99n^2 - 262n - 264}{3}\,2^{n-4}n!.$$

2.2.4 Galled tree-child networks

Deriving closed form-expressions of the number of galled tree-child networks for a small number of reticulations is a direct application of the generating function techniques on galled networks in Section 2.2.2.

Theorem 33. We have

(i) For $k = 2$,

$$\mathrm{GTC}_{n,2} = \frac{n(3n^2+11n+4)}{3}\,(2n-1)!! - n(7n+5)\,2^{n-2}n!. \tag{2.16}$$

(ii) For $k = 3$,

$$\begin{aligned}\mathrm{GTC}_{n,3} = {} & \frac{n\,(35\,n^3+601\,n^2+1018\,n+446)}{105}\,(2n+1)!! \\ & - \frac{n\,(915\,n^3+4262\,n^2+4713\,n+1630)}{3}\,2^{n-7}\,n!.\end{aligned} \tag{2.17}$$

Remark 15. These formulas for galled tree-child networks fora small number of reticulations are first presented here.

The entire proof follows exactly the same framework as that of Section 2.2.2. We define the related generating functions $\tilde{F}_k(z), \tilde{G}(z,v)$ and $\tilde{E}_k(z)$ by

$$\tilde{F}_k(z) := \sum_{n\geq 0} \mathrm{L}_{n+k,k}\,\frac{z^n}{n!}, \quad \tilde{G}(z,v) := \sum_{n\geq 0}\sum_{k\geq 0} \mathrm{GTC}_{n,k}\,v^k\,\frac{z^n}{n!}, \quad \text{and} \quad \tilde{E}_k(z) := \sum_{n\geq k} \mathrm{GTC}_{n,k}\frac{z^n}{n!}.$$

First, we consider the generating function $\tilde{F}_k(z)$ for the components of galled tree-child networks; see the component graph method in Section 1.1.3.4. Recall that $\mathrm{L_{n,k}}$ counts the number of one-component tree-child networks with n leaves and k reticulations where the labels of the leaves below the reticulation vertices are $\{1,\ldots,k\}$ and they are the components for galled tree-child networks; see Definition 1.1.21. Applying Theorem 9 directly, we have $\tilde{F}_k(z)$.

Remark 16. For $k = 0$, $\tilde{F}_0(z)$ collects the phylogenetic trees only, then $\tilde{F}_0(z) = 1-\sqrt{1-2z}$. For $k = 1$ and 2,

$$\tilde{F}_1(z) = \frac{z}{(1-2z)^{3/2}} \quad \text{and} \quad \tilde{F}_2(z) = \frac{3z(2+z)}{(1-2z)^{7/2}}.$$

Note that $F_k(z)$ equals to $\tilde{F}_k(z)$ for $k \leq 1$ but not from $k \geq 2$, where $F_k(z)$'s are defined in Section 2.2.2.

Second, we consider general galled tree-child networks which are built from one-component galled networks and component graphs. This was described in detail in Section 1.1.3.4 and the recursive method there can also be translated into generating functions (because the component graphs are trees). We therefore consider the generating function $\tilde{G}(z,v)$. Then, we have the following result.

Proposition 11. We have,

$$\tilde{G}(z,v) = \sum_{j\geq 0} \tilde{F}_j(z) \frac{(v\tilde{G}(z,v))^j}{j!}. \tag{2.18}$$

Proof. By the decompression process of the component graph method in Section 1.1.3.4, we choose a component graph of galled tree-child networks, whose root has j out-going edges. Then attach a component on the root, that is, $\tilde{F}_j(z)$ and for each outgoing edge of the root, we joint it with a galled tree-child network, i.e., the term $(v\tilde{G}(z,v))^j/j!$, where v counts the number of reticulations. Hence we have a galled tree-child network and it derives the formula. ■

Third, the exponential generating function, $\tilde{E}_k(z$, for the number of galled tree-child networks with k reticulation vertices is obtained from $\tilde{G}(z,v)$ by partial differentiation

$$\tilde{E}_k(z) = \frac{1}{k!}\frac{\partial^k}{\partial v^k}\tilde{G}(z,v)\Big|_{v=0}.$$

From Proposition 11, we obtain a recurrence.

Lemma 34. For $k \geq 1$,

$$\tilde{E}_k(z) = \sum_{j=1}^{k} \frac{\tilde{F}_j(z)}{j!} \sum_{\ell_1+\cdots+\ell_j=k-j} \tilde{E}_{\ell_1}(z)\cdots\tilde{E}_{\ell_j}(z). \tag{2.19}$$

Proof. Differentiating (2.18) k-times gives the formula. ■

Finally, we can now prove Theorem 33.

Proof of Theorem 33. Note that $\tilde{E}_0(z) = 1-\sqrt{1-2z}$. From (2.19) and Remark 16,

$$\tilde{E}_1(z) = \tilde{F}_1(z)\,\tilde{E}_0(z) = \frac{z(1-\sqrt{1-2z})}{(1-2z)^{3/2}}.$$

Moreover, by using (2.19) and Remark 16 once again,

$$\tilde{E}_2(z) = \tilde{F}_1(z)\,\tilde{E}_1(z) + \frac{\tilde{F}_2(z)\,\tilde{E}_0(z)^2}{2} = -\frac{z\left(z^2+4z-6+(2z+6)\sqrt{1-2z}\right)}{(1-2z)^{7/2}}, \quad \text{and}$$

$$\begin{aligned}\tilde{E}_3(z) &= \tilde{F}_1(z)\,\tilde{E}_2(z) + \frac{\tilde{F}_2(z)\,\tilde{E}_1(z)\,\tilde{E}_0(z)}{2} + \frac{\tilde{F}_3(z)\,\tilde{E}_0(z)^3}{6} \\ &= -\frac{z\left(35z^3+221z^2-78z-60-(11z^3+53z^2-138z-60)\sqrt{1-2z}\right)}{(1-2z)^{11/2}}.\end{aligned}$$

Extracting coefficients by using Lemma 32 gives the claimed result (2.16) and (2.17) for $\mathrm{GTC}_{n,2}$ and $\mathrm{GTC}_{n,3}$, respectively. ■

2.3 Exact counting for any k and any n

In Section 1.1.3, we provided exact counting formulas for any k and n for the main classes which have been used for generating tables in previous works; see [9] for tree-child networks, [29] for galled networks, and [13] for galled tree-child networks. In this section, we present concise arithmetic expressions for one-component d-combining tree-child networks, which extend the bi-combining case; see below. We also derive a recurrence which allows computation in polynomial time for d-combining tree-child networks by connecting them to words. The remaining main classes still require further investigation.

A known exacting counting result for one-component tree-child networks were proved in [9] and [19]. 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,

$$\mathrm{OTC}_{n,k} = \binom{n}{k} L_{n,k} \tag{2.20}$$

and

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

Proposition 12 ([9, 19]). (i) We have,

$$\mathrm{OTC}_{n,k} = \binom{n}{k} \frac{(2n-2)!}{2^{n-1}(n-k-1)!}. \tag{2.21}$$

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

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

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

2.3.1 One-component d-combining tree-child networks

We give now the generalization of this formula (2.21) to d-combining networks. Since both Theorem 50 and Theorem 59 are derived from this result, we give two different proofs of it; one below and one in Remark 17 succeeding the proof. The first proof gives a direct combinatorial interpretation of the closed form, while the second one uses a recursion on d and eventually relies on the formula for $\mathrm{OTC}_{n,k}^{(2)}$ in [9].

Theorem 35. The number of one-component d-combining tree-child networks with n leaves and k reticulation nodes for $0 \leq k \leq n-1$ is given by

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

Proof. We give the closed form for one-component networks a direct combinatorial interpretation. For this purpose, we construct all one-component networks as follows:

1. Start with a phylogenetic tree with $n-k$ leaves, i.e., without reticulation nodes.
2. Place dk unary nodes along the $2(n-k)-1$ edges. (Place multiple unary node on the same side is allowed.) Hence, there are $\binom{2(n-k)+dk-2}{dk}$ possibilities.
3. Attach k reticulation nodes that are each connected to d of the unary nodes. Attach to each reticulation node a leaf and label it from 1 to k in the order of creation. Thus, in this step there are $\binom{dk}{d,d,\dots,d}$ possibilities, where the d appears k times.
4. Relabel the n leaves respecting the orders of the $n-k$ initial and k newly created leaves. Hence, this can be done in $\binom{n}{k}$ ways.

Each in that way created network is different and all one-component networks satisfy such a decomposition. Multiplying the number of possibilities for each step gives the claimed formula:

$$\mathrm{OTC}^{(d)}_{n,k} = \mathrm{OTC}^{(d)}_{n-k,0} \cdot \binom{2(n-k)+dk-2}{dk} \cdot \binom{dk}{d,d,\dots,d} \cdot \binom{n}{k}. \tag{2.22}$$

The result follows now by the fact that

$$\mathrm{OTC}^{(d)}_{n-k,0} = (2(n-k)-3)!! = \frac{(2n-2k-2)!}{2^{n-k-1}(n-k-1)!} \tag{2.23}$$

since $\mathrm{OTC}^{(d)}_{n-k,0}$ is the number of phylogenetic trees with $n-k$ leaves. ■

Remark 17 (Alternative way to prove.). A second way to obtain $\mathrm{OTC}^{(d)}_{n,k}$ proceeds by constructing $\mathrm{OTC}^{(d)}_{n,k}$ from $\mathrm{OTC}^{(d-1)}_{n,k}$ as follows. First, let a $(d-1)$-combining one-component tree-child network with n leaves and k reticulation nodes be given. Then, there are $2n+(d-3)k-1$ edges whose end points are not reticulation nodes (i.e. candidate edges). We add k different nodes (each node being the d-th parent of a reticulation node) to these edges. Overall there are

$$\prod_{i=0}^{k-1}(2n+(d-3)k-1+i)$$

ways to do this. Now, if we assign the first one of the k nodes to the first reticulation node, the second one to the second reticulation node, etc., we obtain every one-component d-combining tree-child network with n leaves and k reticulation nodes exactly d^k times. Thus,

$$\frac{\mathrm{OTC}_{n,k}^{(d)}}{\mathrm{OTC}_{n,k}^{(d-1)}} = \frac{\prod_{i=0}^{k-1}(2n+(d-3)k-1+i)}{d^k}.$$

From this, we can get the result for one-component d-combining networks by iteration and using the known result for the bi-combining case from [9].

2.3.2 d-combining tree-child networks

We next turn to general tree-child networks. Here, in contrast to one-component tree-child networks in Theorem 35, we do not have an closed form for $\mathrm{TC}_{n,k}^{(d)}$. However, we introduce a way of encoding these networks by words and this encoding leads to a recursive method for computing $\mathrm{TC}_{n,k}^{(d)}$. Using this method, the values of this sequence for small n, k and d can be computed easily.

Definition 2.3.1. Let $\mathcal{C}_{n,k}^{(d)}$ denote the class of words consisting of the letters $\{\omega_1, \dots, \omega_n\}$ in which k letters occur $d+1$ times and $n-k$ letters occur 2 times and which satisfy the following condition: In every prefix of a word, either a letter has not occurred more than $d-2$ times, or, if it has, then the number of occurrences of ω_i is at least as large as the number of occurrences of ω_j for all $j > i$. Here, for the letters appearing only 2 times, we treat the 0th, 1st, and 2nd occurrence as the $(d-1)$st, d-th, and $(d+1)$st occurrence, respectively. Denote $c_{n,k}^{(d)}$ the cardinality of $\mathcal{C}_{n,k}^{(d)}$.

Remark 18. For $k = n$, we recover the words from [10, Definition 12] which in turn generalized the words from [23, Definition 2] from the bi-combining case.

Remark 19. The words $\mathcal{C}_{n,k}^{(d)}$ can also be encoded by Young tableaux with walls, where a wall between two cells indicates that there are no order constraints between the respective entries; see [4, 5]. The i-th column is associated to the i-th letter ω_i and its size is equal to the number of occurrences of ω_i. Therefore, the corresponding Young tableaux consist of k columns with $d+1$ cells and $n-k$ columns with 2 cells placed next to each other in such a way that the top cells are all side-by-side. We put vertical walls between all cells in rows three to $d+1$. Finally, the cells are filled in increasing order from left to right and bottom to top with the numbers $\{1, 2, \dots, 2n+k(d-1)\}$; see an example in Figure 2.7. The bijection is as follows: Read the words from left to right. A letter ω_i at position j indicates that the value j is put into column i. This generalizes the class analyzed in [5, Section 4] consisting of only one row with walls.

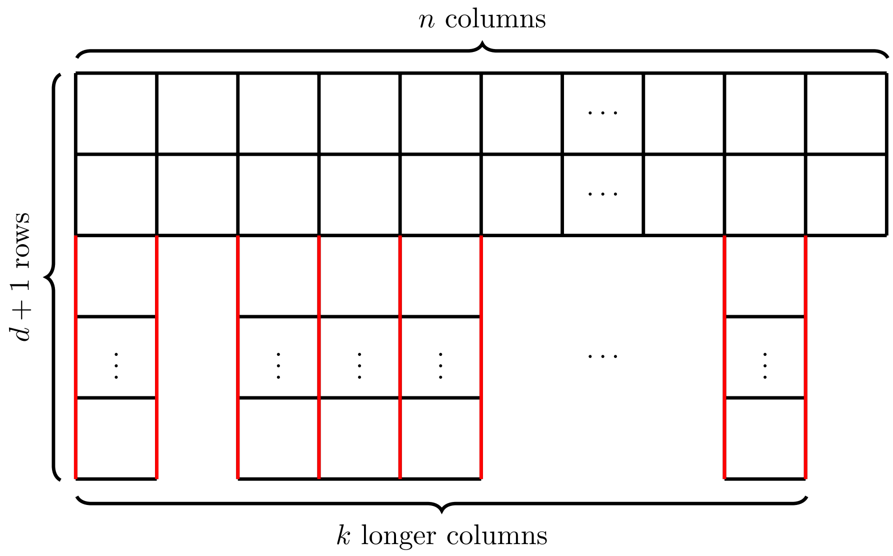


Figure 2.7: A corresponding Young tableau without filled cells for the words in $\mathcal{C}_{n,k}^{(d)}$.

The next result connects tree-child networks and the words from Definition 2.3.1.

Theorem 36. Let $c_{n,k}^{(d)} := |\mathcal{C}_{n,k}^{(d)}|$. Then,

$$\mathrm{TC}_{n,k}^{(d)} = \frac{n!}{2^{n-k-1}} c_{n-1,k}^{(d)}. \tag{2.24}$$

Remark 20. For $d = 2$, [36] proposed an encoding of tree-child networks with n leaves and k reticulation nodes by a (slightly) different class of words. This encoding led to a similar formula for their numbers. However, whereas the formula from [36] is just a conjecture, we provide a rigorous proof of our result.

Before proving the above theorem, we recall some concepts and provide generalizations of results from [23].

Lemma 37. A tree-child network with n leaves and k reticulation nodes has exactly $n-k-1$ free tree nodes and thus $2(n-k-1)$ free edges.

Note that exactly the same result holds in the bi-combining case; see [23, Lemma 1]. For an example demonstrating the last lemma see Figure 2.8.

Proof. Let N be a tree-child network in $\mathcal{TC}_{n,k}^{(d)}$ and N has $n+(d-1)k-1$ tree nodes. Now, observe that the d parents of a reticulation node are (1) tree nodes which are not free due to the stack-free property; (2) different for all reticulation nodes because of tree-sibling property; (3) all non-free tree nodes. Thus, the number of free tree nodes equals

$$(n+(d-1)k-1) - dk = n-k-1. \quad \blacksquare$$

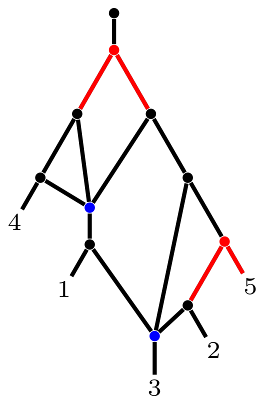


Figure 2.8: A 3-combining tree-child network with 5 leaves and 2 reticulation nodes (blue nodes). Thus, the number of free tree nodes is 2 (red nodes) and there are 4 free edges (red edges).

In [23, Lemma 4.], it carried over to the d-combining case: Every maximally reticulated tree-child network admits a (unique) decomposition into *path-components*, which are maximal paths that start at a node and end at a leaf, where all its intermediate nodes are tree nodes. (See Figure 4 in [23] for an example.)

Now, we are ready to prove Theorem 36.

Proof of Theorem 36. Let N be a tree-child network with n leaves and k reticulation nodes. From Lemma 37, we know that N has $n-k-1$ free tree nodes, which each have 2 free edges. We choose one free edge from each of these pairs of free edges. Then, the number of all networks N together with choices of free edges equals

$$2^{n-k-1}\,\mathrm{TC}_{n,k}^{(d)} = n!\,c_{n-1,k}^{(d)} \tag{2.25}$$

where the equality with the right-hand side is our claim. Thus, in order to prove the claim it suffices to find a bijection between the networks N and a choice of free edges to tuples consisting of a permutation and a word from $\mathcal{C}_{n-1,k}^{(d)}$.

In order to explain this bijection, fix a network N and a choice of free edges For every free tree node, insert $d-1$ nodes on its chosen free edge and a single node on its other free edge. Connect the $d-1$ nodes to the single node, thereby turning this single node into a new reticulation node. Notice that the resulting network is a maximally reticulated tree-child network. The rest of the proof proceeds now as the proof of [23, Proposition 2].

First, we index the path-components as follows: the path-component containing the root gets index 0. Consider all other path-components (which must start with a reticulation node) with parents of the reticulation node already in indexed path components. Index them according to the largest index of the path-component which contains the parents, or, in case of equal largest indices according to whose last parent is encountered first when reading the nodes in the path component from the starting node to the leaf. Repeat this until all path-components have been indexed; see Figure 2.9-(c). Note that one path-component starts with the root, and $n-1$ path-components start with a reticulation node.

Now, label the reticulation node and all its parents of the path-component with index 1 by a, of the path component with index 2 by b, etc. Next, for each chosen free edge, treat the added $d-1$ nodes (which all have the same label) as a single node; see Figure 2.9-(c). Then, a word from $\mathcal{C}_{n-1,k}^{(d)}$ is obtained by reading the labels of nodes of the path-components in increasing order. Finally, a permutation is obtained by reading the labels of each leaf of the path-components in the above order; see Figure 2.9-(d).

Overall, this gives a bijection between N and a fixed choice of free edges for every free tree node to a word from $\mathcal{C}_{n-1,k}^{(d)}$ and a permutation of length n. Thus, we have proved (36) and the claim. ■

Theorem 36 reduces the problem of counting $\mathrm{TC}_{n,k}^{(d)}$ to that of $c_{n,k}^{(d)}$. For the latter sequence, we have the following relation to the sequence $b_{n,k,m}^{(d)}$ which can be computed recursively.

Definition 2.3.2. Denote $b_{n,k,m}^{(d)}$ counts the number of words from $\mathcal{C}_{n,k}^{(d)}$ which having a suffix $\omega_n\omega_m\omega_{m+1}\dots\omega_{n-1}\omega_n$ with $1 \le m \le n$.

Thus, $b_{n,k,m}^{(d)}$ is a refinement of $c_{n,k}^{(d)}$.

Proposition 13. Let $c_{n,k}^{(d)} := |\mathcal{C}_{n,k}^{(d)}|$. Then,

$$c_{n,k}^{(d)} = \sum_{m\ge 1} b_{n,k,m}^{(d)},$$

where $b_{n,k,m}^{(d)}$ $(1 \le m \le n, 0 \le k \le n)$ can be recursively computed as

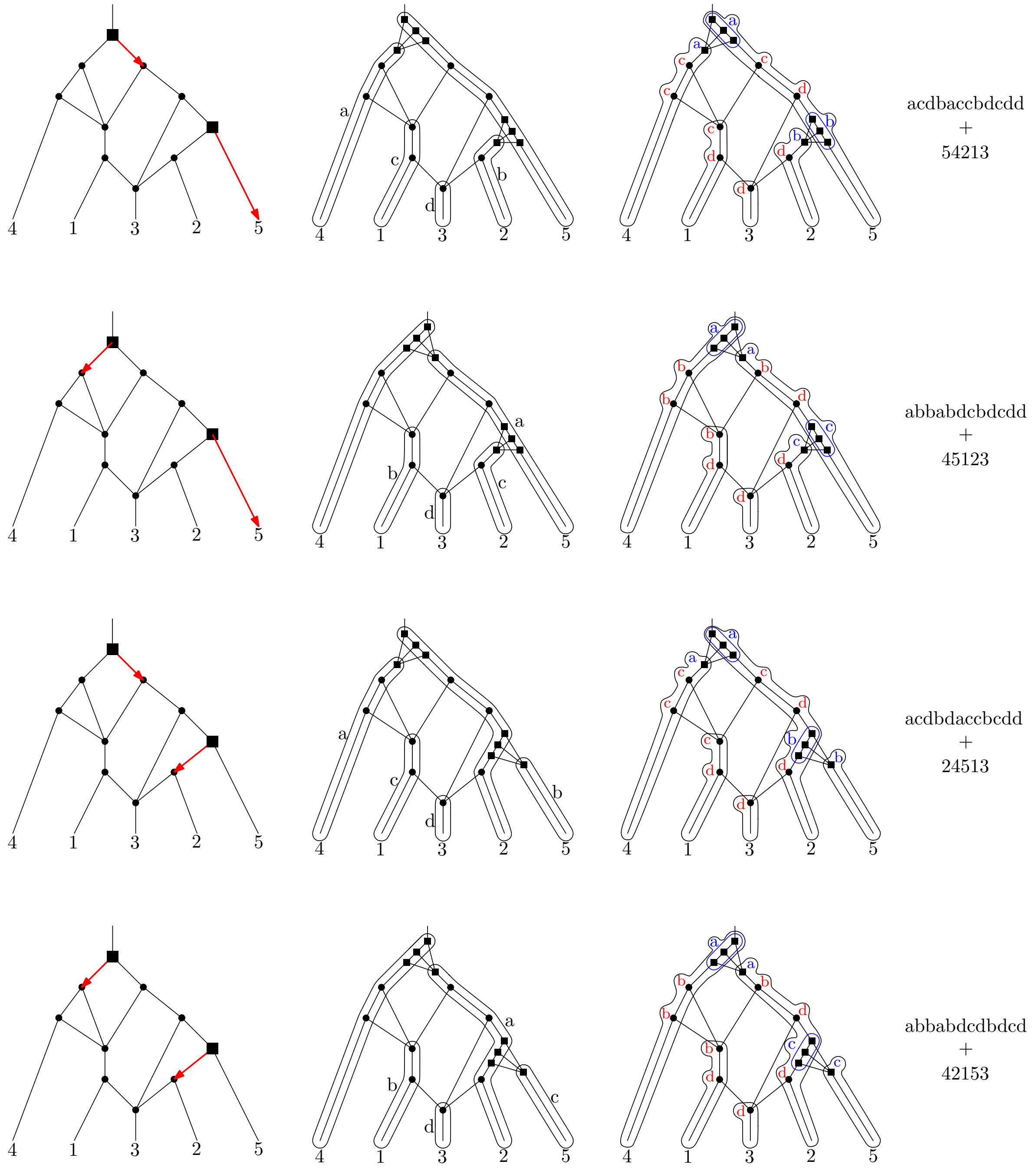
$$b_{n,k,m}^{(d)} = \sum_{j=1}^{m} b_{n-1,k,j}^{(d)} + \binom{n+m+k(d-1)-2}{d-1} \sum_{j=1}^{m} b_{n-1,k-1,j}^{(d)}, \qquad (n \ge 2) \tag{2.26}$$

with initial conditions $b_{n,k,m}^{(d)} = 0$ for $n < m$ of $n < k$, $b_{n,-1,m} = 0$ and $b_{1,0,1}^{(d)} = b_{1,1,1}^{(d)} = 1$.

Proof. We now consider two cases. First, we assume that ω_n is a letter which occurs twice. Then, removing the 2 occurrences of ω_n from these words gives a word of $\mathcal{C}_{n-1,k}^{(d)}$ with suffix $\omega_m\omega_{m+1}\dots\omega_{n-1}$, i.e., it has a suffix $\omega_{n-1}\omega_j\omega_{j+1}\dots\omega_{n-1}$ for $1 \le j \le m$. Reversing this procedure gives the contribution

$$\sum_{j=1}^{m} b_{n-1,k,j}^{(d)}. \tag{2.27}$$

to $b_{n,k,m}^{(d)}$, which is the first term on the right-hand side of (2.26).



(a) (b) (c) (d)

Figure 2.9: (a) The network from Figure 2.8 together with the 4 possible ways of choosing an outgoing free edge for every free tree node; (b) Replacing each free tree node by a reticulation node which results in a maximally reticulated tree-child network whose path-components are indexed; (c) Labeling all internal nodes by labeling reticulation nodes and their parents with the label of their path-component. Note that all nodes on the chosen free edges only receive one label; (d) The word from $\mathcal{C}_{4,2}^{(3)}$ and the permutation corresponding to each network.

Second, if ω_n is a letter which occurs $d+1$ times, we remove the $d+1$ occurrences of this letter. Then, with the same line of reasoning as above, we obtain the second term of the right-hand side of (2.26):

$$\binom{n+m+k(d-1)-2}{d-1}\sum_{j=1}^{m} b^{(d)}_{n-1,k-1,j},$$

where the binomial coefficient counts the number of ways of adding back the $d-1$ occurrences of ω_n after two ω_n's have been added, one before the last ω_m and one at the end of the word. By Definition 2.3.1 these first $d-1$ occurrences of ω_n may be anywhere. ■

The recurrence from the above proposition combined with Theorem 36 allows one to compute values of $\mathrm{TC}^{(d)}_{n,k}$ for small values of n, k, d. Also, like this, we can recover the table for $\mathrm{TC}^{(2)}_{n,k}$ from [9] which was computed in that paper with a much more computation-intensive approach.

$n\backslash k$	0	1	2	3	4	5	6	7
2	1	2						
3	3	21	42					
4	15	228	1272	2544				
5	105	2805	30300	154500	309000			
6	945	39330	696600	6494400	31534200	63068400		
7	10395	623385	16418430	241204950	2068516800	9737380800	19474761600	
8	135135	11055240	405755280	8609378400	113376463200	920900131200	4242782275200	8485564550400

Table 2.2: $\mathrm{TC}^{(2)}_{n,k}$ for $2 \leq n \leq 8$ and $0 \leq k < n$; see also [9].

$n\backslash k$	0	1	2	3	4	5	6
2	1	2					
3	3	33	150				
4	15	492	7908	55320			
5	105	7725	291420	6179940	57939000		
6	945	132030	9603270	430105320	11292075000	132120450000	
7	10395	2471805	307525050	24586633890	1284266876760	40079165452200	560319972030000

Table 2.3: $\mathrm{TC}^{(3)}_{n,k}$ for $2 \leq n \leq 7$ and $0 \leq k < n$.

$n \backslash k$	0	1	2	3	4	5
2	1	2				
3	3	48	546			
4	15	942	45132	1243704		
5	105	18375	2394360	227116260	11351644920	
6	945	375705	107314200	23919407460	3724353682560	291451508298720

Table 2.4: $\mathrm{TC}^{(4)}_{n,k}$ for $2 \leq n \leq 6$ and $0 \leq k < n$.

$n \backslash k$	0	1	2	3	4
2	1	2			
3	3	66	2016		
4	15	1650	242496	28710864	
5	105	39135	17566470	7876446840	2307919133520

Table 2.5: $\mathrm{TC}^{(5)}_{n,k}$ for $2 \leq n \leq 5$ and $0 \leq k < n$.

$n \backslash k$	0	1	2	3	4
2	1	2			
3	3	87	7524		
4	15	2700	1246740	676431360	
5	105	76515	118491090	262058953860	483098464854720

Table 2.6: $\mathrm{TC}^{(6)}_{n,k}$ for $2 \leq n \leq 5$ and $0 \leq k < n$.

2.4 Counting maximally reticulated networks

As a continuation of the topic introduced in Section 2.1 (the maximal number of reticulations), an intuitive counting question to consider is that of enumerating maximally reticulated networks.

The asymptotic result of the number of maximally reticulated reticulation-visible networks and tree-child networks follows from the main result in [23]; see the proof of Theorem 23 which gives the following result.

Theorem 38 ([23])**.** The asymptotic number of maximally reticulated reticulation-visible networks and tree-child networks are given by

$$\mathrm{RV}_{n,3n-3} = \mathrm{TC}_{n,n-1} = \Theta\left(n^{-2/3}\mathrm{e}^{a_1(3n)^{1/3}}\right)\left(\frac{12}{\mathrm{e}^2}\right)^n n^{2n}, \tag{2.28}$$

where a_1 is the largest root of the Airy function of the first kind.

The number of maximally reticulated galled network was provided by Fuchs, Yu, and Zhang in [24].

Theorem 39 ([24])**.** The number of maximally reticulated galled networks is given by

$$\mathrm{GN}_{n,2(n-1)} = 3^{n-1}(2n-3)!! \sim \frac{1}{3\sqrt{2}} n^{-1}\left(\frac{6}{\mathrm{e}}\right)^n n^n.$$

2.4.1 d-combining tree-child networks

The number of maximally reticulated d-combining tree-child networks is highly related to the total number, and they have the same asymptotic results.

Theorem 40. For the number of d-combining tree-child networks, we have, as $n \to \infty$,

$$\mathrm{TC}_n^{(d)} = \Theta\left(\mathrm{TC}_{n,n-1}^{(d)}\right) = \Theta\left((n!)^d\,\gamma(d)^n\, e^{3a_1\beta(d)n^{1/3}} n^{\alpha(d)}\right),$$

where $a_1 = -2.33810741\ldots$ is the largest root of the Airy function of the first kind, defined as the unique function $\mathrm{Ai}(z)$ satisfying $\mathrm{Ai}''(z) = z\mathrm{Ai}(z)$ such that $\lim_{z\to\infty}\mathrm{Ai}(z) = 0$, and

$$\alpha(d) = -\frac{d(3d-1)}{2(d+1)}, \quad \beta(d) = \left(\frac{d-1}{d+1}\right)^{2/3}, \quad \text{and} \quad \gamma(d) = 4\frac{(d+1)^{d-1}}{(d-1)!}.$$

We will later re-state and prove the result together with the total number of d-coming tree-child networks in Section 2.6.2.

2.4.2 Galled tree-child networks

In this section, we enumerate the number of maximally reticulated galled tree-child networks. Since the component graph method (Section 1.1.3.4) says that the maximally reticulated galled tree-child networks are obtained by decompressing a *tree-like structure* which is similar to the counting on galled networks in Section 2.2.2, we may implement generating function techniques (Section 1.2.1) Eventually, by applying Lagrange inversion formula (Section 1.2.4), we have the result below.

Theorem 41. The asymptotic number of maximally reticulated galled tree-child network $\mathrm{GTC}_{n,n-1}$ with n leaves is,

$$\mathrm{GTC}_{n,n-1} \sim \sqrt{\mathrm{e}\pi} n^{-1/2} \left(\frac{2}{\mathrm{e}^2}\right)^n n^{2n}. \tag{2.29}$$

Set

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

where the last line follows from (2.20) and Proposition 12-(i). Then, we have the following result.

Lemma 42. We have,

$$M(z) = z + zL'(M(z)). \tag{2.30}$$

Proof. According to the explanation in the paragraph preceding the lemma 8, 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

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

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 (2.30) is of *Lagrangian type*; see Theorem 17. Thus, we can obtain the asymptotics of $\mathrm{GTC}_{n,n-1}$ by applying Lagrange's inversion formula and Theorem 19 from [6].

Theorem 43. 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 21. For tree-child networks, it was proved in [23] that $\mathrm{TC}_n = \Theta(\mathrm{TC}_{n,n-1})$. (This was a main step in the proof of (48).) The above result together with Theorem 55 shows that the same is not true for galled tree-child networks.

Proof. Applying the Lagrange inversion formula to (2.30) gives

$$\mathrm{GTC}_{n,n-1} = n![z^n]M(z) = (n-1)![\omega^{n-1}](1+L'(\omega))^n. \tag{2.31}$$

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 19 to (2.31) 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. ■

2.5 Asymptotic counting for fixed k as $n \to \infty$

Surveying the previous chapters, our analyses have all focused on exact counts under a constraint on the number of reticulations. Recall the assumptions on k in each section: Section 2.2 considers small fixed k; Section 2.3 allows arbitrary k; Section 2.4 treats maximal k (depending on n). In this section, we turn to the counting problem under the assumption that k is a fixed constant and n tends to infinity.

There are known results for general phylogenetic and tree-child networks. In 2020, Mansouri [34] gave the asymptotics of the number of general phylogenetic networks:

$$\mathrm{PN}_{n,k} \sim d_k\, 2^{3k-1} \left(\frac{2}{\mathrm{e}}\right)^n n^{n+2k-1}, \tag{2.32}$$

where $d_1 = \sqrt{2}/4$, $d_2 = \sqrt{2}/32$ and $d_3 = \sqrt{2}/384$, d_k depends on k. In 2018 and 2020, Fuchs, Gittenberger, and Mansouri [19, 20] proved the same result for tree-child networks (and normal networks) but without the closed-form expression for the leading constant:

$$\mathrm{TC}_{n,k}(\sim \mathrm{NN}_{n,k}) \sim d_k\, 2^{3k-1} \left(\frac{2}{\mathrm{e}}\right)^n n^{n+2k-1}. \tag{2.33}$$

Later in 2022, Fuchs, Huang, and Yu [21] obtained the leading constants d_k:

$$d_k = \frac{\sqrt{2}}{k!4^k}.$$

Thus, the classes of phylogenetic networks between normal networks and general networks have the same asymptotics; see Figure 1.5.

2.5.1 d-combing tree-child networks

In this section, we give another application of the method of component graphs, namely, we derive the first-order asymptotics of $\mathrm{TC}_{n,k}^{(d)}$ for fixed k as $n \to \infty$. This extends the main result from [21] to general d.

The main observation of [21] was that the main asymptotic contribution to the asymptotics of $\mathrm{TC}_{n,k}^{(d)}$ comes from the tree-child networks constructed from the star-component graph with k leaves; see Figure 2.10. (In fact, since component graphs are labeled, there are $k+1$ star-component graphs depending on the label of the root vertex.)

We denote by $S_{n,k}^{(d)}$ the number of tree-child networks arising from the star-component graph(s) with k leaves. Then, we have the following formula for this number which generalizes the formula from [21, Lemma 5] for $d = 2$.

Figure 2.10: Star-component graphs for generating the tree-child networks whose numbers dominate the asymptotics of $\mathrm{TC}_{n,k}^{(d)}$ for fixed k as n tends to infinity. (Left: $d = 2$; Right: $d = 3$. Labels of nodes are removed.)

Lemma 44. We have,

$$S_{n,k}^{(d)} = \frac{n!}{(d!)^k 2^{n-k-1}(k-1)!} \sum_{j=1}^{n-k} \frac{(2j+dk-2)!}{j!(j-1)!} \cdot \frac{(2n-k-2j-1)!}{(n-k-j)!(n-j)!}.$$

Proof. The proof is very similar to the one of [21, Lemma 5]. We give a short sketch.

First, by a combinatorial argument, we have

$$S_{n,k}^{(d)} = \sum_{j=1}^{n-k} \binom{n}{j} \frac{(2j+dk-2)!}{(d!)^k 2^{j-1}(j-1)!} \cdot \frac{1}{k!}(n-j)![z^{n-j}]T(z)^k,$$

where

$$T(z) = \sum_{n\geq 1} (2n-1)!! \cdot \frac{z^n}{n!} = 1 - \sqrt{1-2z}$$

is the exponential generating function of the number of phylogenetic trees.

Briefly, the construction on which this combinatorial argument is based works as follows: first, we pick a one-component tree-child network with $j+k$ leaves where the leaves with labels $\{1, \ldots, k\}$ are the ones below the reticulation nodes. Then, we replace the leaves below reticulation nodes by phylogenetic trees. (For this, we need a forest of k phylogenetic trees.) Finally, we re-label the leaves.

Next, by a standard application of the Lagrange inversion formula, we obtain

$$[z^{n-j}]T(z)^k = \frac{k}{n-j} 2^{j+k-n} \binom{2n-k-2j-1}{n-k-j}.$$

Plugging this into the expression above and straightforward manipulations yield the claimed result. ■

Applying to this the Laplace method, we have the following asymptotic result.

Proposition 14. For fixed k, as $n \to \infty$,

$$S_{n,k}^{(d)} \sim \frac{2^{dk-1}}{(d!)^k\, k! \sqrt{\pi}} n! 2^n n^{dk-3/2}.$$

Proof. The proof is similar to [21, Lemma 6]. Again, we just give a sketch.

First, it suffices to consider the asymptotics of the sum in the expression for $S_{n,k}^{(d)}$ from Lemma 44 as the factor in front of it has already the right shape. Thus, we set

$$\begin{aligned}\Sigma_{n,k} &:= \sum_{j=1}^{n-k} \frac{(2j+dk-2)!}{j!(j-1)!} \cdot \frac{(2n-k-2j-1)!}{(n-k-j)!(n-j)!} \\ &= \sum_{j=0}^{n-k-1} \frac{(2n+(d-2)k-2j-2)!}{(n-k-j)!(n-k-j-1)!} \cdot \frac{(2j+k-1)!}{j!(j+k)!}.\end{aligned}$$

Now, observe that the first term in the last sum is decreasing in j and has the expansion

$$\frac{(2n+(d-2)k-2j-2)!}{(n-k-j)!(n-k-j-1)!} = \frac{2^{(d-2)k}}{\sqrt{\pi}} 4^{n-j-1} n^{dk-3/2} \left(1 + \mathcal{O}\left(\frac{1+j}{n}\right)\right)$$

uniformly in j as $j = o(n)$. Thus, by a standard application of the Laplace method

$$\Sigma_{n,k} \sim \frac{2^{(d-2)k}}{\sqrt{\pi}} \left(\sum_{j=0}^{\infty} \frac{(2j+k-1)!}{j!(j+k)!} 4^{-j}\right) 4^{n-1} n^{dk-3/2} = \frac{2^{(d-1)k}}{k\sqrt{\pi}} 4^{n-1} n^{dk-3/2},$$

where we used [21, Lemma 7] in the last step.

From the above asymptotics multiplied with the factor in front of the sum in the formula for $S_{n,k}^{(d)}$ from Lemma 44, we obtain the claimed result. ■

Finally, using the same arguments as in [21], we can show that also here the contribution from tree-child networks arising from the star-component graph(s) dominates; we leave details to the reader.

Theorem 45. For the number of d-combining tree-child networks with n leaves and k reticulation nodes, we have for fixed k, as $n \to \infty$,

$$\mathrm{TC}_{n,k}^{(d)} \sim \frac{2^{dk-1}}{(d!)^k\, k!\, \sqrt{\pi}} n!\, 2^n\, n^{dk-3/2}.$$

Remark 22. This result was stated in [10, Theorem 8] without proof. (Note that this also corrects two typos in the statement of [10, Theorem 8].)

2.5.2 Reticulation-visible networks

An easy consequence of the asymptotic result of reticulation-visible networks for fixed k as $n \to \infty$ is the following.

Corollary 2.5.1. For fixed k,

$$\mathrm{RV}_{n,k} \sim \frac{2^{k-1}\sqrt{2}}{k!} \left(\frac{2}{e}\right)^n n^{n+2k-1}, \qquad (n \to \infty).$$

Proof. According the the subset relationship, $\mathcal{TC} \subseteq \mathcal{RV} \subseteq \mathcal{PN}$. ■

For the same reason, the classes containing tree-child networks in Figure 1.5 admit the same asymptotic result (2.33).

2.5.3 Galled networks

Galled networks does not be contained in the above mentioned *poset chain* of classes (above the normal or tree-child networks), but surprisingly the same asymptotic result holds.

Theorem 46. For fixed k,

$$\mathrm{GN}_{n,k} \sim \frac{2^{k-1}\sqrt{2}}{k!}\left(\frac{2}{e}\right)^{n} n^{n+2k-1}, \qquad (n \to \infty). \tag{2.34}$$

This will be follow from the following asymptotic result for $E_k(z)$ which is obtained from (2.12) and induction.

Proposition 15. For $k \geq 1$, $E_k(z)$ is Δ-analytic for some Δ-domain at $1/2$ and satisfies, as $z \to 1/2$ in the Δ-domain:

$$E_k(z) \sim \frac{(4k-3)!!}{k!2^k(1-2z)^{2k-1/2}}.$$

Proof. Note that

$$E_0(z) = F_0(z) = 1 - \sqrt{1-2z} \sim 1.$$

Thus, if $k = 0$ is included in the claim, then the power of $1-2z$ in the denominator is $\max\{2k - 1/2, 0\}$.

We use induction on k. For $k = 1$, the claim holds since $E_1(z) \sim F_1(z)$ which has the desired form by (2.8). Thus, we can assume that the claim holds for $k' < k$. We need to prove it for k. Plugging the induction hypothesis into (2.12) and using (2.8), we obtain that for the terms inside the double sum of (2.12):

$$F_j(z)E_{\ell_1}(z)\cdots E_{\ell_j}(z) \sim c(1-2z)^{-(2j-1/2+\max\{2\ell_1-1/2,0\}+\cdots+\max\{2\ell_j-1/2,0\})},$$

where c is a suitable constant and $\ell_1 + \cdots + \ell_j = k - j$. The term inside the bracket is maximized if and only if $j = k$ and thus $\ell_1 = \cdots = \ell_k = 0$. This shows that

$$E_k(z) \sim \frac{F_k(z)}{k!}$$

which by (2.8) gives the claim. ■

Now, we can prove Theorem 46 for $k \geq 1$.

Proof of Theorem 46. From Proposition 2.12 and the transfer theorems of singularity analysis (see Chapter VI in [17]):

$$\mathrm{GN}_{n,k} = n![z^n]E_k(z) \sim n!\frac{(4k-3)!!}{k!2^k\Gamma(2k-1/2)}2^n n^{2k-3/2}.$$

Note that

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

Plugging this into the above expression and using Stirling's formula gives the claimed result. ■

Remark 23. It is easily verified that Theorem 46 holds for $k = 0$, too.

2.5.4 Galled tree-child networks

We next consider $\mathrm{GTC}_{n,k}$ with k small. Here, we have the following result which explains why the asymptotic expansions of $\mathrm{TC}_{n,k}$ in (2.33) and $\mathrm{GN}_{n,k}$ in (2.34) are the same.

Theorem 47. For fixed k, as $n \to \infty$,

$$\mathrm{GTC}_{n,k} \sim \frac{2^{k-1}\sqrt{2}}{k!}\left(\frac{2}{e}\right)^n n^{n+2k-1}. \tag{2.35}$$

The proof of this result uses ideas from [21].

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 [21], the number of these galled tree-child network has the same asymptotics as the one on the right-hand side of (2.35). 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. ■

Note that this re-proves the (surprising) asymptotic result for $\mathrm{GN}_{\mathrm{n,k}}$ in (2.35) from [11].

Alternative proof. First, from the subset-relation $\mathcal{GTC}_{n,k} \subseteq \mathcal{TC}_{n,k}, \mathcal{GN}_{n,k}$ and (2.33) and (2.34), we have

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

for fixed k as $n \to \infty$. On the other hand, also from the equations (2.32), (2.33) and (2.34).

$$\mathrm{TC}_{n,k} + \mathrm{GN}_{n,k} - \mathrm{GTC}_{n,k} = |\mathcal{TC}_{n,k} \cup \mathcal{GN}_{n,k}| \leq \mathrm{PN}_{n,k},$$

which gives a lower bound having the same asymptotic to the upper bound, for $\mathrm{GTC}_{n,k}$. The asymptotic result (2.35) follows. ■

2.6 Asymptotic counting of the total number of networks

Another topic concerns the asymptotic counting of the total number of networks. For essentially every network class, obtaining exact enumeration results is extremely challenging. None of the main classes currently admit closed-form exact counts (compare with the discussion in Section 2.3), and therefore this section focuses on their asymptotic behavior. Moreover, asymptotic enumeration often serves as the first step toward developing probabilistic models and studying random variables on these network classes.

Fuchs et al. [19] proved an asymptotic result for the number of one-component tree-child networks. More precisely, the second result in Theorem 12 (also in [19]) 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 2.6.1 ([19])**.** As $n \to \infty$,

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

The growth terms in the asymptotic result for the number of tree-child networks with n leaves are all determined except the multiplicative constant in [23, Theorem 1].

Theorem 48 ([23])**.** For the number of tree-child networks, we have, as $n \to \infty$,

$$\mathrm{TC}_n = \Theta\left(n^{-2/3} \mathrm{e}^{a_1 (3n)^{1/3}} \left(\frac{12}{\mathrm{e}^2}\right)^n n^{2n}\right), \tag{2.36}$$

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 but the multiplicative constant is well-determined. It was proved in [24] that

Theorem 49 ([24])**.** For the number of galled networks, we have, as $n \to \infty$,

$$\mathrm{GN}_n \sim \frac{\sqrt{2}\,\mathrm{e}^{3/4}}{4} n^{-1} \left(\frac{8}{\mathrm{e}^2}\right)^n n^{2n}. \tag{2.37}$$

The tools used to establish (2.36) and (2.37) were very different: for (2.36), a bijection to a class of words was proved (see Section 2.3.2) and a recurrence for these word was found (see Proposition 13) which could be (asymptotically) analyzed with the approach from [37]; for (2.37), the component graph method introduced in [29] (which was explained in Section 1.1.3.2) together with the Laplace method and Theorem 19 was used.

2.6.1 One-component d-combining tree-child networks

We first look at the total number of the one-component d-combining tree-child networks.

Theorem 50. The following asymptotic equivalences hold for one-component d-combining tree-child networks.

(i) For $d = 2$ (bi-combining case), we have

$$\mathrm{OTC}_n^{(2)} \sim \frac{1}{4\pi\sqrt{e}}(n!)^2 2^n e^{2\sqrt{n}} n^{-9/4};$$

see [23, Theorem 3].

(ii) For $d = 3$, we have

$$\mathrm{OTC}_n^{(3)} \sim I_1(2) \cdot \mathrm{OTC}_{n,n-1}^{(3)} \sim \frac{I_1(2)\sqrt{3}}{9\pi}(n!)^3 \left(\frac{9}{2}\right)^n n^{-3},$$

where

$$I_v(a) = \left(\frac{a}{2}\right)^v \sum_{k=0}^{\infty} \frac{1}{k!\Gamma(k+v+1)} \frac{a^{2k}}{4^k}$$

is the modified Bessel function of the first kind.

(iii) For $d \geq 4$, we have

$$\mathrm{OTC}_n^{(d)} \sim \mathrm{OTC}_{n,n-1}^{(d)} \sim \frac{d!}{d^{d-1/2}(2\pi)^{(d-1)/2}}(n!)^d \left(\frac{d^d}{d!}\right)^n n^{3(1-d)/2}.$$

The case of $d = 2$ is the only case of the three cases above in which we find a stretched exponential in the asymptotics (see [37]). The result and the exact formula for $\mathrm{OTC}_{n,k}^{(d)}$ for any k and n in Theorem 35 also give the distributional result for the number of reticulation nodes which will be shown in Section 2.7. Note that the proof of Theorem 50-(ii) and (iii) will later be proved together with Theorem 59 in Section 2.7.1.

2.6.2 d-combining tree-child networks

To understand the counting of TC_n, we are going to see that the growth of $\mathrm{TC}_{n,k}^{(d)}$ is dominated by $\mathrm{TC}_{n,n-1}^{(d)}$. For the latter sequence, a recurrence used for the computation of its values and the method of [37] (which needs some adaptions because of the dependence on d) yields the following asymptotic counting result for $\mathrm{TC}_n^{(d)}$. For the bi-combining case, this result was proved in [23]. Contrary to the one-component case from Theorem 50, in the general case the stretched exponential appears for all $d \geq 2$.

d	$\alpha(d)$	$\approx$	$\beta(d)$	$\approx$	$\gamma(d)$	$\approx$
2	$-\frac{5}{3}$	-1.67	$(\frac{1}{3})^{2/3}$	0.48	12	12.00
3	-3	-3.00	$(\frac{1}{2})^{2/3}$	0.63	32	32.00
4	$-\frac{22}{5}$	-4.40	$(\frac{3}{5})^{2/3}$	0.71	$\frac{250}{3}$	83.33
5	$-\frac{35}{6}$	-5.83	$(\frac{2}{3})^{2/3}$	0.76	216	216.00
6	$-\frac{51}{7}$	-7.29	$(\frac{5}{7})^{2/3}$	0.80	$\frac{16807}{30}$	560.23
7	$-\frac{35}{4}$	-8.75	$(\frac{3}{4})^{2/3}$	0.83	$\frac{65536}{45}$	1456.36
8	$-\frac{92}{9}$	-10.22	$(\frac{7}{9})^{2/3}$	0.85	$\frac{531441}{140}$	3796.01
9	$-\frac{117}{10}$	-11.70	$(\frac{4}{5})^{2/3}$	0.86	$\frac{625000}{63}$	9920.63
10	$-\frac{145}{11}$	-13.18	$(\frac{9}{11})^{2/3}$	0.87	$\frac{2357947691}{90720}$	25991.49

Table 2.7: Specific values of the asymptotic parameters $\alpha(d)$, $\beta(d)$, and $\gamma(d)$ from Theorem 51.

The theorem below also states the maximally reticulated d-combing tree-child networks in Section 2.4.1 has the same asymptotic growth up to a constant.

Theorem 51. For the number of d-combining tree-child networks, we have, as $n \to \infty$,

$$\mathrm{TC}_n^{(d)} = \Theta\left(\mathrm{TC}_{n,n-1}^{(d)}\right) = \Theta\left((n!)^d\,\gamma(d)^n\,e^{3a_1\beta(d)n^{1/3}}n^{\alpha(d)}\right), \tag{2.38}$$

where $a_1 = -2.33810741\ldots$ is the largest root of the Airy function of the first kind, defined as the unique function $\mathrm{Ai}(z)$ satisfying $\mathrm{Ai}''(z) = z\mathrm{Ai}(z)$ such that $\lim_{z\to\infty}\mathrm{Ai}(z) = 0$, and

$$\alpha(d) = -\frac{d(3d-1)}{2(d+1)}, \qquad \beta(d) = \left(\frac{d-1}{d+1}\right)^{2/3}, \qquad \gamma(d) = 4\frac{(d+1)^{d-1}}{(d-1)!}.$$

The first few specific values of the asymptotic parameters $\alpha(d)$, $\beta(d)$, and $\gamma(d)$ are shown Table 2.7. Also, by performing a finer analysis for k close to n, we obtain the following distributional result for the number of reticulation nodes. (In the case $d \geq 3$, the above mentioned encoding again plays an important role in the proof; the case $d = 2$ needs a different treatment.)

In the following content, we prove Theorem 51, namely, we derive an asymptotic result for $\mathrm{TC}_n^{(d)}$. Recall that

$$\mathrm{TC}_n^{(d)} = \sum_{k=0}^{n-1}\mathrm{TC}_{n,k}^{(d)}.$$

The first main observation is that the last term in this sum dominates; see the first equality in (2.38). More precisely, we have the following.

Lemma 52. For $0 \leq k \leq n-2$, we have

$$\mathrm{TC}_{n,k}^{(d)} \leq \frac{1}{2(n-k-1)} \mathrm{TC}_{n,k+1}^{(d)} \tag{2.39}$$

and consequently,

$$\mathrm{TC}_{n}^{(d)} = \Theta\left(\mathrm{TC}_{n,n-1}^{(d)}\right).$$

Proof. Let N be a tree-child network with n leaves and k reticulation nodes. Recall that N has $2(n-k-1)$ free edges; see Lemma 37.

We can construct tree-child networks with n leaves and $k+1$ reticulation nodes from N by (i) inserting $d-1$ tree nodes into the root edge of N and a reticulation node into a free edge and (ii) connecting the $d-1$ new tree nodes to the new reticulation node. Note that each network built in this way is different. Thus,

$$2(n-k-1)\mathrm{TC}_{n,k}^{(d)} \leq \mathrm{TC}_{n,k+1}^{(d)},$$

which implies the first claim.

Next, by iteration of (2.38), we obtain

$$\mathrm{TC}_{n,k}^{(d)} \leq \frac{1}{2^{n-k-1}(n-k-1)!} \mathrm{TC}_{n,n-1}^{(d)} \tag{2.40}$$

and thus,

$$\mathrm{TC}_{n,n-1}^{(d)} \leq \mathrm{TC}_{n}^{(d)} \leq \left(\sum_{j\geq 0} \frac{1}{2^j j!}\right) \cdot \mathrm{TC}_{n,n-1}^{(d)} = \sqrt{e} \cdot \mathrm{TC}_{n,n-1}^{(d)},$$

which proves the second claim. ■

Remark 24. Note that for $d=2$ and $k=n-2$, equality holds in (2.38) because in this case, (a) the networks constructed from N in the above proof are maximally reticulated and (b) the child of the root of each maximally reticulated network is not free. Thus, the construction from the proof is reversible and we have a bijection. (This was first proved in [9, Proposition 17].)

As a consequence of the last result, we can now entirely concentrate on the maximal reticulated case for which we obtain from Theorem 36:

$$\mathrm{TC}_{n,n-1}^{(d)} = n! c_{n-1}^{(d)},$$

where we have set $c_n^{(d)} := c_{n,n}^{(d)}$. By Proposition 13, this sequence satisfies

$$c_n^{(d)} = \sum_{m\geq 1} b_{n,m}^{(d)},$$

where $b_{n,m}^{(d)} := b_{n,n,m}^{(d)}$ satisfies

$$b_{n,m}^{(d)} = \binom{m+nd-2}{d-1} \sum_{j=1}^{m} b_{n-1,j}^{(d)}. \tag{2.41}$$

The recurrence (2.41) can be brought in a slightly easier form.

Lemma 53. We have,

$$b_{n,m}^{(d)} = \frac{dn+m-2}{dn+m-d-1} b_{n,m-1}^{(d)} + \binom{dn+m-2}{d-1} b_{n-1,m}^{(d)}, \tag{2.42}$$

for $n \geq 2$ and $0 \leq m \leq n$ with initial conditions $b_{1,1}^{(d)} = 1$ and $b_{n,m}^{(d)} = 0$ for (i) $n \geq 2$ and $m = -1$; (ii) $n = 1$ and $m = 0$; and (iii) $n < m$.

Proof. The recursive structure in (2.41) yields

$$\frac{b_{n,m}^{(d)}}{\binom{dn+m-2}{d-1}} - \frac{b_{n,m-1}^{(d)}}{\binom{dn+m-3}{d-1}} = b_{n-1,m}^{(d)}.$$

This gives the claimed recurrence and the initial conditions are easily checked. ■

To this recurrence, we apply now the method from [37]. Due to the similarities, we only discuss the main differences. We start with the following transformation of $(b_{n,m}^{(d)})_{0 \leq m \leq n}$ to $(e_{i,j}^{(d)})_{\substack{0 \leq i \leq j \\ i-j \text{is even}}}$, which changes the indices and captures the exponential and super-exponential terms coming from the binomial coefficient in (2.42).

Lemma 54. We have

$$b_{n,m}^{(d)} = \lambda(d)^n (n!)^{d-1} e_{n+m,n-m}^{(d)} \qquad \text{with} \qquad \lambda(d) = \frac{(d+1)^{d-1}}{(d-1)!},$$

where $e_{n,m}^{(d)}$ satisfies the following recurrence

$$e_{n,m}^{(d)} = \mu_{n,m}^{(d)} e_{n-1,m+1}^{(d)} + \nu_{n,m}^{(d)} e_{n-1,m-1}^{(d)} \tag{2.43}$$

with

$$\mu_{n,m}^{(d)} = 1 + \frac{2(d-1)}{(d+1)n + (d-1)m - 2(d+1)} \quad \text{and} \quad \nu_{n,m}^{(d)} = \prod_{i=2}^{d} \left(1 - \frac{2(m+i)}{(d+1)(n+m)}\right)$$

for $n \geq 3$ and $m \geq 0$, where $e_{n,-1}^{(d)} = e_{2,n}^{(d)} = 0$ except for $e_{2,0}^{(d)} = 1/\lambda(d)$.

Now, we are interested in

$$e_{2n,0}^{(d)} = \frac{b_{n,n}^{(d)}}{\lambda(d)^n (n!)^{d-1}}$$

because by the previous lemmas and (2.41) we have

$$\begin{aligned} \mathrm{TC}_n^{(d)} &= \Theta\left(\mathrm{TC}_{n,n-1}^{(d)}\right) = \Theta\left(n! c_{n-1}^{(d)}\right) \\ &= \Theta\left(n!\, n^{1-d} b_{n,n}^{(d)}\right) = \Theta\left((n!)^d \lambda(d)^n n^{1-d} e_{2n,0}^{(d)}\right). \end{aligned} \tag{2.44}$$

Moreover, observe that for the Theta-result, the initial value of $e_{2,0}^{(d)}$ is irrelevant, as it creates only a constant factor. So we may set it to $e_{2,0}^{(d)} = 1$, or any convenient constant. Note that this recurrence is very similar to that of relaxed trees [37, Equation (2)], yet with more complicated factors. Observe also that this is exactly recurrence [23, Equation (10)] for $d = 2$.

Motivated by experiments for large n, we use the following approach

$$e_{n,m}^{(d)} \approx h(n) f\left(\frac{m+1}{n^{1/3}}\right),$$

where h and f are some "regular" functions. Next, we substitute $s(n) = h(n)/h(n-1)$ and $m = \kappa n^{1/3} - 1$ into (2.43). Then, for $n \to \infty$ we get the expansion

$$f(\kappa)s(n) = 2f(\kappa) + \left(f''(\kappa) - \frac{2(d-1)}{d+1}\kappa f(\kappa)\right) n^{-2/3} + \mathcal{O}\left(n^{-1}\right).$$

Hence, we may assume that

$$s(n) = 2 + c_1 n^{-2/3} + c_2 n^{-1} + \dots$$

and this implies that $f(\kappa)$ satisfies the differential equation

$$f''(\kappa) = \left(c_1 + \frac{2(d-1)}{d+1}\kappa\right) f(\kappa)$$

that is solved by the Airy function Ai of the first kind, as we have $e_{n,m} = 0$ for $m > n$ which corresponds to $\lim_{x\to\infty} f(x) = 0$. Additionally, the boundary conditions allow us to compute c_1 and we get that

$$f(\kappa) = C\mathrm{Ai}\left(a_1 + B^{1/3}\kappa\right), \qquad \text{where} \qquad B := \frac{2(d-1)}{d+1}, \tag{2.45}$$

$a_1 \approx 2.338$ is the largest root of the Airy function Ai, and C is an arbitrary constant. From this we get that $c_1 = a_1 B^{1/3}$. These heuristic arguments guide us to the following results. The proofs are analogous to [37, 14, 23]; for the details we refer to the accompanying Maple worksheet [41]. Note that the next two results generalize [23, Propositions 4 and 5], whose results are recovered by setting $d = 2$.

Proposition 16. For all $n, m \geq 0$ let

$$\tilde{X}_{n,m} := \left(1 - \frac{2d-1}{3(d+1)}\frac{m^2}{n} - \frac{3d^2+12d-11}{6(d+1)}\frac{m}{n}\right) \mathrm{Ai}\left(a_1 + \frac{B^{1/3}(m+1)}{n^{1/3}}\right) \quad \text{and}$$

$$\tilde{s}_n := 2 + \frac{a_1 B^{2/3}}{n^{2/3}} - \frac{3d^2-5d+4}{3(d+1)n} - \frac{1}{n^{7/6}}.$$

Then, for any $\varepsilon > 0$, there exists an $\tilde{n}_0$ such that

$$\tilde{X}_{n,m}\tilde{s}_n \leq \mu_{n,m}^{(d)}\tilde{X}_{n-1,m+1} + \nu_{n,m}^{(d)}\tilde{X}_{n-1,m-1}$$

for all $n \geq \tilde{n}_0$ and for all $0 \leq m < n^{2/3-\varepsilon}$, where $\mu_{n,m}^{(d)}$ and $\nu_{n,m}^{(d)}$ are as in Lemma 54.

Proposition 17. Choose $\eta > \frac{(2d-1)^2}{18(d+1)^2}$ fixed and for all $n, m \geq 0$ let

$$\hat{X}_{n,m} := \left(1 - \frac{2d-1}{3(d+1)}\frac{m^2}{n} - \frac{3d^2+12d-11}{6(d+1)}\frac{m}{n} + \eta\frac{m^4}{n^2}\right) \mathrm{Ai}\left(a_1 + \frac{B^{1/3}(m+1)}{n^{1/3}}\right) \quad \text{and}$$
$$\hat{s}_n := 2 + \frac{a_1 B^{2/3}}{n^{2/3}} - \frac{3d^2-5d+4}{3(d+1)n} + \frac{1}{n^{7/6}}.$$

Then, for any $\varepsilon > 0$, there exists a constant $\hat{n}_0$ such that

$$\hat{X}_{n,m}\hat{s}_n \geq \mu^{(d)}_{n,m}\hat{X}_{n-1,m+1} + \nu^{(d)}_{n,m}\hat{X}_{n-1,m-1}$$

for all $n \geq \hat{n}_0$ and all $0 \leq m < n^{1-\varepsilon}$.

These propositions will be proved in the Section 2.9.2 and they allow us now to prove Theorem 51 on the asymptotics of general d-combining tree-child networks with n leaves.

Proof of Theorem 51. Let us start with the lower bound.

We first define a sequence $X_{n,m} := \max\{\tilde{X}_{n,m}, 0\}$ which satisfies the inequality of Proposition 16 for all $m \leq n$. Then, we define an explicit sequence $\tilde{h}_n := \tilde{s}_n \tilde{h}_{n-1}$ for $n > 0$ and $\tilde{h}_0 = \tilde{s}_0$. From this, we get by induction that $e^{(d)}_{n,m} \geq C_0 \tilde{h}_n X_{n,m}$ for some constant $C_0 > 0$ and all $n \geq \tilde{n}_0$ and all $0 \leq m \leq n$. Hence,

$$\begin{aligned} e^{(d)}_{2n,0} &\geq C_0 \tilde{h}_{2n} X_{2n,0} \\ &\geq C_0 \prod_{i=1}^{2n}\left(2 + \frac{a_1 B^{2/3}}{i^{2/3}} - \frac{3d^2-5d+4}{3(d+1)i} - \frac{1}{i^{7/6}}\right) \mathrm{Ai}\left(a_1 + \frac{B^{1/3}}{(2n)^{1/3}}\right) \\ &\geq C_1 4^n e^{3a_1(B/2)^{2/3}n^{1/3}} n^{\frac{d^2+d-2}{2(d+1)}}. \end{aligned}$$

Finally, combining this with (2.44) we get the lower bound.

The upper bound is similar, yet more technical.

The starting point is Proposition 17 and a function $X_{n,m}$ that is valid for all $0 \leq m \leq n$. For this purpose we define a sequence $\hat{e}^{(d)}_{n,m}$ such that $\hat{e}^{(d)}_{n,m} := e^{(d)}_{n,m}$ for $0 \leq m \leq n^{1-\varepsilon}$ and $\hat{e}^{(d)}_{n,m} := 0$ otherwise; compare with [37, 14]. The missing key step is now to show that $e^{(d)}_{2n,0} = \mathcal{O}(\hat{e}^{(d)}_{2n,0})$. Combining this with the analogous computations performed for the lower bound we get

$$\hat{e}^{(d)}_{2n,0} \leq \hat{C}_1 4^n e^{3a_1(B/2)^{2/3}n^{1/3}} n^{\frac{d^2+d-2}{2(d+1)}}.$$

To complete the prove we show $e^{(d)}_{2n,0} \leq 2\hat{e}^{(d)}_{2n,0}$ using lattice path theory and computer algebra. The argument follows along the same lines as in [23, Appendix]. We start from Equation (2.43) of $e^{(d)}_{n,m}$, which we interpret as a recurrence counting lattice paths. They are composed of steps $(1,1)$ weighted by $\mu^{(d)}_{n,m}$ and $(1,-1)$ weighted by $\nu^{(d)}_{n,m}$ when the respective step ends at (n,m). The total weight of a path is the product of its weights. Now, we are interested in the paths never crossing $y = 0$ and ending at $(2n, 0)$. Let now $p_{\ell,k,2n}$ be the number of such paths starting at (ℓ, k) and ending at $(2n, 0)$. From (2.43) we directly get

$$p_{\ell,k,2n} = \mu^{(d)}_{\ell+1,k-1} p_{\ell+1,k-1,2n} + \nu^{(d)}_{\ell+1,k+1} p_{\ell+1,k+1,2n},$$

with $p_{\ell,-1,2n} = 0$ and $p_{2n,k,2n} = \delta_{k,0}$.

Now, as in [23] we are able to show that

$$\frac{p_{\ell,j,2n}}{(j+1)^2} \geq \frac{p_{\ell,k,2n}}{(k+1)^2}, \tag{2.46}$$

for integers $0 \leq j < k \leq \ell \leq 2n$ such that $2 \mid k - j$. For the technical details, using reverse induction on ℓ, we refer to our accompanying Maple worksheet [41].

Finally, from (2.46) we directly get

$$p_{2x,2y,2n} \leq (2y+1)^2 p_{2x,0,2n}, \tag{2.47}$$

which we need to apply [37, Lemma 4.6] together with the bound $e^{(d)}_{2x,2y} \leq \binom{2x}{x+y}$, which holds due to the same reasons as in [23]: combining the weights of up an down steps gives a weight less than one. This proves $e^{(d)}_{2n,0} \leq 2\hat{e}^{(d)}_{2n,0}$ and ends the proof of Theorem 51. ■

Remark 25. Note that in [37, 14] a stronger result than (2.46) was proved, where the powers in the denominators are 1 instead of 2. However, any polynomial coefficient in (2.47) suffices to get the same result using [37, Lemma 4.6].

With the same strategy as in [37, 14], we could show the stronger result $p_{2x,2y,2n} \leq 2(2y + 1)p_{2x,0,2n}$, however, then other technicalities arise: the value $j = 0$ and the range $0 \leq j < k \leq \ell \leq 2n$ have to be treated separately.

2.6.3 Galled tree-child networks

Again by the component graph method (Section 1.1.3.4), we begin by analyzing its component graphs. By relaxing the condition of how the black arrows can be placed on edges, we obtain an upper bound. Conversely, by restricting our attention to the subclass of component graphs that contributes the dominant term, we obtain a lower bound. Remarkably, both bounds lead to the same asymptotic form, which yields the following result.

Theorem 55. For the number of galled tree-child networks, we have, as $n \to \infty$,

$$\mathrm{GTC}_n \sim \frac{1}{2\sqrt[4]{\mathrm{e}}} n^{-5/4} \mathrm{e}^{2\sqrt{n}} \left(\frac{2}{\mathrm{e}^2}\right) n^{2n}. \tag{2.48}$$

Remark 26. Note that the asymptotics contains a stretched exponential as does the expansion (2.38) for tree-child networks, however, the proof will use the tools which were used in [24] to derive (2.37) for galled networks.

For the proof of Theorem 55, we closely follow the method of proof of (2.37) from [24]. The main idea is to use (1.7) 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 Theorem 4). Since, as explained in Section 1.1.3.4, 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 $\mathrm{GTC}_n \le U_n$. Next, set

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

Then, the definition of U_n implies the following result.

Lemma 56. 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

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

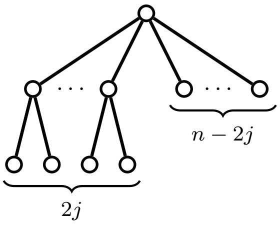
from which the claimed result follows. ■

Now, we can proceed as in the proof of Theorem 24 to obtain the following asymptotic result for U_n.

Proposition 18. 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 (1.7) 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 (not one-component tree-child networks). The decompression procedure from Section 1.1.3.4, then gives the following result.

Lemma 57. We have,

$$\begin{aligned} L_n &= \sum_{j=0}^{\lfloor n/2 \rfloor} \binom{n}{2j} \frac{(2j)!}{j!2^j} \sum_{\ell=0}^{n-2j} \binom{n-2j}{\ell} L_{n-j,j+\ell} \\ &= \sum_{j=0}^{\lfloor n/2 \rfloor} \binom{n}{2j} \frac{(2j)!}{j!2^j} \sum_{\ell=0}^{n-2j} \binom{n-2j}{\ell} \frac{(2n-2j-2)!}{2^{n-j-1}(n-2j-\ell-1)!}. \end{aligned} \tag{2.49}$$

Proof. The first equality is explained as in the proof of Lemma 9 in [24] and the second equality follows from (2.20) and Proposition 12-(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 18) concludes the proof of Theorem 2.48.

Proposition 19. As $n \to \infty$,

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

Sketch of the proof. From Stirling's formula (similar to the proof of Proposition 12-(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 (2.49),

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

which gives the claimed result. ■

Proof of Theorem 55. Note that the inequality

$$L_n \leq \mathrm{GTC}_n \leq U_n,$$

Proposition 18, and Proposition 19 give the result. ■

2.7 The number of reticulations for a random network

This section introduces a probabilistic viewpoint for studying the typical number of reticulations in a random network. Building on the results from Section 2.6 (counting of the total number), our analysis reduces to understanding the enumeration of the number of networks with n leaves and k reticulations $\mathrm{X}_{n,k}$, which was discussed in Section 2.3. By applying the Laplace method (Section 1.2.3) to these enumeration formulas, we obtain the desired asymptotic behavior for the number of reticulations.

We first recall the formal definition of convergence in distribution, which will serve as a foundational notion for analyzing the limiting distributions of random networks in the subsequent sections.

Definition 2.7.1 (Convergence in distribution)**.** A sequence $(X_n)_{n\geq 1}$ of random variables converges in distribution (or in law) to a random variable X (denoted by $X_n \xrightarrow{\mathcal{L}} X$) if

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$$

for each continuity point $x \in \mathbb{R}$ of F_X, where F_{X_n} and F_X are the respective cumulative distribution functions.

In 2022, Fuchs, Yu and Zhang [24] gave the probabilistic results for galled networks.

Theorem 58 ([24])**.** Let I_n be the number of inner reticulations of a random galled networks with n leaves and R_n be the number of (total) reticulations of a random galled networks. The random vector $(I_n, n - R_n)$ weakly tends to a discrete limit distribution (I, R), that is,

$$(I_n, n - R_n) \xrightarrow{\mathcal{L}} (I, R),$$

where the limit law (I, R) is given by

$$\mathbb{P}(I = j, R = k) = \frac{\mathrm{e}^{-7/8}}{16^j\, j!}\left[z^{j-k}\right]\mathrm{e}^{1/(2z)}(1 + 2z + 3z^2)^j,$$

for $j \geq 0, k \geq -j$.

2.7.1 One-component d-combining tree-child network

Continuing from Theorem 36 (the exact counting for arbitrary k and n) and Theorem 50 (the total number), we are now in a position to analyze the number of reticulations in a random one-component d-combining tree-child network.

Theorem 59. Let $R_n^{(d)}$ be the number of reticulation nodes of a one-component d-combining tree-child network picked uniformly at random from the set of all one-component d-combining tree-child networks with n leaves. Then, we have the following limit behavior of $R_n^{(d)}$.

(i) For $d = 2$ (bi-combining case), we have the convergence in distribution result

$$\frac{R_n^{(2)} - n + \sqrt{n}}{\sqrt[4]{n/4}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1),$$

where $\mathcal{N}(0,1)$ denotes the standard normal distribution.

(ii) For $d = 3$, we have the convergence in distribution result

$$n - 1 - R_n^{(3)} \xrightarrow{\mathcal{L}} \text{Bessel}(1,2),$$

where Bessel(v, a) denotes the Bessel distribution, which is defined via $I_v(\alpha)$ from Theorem 50-(ii)

$$\mathbb{P}(\text{Bessel}(1,2) = k) = \frac{1}{I_1(2)k!(k+1)!}, \qquad (k \geq 0).$$

(iii) For $d \geq 4$, the limit law of $n-1-R_n^{(d)}$ is degenerate at 0, i.e., $n-1-R_n^{(d)} \xrightarrow{\mathcal{L}} \text{Dirac}(0)$, where Dirac$(\lambda)$ denotes the Dirac measure at λ.

Note that the above result for $d = 2$ is already contained in the proof of Theorem 50. Thurs, it suffices to prove the cases (ii) and (iii) for both Theorem 50 and Theorem 59.

Remark 27. If t denotes the number of tree nodes and $\tilde{n}$ the total number of nodes, then by the handshaking lemma, we have

$$t = n + (d-1)k - 1 \tag{2.50}$$

and thus,

$$\tilde{n} = 2n + dk.$$

Therefore, we have similar limit distribution results for these numbers as well.

From Theorem 36, we can now deduce Theorems 50 and 59 by the Laplace method. (A standard method of asymptotic analysis; see, e.g., Example 1.2.7 and [17, Chapter 4.7].) The number of reticulation $R_n^{(d)}$ for a random one-component d-combining tree-child network satisfy

$$\mathbb{P}\left(R_n^{(d)} = k\right) = \frac{\text{OTC}_{n,k}^{(d)}}{\text{OTC}_n^{(d)}}.$$

Proof of Theorem 50 and Theorem 59. Since the results for $d = 2$ are already contained in [23] (see also [22]), we can focus on the cases $d \geq 3$. We start with the case $d = 3$. Note that

$$\mathrm{OTC}_{n,k}^{(3)} = \binom{n}{k} \frac{(2n+k-2)!}{3^k 2^{n-1}(n-k-1)!}, \qquad (0 \leq k \leq n-1)$$

and this sequence is increasing in k. (This is in contrast to $d = 2$ where this sequence increases until its maximum at $k = n - \sqrt{n+1}$ and then decreases; see [23].) By replacing k by $n-1-k$ and using Stirling's formula, we obtain

$$\mathrm{OTC}_{n,n-1-k}^{(3)} = \frac{1}{k!(k+1)!} \cdot \frac{n(3n-3)!}{6^{n-1}} \left(1 + \mathcal{O}\left(\frac{1+k^2}{n}\right)\right) \tag{2.51}$$

uniformly for k with $k = o(\sqrt{n})$. Thus, by a standard application of the Laplace method, we get

$$\mathrm{OTC}_{n}^{(3)} \sim \left(\sum_{k\geq 0} \frac{1}{k!(k+1)!}\right) \cdot \frac{n(3n-3)!}{6^{n-1}} = I_1(2) \cdot \frac{n(3n-3)!}{6^{n-1}},$$

which is the first claim from Theorem 50, (ii); the second follows from this by another application of Stirling's formula. Moreover, since

$$\mathbb{P}(R_n^{(3)} = n-1-k) = \frac{\mathrm{OTC}_{n,n-1-k}^{(3)}}{\mathrm{OTC}_{n}^{(3)}},$$

the result from Theorem 59, (ii) follows from the above two expansions, too.

Next, we consider the case $d \geq 4$. The details of the proof are the same as above, with the main difference that the expansion (2.51) now becomes

$$\mathrm{OTC}_{n,n-1-k}^{(d)} = \left(\frac{d^2 d!}{2d^d}\right)^k \frac{1}{k!(k+1)!} \cdot n^{(3-d)k} \cdot \frac{n(dn-d)!}{(d!)^{n-1}} \left(1 + \mathcal{O}\left(\frac{1+k^2}{n}\right)\right)$$

uniformly for k with $k = o(\sqrt{n})$. For $d \geq 4$, this expansion contains the (non-trivial decreasing) factor $n^{(3-d)k}$, and therefore $\mathrm{OTC}_{n}^{(d)}$ is now asymptotically dominated by $\mathrm{OTC}_{n,n-1}^{(d)}$ (proving Theorem 50, (iii)) and the limiting distribution of $n-1-R_n^{(d)}$ is degenerate (proving Theorem 59, (iii)). ■

2.7.2 d-combing tree-child networks

This section contains the proof of Theorem 60; see below. Since the proof for $d = 2$ and $d \geq 3$ is different, we split it into two subsections (Section 2.7.2.1 and Section 2.7.2.2 below). Theorem 60 also leads to two corollaries in the Section 2.9.1.

Theorem 60. Let $R_n^{(d)}$ be the number of reticulation nodes of a d-combining tree-child network picked uniformly at random from the set of all d-combining tree-child networks with n leaves. Then, we have the following limit behavior of $R_n^{(d)}$.

(i) For $d = 2$ (bi-combining case), we have the convergence in distribution result

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

where $\text{Poisson}(\alpha)$ denotes the Poisson distribution.

(ii) For $d \geq 3$, the limit distribution of $n - 1 - R_n^{(d)}$ is degenerate at 0.

This result is new even in the case $d = 2$. In fact, as far as we are aware of, it constitutes the first limit law result for a shape parameter in random tree-child networks. Also, the result for $d = 2$ improves [35, Theorem 1.5, (iii)], which states that the number of reticulation nodes of almost all bi-combining tree-child networks with n leaves is asymptotic to n.

2.7.2.1 Bi-combining networks

In this subsection, we consider the case $d = 2$.

For convenience, we drop the super-index in the notation. We start with the following bounds for $\text{TC}_{n,k}$.

Lemma 61. For $1 \leq k \leq n - 1$,

$$\frac{n-k}{k(3n-k-3)}\text{TC}_{n,n-k} \leq \text{TC}_{n,n-1-k} \leq \frac{1}{2k}\text{TC}_{n,n-k}. \tag{2.52}$$

Proof. The upper bound follows from (2.39). For the lower bound, we generalize the argument from [23, Lemma 3]. Therefore, consider a tree-child network N with n leaves and $n - 1 - k$ reticulation nodes. From Lemma 37, we know that N has $2k$ free edges. Moreover, N has $3n - k - 2$ edges which do not end in a reticulation node; see (2.50). Now, by inserting a node into a tree edge and connecting it to a node which is inserted into a free edge, we obtain at most $2k(3n - k - 3)\text{TC}_{n,n-1-k}$ tree-child networks with n leaves and $n - k$ reticulation nodes (as those with cycles have to be discarded). On the other hand, each network is created from a latter network exactly $2(n - k)$ times. Thus,

$$2(n-k)\text{TC}_{n,n-k} \leq 2k(3n-k-3)\text{TC}_{n,n-1-k}.$$

which proves the lower bound. ■

From the above bounds, we deduce the following lemma.

Lemma 62. We have,

$$\frac{1}{3^k k!}(1 + o(1))\text{TC}_{n,n-1} \leq \text{TC}_{n,n-1-k} \leq \frac{1}{2^k k!}\text{TC}_{n,n-1}$$

uniformly in $k = o(\sqrt{n})$.

Proof. The upper bound follows from iterating the upper bound of (2.52); see also (2.39).

For the lower bound, observe that

$$\frac{n-k}{k(3n-k-3)} = \frac{1}{3k}\left(1+\mathcal{O}\left(\frac{k}{n}\right)\right).$$

Thus, by iterating the lower bound in (2.52):

$$\frac{1}{3^k k!}\left(1+\mathcal{O}\left(\frac{k}{n}\right)\right)^k \mathrm{TC}_{n,n-1} \le \mathrm{TC}_{n,n-1-k}.$$

From this the result follows since for the indicated range of k, we have

$$\left(1+\mathcal{O}\left(\frac{k}{n}\right)\right)^k = 1+\mathcal{O}\left(\frac{k^2}{n}\right) = 1+o(1).$$

This concludes the proof of the lemma. ■

We next denote by $\mathrm{F}_{n,k}$ (resp. $\mathrm{NF}_{n,k}$) the number of tree-child networks with n leaves and k reticulation nodes whose child of the root is free (resp. not free). For all n and k,

$$\mathrm{TC}_{n,k} = \mathrm{F}_{n,k} + \mathrm{NF}_{n,k}.$$

We start with an easy observation.

Lemma 63. For $1 \le k \le n-1$, we have $(2k)\mathrm{TC}_{n,n-1-k} = \mathrm{NF}_{n,n-k}$.

Proof. A tree-child network with n leaves and $n-1-k$ reticulation nodes has $2k$ free edges; see Lemma 37. Taking any of these free edges and the root edge, inserting nodes in both edges and connecting the node inserted into the root edge with the other node gives a tree-child network with n leaves and $n-k$ reticulation nodes that is not free. Moreover, this construction is clearly reversible.

■

Remark 28. Note that $\mathrm{NF}_{n,n-1} = \mathrm{TC}_{n,n-1}$. Thus, for $k = 1$, the above result shows that equality in (2.39) holds for $k = n-2$; compare with Remark 24.

Another easy observation is the following.

Lemma 64. For $1 \le k \le n-1$,

$$\mathrm{F}_{n,n-1-k} \le \frac{1}{2^{k-1}(k-1)!}\mathrm{F}_{n,n-2}.$$

Proof. A tree-child network with n leaves and $n-1-k$ reticulation nodes whose child of the root is free has $2k$ free edges of which $2k-2$ are not the edges from the child of the root to its children. By picking one of the latter two edges, one of the remaining $2k-2$ edges, inserting nodes and connecting the former edge to the latter, we obtain a tree-child network with n leaves and $n-k$ reticulation nodes whose child of the root is again free. Conversely, every such network is obtained by this construction at most 2 times. Thus,

$$2(2k-2)\mathrm{F}_{n,n-1-k} \leq 2\mathrm{F}_{n,n-k}$$

or

$$\mathrm{F}_{n,n-1-k} \leq \frac{1}{2(k-1)}\mathrm{F}_{n,n-k}.$$

Iterating this gives the claimed result. ■

The final result we need is the following.

Lemma 65. We have,

$$\mathrm{F}_{n,n-2} = \mathcal{O}\left(\frac{\mathrm{TC}_{n,n-1}}{n^{2/3}}\right).$$

Proof. Let N be a network with n leaves and $n-2$ reticulation nodes whose child of the root is free. Note that the two words constructed from N in the proof of Theorem 36 both start with a and that this is the sole letter which occurs only twice. Conversely, all words with this property arise from networks with n leaves and $n-2$ reticulation nodes whose child of the root is free. Thus, with the same arguments as in the proof of Theorem 36, we have

$$\mathrm{F}_{n,n-2} = \frac{n!}{2}g_{n-1},$$

where g_{n-1} is the number of words in $\mathcal{C}_{n-1,n-2}$ which start with a and this is the sole letter which occurs twice. Next, with the same arguments as used in the proof of Proposition 13:

$$g_n = \sum_{m\geq 1} h_{n,m}, \qquad \text{where} \qquad h_{n,m} = (n+m+n-4)\sum_{j=1}^{m} h_{n-1,j}.$$

We now apply to this sequence the same method as in the last section, where we only need an upper bound. This gives

$$g_n = \mathcal{O}\left(n!12^n e^{a_1(3n)^{1/3}} n^{-4/3}\right)$$

and thus,

$$\mathrm{F}_{n,n-2} = \mathcal{O}\left((n!)^2 12^n e^{a_1(3n)^{1/3}} n^{-7/3}\right).$$

Comparing with the Theta-result for $\mathrm{TC}_{n,n-1}$ from Theorem 51 (which was also the main result of [23]) gives the claimed result. ■

Now, we can prove the following proposition.

Proposition 20. We have,

$$\mathrm{TC}_{n,n-1-k} = \frac{1}{2^k k!}(1+o(1))\mathrm{TC}_{n,n-1}$$

uniformly for $0 \le k \le c\log n$ where $c = 1/(3\log(3/2))$.

Proof. First note that

$$\mathrm{TC}_{n,n-k} = \mathrm{NF}_{n,n-k} + \mathrm{F}_{n,n-k} = (2k)\mathrm{TC}_{n,n-1-k} + \mathrm{F}_{n,n-k}, \tag{2.53}$$

where we used Lemma 63.

Next, by using Lemma 64, Lemma 65 and the lower bound in Lemma 62, we obtain that

$$\mathrm{F}_{n,n-k} = \mathcal{O}\left(\frac{\mathrm{F}_{n,n-2}}{2^k(k-2)!}\right) = \mathcal{O}\left(\frac{\mathrm{TC}_{n,n-1}}{2^k(k-2)!n^{2/3}}\right) = \mathcal{O}\left(\left(\frac{3}{2}\right)^k \frac{k}{n^{2/3}} \times \mathrm{TC}_{n,n-k}\right)$$

for $1 \le k \le n^{1/4}$ (which is within the range of applicability of Lemma 62). Thus, for $1 \le k \le c\log n$,

$$\mathrm{F}_{n,n-k} = \mathcal{O}\left(\frac{\log n}{n^{1/3}} \times \mathrm{TC}_{n,n-k}\right).$$

Plugging this into (2.53), we obtain

$$\mathrm{TC}_{n,n-1-k} = \frac{1}{2k}\left(1+\mathcal{O}\left(\frac{\log n}{n^{1/3}}\right)\right)\mathrm{TC}_{n,n-k}$$

and by iteration

$$\mathrm{TC}_{n,n-1-k} = \frac{1}{2^k k!}\left(1+\mathcal{O}\left(\frac{\log n}{n^{1/3}}\right)\right)^k \mathrm{TC}_{n,n-1}$$

from which the result follows since

$$\left(1+\mathcal{O}\left(\frac{\log n}{n^{1/3}}\right)\right)^k = 1+\mathcal{O}\left(\frac{\log^2 n}{n^{1/3}}\right) = 1+o(1).$$

This completes the proof. ■

Remark 29. Note that the last result improves the bounds of Lemma 62, but for a smaller range of k. (The value of c in the proposition is not best possible; however, it is sufficient for our purpose.)

Proof of Theorem 60. We first consider the number of tree-child networks with n leaves. Recall that

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

Let $k^* = c\log n$ with c from the last proposition. Then,

$$\mathrm{TC}_n = \sum_{k\le k^*} \mathrm{TC}_{n,n-1-k} + \sum_{k^*<k\le n-1} \mathrm{TC}_{n,n-1-k}.$$

Using the upper bound in (62), we obtain

$$\sum_{k^*<k\le n-1} \mathrm{TC}_{n,n-1-k} \le \mathrm{TC}_{n,n-1} \sum_{k>k^*} \frac{1}{2^k k!} = o(\mathrm{TC}_{n,n-1}).$$

On the other hand, by the last proposition:

$$\begin{aligned}\sum_{k\le k^*} \mathrm{TC}_{n,n-1-k} &= (1+o(1))\mathrm{TC}_{n,n-1} \sum_{k\le k^*} \frac{1}{2^k k!}\\ &= (1+o(1))\mathrm{TC}_{n,n-1} \sum_{k=0}^{\infty} \frac{1}{2^k k!}\\ &= (1+o(1))\mathrm{TC}_{n,n-1} e^{1/2}.\end{aligned}$$

Combining the last two displays gives

$$\mathrm{TC}_n \sim e^{1/2}\mathrm{TC}_{n,n-1}, \qquad (n\to\infty). \tag{2.54}$$

Thus, for all fixed k,

$$\mathbb{P}(n-1-R_n=k) = \frac{\mathrm{TC}_{n,n-1-k}}{\mathrm{TC}_n} \longrightarrow \frac{e^{-1/2}}{2^k k!}, \qquad (n\to\infty)$$

which proves the claimed Poisson limit law. ■

Remark 30. With the same arguments as used to prove (2.54), we can also show that

$$\sum_{k=0}^{n-1} k\mathrm{TC}_{n-1-k} \sim \frac{1}{2} e^{1/2}\mathrm{TC}_{n,n-1}$$

and thus $\mathbb{E}(n-1-R_n) \sim 1/2$. Moreover, in a similar way, higher moments of $n-1-R_n$ can be shown as well to converge to those of $\mathrm{Poisson}(1/2)$.

2.7.2.2 d-combining networks with $d \ge 3$

Here, we prove Theorem 60 for $d \ge 3$. For the sake of simplicity, we restrict ourselves to the case $d=3$; the case of larger d follows along similar lines.

Recall the class of words $\mathcal{C}_n^{(d)}$ from Definition 2.3.1. Note that in Proposition 36, we constructed a bijection f from

$$f:\ \underbrace{\{0,1\}^{n-k-1} \times \mathcal{TC}_{n,k}^{(d)}}_{=:\ 2^{n-k-1}\times \mathcal{TC}_{n,k}^{(d)}} \longmapsto \mathcal{C}_{n-1,k}^{(d)} \times \mathcal{S}_n,$$

where $\mathcal{S}_n$ denotes the symmetric group of order n.

In particular, this bijection implies that $\mathcal{TC}_{n,n-1}^{(d)}$ is in bijection with $\mathcal{C}_{n-1,n-1}^{(d)} \times \mathcal{S}_n$ for $d=2,3$ ($k=n-1$) and $2\times\mathcal{TC}_{n,n-2}^{(d)}$ is in bijection with $\mathcal{C}_{n-1,n-2}^{(d)} \times \mathcal{S}_n$ for $d=2,3$ ($k=n-2$); see the

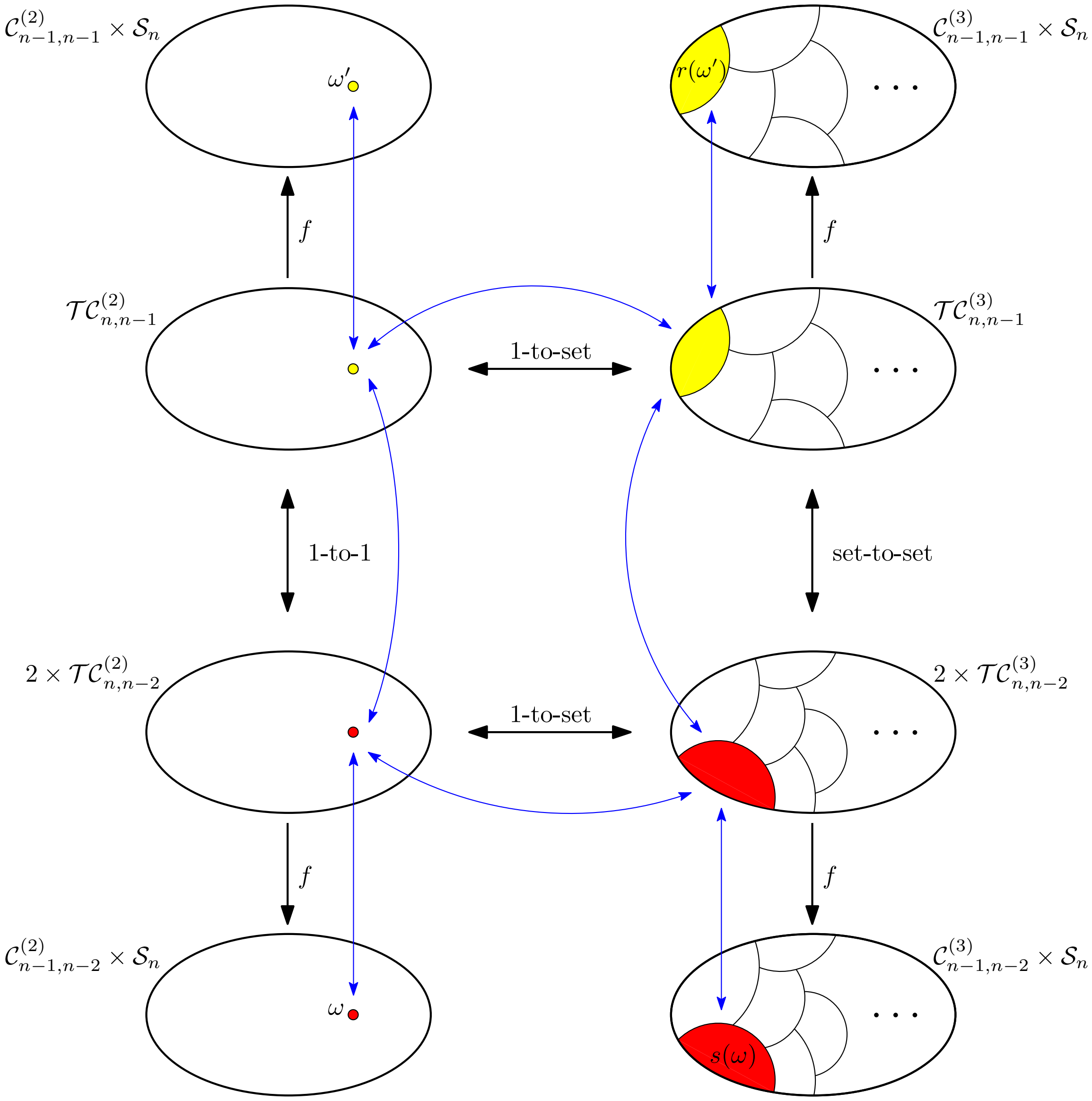


Figure 2.11: The constructions and maps which are used in the proof of Lemma 66.

upper half and lower half of Figure 2.11. Also, from Remark 24, we have a bijection from $\mathcal{TC}^{(2)}_{n,n-1}$ to $2 \times \mathcal{TC}^{(2)}_{n,n-2}$ (see the middle arrow in the left half of Figure 2.11) which gives a 2-to-1 map from $\mathcal{TC}^{(2)}_{n,n-1}$ to $\mathcal{TC}^{(2)}_{n,n-2}$ (take a maximally reticulated network and remove the reticulation node associated with the non-free edge of the child of the root, the non-free edge and its initial node; see Remark 24) followed by a 1-to-2 map from $\mathcal{TC}^{(2)}_{n,n-2}$ to $2 \times \mathcal{TC}^{(2)}_{n,n-2}$ (by picking the newly created free edge which contained the reticulation node in the previous construction).

Using this, we can now prove the following result.

Lemma 66. We have,

$$\mathrm{TC}^{(3)}_{n,n-2} = o(\mathrm{TC}^{(3)}_{n,n-1}).$$

Proof. We first explain a construction of $\mathcal{TC}^{(3)}_{n,n-1}$ from $\mathcal{TC}^{(2)}_{n,n-1}$: for a network from $\mathcal{TC}^{(2)}_{n,n-1}$, we add a new parent (and corresponding edge) for each reticulation node to an edge on the path-components we pass before we read the first parent of the reticulation node in the construction of the word and permutation from the proof of Theorem 36, where the reticulation nodes are processed consecutively (in any order). Depending on the choice of the edges, we get several networks in $\mathcal{TC}^{(3)}_{n,n-1}$ which all have essentially the same path-component structure as the network from $\mathcal{TC}^{(2)}_{n,n-1}$ (with respect to the encoding from Section 2.3.2). Conversely, removing the first parent (and corresponding edge) of each reticulation node of a network in $\mathcal{TC}^{(3)}_{n,n-1}$ gives a network in $\mathcal{TC}^{(2)}_{n,n-1}$. Thus, this gives a bijection between networks in $\mathcal{TC}^{(2)}_{n,n-1}$ and classes of networks from $\mathcal{TC}^{(3)}_{n,n-1}$, where these classes form a partition of $\mathcal{TC}^{(3)}_{n,n-1}$; see the second row in the top half of Figure 2.11.

Equivalently, when viewing networks as words and permutations (where the permutation however is here irrelevant because it does not change in the construction), any word in $\mathcal{C}^{(3)}_{n-1,n-1}$ can be obtained by adding a new ω_i at any position before the first occurrence of ω_i in a word from $\mathcal{C}^{(2)}_{n-1,n-1}$, where ω_i runs through all letters.

For example: $baaabb$ (a word in $\mathcal{C}^{(2)}_{2,2}$) leads to the class $\{bbaaaabb, babaaabb, abbaaabb\}$ (words in $\mathcal{C}^{(3)}_{2,2}$); see Figure 2.12 for the corresponding networks.

In the next step, we construct $2 \times \mathcal{TC}^{(3)}_{n,n-2}$ from $2 \times \mathcal{TC}^{(2)}_{n,n-2}$ in a similar way (where we turn networks into maximally reticulated networks as in the proof of Theorem 36). In particular, since every network in $2 \times \mathcal{TC}^{(2)}_{n,n-2}$ corresponds to a word from $\mathcal{C}^{(2)}_{n-1,n-2}$ and a permutation (which is however again irrelevant), we apply to the words from $\mathcal{C}^{(2)}_{n-1,n-2}$ the same construction as above with the only difference that we only use the ω_i's which are repeated 3 times. Like this, we obtain all words from $\mathcal{C}^{(3)}_{n-1,n-2}$.

For example, take $abbab + 132$ and $aabbb + 312$ (which encode the same network in $\mathcal{TC}_{3,1}$); $abbab$ leads to the words $\{babbab, abbbab\}$ from $\mathcal{C}^{(3)}_{2,1}$ and $aabbb$ leads to the words $\{baabbbb, ababbb, aabbbb\}$ from $\mathcal{C}^{(2)}_{2,1}$; see Figure 2.13 for a plot of the corresponding networks from $\mathcal{TC}^{(3)}_{3,1}$.

Next, as mentioned in the paragraph before the lemma, there is a bijection between $\mathcal{TC}^{(2)}_{n,n-1}$ and $2 \times \mathcal{TC}^{(2)}_{n,n-2}$. This bijection gives rise to a bijection between $\mathcal{C}^{(2)}_{n-1,n-1} \times \mathcal{S}_n$ and $\mathcal{C}^{(2)}_{n-1,n-2} \times \mathcal{S}_n$ (which just removes the first letter from a word ω' of the former to obtain a word ω of the latter); see left half of Figure 2.11. (The permutation remains unchanged.) Note that ω' is bijectively mapped onto a set of words from $\mathcal{C}^{(3)}_{n,n-1}$ and ω onto a set of words from $\mathcal{C}^{(2)}_{n,n-2}$; see the top and bottom half of Figure 2.11. Denote the cardinality of these sets by $r(\omega')$ and $s(\omega)$, respectively. Then, to show that $\mathrm{TC}^{(3)}_{n,n-2} = o(\mathrm{TC}^{(3)}_{n,n-1})$, it suffices to show that $s(\omega) = o(r(\omega'))$ uniformly over all ω in $\mathcal{C}^{(2)}_{n,n-2}$, or equivalently, we have to find a uniform lower bound of the ratio $r(\omega')/s(\omega)$ that tends to infinity. We consider this ratio next.

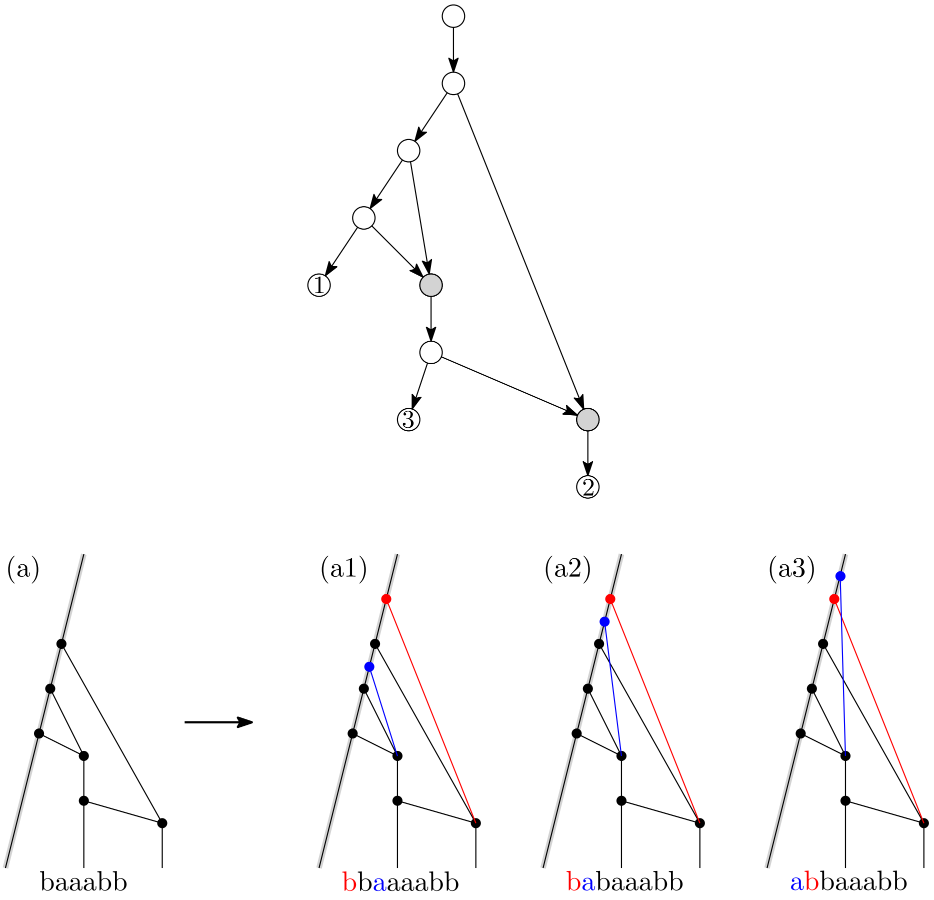


Figure 2.12: The construction of $\mathcal{TC}_{n,n-1}^{(3)}$ from $\mathcal{TC}_{n,n-1}^{(2)}$. Top: a network from $\mathcal{TC}_{n,n-1}^{(2)}$. Bottom: the three networks from $\mathcal{TC}_{n,n-1}^{(3)}$ constructed from the top network; the corresponding words are below each network. (The permutation in each case is 132; that is why we refrained from indicating it.)

For example, if $\omega = abbccabc$, then $\omega' = aabbccabc$ and the ratio becomes

$$\frac{r(aabbccabc)}{s(abbccabc)} = \frac{1\cdot 4\cdot 7}{2\cdot 5}.$$

More generally,

$$\frac{r(\omega')}{s(\omega)} = \frac{k_1\cdot k_2\cdots k_n}{(k_2-2)\cdots(k_n-2)},$$

where i denotes the ith first parent of the letters in ω' and k_i indicates the number of possibilities of adding an additional parent for i by the above method. Note that $k_1 = 1$ and the k_i's increase, thus, the $k_i/(k_i-2)$'s decrease. Moreover, note that for each k_i, we have $k_i \le 4(i-1)+1$ since the upper bound is the extremal case, i.e., the case that each of the previous $i-1$ letters occur 4 times. Consequently,

$$\frac{r(\omega')}{s(\omega)} = \frac{k_2\cdots k_n}{(k_2-2)\cdots(k_n-2)} \ge \frac{5\cdot 9\cdots(4n-3)}{3\cdot 7\ldots(4n-5)} = \Theta\left(\frac{\Gamma(n+\frac{1}{4})}{\Gamma(n-\frac{1}{4})}\right) = \Theta(n^{1/2}),$$

where we used Stirling's formula for the gamma function in the last step. This implies that, $s(\omega) = o(r(\omega'))$ which (as explained above) in turn implies that $\mathrm{TC}_{n,n-2}^{(3)} = o(\mathrm{TC}_{n,n-1}^{(3)})$. This is the claimed result. ■

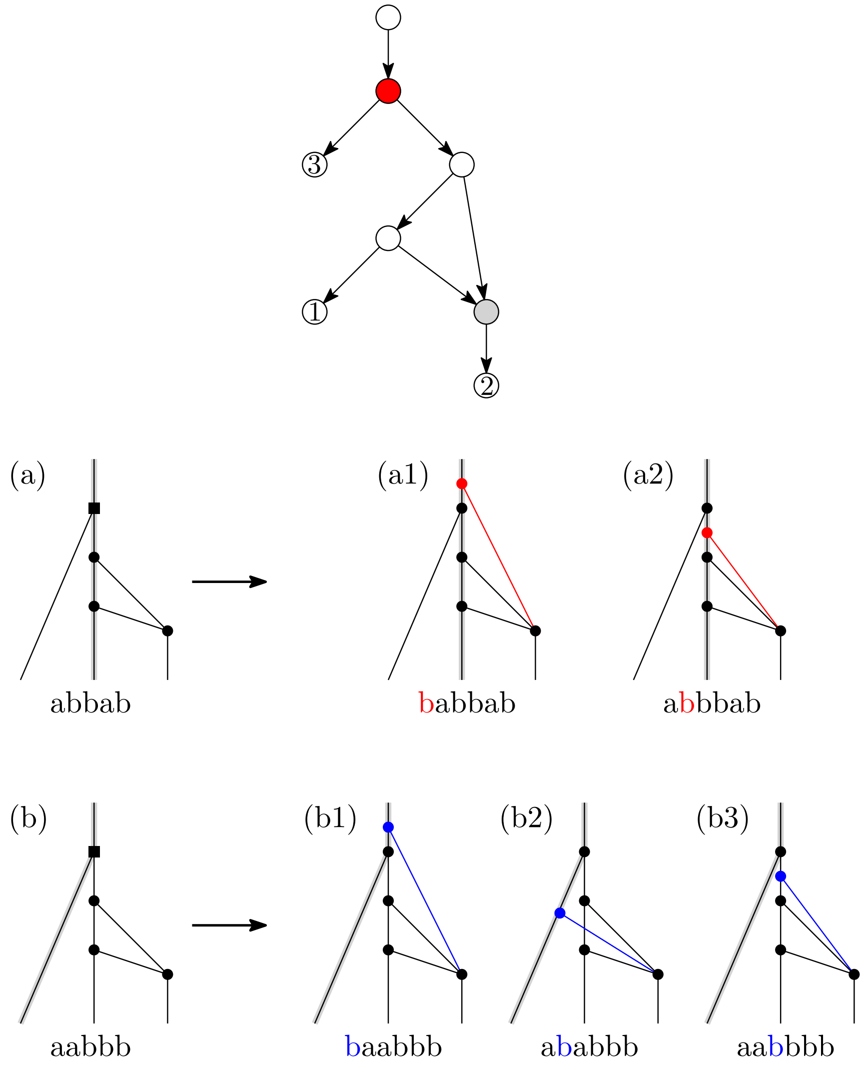


Figure 2.13: The construction of $\mathcal{TC}^{(3)}_{n,n-2}$ from $\mathcal{TC}^{(2)}_{n,n-2}$. Top: a network from $\mathcal{TC}^{(2)}_{n,n-2}$. Bottom left: the two corresponding networks from $2\times\mathcal{TC}^{(2)}_{n,n-2}$ where the first path-component (using the indexing from the proof of Theorem 36) is in bold. Bottom right: the networks from $\mathcal{TC}^{(3)}_{n,n-2}$ constructed from each of the two networks; the corresponding words are below the networks.

We can now prove the second case of Theorem 60.

Proof of Theorem 60-(ii). First, observe that from (2.39) and the previous lemma, we have

$$\mathrm{TC}^{(3)}_{n,n-1-k} = o(\mathrm{TC}^{(3)}_{n,n-1}) \tag{2.55}$$

for all fixed $k \geq 1$.

Next, by iterating (2.39),

$$\mathrm{TC}^{(3)}_{n,n-1-k} \leq \frac{1}{2^{k-1}k!}\mathrm{TC}^{(3)}_{n,n-2}$$

for all $1 \leq k \leq n-1$. Consequently,

$$\sum_{k=2}^{n-1} \mathrm{TC}^{(3)}_{n,n-1-k} = \mathcal{O}(\mathrm{TC}^{(3)}_{n,n-2}) = o(\mathrm{TC}^{(3)}_{n,n-1}).$$

Thus,

$$\mathrm{TC}_n^{(3)} = \mathrm{TC}_{n,n-1}^{(3)} + \mathrm{TC}_{n,n-2}^{(3)} + \sum_{k=2}^{n-1} \mathrm{TC}_{n,n-1-k}^{(3)} \sim \mathrm{TC}_{n,n-1}^{(3)}. \tag{2.56}$$

Now, we can prove the claim:

$$\mathbb{P}(n-1-T_n^{(3)} = k) = \mathbb{P}(T_n^{(3)} = n-1-k) = \frac{\mathrm{TC}_{n,n-1-k}^{(3)}}{\mathrm{TC}_n^{(3)}} \longrightarrow \begin{cases} 1, & \text{if } k = 0; \\ 0, & \text{if } k \geq 1, \end{cases}$$

where the last step follows from (2.55) and (2.56). ■

Remark 31. Similar as in Remark 30, we can prove that all moments of $n-1-T_n^{(3)}$ converge to 0.

2.7.3 Galled tree-child networks

In contrast to bi-combining tree-child networks and galled networks, the limit law of R_n (suitably scaled) for galled tree-child networks is continuous.

Theorem 67. The number of reticulation nodes R_n of a random galled tree-child networks satisfies, as $n \to \infty$,

$$\frac{R_n - \mathbb{E}(R_n)}{\sqrt{\mathrm{Var}(R_n)}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1),$$

where $\mathcal{N}(0,1)$ denotes the standard normal distribution. Moreover,

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

The above results show that galled tree-child networks behave quite different from both bi-combining tree-child networks (Theorem 60) and galled networks (Theorem 58). That is one reason why we find them interesting.

By refining the method of proving Theorem 55 (see Section 6 of [24] where the same was done for galled networks), we obtain the following result which implies Theorem 67.

Theorem 68. Let I_n be the number of inner reticulations of a random galled tree-child network of size n and R_n be the number of reticulation vertices. Then, as $n \to \infty$,

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

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

2.8 Sackin Index

In this section, we investigate a shape parameter for random one-component d-combining tree-child networks, namely, a Sackin index. Until the end of this section, we assume that N is a one-component d-combining tree-child network with n leaves and k reticulation nodes. Let us first state the definition of the Sackin index from [43].

Definition 2.8.1 (Sackin index)**.** The Sackin index of N, in symbols $S(N)$, is defined as the sum over all leaves of the lengths of the longest paths to the leaves.

For phylogenetic trees, the investigation of this index has a long history and many results have been proved; see [16, 15] for a summary of some of these results. For (bi-combining) one-component tree-child networks, [43] gave the first generalization of the Sackin index to networks. We simplify this approach and extend the analysis to the d-combining case. In 2025, Fuchs and Gittenberger [18] obtained results for galled-tree networks. For instance, they proved the following.

Theorem 69 ([18])**.** Let S_n be the Sackin index of a random galled-tree network with n leaves. Then,

$$\mathbb{E}(S_n) \sim \sqrt{\pi}\left(\frac{47}{32}+\frac{38\sqrt{17}}{544}\right)n^{3/2}.$$

2.8.1 One-component d-combining tree-child networks

We now turn to one-component d-combining tree-child networks. For the analysis, following [43], we define a second index.

Definition 2.8.2 (Top Tree Component)**.** The *top tree component* of N, in symbols $\mathrm{T}_{\mathrm{top}}(N)$, is defined as the tree obtained from N by deleting all reticulation nodes together with their incident edges and their leaves below.

Note that we retain the parents of all reticulation nodes and thus the top tree component is *not* a binary tree, it also has unary nodes. We denote by $P(N)$ the total path length of $\mathrm{T}_{\mathrm{top}}(N)$, i.e., the sum over all root-distances of all vertices. This index is related to the Sackin index as follows.

Lemma 70. For all one-component tree-child networks N with at least two leaves, we have for $d = 2$,

$$S(N) \le P(N) + 1 \le 2S(N)$$

and for $d \geq 3$,

$$S(N) \le P(N) \le dS(N). \tag{2.57}$$

Consequently, regardless of the value of d, $S(N) = \Theta(P(N))$.

Proof. The result for $d = 2$ was given in [43] and the ideas of the proof in [43] can be used to handle also the case $d \geq 3$. For the readers convenience, we give some of the details.

First, we define the following sets of vertices of N:

(i) L_T collects leaves of N which are not below a reticulation node;

(ii) L_R collects leaves of N which are below a reticulation node;

(iii) P collects the parents of all reticulation nodes;

(iv) $R = V(\mathrm{T}_{\text{top}}(N)) \setminus (L_T \cup P)$ collects the remaining vertices in $\mathrm{T}_{\text{top}}(N)$. (Here, $V(\mathrm{T}_{\text{top}}(N))$ denotes the set of vertices of $\mathrm{T}_{\text{top}}(N)$.)

Then,

$$S(N) = \sum_{v \in L_T \cup L_R} \mathrm{depth}(v),$$

where the depth of v is the longest distance to the root.

Now, in order to prove the lower bound of $P(N)$, note that for $v \in L_R$, we have

$$\mathrm{depth}(v) = 2 + \max\{\mathrm{depth}(p) \ : \ p \text{ is a grandparent of } v\} \leq \sum_{p \text{ is a grandparent of } v} \mathrm{depth}(p).$$

Here, for the last inequality, we used that $d \geq 3$. Thus,

$$S(N) \leq \sum_{v \in L_T \cup P} \mathrm{depth}(v) \leq \sum_{v \in L_T \cup P \cup R} \mathrm{depth}(v) = P(N)$$

which shows the first part of the claim in (2.57).

Next, in order to show the upper bound of $P(N)$, we first recall that there exists a bijection ϕ from R to L_T such that w is an ancestor of $\phi(w)$ for all $w \in R$; see the appendix of [43] for details. Consequently, $\mathrm{depth}(w) \leq \mathrm{depth}(\phi(w))$ and thus,

$$\begin{aligned} P(N) &\leq \sum_{v \in L_T} \mathrm{depth}(v) + d \cdot \sum_{v \in L_R} \mathrm{depth}(v) + \underbrace{\sum_{w \in R} \mathrm{depth}(\phi(w))}_{= \sum_{v \in L_T} \mathrm{depth}(v)} \\ &= 2 \sum_{v \in L_T} \mathrm{depth}(v) + d \sum_{v \in L_R} \mathrm{depth}(l) \leq d\, S(N). \end{aligned}$$

This completes the proof. ■

We next analyze $P(N)$. (Recall that, in Section 1.1.3.4, we denote by $L_{n,k}$ the number of one-component tree-child networks of size n and k reticulations .) Denote by $\mathcal{L}_{n,k}^{(d)}$ the set of all one-component d-combining tree-child networks with n leaves and k reticulation nodes where the leaves below the reticulation nodes are labeled by the k *largest* labels from $\{1,\dots,n\}$ and $\mathrm{L}_{n,k}^{(d)}$ the cardinality of $\mathcal{L}_{n,k}^{(d)}$. Note that

$$\mathrm{OTC}_{n,k}^{(d)}=\binom{n}{k}\mathrm{L}_{n,k}^{(d)},$$

since all one-component tree-child networks are obtained from the networks in $\mathcal{L}_{n,k}^{(d)}$ by re-labeling the leaves. Also, set

$$P(\mathcal{L}_{n,k}^{(d)}):=\sum_{N\in\mathcal{L}_{n,k}^{(d)}}P(N).$$

For this quantity, we obtain a (surprisingly) simple recurrence which then yields an exact formula; see [43, Theorem 3] for the bi-combining case.

We first prove this recurrence with a computational approach (which extends and simplifies the one from [43]). Then, we give a simpler and more elegant combinatorial proof of the solution of the recurrence in two (long) remarks below the proof (Remark 32 and Remark 33). (This second proof however requires the knowledge of the final result which we found with the first proof.) The reader who is only interested in the combinatorial proof might immediately skip to these remarks as the arguments in the second proof are independent from the arguments in the first proof.

Proposition 21. $P(\mathcal{L}_{n,k}^{(d)})$ satisfies the recurrence

$$P(\mathcal{L}_{n,k}^{(d)})=\binom{2n+(d-2)k}{d}P(\mathcal{L}_{n-1,k-1}^{(d)})$$

with $P(\mathcal{L}_{n,0}^{(d)})=(2n)!!-(2n-1)!!$ as initial condition. Consequently,

$$P(\mathcal{L}_{n,k}^{(d)})=\frac{(2n+(d-2)k)!}{(d!)^k(2n-2k)!}\big((2n-2k)!!-(2n-2k-1)!!\big). \tag{2.58}$$

Proof. Let N be a network in $\mathcal{L}_{n-1,k-1}^{(d)}$. We subsequently denote by $V(\mathrm{T}_{\mathrm{top}}(N))$ and $E(\mathrm{T}_{\mathrm{top}}(N))$ the vertex set and edge set of the top tree component, respectively. Moreover, for a $v\in V(\mathrm{T}_{\mathrm{top}}(N))$, we denote by $\delta_{\mathrm{T}_{\mathrm{top}}(N)}(v)$ the number of descendants and by $\alpha_{\mathrm{T}_{\mathrm{top}}(N)}(v)$ the number of ascendants of v in $\mathrm{T}_{\mathrm{top}}(N)$. Note that the number of ascendants of v in N and $\mathrm{T}_{\mathrm{top}}(N)$ is the same, i.e., $\alpha_{\mathrm{T}_{\mathrm{top}}(N)}(v)=\alpha_N(v)$. Finally, we denote by $\mathcal{S}$ the set of multi-sets of d edges of $\mathrm{T}_{\mathrm{top}}(N)$ (where edges are counted with repetition). Every $S\in\mathcal{S}$ corresponds to a set of edges where nodes are inserted in order to add an additional reticulation node with label n. This notion allows us to construct all networks of $\mathcal{L}_{n,k}^{(d)}$ from those of $\mathcal{L}_{n-1,k-1}^{(d)}$. More precisely, for $S\in\mathcal{S}$, we construct a network $N'=N'(S)$ by inserting nodes into the d edges from S and connecting them with a new reticulation node whose child has label n. (Note that this is a similar construction to the one we used in the proof of Theorem 35.)

We consider now $P(N')$ which by definition is given by:

$$P(N') = \sum_{v \in V(\mathrm{T}_{\mathrm{top}}(N'))} \alpha_{N'}(v).$$

Partitioning vertices of $\mathrm{T}_{\mathrm{top}}(N')$ into vertices of $\mathrm{T}_{\mathrm{top}}(N)$ and the d new ones which have been added to N to construct N', we have

$$P(N') = \sum_{v \in V(\mathrm{T}_{\mathrm{top}}(N))} \alpha_{N'}(v) + \sum_{v \in D} \alpha_{N'}(v),$$

where $D = V(\mathrm{T}_{\mathrm{top}}(N')) \setminus V(\mathrm{T}_{\mathrm{top}}(N))$ is the set of new vertices.

Now, to find a relation to $P(N)$, we replace $\alpha_{N'}$ by α_N and get

$$P(N') = \sum_{v \in V(\mathrm{T}_{\mathrm{top}}(N))} \alpha_N(v) + \sum_{v \in V(\mathrm{T}_{\mathrm{top}}(N))} \beta(v) + \sum_{v \in D} \alpha_N(v^{\downarrow}) + \sum_{v \in D} \gamma(v), \tag{2.59}$$

where

(i) $\beta(v)$ denotes the number of nodes in D that are ascendants of v in $\mathrm{T}_{\mathrm{top}}(N')$;

(ii) $v^{\downarrow}$ is closest descendant of v in $\mathrm{T}_{\mathrm{top}}(N')$ that belongs to $V(\mathrm{T}_{\mathrm{top}}(N))$;

(iii) $\gamma(v)$ denotes the number of nodes in D that are ascendants of v in $\mathrm{T}_{\mathrm{top}}(N')$.

Observe that the second sum can be rewritten as:

$$\sum_{v \in V(\mathrm{T}_{\mathrm{top}}(N))} \beta(v) = \sum_{v \in D} \delta_{\mathrm{T}_{\mathrm{top}}(N)}(v^{\uparrow}),$$

where $v^{\uparrow}$ is the closest ascendant of v in $\mathrm{T}_{\mathrm{top}}(N')$ which belongs to $V(\mathrm{T}_{\mathrm{top}}(N))$. Thus, (2.59) can be rewritten into:

$$P(N') = P(N) + \sum_{v \in D} \left(\delta_{\mathrm{T}_{\mathrm{top}}(N)}(v^{\uparrow}) + \alpha_N(v^{\downarrow})\right) + \sum_{v \in D} \gamma(v).$$

Next, we sum both sides over $S \in \mathcal{S}$, which on the right-hand side gives three sums which we denote by Σ_1, Σ_2 and Σ_3, respectively.

First, for Σ_1, note that $E := |E(\mathrm{T}_{\mathrm{top}}(N))| = 2(n-k) - 1 + dk$. Thus,

$$\Sigma_1 = \sum_{S \in \mathcal{S}} P(N) = |\mathcal{S}| P(N) = \binom{E+d-1}{d} P(N)$$

since

$$|\mathcal{S}| = \#\{x_1 + \cdots + x_E = d \ : \ x_i \text{ non-negative integers}, \ 1 \le i \le E\},$$

where each x_i corresponds to an edge from $E(\mathrm{T}_{\mathrm{top}}(N))$ (and counts how many times that edge occurs in the set S).

Next, for Σ_2, set $S = \{e_1, \dots, e_d\}$ and $B(e) := \delta_{\mathrm{T}_{\mathrm{top}}(N)}(v^{\uparrow}) + \alpha_N(v^{\downarrow})$, where e is an edge in S into which v is inserted and $v^{\uparrow}$ and $v^{\downarrow}$ are initial and end point of e, respectively. Then,

$$\Sigma_2 = \sum_{S \in \mathcal{S}} \left(B(e_1) + \cdots + B(e_d) \right).$$

We fix an edge e in $\mathrm{T}_{\mathrm{top}}(N)$ and count the number of times $B(e)$ occurs in the above sum. This gives,

$$\Sigma_2 = \sum_{e \in E(\mathrm{T}_{\mathrm{top}}(N))} \left(\sum_{x_1 + \cdots + x_E = d} x_i \right) B(e) = 2 \binom{E+d-1}{d-1} P(N),$$

where x_i inside the second sum is the term corresponding to edge e and for the third expression, we used that

$$\sum_{x_1 + \cdots + x_E = d} x_i = \frac{1}{E} \sum_{x_1 + \cdots + x_E = d} x_1 + \cdots + x_E = \binom{E+d-1}{d-1},$$

which holds for all $1 \le i \le E$ by symmetry. Moreover, we used that

$$\sum_{e \in E(\mathrm{T}_{\mathrm{top}}(N))} B(e) = 2\, P(N).$$

Finally, for Σ_3, we first give an explicit formula

$$\Sigma_3 = \sum_{v \in \mathrm{T}_{\mathrm{top}}(N) \setminus \{\rho\}} \sum_{0 \le i+j \le d} \left(ij + \frac{i(i-1)}{2} \right) \binom{\alpha_N(v) - 2 + j}{j} \binom{E - \alpha_N(v) - 1 + d - i - j}{d - i - j},$$

by the following combinatorial steps:

1. Choose a vertex v in $\mathrm{T}_{\mathrm{top}}(N)$ which is not the root vertex ρ. We consider the incoming edge e of v;

2. Choose pair (i, j) with $0 \le i + j \le d$. Here, the meaning of i is that i nodes are inserted into e and j nodes are inserted into an edge which lies on the path from ρ to v (excluding e);

3. So far, the contribution of v to Σ_3 is
$$ij + \sum_{\ell=1}^{i-1} \ell = ij + \frac{i(i-1)}{2};$$

4. Next, there are $\binom{\alpha_N(v) - 2 + j}{j}$ ways of choosing the edges into which the j nodes are inserted;

5. Finally, the second binomial coefficient counts the number of ways that the remaining $d - i - j$ new nodes are inserted into edges which are not on the path from ρ to v.

In order to simplify the formula, we set

$$\alpha := \alpha_N(v), \qquad A := \alpha - 2, \qquad B := E - \alpha - 1, \qquad M := d - i.$$

Then, for the inner sum Σ_3, we have

$$\begin{aligned}
&\sum_{i=0}^{d}\left(i\sum_{j=0}^{M} j\binom{A+j}{j}\binom{B+M-j}{M-j} + \frac{i(i-1)}{2}\sum_{j=0}^{M}\binom{A+j}{j}\binom{B+M-j}{M-j}\right)\\
&= \sum_{i=0}^{d}\left(i(\alpha-1)[z^{M-1}](1-z)^{-A-2}(1-z)^{-B-1} + \frac{i(i-1)}{2}[z^{M}](1-z)^{-A-1}(1-z)^{-B-1}\right)\\
&= (\alpha-1)\sum_{i=0}^{d} i\binom{A+B+1+M}{M-1} + \sum_{i=0}^{d}\frac{i(i-1)}{2}\binom{A+B+1+M}{M}\\
&= (\alpha-1)[z^{d-2}](1-z)^{-A-B-5} + [z^{d-2}](1-z)^{-A-B-5}\\
&= \alpha\binom{E+d-1}{d-2}.
\end{aligned}$$

Plugging this into the above formula for Σ_3 gives

$$\Sigma_3 = \binom{E+d-1}{d-2}\sum_{v\in V(\mathrm{T}_{\mathrm{top}}(N))\setminus\{\rho\}} \alpha_N(v) = \binom{E+d-1}{d-2}P(N).$$

Finally, summing the above expressions for Σ_1, Σ_2 and Σ_3 and summing over all N in $\mathcal{L}_{n-1,k-1}^{(d)}$ gives the surprisingly simple recurrence:

$$P(\mathcal{L}_{n,k}^{(d)}) = \binom{2n+(d-2)k}{d} P(\mathcal{L}_{n-1,k-1}^{(d)}). \tag{2.60}$$

The initial condition, namely, the formula for $P(\mathcal{L}_{n,0}^{(d)}) = (2n)!! - (2n-1)!!$ is well-known; see, e.g., [43]. From this, (2.58) is obtained by iteration. This concludes the proof. ■

Remark 32. A *unary-binary tree* is a rooted tree whose nodes are either leaves (out-degree 0), unary (out-degree 1), or binary (out-degree 2). Then, we observe that the set $\mathrm{T}_{\mathrm{top}}(\mathcal{L}_{n,k}^{(d)})$ of top trees obtained from all elements of $\mathcal{L}_{n,k}^{(d)}$ is equal to the set of all unary-binary trees with $n-k$ labeled leaves and dk unary nodes. Hence, (2.58) nearly gives the path length of unary-binary trees, up to an overcount (which is actually a factor) that is related to reticulation nodes. Now, we adapt the construction of the proof of Theorem 35 to build $\mathcal{L}_{n,k}^{(d)}$ in order to determine this factor. Steps (1) and (2) remain the same, and step (4) is not performed as the leaves of reticulation nodes have maximal labels in $\mathcal{L}_{n,k}^{(d)}$. It remains to consider step (3), which needs to be adapted. Observe that adding reticulation nodes and unary edges corresponds to the factor $\binom{dk}{d,d,\ldots,d}$; see Figure 2.14. Thus, dividing (2.58) by this factor, we get that the path length of unary-binary trees with $n-k$ labeled leaves and dk unary nodes is equal to

$$\big((2(n-k))!! - (2(n-k)-1)!!\big) \cdot \binom{2(n-k)+dk}{dk}. \tag{2.61}$$

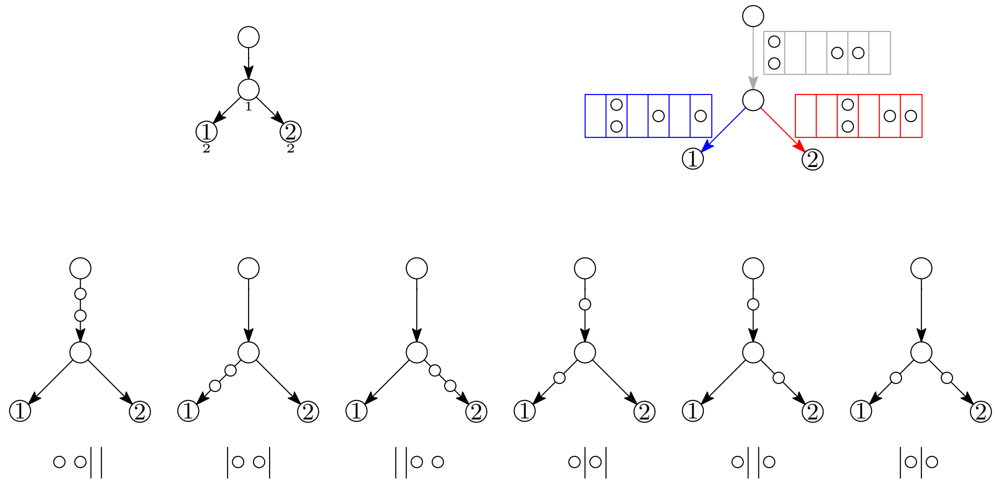


Figure 2.14: Construction of $\mathcal{L}^{(2)}_{3,1}$ in Remark 33 ($K=2$, $M=3$). **Top, left**: The only phylogenetic tree with 2 leaves; path lengths below each node and total path length 5; **Bottom**: $\binom{4}{2}=6$ top trees $\mathcal{L}^{(2)}_{2,1}$ created from it after adding 2 unary nodes and the respective ”balls and bars” diagrams; **Top, right**: Superposition of all 6 top trees.

Remark 33. The surprising simplifications of Proposition 21 also have a direct combinatorial explanation. By Remark 32, it suffices to show that the path length of unary-binary trees with $n-k$ labeled leaves and dk unary nodes is equal to (2.61).

The main idea is to use again the construction from the proof of Theorem 35 to build the networks in $\mathcal{L}^{(d)}_{n,k}$ bottom-up. We start in step (1) with a phylogenetic tree (i.e., without reticulation nodes) with $\ell := n-k$ leaves that is weighted by the path length. In the sequel, we call the nodes/edges of this phylogenetic tree, the original nodes/edges. As noted before, the total weight of such trees is $P(\mathcal{L}^{(d)}_{\ell,0}) = (2\ell)!! - (2\ell-1)!!$. Then, we place $K := dk$ unary nodes along the $M := 2\ell-1$ edges. This gives a new structure enumerated by

$$P(\mathcal{L}^{(d)}_{\ell,0})\binom{K+M-1}{K}. \tag{2.62}$$

Note that this is nearly equal to what we want to prove in (2.61), except that the binomial coefficient should be $\binom{K+M+1}{K}$. What remains to be done, is to properly change the weights, as they do not correspond to the path lengths anymore, due to the additional unary nodes.

Let us start with a simple observation: Set all edge weights to w in a phylogenetic tree with ℓ leaves. Then the path length is equal to w times the path length of the unweighted tree.

Having this observation in mind, we interpret unary nodes as weights on the original edges. We do this in three steps; the process is visualized in Figures 2.14 and 2.15 for $\mathcal{L}_{3,1}^{(2)}$. The main idea is to superimpose all $\binom{K+M-1}{K}$ created trees and group them cleverly.

First, we "forget" about the unary nodes, and interpret the $\binom{K+M-1}{K}$ new instances of a phylogenetic tree as the same phylogenetic tree. Hence, the weights of all edges change to $\binom{K+M-1}{K}$ and the total number of these weighted phylogenetic trees is (2.62). In the sequel, the following direct interpretation of the binomial coefficient using "balls and bars" is useful: The K unary nodes correspond to balls and the edges M to bins, which are modeled by $M-1$ bars. Then each distribution of unary nodes on the edges corresponds to choosing K out of M bins with repetitions allowed; see Figure 2.14.

Second, we correct the weights of all original edges. Note that each unary node splits an edge into two and thereby increases the path length of any node below by one, or, equivalently, increases the weight of the associated original edge by one. Hence, the weight of an edge increases by the number of unary nodes it contains. In the superposition of all trees, all edges have the same number of unary nodes, as each edge is chosen equally often. The weight is equal to the sum of all unary nodes placed onto a fixed edge among the $\binom{K+M-1}{K}$ configurations. As there are M edges and K unary nodes (all equally distributed) we need a factor of K/M. Thus,

$$P(\mathcal{L}_{\ell,0}^{(d)})\binom{K+M-1}{K}\left(1+\frac{K}{M}\right) = P(\mathcal{L}_{\ell,0}^{(d)})\binom{K+M}{K}.$$

A direct interpretation of the last formula is as follows: Fix the root edge. Mark one of the unary nodes of the root edge. Like this, we create as many instances (with different markers) as there are unary nodes on the root edge in all configurations, i.e., we count the total number of unary nodes on the root edge. We model this marker, by splitting the root edge just after this marker into two parts. Hence, we have now $M+1$ edges to distribute K unary nodes. We interpret the case when the first part of the root edge is empty, as the $\binom{K+M-1}{K}$ created weighted trees from placing unary nodes. Thus, there are $\binom{K+M}{K}$ many choices.

Third, we assign weights for the new unary nodes. Consider an original edge with i unary nodes. To give the lowest one its correct weight, we remove the original node just below and replace it by the lowest unary node. Then, we may use the idea of the previous step: The weight of the edge above is now i, as $i-1$ unary nodes remain. In the superposition, each such sequence of i unary nodes appears on each original edge the same number of times. Hence, we collect a weight i on each edge, and the total weight gives the correct weights for all lowest unary nodes. Then, we repeat this process for the next unary node, with an edge weight $i-1$, etc.

In total, an original edge with i unary nodes gives rise to a total weight of $1 + 2 + \cdots + i$. For an arbitrary edge, this is now, analogously to before, equal to the sum of unary nodes that are before the marked node from the previous step. Thus, this gives a factor $K/(M+1)$ and we get

$$P(\mathcal{L}_{\ell,0}^{(d)})\binom{K+M}{K}\left(1+\frac{K}{M+1}\right) = P(\mathcal{L}_{\ell,0}^{(d)})\binom{K+M+1}{K}.$$

Note, as before, this has a direct combinatorial interpretation, where we split the first part of the root edge again. Hence, there are now $M+2$ edges to distribute K unary nodes, as claimed. Moreover, note that the last formula also gives the path length of unary binary trees with ℓ labeled leaves and K unary nodes ($M = 2\ell - 1$); compare with Remark 32.

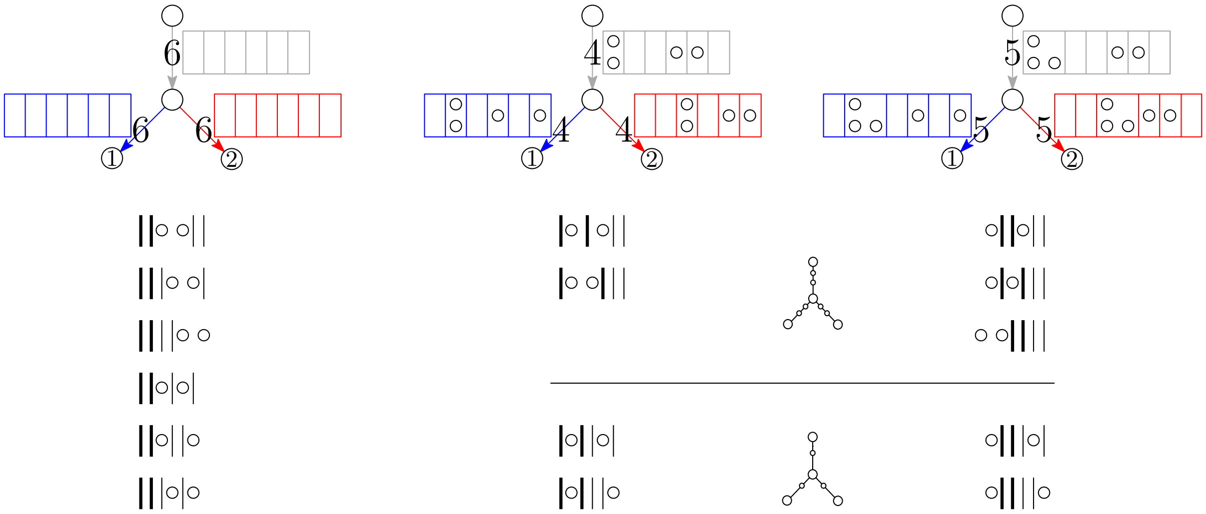


Figure 2.15: Enumeration of $P(\mathcal{L}_{3,1}^{(2)})$ in Remark 33 using superposition ($K = 2$, $M = 3$); see Figure 2.14. **Left** Step 1: Each new instance increases edge weight by one; total $\binom{4}{2} = 6$; **Middle** Step 2: Correct path length of original nodes; sum unary nodes per edge; **Right** Step 3: Add path length for unary nodes; for i nodes weight $1 + 2 + \cdots + i$; **Bottom** "Balls and bars" corresponding to each step; the two added bars are shown bold.

Using Proposition 21, we are now ready to complete our analysis. Let $P_n^{(d)}$ be the path length of the top tree component of a random one-component tree-child network with n leaves. Then,

$$\mathbb{E}(P_n^{(d)}) = \frac{\sum_{k=0}^{n-1}\binom{n}{k}P(\mathcal{L}_{n,k}^{(d)})}{\mathrm{OTC}_n^{(d)}}.$$

Applying the Laplace method to the numerator and using Theorem 50 gives the following result.

Proposition 22. (i) For $d = 2$ (bi-combining case), we have

$$\mathbb{E}(P_n^{(2)}) \sim 2\sqrt{\pi} n^{7/4}.$$

(ii) For $d = 3$, we have

$$\mathbb{E}(P_n^{(3)}) \sim \frac{9(\cosh(2) - I_0(2))}{2I_1(2)} n^2,$$

where the constant is approximately 4.19438713.

(iii) For $d \geq 4$, we have

$$\mathbb{E}(P_n^{(d)}) \sim \frac{d^2}{2} n^2.$$

Proof. We start with the bi-combining case ($d = 2$). Using the result from the Proposition 21 yields

$$\begin{aligned}\sum_{k=0}^{n-1} \binom{n}{k} P(\mathcal{L}_{n,k}^{(2)}) &= \sum_{k=0}^{n-1} \binom{n}{k+1} P(\mathcal{L}_{n,n-1-k}^{(2)}) \\ &= \sum_{k=0}^{n-1} \binom{n}{k+1} \frac{(2n)!}{2^{n-1-k}(2k+2)!} \left(2^{k+1}(k+1)! - \frac{(2k+1)!}{2^k k!} \right).\end{aligned}$$

We break the last sum S into two sums, i.e., $S = S_1 + S_2$ according to the two terms in the bracket. Thus,

$$S_1 = \frac{n!(2n)!}{2^{n-2}} \sum_{k=0}^{n-1} \frac{4^k}{(2k+2)!(n-1-k)!}$$

and we have a similar expression for S_2. Note that the terms inside the sum increase until a positive integer k^* with $k^* = \sqrt{n} + \mathcal{O}(1)$ and decrease afterwards. Moreover, by using Stirling's formula, we see that

$$\frac{4^k}{(2k+2)!(n-1-k)!} = \frac{1}{8\pi\sqrt{2e}} n^{-3/4} e^{2\sqrt{n}} e^n n^{-n} e^{-x^2/\sqrt{n}} \left(1 + \mathcal{O}\left(\frac{1+|x|}{\sqrt{n}} + \frac{x^3}{n} \right) \right)$$

uniformly for $|x| \leq n^{3/10}$ where $k = \sqrt{n} + x$. Consequently, from a standard application of the Laplace method:

$$\begin{aligned}\sum_{k=0}^{n-1} \frac{4^k}{(2k+2)!(n-1-k)!} &\sim \frac{1}{8\pi\sqrt{2e}} n^{-3/4} e^{2\sqrt{n}} e^n n^{-n} \int_{-\infty}^{\infty} e^{-x^2/\sqrt{n}} \mathrm{d}x \\ &= \frac{1}{8\sqrt{2e\pi}} n^{-1/2} e^{2\sqrt{n}} e^n n^{-n}\end{aligned}$$

and thus,

$$S_1 \sim \frac{n!(2n)!}{2^{n-2}} \cdot \frac{1}{8\sqrt{2e\pi}} n^{-1/2} e^{2\sqrt{n}} e^n n^{-n} \sim \frac{1}{2\sqrt{e}} (2n)! 2^{-n} e^{2\sqrt{n}},$$

where we again used Stirling's formula. Similarly, we can derive the asymptotics of S_2 which shows that S_2 is of a smaller asymptotic order, i.e., $S_2 = o(S_1)$. Consequently, $S \sim S_1$. Finally, dividing by the asymptotics of $\mathrm{OTC}_n^{(2)}$ from Theorem 50-(i) and using (once more) Stirling's formula gives the claimed result.

Next, for $d = 3$, we first note that

$$\binom{n}{k} P(\mathcal{L}_{n,k}^{(3)}) = \binom{n}{k} \frac{(2n+k)!}{6^k(2n-2k)!} \left(2^{n-k}(n-k)! - \frac{(2n-2k-1)!}{2^{n-k-1}(n-k-1)!} \right)$$

is increasing in k with $0 \le k \le n-1$. By replacing k by $n-1-k$ and using Stirling's formula, we obtain that

$$\binom{n}{k+1} P(\mathcal{L}_{n,n-1-k}^{(3)}) = \left(\frac{4^{k+1}}{(2k+2)!} - \frac{1}{(k+1)!^2} \right) \frac{(3n)!}{6^n} \left(1 + \mathcal{O}\left(\frac{1+k^2}{n} \right) \right),$$

uniformly for k with $k = o(\sqrt{n})$. Thus, by an another application of the Laplace method,

$$\sum_{k=0}^{n-1} \binom{n}{k+1} P(\mathcal{L}_{n,n-1-k}^{(3)}) \sim \left(\sum_{k\ge 1} \frac{4^k}{(2k)!} - \frac{1}{k!^2} \right) \frac{(3n)!}{6^n} = \big(\cosh(2) - I_0(2)\big) \frac{(3n)!}{6^n},$$

where $I_0(2)$ is the modified Bessel function (see Theorem 50-(ii)). Dividing by the asymptotics of $\mathrm{OTC}_n^{(3)}$ (again see Theorem 50-(ii)), we have

$$\mathbb{E}(P_n^{(3)}) \sim \frac{9(\cosh(2) - I_0(2))}{2I_1(2)} n^2.$$

This proves the claim in this case. (Note the similarity of this proof to the one of Theorem 59-(ii).)

For $d \ge 4$, with similar arguments as in the proof of part (iii) of Theorem 59:

$$\sum_{k=0}^{n-1} \binom{n}{k+1} P(\mathcal{L}_{n,n-1-k}^{(d)}) \sim n\, P(\mathcal{L}_{n,n-1}^{(d)}) = \frac{n\,(dn-d+2)!}{2\,(d!)^{n-1}}.$$

Dividing by the asymptotics of $\mathrm{OTC}_n^{(d)}$ from part (iii) in Theorem 59, we have

$$\mathbb{E}(P_n^{(d)}) \sim \frac{d^2}{2} n^2,$$

which proves the claimed result also in this case. ■

Remark 34. A (slightly) weaker version of the above result for the bi-combining case was derived in [43, Proposition 2].

Denote by $S_n^{(d)}$ the Sackin index of a random one-component tree-child network with n leaves. Then, by combining the last proposition with Lemma 70, we obtain the main result of this subsection. (This result for $d = 2$ was also the main result of [43].)

Theorem 71. (i) For $d = 2$ (bi-combining case), we have

$$\mathbb{E}(S_n^{(d)}) = \Theta(n^{7/4}).$$

(ii) For $d \geq 3$, we have

$$\mathbb{E}(S_n^{(d)}) = \Theta(n^2).$$

2.9 Other

This section gathers some minor topics and proofs that do not fit well into the previous subsections.

2.9.1 Corollaries of Theorem 60 (d-combining tree-child network)

Theorem 60 can be used to improve and extend [35, Proposition 1.6, (ii)], which was concerned with the number of *twigs* of (bi-combining) tree-child networks. A twig is a tree node which is contained in a *pendant subtree*, i.e., a tree node that has no reticulation node as descendant. In [35], it was proved that the number of twigs in a random bi-combining tree-child network is $o(n)$. In fact, twigs are even rarer than that.

Corollary 2.9.1. Let $W_n^{(d)}$ be the number of twigs of a d-combining tree-child network picked uniformly at random from the set of all d-combining tree-child networks with n leaves.

(i) For $d = 2$ (bi-combining), we have

$$\mathbb{E}(W_n^{(d)}) = \mathcal{O}(1).$$

(ii) For $d \geq 3$, the limit law of $W_n^{(d)}$ is degenerate at 0. More precisely,

$$\mathbb{E}(W_n^{(d)}) \longrightarrow 0.$$

Proof. Note that

$$\mathbb{E}(W_n^{(d)}) = \sum_{\ell\geq 0} \mathbb{P}(W_n^{(d)} > \ell) \leq \sum_{\ell\geq 0} \mathbb{P}(n-1-T_n^{(d)} > \ell) = \mathbb{E}(n-1-T_n^{(d)}).$$

where $T_n^{(d)}$ is the number of reticulations of a random d-combining tree-child network and the inequality follows from the fact that each twig is a free tree node and the number of free tree nodes is given by $n-1-k$ where k is the number of reticulation nodes; see Lemma 37. The result follows now from:

$$\mathbb{E}(n-1-T_n^{(d)}) \longrightarrow \begin{cases} 1/2, & \text{if } d = 2; \\ 0, & \text{if } d \geq 3; \end{cases}$$

see Remark 30 and Remark 31. ■

This result also shows that the expected number of cherries, i.e., tree nodes with both children leaves, of a random d-combining tree-child network is bounded, too (since cherries are clearly twigs). Note that the number of cherries is a popular parameter in phylogenetics and has been extensively studied for *phylogenetic trees* (which are bi-combining networks without reticulation nodes).

Finally, Theorem 60 also implies an improvement of the first equality in (2.38).

Corollary 2.9.2. The following asymptotic equivalences hold for d-combining tree-child networks.

(i) For $d = 2$ (bi-combining case), we have

$$\mathrm{TC}_n^{(2)} \sim \sqrt{e} \cdot \mathrm{TC}_{n,n-1}^{(2)}.$$

(ii) For $d \geq 3$, we have

$$\mathrm{TC}_n^{(d)} \sim \mathrm{TC}_{n,n-1}^{(d)}.$$

Remark 35. Note that even with the above result, it is still not possible to give the first-order asymptotics of $\mathrm{TC}_n^{(d)}$ since the approach of [37] (which we are going to use in order to prove Theorem 51) gives only a Theta-result.

2.9.2 Proofs of Propositions 16 and 17 (d-combining tree-child network)

The proofs follow nearly verbatim the steps of [37, Lemmas 4.2 and 4.4] as well as [14, Lemmas 7 and 9]. For the convenience of the reader, we repeat the main steps for Proposition 16 and point out the main differences. Full details of all computations are given in our accompanying Maple worksheet [41].

We start by defining the following sequence

$$P_{n,m} := -Z_{n,m} s_n + \mu_{n,m}^{(d)} Z_{n-1,m+1} + \nu_{n,m}^{(d)} Z_{n-1,m-1},$$

where we use the approach

$$s_n := \sigma_0 + \frac{\sigma_1}{n^{1/3}} + \frac{\sigma_2}{n^{2/3}} + \frac{\sigma_3}{n} + \frac{\sigma_4}{n^{7/6}}$$
$$Z_{n,m} := \left(1 + \frac{\tau_2 m^2 + \tau_1 m}{n}\right) \mathrm{Ai}\left(a_1 + \frac{B^{1/3}(m+1)}{n^{1/3}}\right),$$

with parameters $\sigma_i, \tau_j \in \mathbb{R}$, and B from (2.45). Note that in [14, 37] we had $B = 2$, whereas here B depends on d. Then the claimed inequality is equivalent to $P_{n,m} \geq 0$ with the parameters chosen accordingly.

To prove it, we expand the Airy function $\mathrm{Ai}(z)$ in a neighborhood of

$$\alpha = a_1 + \frac{B^{1/3} m}{n^{1/3}}, \tag{2.63}$$

and, due to the defining differential equation $\mathrm{Ai}''(x) = x\mathrm{Ai}(x)$, we get the following expansion

$$P_{n,m} = p_{n,m}\mathrm{Ai}(\alpha) + p'_{n,m}\mathrm{Ai}'(\alpha),$$

where $p_{n,m}$ and $p'_{n,m}$ are functions of m and n^{-1} and may be expanded as power series in $n^{-1/6}$ whose coefficients are polynomials in m. As the Airy function is entire, these series converge absolutely for $n > 1$ and $m < n$.

Now we proceed with the technical analysis. First, we show that $[m^i n^j]P_{n,m} = 0$ for $i + j > 1$, $i, j \in \mathbb{Q}$. We omit this step, as it is analogous to the previous cases. Second, we use computer algebra to strengthen this result by choosing suitable values σ_i and τ_j to eliminate more terms; see Figure 2.16. A solid diamond at (i, j) indicates that the coefficient $[m^i n^j]P_{n,m}$ is non-zero for generic values of σ_i and τ_j; an empty diamond indicates that the specific choice of σ_i and τ_j makes it vanish. The convex hull is formed by the following three lines

$$L_1 : j = -\frac{7}{6} - \frac{7i}{18}, \qquad L_2 : j = -\frac{1}{3} - \frac{2i}{3}, \qquad L_3 : j = 1 - i.$$

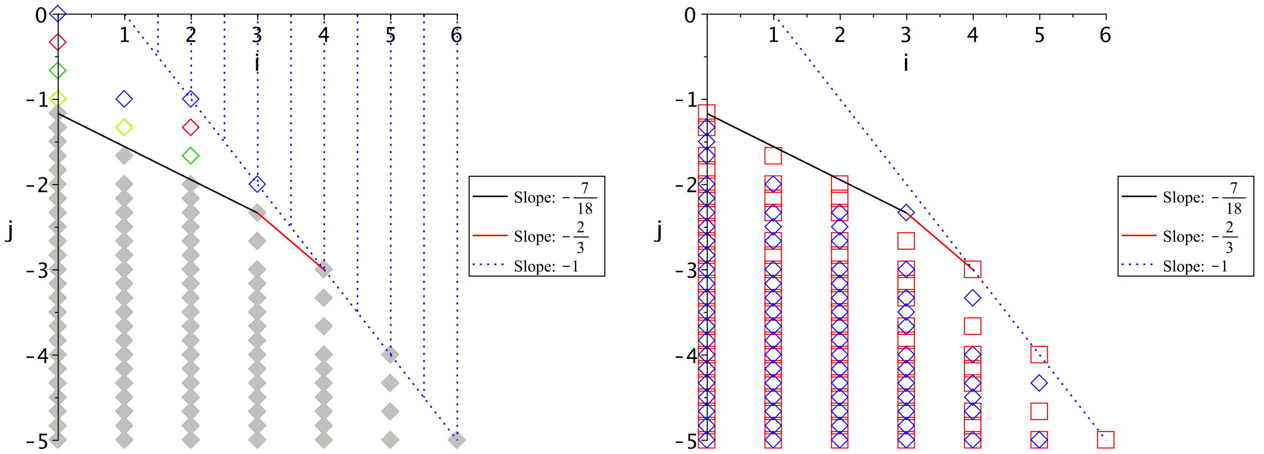


Figure 2.16: (Left) Non-zero coefficients of $P_{n,m} = \sum a_{i,j} m^i n^j$ shown by diamonds for $s_n := \sigma_0 + \frac{\sigma_1}{n^{1/3}} + \frac{\sigma_2}{n^{2/3}} + \frac{\sigma_3}{n} + \frac{\sigma_4}{n^{7/6}}$ and $Z_{n,m} := \left(1 + \frac{\tau_2 m^2 + \tau_1 m}{n}\right) \mathrm{Ai}\left(a_1 + \frac{2^{1/3}(m+1)}{n^{1/3}}\right)$. There are no terms in the blue dashed area. The blue terms vanish for $\sigma_0 = 2$, the red terms vanish for $\sigma_1 = 0$, the green terms vanish for $\sigma_2 = B^{2/3}a_1$, and the yellow terms vanish for $\sigma_3 = -\frac{3d^2-5d+4}{3(d+1)}$ and $\tau_2 = -\frac{2d-1}{3(d+1)}$. The black, red, and blue lines represent the parts L_1, L_2, and L_3, respectively, of the convex hull. (Right) The solid gray diamonds are decomposed into the coefficients $p_{n,m}$ of $\mathrm{Ai}(\alpha)$ (red boxes) and $p'_{n,m}$ of $\mathrm{Ai}'(\alpha)$ (blue diamonds).

In a final step, we distinguish between $p_{n,m}$ and $p'_{n,m}$; see Figure 2.17. The expansions for n tending to infinity start as follows, where the elements on the convex hull are written in color:

$$\begin{aligned} P_{n,m} = \mathrm{Ai}(\alpha) \Bigg(& -\frac{\sigma_4}{n^{7/6}} - \frac{B^{5/3}a_1 m}{3n^{5/3}} - \frac{(23d^2 - 14d + 5)m^2}{9(d+1)^2 n^2} - \frac{2(2d-1)(3d-1)B^{5/3}a_1 m^3}{9(d+1)^2 n^{8/3}} \\ & - \frac{(2d-1)(23d-9)Bm^4}{18(d+1)^2 n^3} + \frac{(2d-1)(209d^2 - 258d + 129)Bm^5}{270(d+1)^3 n^4} + \dots \Bigg) + \\ \mathrm{Ai}'(\alpha) \Bigg(& -\frac{B^{1/3}\sigma_4}{n^{3/2}} \frac{4(d-2)Ba_1 m}{9(d+1)n^2} - \frac{2(9d^3 + 50d^2 - 67d + 21)B^{1/3}m^2}{9(d+1)^2 n^{7/3}} - \frac{4(2d-1)^2 B^{1/3} m^3}{9(d+1)^2 n^{7/3}} \end{aligned}$$

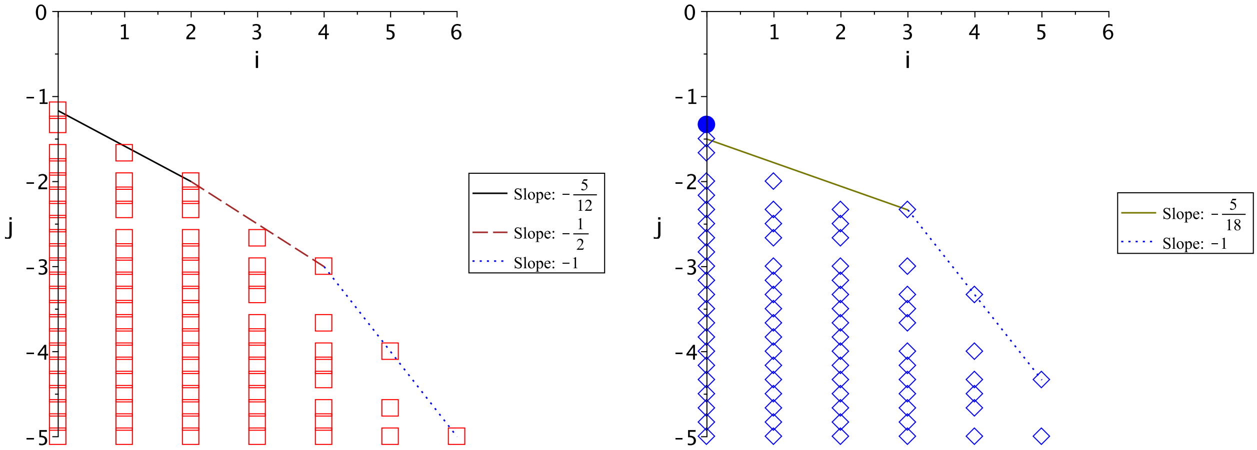


Figure 2.17: Non-zero coefficients $p_{n,m} = \sum \tilde{a}_{i,j} m^i n^j$ (red) and $p'_{n,m} = \sum \tilde{a}'_{i,j} m^i n^j$ (blue) of the expansion (2.63) for $P_{n,m}$. The coefficient of $n^{-4/3}$ in the right picture depicted as a solid blue circle disappears for $\tau_1 = \frac{3d^2+12d-11}{6(d+1)}$.

$$- \frac{(2d-1)(5d-1)B^{4/3}m^4}{18(d+1)^2 n^{10/3}} - \frac{(2d-1)(119d^2+32d-51)B^{4/3}m^5}{270(d+1)^3 n^{13/3}} + \dots \Big) .$$

We now choose $\sigma_4 = -1$ which leads to a positive term $\text{Ai}(\alpha) n^{-7/6}$. Next, for fixed (large) n we prove that for all m the dominant contributions in $P_{n,m}$ are positive. Motivated by Figures 2.16 and 2.17, we consider three different regimes:

$$\text{(i)} \quad m \leq Cn^{1/3}; \qquad \text{(ii)} \quad Cn^{1/3} < m \leq n^{7/18}; \qquad \text{(iii)} \quad n^{7/18} < m < n^{2/3-\epsilon}$$

for a suitable constant $C > 0$. This part is analogous to the one of [37, Lemma 4.2], which is why we omit the technical details. In the end we get that there exists an $N > 0$ such that all terms are positive for $n > N$ and all $m < n^{2/3}$, which ends the proof of Proposition 16.

The proof of Proposition 17 follows analogously.

Chapter 3

Conclusion and Outlook

In the preceding chapters, we have obtained asymptotic and exact results for the main classes of phylogenetic networks. Nevertheless, each section still leaves a number of major classes unaddressed. For instance, the exact enumeration of one-component d-combining networks has been established, but the multi-component and fully general combining structures remain open. Likewise, although the component graph method (Section 1.1.3) yields asymptotic results for networks with fixed or maximal reticulation number, the corresponding exact formulas have yet to be determined.

Beyond these specific gaps, each topic suggests natural directions for further development. One possibility is to extend our techniques to additional network families, such as level-k networks, orchard networks, or time-consistent models, whose combinatorial behavior is not yet well understood. Another promising direction is to broaden the probabilistic perspective initiated in the Section 2.7. In particular, studying additional random variables (e.g., height, *balanced indices*) or establishing limit theorems may lead to a more refined understanding of the global shape of random networks.

Taken together, these open classes and potential extensions highlight a much larger landscape that remains unexplored. The methods mentioned in this thesis, like component graph method, singularity analysis, singular inversion, and encoding-based recurrence derivations, provide a foundation for future progress, and we expect that many of the unresolved classes can be approached using these tools, possibly with suitable adaptations.

The main future objective is to complete the analyses initiated in Chapter 2 for all main network classes. In what follows, we summarize, for each topic, which main classes have not yet been addressed. We also outline possible directions for further development, including the investigation of additional network classes and various combinatorial or probabilistic extensions.

- Section 2.1 (The maximal number of reticulation): This fundamental question has already been resolved, regardless of whether one considers the main classes or not. A natural direction for further investigation is to develop more intuitive and concise proofs for some of the results (for others, this has already been achived here and elsewhere).

- Section 2.2 (Closed-form expressions for small k and arbitrary n): This has been done for all main classes. What remains is to find more effective ways of computing closed-form expressions for larger values of k.

- Section 2.3 (Exact counting for any k and any n): Galled, reticulation-visible, and galled tree-child networks remain.

  The term *exact counting* results is, in fact, somewhat imprecise. Through the component graph method (CGM), it is theoretically possible to compute exact counts for any k and any n. However, the computational complexity rapidly becomes enormous, well beyond polynomial time, making such calculations infeasible in practice. Thus, within this context, an exact counting result refers to any method that improves the computational complexity and yields a tractable procedure. Ideally, such a result takes the form of a closed formula, as in the case of one-component d-combining tree-child networks; see Theorem 35. Alternatively, a result may provide an explicit arithmetic recurrence, such as for general d-combining tree-child networks, which still represents a substantial improvement over the raw CGM computation; see Theorem 36.

- Section 2.4 (Counting maximally reticulated networks): At present, all established results except that for galled networks are asymptotic rather than exact. The next stage is to obtain an exact enumeration of maximally reticulated networks.

- Section 2.5 (Asymptotic counting for fixed k as $n \to \infty$): All main classes are done.

  Further classes have also been considered. For instance, recently in 2024, Agranat-Tamir, Fuchs, Gittenberger and Rosenberg gave an asymptotic result for the number of unlabeled normal galled-tree networks, say $\mathcal{UNGT}$.

Theorem 72 ([1])**.** For the number of unlabeled normal galled-tree networks with n leaves and k reticulation nodes, we have for fixed k, as $n \to \infty$,

$$\mathrm{UNGT}_{n,k} \sim \frac{2^{2k-1}}{(2k)!\,\gamma^{4k-1}\sqrt{\pi}}\, n^{2k-3/2}\,\rho^{-n},$$

where ρ is the radius of convergence of the generating function of rooted binary unlabeled trees $U(z)$ which is defined by

$$U(z) = z + \frac{1}{2}U(z)^2 + \frac{1}{2}U(z^2),$$

and γ is a constant satisfying

$$U(z) \sim 1 - \gamma\sqrt{1 - \frac{z}{\rho}} \quad \text{as } z \to \rho^-.$$

Note that $\rho \approx 0.4027$ and $\gamma \approx 1.1300$.

In 2025, Agranat-Tamir et al. [2] gave the results for (leaf-labeled) normal galled-tree networks, say $\mathcal{NGT}$, which is equivalent to time-consistent galled trees.

Theorem 73. For the number of normal galled-tree networks with n leaves and k reticulation nodes, we have for fixed k, as $n \to \infty$,

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

Remark 36. The result differs from the asymptotic for other classe of phylogenetic networks; the difference arises in the constant factors, including those depending on k.

Also, Agranat-Tamir et al. [3] showed that galled-tree networks exhibit the same asymptotic behavior as in the two theorems above.

Theorem 74. Denote by GT (UGT) the number of (unlabeled) galled-tree networks with n leaves and k reticulation nodes, we have for fixed k, as $n \to \infty$,

$$\mathrm{UGT}_{n,k} \sim \frac{2^{2k-1}}{(2k)!\,\gamma^{4k-1}\sqrt{\pi}}\, n^{2k-3/2}\rho^{-n},$$

and

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

The remaining classes in Figure 1.5 are level-k, spread-k, and k-reticulated networks.

- Section 2.6 (Asymptotic counting of the total number of networks): The remaining main class is reticulation-visible networks. The next stage is to obtain an exact enumeration of the total number of each main class of networks.

- Section 2.7. (The number of reticulations for a random network): Only reticulation-visible networks remains.

- Section 2.8. (Sackin Index): Only one-component d-combining tree-child networks are done and only a Theta result $\Theta(\cdot)$ for the mean was derived. The next stage is to find its constant factor, higher moments and limit laws, or any result for the remaining main classes. Also, there are more statistic parameters can be explored.

Appendix A

Trip-trees

In the Section 1.1.4, we counted the number of dup-trees. In this section, we count the number of trip-trees; see Definition 1.1.7.

Definition A.0.1.

1. $\mathrm{UT}_{n,i}$ ($\mathrm{RT}_{n,i}$) counts the number of un-rooted (rooted) twin-cherry-free trip-trees labeled by $[n] \uplus [i] \uplus [i]$;
2. $\mathrm{C}^{(j)}_{n,i}$ counts the number of un-rooted trip-trees labeled by $[n] \uplus [i] \uplus [i]$ with exactly j twin-cherries;
3. $n_{n,i}$ counts the number of edges in $\mathrm{UT}_{n,i}$ (or $\mathrm{C}^{(j)}_{n,i}$);
4. $\mathrm{UT}^{+}_{n,i}$ counts the number of un-rooted trip-trees labeled by $[n] \uplus [i] \uplus [i]$, where twin-cherries are allowed;

Here we reuse this name UT which referred to un-rooted phylogenetic trees when counting dup-trees. We have some simple observations.

Proposition 23. For $0 \leq i \leq n$, we have

1. $\mathrm{RT}_{n,i} = \mathrm{UT}_{n+1,i}$ and $\mathrm{UT}_{n,i} = \mathrm{C}^{(0)}_{n,i}$;
2. $\mathrm{C}^{(j)}_{n,i} = \binom{i}{j} \mathrm{UT}_{n+j,i-j}$ for all $j \geq 0$;
3. $n_{n,i} = \max(2(n+2i)-3, 0)$ for $0 \leq i \leq n$;
4. $\mathrm{UT}^{+}_{n,i} = \sum_{j=0}^{i} \mathrm{C}^{(j)}_{n,i} = \sum_{j=0}^{i} \binom{i}{j} \mathrm{UT}_{n+j,i-j}$.

Proof. The first statement is trivial by taking the leaf $n+1$ as the root. The third statement is also trivial by definition and the fourth statement is a refinement. It suffices to show the second statement.

Let $j \leq i$. We choose j labels $\mathcal{J} = \{\ell_1, \ell_2, \dots, \ell_j\}$ from $[i]$ with ℓ_k's are increasing and choose a trip-tree T in $\mathcal{UT}_{n+j,i-j}$. For each k from 1 to j,

1. Shift the leaf label m by $m+1$ if $m \geq \ell_k$. T has no leaf labeled by ℓ_k by now. The current largest label is $n+j+2-k$.
2. Replace the leaf with the current largest label with a twin-cherry labeled by ℓ_1.
3. Replace the leaf with the current largest label with a leaf labeled by ℓ_1. The current largest label is $n+j-k$.

We have a trip-tree in $\mathcal{C}_{n,i}^{(j)}$. ■

1. Find the classes $\mathcal{O}_{m,j}$ for $\mathcal{UT}_{n,i+1}$ We implement the same classification steps of the dup-trees on the trip-trees $\mathcal{UT}_{n,i+1}$. Eventually, there are 32 classes in total with their multiplicity. For each $\mathcal{O}_{m,j}$, the multiplicity m can be obtained by

1. deleting $\underline{\text{two}}$ red dots from $\mathcal{O}_{m,j}$, we have some at most three resulting graphs $P_{m,j}^k$,
2. find the multiplicity of $P_{m,j}^k$ by adding back $\underline{\text{two}}$ red dots and sum them up.

There are 22 classes of them do not contains a symmetrical structure (without b-edges and red dots) consisting of 3 identical parts; see Figure A.1. The remaining 10 classes contain a symmetrical structure (without b-edges and red dots) consisting of 3 identical parts; see Figure A.2. The graphs $P_{m,j}^{(k)}$'s are temporarily used to calculate the multiplicity value, so we hide them in the above two figures, otherwise, they are too large. These classes form a partition of $\mathcal{UT}_{n,i}$ and it gives the formula.

$$\mathrm{UT}_{n,i+1} = \sum_{m,j} \mathrm{O}_{m,j}. \tag{A.1}$$

2. Establish the formula between $\mathcal{UT}_{n,i+1}$ and $\mathcal{UT}_{n,i}$. For a trip-tree in $\mathcal{UT}_{n,i+1}$, we choose two leaves labeled by $i+1$ (there are three choices), delete them and suppress the unary nodes. It may produces at most two twin-cherries, that is, the resulting trip-trees fall into $\mathcal{UT}_{n,i}, \mathcal{C}_{n,i}^{(1)}$ or $\mathcal{C}_{n,i}^{(2)}$. Next, in each case, we adding back two leaves $i+1$ (two red dots) without creating twin-cherries. Recall that $n_{n,i} = \max(2(n+2i)-3, 0)$ is the number of edges of a trip-tree in $\mathcal{UT}_{n,i}, \mathcal{C}_{n,i}^{(1)}$ or $\mathcal{C}_{n,i}^{(2)}$.

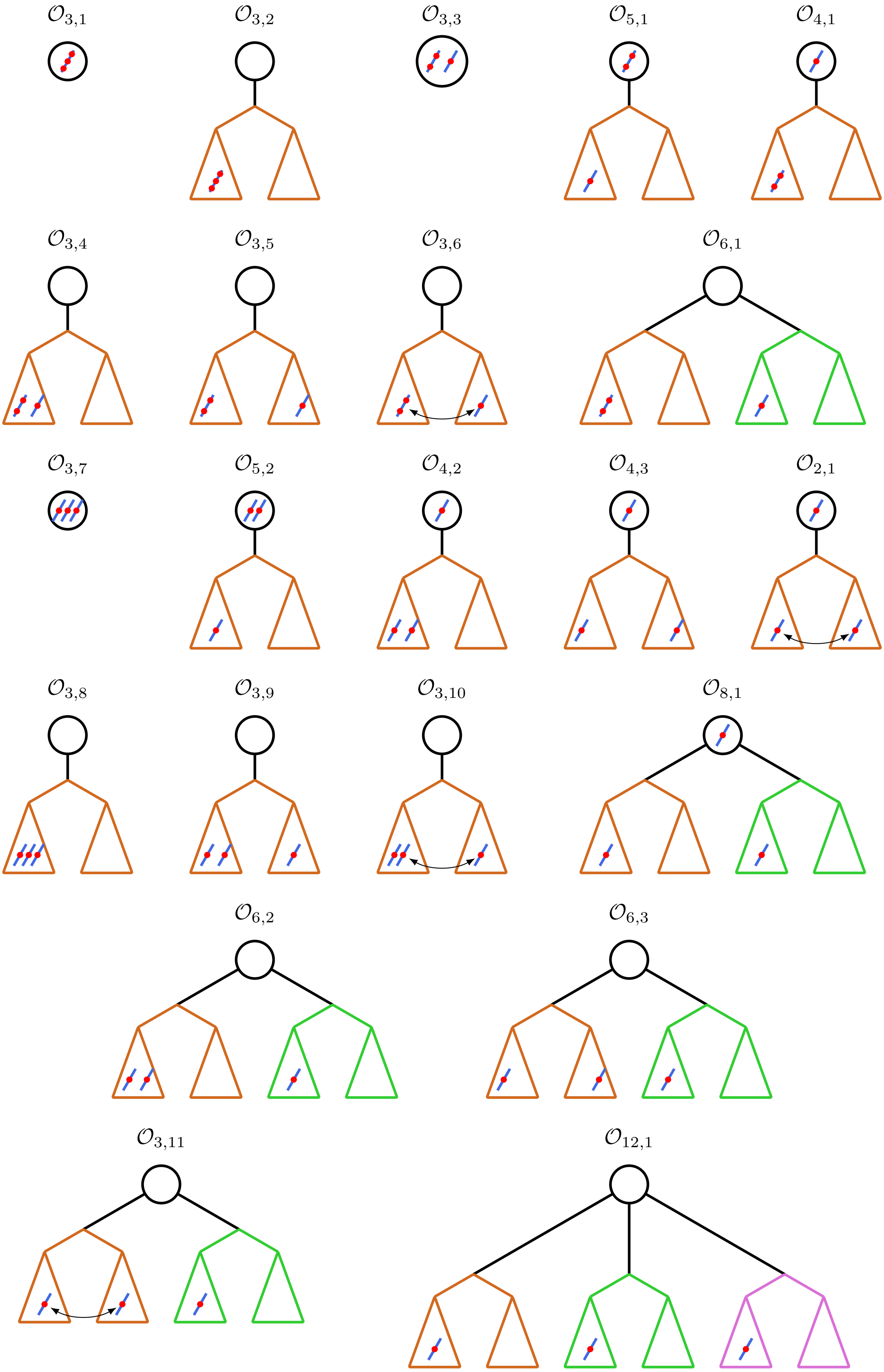


Figure A.1: There are 22 classes of $\mathcal{UT}_{n,i+1}$ having no symmetric structures of 3 parts.

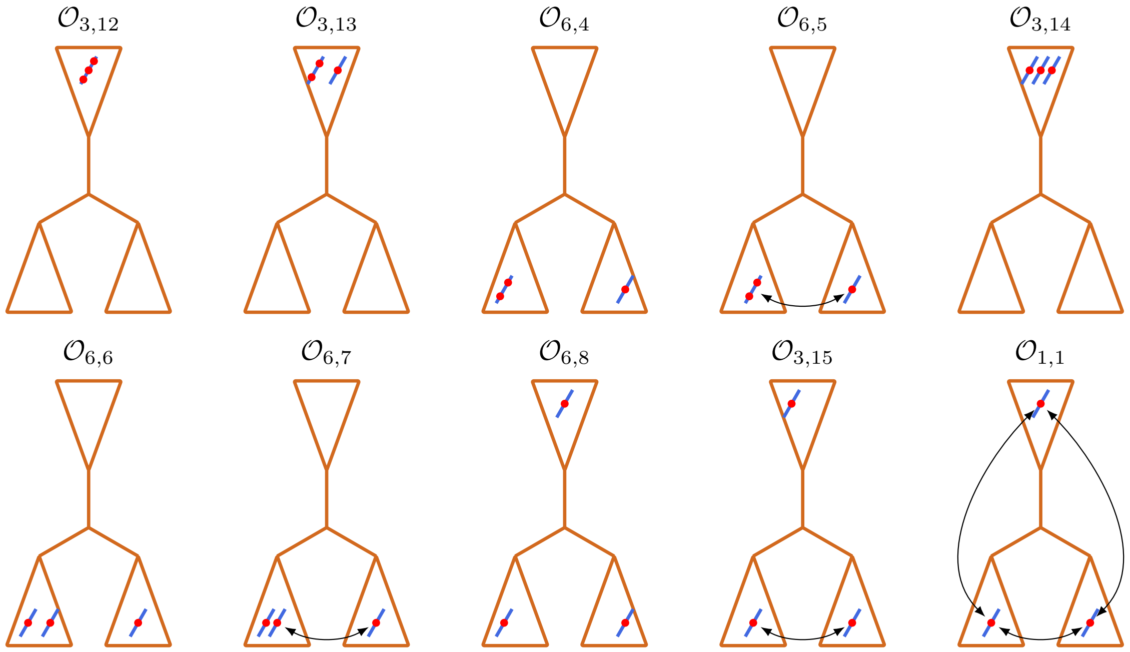


Figure A.2: There are 10 classes of $\mathcal{UT}_{n,i+1}$ having symmetric structures of 3 parts and they exists only when $n = i + 1$.

1. For a trip-tree in $\mathcal{UT}_{n,i}$, we can't choose the leaf-edge of the leaf $i + 1$. We may attach them on the same or different remaining edge(s). There are $\binom{n_{n,i}}{2}$ ways.
$$\binom{n_{n,i}-1}{1} + \binom{n_{n,i}-1}{2} = \binom{n_{n,i}}{2}.$$
2. For a trip-tree in $\mathcal{C}^{(1)}_{n,i}$, at least one new leaf must attach on the twin-cherry. There are $2n_{n,i} - 3$ ways by the inclusion–exclusion principle.
$$\binom{n_{n,i}}{2} - \binom{n_{n,i}-2}{2} = 2n_{n,i} - 3.$$
3. For a trip-tree in $\mathcal{C}^{(2)}_{n,i}$, there are only 4 ways.

To sum up, the formula is provided by

$$\begin{aligned}\binom{n_{n,i}}{2} \mathrm{UT}_{n,i} + (2n_{n,i} - 3)\, \mathrm{C}^{(1)}_{n,i} + 4\, \mathrm{C}^{(2)}_{n,i} &= \sum_{m,j} m\, \mathrm{O}_{m,j} \\ &= 3\, \mathrm{UT}_{n,i+1} + \sum_{m,j} (m-3)\ \mathrm{O}_{m,j}.\end{aligned} \tag{A.2}$$

3. Compute the remaining terms. The purpose here is to compute the summation

$$\sum_{m,j} (m-3)\ \mathrm{O}_{m,j} = -2\, \mathrm{O}_{1,1} - \mathrm{O}_{2,1} + \sum_{j=1}^{3} \mathrm{O}_{4,j} + 2 \sum_{j=1}^{2} \mathrm{O}_{5,j} + 3 \sum_{j=1}^{8} \mathrm{O}_{6,j} + 5\, \mathrm{O}_{8,1} + 9\, \mathrm{O}_{12,1}, \tag{A.3}$$

which contains many minor terms. An intuitive approach is to calculate the number for each $\mathrm{O}_{m,j}$ independently. However, this would make the calculations between many terms overly complicated. We define slightly larger classes, called type, focusing on a fixed structure that can be directly calculated from $\mathrm{UT}_{n,i}$, $\mathrm{C}^{(1)}_{n,i}$ or $\mathrm{C}^{(2)}_{n,i}$. It corresponds to some specific $\mathcal{O}_{m,j}$'s for each type.

Definition A.0.2 (type a)**.** Define the set $\mathcal{O}_a^{(j)}$ collects the trip-trees labeled by $[n] \uplus [i+1] \uplus [i]$ such that

1. it contains exactly j twin-cherries;
2. the two b-edges with one red dot (the leaf labeled by $i+1$) locate at the same positions of different triangles of a symmetry structures S_a;
3. the third red dot "will" be add on somewhere outside S_a,

and $\mathcal{O}_a = \cup_{j=0}^{1} \mathcal{O}_a^{(j)}$.

A definition of a type defines not only the sets $\mathcal{O}_a^{(j)}$'s, but also defines a combinatorial way to "completing" $\mathcal{O}_a$ to twin-cherry-free trip-trees labeled by $[n] \uplus [i+1] \uplus [i+1]$. From this perspective, a type must define both a fixed structure and a combination step.

By two-way counting between $\mathcal{O}_a$ and resulting trip-trees, we have the following proposition.

Proposition 24 (type a)**.**

$$\sum_{1\le d\le i} \binom{i}{d} (2d-1)!! \left(n_{n,i-d}\, \mathrm{UT}_{n,i-d} + 2\, \mathrm{C}^{(1)}_{n,i-d} \right) = \mathrm{O}_{2,1} + 2\, \mathrm{O}_{3,11} + \mathrm{O}_{3,15} + \mathrm{O}_{1,1}. \quad \text{(A.4)}$$

Proof. Count the number of edges of the bipartite $G = (\mathcal{A}, \mathcal{B}; \mathcal{E})$, where

1. $\mathcal{A} = \mathcal{O}_a^{(0)} \cup \mathcal{O}_a^{(1)}$,
2. $\mathcal{B} = \mathcal{O}_{2,1} \cup \mathcal{O}_{3,11} \cup \mathcal{O}_{3,15} \cup \mathcal{O}_{1,1}$,
3. $\mathcal{E}$ collect (A, e, B) such that $A \in \mathcal{A}$, $B \in \mathcal{B}$ and B is obtained by adding a red dot on the edge e of A.

First, we count the number of edges that adjacent to $\mathcal{O}_a^{(0)}$ by construction.

1. Choose a phylogenetic tree T on $\mathcal{D}$. (There are $(2d-3)!!$ ways.)
2. Add a leaf $i+1$ on an edge of T. (There are $2d-1$ ways.)
3. Duplicate T and joint their roots, say the joint node u. (the structure is what type a fixes.)

4. Choose a trip-tree T' in $\mathcal{UT}_{n,i-d}$. Joint the leaf $i+1$ to the node u (there's no leaf $i+1$ by now).

5. Add a leaf $i+1$ on an edge of T'. (There are $n_{n,i-d}$ choices.)

6. Relabel the indices of T and the repeated indices of T'. (There are $\binom{i}{d}$ ways.)

Eventually, it gives

$$\sum_{1\le d\le i} \binom{i}{d} (2d-1)!!\, n_{n,i-d}\, \mathrm{UT}_{n,i-d} = \mathrm{O}_{2,1} + 2 \times (\text{part of } \mathrm{O}_{3,11}) + \mathrm{O}_{3,15} + \mathrm{O}_{1,1},$$

where "$2 \times$ (part of $\mathrm{O}_{3,11}$)" means that some resulting graphs fall into $\mathcal{O}_{3,11}$ if the third red dot is added on a symmetry structure. Paired of resulting trip-trees are identical due to the symmetry and it explains the multiplicative constant of $\mathrm{O}_{3,11}$ is 2. The text "part of" shows that the third red dot locates at in a symmetry structure but size is greater than 1 of a trip-tree in $\mathcal{O}_{3,11}$. The next paragraph includes the ones with size 1 (i.e. twin-cherry) for $\mathcal{O}_{3,11}$.

Second, we count the number of edges that adjacent to $\mathcal{O}_a^{(1)}$ by construction. The first 4 steps are the same as the above steps.

5' Choose a trip-tree T' in $\mathcal{C}^{(1)}_{n,i-d}$ and do the same relabel process.

6' Insert a leaf $i+1$ into T' by reducing the twin-cherry. (There are 2 choices.)

$$\sum_{1\le d\le i} \binom{i}{d} (2d-1)!!\, 2\, \mathrm{C}^{(1)}_{n,i-d} = (\text{remaining part of } 2 \times \mathrm{O}_{3,11})\,.$$

Sum up the above two equations gives the LHS of the result. ■

Definition A.0.3 (type b, c, d, e, f and g)**.** For $x \in \{b, c, d, e\}$ and $j \ge 0$, define the set $\mathcal{O}_x^{(j)}$ collects the trip-trees labeled by $[n] \uplus [i+1] \uplus [i]$ such that

1. it contains exactly j twin-cherries;

2. $P(x)$ holds, where $P(x)$ means the two red dots (two leaves $i+1$) belong to …

   ($x = b$) one b-edge locating in one triangle of a symmetry structure S_b (Like $\mathcal{O}_{4,1}$ without the b-edge in the black circle);

   ($x = c$) two b-edges resp. locating at different positions in one triangle of a symmetry structure S_c (Like $\mathcal{O}_{4,2}$ without the b-edge in the black circle);

   ($x = d$) two b-edges resp. locating at different positions in different triangles of a symmetry structure S_d (Like $\mathcal{O}_{4,3}$ without the b-edge in the black circle);

   ($x = e$) two b-edges resp. locating in different triangles of different symmetry structures S_{e_1} and S_{e_2} (Like $\mathcal{O}_{8,1}$ without the b-edge in the black circle),

3. the third red dot "will" be add on somewhere outside S_x; S_{e_1} and S_{e_2} if $x = e$.

and $\mathcal{O}_x = \cup_{j=0}^{1} \mathcal{O}_x^{(j)}$.

For $x \in f, g$, define the set $\mathcal{O}_x^{(j)}$ collects the trip-trees labeled by $[n] \uplus [i] \uplus [i]$ such that

1. it contains exactly j twin-cherries;
2. the only b-edge with one red dot locates in a triangle of a symmetric structure S_x;
3. the second and third red dots "will" be add on somewhere …

   ($x = f$) outside S_x but in the same symmetric structure;

   ($x = g$) outside S_x,

   (may or may not locate at the same position, i.e. a b-edge with two red dots),

and $\mathcal{O}_f = \mathcal{O}_f^{(0)}$ and $\mathcal{O}_g = \cup_{j=0}^{2} \mathcal{O}_g^{(j)}$.

Types f and g are relatively special that it contains only one leaf $i+1$ while each other type contains two leaves $i+1$. In fact, there are other types that can be explored basing on the number and position of b-edges with some red dots, but the types mentioned above are sufficient to reach a conclusion.

Proposition 25 (type b, c, d, e, f and g). We have

$$\sum_{1 \le d \le i} \binom{i}{d} (2d-1)!! \left(n_{n,i-d} \,\mathrm{UT}_{n,i-d} + 2\,\mathrm{C}_{n,i-d}^{(1)} \right) = \mathrm{O}_{4,1} + 2\,\mathrm{O}_{6,1} + \mathrm{O}_{6,4} + \mathrm{O}_{6,5}, \quad \text{(A.5)}$$

$$\sum_{1 \le d \le i} \binom{i}{d} (2d-3)!! \binom{2d-1}{2} \left(n_{n,i-d} \,\mathrm{UT}_{n,i-d} + 2\,\mathrm{C}_{n,i-d}^{(1)} \right) = \mathrm{O}_{4,2} + 2\,\mathrm{O}_{6,2} + \mathrm{O}_{6,6} + \mathrm{O}_{6,7} \quad \text{(A.6)}$$

$$\sum_{1 \le d \le i} \binom{i}{d} (2d-3)!! \binom{2d-1}{2} \left(n_{n,i-d} \,\mathrm{UT}_{n,i-d} + 2\,\mathrm{C}_{n,i-d}^{(1)} \right) = \mathrm{O}_{4,3} + 2\,\mathrm{O}_{6,3} + 3\,\mathrm{O}_{6,8} + \mathrm{O}_{3,15}, \quad \text{(A.7)}$$

$$\sum_{\substack{1 \le d_1 + d_2 \le i, \\ d := d_1 + d_2}} \binom{i}{d_1, d_2, i-d} (2d_1-1)!!\,(2d_2-1)!! \left(n_{n+1,i-d} \,\mathrm{UT}_{n+1,i-d} + 2\,\mathrm{C}_{n+1,i-d}^{(1)} \right) = 2\,\mathrm{O}_{8,1} + 12\,\mathrm{O}_{12,1}, \quad \text{(A.8)}$$

$$\sum_{\substack{1\le d_1+d_2\le i,\\ d:=d_1+d_2}} \binom{i}{d_1, d_2, i-d} (2d_1-1)!!\,(2d_2-3)!!(2d_2-1)^2\,\mathrm{UT}_{n+1,i-d}$$
$$= \mathrm{O}_{6,1}+\mathrm{O}_{6,2}+\mathrm{O}_{6,3}, \quad \text{(A.9)}$$

$$\sum_{1\le d\le i} \binom{i}{d} (2d-1)!! \left[\binom{n_{n,i-d}+1}{2}\mathrm{UT}_{n,i-d}+(2\,n_{n,i-d}-1)\ \mathrm{C}^{(1)}_{n,i-d}+4\,\mathrm{C}^{(2)}_{n,i-d}\right]$$
$$=\mathrm{O}_{5,1}+2\,\mathrm{O}_{6,1}+\mathrm{O}_{5,2}+4\,\mathrm{O}_{8,1}+2\,\mathrm{O}_{6,2}+2\,\mathrm{O}_{6,3}+\mathrm{O}_{3,11}+12\,\mathrm{O}_{12,1}+$$
$$\mathrm{O}_{6,4}+\mathrm{O}_{6,5}+\mathrm{O}_{6,6}+\mathrm{O}_{6,7}. \quad \text{(A.10)}$$

Proof. (Type b) We classify the process into three main parts, with relabeling placed last.

1. Choose a rooted phylogenetic trees T of size d, duplicate T as T'', attach the b-edge with two red dots on T's edges, and joint their roots, say u; (There are $(2d-1)!!$ ways.)
2. Choose a un-rooted trip-tree U in $\mathcal{UT}_{n,i-d}$ (or $\mathcal{C}^{(1)}_{n,i-d}$), joint the leaf $i+1$ with the node u, and attach a new leaf $i+1$ on U's edges by reducing all twin-cherries if they exist. (There are $n_{n,i-d}\mathrm{UT}_{n,i-d}$ or $2\mathrm{C}^{(1)}_{n,i-d}$ ways.)
3. Relabel the duplicated labels in the first and second steps. (There are $\binom{i}{d}$ ways.)

(Type c) It's very similar to type b. The different part is in the step 1 that we attach two b-edges with one red dot resp. on the "different" edges of T; if we choose the same edge, then it becomes the case of one b-edge with two red dots. Hence, there are $\binom{2d-1}{2}$ ways for attaching.

(Type d) There is an over-counting problem here.

1. Choose a rooted phylogenetic trees T_1 of size d_1, duplicate it as T_1', attach the b-edge with one red dot on T_1's edges, and joint their roots, say u_1; (There are $(2d_1-1)!!$ ways.)
2. Again, choose a rooted phylogenetic trees T_2 of size d_2, duplicate it as T_2', attach the b-edge with one red dot on T_2's edges, and joint their roots, say u_2; (There are $(2d_2-1)!!$ ways.)
3. Choose a un-rooted trip-tree U in $\mathcal{UT}_{n+1,i-d}$ (or $\mathcal{C}^{(1)}_{n+1,i-d}$), joint the leaf $n+1$ with the node u_1, joint the leaf n with the node u_2, shift the label j by $j+1$ for $j\ge i+1$ (there's no leaf $i+1$ by now), and attach a new leaf $i+1$ on U's edges by reducing all twin-cherries if they exist. (There are $n_{n+1,i-d}\mathrm{UT}_{n+1,i-d}$ or $2\mathrm{C}^{(1)}_{n+1,i-d}$ ways.)
4. We classify the resulting graphs by the edge that the new leaf is attached on, it is either (1) not involved in a symmetry structure or (2) involved in a symmetry structure, say S_3.

(1) The graphs belongs to $\mathcal{O}_{8,1}$ and each pair of them is identical because the order in which the symmetry structures S_1 and S_2, consisting of T_1 and T_2 respectively, are chosen can be arranged. The multiplicity value is 2.

(2) The graphs belongs to $\mathcal{O}_{12,1}$. The multiplicity value is 12 because the symmetry structures S_1, S_2 and S_3 are arrangeable and there are two edges in S_3 in step 3 that making two resulting graphs identical.

5. Relabel the duplicated labels in the first three steps. (There are $\binom{i}{d_1, d_2, i-(d_1+d_2)}$ ways.)

■

Therefore, we may build the remaining terms (A.3) from the above types with little terms left. Denote by (x) the RHS of the equation in Proposition 24 and Proposition 25 for $x = a, b, \ldots, g$.

$$(A.3) = -(a) + (b) + (c) + (d) - \frac{3}{2}(e) - 3(f) + 2(g) + 3\,\mathrm{O}_{12,1} - \mathrm{O}_{1,1}.$$

Finally we directly compute the remaining two terms.

Proposition 26. We have,

$$\mathrm{O}_{12,1} = \frac{1}{3!} \sum_{\substack{1 \le d_1+d_2+d_3 \le i,\\ d := d_1+d_2+d_3}} \binom{i}{d_1, d_2, d_3, i-d} (2d_1-1)!!\,(2d_2-1)!!(2d_3-1)!!\,\mathrm{UT}_{n+2,i-d}$$

and

$$\mathrm{O}_{1,1} = \begin{cases} (2i-1)!!, & \text{if } n = i+1 \text{ and } i \ge 1; \\ 0, & \text{otherwise.} \end{cases}$$

Note that in $\mathrm{O}_{12,1}$, the leaves $i+1, n+1$ and $n+2$ of a trip-tree in $\mathrm{UT}_{n+2,i-d}$ are used to joint the three symmetry structures. In $\mathrm{O}_{1,1}$, the condition requires $i \ge 1$ for the existence of a rooted phylogenetic tree of size i (and then attach the leaf $i+1$ on it).

Proposition 27. We have, $\mathrm{UT}_{n,i+1} =$

$$\frac{1}{3}\Big\{_{(\alpha)} \binom{n_{n,i}}{2} \mathrm{UT}_{n,i} + (2n_{n,i}-3)\,\mathrm{C}^{(1)}_{n,i} + 4\,\mathrm{C}^{(2)}_{n,i}$$

$$-\Big\{_{(\beta)}\, 2 \sum_{d=1}^{i} \binom{i}{d} (2d-1)!!\,(d-1)\,\Big[n_{n,i-d}\,\mathrm{UT}_{n,i-d} + 2\,\mathrm{C}^{(1)}_{n,i-d}\Big]$$

$$+\,2 \sum_{d=1}^{i} \binom{i}{d} (2d-1)!!\,\left[\binom{n_{n,i-d}+1}{2} \mathrm{UT}_{n,i-d} + (2\,n_{n,i-d}-1)\,\mathrm{C}^{(1)}_{n,i-d} + 4\,\mathrm{C}^{(2)}_{n,i-d}\right]$$

$$
\begin{aligned}
&- \frac{3}{2} \sum_{\substack{1\le d_1+d_2\le i\\ d:=d_1+d_2}} \binom{i}{d_1, d_2} (2d_1-1)!!\,(2d_2-1)!!\,\Big[n_{n+1,i-d}\,\mathrm{UT}_{n+1,i-d} + 2\,\mathrm{C}^{(1)}_{n+1,i-d}\Big] \\
&- 3 \sum_{\substack{1\le d_1+d_2\le i\\ d:=d_1+d_2}} \binom{i}{d_1, d_2} (2d_1-1)!!\,(2d_2-1)!!\,(2d_2-1)\,\mathrm{UT}_{n+1,i-d} \\
&+ 3 \sum_{\substack{1\le d_1+d_2+d_3\le i\\ d:=d_1+d_2+d_3}} \binom{i}{d_1, d_2, d_3} \frac{(2d_1-1)!!\,(2d_2-1)!!\,(2d_3-1)!!}{3!}\,\mathrm{UT}_{n+2,i-d} \\
&- \mathbf{1}_{\substack{n=i+1\\ \text{and} i\ge 1}}\,(2i-1)!! \Big\}_{(\beta)} \Bigg\}_{(\alpha)},
\end{aligned}
$$

for $0 \le i \le n-1$ with initials $\mathrm{UT}_{n,0} = (2n-5)!!$ for $n \ge 3$, $\mathrm{UT}_{2,0} = \mathrm{UT}_{1,0} = 1$ and $\mathrm{UT}_{n,i} = 0$ for undefined range.

Here are the tables for $\mathrm{UT}_{n,i}$ and $\mathrm{UT}^{+}_{n,i}$.

$n\backslash i$	0	1	2	3	4	5	6
1	1	0					
2	1	0	2				
3	1	1	14	318			
4	3	10	165	5115	267570		
5	15	105	2310	92715	5940270	554642760	
6	105	1260	36855	1859235	143313135	15629853225	2287933545240

Table A.1: $\mathrm{UT}_{n,i}$ for $1 \le n \le 6$ and $0 \le i \le n$.

$n\backslash i$	0	1	2	3	4	5	6
0	0						
1	1	1					
2	1	1	7				
3	1	4	49	1233			
4	3	25	480	16770	939255		
5	15	210	5775	265650	18543525	1835552985	
6	105	2205	81900	4780230	405762630	47374167750	7294858389315

Table A.2: $\mathrm{UT}^{+}_{n,i}$ for $1 \le n \le 6$ and $0 \le i \le n$.

Bibliography

[1] Lily Agranat-Tamir, Michael Fuchs, Bernhard Gittenberger, and Noah A Rosenberg. Asymptotic enumeration of rooted binary unlabeled galled trees with a fixed number of galls. In 35th International Conference on Probabilistic, Combinatorial and Asymptotic Methods for the Analysis of Algorithms (AofA 2024), pages 27–1. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2024.

[2] Lily Agranat-Tamir, Michael Fuchs, Bernhard Gittenberger, and Noah A Rosenberg. Enumerative combinatorics of unlabeled and labeled time-consistent galled trees. arXiv preprint arXiv:2504.16302, 2025.

[3] Lily Agranat-Tamir, Karthik V. Seetharaman, Michael Fuchs, Bernhard Gittenberger, and Noah A. Rosenberg. Combinatorial comparison of general galled trees, time-consistent galled trees, and simplex time-consistent galled trees. Submitted, 2025.

[4] Cyril Banderier, Philippe Marchal, and Michael Wallner. Rectangular young tableaux with local decreases and the density method for uniform random generation (short version). In L. Ferrari and M. Vamvakari, editors, Proceedings of the GASCom 2018 Workshop, CEUR Workshop Proceedings, Vol. 2113, volume 2113 of CEUR Workshop Proceedings, pages 60–68, 2018. Short version; also available on arXiv:1805.09017. URL: `https://ceur-ws.org/Vol-2113/paper5.pdf`.

[5] Cyril Banderier and Michael Wallner. Young tableaux with periodic walls: counting with the density method. Sém. Lothar. Combin. B, 85, 2021.

[6] Edward A Bender and L Bruce Richmond. An asymptotic expansion for the coefficients of some power series ii: Lagrange inversion. Discrete Mathematics, 50:135–141, 1984.

[7] Magnus Bordewich and Charles Semple. Reticulation-visible networks. Advances in Applied Mathematics, 78:114–141, 2016.

[8] Gabriel Cardona, Francesc Rosselló, and Gabriel Valiente. Comparison of tree-child phylogenetic networks. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 6(4):552–569, 2008.

[9] Gabriel Cardona and Louxin Zhang. Counting and enumerating tree-child networks and their subclasses. Journal of Computer and System Sciences, 114:84–104, 2020.

[10] Y.-S. Chang, M. Fuchs, H. Liu, M. Wallner, and G.-R. Yu. Enumeration of d-combining tree-child networks. In Proceedings of the 33rd International Meeting on Probabilistic, Combinatorial and Asymptotic Methods for the Analysis of Algorithms (LIPICS), volume 225 of LIPIcs. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2022. Paper No. 5.

[11] Yu-Sheng Chang and Michael Fuchs. Counting phylogenetic networks with few reticulation vertices: Galled and reticulation-visible networks. Bull. Math. Biol., 2024. Originally appeared as a preprint on April 25 , 2024 [4]; to appear in Bull. Math. Biol.

[12] Yu-Sheng Chang, Michael Fuchs, Hexuan Liu, Michael Wallner, and Guan-Ru Yu. Enumerative and distributional results for d-combining tree-child networks. Adv. in Appl. Math., 157:Paper No. 102704, 2024. Originally appeared as a preprint on March 25, 2024 [2]. doi:10.1016/j.aam.2024.102704.

[13] Yu-Sheng Chang, Michael Fuchs, and Guan-Ru Yu. Galled tree-child networks. In Cécile Mailler and Sebastian Wild, editors, 35th International Conference on Probabilistic, Combinatorial and Asymptotic Methods for the Analysis of Algorithms (AofA 2024), volume 2 of Leibniz International Proceedings in Informatics (LIPIcs), pages 2:1–2:13. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany, 2024. doi:10.4230/LIPIcs.AofA.2024.2.

[14] Andrew Elvey Price, Wenjie Fang, and Michael Wallner. Asymptotics of minimal deterministic finite automata recognizing a finite binary language. In 31st International Conference on Probabilistic, Combinatorial and Asymptotic Methods for the Analysis of Algorithms (AofA 2020), pages 11–1. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2020.

[15] Mareike Fischer, Lina Herbst, Sophie Kersting, Annemarie Luise Kühn, and Kristina Wicke. Tree Balance Indices: A Comprehensive Survey. Springer Cham, 2023. doi:10.1007/978-3-031-39800-1.

[16] Mareike Fischer, Lina Herbst, Sophie Kersting, Luise Kühn, and Kristina Wicke. Tree balance indices: a comprehensive survey. arXiv preprint arXiv:2109.12281, 2021.

[17] Philippe Flajolet and Robert Sedgewick. Analytic combinatorics. cambridge University press, 2009.

[18] Michael Fuchs and Bernhard Gittenberger. Sackin indices for labeled and unlabeled classes of galled trees. Journal of Mathematical Biology, 90(4):1–34, 2025.

[19] Michael Fuchs, Bernhard Gittenberger, and Marefatollah Mansouri. Counting phylogenetic networks with few reticulation vertices: tree-child and normal networks. arXiv preprint arXiv:1803.11325, 2018.

[20] Michael Fuchs, Bernhard Gittenberger, and Marefatollah Mansouri. Counting phylogenetic networks with few reticulation vertices: exact enumeration and corrections. arXiv preprint arXiv:2006.15784, 2020.

[21] Michael Fuchs, En-Yu Huang, and Guan-Ru Yu. Counting phylogenetic networks with few reticulation vertices: a second approach. Discrete Applied Mathematics, 320:140–149, 2022.

[22] Michael Fuchs, Hexuan Liu, and Guan-Ru Yu. A short note on the exact counting of tree-child networks. arXiv preprint arXiv:2110.03842, 2021.

[23] Michael Fuchs, Guan-Ru Yu, and Louxin Zhang. On the asymptotic growth of the number of tree-child networks. European Journal of Combinatorics, 93:103278, 2021.

[24] Michael Fuchs, Guan-Ru Yu, and Louxin Zhang. Asymptotic enumeration and distributional properties of galled networks. Journal of Combinatorial Theory, Series A, 189:105599, 2022.

[25] Philippe Gambette, Andreas DM Gunawan, Anthony Labarre, Stéphane Vialette, and Louxin Zhang. Locating a tree in a phylogenetic network in quadratic time. In Research in Computational Molecular Biology: 19th Annual International Conference, RECOMB 2015, Warsaw, Poland, April 12–15, 2015, Proceedings, volume 9029 of Lecture Notes in Computer Science, pages 96–107. Springer, 2015. `doi:10.1007/978-3-319-16706-0_9`.

[26] Philippe Gambette, Andreas DM Gunawan, Anthony Labarre, Stéphane Vialette, and Louxin Zhang. Solving the tree containment problem in linear time for nearly stable phylogenetic networks. Discrete Applied Mathematics, 246:62–79, 2018.

[27] Ronald L. Graham, Donald E. Knuth, and Oren Patashnik. Concrete Mathematics: A Foundation for Computer Science. Addison-Wesley, 2nd edition, 1994.

[28] Andreas DM Gunawan, Bhaskar DasGupta, and Louxin Zhang. A decomposition theorem and two algorithms for reticulation-visible networks. Information and Computation, 252:161–175, 2017.

[29] Andreas DM Gunawan, Jeyaram Rathin, and Louxin Zhang. Counting and enumerating galled networks. Discrete Applied Mathematics, 283:644–654, 2020.

[30] Andreas DM Gunawan, Hongwei Yan, and Louxin Zhang. Compression of phylogenetic networks and algorithm for the tree containment problem. Journal of Computational Biology, 26(3):285–294, 2019.

[31] Andreas DM Gunawan and Louxin Zhang. Bounding the size of a network defined by visibility property. arXiv preprint arXiv:1510.00115, 2015.

[32] Daniel H Huson and Tobias H Klöpper. Beyond galled trees-decomposition and computation of galled networks. In Annual international conference on research in computational molecular biology, pages 211–225. Springer, 2007.

[33] Sungsik Kong, Joan Carles Pons, Laura Kubatko, and Kristina Wicke. Classes of explicit phylogenetic networks and their biological and mathematical significance. Journal of Mathematical Biology, 84(6):47, 2022.

[34] Marefatollah Mansouri. Counting general phylogenetic networks. arXiv preprint arXiv:2005.14547, 2020.

[35] Colin McDiarmid, Charles Semple, and Dominic Welsh. Counting phylogenetic networks. Annals of Combinatorics, 19(1):205–224, 2015.

[36] Miquel Pons and Josep Batle. Combinatorial characterization of a certain class of words and a conjectured connection with general subclasses of phylogenetic tree-child networks. Scientific reports, 11(1):21875, 2021.

[37] Andrew Elvey Price, Wenjie Fang, and Michael Wallner. Compacted binary trees admit a stretched exponential. Journal of Combinatorial Theory, Series A, 177:105306, 2021.

[38] Charles Semple. Size of a phylogenetic network. Discrete Applied Mathematics, 217:362–367, 2017.

[39] Mike Steel. Phylogeny: discrete and random processes in evolution. SIAM, 2016.

[40] Erlang Surya and Lutz Warnke. Lagrange inversion formula by induction. The American Mathematical Monthly, 130:1–5, 09 2023. `doi:10.1080/00029890.2023.2251344`.

[41] Michael Wallner. Ancillary files of arxiv:2203.07619. `http://dmg.tuwien.ac.at/mwallner`, 2022. DOI: 10.48550/arXiv.2203.07619.

[42] Stephen J Willson. Properties of normal phylogenetic networks. Bulletin of mathematical biology, 72(2):340–358, 2010.

[43] Louxin Zhang. The sackin index of simplex networks. In RECOMB International Workshop on Comparative Genomics, pages 52–67. Springer, 2022.