AUSTRALASIAN JOURNAL OF COMBINATORICS
Volume **83(1)** (2022), Pages 40–86

Counting general phylogenetic networks

Marefatollah Mansouri*

Department of Discrete Mathematics and Geometry
Technische Universität Wien, Wiedner Hauptstraße 8-10/104
A-1040 Wien, Austria
`marefatollah.mansouri@tuwien.ac.at`

Abstract

We provide precise asymptotic estimates for the number of general phylogenetic networks by using analytic combinatorial methods. Recently, this approach has been studied by Fuchs, Gittenberger and the author himself, to count networks with few reticulation vertices for two subclasses: tree-child and normal networks. We follow this line of research to show how to obtain results on the enumeration of general phylogenetic networks.

1 Introduction and Results

A phylogenetic network is defined as a directed acyclic graph (DAG) which is connected and consists of the following vertices:

(1) *leaves* which have out-degree 0 and in-degree 1;

(2) *tree vertices* which have out-degree 2 and in-degree 1;

(3) *reticulation vertices* which have out-degree 1 and in-degree 2;

(4) and the *root node* with out-degree 2 and in-degree 0.

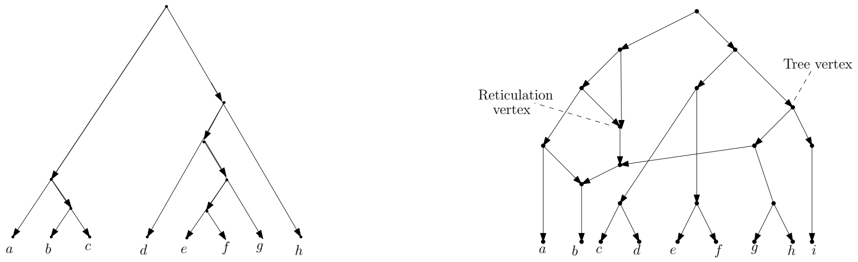
A phylogenetic network is a generalization of a phylogenetic tree which can be used to describe the evolutionary history of a set of species that is non-tree like. Phylogenetic trees are widely used to address this task and are usually computed from molecular sequences. Phylogenetic trees provide a useful representation of many evolutionary relationships, and have been well studied (see, for example [5, 18, 24, 25, 28]). However, these trees are less suited to model mechanisms of *reticulate evolution* [29], such as hybridization, recombination, or reassortment. Phylogenetic networks provide an alternative to phylogenetic trees when analyzing data sets whose evolution involves reticulate events (for more details see, [2, 17, 19]).

For a phylogenetic network there are often two kinds of labelings: *leaf-labeled* and *vertex-labeled*. In the latter case, all the vertices take different labels, and in

---

[*] This research has been supported by bilateral Austrian-Taiwanese project FWF-MOST, grant I 2309-N35.

ISSN: 2202-3518



(i) (ii)

Figure 1: (i) A rooted phylogenetic tree. (ii) A general phylogenetic network.

the former case, leaves are labeled but non-leaf vertices are unlabeled. Phylogenetic networks, as explicit representations of reticulate evolution—in particular, the formation of hybrid species—play an important role in biology. However, the process of evolution is driven by specific principles which add further restrictions on phylogenetic networks. Thus, biologists have defined many subclasses of the class of phylogenetic networks. Recently, people studying phylogenetic networks or related structures have become more and more interested in enumerative aspects. We mentioned already the shape analysis of phylogenetic trees [3, 4, 12, 13] and the bounds for the counting sequences of some classes of phylogenetic networks [21]. But other counting problems have been studied in [1, 6, 7, 8, 9, 10, 15, 20, 23, 26, 28, 31]. Combinatorial counting problems are often amenable to the rich tool box of analytic combinatorics [11], and generating functions have received much attention recently in phylogenetic enumeration problems ([14, 16, 22]).

There are quite few research studies on general phylogenetic networks. This paper is concerned with the counting of general phylogenetic networks, a basic and fundamental question which is of interest in mathematical biology [21]. The combinatorics of general phylogenetic networks, on the other hand, remains a challenge and only a few papers have addressed it. The goal of this study is to develop a more rigorous understanding of the counting problem for general phylogenetic networks with a fixed number of reticulation events. Here we come back to the open problems of [15] which are left for general networks and show that *sparsened skeleton decomposition* is a powerful tool for enumeration problems in general phylogenetic networks. The purpose of the current study is to show how the present method in [15] for *tree-child* networks can be extended for general networks. Before stating our results in more detail, we recall some definitions and previous work.

Recall that a phylogenetic network is called *tree-child* network if for every non-leaf node at least one of its children is a tree node or a leaf. Equivalently, every tree vertex must have at least one child which is not a reticulation vertex and every reticulation vertex is not directly followed by another reticulation vertex. Normal networks, on the other hand, form a subclass of the class of tree-child networks with the additional requirement that evolution does not take shortcuts. For more details about normal networks see [15, 30].

Note that variations on the definition of rooted binary phylogenetic networks are around in the literature. In general phylogenetic networks, as defined above, multiple edges are not explicitly forbidden. Our goal is indeed to study the most general model of binary phylogenetic networks that could be counted if their number of reticulation vertices are fixed. This paper attempts to provide a more detailed investigation regarding enumeration properties of general networks with multiple edges or not.

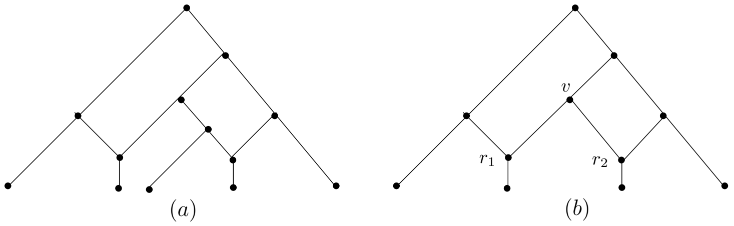


Figure 2: Two general phylogenetic networks, where only (a) is a tree-child network. Edges are directed downwards.

Now, denote by $G_{k,n}$ the number of general vertex-labeled networks with k reticulation vertices and a total of n vertices. Note that the contribution of multiple edges is considered as well. In the same way, $\tilde{G}_{k,\ell}$ is the number of general leaf-labeled networks with k reticulation vertices with ℓ leaves. In contrast to earlier findings in [15], for tree-child and normal networks, we do not have a simple connection between the vertex-labeled and the leaf-labeled for general networks; howeveri, in Section 4.0.2 we will argue in what situation someone can find a direct correspondence between them, for a subclass of general networks. The results of this study indicate:

Theorem 1.1. *For the number $G_{k,n}$ of vertex-labeled phylogenetic networks with $k \geq 1$ reticulation vertices, there is a positive constant d_k such that*

$$G_{k,n} \sim d_k \left(1-(-1)^n\right) \left(\frac{\sqrt{2}}{e}\right)^n n^{n+2k-1}, \qquad (n \to \infty).$$

In particular,

$$d_1 = \frac{\sqrt{2}}{4}; \qquad d_2 = \frac{\sqrt{2}}{32}; \qquad d_3 = \frac{\sqrt{2}}{384}.$$

The result reveals that the first and second order asymptotics are the same as the ones for vertex-labeled tree-child networks (see [15]). In other words, we can show that for the general networks with a fixed number of reticulation vertices, the additional networks that do not satisfy the tree-child conditions are asymptotically negligible as $n \to \infty$. Also, this approach leads us to the following result.

Theorem 1.2. *For the numbers $\tilde{G}_{k,\ell}$ of leaf-labeled general networks with $k \geq 1$ reticulation vertices, we have*

$$\tilde{G}_{k,\ell} \sim 2^{3k-1} d_k \left(\frac{2}{e}\right)^{\ell} \ell^{\ell+2k-1}, \qquad (\ell \to \infty)$$

where d_k is as in Theorem 1.1.

Remark. Note that this result only holds for fixed k as ℓ goes to infinity. The case when k approaches to infinity cannot be done in this way.

The other objective of this paper is to study procedures that can be used to extract explicit formulas for the number of phylogenetic networks with a fixed reticulation vertices. So points of the presented argumentation can shed some light on open questions that are left in [31] and [8] about an explicit formula for the count of phylogenetic networks. For some fixed number of reticulation vertices we have the following results.

Type of networks	Formulas
Vertex-labeled general networks with one reticulation vertex	Eq. 5
Leaf-labeled general networks with one reticulation vertex	Eq. 7
Vertex-labeled general networks with two reticulation vertices	Prop. 4.4
Leaf-labeled general networks with two reticulation vertices	Prop.4.5
Vertex-labeled general networks with three vertices	Prop.4.7
Leaf-labeled general networks with three reticulation vertices	Eq. 35

Table 1: Explicit formulas for the count of general phylogenetic networks, with fixed number of reticulation vertices. Note that to get them, in the first step we will show how to set up exponential generating functions for general networks with no multiple edges, and then we will add the contribution of other networks with multiple edges.

2 Generating functions and methods from Analytic combinatorics

This section summarizes some basic concepts on combinatorial classes and their generating functions that will be used in our work. Our presentation follows closely [11] (although with much less detail), and the reader who is interested to know more on the topic is referred to [11, mainly Chapters I.5, II.1, II.5, VI.3, VII.3, VII.4].

2.1 (Univariate) generating functions and counting

Generally speaking, a *combinatorial class* is a collection $\mathcal{C}$ of objects of a similar kind (e.g. words, trees, graphs), endowed with a suitable notion of size or weight (which is a function $f : \mathcal{C} \longrightarrow \mathbb{N}$) in a way that there are only finitely many objects of each size. We denote by $\mathcal{C}_n$ the set of objects of size n in $\mathcal{C}$, and by c_n the cardinality of $\mathcal{C}_n$. Specifically in this paper, each combinatorial class we consider is a family of general phylogenetic networks, and the size of such a network is its number of vertices or leaves.

Objects of size n in $\mathcal{C}$ can be seen as an arrangement (following some rules to be precise) of n atoms, which are objects of size 1. In our context, these atoms are the vertices (or leaves) of the networks. In general, combinatorial objects may or may not be labeled, depending on whether the atoms constituting an object are distinguishable from one another (*labeled* case) or not (*unlabeled* case). Here, our networks will be labeled combinatorial objects.

To deal with a labeled combinatorial class $\mathcal{C}$, we introduce *exponential generating function* $C(z) = \sum_{n\geq 0} c_n \frac{z^n}{n!}$, which is a formal power series in z displays the entire counting sequence of $\mathcal{C}$. The neutral class $\mathcal{E}$ is made of a single object of size 0, and its associated generating function is $E(z) = 1$. The atomic class $\mathcal{Z}$ is made of a single object of size 1, and its associated generating function is $Z(z) = z$.

We now turn our attention to recursive *specifications* of a combinatorial class. For instance, trees are best described recursively. Note that in the next section we are going to describe a decomposition of a phylogenetic network that is based on tree structure, which will then be translated into a functional equation involving their associated exponential generating functions.

Example 2.1. A rooted plane tree is a root to which is attached a (possibly empty) sequence of trees. In other words, the class $\mathcal{T}$ of rooted plane trees is definable by the recursive equation: $\mathcal{T} = \mathcal{Z} + \mathcal{Z} \times \mathcal{T} + \mathcal{Z} \times \mathcal{T}^2 + \mathcal{Z} \times \mathcal{T}^3 + \cdots = \mathcal{Z} \times SEQ(\mathcal{T})$.

Note that $SEQ(\mathcal{T})$ correspond to sequences (i.e., m-tuples of objects of a class $\mathcal{T}$, for any $m \geq 0$) to the quasi-inverse $\frac{1}{1-T(z)}$. This holds for exponential generating functions of labeled objects. For labeled classes, the precise statement that we refer to is [11, Theorem II.1]. Here we get that the corresponding generating function satisfies $T(z) = \frac{z}{1-T(z)}$. The next step is to have access to the enumeration sequence (t_n) of a class $\mathcal{T}$ from an equation satisfied by the generating function $T(z)$ of $\mathcal{T}$. To state it, we introduce the notation $n! \cdot [z^n]T(z)$ to denote the n-th coefficient of the series $T(z)$; that is to say, writing $T(z) = \sum_{n\geq 0} t_n \frac{z^n}{n!}$, we have $t_n := n! \cdot [z^n]T(z)$. From this point on, basic algebra does the rest. First the original equation is equivalent to $T - T^2 - z = 0$. Solving this quadratic equation gives

$$\begin{aligned} T(z) &= \frac{1}{2}(1 - \sqrt{(1-4z)}) \\ &= z + z^2 + 2z^3 + 5z^4 + 14z^5 + 42z^6 + 132z^7 + 429z^8 + \cdots \\ &= \sum_{n\geq 1} \frac{1}{n}\binom{2n-2}{n-1} z^n, \end{aligned}$$

and consequently, $t_n = n! \cdot \dfrac{1}{n}\dbinom{2n-2}{n-1}$.

The other possible way, especially in the case of tree-like objects, is to appeal to the *transfer theorem* (see [11], VI.1). Before going ahead, first we illustrate some concepts which help us to clarify the details. A *singularity* of an analytic function $f(z)$ is a point z_0 on the boundary of its region of analycity for which $f(z)$ is not analytically continuable. Singularities of a function analytic at 0, which lie on the boundary of the disc of convergence, are called dominant singularities. In this case, a dominant singularity is a singularity with smallest modulus. From *Pringsheim*'s

Theorem ([11], Theorem IV.6) we know that if $f(z)$ is representable at the origin by a series expansion that has non-negative coefficients and radius of convergence ρ, then the point $z=\rho$ is a singularity of $f(z)$. The idea behind the *transfer theorem* is that if $A(z)$ and $B(z)$ are two generating functions with the same positive real number ρ as dominant singularity, so when $z \to \rho$, we can write $A(z) \sim B(z)$. We obtain the asymptotic expansion of $[z^n]A(z)$ by transferring the behaviour of $A(z)$ around its dominant singularity from a simpler function $B(z)$, from which we know the analytic behaviour.

Theorem 2.1 (Transfer Theorem). *If the generating function $A(z)$ admits an expansion of the form $A(z) \sim c \cdot (1 - \frac{z}{\rho})^{-\alpha}$ as $n \to \infty$, around its (unique) dominant singularity ρ, then we have*

$$n! \cdot [z^n]A(z) \sim n! \cdot c \cdot \frac{n^{\alpha-1}}{\Gamma(\alpha)} \cdot \rho^{-n},$$

as $n \to \infty$.

Remark. Here $A(z)$ is delta-analytic in the disk of radius ρ centered at the origin.

Recall that $[z^n]A(z)$ is the coefficient of z^n in $A(z)$, and so it is $\frac{c_n}{n!}$ (respectively, c_n), when $A(z)$ is an exponential (respectively, ordinary) generating function. Note that the location of a dominant singularity will give the exponential growth of the sequence, and the nature of this singularity the subexponential term.

These methods are fundamental results from complex analysis that allow the setting up of generating functions in its disk of convergence, but not always. In particular, the *transfer theorem* (Theorem VI.1 of [11]) is one of the suitable tools, which allows us to derive asymptotic estimates of the coefficients of generating functions. The next example is to have access to the enumeration formula of non-plane binary trees (phylogenetic trees).

Example 2.2. A class $\mathcal{P}$ of rooted leaf-labeled non-plane binary tree can be recursively described as a root followed by an unordered 2-tuple of labeled rooted trees. This recursive description is then translated to specification

$$P(z) = \frac{P(z)^2}{2} + z.$$

The term "$+z$" accounts for the case where we have just a single isolated root vertex; the factor $\frac{1}{2}$ accounts for the fact that there are two ways to designate children of the root of the tree. Solving this equation, we have $P(z) = 1 - \sqrt{1-2z}$. Therefore,

$$\begin{aligned} P(z) &= (1 - \sqrt{1-2z}) \\ &= \left(1 - \sum_{n \geq 0} \binom{\frac{1}{2}}{n} (-2z)^n \right) \end{aligned}$$

and consequently, $p_n = n! \cdot (-1) \binom{\frac{1}{2}}{n} (-2)^n = (2n-3)!!$.

2.2 Additive parameters and bivariate generating functions

It is sometimes interesting to analyze the behaviour of parameters other than size. For example, interesting parameters for plane trees can be: height, number of leaves, path length, etc. These parameters are important for algorithm analysis as they correspond to the performance of algorithms that compute with or are modeled by plane trees. We now consider *multivariate* generating functions, where additional variables $(x,\ y,\ \dots)$ record the value of other parameters of our objects. One variable is used to track the size of the structure (e.g. number of nodes in a plane tree) and the other is used to track the parameter of interest.

In our cases, we will consider one more such parameter, which is numbers of certain “unary nodes” occuring in our objects. Namely, denoting i by $c_{n,\ell}$ the number of objects of size n in the combinatorial class $\mathcal{C}$ such that the parameter has value ℓ, the multivariate exponential generating function we consider is $C(z, y) = \sum_{n,\ell} c_{n,\ell} y^{\ell} \frac{z^n}{n!}$.

For instance, in the previous example of rooted plane trees, consider one additional parameter, which is the number of leaves nodes. The coefficient of $z^n y^{\ell}$ in the generating function $T(z, y)$ is then the number of rooted plane trees with n nodes and exactly ℓ leaves, divided by $n!$.

The “dictionary” translating combinatorial specifications to equations satisfied by the generating function extends to multivariate series, and our specification that shows any such tree is leaf or sequences (≥ 1) of trees that are attached to the root node. This gives $T(z, y) = zy + \dfrac{zT(z, y)}{1 - T(z, y)}$.

3 Decomposing binary phylogenetic networks

In order to count general phylogenetic networks, we will adjust the procedure of sparsened skeleton decomposition for general networks. This method is well-studied for tree-child networks, with k reticulation vertices, in [15]. We will use the decomposition to obtain a reduction which can be easily analyzed by means of generating functions. Consider a general phylogenetic network having k reticulation vertices. Then each such vertex has two incoming edges. If one edge is removed for each of the k reticulation vertices, then the remaining graph is again a Motzkin tree (labeled and nonplane). Depending upon our choice of removed edges, this Motzkin tree has at most $2k$ unary nodes. Recall that for tree-child networks this method gives exactly $2k$ unary nodes.

Now consider the following procedure: start with a Motzkin tree T with not more than $2k$ unary nodes and n vertices in total.

- Add directed edges such that each edge connects two unary nodes and no two edges have a vertex in common. Color the started vertices of the added directed edges green and their end vertices red. Note that if a Motzkin tree has exactly $2k$ unary nodes, then the coloring procedure imposes that there will be equal k colored green and k colored red nodes (see Figure 3, (1)).

- Let R_{r_g} denote a subset of $i \geq 1$ leaves in the Motzkin tree T. Now, consider two unary vertices and make a directed path that connects these unary vertices such that it goes through all leaves in R_{rg}. As before, similarly color the first unary vertex green and consider red color for second ones; then mark (color) all leaves (leaf) in R_{r_g} *red-green* (Figure 3, (2)).

- Consider a leaf g_g with two different target unary vertices in the Motzkin tree T. Now, make two distinct directed paths that connect g_g to the mentioned unary vertices. Note that the created paths may pass through some of the leaves as well. A leaf g_g is called *double-green*. Now color target unary vertices red and then consider red-green color for all the leaves on the paths of g_g to the unary vertices (see Figure 3 (3)).

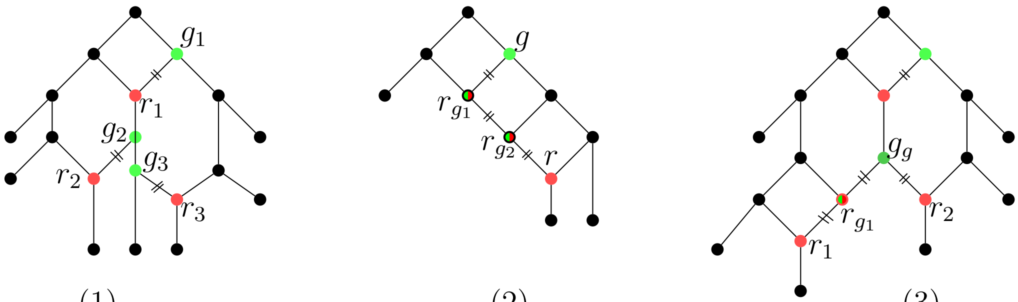
Note that in the above procedure the resulting graph must be a general phylogenetic network $\mathcal{G}$. We say then that T (keeping the colors from the above generation of $\mathcal{G}$, but not the edges) is a *colored Motzkin skeleton* (or simply Motzkin skeleton) of $\mathcal{G}$.

Proposition 3.1. *Every general network $\mathcal{G}$ with k reticulation vertices has at most 2^k different Motzkin skeletons.*

Proof. To see this, first consider two sets, not necessarily disjoint, of colored vertices obtained from the above procedure. The members of the first set are all colored vertices with starting outgoing edges, and for the second ones consider vertices with ingoing edges. Call them *pointer* and *target* sets, respectively. In this way, red-green vertices are considered in both pointer and target sets. It is not hard to see that the size of target set vertices is correspondent with the number of reticulation nodes on a general phylogenetic network. Note that in this procedure any general network with no multiple edges and n vertices is generated and each of them exactly 2^k times, so in this case every network $\mathcal{G}$ with k reticulation vertices has exactly 2^k different Motzkin skeletons. Note that as opposed to tree-child networks, multiple edges (reticulation vertex with one parent) are allowed to be in general networks. So for a reticulation vertex with just one parent, any arbitrary choice and removing of multiple edges, causes the same Motzkin skeleton. □

In the first step, we will set up exponential generating functions for general networks with no multiple edges and in the end we add the contribution of other networks with at least one multiple edge.

In order to set up generating functions for the number of general phylogenetic networks, we will construct them as follows. For a given network $\mathcal{G}$, fix one of its possible Motzkin tree skeletons, that shows us how the pointer set vertices are distributed within $\mathcal{G}$ (for instance consider networks in Figure 3 without marked edges). Now look for sparsened skeletons of $\mathcal{G}$ which contain all pointer set vertices and contract all paths between any two vertices which are either pointer vertices or an ancestor of them to one edge. Note that this ancestor may be pointer vertices itself (also see [15]). In order to construct general networks with k target vertices (reticulations), we consider a sparsened skeleton having no more than k pointer



(1) (2) (3)

Figure 3: Three general phylogenetic networks, where (1) after adding marked edges there is a bijection between green and red (reticulation) vertices in the Motzkin skeleton. General networks are depicted in (2) and (3) in which the red-green and double-green vertices appear after adding the marked edges. Edges are directed downwards.

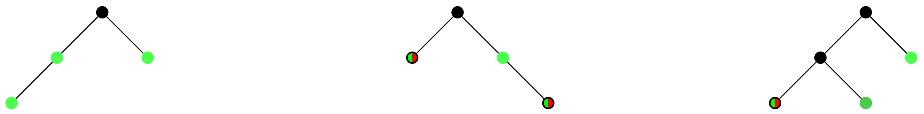


Figure 4: Corresponding sparsened skeletons of Figure 3 networks.

vertices. Then we replace all edges by paths that are made of red vertices or binary vertices with a Motzkin tree (whose unary vertices are all colored red) as second child and add a path of the same type on top of the root of the sparsened skeleton. Moreover, we attach a Motzkin tree (again with all unary vertices colored red) only to those leaves of the sparsened skeleton which are just colored green (not red-green or double green). Note that red-green and double-green nodes lie on leaves of sparsened skeleton. Do all of the above in such a way that the new structure has k target vertices (red and red-green) altogether. What we obtain so far is a Motzkin skeleton of a phylogenetic network. Finally, add edges connecting the pointer vertices to the target ones in such a way that the general phylogenetic networks condition is respected. As an advantage, a similar procedure can be used to set up generating functions for other kinds of phylogenetic networks, with a fixed number of reticulation vertices, such as “stack-free” and “galled” networks that are defined in [27, 18].

Let us set up the exponential generating function for the Motzkin trees which appear in the above construction. After all, the unary vertices in those trees will be the red nodes of our network.

Denote by $M_{\ell,n}$ the number of all vertex-labeled Motzkin trees with n vertices and ℓ unary vertices. Furthermore, let $\mathcal{M}$ be the set of all these Motzkin trees. The exponential generating function associated with $\mathcal{M}$ is

$$M(z,y)=\sum_{n\geq 1}\sum_{\ell\geq 0}M_{\ell,n}y^{\ell}\frac{z^n}{n!}.$$

Furthermore, let $M(z,y)$ denote the generating function, that is, exponential in z

but ordinary in y, associated to all Motzkin trees in $\mathcal{M}$ whose root is a unary node or a binary node, so we have

$$M(z,y) = z + zyM(z,y) + \frac{z}{2}M^2(z,y).$$

and thus

$$M(z,y) = \frac{1 - zy - \sqrt{1 + (y^2-2)z^2 - 2zy}}{z}. \tag{1}$$

The first few coefficients can be seen from

$$z + 2y\frac{z^2}{2!} + (6y^2+3)\frac{z^3}{3!} + (24y^3+36y)\frac{z^4}{4!} + \cdots .$$

In order to count general networks with a fixed number of reticulation vertices we use Motzkin trees from the class $\mathcal{M}$, which have generating function (1), and (sparsened) skeletons, as described above. We delete one of the two incoming edges of the reticulation vertex which then gives a unary-binary tree satisfying the general network properties. Conversely, we can start with the general tree or even the sparsened skeleton and then construct the network from this.

4 Counting Vertex And Leaf-Labeled General Phylogenetic Networks

In this section, we will count (vertex-labeled) general phylogenetic networks with a fixed number k of reticulation vertices. To clear up the methods, we start with simple cases, and determine the asymptotic number of general phylogenetic networks with up to 3 reticulation nodes. At the next step, we will show how corresponding generating functions of each case lead to present explicit formulas for the exact number of vertices and leaf-labeled of these. Finally, for proving Theorems 1.1 and 1.2, we will show how this pattern continues to hold for the exponential generating function of general phylogenetic networks with k reticulation vertices. As a warm-up, consider a general phylogenetic network with only one reticulation node; we use the procedure to obtain (1) and the (sparsened) skeleton, as described in the previous section. Consider a general network with no multiple edges, and delete one of the two incoming edges of the reticulation node, which then gives a unary-binary tree with exactly two unary nodes which are colored green and red (we will consider general networks with multiple edges separately). Conversely, we can start with the general tree or even the sparsened skeleton and then construct the network from this. More explicitly, let $G_i^{\nparallel}(z)$, respectively, $G_i^{\parallel}(z)$, denote exponential generating functions for networks with no multiple edges (respectively, with multiple edges) and i reticulation vertices.

Proposition 4.1. *The exponential generating function for vertex-labeled general phylogenetic networks with one reticulation node is*

$$G_1(z) = G_1^{\nparallel}(z) + G_1^{\parallel}(z) = z\frac{\tilde{a}_1(z^2) - \tilde{b}_1(z^2)\sqrt{1-2z^2}}{(1-2z^2)^{\frac{3}{2}}}, \tag{2}$$

where

$$\tilde{a}_1(z) = \tilde{b}_1(z) = 1 - z. \tag{3}$$

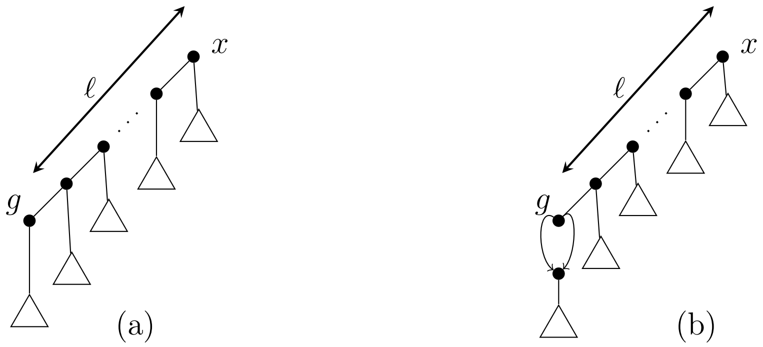


Figure 5: (a) The structure of Motzkin skeletons of networks with one reticulation vertex. It originates from a sparsened skeleton which consists of only one green vertex. It has one green vertex, denoted by g, and one red vertex which is hidden within the forest made of the triangles in the picture, which are attached to g and all the vertices on the path of length ℓ. Note that the position of the red vertex in this forest is restricted by the general condition. (b) There is a multiple edge when the root of the subtree which is attached to g is the red node.

Proof. We start with the general tree as depicted in Figure 5 (a) and add an edge starting from g and ending at a red vertex. Note that for all phylogenetic networks, this edge is not allowed to point to a node on the path from g to the root (since the network must be a DAG). Thus, when starting from the sparsened skeleton, i.e., the single green vertex g, we must add a sequence of trees on top of g which consist of a root (these vertices make the path from g to the root of the network) to which either a leaf or a binary node with two trees is attached. The red vertex must be contained in the forest made by this sequence or the tree attached to g. Note that the second expression refers to the depicted structure (b) which is for general networks with a multiple edge. In terms of generating functions evaluating the partial derivative of M (with respect to y) gives

$$G_1(z) = \frac{1}{2}\frac{\partial}{\partial y}\frac{z\tilde{M}(z,y)}{1 - zM(z,y)}\Big|_{y=0} + \frac{z^2M(z,0)}{1 - zM(z,0)},$$

where

$$\tilde{M}(z,y) = M(z,y) - zyM(z,y) = (1 - zy)M(z,y). \tag{4}$$

The factor 1/2 makes up for the fact that each network in case (a) is counted by the above procedure exactly twice. □

From this result we can now easily obtain the asymptotic number of networks.

Proposition 4.2. *Let $G_{1,n}$ denote the number of vertex-labeled general phylogenetic networks with n vertices and one reticulation vertex. If n is even then $G_{1,n}$ is zero, otherwise*

$$G_{1,n} = n! \cdot [z^n]G_1(z) = \left(\frac{\sqrt{2}}{e}\right)^n n^{n+1}\left(\frac{\sqrt{2}}{2} - \frac{\sqrt{\pi}}{2} \cdot \frac{1}{\sqrt{n}} + \mathcal{O}\left(\frac{1}{n}\right)\right),$$

as $n \to \infty$.

Proof. The function (2) has two dominant singularities, namely at $\pm 1/\sqrt{2}$, with singular expansions

$$G_1(z) \underset{z\to 1/\sqrt{2}}{\sim} \frac{1}{8(1-\sqrt{2}z)^{3/2}}, \qquad G_1(z) \underset{z\to -1/\sqrt{2}}{\sim} -\frac{1}{8(1+\sqrt{2}z)^{3/2}}.$$

Applying a transfer lemma for these two singularities completes the proof. ☐

4.0.1 Exact value of vertex-labeled general phylogenetic networks with one reticulation vertex

First, set $n = 2m+1$. Then, from (2) we obtain

$$[z^n]G_1(z) = [z^m]\bar{G}_1(z)$$

with

$$\bar{G}_1(z) = \frac{\tilde{a}_1(z) - \tilde{b}_1(z)\sqrt{1-2z}}{(1-2z)^{3/2}},$$

where $\tilde{a}_1(z)$ and $\tilde{b}_1(z)$ are as in (3). So we have

$$[z^m]\bar{G}_1(z) = [z^m]\frac{\tilde{a}_1(z)}{(1-2z)^{\frac{3}{2}}} - [z^m]\frac{\tilde{b}_1(z)}{(1-2z)}.$$

After some computation we have

$$[z^m]\bar{G}_1(z) = 2^m\big((m+1)\frac{\binom{2m}{m}}{4^m} - \frac{1}{2}\big).$$

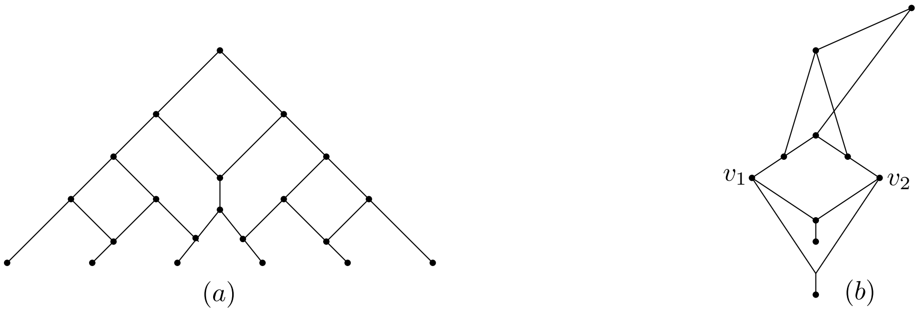
By replacing $m = (n-1)/2$ this implies

$$G_{1,n} = n! \cdot 2^{(n-3)/2}\big((n+1)\frac{\binom{n-1}{(n-1)/2}}{2^{n-1}} - 1\big). \tag{5}$$

4.0.2 Counting leaf-labeled general phylogenetic networks

Let $G_{n,k}$ (respectively, $\tilde{G}_{\ell,k}$) denote the number of vertex-labeled (respectively, leaf-labeled) general phylogenetic networks with n vertices (ℓ leaves) and k reticulation nodes. It is well-studied in [15], that for all subclasses of general networks containing only networks in which any two vertices have different sets of descendants, we have the following equation:

$$G_{k,n} = \binom{n}{\ell}(n-\ell)!\,\tilde{G}_{k,\ell}. \tag{6}$$



(a) (b)

Figure 6: Two general phylogenetic networks, where (a) is a general network in which there are no vertices with the same set of descendents, and (b) is a general network in which vertices v_1 and v_2 have a set of the same descendents.

To see this, first recall from [21] that for any phylogenetic network with ℓ leaves, k reticulation vertices and n vertices, we have $\ell + k = \frac{n+1}{2}$ (recall that n is always odd). Now all vertex-labeled general networks with n vertices and k reticulation vertices can be constructed as follows: start with a (fixed) leaf-labeled general network with ℓ leaves and k reticulation vertices. Then, choose ℓ labels from the set of n labels and re-label the leaves of the fixed network such that the order is preserved. Finally, label the remaining $n-\ell$ vertices by any permutation of the set of remaining $n-\ell$ labels. By the above structure, in this way every vertex-labeled general network is obtained exactly once.

But for classes of networks where not all networks have the above mentioned property, it is difficult to obtain a simple connection between the vertex-labeled and leaf-labeled phylogenetic networks. For that, we have to cope with symmetry in some of the generated general networks. Here, we will present complete details to show how to deal with symmetry for general networks with up to 3 fixed reticulation vertices. However, it will later be shown that as n goes infinity (respectively, ℓ), the family of general networks that need to deal with symmetry is asymptotically negligible and thus one again expects $\tilde{G}_{k,\ell} \sim \frac{\ell!}{n!}G_{k,n}$, to be a good asymptotics approximation for all leaf-labeled general networks with a fixed number of reticulation vertices when n goes to infinity.

As a warm up, we are going to take the exact formula for leaf-labeled general phylogenetic networks with one reticulation vertex. By the above points we get

$$\tilde{G}_{1,\ell} = \frac{\ell!}{n!} G_{1,n}.$$

After setting $n = 2\ell + 1$ in (5) we have

$$\tilde{G}_{1,l} = \ell!\ 2^{\ell}\Big((\ell+1)\frac{\binom{2\ell}{\ell}}{4^{\ell}} - \frac{1}{2}\Big). \tag{7}$$

4.0.3 Relationship to tree-child networks

In [15], the authors counted tree-child networks which are vertex-labeled or leaf-labeled. On the other hand, general phylogenetic networks with exactly one reticulation vertex and no multiple edges are tree-child networks. This means that $G_1^{\mathrm{H}}(z)$ exactly corresponds to a generating function for a vertex-labeled tree-child network with one reticulation vertex. This translates into

$$T_1(z) = G_1^{\mathrm{H}}(z) = \frac{1}{2}\frac{\partial}{\partial y}\frac{z\tilde{M}(z,y)}{1 - zM(z,y)}\Big|_{y=0}.$$

Solving this equation gives

$$T_1(z) = G_1^{\mathrm{H}}(z) = \frac{z^3(1 - \sqrt{1-2z^2})}{2(1-2z^2)^{3/2}}$$

as it must be. In the same way as before, the mentioned approaches immediately imply that

$$T_{1,n} = G_{1,n}^{\mathrm{H}} = n!\ 2^{(n-3)/2}\Big((n-1)\frac{\binom{n-1}{(n-1)/2}}{2^{n-1}} - 1\Big), \tag{8}$$

and for leaf-labeled,

$$\tilde{T}_{1,\ell} = G_{1,\ell}^{\mathrm{H}} = \ell!\ 2^{\ell}\ \Big(\ell\frac{\binom{2\ell}{\ell}}{4^{\ell}} - \frac{1}{2}\Big). \tag{9}$$

This approach for the leaf labeled case can be seen in [31] with different methods.

4.1 General phylogenetic network with two reticulation vertices

Now we expand the methods for general phylogenetic networks with two reticulation nodes. For this case, we use some variables $y_1, y_2, y_{r_g}, y_{g_g}$ to express the possible pointing of the pointer set vertices of the Motzkin skeletons. Furthermore, we now have more complicated paths (and attached trees) which replace the edges of the sparsened skeleton and thus we first set up the generating function corresponding to theses paths. To govern the situation where an edge from one of the two pointer set vertices must not point to a certain vertex on the paths in order to avoid multiple edges in the first step, we distinguish three types of unary vertices, which are the red vertices of our construction. We will define a class $\mathcal{P}$ of paths which serve as the essential building blocks for Motzkin skeletons. In this class the rules for pointing to particular red vertices differ, depending on whether (i) the red vertex lies on the path itself, (ii) it is one of the vertices of one of the attached subtrees, or (iii) the red vertex is the first vertex of the path. We will mark the red vertices of type (i) with the variable y, those of type (ii) with $\tilde{y}$, and the vertex of type (iii) with $\hat{y}$, which is introduced in order to manage structures analysis of multiple edges in phylogenetic networks.

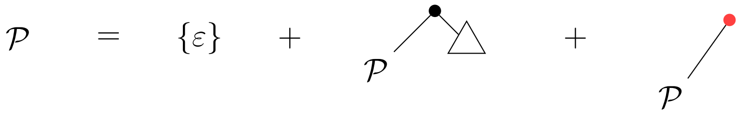


Figure 7: The specification of the class $\mathcal{P}$.

To simplify the explanation, let us use the following conventions where ε denotes the empty tree. Each path is a sequence of vertices with trees attached. Note that each vertex may be red that belongs to different categories (if it is first vertex of a path marked with $\hat{y}$, otherwise with y). In our analysis the variables y, $\tilde{y}$ and $\hat{y}$ will be replaced by a sum of variables y_i for $i \in \{1, 2, r_g, g_g\}$, where the presence of a particular y_i indicates that the corresponding g_i is allowed to point, its absence that pointing is forbidden. In particular, y represents the permission to point to vertices of the path (except its first vertex) and $\tilde{y}$ describes the permission to point to vertices of attached trees and $\hat{y}$ allows pointing to the first vertex of the path. We specify $\mathcal{P}$ as

$$\mathcal{P} = \frac{1 - zy}{1 - z(y + M(z, \tilde{y}))} + \frac{z\hat{y}}{1 - z(y + M(z, \tilde{y}))}.$$

This leads to the generating function

$$P(z, y, \tilde{y}, \hat{y}) = \frac{1 - zy + z\hat{y}}{1 - z(y + M(z, \tilde{y}))},$$

after all. Start with this assumption that in the Motzkin skeletons added directed edges are not allowed to make multiple edges (see Figure 8). The cases (a) and (b) are possible structures of Motzkin skeletons of networks with two reticulation vertices that originate from the two possible sparsened skeletons made of two green vertices (also see Figure 3 case (1)). The path of length one gives rise to the (a) Motzkin skeleton, and the cherry leads to the (b) Motzkin skeleton. Similarly, the cases (c) and (e) correspond to Motzkin skeletons that have one red-green vertex in a network (Figure 3 case (2)). Finally, (d) is a Motzkin skeleton with one double-green vertex.

Finally, we add the contribution of all other possible Motzkin tree skeletons with multiple edges as shown in Figure 9. Now we are ready to state the following result.

Proposition 4.3. *The exponential generating function for vertex-labeled general phylogenetic networks with two reticulation nodes is*

$$G_2(z) = G_2^{\#}(z) + G_2^{\parallel}(z) = z \cdot \frac{\tilde{a}_2(z^2) - \tilde{b}_2(z^2)\sqrt{1-2z^2}}{(1-2z^2)^{7/2}},$$

where

$$\tilde{a}_2(z) = z^4 - 2z^3 - \frac{1}{2}z^2 + \frac{5}{2}z \quad \textit{and} \quad \tilde{b}_2(z) = -z^2 + \frac{5}{2}z.$$

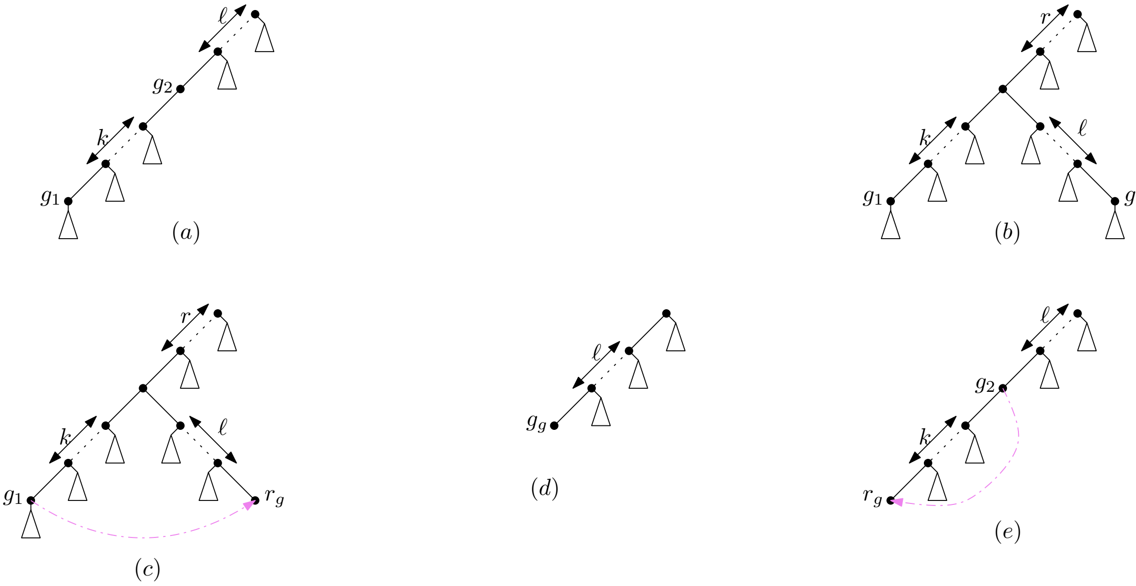


Figure 8: The structures of the Motzkin skeletons of general phylogenetic networks with two reticulation nodes such that added edges are not allowed to make multiple edges.

Proof. Consider the general phylogenetic networks arising from the Motzkin skeleton on Figure 8 (a) and complete the Motzkin skeleton by adding two edges having start vertex g_1 and g_2, respectively. Due to this, note that pointings of the green vertices do not violate the general phylogenetic network properties by making a directed cyclic component. Also, to avoid multiple edges, set up the generating function $\tilde{M}_1(z, y_1 + y_2)$ for the tree attached to the green vertex g_1. In general $\tilde{M}_i(z, y_1 + y_2) = (1 - zy_i)M(z, y_1 + y_2)$ is the specification of unary root Motzkin trees such that the pointer vertex, which is already marked by variable y_i, is not allowed to point to the root vertex. So this means pointing to the root of this tree is forbidden for g_1 but not for g_2. For all the other trees there is no pointing restriction. The analysis of the vertices on the paths is done path by path.

- Path ℓ: No green vertex is allowed to point to the vertices of that path.

- Path k: Except for the first node, pointing to all vertices is allowed for g_2, but g_1 may not point to that path at all. So we have

$$G_a^{\text{H}}(z) = \partial_{y_1}\partial_{y_2} z^2 \tilde{M}_1(z, y_1 + y_2) P(z, y_2, y_1 + y_2, 0) P(z, 0, y_1 + y_2, 0)\Big|_{y_1=0, y_2=0}.$$

Now, consider the Motzkin skeleton (b). For the trees attached to the green vertices, only pointing to the root is forbidden for parent vertices, and for all the other trees there is no pointing restriction. The analysis of the vertices on the paths is done path by path.

- Path r: No green vertex is allowed to point to the vertices of that path.
- Path k: Pointing to all vertices is allowed for g_2, but g_1 may not point to that path at all.
- The situation for path ℓ is symmetric.

In this way, Motzkin skeletons which are not respecting the general condition are generated as well. Indeed, g_1 may point to the vertex of ℓ and g_2 to the vertex of k, such that ithe generality condition is violated by making a directed cyclic component. The factor $\frac{1}{2}$ in the beginning of the expression comes from the "horizontal symmetry" (this can be briefly shown by H-S) of the Motzkin skeleton. This yields the generating function

$$\begin{aligned} G_b^{\text{H}}(z) = \frac{1}{2}\partial_{y_1}\partial_{y_2} \Bigg(&\frac{z^3 \tilde{M}_1(z, y_1 + y_2) \tilde{M}_2(z, y_1 + y_2)}{1 - zM(z, y_1 + y_2)} P(z, y_2, y_1 + y_2, y_2) P(z, y_1, y_1 + y_2, y_1) \\ &- \frac{z^3 M(z, 0)^2}{1 - zM(z, 0)} P(z, y_2, 0, y_2) P(z, y_1, 0, y_1)\Big|_{y_1=0, y_2=0} \Bigg). \end{aligned}$$

The other case of general networks has the Motzkin skeleton as shown in Figure 8(c). The property of red-green leaf entails first one added directed edge connecting g_1 to r_g. After that, there is no restriction for pointing of r_g except the vertices on the paths. This gives

$$G_c^{\text{H}}(z) = \partial_{y_r} \frac{z^3 M(z, y_r)}{(1 - zM(z, y_r))^3}\Big|_{y_{r_g}=0}.$$

Now, consider the Motzkin skeleton (d) of Figure 8. The double-green vertex g_g can point to all vertices (the pointing order does not matter, so we divide by 2) in the attached subtrees.

$$G_d^{\text{H}}(z) = \frac{1}{2}(\partial_{y_g})^2 \frac{z}{1 - zM(z, y_g)}\Big|_{y_g=0}.$$

For the final case, consider the Motzkin skeleton (e). The generality condition entails that r_g be the only possible target vertex for pointing of g_2. For all the other trees there is no pointing restriction for r_g. To avoid multiple edges, the path k cannot be a simple edge. To do that, set the generating function

$$P^{\star}(z, y, \tilde{y}, \hat{y}) = P(z, y, \tilde{y}, \hat{y}) - 1 = \frac{zM(z, \tilde{y}) + z\hat{y}}{1 - z(y + M(z, \tilde{y}))},$$

for a nonempty path. Then we get

$$G_e^{\nparallel}(z) = \partial_{y_r} \frac{z}{1 - zM(z, y_r)} P^{\star}(z, 0, y_r, 0)\Big|_{y_r=0} = \partial_{y_r} \frac{z^2 M(z, y_r)}{(1 - zM(z, y_r))^2}\Big|_{y_r=0}.$$

The exponential generating function for vertex-labeled general networks (with no multiple edges) is obtained as $G_2^{\nparallel}(z) = G_a^{\nparallel}(z) + G_b^{\nparallel}(z) + G_c^{\nparallel}(z) + G_d^{\nparallel}(z) + G_e^{\nparallel}(z)/4$ after all. This gives the following result.

$$G_2^{\nparallel}(z) = z \cdot \frac{a_2^{\nparallel}(z^2) - b_2^{\nparallel}(z^2)\sqrt{1 - 2z^2}}{(1 - 2z^2)^{7/2}}, \tag{10}$$

where

$$a_2^{\nparallel}(z) = z^4 + \frac{1}{2}z^2 + \frac{3}{2}z \quad \text{and} \quad b_2^{\nparallel}(z) = z^2 + \frac{3}{2}z. \tag{11}$$

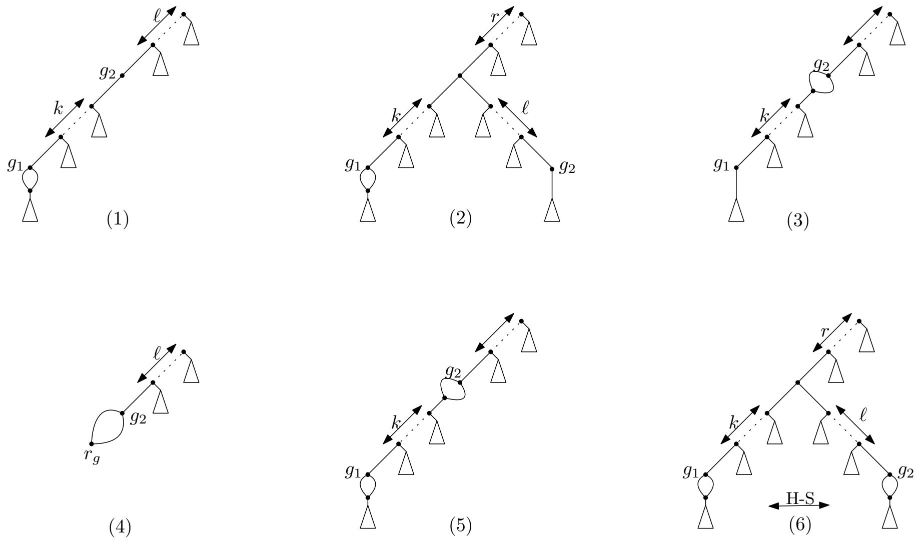


Figure 9: The structures of the Motzkin skeletons of phylogenetic networks with two reticulation nodes with all possible multiple edges contributions.

Next we will set up the exponential generating function for general networks with at least one multiple edge on their structure (see Figure 9). Altogether, we obtain

$$G_2^{\shortparallel}(z) = \frac{1}{2}\Bigg(\partial_{y_2} z^3 M(z, y_2) \frac{P(z, y_2, y_2, 0)}{1 - zM(z, y_2)}\Big|_{y_2=0} + \partial_{y_2} \frac{z^4 M(z, y_2)\tilde{M}_2(z, y_2) P(z, y_2, y_2, y_2)}{(1 - zM(z, y_2))^2}\Big|_{y_2=0}$$

$$+\partial_{y_1}\frac{z^3\tilde{M}_1(z,y_1)}{(1-zM(z,y_1))^2}\bigg|_{y_1=0}+\partial_{y_r}\frac{z^2}{(1-zM(z,y_r))}\bigg|_{y_r=0}\Bigg)$$
$$+\frac{z^4M(z,0)}{(1-zM(z,0))^2}+\frac{1}{2}\frac{z^5M^2(z,0)}{(1-zM(z,0))^3},$$

where the factor 2 appears for the expression of (1) to (4), because in these cases each general phylogenetic network is generated two times. Note that there is just a unique general network which arises from the case 5. Also, the factor 2 appears in the last term, because of horizontal symmetry. □

So the exponential generating function for vertex-labeled general phylogenetic networks with two reticulation nodes is then $G_2(z) = G_2^{\nparallel}(z) + G_2^{\parallel}(z)$. As an easy consequence, we obtain the asymptotic number of networks.

Corollary. *Let $G_{2,n}$ denote the number of vertex-labeled general phylogenetic networks with n vertices and exactly two reticulation vertices. If n is even then $G_{2,n}$ is zero, otherwise*

$$G_{2,n} = n!\cdot[z^n]G_2(z) = \left(\frac{\sqrt{2}}{e}\right)^n n^{n+3}\left(\frac{\sqrt{2}}{16}-\frac{\sqrt{\pi}}{8}\cdot\frac{1}{\sqrt{n}}+\mathcal{O}\left(\frac{1}{n}\right)\right),$$

as $n\to\infty$.

Proof. This follows by singularity analysis as before. □

4.1.1 Explicit formula for vertex-labeled general networks with two reticulation vertices

We can use generating functions $G_2^{\nparallel}(z)$ and $G_2^{\parallel}(z)$ to extract closed formulas for vertex-labeled general networks. To see this, consider the contribution of each of them separately. Start with the exponential generating function $G_2^{\nparallel}(z)$ for general networks that do not have double edges in their own structures.

First, set $n = 2m + 1$. Then, from (10) we obtain

$$[z^n]G_2^{\nparallel}(z) = [z^m]\bar{G}_2^{\nparallel}(z)$$

with

$$\bar{G}_2^{\nparallel}(z) = \frac{a_2^{\nparallel}(z) - b_2^{\nparallel}(z)\sqrt{1-2z}}{(1-2z)^{7/2}},$$

where $a_2^{\nparallel}(z)$ and $b_2^{\nparallel}(z)$ are as in (11). So we have

$$[z^m]\bar{G}_2^{\nparallel}(z) = [z^m]\frac{a_2^{\nparallel}(z)}{(1-2z)^{\frac{7}{2}}} - [z^m]\frac{b_2^{\nparallel}(z)}{(1-2z)^3}.$$

After some computation we have

$$[z^m]\bar{G}_2^{\nparallel}(z) = 2^{m-2}\Big(P_1(m)\frac{2m\binom{2m}{m}}{15(2m-1)4^m} - P_2(m)\Big),$$

where

$$P_1(m) = 30m^3 + 20m^2 + 15m - 20 \quad \text{and} \quad P_2(m) = 2m^2 + m. \tag{12}$$

Let $G_{2,n}^{\nparallel}$ (resp. $G_{2,n}^{\parallel}$) denote the number of vertex-labeled general phylogenetic networks with n vertices, two reticulation nodes and no multiple edges (respectively, with multiple edges). By substituting $m = (n-1)/2$ this implies

$$G_{2,n}^{\nparallel} = n!2^{(n-5)/2}\Big(P_1((n-1)/2)\frac{(n-1)\binom{n-1}{(n-1)/2}}{15(n-2)2^{n-1}} - P_2((n-1)/2)\Big). \tag{13}$$

Note that the correspondent generating function for general networks with multiple edges is

$$G_2^{\parallel}(z) = z \cdot \frac{a_2^{\parallel}(z^2) - b_2^{\parallel}(z^2)\sqrt{1-2z^2}}{(1-2z^2)^{1/2}},$$

such that

$$a_2^{\parallel}(z) = z^2 + z \quad \text{and} \quad b_2^{\parallel}(z) = z.$$

In the same way, it can be used to get an exact formula for vertex-labeled general networks that belong to this subclass. We refrain from giving details and just list the obtained expressions. The reader is invited to derive them herself.

$$G_{2,n}^{\parallel} = n!2^{(n-3)/2}(n-1)\Big(\frac{(n-1)\binom{n-1}{(n-1)/2}}{2^n} - \frac{1}{2})\Big). \tag{14}$$

Finally, by summing up (13) and (14) we are ready to state the following result.

Proposition 4.4. *The explicit formula for the number of vertex-labeled general networks with two reticulation vertices is*

$$G_{2,n} = G_{2,n}^{\nparallel} + G_{2,n}^{\parallel} = n!2^{(n-3)/2}\Big(A((n-1)/2)\frac{(n-1)\binom{n-1}{(n-1)/2}}{15(n-2)2^{n-1}} - B((n-1)/2)\Big), \tag{15}$$

where

$$A(m) = 30m^3 + 80m^2 - 15m - 20 \quad \textit{and} \quad B(m) = m^2 + \frac{3}{2}m. \tag{16}$$

4.1.2 Explicit formula for leaf-labeled general networks with two reticulation vertices

Note that equation (6) which comes from the described procedure in Section (4.0.2) for construction of all vertex-labeled networks from fixed leaf-labeled ones does not work anymore. This is because by applying the method there are some leaf-labeled networks which generate some vertex-labeled networks more than once (here, twice). Thus for normalization, and to deal with symmetry, the corresponding generating functions of such networks can be considered separately (see Figure 10).

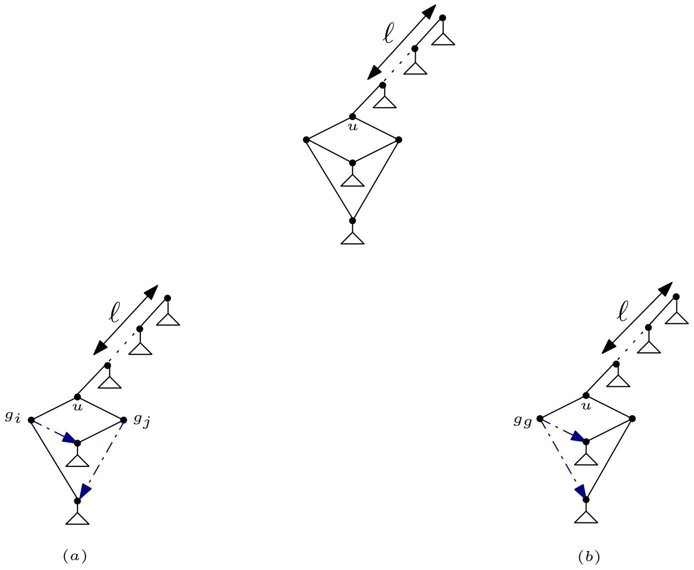


Figure 10: General network with two reticulation vertices such that two vertices have the same set of descendants which may generate from (a) by pointing green vertices to the root of each other attached trees or (b) a double-green vertex points to unary vertices with same parents.

So we have

$$G_2^{\natural}(z) = \dot{G}_2^{\natural}(z) + \ddot{G}_2^{\natural}(z),$$

such that $\dot{G}_2^{\natural}(z)$ is the corresponding generating function for general networks in which no two vertices have the same set of descendants. So

$$\dot{G}_2^{\natural}(z) = z \cdot \frac{\dot{a}_2^{\natural}(z^2) - \dot{b}_2^{\natural}(z^2)\sqrt{1-2z^2}}{(1-2z^2)^{7/2}},$$

where

$$\dot{a}_2^{\natural}(z) = -4z^5 + 11z^4 - 9z^3 + 4z^2 + z \quad \text{and} \quad \dot{b}_2^{\natural}(z) = 4z^4 - 6z^3 + 4z^2 + z.$$

For this set of general networks, we can directly use equation (6). Thus, the same procedure as before gives us

$$\dot{G}_{2,\ell}^{\natural} = \ell! 2^{\ell-1}\Big((6\ell^4 + 19\ell^3 + 18\ell^2 - 7\ell - 3)\frac{(\ell+1)\binom{2\ell+2}{\ell+1}}{(6\ell-3)(2\ell+1)4^{\ell}} - (2\ell^2 + 5\ell + 3)\Big).$$

Now we set up generating functions for the family of networks which are shown in the top of Figure (10). It is not hard to see that, by using the previous methods, each related (fixed) leaf-labeled general network can construct a vertex-labeled general network exactly twice (because of symmetry). For this case the equation (6) can be modified as $\ddot{G}^{\nparallel}_{2,\ell} = 2\frac{\ell!}{n!}\ddot{G}^{\nparallel}_{2,n}$. The generating function for this subfamily of general networks is

$$\ddot{G}^{\nparallel}_{2}(z) = z \cdot \frac{\ddot{a}^{\nparallel}_{2}(z^2) - \ddot{b}^{\nparallel}_{2}(z^2)\sqrt{1-2z^2}}{(1-2z^2)^{1/2}},$$

where

$$\ddot{a}^{\nparallel}_{2}(z) = -\frac{1}{2}z^2 + \frac{1}{2}z \quad \text{and} \quad \ddot{b}^{\nparallel}_{2}(z) = \frac{1}{2}z.$$

After some manipulation we get

$$\ddot{G}^{\nparallel}_{2,\ell} = \ell! 2^{\ell-1}\Big((\ell-1)\frac{(\ell+1)\binom{2\ell+2}{\ell+1}}{(2\ell-1)(2\ell+1)4^{\ell}}\Big).$$

The explicit formula for leaf-labeled general networks with no multiple edges and two reticulation vertices is

$$\begin{aligned} G^{\nparallel}_{2,\ell} &= \dot{G}^{\nparallel}_{2,\ell} + \ddot{G}^{\nparallel}_{2,\ell} \\ &= \ell! 2^{\ell-1}\Big((6\ell^4+19\ell^3+18\ell^2-4\ell-6)\frac{(\ell+1)\binom{2\ell+2}{\ell+1}}{(6\ell-3)(2\ell+1)4^{\ell}} - (2\ell^2+5\ell+3)\Big). \end{aligned} \tag{17}$$

To complete the details, we can extract the number of leaf-labeled networks that are generated by sparsened skeletons as depicted in Figure 9. Note that for this case, all generated networks belong to the first subclass of general networks in which Equation (6) can be used directly. So we have

$$G^{\shortparallel}_{2,\ell} = \ell! 2^{\ell-1}\Big((\ell+1)\frac{(\ell+1)\binom{2\ell+2}{\ell+1}}{4^{\ell}} - 2(\ell+1)\Big).$$

Finally, we get following result.

Proposition 4.5. *The explicit formula for the number of leaf-labeled general networks with two reticulation vertices is*

$$\tilde{G}_{2,\ell} = G^{\nparallel}_{2,\ell} + G^{\shortparallel}_{2,\ell} = \ell! 2^{\ell-1}\Big(A(\ell)\frac{(\ell+1)\binom{2\ell+2}{\ell+1}}{(6\ell-3)(2\ell+1)4^{\ell}} - B(\ell)\Big)\Big), \tag{18}$$

where

$$A(\ell) = 6\ell^4 + 31\ell^3 + 30\ell^2 - 10\ell - 3 \quad \textit{and} \quad B(\ell) = 2\ell^2 + 7\ell + 5. \tag{19}$$

In the same way, the methods can be used for the study of specifications for general phylogenetic networks with $k \geq 3$ reticulation nodes. It is obvious by increasing the number of reticulation nodes that we have to consider various numbers of the Motzkin skeletons to cover all possible cases. For more understanding, we invite the reader to look at the Appendix section to see all the Motzkin skeletons and related specifications for $k = 3$. Similarly to $k = 2$, first we focus on structures with no multiple edges and then for each Motzkin skeleton we consider possible contributions of double edges on the structures and add them to the results. After all this gives $G_3(z) = G_3^{\nparallel}(z) + G_3^{\parallel}(z)$ (see Appendix A for more details) which causes the following results for vertex-labeled general networks with fixed 3 reticulation vertices.

Proposition 4.6. *The exponential generating function for vertex-labeled general phylogenetic networks with three reticulation nodes is*

$$G_3(z) = G_3^{\nparallel}(z) + G_3^{\parallel}(z) = z \cdot \frac{a_3(z^2) - b_3(z^2)\sqrt{1-2z^2}}{(1-2z^2)^{11/2}},$$

where

$$a_3(z) = z^6 + 5z^5 - 10z^4 - \frac{23}{2}z^3 + \frac{109}{4}z^2,$$

and

$$b_3(z) = z^5 - \frac{7}{2}z^4 - 5z^3 + \frac{109}{4}z^2.$$

Corollary. *Let $G_{3,n}$ denote the number of vertex-labeled general phylogenetic networks with n vertices and exactly three reticulation vertices. If n is even then $G_{3,n}$ is zero, otherwise*

$$G_{3,n} = n![z^n]G_3(z) = \left(\frac{\sqrt{2}}{e}\right)^n n^{n+5}\left(\frac{\sqrt{2}}{192} - \frac{\sqrt{\pi}}{64}\cdot\frac{1}{\sqrt{n}} + \mathcal{O}\left(\frac{1}{n}\right)\right), \quad as \quad n \to \infty.$$

Also, consequently, similar to before we can take the explicit formulas for vertex and leaf-labeled general networks with 3 reticulation vertices. For the vertex labeled case, as before we set $n = 2m + 1$, so we have

$$[z^n]G_3(z) = [z^m]\bar{G}_3(z),$$

such that

$$[z^m]\bar{G}_3(z) = [z^m]\frac{a_3(z)}{(1-2z)^{\frac{11}{2}}} - [z^m]\frac{b_3(z)}{(1-2z)^5}.$$

By substituting $m = (n-1)/2$, we have the following result.

Proposition 4.7. *The explicit formula for the number of vertex-labeled general networks with three reticulation vertices is*

$$G_{3,n} = n! \cdot \mathcal{F}((n-1)/2),$$

where

$$\mathcal{F}(m) := [z^m]\bar{G}_3(z) = \frac{2^{m-6}}{3}\Big(A_1(m)\frac{m(m-1)\binom{2m}{m}}{35(2m-1)4^{m-2}} - B_1(m)\Big),$$

and

$$\begin{aligned} A_1(m) &= 104m^4 + 836m^3 + 876m^2 - 454m - 79, \\ \text{and } B_1(m) &= 48m^4 + 127m^3 - 60m^2 - 121m + 6. \end{aligned} \tag{20}$$

We need some more arguments to extract an explicit formula for the leaf-labeled case. It is because of the symmetry that we can see in some of the generated networks. For someone who is interested, complete details of steps can be found in the Appendix.

Now, the defined structure for paths of sparsened skeletons with the well-defined generating function (1) for attached trees, enables us to prove Theorem 1.1.

Proof of Theorem 1.1. In particular note that the function $G(z, y)$ is in the form zM (1), where z refers to vertices which lie on the paths of the sparsened skeleton. We have

$$G(z, y) = a(z, y) - b(z, y)\sqrt{1 + (y^2 - 2)z^2 - 2zy}, \tag{21}$$

where $a(z, y)$, $b(z, y)$ are polynomials in z and y with $a(z, 0) = b(z, 0) = 1$. In summary, we have an exponential generating function G_k for a phylogenetic network in the sum of terms of the form

$$\partial_{y_1} \cdots \partial_{y_k} \frac{G_1(z, y) \cdots G_s(z, y)}{(1 - G_{s+1}(z, y)) \cdots (1 - G_{s+t}(z, y))}\Big|_{y_1=0,\dots,y_k=0}. \tag{22}$$

Note that in this expression, the numerator refers to the generating function of subtrees which are rooted at green vertices. The denominator refers to sequences of subtrees which are rooted at the vertices on the paths of the sparsened skeleton. Also here the number of functions $G_{s+i}(z, y)$ is bounded by the number of edges of the sparsened skeleton increased by one (for the sequence of trees is added above the root when constructing the Motzkin skeletons). Now, the following lemma from [15] can be used for any similar structures such as $G(z, y)$.

Lemma 4.1 (see [15]). *(a) For all $\ell \geq 1$,*

$$\frac{\partial^\ell}{\partial y^\ell} G(z, y)\Big|_{y=0} = \frac{c_\ell(z) - d_\ell(z)\sqrt{1 - 2z^2}}{(1 - 2z^2)^{\ell - 1/2}},$$

where $c_\ell(z)$ and $d_\ell(z)$ are suitable polynomials.

(b) For all $\ell \geq 0$,

$$\frac{\partial^\ell}{\partial y^\ell} \frac{1}{1 - G(z, y)}\Big|_{y=0} = \frac{e_\ell(z) - f_\ell(z)\sqrt{1 - 2z^2}}{(1 - 2z^2)^{\ell + 1/2}},$$

where $e_\ell(z)$ and $f_\ell(z)$ are suitable polynomials.

We can apply the above lemma after expanding (22) and obtain:

$$G_k(z) = \frac{a_k(z) - b_k(z)\sqrt{1-2z^2}}{(1-2z^2)^p}. \tag{23}$$

We proceed to show that $p = 2k - 1/2$. For this, observe that (22) without the derivatives is of the general form given in (23), with the exponent of the denominator equal to $t/2$, which reaches its maximum for the sparsened skeleton with the maximal number of edges and is thus at most $k - 1/2$. Also, from the above lemma, we see that each differentiation increases the exponent by 1. Thus, the exponent of (22) when written as (23) is at most $2k - 1/2$. Adding up these terms gives

$$G_k(z) = \frac{a_k(z) - b_k(z)\sqrt{1-2z^2}}{(1-2z^2)^{2k-1/2}},$$

where $a_k(z)$ and $b_k(z)$ are suitable polynomials. Let $G_{k,n}$ denote the number of vertex-labeled general phylogenetic networks with n vertices and k reticulation vertices. If n is even then $G_{k,n}$ is zero; otherwise there is a positive constant d_k such that, as $n \to \infty$,

$$G_{k,n} = n![z^n]G_k(z) \sim d_k \left(\frac{\sqrt{2}}{e}\right)^n n^{n+2k-1},$$

where, by singularity analysis and Stirling's formula, we get

$$d_k = \frac{2\sqrt{2\pi}a_k(1/\sqrt{2})}{4^k\Gamma(2k-1/2)}.$$

□

Remark. For the positivity claim, we already see in [15] that the corresponding constant $\tilde{d}_k$ for normal and tree-child networks is positive, which is the lower bound of d_k for general networks.

Proposition 4.8. *For fixed k, the numbers of vertex-labeled general phylogenetic networks $G_{k,n}$ and vertex-labeled tree-child networks $T_{k,n}$, are*

$$G_{k,n} = T_{k,n}\left(1 + \mathcal{O}(\frac{1}{n})\right), \tag{24}$$

as $n \to \infty$.

Proof. First, observe that $G_{k,n} - T_{k,n}$ is bounded by the number of networks which arise from all types of Motzkin skeletons, where for each green vertex we consider all possibilities of adding an edge such that the tree-child condition is violated. Similar to the previous argument for Theorem 1, the largest number will come from the Motzkin skeletons where all pointer vertices are the leaves. Now, fix such a type of Motzkin skeleton and one of its green vertices. Then, for this vertex, we will have the following options.

- The green vertex points to the root of the subtree which is attached to one of the green vertices in the Motzkin skeleton. Note that if it points to the root of its subtree, tree-child condition is violated by making a multiple edge. For the exponential generating function this gives

  $$\partial_{y_2}\cdots\partial_{y_k}\frac{G_1(z,y)\cdots G_s(z,y)}{(1-G_{s+1}(z,y))\cdots(1-G_{s+2k-1}(z,y))}\bigg|_{y_2=0,\ldots,y_k=0},$$

  Here, and below, y is the sum of y_i's with $2 \leq i \leq k$ and not all of the y_i's must be present; also those which are present can differ from one occurrence to the next.

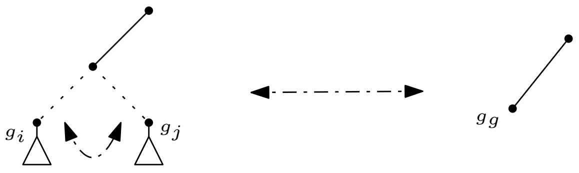
- There is a red-green vertex on the Motzkin skeleton. Note that the red-green property entails that one other pointer vertex joins to this leaf by adding a directed edge which reduces the number of the derivative by one. Then we get

  $$\partial_{y_2}\cdots\partial_{y_k}\frac{G_1(z,y)\cdots G_{s-1}(z,y)}{(1-G_{s+1}(z,y))\cdots(1-G_{s+2k-1}(z,y))}\bigg|_{y_2=0,\ldots,y_k=0}.$$

- There is a double-green vertex in the Motzkin skeleton that points to the branches of the sparsened skeleton. Then, we have

  $$\partial_{y_3}\cdots\partial_{y_k}\frac{G_1(z,y)\cdots G_{s-2}(z,y)}{2\cdot(1-G_{s+1}(z,y))\cdots(1-G_{s+2k-1}(z,y))}\bigg|_{y_2=0,\ldots,y_k=0}.$$

  The existence of a double green node in the considered skeleton is like two green vertices merged to each other. Consequently, the number of edges reduces by two, which also leads to a contribution of smaller order.



The exponential generating function of all networks arising from these Motzkin skeletons and the pointer vertices are a sum of generating functions of the above three types. Thus, we find that this generating function has the form

$$\frac{c(z)-d(z)\sqrt{1-2z^2}}{(1-2z^2)^p},$$

where $c(z)$ and $d(z)$ are suitable polynomials and the maximum of p is as follows: note that without the derivatives in the above expressions, p would be at most $k-1/2$. Also, because of the above lemma, each derivative increases this bound by one. Thus, p is at most $2k-\frac{3}{2}$.

Now, we obtain that the exponential generating function of the above number has the form

$$\frac{c(z) - d(z)\sqrt{1-2z^2}}{(1-2z^2)^{2k-\frac{3}{2}}},$$

where $c(z)$ and $d(z)$ are suitable polynomials. Singularity analysis then gives the bound

$$\mathcal{O}\left(\left(\frac{\sqrt{2}}{e}\right)^n n^{n+2k-2}\right).$$

Summing over all possible type of Motzkin skeletons and all green vertices, we obtain the claimed result. ☐

Vertex Labeled Phylogenetic Networks	$k=1$		$k=2$		$k=3$	
	c_1	c_1'	c_2	c_2'	c_3	c_3'
$N_{k,n}$	$\frac{\sqrt{2}}{2}$	$-\frac{3\sqrt{\pi}}{2}$	$\frac{\sqrt{2}}{16}$	$-\frac{3\sqrt{\pi}}{8}$	$\frac{\sqrt{2}}{192}$	$-\frac{3\sqrt{\pi}}{64}$
$T_{k,n}$	$\frac{\sqrt{2}}{2}$	$-\frac{\sqrt{\pi}}{2}$	$\frac{\sqrt{2}}{16}$	$-\frac{\sqrt{\pi}}{8}$	$\frac{\sqrt{2}}{192}$	$-\frac{\sqrt{\pi}}{64}$
$G_{k,n}$	$\frac{\sqrt{2}}{2}$	$-\frac{\sqrt{\pi}}{2}$	$\frac{\sqrt{2}}{16}$	$-\frac{\sqrt{\pi}}{8}$	$\frac{\sqrt{2}}{192}$	$-\frac{\sqrt{\pi}}{64}$

Table 2: The first two asymptotic orders (c_i and c_i') of normal, tree-child and general phylogenetic networks with $i = 1, 2$ and 3 reticulation vertices. For all of them the first coefficient is same.

4.2 Asymptotic counting of leaf-labeled general phylogenetic networks

In this part we want to prove Theorem 1.2 and argue that for the number of leaf-labeled general phylogenetic networks with $k \geq 1$ reticulation vertices (as like leaf-labeled tree-child ($\tilde{T}_{k,\ell}$) and normal networks, see [15]) we can use

$$\tilde{G}_{k,\ell} \sim 2^{3k-1} d_k \left(\frac{2}{e}\right)^{\ell} \ell^{\ell+2k-1}, \qquad (\ell \to \infty) \tag{25}$$

as a relative precise estimate of a leaf-labeled general phylogenetic network, where d_k is as in Theorem 1.1.

It is enough to show that existence of a subfamily of general networks with two vertices that have the same set of descendants is rare. Note that the number of these networks is an upper bound for the number of leaf-labeled networks that have the property that different vertex labelings might generate the same vertex-labeled network. In other words, being a pair of vertices with a set of the same descendants is a necessary but not sufficient condition to generate vertex-labeled networks twice or more. For instance, consider a leaf which is attached to edge (u, g_i) in Figure 11 (a). Nevertheless, g_1 and g_2 have a set of the same descendants but applying the procedure (4.0.2) generates each vertex-label uniquely.

(a) (b)

Figure 11: The structures of general phylogenetic networks where pairs of vertices have a same descendants set after adding the directed edges in Motzkin skeletons.

Proof of Theorem 1.2. Consider a subfamily of general networks with vertices some of which have a set of the same descendants. It is sufficient for our purposes to show that when $\ell \to \infty$, the number of these networks is asymptotically negligible. Assume, without loss of generality, that these networks have no multiple edges, because each of them reduces the number of differentiations in the expression for the exponential generating function by one, which causes the contribution of lower-order.

A precise look at the structure of Motzkin skeletons shows these networks are generated when two green vertices point to the child vertices of each other (Figure 11 (a)) or a double-green vertex points to unary vertices with the same parent (b). In the former case two green vertices and in the latter case a double-green vertex with vertex v have a set of the same descendants. Note that in each of described cases, the number of derivatives and consequently, the power of denominator in the exponential generating function, will be reduced by two. So the first two asymptotic orders are as in Theorem 1.1. This implies

$$G_{k,2\ell+2k-1} \sim \binom{2\ell+2k-1}{\ell}(\ell+2k-1)!\tilde{G}_{k,\ell}. \tag{26}$$

Now we have $\tilde{G}_{k,\ell} \sim \dfrac{\ell!}{(2\ell+2k-1)!}G_{k,2\ell+2k-1}$, from which an asymptotic result (25) follows by Theorem 1.1 and Stirling's formula. □

Appendix A
General Phylogenetic Network With Three Reticulation Nodes

This section presents a theoretical extension of the studied procedure for general phylogenetic networks with three reticulation vertices. As before, we decompose the network according to how the reticulation vertices are distributed in the networks. More explicitly, first consider the Motzkin skeletons with just green vertices (Figure 12). We can use them to figure out the rest of the Motzkin skeletons with red-green and double-green vertices as well. In the end, we add the contribution of the Motzkin skeletons with multiple edges.

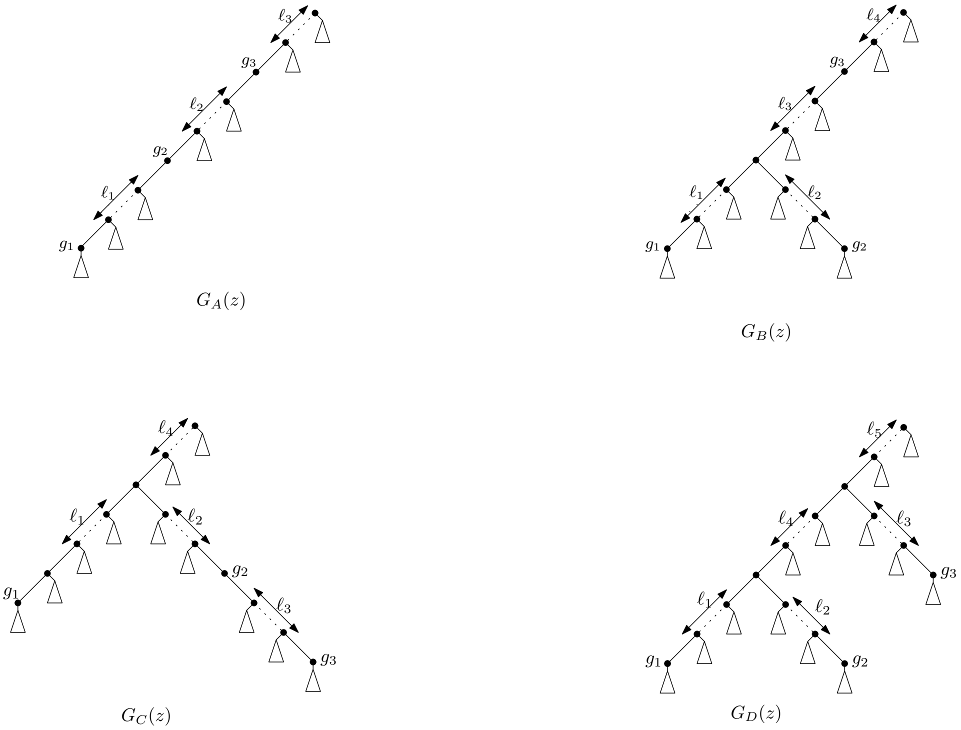
For $i,\ j \in \{1,2,3,r,g\}$, use $Y_{i,\dots,j}$ to denote the operator differentiating with respect to $y_i, \dots,\ y_j$ and setting $y_i = \cdots = y_j = 0$ afterwards, i.e., $Y_{i,\dots,j} f(z, y_i, \dots, y_j) = \left(\partial_{y_i} \dots \partial_{y_j} f\right)(z, 0, \dots, 0)$.

Now we investigate the details of extracting generating function for cases in Figure 12, We follow the same procedure that was used for general phylogenetic networks with two reticulation vertices. Start with the simple case where the three green vertices lie on one path, i.e., one green vertex is an ancestor of another, which itself is an ancestor of the third one. All possibilities for the pointings of the edges starting at g_1, g_2 and g_3 may target any vertex in all the other trees. Concerning the vertices on the spine, we have some restrictions. The edge from g_1 may not end at any vertex from ℓ_1, ℓ_2 and the root of its attached subtree. The edges from g_2 may not point to the first vertex of ℓ_1 (to avoid multiple edges) nor to any vertex of ℓ_2. Finally, no green vertex may point to the vertex ℓ_3. Note that the contribution of multiple edges will be considered in later cases. Overall, this yields the generating function

$$G_A(z) =$$
$$Y_{1,2,3}\left(\frac{z^3 \tilde{M}_1(z, y_1+y_2+y_3)}{1 - zM(z, y_1+y_2+y_3)} P(z, y_3, y_1+y_2+y_3, 0) P(z, y_2+y_3, y_1+y_2+y_3, y_3)\right).$$

Next we will determine the generating function of all general networks belonging to case that one green vertex is a common ancestor of the other two, but none of those two is an ancestor of the other one. As in the previous section we analyse the substructures. There are four vertices in the sparsened skeleton, yielding a factor z^4 . Any non-root red vertices in the subtree attached to g_1 may be targets of the edge coming from any green nodes and for root one, pointing is allowed for g_2 and g_3 (but not g_1 to avoid multiple edges) for the subtree attached to g_2 and vice versa.

- Paths ℓ_3 and ℓ_4: These paths are sequences of vertices, each with a subtree attached to it. For ℓ_4 each green vertex is allowed to point at the red vertices in these subtrees. Pointing to the vertices of the path is not allowed. Likewise, just the corresponding vertices on the path of ℓ_3 are forbidden for g_1 and g_2 by the generality condition but g_3 may point to the non-first vertex of that as well.



$G_A(z)$ $G_B(z)$

$G_C(z)$ $G_D(z)$

Figure 12: The structure of Motzkin skeletons of networks with 3 reticulation vertices. All of them originate from a sparsened skeleton which consists of only green vertices.

- Paths ℓ_1 and ℓ_2: They are symmetric, so we discuss ℓ_1. The vertices of the subtrees are allowed targets for the edge from all green vertices. The edge from g_2 and g_3 may end at each vertex of the path.

Note that the generality condition will be violated by making a cyclic component, if g_2 points to a red vertex on the path ℓ_1, and g_1 vice versa. We subtract these cases from the result. Overall, this gives, again using the operator $Y_{i,\ldots,j}$ defined above, the generating function:

$$\begin{aligned} G_B(z) =& \frac{1}{2} Y_{1,2,3}\Bigg(\frac{z^4 \tilde{M}_1(z,y_1+y_2+y_3)\tilde{M}_2(z,y_1+y_2+y_3)}{1-zM(z,y_1+y_2+y_3)} P(z,y_1+y_3,y_1+y_2+y_3,y_1+y_3) \\ & \times P(z,y_3,y_1+y_2+y_3,0)P(z,y_3+y_2,y_1+y_2+y_3,y_3+y_2) \\ & - \frac{z^4 M(z,y_3)^2}{(1-zM(z,y_3))} P(z,y_1+y_3,y_3,y_1+y_3)P(z,y_2+y_3,y_3,y_2+y_3)P(z,y_3,y_3,0)\Bigg). \end{aligned}$$

Next we pay attention to the case that one green vertex is an ancestor of another one, but not of both of them, and the third one is not an ancestor of any other green vertex, in Figure 12. The sparsened skeleton has 4 vertices and the subtrees attached to g_1 and g_3. The red vertices of the subtree of g_1 and g_3 may be targeted by any edges starting from green vertices. Note that if g_1 and g_3 have the red root attached to subtrees, they are not allowed to point at their own attached red root vertex respectively to avoid multiple edges. Next we inspect the paths:

- Path ℓ_4: All green vertices may point to the vertices of the subtrees. Pointing to the path itself is not allowed.
- Path ℓ_3: The edge starting at g_3 may point to vertices of the subtrees, but not to the vertices of the path itself. All but the first vertex for g_2 of the path as well as all tree vertices can be the end point of the edge starting at g_1 and g_2.
- Path ℓ_1: Similar to ℓ_3. The edges from g_2 and g_3 may point anywhere in the path. The vertices of the subtrees may be targeted by g_1 as well.
- Path ℓ_2: All green vertices may point to the vertices of the subtrees. To point at the vertices on the path is only allowed for g_1.

Altogether, we obtain the generating function $G_C(z)$ with the expression:

$$\begin{aligned}G_C(z) = Y_{1,2,3}\Bigg(&\frac{z^4\tilde{M}_3(z,y_1+y_2+y_3)\tilde{M}_1(z,y_1+y_2+y_3)}{1-zM(z,y_1+y_2+y_3)}P(z,y_2+y_3,y_1+y_2+y_3,y_2+y_3)\\&P(z,y_1,y_1+y_2+y_3,y_1)P(z,y_1+y_2,y_1+y_2+y_3,y_1)\\&-\frac{z^4M(z,0)^2}{1-zM(z,0)}P(z,y_1,0,y_1)^2P(z,y_2+y_3,0,y_2+y_3)\\&-\frac{z^4\tilde{M}_3(z,y_3)M_3(z,y_3)}{(1-zM(z,y_3))^2}P(z,y_1,y_3,y_1)P(z,y_2,y_3,y_2)\\&-\frac{z^4M_2(z,y_2)^2}{1-zM(z,y_2)}P(z,y_1,y_2,y_1)P(z,y_1+y_2,y_2,y_1)P(z,y_3,y_2,y_3)\Bigg).\end{aligned}$$

In this way, Motzkin skeletons which are not respecting the generality condition are generated as well; indeed, g_1 may point to the vertex on the paths ℓ_2 or ℓ_3 when both or one of g_2 and g_3 points to the vertex of ℓ_1, such that it makes a directed cyclic component.

The last case of general networks has Motzkin skeletons as shown in Figure 12. The restriction for the target vertex of the edges to be added at g_1, g_2 and g_3 follow the analogous rules in order to meet the generality constraint. Setting up the generating function follows the same pattern as before. We omit the details and get from path analysis after all:

$$\begin{aligned}G_D(z) = &\frac{1}{2}Y_{1,2,3}\\&\Bigg(\frac{z^5\tilde{M}_3(z,y_1+y_2+y_3)\tilde{M}_2(z,y_1+y_2+y_3)\tilde{M}_1(z,y_1+y_2+y_3)}{1-zM(z,y_1+y_2+y_3)}P(z,y_1+y_2,y_1+y_2+y_3,y_1+y_2)\\&\times P(z,y_1+y_3,y_1+y_2+y_3,y_1+y_3)P(z,y_2+y_3,y_1+y_2+y_3,y_2+y_3)P(z,y_3,y_1+y_2+y_3,y_3)\\&-\frac{z^5\tilde{M}_3(z,y_3)M(z,y_3)^2}{(1-zM(z,y_3))^2}P(z,y_1+y_3,y_3,y_1+y_3)P(z,y_2+y_3,y_3,y_2+y_3)P(z,y_3,y_3,y_3)\Bigg)\\&-Y_{1,2,3}\Bigg(\frac{z^5\tilde{M}_1(z,y_3)M(z,y_1)^2}{(1-zM(z,y_1))^2}P(z,y_1+y_3,y_1,y_1+y_3)P(z,y_2+y_1,y_1,y_2+y_1)P(z,y_3,y_1,y_3)\\&-\frac{z^5M(z,0)^3}{(1-zM(z,0))^2}P(z,y_1,0,y_1)P(z,y_2,0,y_2)P(z,y_3,0,y_3)\Bigg).\end{aligned}$$

So far we have just considered the Motzkin skeletons in Figure 12 with three reticulation vertices such that only green vertices are pointer set vertices. Now we consider the structure of the Motzkin skeletons with red-green and double-green vertices and set up generating functions for them separately. Note that the crucial point is that distribution of pointer nodes on the Motzkin skeleton must be in such a way that, after adding directed edges, we get a general phylogenetic network with 3 reticulation vertices. Recall that, for any red-green leaf first we consider another pointer vertex such that it connects to these nodes by adding a directed edge. Let us start with the Motzkin skeletons that contain at least one red-green vertex. Consider a case with three pointer vertices lying on a path (two green colored vertices with a red-green leaf), in such a way that a red-green one lies on the bottom of the path (left of Figure 13). Note that we have two different expressions depending on our choice of which green vertex (g_2 or g_3) is considered first to point to a red-green leaf.

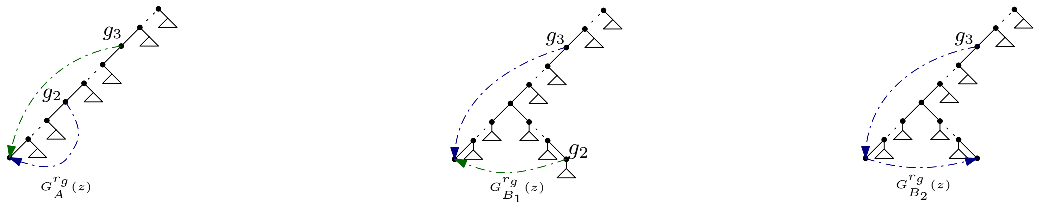


Figure 13: The structure of Motzkin skeletons with all possible red-green vertices that can be generated from the Motzkin skeletons $G_A(z)$ and $G_B(z)$ by replacing some green vertices with red-green leaves.

Now

$$G^{rg}_A(z) = Y_{r,3}\frac{z^3P^\star(z,y_3,y_r+y_3,y_3)P(z,y_3,y_r+y_3,0)}{1-zM(z,y_r+y_3)} + Y_{r,2}\frac{z^3P(z,y_2,y_r+y_2,0)}{(1-zM(z,y_r+y_2))^2}.$$

Note that to avoid multiple edges, the path between g_2 and the red-green vertex cannot be an empty edge, in the case of an added directed edge connecting g_2 to the red-green leaf.

Similarly to before, there are two possible cases for the general networks arising from the Motzkin skeletons depicted in the middle of Figure 13. In the first case, if we fix an added directed edge from g_2 to the red-green leaf, the only restriction for pointing of g_3 will be the vertices on the path that connects it to the root and its first child vertex (to avoid multiple edges). The red-green vertex may point to any non-path vertex. The second term regards the situation that a shortcut connects g_3 to the red-green vertex. After subtracting Motzkin skeletons which are not respecting the general network condition, we obtain:

$$G^{rg}_{B_1}(z) = Y_{r,3}\left(\frac{z^4M(z,y_r+y_3)}{1-zM(z,y_r+y_3)}P(z,y_3,y_r+y_3,0)P(z,y_3,y_r+y_3,y_3)^2\right)$$

$$
\begin{aligned}
&+ Y_{r,2}\left(\frac{z^4\tilde{M}_2(z,y_2+y_r)}{(1-zM(z,y_r+y_2))^2}P(z,y_2,y_r+y_2,y_2)P(z,y_r,y_r+y_2,y_r)\right)\\
&- Y_{r,2}\left(\frac{z^4M(z,0)}{(1-zM(z,0))^2}P(z,y_2,0,y_2)P(z,y_r,0,y_r)\right).
\end{aligned}
$$

Another case, such that one green vertex is a common ancestor of the other two red-green vertices, is depicted in the right of Figure 13. First, g_3 points to the one of the red-green leaves, then another directed edge connects this leaf to a second red-green leaf in the Motzkin skeleton. The edge starting at the latter red-green leaf may point to any vertex except ones on the paths. This yields the generating function:

$$
G^{rg}_{B_2}(z) = Y_r\left(\frac{z^4}{(1-zM(z,y_r))^4}\right).
$$

Consider the Motzkin skeletons depicted in Figure 14. For the first one, the generating function is given by

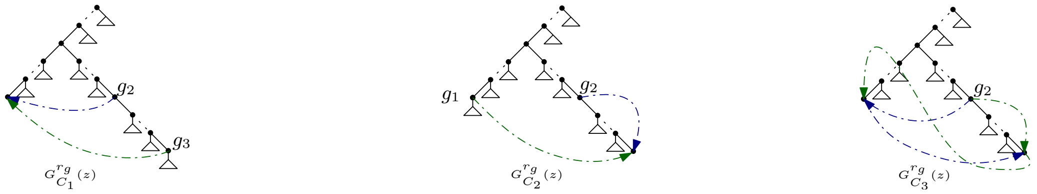


Figure 14: The structure of Motzkin skeletons that are corresponding with the Motzkin skeleton $G_C(z)$ by replacing all possible green vertices with red-green leaves.

$$
\begin{aligned}
G^{rg}_{C_1}(z) = Y_{r,3}&\left(\frac{z^4\tilde{M}_3(z,y_r+y_3)}{(1-zM(z,y_r+y_3))^2}P(z,y_r,y_r+y_3,y_r)P(z,y_3,y_r+y_3,y_3)\right.\\
&\left.-\frac{z^4M(z,0)}{(1-zM(z,0))^2}P(z,y_r,0,y_r)P(z,y_3,0,y_3)\right)\\
&+ Y_{r,2}\left(\frac{z^4M(z,y_2+y_r)}{(1-zM(z,y_r+y_2))}P(z,y_2,y_r+y_2,y_2)P(z,y_2,y_r+y_2,0)\right).
\end{aligned}
$$

For the Motzkin skeletons in the middle of Figure 14, we obtain

$$
\begin{aligned}
G^{rg}_{C_2}(z) = Y_{r,2}&\left(\frac{z^4M(z,y_r+y_2)}{(1-zM(z,y_r+y_2))^2}P(z,y_2,y_r+y_2,y_2)P(z,y_2,y_r+y_2,0)\right)\\
&+ Y_{r,1}\left(\frac{z^4\tilde{M}_1(z,y_1+y_r)}{1-zM(z,y_r+y_1)}P(z,y_1,y_r+y_1,y_1)P^{\star}(z,y_1,y_r+y_1,y_1)P(z,y_r,y_r+y_1,y_r)\right.\\
&\left.-\frac{z^4M(z,0)}{1-zM(z,0)}P(z,y_r,0,y_r)P(z,y_1,0,y_1)P^{\star}(z,y_1,0,y_1)\right).
\end{aligned}
$$

For the right one, we will take two terms for the exponential generating function depending on which red-green leaf is pointed by g_2 first. After all, we get from path analysis:

$$G^{rg}_{C_3}(z) = Y_r \left(\frac{z^4}{(1 - zM(z, y_r))^3} P^{\star}(z, 0, y_r, 0) + \frac{z^4}{(1 - zM(z, y_r))^4} \right).$$

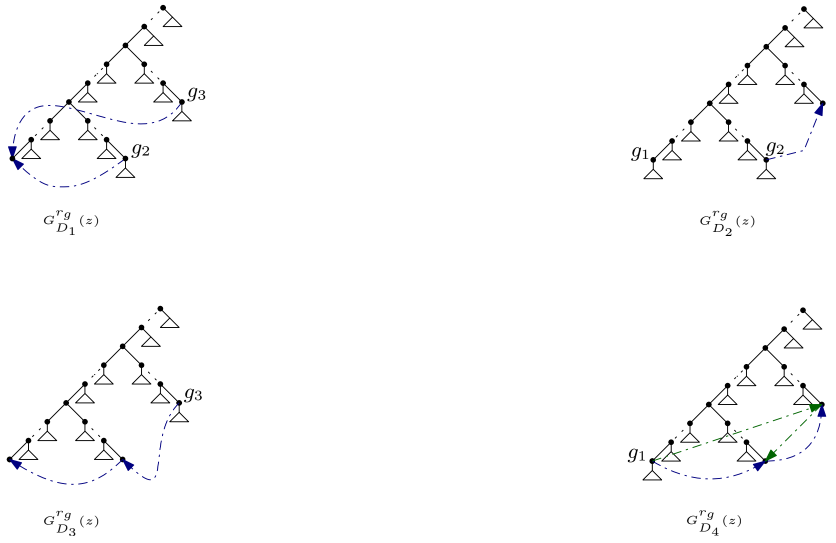


Figure 15: The Motzkin skeletons which arise from the $G_D(z)$ by considering distribution of all possible red-green leaves.

The last case of general networks with at least one red-green vertex has Motzkin skeletons as shown in Figure 15. The restriction for the target vertex of the edges to be added at pointer set vertices follows the analogous rules in order to meet the generality constraint. Setting up the generating function follows the same pattern as before. We omit the details, and get from path analysis after all:

$$\begin{aligned}
G^{rg}_{D_1}(z) = Y_{r,3} &\left(\frac{z^5 \tilde{M}_3(z, y_r{+}y_3) M(z, y_r{+}y_3)}{1 - zM(z, y_r + y_3)} P(z, y_r, y_r{+}y_3, y_r) P(z, y_3, y_r{+}y_3, y_3)^3 \right. \\
&\left. - \frac{z^5 M(z,0)^2}{(1 - zM(z,0))} P(z, y_r, 0, y_r) P(z, y_3, 0, y_3)^3 \right) \\
&+ Y_{r,2} \left(\frac{z^5 \tilde{M}_2(z, y_2{+}y_r) M(z, y_2{+}y_r)}{(1 - zM(z, y_r + y_2))^2} P(z, y_2, y_r{+}y_2, y_2)^2 P(z, y_r, y_r{+}y_2, y_r) \right. \\
&\left. - \frac{z^5 M(z,0)^2}{(1 - zM(z,0))^2} P(z, y_2, 0, y_2)^2 P(z, y_r, 0, y_r) \right);
\end{aligned}$$

$$G^{rg}_{D_2}(z) = Y_{r,1} \left(\frac{z^5 \tilde{M}_1(z, y_r{+}y_1) M(z, y_r{+}y_1)}{(1 - zM(z, y_r + y_1))^2} P(z, y_r, y_r{+}y_1, y_r) P(z, y_1, y_r{+}y_1, y_1)^2 \right.$$

$$
\begin{aligned}
&\qquad - \frac{z^5 M(z,0)^2}{(1-zM(z,0))^2} P(z,y_r,0,y_r)P(z,y_1,0,y_1)^2 \Bigg) ;\\
G^{rg}_{D_3}(z) &= Y_r \left(\frac{z^5 M(z,y_r)}{(1-zM(z,y_r))^5} \right) ;\\
G^{rg}_{D_4}(z) &= Y_r \left(\frac{z^5 M(z,y_r)}{(1-zM(z,y_r))^5} + \frac{z^5 M(z,y_r)}{(1-zM(z,y_r))^5} \right) .
\end{aligned}
$$

Finally, we consider the Motzkin skeletons with the contribution from double-green vertices as depicted in Figure 16. Note that the extra factor $\frac{1}{2}$ appears in the expression of $G^2_E(z)$ and $G^3_E(z)$, because the order of pointing for double-green vertex does not matter. After normalization we obtain:

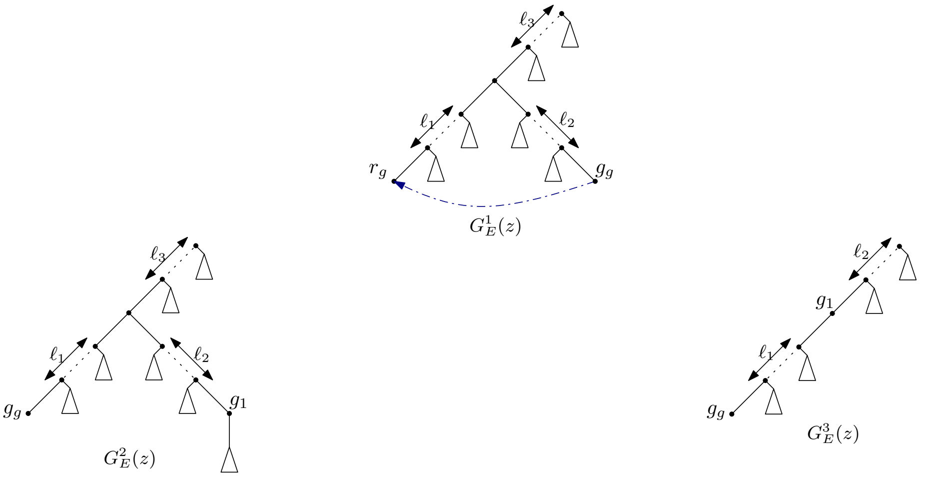


Figure 16: Three possible structures of Motzkin skeletons of networks with double-green vertices.

$$
\begin{aligned}
G^1_E(z) &= Y_{g,r} \frac{z^3 P(z,y_g,y_r+y_g,y_g)}{(1-zM(z,y_g+y_r))^2};\\
G^2_E(z) &= \frac{1}{2}(Y_g)^2 Y_1 \frac{z^3 \tilde{M}_1(z,y_1+y_g)}{1-zM(z,y_1+y_g)} P(z,y_g,y_1+y_g,y_g)P(z,y_1,y_g+y_r,y_1)\\
&\quad - \frac{1}{2}(Y_g)^2 Y_1 \frac{z^3 M(z,0)}{1-zM(z,0)} P(z,y_1,0,y_1)P(z,y_g,0,y_g)\\
&\quad - Y_g \frac{z^5 M(z,y_g)}{(1-zM(z,y_g))^5};\\
G^3_E(z) &= \frac{1}{2}(Y_g)^2 Y_1 \frac{z^2 P(z,y_1,y_1+y_g,0)}{1-zM(z,y_1+y_g)}.
\end{aligned}
$$

Now, we sum up all the generating functions obtained so far. For normalization, the result must be divided by 8, since the procedure will generate each general network eight times. Overall, by collecting everything, the exponential generating function for vertex-labeled general phylogenetic networks with three reticulation nodes is

$$G_3^{\text{ᚺ}}(z) = z \cdot \frac{a_3^{\text{ᚺ}}(z^2) - b_3^{\text{ᚺ}}(z^2)\sqrt{1-2z^2}}{(1-2z^2)^{11/2}},$$

where

$$a_3^{\text{ᚺ}}(z) = 3z^6 + 2z^5 + 4z^4 + 2z^3 + \frac{69}{4}z^2,$$

and

$$b_3^{\text{ᚺ}}(z) = z^5 + \frac{9}{2}z^4 + 11z^3 + \frac{69}{4}z^2.$$

Also, consequently, similar to before, we can take the explicit formulas for vertex and leaf-labeled general networks with 3 reticulation vertices. To see them set $n = 2m+1$, so we have

$$[z^n]G_3^{\text{ᚺ}}(z) = [z^m]\bar{G}_3^{\text{ᚺ}}(z),$$

such that

$$[z^m]\bar{G}_3^{\text{ᚺ}}(z) = [z^m]\frac{a_3^{\text{ᚺ}}(z)}{(1-2z)^{\frac{11}{2}}} - [z^m]\frac{b_3^{\text{ᚺ}}(z)}{(1-2z)^5}.$$

This gives

$$\mathcal{F}^{\text{ᚺ}}(m) := [z^m]\bar{G}^{\text{ᚺ}}{}_3(z) = \frac{2^{m-6}}{3}\Big(A_3^{\text{ᚺ}}(m)\frac{m(m-1)\binom{2m}{m}}{35(2m-1)4^{m-2}} - B_3^{\text{ᚺ}}(m)\Big),$$

where

$$A_3^{\text{ᚺ}}(m) = 104m^4 + 416m^3 + 596m^2 - 384m + 61,$$
$$B_3^{\text{ᚺ}}(m) = 48m^4 + 31m^3 - 12m^2 - 73m + 6.$$

By substituting $m = (n-1)/2$, we have $G_{3,n}^{\text{ᚺ}} = n! \cdot \mathcal{F}^{\text{ᚺ}}((n-1)/2)$.

With some more steps, but similar to before, we can present explicit formulas for the number of leaf-labeled general networks with three reticulation vertices. Let $\dot{G}_3^{\text{ᚺ}}(z)$ denote the corresponding generating function for general networks that holds the situation of equation (6) and let $\ddot{G}_3^{\text{ᚺ}}(z)$ be the generating function for general networks which arise from the Motzkin skeletons in Figure 17. We have $G_3^{\text{ᚺ}}(z) = \dot{G}_3^{\text{ᚺ}}(z) + \ddot{G}_3^{\text{ᚺ}}(z)$. So for the first subfamily (for $m > 3$) we get

$$\dot{\mathcal{F}}^{\text{ᚺ}}(m) = \frac{2^{m-5}}{3}\Big(\dot{A}_3^{\text{ᚺ}}(m)\frac{m(m-1)\binom{2m}{m}}{35(2n-5)(2n-3)(2m-1)4^{m-2}} - \dot{B}_3^{\text{ᚺ}}(m)\Big), \tag{27}$$

where

$$\dot{A}_3^{\text{ᚺ}}(m) = 280m^6 - 288m^5 - 1086m^4 - 2626m^3 + 9239m^2 - 7463m + 4290,$$

and

$$\dot{B}_3^{\text{ᚺ}}(m) = 24m^4 - \frac{31}{2}m^3 + 6m^2 + \frac{85}{2}m - 21.$$

Also for $m = 3$ ($\ell = 1$), we have $\dot{\mathcal{F}}^{\text{ᚺ}}(3) = 8$.

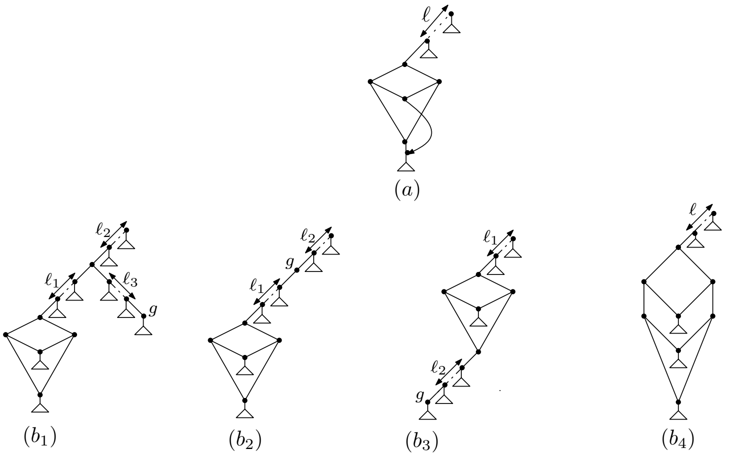


Figure 17: The family of general networks with generating function $\ddot{G}^{\nparallel}_3(z)$ such that each fixed leaf-labeled general network of (a) can construct corresponding vertex-labeled general network four times. For general network arising from second row shapes it will be exactly two times. Note that for b_1, b_2 and b_3 first we complete the structure by adding one more directed edge from green vertex to unary red vertex.

Now we consider the family of general networks with 3 reticulation vertices such that there is a pair of vertices that have the set of same descendent and applying the procedure (4.0.2) needs to cope with symmetry for them ; see Figure 17. First, we set up generating function, let's show it $G^{\nparallel}_{s_1}(z)$, for case (a) as shown at the top of Figure 17. Each fixed leaf-labeled general network which is arisen from this structure can generate corresponding vertex-labeled networks four times. So for this case we normalize equation 6 by considering $G^{\nparallel}{}_{s_1,\ell} = 4\frac{\ell!}{n!}G^{\nparallel}_{s_1,n}$. Let $G^{\nparallel}_{s_2}(z)$ denote the corresponding generating function for second row structures of Figure 17. Note that each fixed leaf-labeled network which belongs to this family can construct vertex-labeled network two times, so we get $G^{\nparallel}{}_{s_2,\ell} = 2\frac{\ell!}{n!}G^{\nparallel}_{s_2,n}$. Overall, we obtain the $\ddot{G}^{\nparallel}_3(z) = G^{\nparallel}_{s_1}(z) + G^{\nparallel}_{s_2}(z)$, where

$$G^{\nparallel}_{s_1}(z) = \frac{1}{4}\frac{z^6 M(z,0)}{1 - zM(z,0)},$$

and then we get

$$\ddot{\mathcal{F}}^{\nparallel}_{s_1}(m) := [z^m]\bar{G}^{\nparallel}_{s_1}(z) = 2^{m-2}\Big(\frac{m(m-1)\binom{2m}{m}}{(2n-3)(2m-1)4^m}\Big). \tag{28}$$

Also we have

$$\begin{aligned} G^{\nparallel}_{s_2}(z) = \frac{1}{2}\partial y\frac{z^7\tilde{M}(z,y)M(z,y)^2}{4(1 - zM(z,y))^2}P(z,y,y,y) \\ + \frac{1}{2}\partial y\frac{z^6 M(z,y)^2}{4(1 - zM(z,y))}P(z,y,y,0) \end{aligned}$$

$$+\frac{1}{2}\partial y\frac{z^6\tilde{M}(z,y)M(z,y)}{2(1-zM(z,y))^2}$$
$$+\frac{1}{4}\frac{z^8M(z,0)^3}{1-zM(z,0)}.$$

such that for $m > 3$ ($\ell > 1$) we have

$$\ddot{\mathcal{F}}^{\sharp}_{s_2}(m) := [z^m]\bar{G}^{\sharp}_{s_2}(z)$$
$$= 2^{m-3}\Big((2m^3-15m^2+38m-34)\frac{m(m-1)\binom{2m}{m}}{(2m-5)(2n-3)(2m-1)4^{m-1}} - \frac{1}{2}(m-3)\Big), \quad (29)$$

and $\ddot{\mathcal{F}}^{\sharp}_{s_2}(3) = 0$. Obviously, this means there is no general network with one leaf which can be generated by second row structures of Figure 17. Overall, by collecting everything, we have $G^{\sharp}_{3,1} = 51$ and for $\ell > 1$ we have

$$G^{\sharp}_{3,\ell} = \ell! \cdot \Big(\dot{\mathcal{F}}^{\sharp}(\ell+2) + 4\ddot{F}^{\sharp}_{s_1}(\ell+2) + 2\ddot{F}^{\sharp}_{s_2}(\ell+2)\Big)$$
$$= \ell! \cdot 2^{\ell}\Big(\frac{(\ell+1)(\ell+2)(280\ell^6+3072\ell^5+12834\ell^4+22386\ell^3+10949\ell^2-5211\ell-3990)}{840(2\ell+3)(2\ell+1)(2\ell-1)4^{\ell}}\binom{2\ell+4}{\ell+2}$$
$$-\frac{48\ell^4+415\ell^3+1326\ell^2+1799\ell+816}{768}\Big), \quad (30)$$

for the number of leaf-labeled general networks with three reticulation vertices and no multiple edges.

In the following, we want to set up exponential generating functions for general networks with three reticulation vertices and at least one multiple edge. This can be done by a case by case analysis of each of the sparsened skeletons which are depicted in Figures 18 to 23. Note that each factor of the expression makes up for the fact that each network is generated however many times. So we use them to normalize counting values of each case separately.

$$G^1_A(z) = \frac{1}{4}Y_{2,3}\frac{z^4M(z,y_2+y_3)}{1-zM(z,y_2+y_3)}P(z,y_2+y_3,y_2+y_3,y_3)P(z,y_3,y_2+y_3,0).$$

$$G^2_A(z) = \frac{1}{4}Y_{1,3}\frac{z^4\tilde{M}_1(z,y_1+y_3)}{1-zM(z,y_1+y_3)}P(z,y_3,y_1+y_3,y_3)P(z,y_3,y_1+y_3,0).$$

$$G^3_A(z) = \frac{1}{4}Y_{1,2}\frac{z^4\tilde{M}_1(z,y_1+y_2)}{(1-zM(z,y_1+y_2))^2}P(z,y_2,y_1+y_2,0).$$

$$G^4_A(z) = \frac{1}{2}Y_3\frac{z^5M(z,y_3)}{1-zM(z,y_3)}P(z,y_3,y_3,y_3)P(z,y_3,y_3,0).$$

$$G^5_A(z) = \frac{1}{2}Y_2\frac{z^5M(z,y_2)}{(1-zM(z,y_2))^2}P(z,y_2,y_2,0).$$

$$G^6_A(z) = \frac{1}{2}Y_1\frac{z^5\tilde{M}(z,y_1)}{(1-zM(z,y_1))^3}.$$

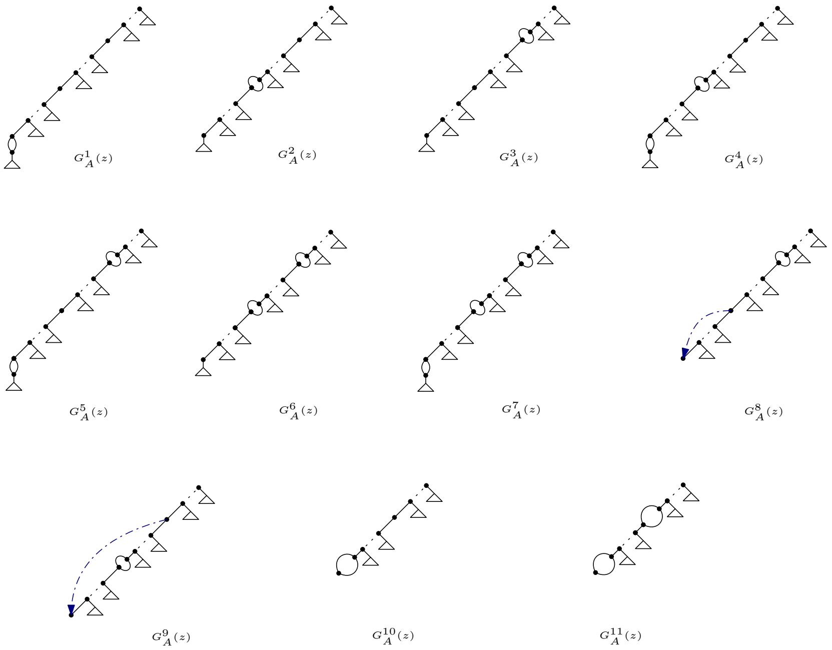


Figure 18: The structures of the Motzkin skeletons of general phylogenetic networks with at least one multiple edge which arise from $G_A(z)$.

$$G_A^7(z) = \frac{z^6 M(z,0)}{(1-zM(z,0))^3}.$$

$$G_A^8(z) = \frac{1}{4} Y_r \frac{z^4 P^\star(z,0,y_r,0)}{(1-zM(z,y_r))^2}.$$

$$G_A^9(z) = \frac{1}{4} Y_r \frac{z^4}{(1-zM(z,y_r))^3}.$$

$$G_A^{10}(z) = \frac{1}{4} Y_{r,3} \frac{z^3 P(z,y_3,y_r+y_3,0)}{1-zM(z,y_3+y_r)}.$$

$$G_A^{11}(z) = \frac{1}{2} Y_r \frac{z^4}{(1-zM(z,y_r))^2}.$$

$$\begin{aligned} G_B^1(z) = & \frac{1}{8} Y_{1,2} \Bigg(\frac{z^5 \tilde{M}_1(z,y_1+y_2) \tilde{M}_2(z,y_1+y_2)}{(1-zM(z,y_1+y_2))^2} P(z,y_1,y_1+y_2,y_1) P(z,y_2,y_1+y_2,y_2) \\ & - \frac{z^5 M(z,0)^2}{(1-zM(z,0))^2} P(z,y_1,0,y_1) P(z,y_2,0,y_2) \Bigg). \\ G_B^2(z) = & \frac{1}{4} Y_{2,3} \Bigg(\frac{z^5 M(z,y_2+y_3) \tilde{M}_2(z,y_2+y_3)}{1-zM(z,y_2+y_3)} P(z,y_2+y_3,y_2+y_3,y_2+y_3) \\ & \times P(z,y_3,y_2+y_3,0) P(z,y_3,y_2+y_3,y_3) \Bigg). \end{aligned}$$

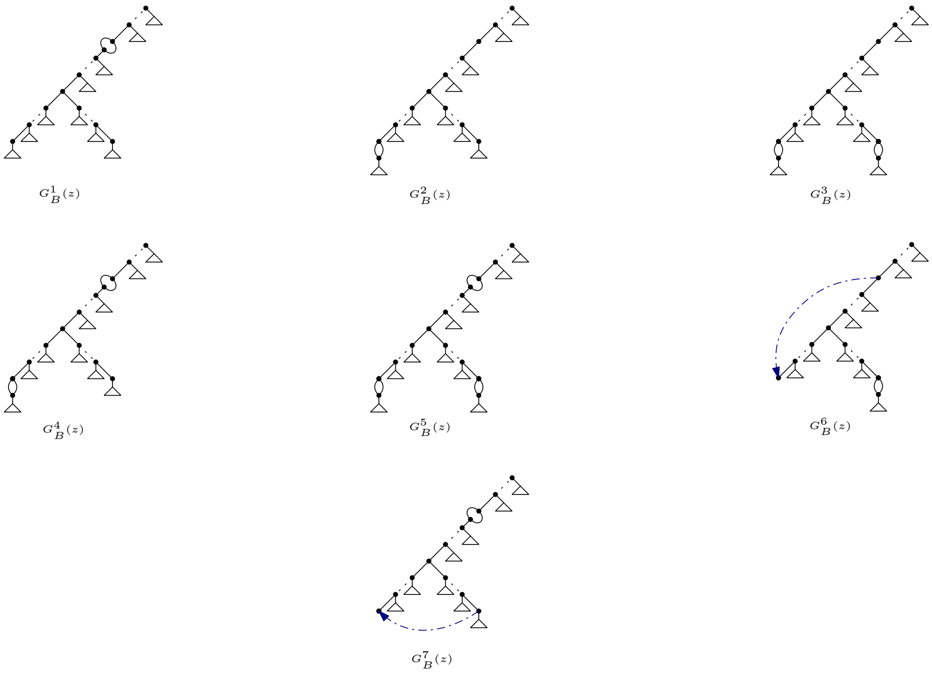


Figure 19: The structures of the Motzkin skeletons of general phylogenetic networks with at least one multiple edge which arise from $G_B(z)$.

$$G_B^3(z) = \frac{1}{4} Y_3 \frac{z^6 M(z,y_3)^2}{1 - zM(z,y_3)} P(z,y_3,y_3,y_3)^2 P(z,y_3,y_3,0).$$

$$G_B^4(z) = \frac{1}{2} Y_2 \frac{z^6 \tilde{M}_2(z,y_2) M(z,y_2)}{(1 - zM(z,y_2))^3} P(z,y_2,y_2,y_2).$$

$$G_B^5(z) = \frac{1}{2} \frac{z^7 M(z,0)^2}{(1 - zM(z,0))^4}.$$

$$G_B^6(z) = \frac{1}{4} Y_r \frac{z^5 M(z,y_r)}{(1 - zM(z,y_r))^3} P(z,y_r,y_r,y_r).$$

$$G_B^7(z) = \frac{1}{4} Y_r \frac{z^5 M(z,y_r)}{(1 - zM(z,y_r))^4}.$$

$$G_C^1(z) = \frac{1}{4} Y_{2,3} \frac{z^5 M(z,y_2+y_3) \tilde{M}_3(z,y_2+y_3)}{(1 - zM(z,y_2+y_3))^2} P(z,y_2+y_3,y_2+y_3,y_2+y_3) P(z,y_2,y_2+y_3,0).$$

$$\begin{aligned} G_C^2(z) = \frac{1}{4} Y_{1,3} \Bigg(& \frac{z^5 \tilde{M}_1(z,y_3+y_1) \tilde{M}_3(z,y_3+y_1)}{1 - zM(z,y_3+y_1)} P(z,y_1,y_3+y_1,y_1)^2 P(z,y_3,y_3+y_1,y_3) \\ & - \frac{z^5 M(z,0)^2}{1 - zM(z,0)} P(z,y_1,0,y_1)^2 P(z,y_3,0,y_3) \Bigg). \end{aligned}$$

$$\begin{aligned} G_C^3(z) = \frac{1}{4} Y_{1,2} \Bigg(& \frac{z^5 \tilde{M}_1(z,y_2+y_1) M(z,y_2+y_1)}{1 - zM(z,y_2+y_1)} P(z,y_1,y_2+y_1,y_1) \\ & \times P(z,y_1+y_2,y_1+y_2,y_1) P(z,y_2,y_2+y_1,y_2) \\ & - \frac{z^5 M(z,0)^2}{(1 - zM(z,0))^2} P(z,y_1,0,y_1) P(z,y_2,0,y_2) \Bigg). \end{aligned}$$

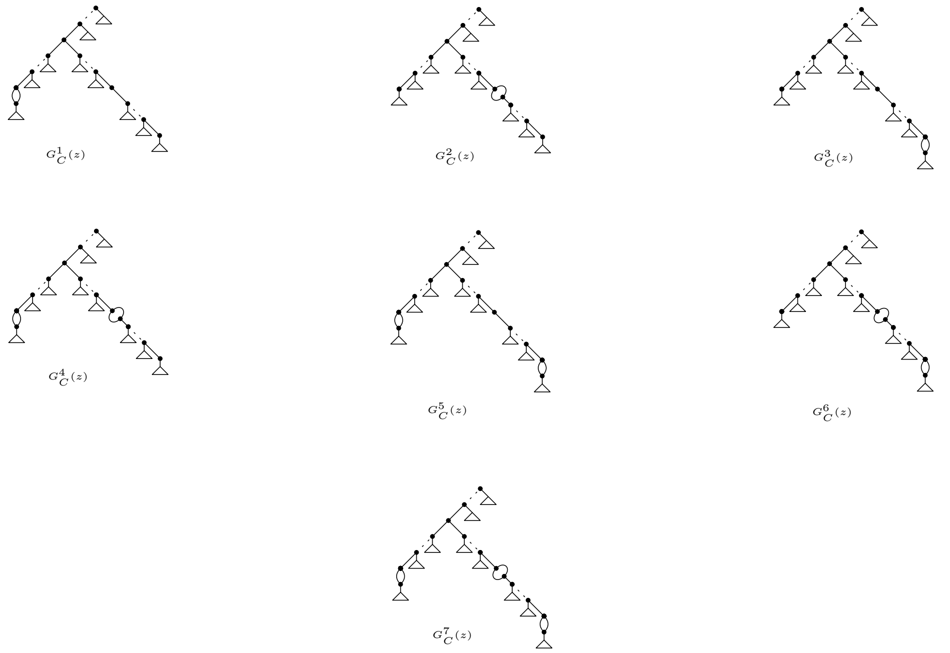


Figure 20: The structures of the Motzkin skeletons of general phylogenetic networks with at least one multiple edge which arise from $G_C(z)$.

$$G_C^4(z) = \frac{1}{2} Y_3 \frac{z^6 \tilde{M}_3(z, y_3) M(z, y_3)}{(1 - zM(z, y_3))^3} P(z, y_3, y_3, y_3).$$

$$G_C^5(z) = \frac{1}{2} Y_2 \frac{z^6 M(z, y_2)^2}{(1 - zM(z, y_2))^2} P(z, y_2, y_2, y_2) P(z, y_2, y_2, 0).$$

$$G_C^6(z) = \frac{1}{2} Y_1 \frac{z^6 \tilde{M}_1(z, y_1) M(z, y_1)}{(1 - zM(z, y_1))^2} P(z, y_1, y_1, y_1)^2.$$

$$G_C^7(z) = \frac{z^7 M(z, 0)^2}{(1 - zM(z, 0))^4}.$$

$$G_C^9(z) = \frac{1}{4} Y_r \frac{z^5 M(z, y_r)}{(1 - zM(z, y_r))^3} P(z, y_r, y_r, y_r).$$

$$G_C^{10}(z) = \frac{1}{4} Y_r \frac{z^5 M(z, y_r)}{(1 - zM(z, y_r))^4}.$$

$$G_C^{11}(z) = \frac{1}{4} Y_r \frac{z^5 M(z, y_r)}{(1 - zM(z, y_r))^2} P(z, y_r, y_r, y_r) P^{\star}(z, 0, y_r, 0).$$

$$G_C^{12}(z) = \frac{1}{4} Y_r \frac{z^5 M(z, y_r)}{(1 - zM(z, y_r))^4}.$$

$$G_C^{13}(z) = \frac{1}{4} Y_{r,1} \left(\frac{z^4 \tilde{M}_1(z, y_1 + y_r)}{1 - zM(z, y_1 + y_r)} P(z, y_1, y_1 + y_r, y_1) P(z, y_r, y_1 + y_r, y_r) \right.$$

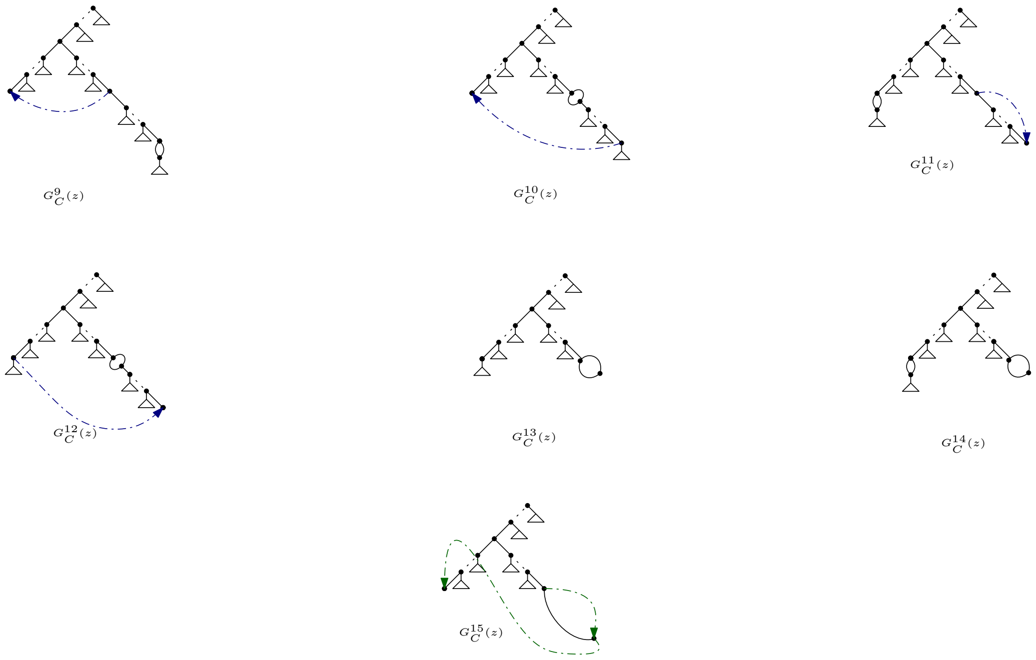

$G_C^9(z)$ $G_C^{10}(z)$ $G_C^{11}(z)$ $G_C^{12}(z)$ $G_C^{13}(z)$ $G_C^{14}(z)$ $G_C^{15}(z)$

Figure 21: The structures of the Motzkin skeletons of general phylogenetic networks with at least one multiple edge which arise from $G_C(z)$.

$$\left. - \frac{M(z,0)}{1-zM(z,0)} P(z,y_1,0,y_1)P(z,y_r,0,y_r)\right).$$

$$G_C^{14}(z) = \frac{1}{2}Y_r\left(\frac{z^5M(z,y_r)}{(1-zM(z,y_r))^2}P(z,y_r,y_r,y_r)\right).$$

$$G_C^{15}(z) = \frac{1}{4}Y_r\frac{z^4}{(1-zM(z,y_r))^3}.$$

$$G_D^1(z) = \frac{1}{4}Y_{2,3}\left(\frac{z^6M(z,y_2+y_3)\tilde{M}_2(z,y_2+y_3)\tilde{M}_3(z,y_2+y_3)}{1-zM(z,y_2+y_3)}P(z,y_2+y_3,y_2+y_3,y_2+y_3)\right.$$
$$\left.\times P(z,y_2,y_2+y_3,y_2)P(z,y_3,y_2+y_3,y_3)^2 - \frac{z^6M(z,0)^3}{(1-zM(z,0))^2}P(z,y_2,0,y_2)P(z,y_3,0,y_3)^2\right).$$

$$G_D^2(z) = \frac{1}{8}Y_{1,2}\left(\frac{z^6M(z,y_1+y_2)\tilde{M}_1(z,y_1+y_2)\tilde{M}_2(z,y_1+y_2)}{(1-zM(z,y_1+y_2))^2}P(z,y_1+y_2,y_1+y_2,y_1+y_2)\right.$$
$$\left.\times P(z,y_1,y_1+y_2,y_1)P(z,y_2,y_1+y_2,y_2) - \frac{z^6M(z,0)^3}{(1-zM(z,0))^3}P(z,y_1,0,y_1)P(z,y_2,0,y_2)\right).$$

$$G_D^3(z) = \frac{1}{4}Y_3\frac{z^7M(z,y_3)^2\tilde{M}_3(z,y_3)}{(1-zM(z,y_3))^2}P(z,y_3,y_3,y_3)^3.$$

$$G_D^4(z) = \frac{1}{2}Y_2\frac{z^7M(z,y_2)^2\tilde{M}_2(z,y_2)}{(1-zM(z,y_2))^3}P(z,y_2,y_2,y_2)^2.$$

$$G_D^5(z) = \frac{1}{2}\frac{z^8M(z,0)^3}{(1-zM(z,0))^5}.$$

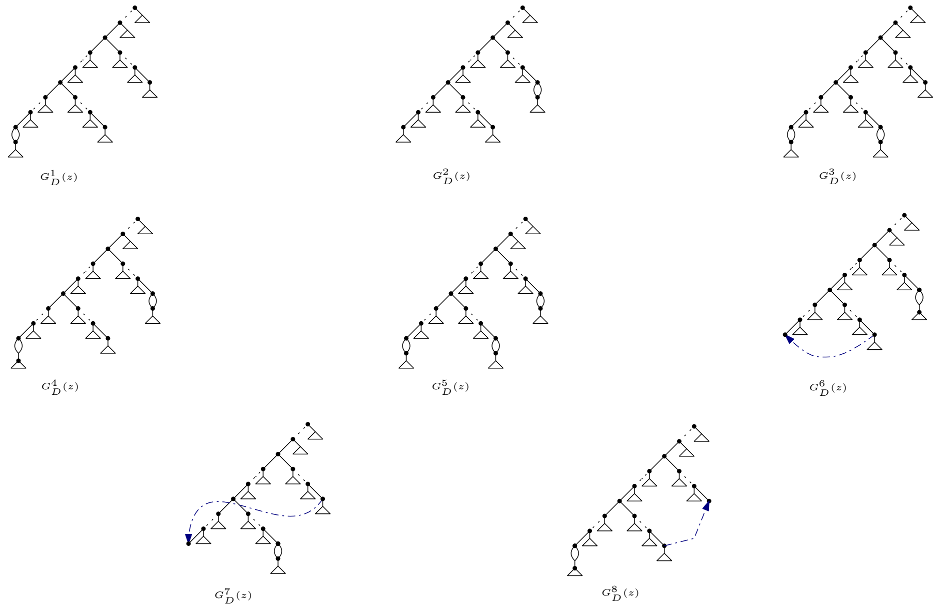

$G^1_D(z)$ $G^2_D(z)$ $G^3_D(z)$ $G^4_D(z)$ $G^5_D(z)$ $G^6_D(z)$ $G^7_D(z)$ $G^8_D(z)$

Figure 22: The structures of the Motzkin skeletons of general phylogenetic networks with at least one multiple edge which arise from $G_D(z)$.

$G'_E(z)$ $G''_E(z)$

Figure 23: The structures of the Motzkin skeletons of general phylogenetic networks with at least one multiple edge which arise from $G^2_E(z)$ and $G^3_E(z)$.

$$G^6_D(z) = \frac{1}{4}\left(Y_r \frac{z^6 M(z,y_r)^2}{(1-zM(z,y_r))^4} P(z,y_r,y_r,y_r)\right).$$
$$G^7_D(z) = \frac{1}{4}\left(Y_r \frac{z^6 M(z,y_r)^2}{(1-zM(z,y_r))^4} P(z,y_r,y_r,y_r)\right).$$
$$G^8_D(z) = \frac{1}{4}Y_r \left(\frac{z^6 M(z,y_r)^2}{(1-zM(z,y_r))^4} P(z,y_r,y_r,y_r)\right).$$

$$G'_E(z) = \frac{1}{8}(Y_g)^2 \frac{z^4 M(z,y_g)}{(1-zM(z,y_g))^2} P(z,y_g,y_g,y_g).$$
$$G''_E(z) = \frac{1}{8}(Y_g)^2 \frac{z^3}{(1-zM(z,y_g))^2}.$$

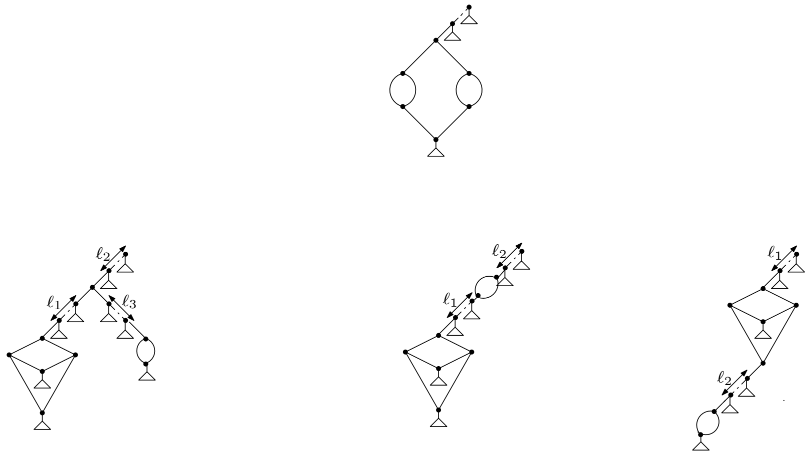


Figure 24: General networks with multiple edges and corresponding generating function $\ddot{G}^{\shortparallel}_3(z)$ such that any fixed leaf-labeled one of them can generate all-vertex labeled exactly twice.

Overall, by collecting everything, we obtain the following result.

$$G^{\shortparallel}_3(z) = z \cdot \frac{a^{\shortparallel}_3(z^2) - b^{\shortparallel}_3(z^2)\sqrt{1-2z^2}}{(1-2z^2)^{11/2}},$$

where

$$a^{\shortparallel}_3(z) = z^5 - z^4 + \frac{13}{2}z^3 + 10z^2 \quad \text{and} \quad b^{\shortparallel}_3(z) = 4z^3 + 10z^2.$$

After some computation, it gives

$$\mathcal{F}^{\shortparallel}(m) := [z^m]\bar{G}^{\shortparallel}{}_3(z) = 2^{m-1}\Big(A^{\shortparallel}_3(m)\frac{m(m-1)\binom{2m}{m}}{3(2m-1)4^m} - B^{\shortparallel}_3(m)\Big), \tag{31}$$

where

$$A^{\shortparallel}_3(m) = 6m^3 + 4m^2 - m - 2 \quad \text{and} \quad B^{\shortparallel}_3(m) = m^3 - \frac{1}{2}m^2 - \frac{1}{2}.$$

By substituting $m = (n-1)/2$ we have $G^{\shortparallel}_{3,n} = n! \cdot \mathcal{F}^{\shortparallel}((n-1)/2)$ for the number of vertex-labeled general phylogenetics with 3 reticulation vertices and at least one multiple edge in their structures.

Now we set up a generating function for the leaf-labeled. We consider $G^{\shortparallel}{}_3(z) = \dot{G}^{\shortparallel}_3(z) + \ddot{G}^{\shortparallel}_3(z)$; the right side of this equation denotes generating functions for two subfamilies of this class (general networks with multiple edges) that can be used from the equation directly or not (needs to cope with symmetry); see Figure 24. For the first subfamily we get

$$\dot{\mathcal{F}}^{\shortparallel}(m) := [z^m]\dot{\bar{G}}^{\sharp}_3(z) = 2^{m-2}\Big(\dot{A}^{\sharp}_3(m)\frac{m(m-1)\binom{2m}{m}}{3(2n-3)(2m-1)4^{m-1}} - \dot{B}^{\sharp}_3(m)\Big),$$

where

$$\dot{A}_3^{\shortparallel}(m) = 6m^4 - 5m^3 - 7m^2 - 2m + 6 \quad \text{and} \quad \dot{B}_3^{\shortparallel}(m) = 2m^3 - m^2 - m. \tag{32}$$

Also the generating function corresponding to the general networks in Figure 24 is

$$\begin{aligned}\ddot{G}_3^{\shortparallel}(z) &= \frac{1}{2}\frac{z^6 M(z,0)}{1 - zM(z,0)} + \frac{1}{4}\frac{z^8 M(z,0)^3}{(1 - zM(z,0))^3} + \frac{1}{4}\frac{z^7 M(z,0)^2}{(1 - zM(z,0))^2} + \frac{1}{2}\frac{z^7 M(z,0)^2}{(1 - zM(z,0))^2}\\ &= \frac{1}{2}\frac{z^3}{(1-2z^2)^{\frac{3}{2}}},\end{aligned}$$

such that

$$\ddot{\mathcal{F}}^{\shortparallel}(m) := [z^m]\ddot{\tilde{G}}_3^{\nparallel}(z) = 2^{m-1}m(m-1)(m-2)\Big(\frac{\binom{2m}{m}}{(2n-3)(2m-1)4^m}\Big). \tag{33}$$

Note that every member of leaf-labeled general networks arising from Figure 24 constructs corresponding vertex-labeled networks twice. Overall, by substituting $m = \ell + 2$, we have

$$\begin{aligned}G_{3,\ell}^{\shortparallel} &= \ell! \cdot \Big(\dot{\mathcal{F}}^{\shortparallel}(\ell+2) + 2\ddot{\mathcal{F}}^{\shortparallel}(\ell+2)\Big)\\ &= \ell! \cdot 2^{\ell} \cdot \Big(\frac{(\ell{+}1)(\ell{+}2)^2(6\ell^3{+}31\ell^2{+}45\ell{+}15)\binom{2\ell+4}{\ell+2}}{3(2\ell+1)(2\ell+3)4^{\ell+1}} - (2\ell^3{+}11\ell^2{+}19\ell{+}10)\Big),\end{aligned} \tag{34}$$

for the number of leaf-labeled general networks with three reticulation vertices and at least one multiple edge. Finally, we have

$$\tilde{G}_{3,\ell} = G_{3,\ell}^{\nparallel} + G_{3,\ell}^{\shortparallel} \tag{35}$$

for the number of all general phylogenetic networks with three reticulation vertices.

Acknowledgments

I would like to thank Bernhard Gittenberger and Michael Fuchs for critically reading the manuscript and making suggestions that led to significant improvements to the content and clarity of the paper. Moreover, useful comments from the editor and the three anonymous reviewers are also acknowledged.

References

[1] N. Alexeev and M. A. Alekseyev, Combinatorial scoring of phylogenetic networks, in: *Computing and Combinatorics*, *Lec. Notes Comp. Sci.* 9797, Springer (2016), 560–572.

[2] H.-J. Bandelt, Phylogenetic Networks, *Verhandlungen des Naturwissenschaftlichen Vereins Hamburg* **34** (1994), 51–71.

[3] M. Bóna, On the number of vertices of each rank in k-phylogenetic trees, *Discrete Math. Theor. Comp. Sci.* **18** (3) (2016), Paper No. 7.

[4] M. Bóna and P. Flajolet, Isomorphism and symmetries in random phylogenetic trees, *J. Appl. Probab.* **46** (4) (2009), 1005–1019.

[5] M. Bordewich and C. Semple, Determining phylogenetic networks from inter-taxa distances, *J. Math. Biol.* **73** (2) (2016), 283–303.

[6] M. Bouvel, P. Gambette and M. Mansouri, Counting Phylogenetic Networks of level 1 and 2, *J. Math. Biol.* **81** (2020), 1357–1395.

[7] G. Cardona, J. C. Pons and C. Scornavacca, Correction: Generation of Binary Tree-Child phylogenetic networks, *PLOS Comput. Biol.* **15** (2019), Paper No. 1.

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

[9] É. Czabarka, P. L. Erdős, V. Johnson and V. Moulton, Generating functions for multi-labeled trees, *Discrete Appl. Math.* **161** (2) (2013), 107–117.

[10] F. Disanto and N. A. Rosenberg, Enumeration of ancestral configurations for matching gene trees and species trees, *J. Comput. Biol.* **24** (9) (2017), 831–850.

[11] P. Flajolet and R. Sedgewick, *Analytic Combinatorics*, Cambridge University Press, Cambridge, 2009.

[12] L. R. Foulds and R. W. Robinson, Determining the asymptotic number of phylogenetic trees, in: *Combinatorial Mathematics, VII (Proc. Seventh Australian Conf., Univ. Newcastle, Newcastle, 1979)*, *Lec. Notes in Math.* 829, Springer, Berlin, 1980, 110–126.

[13] L. R. Foulds and R. W. Robinson, Enumerating phylogenetic trees with multiple labels, In *Discrete Math.* **72** (1988), 129–139.

[14] M. Fuchs, B. Gittenberger and M. Mansouri, Counting phylogenetic networks with few reticulation vertices: exact enumeration and corrections, *Australas. J. Combin.* **81** (2) (2021), 257–282.

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

[16] M. Fuchs, G.-R. Yu and L. Zhang, Asymptotic Enumeration and Distributional Properties of Galled Networks, `arXiv:2010.13324` (2021).

[17] D. Gusfield, S. Eddhu and C. Langley, Efficient reconstruction of phylogenetic networks with constrained recombination, *IEEE Comput. Soc. Bioinform. Conf.* **2** (2003), 363–374.

[18] D. H. Huson, R. Rupp and C. Scornavacca, Phylogenetic Networks: Concepts, Algorithms and Applications, *Cambridge University Press* (2010).

[19] C. R. Linder and L. H. Rieseberg, Reconstructing patterns of reticulate evolution in plants, *Amer. J. Botany* **91** (10) (2004), 1700–1708.

[20] S. Linz, K. St. John and C. Semple, Counting trees in a phylogenetic network is #P-complete, *SIAM J. Comput.* **42** (4) (2013), 1768–1776.

[21] C. McDiarmid, C. Semple and D. Welsh, Counting phylogenetic networks, *Ann. Comb.* **19** (1) (2015), 205–224.

[22] M. Pons and J. Batle, On The Exact Counting of Tree-Child Networks, *Scientific Reports* **10** (2021).

[23] N. A. Rosenberg, Counting coalescent histories, *J. Comput. Biol.* **14** (3) (2007), 360–377.

[24] E. Schröder, Vier kombinatorische Probleme, *Z. Math. Phys.* **15** (1870), 361–376.

[25] C. Semple, Phylogenetic networks with every embedded phylogenetic tree a base tree, *Bull. Math. Biol.* **78** (1) (2016), 132–137.

[26] C. Semple, Size of a phylogenetic network, *Discrete Appl. Math.* **217** (2) (2017), 362–367.

[27] C. Semple, When is a Phylogenetic Network Simply an Amalgamation of Two Trees?, *Bull. Math. Biol.* **80** (9) (2018), 1700–1708.

[28] C. Semple and M. Steel, Unicyclic networks: compatibility and enumeration, *IEEE/ACM Trans. Comput. Biology Bioinform.* **3** (2016), 84–91.

[29] P. H. A. Sneath, Cladistic Representation of Reticulate Evolution, *Systematic Zoology* **24** (3) (1975), 360–368.

[30] S. J. Willson, Properties of normal phylogenetic networks, *Bull. Math. Biol.* **72** (2) (2010), 340–358.

[31] L. Zhang, Generating normal networks via leaf insertion and nearest neighbor interchange, *BMC Bioinformatics* **20** (20) (2019), 642.

(Received 14 Dec 2020; revised 22 Nov 2021, 16 Mar 2022)