



# Semi-Simplex Phylogenetic Networks: Tree-Child Networks and Galled Trees

Michael Fuchs  

Department of Mathematical Sciences, National Chengchi University, Taipei, 116, Taiwan

Tsan-Cheng Yu  

Department of Mathematics, Fu Jen Catholic University, New Taipei City, 242062, Taiwan

## Abstract

Understanding the size of phylogenetic network classes and the typical shape of a random network from a fixed class has been one of the major research focuses in phylogenetics over the last couple of years. In this extended abstract, we consider two subclasses of the (recently introduced) class of semi-simplex phylogenetic networks, namely, semi-simplex tree-child networks and semi-simplex galled trees. We clarify their sizes relative to the (known) sizes of general tree-child networks and galled trees, respectively, and prove limit laws for parameters of random networks from these classes. Additional classes of semi-simplex networks will be considered in the full version of this paper.

**2012 ACM Subject Classification** Mathematics of computing → Enumeration; Mathematics of computing → Distribution functions

**Keywords and phrases** Semi-simplex networks, tree-child networks, galled trees, enumeration, limit laws, Sackin index

**Digital Object Identifier** 10.4230/LIPIcs.AofA.2026.27

**Funding** *Michael Fuchs*: NSTC-113-2115-M-004-004-MY3

*Tsan-Cheng Yu*: NSCT-114-2115-M-030-006

**Acknowledgements** We thank the reviewers for helpful comments.

## 1 Introduction

Phylogenetic networks are used to model reticulate evolution, as they can accommodate the merging of two lineages into one. However, the space of phylogenetic networks is extremely large. Consequently, biologists, bioinformaticians, and mathematicians have introduced various restricted subspaces that allow for feasible mathematical analysis and tractable algorithmic problems; see [18]. In this paper, we study two classes of semi-simplex networks, namely semi-simplex tree-child networks and semi-simplex galled trees. We start with general definitions and then explain the motivation for focusing on these two classes.

► **Definition 1.** *A (binary, rooted) phylogenetic network of size  $n$  is a simple, directed, acyclic graph (simple DAG) whose vertices fall into one of the following categories:*

- (i) *A (unique) root which is a node with indegree 0 and outdegree 2;*
- (ii)  *$n$  leaves which are nodes of indegree 1 and outdegree 0. In addition, the leaves are bijectively labeled with labels from the set  $\{1, \dots, n\}$ ;*
- (iii) *Tree nodes which are nodes of indegree 1 and outdegree 2;*
- (iv) *Reticulation nodes which are nodes of indegree 2 and outdegree 1.*

If no reticulation nodes are present, phylogenetic networks become *phylogenetic trees* which have been used as basic models in evolution for centuries; see [21, 22].

We also need the following three subclasses of phylogenetic networks.

► **Definition 2.** (i) *A phylogenetic network is called a tree-child network if every non-leaf node has at least one child which is not a reticulation node.*



© Michael Fuchs and Tsan-Cheng Yu;

licensed under Creative Commons License CC-BY 4.0

37th International Conference on Probabilistic, Combinatorial and Asymptotic Methods for the Analysis of Algorithms (AofA 2026).

Editor: Konstantinos Panagiotou; Article No. 27; pp. 27:1–27:17



Leibniz International Proceedings in Informatics

Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

- 43 (ii) A phylogenetic network is called a galled network if every reticulation node is in a tree  
 44 cycle, i.e., in a set of two edge-disjoint paths that start with the same tree node, end  
 45 with the same reticulation node, and whose remaining nodes are all tree nodes.  
 46 (iii) A phylogenetic network is called a galled tree if it is a galled network with the additional  
 47 requirement that the tree-cycles of any two reticulation nodes are edge-disjoint.

48 Enumeration and stochastic properties of all three classes have been studied in previous  
 49 papers; see [4, 15, 16, 20, 23] and references therein. Here, we will study their subclasses of  
 50 semi-simplex networks.

- 51 ► **Definition 3.** (i) A phylogenetic network is called simplex if every reticulation node is  
 52 directly followed by a leaf.  
 53 (ii) A phylogenetic network is called semi-simplex if the substructure rooted at the child of  
 54 every reticulation node is a tree.

55 Simplex networks, also called *one-component networks* [5] or *simplicial networks* [14],  
 56 have played an important role in the construction of networks as they are basic building  
 57 blocks; see [5, 6, 17]. Their enumerative properties for simplex tree-child networks have been  
 58 studied in [5, 15], for simplex galled networks in [16, 17], and for simplex galled trees in [12].

59 Semi-simplex networks, on the other hand, have recently been introduced in [14] and  
 60 counted only for galled networks. In this paper, we are interested in enumerating two more  
 61 classes of semi-simplex networks, namely, semi-simplex tree-child networks and semi-simplex  
 62 galled trees.

63 First, let us consider semi-simplex tree-child networks. This class has actually already  
 64 played an important role in the counting of general tree-child networks with a fixed number  
 65 of reticulation nodes; see [13]. More precisely, one crucial step in solving this counting  
 66 problem was to show that when the number of reticulation nodes is fixed, almost every  
 67 general tree-child network is a semi-simplex tree-child network. On the other hand, the  
 68 number of simplex tree-child networks is smaller by a factor of  $2^k$ . Overall, this gives for the  
 69 number  $\text{TC}_{n,k}$  of general tree-child networks, the number  $\text{SSTC}_{n,k}$  of semi-simplex tree-child  
 70 networks, and the number  $\text{STC}_{n,k}$  of simplex tree-child networks of size  $n$  and  $k$  reticulation  
 71 nodes the following chain of equivalences

$$72 \quad 2^k \text{STC}_{n,k} \sim \text{SSTC}_{n,k} \sim \text{TC}_{n,k} \sim \frac{2^{k-1}}{k! \sqrt{\pi}} n^{2k-3/2} 2^n n! \quad (1)$$

73 for fixed  $k$  as  $n \rightarrow \infty$ ; see [13] for the last two asymptotic equivalences and the formula from  
 74 Theorem 13 in [5] from which the first follows.

75 How about the number  $\text{STC}_n$  of simplex tree-child networks and the number  $\text{TC}_n$  of  
 76 general tree-child networks of size  $n$ ? First, it is easy to see that  $k \leq n - 1$  and thus

$$77 \quad \text{STC}_n = \sum_{k=0}^{n-1} \text{STC}_{n,k} \quad \text{and} \quad \text{TC}_n = \sum_{k=0}^{n-1} \text{TC}_{n,k}.$$

78 Moreover, it was shown in [15] that

$$79 \quad \text{STC}_n \sim \frac{1}{4\pi\sqrt{e}} n^{-9/4} e^{2\sqrt{n}} 2^n n!^2, \quad (n \rightarrow \infty)$$

80 and

$$81 \quad \text{TC}_n = \Theta \left( n^{-5/3} e^{a_1(3n)^{1/3}} 12^n n!^2 \right), \quad (n \rightarrow \infty),$$

82 where  $a_1$  is the largest root of the Airy function of first kind. This raises the question of  
 83 how the number  $SSTC_n$  of semi-simplex tree-child networks of size  $n$  relates to these two  
 84 numbers? Is it closer to  $TC_n$  (as for fixed  $k$  in (1)) or closer to  $STC_n$ ? This is answered by  
 85 our first main result.

86 ► **Theorem 4.** *The number  $SSTC_n$  of semi-simplex tree-child networks of size  $n$  behaves*  
 87 *asymptotically as*

$$88 \quad SSTC_n \sim \frac{1}{4\pi \sqrt[4]{e}} n^{-9/4} e^{2\sqrt{n}} 2^n n!^2, \quad (n \rightarrow \infty).$$

89 ► **Remark 5.** This result implies that  $SSTC_n \sim \sqrt[4]{e} STC_n$  and that  $SSTC_n$  is much smaller  
 90 than  $TC_n$ .

91 Next, how about limit laws for the number of reticulation nodes? Here, it was proved  
 92 in [7] that for the number  $SX_n$  of reticulation nodes of a simplex tree-child network picked  
 93 uniformly at random from the class of all simplex tree-child networks of size  $n$ :

$$94 \quad \frac{SX_n - n + \sqrt{n}}{\sqrt[4]{n/4}} \xrightarrow{d} N(0, 1), \quad (n \rightarrow \infty), \quad (2)$$

95 where  $\xrightarrow{d}$  denotes convergence in distribution and  $N(0, 1)$  is the standard normal distribution.  
 96 On the other hand, for the number  $X_n$  of reticulation nodes of a general tree-child networks  
 97 picked uniformly at random from the class of all general tree-child networks of size  $n$ :

$$98 \quad n - 1 - X_n \xrightarrow{d} \text{Poisson}(1/2), \quad (n \rightarrow \infty),$$

99 where  $\text{Poisson}(\lambda)$  denotes a Poisson random variable with parameter  $\lambda$ ; see also [7]. Again,  
 100 the random number  $SSX_n$  of reticulation nodes behaves similarly to the corresponding  
 101 random number for simplex tree-child networks.

102 ► **Theorem 6.** *The number  $SSX_n$  of reticulation nodes of a semi-simplex tree-child network*  
 103 *picked uniformly at random from the class of all semi-simplex tree-child networks of size  $n$*   
 104 *satisfies*

$$105 \quad \frac{SSX_n - n + \sqrt{n}}{\sqrt[4]{n/4}} \xrightarrow{d} N(0, 1), \quad (n \rightarrow \infty).$$

106 Finally, we consider the *Sackin index* of tree-child networks which was introduced in [25]  
 107 as an extension of the classical Sackin index of phylogenetic trees [9]. The definition is as  
 108 follows. The Sackin index is defined as the sum of heights of the leaves with the *height* of a  
 109 leaf being the length of a longest path from the root to that leaf. Denote by  $SS_n$  the Sackin  
 110 index of a random simplex tree-child network of size  $n$ . Then, it was proved in [25] that the  
 111 expected value of  $SS_n$  has order  $n^{7/4}$ . This generalizes to semi-simplex tree-child networks.

112 ► **Theorem 7.** *Let  $SSS_n$  be the Sackin index, i.e., the sum of heights of all leaves of a*  
 113 *semi-simplex tree-child network picked uniformly at random from the class of all semi-simplex*  
 114 *tree-child networks of size  $n$ . Then,*

$$115 \quad \mathbb{E}(SSS_n) = \Theta(n^{7/4}), \quad (n \rightarrow \infty).$$

116 The above results will be proved in the next section (where for Theorem 7, we only give a  
 117 sketch of the proof). Then, in Section 3, we will prove corresponding results for galled trees

## 27:4 Semi-Simplex Tree-Child Networks and Galled Trees

118 for whose numbers, we will use the same notation as above only with TC replaced by GT. It  
 119 is known from [2, 4, 12] that

$$120 \quad \text{GT}_{n,k} \sim \frac{2^{2k-1}}{(2k)!\sqrt{\pi}} n^{2k-3/2} 2^n n! \quad (3)$$

121 for fixed  $k$ , as  $n \rightarrow \infty$ , and

$$122 \quad \text{SGT}_n \sim \frac{\sqrt{2-\sqrt{2}}}{4\sqrt{\pi}} n^{-3/2} (3+2\sqrt{2})^n n!, \quad (n \rightarrow \infty), \quad (4)$$

123 and

$$124 \quad \text{GT}_n \sim \frac{17-\sqrt{17}}{136\sqrt{\pi}} n^{-3/2} 8^n n!, \quad (n \rightarrow \infty). \quad (5)$$

125 We will add asymptotic results for the remaining numbers, namely,  $\text{SGT}_{n,k}$ ,  $\text{SSGT}_{n,k}$ , and  
 126  $\text{SSGT}_n$ , and also derive stochastic results for the number of reticulation nodes and the Sackin  
 127 index which will be compared with the known results for general galled trees and simplex  
 128 galled trees [4, 12]. Finally, we will give some concluding remarks in Section 4.

### 2 Semi-Simplex Tree-Child Networks

130 In this section, we are going to prove our results for semi-simplex tree-child networks. We  
 131 start by deriving a closed-form expression for  $\text{SSTC}_{n,k}$  from which the claimed asymptotic  
 132 result in Theorem 4 will follow by a double application of the Laplace method.

133 First, recall the exact counting formula for  $\text{STC}_{n,k}$  from [5].

134 ► **Proposition 8** (Theorem 13 in [5]). *For  $0 \leq k \leq n-1$ , we have*

$$135 \quad \text{STC}_{n,k} = \binom{n}{k} \frac{(2n-2)!}{2^{n-1}(n-k-1)!}.$$

136 ► **Remark 9.** For  $k=0$ , we have

$$137 \quad T_n := \text{STC}_{n,0} = \frac{(2n-2)!}{2^{n-1}(n-1)!} = (2n-3)!! \sim \frac{1}{2\sqrt{\pi}} n^{-3/2} 2^n n! \quad (6)$$

138 which is the (well-known) formula for the number of phylogenetic trees; see [21]. Thus, for  
 139 the exponential generating function (EGF) of  $T_n$ , we obtain that

$$140 \quad T(z) := \sum_{n \geq 0} T_n \frac{z^n}{n!} = 1 - \sqrt{1-2z}. \quad (7)$$

141 Alternatively, this result can be derived by the symbolic method; see Note II.19 in [10].

142 In addition, we need the following lemma for the number of *phylogenetic forests* which  
 143 are leaf-labeled sets of rooted binary trees with label set  $\{1, \dots, n\}$ .

144 ► **Lemma 10.** *The number of phylogenetic forests consisting of  $k$  trees whose leaves are*  
 145 *labeled with labels from the set  $\{1, \dots, n\}$  equals*

$$146 \quad \frac{(2n-k-1)!}{(n-k)!(k-1)!2^{n-k}}$$

147 *provided that  $n \geq k \geq 1$  and 0 otherwise.*

148 **Proof.** The number of such forests is given by  $[z^n]T(z)^k/k!$ . Note that  $T(z) = z + T(z)^2/2$ .  
 149 Thus, by the Lagrange inversion formula:

$$150 \quad n! [z^n] \frac{T(z)^k}{k!} = \frac{(n-1)!}{(k-1)!} [w^{n-k}] (1 - \omega/2)^{-n} = \frac{(n-1)!}{(k-1)! 2^{n-k}} \binom{2n-k-1}{n-k}.$$

151 This is the claimed result. ◀

152 We can now state the following exact formula for  $SSTC_{n,k}$ .

153 **► Proposition 11.** For  $1 \leq k \leq n-1$ ,

$$154 \quad SSTC_{n,k} = 2^{1-n} \sum_{j=1}^{n-k} \binom{n}{j} \frac{(2j+2k-2)!(2n-2j-k-1)!}{(j-1)!(n-j-k)!(k-1)!}$$

155 and  $SSTC_{n,0} = (2n-3)!!$

156 **Proof.** The result for  $k=0$  follows by Remark 9. Thus, we may assume that  $1 \leq k \leq n-1$ .  
 157 According to Proposition 8, there are

$$158 \quad \frac{(2j+2k-2)!}{2^{j+k-1}(j-1)!} \tag{8}$$

159 simplex tree-child networks with  $k$  reticulation nodes and  $j+k$  leaves which are labeled by  
 160  $\{1, \dots, j+k\}$  where the leaves below the  $k$  reticulation nodes are labeled by  $\{1, \dots, k\}$ . Now,  
 161 in order to construct all semi-simplex tree-child networks, we have to replace these leaves by  
 162 a phylogenetic forest consisting of  $k$  trees whose leaves are labeled by  $\{1, \dots, n-j\}$  (so, that  
 163 we obtain  $n$  leaves in total). According to Lemma 10, the number of such forests is given by

$$164 \quad \frac{(2n-2j-k-1)!}{(n-j-k)!(k-1)! 2^{n-j-k}}$$

165 for  $\ell-j \geq k$ . Multiplying this expression with (8), relabeling the leaves, and summing over  
 166  $1 \leq j \leq \ell-k$  gives the claim. ◀

167 Consequently,

$$168 \quad SSTC_n = (2n-3)!! + 2^{1-n} \sum_{k=1}^{n-1} \sum_{j=1}^{n-k} \binom{n}{j} \frac{(2j+2k-2)!(2n-2j-k-1)!}{(j-1)!(n-j-k)!(k-1)!} \tag{9}$$

$$169 \quad = (2n-3)!! + 2^{1-n} \sum_{j=1}^{n-1} \binom{n}{j} \frac{1}{(j-1)!} \sum_{k=1}^{n-j} \frac{(2j+2k-2)!(2n-2j-k-1)!}{(n-j-k)!(k-1)!}. \tag{10}$$

170 The terms in the inner sum of (9) are unimodal with the maximum hard to locate. However,  
 171 rather surprisingly, the terms in the inner sum of (10) are monotonic.

172 **► Lemma 12.** For fixed  $n$  and  $1 \leq j \leq n-1$ , the sequence

$$173 \quad \left\{ \frac{(2j+2k-2)!(2n-2j-k-1)!}{(n-j-k)!(k-1)!} \right\}_{1 \leq k \leq n-j}$$

174 is increasing.

## 27:6 Semi-Simplex Tree-Child Networks and Galled Trees

175 **Proof.** The fraction of two consecutive terms of the sequence is given by:

$$176 \quad \frac{2(2j+2k-1)(j+k)(n-j-k)}{(2n-2j-k-1)k} = \frac{2j+2k-1}{k} \cdot \frac{2(j+k)(n-j-k)}{2n-2j-k-1}.$$

177 The first term on the right-hand side of this expression is clearly  $\geq 1$ . For the second term,  
178 we have

$$179 \quad \frac{2(j+k)(n-j-k)}{2n-2j-k-1} \geq \frac{2(j+k)(n-j-k)}{2n-j-k-1}.$$

180 Set  $x := j+k$  and consider the function

$$181 \quad f(x) := \frac{2x(n-x)}{2n-x-1}$$

182 with  $1 \leq x \leq n-1$ . This function is easily seen to be first increasing and then decreasing.  
183 Thus, to show that it is  $\geq 1$  (which implies that the second term above is also  $\geq 1$ ), we only  
184 need to evaluate it at the two boundary points. At  $x=1$ , we have  $f(1)=1$ . At  $x=n-1$ ,  
185 we have  $f(n-1)=(2n-1)/n$  which clearly is  $\geq 1$ . Thus, the claim is proved.  $\blacktriangleleft$

186 We derive now Theorem 4 from (10) in two steps: first we treat the inner sum and then  
187 the outer sum, where both times we apply the Laplace method.

188  $\blacktriangleright$  **Remark 13.** It would be slightly harder to apply the same strategy of proof to (9) due to  
189 the more involved behavior of the terms of the inner sum.

190  $\blacktriangleright$  **Lemma 14.** For  $j = \sqrt{n} + x$ , we have

$$191 \quad \sum_{k=1}^{n-j} \frac{(2j+2k-2)!(2n-2j-k-1)!}{(n-j-k)!(k-1)!} \sim \frac{\sqrt[4]{e}}{4\sqrt{\pi}} n^{-5/2} 4^n n!^2, \quad (n \rightarrow \infty).$$

192 uniformly for  $x = \mathcal{O}(n^{1/4+\epsilon})$  with  $\epsilon > 0$ .

193 **Proof.** Call the sum in the lemma  $\Sigma(n)$ . Due to Lemma 12, the terms close to  $n-j$   
194 contribute most. Thus, we first change the order of summation:

$$195 \quad \Sigma(n) = \sum_{k=0}^{n-j-1} \frac{(2n-2k-2)!(n-j+k-1)!}{k!(n-j-k-1)!}.$$

196 Next, we set  $j = \sqrt{n} + x$  and expand the terms inside the sum which gives

$$197 \quad \frac{(2n-2k-2)!(n-j+k-1)!}{k!(n-j-k-1)!} \sim \frac{1}{\sqrt{\pi} k! 4^{k+1}} n^{-5/2} 4^n n!^2 \quad (11)$$

198 uniformly for  $x = \mathcal{O}(n^{1/4+\epsilon})$  and  $k = \mathcal{O}(\sqrt[4]{n})$ . We now break  $\Sigma(n)$  into two parts:

$$199 \quad \Sigma(n) = \sum_{k < \sqrt[4]{n}} + \sum_{k \geq \sqrt[4]{n}} =: \Sigma_1(n) + \Sigma_2(n).$$

200 For the second part, by (11) and Lemma 12:

$$201 \quad \Sigma_2(n) = \mathcal{O}\left(\frac{1}{[\sqrt[4]{n}]! 4^{\sqrt[4]{n}}} n^{-3/2} 4^n n!^2\right)$$

202 which is much smaller than the contribution of  $\Sigma_1(n)$  which again by (11) is given by:

$$203 \quad \Sigma_1(n) \sim \frac{1}{4\sqrt{\pi}} \left( \sum_{k < \sqrt[4]{n}} \frac{1}{k! 4^k} \right) n^{-5/2} 4^n n!^2 \sim \frac{\sqrt[4]{e}}{4\sqrt{\pi}} n^{-5/2} 4^n n!^2$$

204 uniformly for  $x = \mathcal{O}(n^{1/4+\epsilon})$ .  $\blacktriangleleft$

205 We are now ready to proof Theorem 4.

206 **Proof of Theorem 4.** Set:

$$207 \quad \Sigma(n) := \sum_{j=1}^{n-1} \binom{n}{j} \frac{2^{1-n}}{(j-1)!} \sum_{k=1}^{n-j} \frac{(2j+2k-2)!(2n-2j-k-1)!}{(n-j-k)!(k-1)!}$$

208 which is the second term of (10). By Lemma 12, we have

$$209 \quad \Sigma(n) \leq \sum_{j=1}^{n-1} \binom{n}{j} \frac{2^{1-n}(2n-2)!(n-j)}{(j-1)!}.$$

210 The terms of this sum are increasing for  $j < \sqrt{n} - 1$  and decreasing otherwise. Moreover, for  
211  $j = \sqrt{n} + x$ ,

$$212 \quad \binom{n}{j} \frac{2^{1-n}(2n-2)!(n-j)}{(j-1)!} \sim \frac{1}{4\pi\sqrt{\pi e}} n^{-3/2} e^{2\sqrt{n}} 2^n n!^2 e^{-x^2/\sqrt{n}}, \quad (n \rightarrow \infty) \quad (12)$$

213 uniformly for  $x = \mathcal{O}(n^{1/4+\epsilon})$  with  $\epsilon > 0$ . Now, as in the proof of the previous lemma, we  
214 again break  $\Sigma(n)$  into two parts:

$$215 \quad \Sigma(n) = \sum_{|j-\sqrt{n}| < n^{1/4+\epsilon}} + \sum_{|j-\sqrt{n}| \geq n^{1/4+\epsilon}} =: \Sigma_1(n) + \Sigma_2(n).$$

216 For the second part, by the above estimate and (12):

$$217 \quad \Sigma_2(n) = \mathcal{O}\left(n^{-1/2} e^{2\sqrt{n}-n^{2\epsilon}} 2^n n!^2\right).$$

218 For the first part, by using Lemma 14 and again (12):

$$219 \quad \Sigma_1(n) \sim \frac{1}{16\pi^2 \sqrt[4]{e}(2n-2)!n} n^{-4} e^{2\sqrt{n}} 8^n n!^4 \left( \sum_{|j-\sqrt{n}| < n^{1/4+\epsilon}} e^{-x^2/\sqrt{n}} \right)$$

$$220 \quad \sim \frac{1}{4\pi\sqrt{\pi}\sqrt[4]{e}} n^{-5/2} e^{2\sqrt{n}} 2^n n!^2 \int_{-\infty}^{\infty} e^{-x^2/\sqrt{n}} dx \sim \frac{1}{4\pi\sqrt[4]{e}} n^{-9/4} e^{2\sqrt{n}} 2^n n!^2.$$

221 Combining this with the above estimate for  $\Sigma_2(n)$  and noting that  $\text{SSTC}_n \sim \Sigma(n)$  since  
222  $(2n-3)!!$  has a smaller asymptotic order (see (6)) gives the claimed result. ◀

223 We next prove Theorem 6. For this, we return to Proposition 11 whose terms, as  
224 mentioned above, are in general unimodal. However, in the major range of  $k$ , we again have  
225 monotonicity.

226 ▶ **Lemma 15.** For sufficiently large  $n$  and  $n - 2\sqrt{n} \leq k \leq n - 1$ , the sequence

$$227 \quad \left\{ \binom{n}{j} \frac{(2j+2k-2)!(2n-2j-k-1)!}{(j-1)!(n-j-k)!(k-1)!} \right\}_{1 \leq j \leq n-k}$$

228 is increasing.

229 **Proof.** The fraction of two consecutive terms of the sequence is given by

$$230 \quad \frac{(2j+2k)(2j+2k-1)(n-j)(n-j-k)}{j(j+1)(2n-2j-k-1)(2n-2j-k-2)} \quad (13)$$

## 27:8 Semi-Simplex Tree-Child Networks and Galled Trees

231 and we have to show that this is  $\geq 1$  for  $1 \leq j \leq n - k - 1$ . Replace  $j = x$  and let  $f(x)$  be  
 232 the logarithm of the above expression on  $(0, n - k)$ . Taking derivatives gives:

$$233 \quad f'(x) = \frac{2}{2x + 2k} + \frac{2}{2x + 2k - 1} + \frac{2}{2n - 2x - k - 1} + \frac{2}{2n - 2x - k - 2}$$

$$234 \quad - \frac{1}{x} - \frac{1}{x + 1} - \frac{1}{n - x} - \frac{1}{n - x - k}.$$

235 Note that

$$236 \quad 2x + 2k \geq 2x + 2k - 1 \geq 2k - 1 \geq 2n - 4\sqrt{n} - 1$$

237 and

$$238 \quad 2n - 2x - k - 1 \geq 2n - 2x - k - 2 \geq k - 2 \geq n - 2\sqrt{n} - 2.$$

239 Thus,

$$240 \quad f'(x) \leq \frac{8}{n - 2\sqrt{n} - 2} - \frac{1}{x} \leq \frac{8}{n - 2\sqrt{n} - 2} - \frac{1}{n - k} \leq \frac{8}{n - 2\sqrt{n} - 2} - \frac{1}{2\sqrt{n}}$$

241 which is negative for sufficiently large  $n$ . This implies that  $f(x)$  is decreasing and hence we  
 242 have only to show that (13) is  $\geq 1$  for  $j = n - k - 1$ . In this case (13) becomes

$$243 \quad \frac{(2n - 2)(2n - 3)}{(n - k - 1)(n - k)k} \geq \frac{2n - 3}{(2\sqrt{n} - 1)\sqrt{n}} = 1 + \frac{1}{2\sqrt{n}} + \mathcal{O}\left(\frac{1}{n}\right)$$

244 which proves the claim for sufficiently large  $n$ .  $\blacktriangleleft$

245 Using this lemma, we can prove the following.

246 **► Proposition 16.** For  $k = n - \sqrt{n} + x$ , we have

$$247 \quad \text{SSTC}_{n,k} \sim \frac{1}{4\pi\sqrt{\pi}\sqrt[4]{e}} n^{-5/2} e^{2\sqrt{n}} 2^n n!^2 e^{-x^2/\sqrt{n}}, \quad (n \rightarrow \infty)$$

248 uniformly for  $x = \mathcal{O}(n^{1/4+\epsilon})$  with  $\epsilon > 0$ .

249 **Proof.** By the same arguments as in the proof of Lemma 14, for  $k = n - \sqrt{n} + x$

$$250 \quad \sum_{j=1}^{n-k} \binom{n}{j} \frac{(2j + 2k - 2)!(2n - 2j - k - 1)!}{(j - 1)!(n - j - k)!(k - 1)!} \sim \frac{1}{8\pi\sqrt{\pi}\sqrt[4]{e}} n^{-3} e^{2\sqrt{n}} 4^n n!^2 e^{-x^2/\sqrt{n}}, \quad (n \rightarrow \infty)$$

251 uniformly for  $x = \mathcal{O}(n^{1/4+\epsilon})$  with  $\epsilon > 0$ . Multiplying by  $2^{1-n}$  gives the claimed result.  $\blacktriangleleft$

252 Now, we are ready to proof our second main result.

253 **Proof of Theorem 6.** Combining Proposition 16 with Theorem 4 gives the following local  
 254 limit theorem for  $SSX_n$ :

$$255 \quad \mathbb{P}(SSX_n = n - \sqrt{n} + x) \sim \frac{1}{\sqrt{\pi}\sqrt[4]{n}} e^{-x^2/\sqrt{n}}, \quad (n \rightarrow \infty)$$

256 uniformly for  $x = \mathcal{O}(n^{1/4+\epsilon})$  with  $\epsilon > 0$  where  $n - \sqrt{n} + x$  is an integer. This is stronger  
 257 than the claimed central limit theorem from Theorem 6 which follows from it by standard  
 258 arguments.  $\blacktriangleleft$

259 What is left is the proof of Theorem 7. Here, we will work with the *depth* of a random  
 260 tree-child network instead of the Sackin index. The depth, a standard parameter in computer  
 261 science (see [19]), is defined as the height of a random leaf. Let  $SD_n$  and  $SSD_n$  denote the  
 262 depth of a random simplex tree-child networks and random semi-simplex tree-child networks  
 263 of size  $n$ , respectively. Then,

$$264 \quad \mathbb{E}(SSS_n) = n\mathbb{E}(SSD_n), \quad \mathbb{E}(SS_n) = n\mathbb{E}(SD_n). \quad (14)$$

265 Note that from the latter and the result of [25], we obtain that

$$266 \quad \mathbb{E}(SD_n) = \Theta(n^{3/4}), \quad (n \rightarrow \infty). \quad (15)$$

267 In addition, the mean value of the depth of a random phylogenetic tree is known as well.

268 ► **Theorem 17** (Theorem 4 in [3]). *Let  $D_n$  denote depth of a phylogenetic tree picked uniformly*  
 269 *at random from the class of all phylogenetic trees of size  $n$ . Then,*

$$270 \quad \mathbb{E}(D_n) \sim \sqrt{\pi n}, \quad (n \rightarrow \infty).$$

271 ► **Remark 18.** In [3], the above result was formulated for the Sackin index  $S_n$  of random  
 272 phylogenetic trees of size  $n$ . However, as above, we have the relation:

$$273 \quad \mathbb{E}(S_n) = n\mathbb{E}(D_n).$$

274 Apart from these tools, we need another ingredient in the proof of Theorem 7 which will  
 275 be stated in the below sketch of proof (and rigorously shown in the journal version of the  
 276 paper).

277 **Sketch of Proof of Theorem 7.** First, denote by  $E_n$  the event that a random semi-simplex  
 278 tree-child network is a simplex tree-child network. Then, by Remark 5:

$$279 \quad \mathbb{P}(E_n) = \frac{\text{STC}_n}{\text{SSTC}_n} \rightarrow e^{-1/4}, \quad (n \rightarrow \infty).$$

280 Consequently,

$$281 \quad \mathbb{E}(SSD_n) \geq \mathbb{E}(SSD_n|E_n)\mathbb{P}(E_n) = \mathbb{E}(SD_n)\mathbb{P}(E_n) = \Omega(n^{3/4}),$$

282 where the last line follows from (15). Thus, we only need to prove a corresponding upper  
 283 bound. For this, recall that  $SSX_n$  denotes the number of reticulation nodes of a random  
 284 semi-simplex tree-child network of size  $n$ . In addition, given that  $SSX_n = k$ , denote by  $SSY_n$   
 285 the vector of sizes of the  $k$  subtrees below the reticulation nodes and by  $SSZ_n$  the number  
 286 of leaves which are not below reticulation nodes. Thus, if  $SSX_n = k$ ,  $SSY_n = (s_1, \dots, s_k)$ ,  
 287 and  $SSZ_n = j$ , then  $s_1 + \dots + s_k = n - j$  and  $1 \leq j \leq n - k$ . Now, we have

$$288 \quad \mathbb{E}(SSD_n) = \sum_{j,k,s_1,\dots,s_k} \mathbb{E}(SSD_n|SSX_n = k, SSY_n = (s_1, \dots, s_k), SSZ_n = j) \\ 289 \quad \times \mathbb{P}(SSX_n = k, SSY_n = (s_1, \dots, s_k), SSZ_n = j). \quad (16)$$

290 Then, since a random semi-simplex tree-child network with  $SSX_n = k$ ,  $SSY_n = (s_1, \dots, s_k)$ ,  
 291 and  $SSZ_n = j$  consists of a simplex tree-child network of size  $j + k$  with trees of sizes  
 292  $s_1, \dots, s_k$  below the  $k$  reticulation nodes, we have

$$293 \quad \mathbb{E}(SSD_n|SSX_n = k, SSY_n = (s_1, \dots, s_k), SSZ_n = j) \\ 294 \quad = \max_{1 \leq i \leq k} \{s_i\} \mathcal{O}((j+k)^{3/4}) + \mathcal{O}((n-j)^{1/2}) = \max_{1 \leq i \leq k} \{s_i\} \mathcal{O}(n^{3/4}),$$

## 27:10 Semi-Simplex Tree-Child Networks and Galled Trees

295 where the first  $\mathcal{O}$ -term above is due to (15) and the second follows from Theorem 17. Plugging  
 296 this into (16) gives:

$$\begin{aligned}
 297 \quad \mathbb{E}(SSD_n) &= \mathcal{O}(n^{3/4}) \sum_{j,k,s_1,\dots,s_k} \max_{1 \leq i \leq k} \{s_i\} \mathbb{P}(SSX_n = k, SSY_n = (s_1, \dots, s_k), SSZ_n = j) \\
 298 \quad &= \mathcal{O}(n^{3/4}),
 \end{aligned}$$

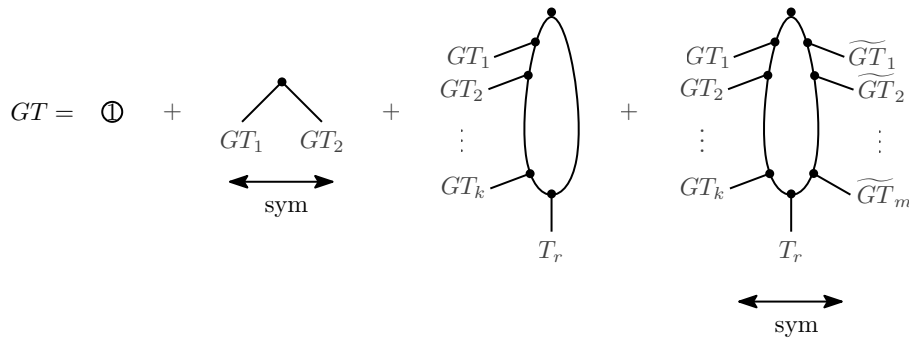
299 where the last line follows from

$$300 \quad \sum_{k,s_1,\dots,s_k} \max_{1 \leq i \leq k} \{s_i\} \mathbb{P}(SSX_n = k, SSY_n = (s_1, \dots, s_k)) = \mathcal{O}(1),$$

301 i.e., the maximal size of a subtree below a reticulation node in a random semi-simplex  
 302 tree-child networks is bounded. (This will be proved in the journal version of the paper.)  
 303 Overall,

$$304 \quad \mathbb{E}(SSD_n) = \Theta(n^{3/4}), \quad (n \rightarrow \infty)$$

305 which together with (14) implies the claimed result. ◀



■ **Figure 1** Specification of semi-simplex galled trees. Note that if the root is in a tree cycle (cases three and four on the right-hand side of the equality), then the reticulation node is followed by a tree  $T$ ; in all other cases galled trees  $GT$  are attached as children.

### 306 **3 Semi-Simplex Galled Trees**

307 In this section, we derive corresponding results for galled trees. Note that galled trees  
 308 are easier to handle than tree-child networks as they possess a straightforward recursive  
 309 decomposition as follows. The root is either not contained in a tree cycle in which case it has  
 310 two subtrees which are again galled trees (second case on the right-hand side of the equality  
 311 in Figure 1) or it is contained in a tree cycle (third and fourth case in Figure 1 where the  
 312 sequences of galled trees on the left and right of the tree cycle are non-empty). In addition,  
 313 note that we have a symmetry along the vertical line through the root since swapping the  
 314 galled tree along this line just yields the same tree (except in the third case in Figure 1).

315 Due to this decomposition, we can now use exponential generating functions (EGFs) to  
 316 enumerate semi-simplex galled trees as well as study parameters of random semi-simplex  
 317 galled trees. We start with the enumeration problem.

318 ▶ **Lemma 19.** Let  $\text{SSGT}_n$  denote the number of semi-simplex galled trees of size  $n$  with  
 319 EGF:

$$320 \quad A(z) := \sum_{n \geq 1} \text{SSGT}_n \frac{z^n}{n!}.$$

321 Then,

$$322 \quad A(z) = z + \frac{A(z)^2}{2} + T(z) \frac{A(z)}{1 - A(z)} + \frac{T(z)}{2} \frac{A(z)^2}{(1 - A(z))^2}, \tag{17}$$

323 where  $T(z)$  denotes the EGF of the number phylogenetic trees; see (7).

324 **Proof.** From the symbolic method (see Chapter I in [10]), we immediately obtain (17) since  
 325  $A(z)/(1 - A(z))$  is the generating function of a non-empty sequence of galled trees and every  
 326 reticulation node is followed by a tree (which explains the factor  $T(z)$  in the third and fourth  
 327 term on the right-hand side of equation (17)). Also, note that the factor  $1/2$  in the second  
 328 and fourth term on the right-hand side of equation (17) arises from the symmetry explained  
 329 in the paragraph prior to the lemma (and is indicated by horizontal arrows in Figure 1). ◀

330 The equation for  $A(z)$  can be solved and the asymptotics of the coefficient derived with  
 331 singularity analysis (see Chapter VI in [10]).

332 ▶ **Theorem 20.** The number  $\text{SSGT}_n$  of semi-simplex galled trees of size  $n$  behaves asymptot-  
 333 ically as

$$334 \quad \text{SSGT}_n \sim \frac{1}{4} \sqrt{\frac{2\alpha(5 - 4\alpha) + 4\alpha(3\alpha - 1)(1 - 2\alpha)^{-1/2}}{(4 - 2\sqrt{1 - 2\alpha} - 4\alpha)\pi}} n^{-3/2} \alpha^{-n} n!, \quad (n \rightarrow \infty), \tag{18}$$

335 where  $\alpha = (1 - \beta^2)/2 \approx 0.16486$  and  $\beta \approx 0.81871$  is the smallest positive real root of the  
 336 equation  $z^4 - 2z^3 + 3z^2 + 2z - 3 = 0$ .

337 **Proof.** Solving (17) gives:

$$338 \quad A(z) = 1 - \frac{1}{2} \sqrt{4 + 2\sqrt{1 - 10z + 4z\sqrt{1 - 2z} + 4z^2} - 2\sqrt{1 - 2z} - 4z}.$$

339 We next have to find the singularities with smallest absolute value (the *dominant singularities*)  
 340 of  $A(z)$ . It is easy to see that there is only one that is obtained by solving  $1 - 10z + 4z\sqrt{1 - 2z} +$   
 341  $4z^2 = 0$  with  $z \in \mathbb{R}$ . Set  $\sqrt{1 - 2\alpha} =: \beta$  which gives the equation  $\beta^4 - 2\beta^3 + 3\beta^2 + 2\beta - 3 = 0$   
 342 that can be solved by Ferrari's method:

$$343 \quad \gamma_1 = \frac{\sqrt[3]{89 + 6\sqrt{159}} + \sqrt[3]{89 - 6\sqrt{159}} - 1}{6},$$

$$344 \quad \gamma_2 = \frac{\sqrt{2\gamma_1 + 1} - \sqrt{1 - 2\gamma_1 + 4\sqrt{\gamma_1^2 - 1}}}{2},$$

$$345 \quad \beta = 1 - \gamma_2^2$$

$$346 \quad = 1 - \left( \frac{1}{6} \sqrt{3(89 + 6\sqrt{159})^{1/3} + 3(89 - 6\sqrt{159})^{1/3} + 6} - \frac{1}{6} \left( 12 - 3(89 + 6\sqrt{159})^{1/3} \right. \right.$$

$$347 \quad \left. \left. - 3(89 - 6\sqrt{159})^{1/3} + 36 \sqrt{\left( \frac{1}{6} (89 + 6\sqrt{159})^{1/3} + \frac{1}{6} (89 - 6\sqrt{159})^{1/3} - \frac{1}{6} \right)^2 - 1} \right)^{1/2} \right)^2$$

$$348 \quad \alpha = (1 - \beta^2)/2,$$

## 27:12 Semi-Simplex Tree-Child Networks and Galled Trees

349 where  $\alpha \approx 0.16486$ . Expanding  $A(z)$  as  $z \rightarrow \alpha^-$  gives:

$$350 \quad A(z) \sim 1 - \frac{1}{2}\sqrt{4 - 2\sqrt{1 - 2\alpha} - 4\alpha} - \frac{1}{2}\sqrt{\frac{2\alpha(5 - 4\alpha) + 4\alpha(3\alpha - 1)(1 - 2\alpha)^{-1/2}}{4 - 2\sqrt{1 - 2\alpha} - 4\alpha}}\sqrt{1 - \frac{z}{\alpha}}.$$

351 Consequently, by the transfer theorems of singularity analysis

$$352 \quad \text{SSGT}_n = n![z^n]A(z) \sim -\frac{1}{2\Gamma(-1/2)}\sqrt{\frac{2\alpha(5 - 4\alpha) + 4\alpha(3\alpha - 1)(1 - 2\alpha)^{-1/2}}{4 - 2\sqrt{1 - 2\alpha} - 4\alpha}}n^{-3/2}\alpha^{-n}n! \\ 353 \quad \sim \frac{1}{4}\sqrt{\frac{2\alpha(5 - 4\alpha) + 4\alpha(3\alpha - 1)(1 - 2\alpha)^{-1/2}}{(4 - 2\sqrt{1 - 2\alpha} - 4\alpha)\pi}}n^{-3/2}\alpha^{-n}n! \quad (19)$$

354 which is the claimed result.  $\blacktriangleleft$

355 **► Remark 21.** In contrast to tree-child networks, the asymptotics of the numbers of semi-  
356 simplex galled trees (18) and simplex galled trees (4) differ already in the exponential growth  
357 term. The former is  $\alpha^{-n} \approx 6.06574^n$  and the latter  $(3 + 2\sqrt{2})^n \approx 5.82843^n$ . Also, note that  
358 both terms are much smaller than the exponential growth term of general galled trees in (5).

359 In order to study the number of reticulation nodes of galled trees, we use bivariate  
360 generating functions.

361 **► Lemma 22.** Let  $\text{SSGT}_{n,k}$  denote number of semi-simplex galled trees of size  $n$  and  $k$   
362 reticulation nodes with bivariate generating function:

$$363 \quad A(z, u) := \sum_{n,k} \text{SSGT}_{n,k} \frac{z^n}{n!} u^k.$$

364 Then,

$$365 \quad A(z, u) = z + \frac{A(z, u)^2}{2} + uT(z)\frac{A(z, u)}{1 - A(z, u)} + \frac{uT(z)}{2}\frac{A(z, u)^2}{(1 - A(z, u))^2}, \quad (20)$$

366 where  $T(z)$  denotes the EGF of the number of phylogenetic trees; see (7).

367 **Proof.** The proof is similar to the proof of Lemma 19. More precisely, note that the number  
368 of reticulation nodes is an *additive* parameter, i.e., it can be computed by simply adding up  
369 contributions from subtrees. Consequently, the second term on the right-hand side of (20),  
370 which corresponds to the second case in the specification in Figure 1, is just  $A(z, u)^2/2$  and  
371 similarly the other two terms are explained where the additional factor  $u$  takes into account  
372 the reticulation node in the tree cycle of the root in the third and fourth case of Figure 1.  $\blacktriangleleft$

373 We can now apply the following theorem from [8] to prove a central limit theorem for the  
374 number of reticulation nodes of a random galled tree.

375 **► Theorem 23** (Theorem 2.23 in [8]). Assume that  $F(z, y, u)$  is an analytic function at  
376  $(z, y, u) = (0, 0, 1)$  with  $F(0, y, u) \equiv 0$ ,  $F(z, 0, u) \not\equiv 0$  and all Maclaurin coefficients of  
377  $F(z, y, 1)$  are real and non-negative. In addition, assume that the region of convergence of  
378  $F(z, y, u)$  contains a non-negative solution  $(z, y) = (z_0, y_0)$  of the system

$$379 \quad y = F(z, y, 1) \quad \text{and} \quad 1 = F_y(z, y, 1). \quad (21)$$

380 which satisfies  $F_z(z_0, y_0, 1) \neq 0$  and  $F_{yy}(z_0, y_0, 1) \neq 0$ . Finally, assume that the Maclaurin  
381 coefficients of the (unique) solution  $y(z, u)$  with non-negative Maclaurin coefficients of

382  $y = F(z, y, u)$  with  $y(0, u) = 0$  are eventually positive. Then, the random variable  $X_n$  with  
 383 probability generating function

$$384 \quad \mathbb{E}(u^{X_n}) = \frac{[z^n]y(z, u)}{[z^n]y(z, 1)}$$

385 satisfies

$$386 \quad \mathbb{E}(X_n) \sim \mu n \quad \text{and} \quad \text{Var}(X_n) \sim \sigma^2 n$$

387 with

$$388 \quad \begin{aligned} \mu &= \frac{F_u}{z_0 F_z} \\ \sigma^2 &= \mu + \mu^2 + \frac{1}{z_0 F_z^3 F_{yy}} \left( F_z^2 (F_{yy} F_{uu} - F_{yu}^2) - 2F_z F_u (F_{yy} F_{zu} - F_{yz} F_{yu}) \right. \\ &\quad \left. + F_u^2 (F_{zz} F_{yy} - F_{yz}^2) \right), \end{aligned} \tag{22}$$

389 where all derivatives are evaluated at  $(z, y, u) = (z_0, y_0, 1)$ . Moreover, if  $\sigma^2 > 0$ , then

$$390 \quad \frac{X_n - \mu n}{\sigma \sqrt{n}} \xrightarrow{d} N(0, 1), \quad (n \rightarrow \infty).$$

391 ► **Theorem 24.** The number  $SSX_n$  of reticulation nodes of a semi-simplex galled tree picked  
 392 uniformly at random from the class of all semi-simplex galled trees of size  $n$  satisfies

$$393 \quad \frac{SSX_n - \mu_R n}{\sigma_R \sqrt{n}} \xrightarrow{d} N(0, 1), \quad (n \rightarrow \infty),$$

394 where  $(\mu_R, \sigma_R^2) \approx (0.40662, 0.11657)$ .

395 **Proof.** Let

$$396 \quad F(z, y, u) = \frac{z}{1 - y/2 - uT(z)/(1 - y) - uT(z)y/(2(1 - y)^2)}.$$

397 Then, because of (20),  $A(z, y, u)$  is a solution of  $y = F(z, y, u)$ . Next, note that all the  
 398 assumptions of Theorem 23 are satisfied, e.g.,  $z_0 = \alpha$  and  $y_0 = A(z_0)$  is a solution of (21)  
 399 lying inside the domain of convergence of  $F(z, y, u)$ . Thus, we obtain the claimed result  
 400 with the values for the constants of mean and variance which are computed from (22) by  
 401 straightforward computation. ◀

402 ► **Remark 25.** A similar central limit theorem has been proved for the number of reticulation  
 403 nodes of random general galled trees [4] with the corresponding constants for mean value  
 404 and variance approximate equal to  $(\mu_R, \sigma_R^2) \approx (0.56155, 0.17736)$ . Also, by a variation of the  
 405 method, a similar result can also be proved for the number of reticulation nodes of random  
 406 simplex galled trees where  $(\mu_R, \sigma_R^2) \approx (0.41421, 0.12132)$ . Note that from this, we see that  
 407 on average, semi-simplex galled trees have the smallest number of reticulation nodes among  
 408 these three classes of networks.

409 Next, for the Sackin index, we again use bivariate generating functions. We denote  
 410 the number of phylogenetic trees of size  $n$  and Sackin index equal to  $k$  by  $ST_{n,k}$  and  
 411  $T(z, u) := \sum_{n,k} ST_{n,k} \frac{z^n}{n!} u^k$ .

## 27:14 Semi-Simplex Tree-Child Networks and Galled Trees

412 ► **Lemma 26.** Let  $SSS_{n,k}$  be the number of semi-simplex galled trees of size  $n$  and Sackin  
 413 index equal to  $k$  with bivariate generating function:

$$414 \quad B(z, u) := \sum_{n,k} SSS_{n,k} \frac{z^n}{n!} u^k.$$

415 Then,

$$416 \quad B(z, u) = z + \frac{1}{2} B(zu, u)^2 + \sum_{k>0} T(zu^{k+2}, u) \prod_{i=1}^k B(zu^{i+1}, u)$$

$$417 \quad + \sum_{m \geq 1} \sum_{k>m} T(zu^{k+2}, u) \prod_{i=1}^k B(zu^{i+1}, u) \prod_{j=1}^m B(zu^{j+1}, u)$$

$$418 \quad + \frac{1}{2} \sum_{k \geq 1} T(zu^{k+2}, u) \prod_{i=1}^k B(zu^{i+1}, u)^2. \quad (23)$$

419 **Proof.** The proof again follows from the symbolic method where the second term is explained  
 420 by additivity of the Sackin index, i.e., it can be computed for subtrees separately and then  
 421 added up where one has to add the root edge for each leaf (this explains why the first  
 422 argument is  $zu$  instead of  $z$ ). Similarly the third and last two terms are explained which  
 423 correspond to the third and fourth case on the right-hand side of the specification in Figure 1.  
 424 Here, note that for the tree below the reticulation node we have to choose the longer path  
 425 from the root (which explains the term  $zu^{k+2}$ ). In addition, we have splitted the fourth case  
 426 in Figure 1 into two subcases according to whether  $k > m$  or  $k = m$ ; only in the second  
 427 subcase, we need the additional factor  $1/2$  due to symmetry. ◀

428 From this result, we obtain the following.

429 ► **Theorem 27.** Let  $SSS_n$  be the Sackin index of a semi-simplex galled tree picked uniformly  
 430 at random from the class of all semi-simplex galled trees of size  $n$ . Then,

$$431 \quad \mathbb{E}(SSS_n) \sim \sqrt{\pi} \mu_S n^{3/2}, \quad (n \rightarrow \infty),$$

432 where

$$433 \quad \mu_S = \sqrt{\frac{4 - 2\sqrt{1 - 2\alpha} - 4\alpha}{2\alpha(5 - 4\alpha) + 4\alpha(3\alpha - 1)(1 - 2\alpha)^{-1/2}}} \approx 1.28962.$$

434 **Proof.** Let  $S(z) := \frac{\partial}{\partial u} B(z, u)|_{u=1}$ . Then,

$$435 \quad \mathbb{E}(SSS_n) = \frac{[z^n] S(z)}{[z^n] A(z)},$$

436 where the asymptotics of the denominator is known by (19). Consequently, we only have  
 437 to find the asymptotics of the numerator. For this purpose, note that  $\frac{\partial}{\partial u} B(zu^i, u)|_{u=1} =$   
 438  $izA'(z) + S(z)$ . Thus, by taking partial derivatives on both sides of equation (23) with  
 439 respect to  $u$  and letting  $u = 1$ , we obtain

$$440 \quad S(z) = \frac{zT'(z)f_1(A(z))}{1 - A(z) - T(z)g(A(z))} + \frac{zA'(z)f_2(A(z))}{1 - A(z) - T(z)g(A(z))}$$

$$441 \quad + \frac{zA'(z)T(z)f_3(A(z))}{1 - A(z) - T(z)g(A(z))} + \frac{\frac{\partial}{\partial u} T(z, u)|_{u=1} f_4(A(z))}{1 - A(z) - T(z)g(A(z))},$$

442 where  $g(t) = 1/(1-t)^3$  and

$$443 \quad f_1(t) = \frac{t(2t^3 - 4t^2 - t + 6)}{2(1-t)^2(1-t^2)}, \quad f_2(t) = \frac{t(t^4 - 4t^3 + 6t^2 - 4t + 1)}{(1-t)^4},$$

$$444 \quad f_3(t) = \frac{2-t}{(1-t)^4}, \quad f_4(t) = \frac{t(2-t)}{2(1-t)^2}.$$

445 Also, with the same argument as in the proof of Lemma 26,

$$446 \quad T(z, u) = z + \frac{1}{2}T(zu, u)^2.$$

447 Consequently, by again taking partial derivatives

$$448 \quad \frac{\partial}{\partial u}T(z, u)|_{u=1} = \frac{zT'(z)T(z)}{1-T(z)}$$

449 which can be plugged into the above expression for  $S(z)$ . Next, we have to find the dominant  
 450 singularities of  $S(z)$ . Therefore, note that  $1 - A(\alpha) - T(\alpha)g(A(\alpha)) = 0$  and  $1 - A(z) -$   
 451  $T(z)g(A(z)) \neq 0$ , when  $0 \leq z < \alpha$ . Also, we have  $A(\alpha) = 1 - \frac{1}{2}\sqrt{4 - 2\sqrt{1 - 2\alpha} - 4\alpha} \approx$   
 452  $0.34748$ , and  $f_i(t) \neq 0$ , for all  $i = 1, 2, 3, 4$ , when  $0 < t < 0.35$ . Thus, the dominant singularity  
 453 of  $S(z)$  is unique and again given by  $\alpha$ . What is left is to expand  $S(z)$  as  $z \rightarrow \alpha^-$ . This is  
 454 best done with the assistance of a computer algebra program such as Maple. We obtain:

$$455 \quad S(z) \sim \frac{1}{4(1-z/\alpha)}.$$

456 Thus, by the transfer theorems of singularity analysis,

$$457 \quad [z^n]S(z) \sim 4^{-1}\alpha^{-n}$$

458 and consequently

$$459 \quad \mathbb{E}(SSS_n) = \frac{[z^n]S(z)}{[z^n]A(z)} \sim \sqrt{\frac{(4 - 2\sqrt{1 - 2\alpha} - 4\alpha)\pi}{2\alpha(5 - 4\alpha) + 4\alpha(3\alpha - 1)(1 - 2\alpha)^{-1/2}}} n^{3/2}.$$

460 Numerical evaluation of the constant gives the claimed result. ◀

461 ▶ **Remark 28.** (i) Similar results have been proved in [12] for the Sackin index of random  
 462 general galled trees and simplex galled trees for which we have  $\mu_S \approx 1.76434$  in the  
 463 former case and  $\mu_S \approx 1.30656$  in the latter case, respectively. Again, the mean constant  
 464 is smallest for semi-simplex galled trees; compare with Remark 25.

465 (ii) Using the method from [12], the first-order asymptotics of higher moments of  $SSS_n$   
 466 can be computed as well. Moreover, we expect that  $SSS_n/(\mu_S n^{3/2})$  again converges  
 467 weakly to the Airy distribution; see Theorem 2 in [12].

468 Finally, one wonders if again a relation similar to (1) holds for the number  $\text{SGT}_{n,k}$  of  
 469 simplex galled trees,  $\text{SSGT}_{n,k}$  of semi-simplex galled trees, and  $\text{GT}_{n,k}$  of general galled trees  
 470 of size  $n$  and  $k$  reticulation nodes? In particular, is again almost every general galled tree  
 471 with a fixed number of reticulation nodes a semi-simplex galled tree? We expect that this is  
 472 true and that we again have the following chain of equivalences

$$473 \quad 2^k \text{SGT}_{n,k} \sim \text{SSGT}_{n,k} \sim \text{GT}_{n,k} \sim \frac{2^{2k-1}}{(2k)! \sqrt{\pi}} n^{2k-3/2} 2^n n!$$

474 for fixed  $k$  as  $n \rightarrow \infty$ , where the last asymptotic equivalence follows from (3). This conjecture  
 475 can be proved with the tools developed in [1, 2] and will be presented in the journal version  
 476 of this paper.

477 **4 Conclusion**

478 In this extended abstract, we have enumerated and studied stochastic properties of two  
 479 classes of semi-simplex phylogenetic networks, namely, semi-simplex tree-child networks and  
 480 semi-simplex galled trees. For the former class, we have proved that the numbers of simplex  
 481 and semi-simplex tree-child networks just differ by a multiplicative factor whereas both are  
 482 much smaller than the the number of general tree-child networks. On the other hand, the  
 483 asymptotic expansions of the numbers of simplex galled trees and semi-simplex galled trees  
 484 already differ in their exponential growth term (but are again much smaller than the number  
 485 of general galled trees). In addition, we have studied shape parameters of these two classes of  
 486 semi-simplex networks and obtained limit distribution results for the number of reticulation  
 487 nodes and asymptotic results for the mean of the Sackin index.

488 Which other classes of semi-simplex phylogenetic networks could be studied? First, for  
 489 semi-simplex galled networks, only the enumeration problem has been solved so far. One  
 490 can ask for stochastic results for shape parameters and such results will be presented in the  
 491 full version of this paper. Another interesting class which could be studied are semi-simplex  
 492 normal networks, where a *normal* network ([11, 24]) is a tree-child network without *shortcuts*,  
 493 i.e., without edges from a node  $u$  to a node  $v$  so that  $v$  can be reached from  $u$  also by a path  
 494 of length at least 2. How many semi-simplex normal networks of size  $n$  are there? What  
 495 can be said of about the stochastic behavior of shape parameters of random semi-simplex  
 496 normal networks? These are interesting questions, however, their study might be much  
 497 harder compared to the corresponding questions for semi-simplex tree-child networks and  
 498 semi-simplex galled trees studied in the current paper as even simplex normal networks of  
 499 size  $n$  have not yet been successfully enumerated.

500 — **References** —

- 501 1 Lily Agranat-Tamir, Michael Fuchs, Bernhard Gittenberger, and Noah A. Rosenberg. Asymp-  
 502 totic enumeration of rooted binary unlabeled galled trees with a fixed number of galls. In  
 503 *35th International Conference on Probabilistic, Combinatorial and Asymptotic Methods for the*  
 504 *Analysis of Algorithms*, volume 302 of *LIPICs. Leibniz Int. Proc. Inform.*, pages 14, Art. No. 27.  
 505 Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2024. doi:10.4230/lipics.aofa.2024.27.
- 506 2 Lily Agranat-Tamir, Michael Fuchs, Bernhard Gittenberger, Noah A. Rosenberg, and Karth-  
 507 nik V. Seetharaman. Combinatorial comparison of general galled trees, time-consistent galled  
 508 trees, and simplex time-consistent galled trees. *Submitted*, 2026+.
- 509 3 Michael G. B. Blum, Olivier François, and Svante Janson. The mean, variance and limiting  
 510 distribution of two statistics sensitive to phylogenetic tree balance. *Ann. Appl. Probab.*,  
 511 16(4):2195–2214, 2006. doi:10.1214/105051606000000547.
- 512 4 Mathilde Bouvel, Philippe Gambette, and Marefatollah Mansouri. Counting phylogen-  
 513 etic networks of level 1 and 2. *J. Math. Biol.*, 81(6-7):1357–1395, 2020. doi:10.1007/  
 514 s00285-020-01543-5.
- 515 5 Gabriel Cardona and Louxin Zhang. Counting and enumerating tree-child networks and their  
 516 subclasses. *J. Comput. System Sci.*, 114:84–104, 2020. doi:10.1016/j.jcss.2020.06.001.
- 517 6 Yu-Sheng Chang and Michael Fuchs. Counting phylogenetic networks with few reticulation  
 518 vertices: galled and reticulation-visible networks. *Bull. Math. Biol.*, 86(7):Paper No. 76, 23,  
 519 2024. doi:10.1007/s11538-024-01309-w.
- 520 7 Yu-Sheng Chang, Michael Fuchs, Hexuan Liu, Michael Wallner, and Guan-Ru Yu. Enumerative  
 521 and distributional results for  $d$ -combining tree-child networks. *Adv. in Appl. Math.*, 157:Paper  
 522 No. 102704, 58, 2024. doi:10.1016/j.aam.2024.102704.
- 523 8 Michael Drmota. *Random Trees*. SpringerWienNewYork, Vienna, 2009. An interplay between  
 524 combinatorics and probability. doi:10.1007/978-3-211-75357-6.

- 525 9 Mareike Fischer, Lena Herbst, Sophie J. Kersting, Annemarie Luise Kühn, and Kristina Wicke.  
526 *Tree Balance Indices: A Comprehensive Survey*. Springer International Publishing, Cham,  
527 Switzerland, 2023. doi:10.1007/978-3-031-39800-1.
- 528 10 Philippe Flajolet and Robert Sedgewick. *Analytic Combinatorics*. Cambridge University Press,  
529 Cambridge, 2009. doi:10.1017/CB09780511801655.
- 530 11 Andrew Francis. “Normal” phylogenetic networks emerge as the leading class.  
531 *ArXiv:2107.10414*, 2021.
- 532 12 Michael Fuchs and Bernhard Gittenberger. Sackin indices for labeled and unlabeled classes of  
533 galled trees. *J. Math. Biol.*, 90(4):Paper No. 42, 34, 2025. doi:10.1007/s00285-025-02205-0.
- 534 13 Michael Fuchs, En-Yu Huang, and Guan-Ru Yu. Counting phylogenetic networks with  
535 few reticulation vertices: a second approach. *Discrete Appl. Math.*, 320:140–149, 2022.  
536 doi:10.1016/j.dam.2022.03.026.
- 537 14 Michael Fuchs and Mike Steel. A dichotomy law for certain classes of phylogenetic networks.  
538 *Bull. Math. Biol.*, 87(11):Paper No. 153, 17, 2025. doi:10.1007/s11538-025-01536-9.
- 539 15 Michael Fuchs, Guan-Ru Yu, and Louxin Zhang. On the asymptotic growth of the number of  
540 tree-child networks. *European J. Combin.*, 93:Paper No. 103278, 20, 2021. doi:10.1016/j.  
541 ejc.2020.103278.
- 542 16 Michael Fuchs, Guan-Ru Yu, and Louxin Zhang. Asymptotic enumeration and distributional  
543 properties of galled networks. *J. Combin. Theory Ser. A*, 189:Paper No. 105599, 28, 2022.  
544 doi:10.1016/j.jcta.2022.105599.
- 545 17 Andreas D. M. Gunawan, Jeyaram Rathin, and Louxin Zhang. Counting and enumerating  
546 galled networks. *Discrete Appl. Math.*, 283:644–654, 2020. doi:10.1016/j.dam.2020.03.005.
- 547 18 Sungsik Kong, Joan Carles Pons, Laura Kubatko, and Kristina Wicke. Classes of explicit  
548 phylogenetic networks and their biological and mathematical significance. *J. Math. Biol.*,  
549 84(6):Paper No. 47, 44, 2022. doi:10.1007/s00285-022-01746-y.
- 550 19 Hosam M. Mahmoud. *Evolution of random search trees*. Wiley-Interscience Series in Discrete  
551 Mathematics and Optimization. John Wiley & Sons, Inc., New York, 1992. A Wiley-Interscience  
552 Publication.
- 553 20 Colin McDiarmid, Charles Semple, and Dominic Welsh. Counting phylogenetic networks. *Ann.*  
554 *Comb.*, 19(1):205–224, 2015. doi:10.1007/s00026-015-0260-2.
- 555 21 Charles Semple and Mike Steel. *Phylogenetics*, volume 24 of *Oxford Lecture Series in Mathem-*  
556 *atics and its Applications*. Oxford University Press, Oxford, 2003.
- 557 22 Mike Steel. *Phylogeny—Discrete and Random Processes in Evolution*, volume 89 of *CBMS-*  
558 *NSF Regional Conference Series in Applied Mathematics*. Society for Industrial and Applied  
559 Mathematics (SIAM), Philadelphia, PA, 2016. doi:10.1137/1.9781611974485.ch1.
- 560 23 Benedikt Stuffer. A branching process approach to level- $k$  phylogenetic networks. *Random*  
561 *Structures Algorithms*, 61(2):397–421, 2022. doi:10.1002/rsa.21065.
- 562 24 Stephen J. Willson. Unique determination of some homoplasies at hybridization events. *Bull.*  
563 *Math. Biol.*, 69(5):1709–1725, 2007. doi:10.1007/s11538-006-9187-4.
- 564 25 Louxin Zhang. The Sackin index of simplex networks. In *Comparative genomics*, volume  
565 13234 of *Lecture Notes in Comput. Sci.*, pages 52–67. Springer, Cham, 2022. URL: [https://doi.org/10.1007/978-3-031-06220-9\\_4](https://doi.org/10.1007/978-3-031-06220-9_4), doi:10.1007/978-3-031-06220-9\\_4.
- 566