Two Remarks on The Sackin and Colless Indices of Trees

Gary Goh\textsuperscript{a}, Michael Fuchs\textsuperscript{b}, Louxin Zhang\textsuperscript{a}\textsuperscript{∗}
\textsuperscript{a} Department of Mathematics, National University of Singapore, Singapore, 119764
\textsuperscript{b} Department of Mathematical Sciences, National Chengchi University, Taipei, 116, Taiwan

Abstract

The Sackin and Colless indices are two widely-used metrics for measuring the balance of trees and for testing evolutionary models in phylogenetics. This short paper provides an asymptotic analysis of the expected Sackin and Colless indices on tree shapes (which are full binary rooted unlabelled trees) under the uniform model where tree shapes are sampled with equal probability. It also presents a short elementary proof of the closed formula for the expected Sackin index of phylogenetic trees (which are full binary rooted trees with leaves being labelled) under the uniform model. The new derivation does not even use the Catalan numbers.

Keywords: Phylogenetics, Tree balance, Sackin index, Colless index, Singularity analysis

1 Introduction

The Sackin \cite{1, 2} and Colless \cite{3} indices are two widely-used metrics for measuring the balance of phylogenetic trees and testing evolutionary models \cite{4, 5, 6, 7, 8, 9}. For a phylogenetic tree, its Sackin index is defined as the sum over its internal nodes of the number of leaves below that node, whereas its Colless index is defined as the sum over its internal nodes of the balance of that node, where the balance of a node is the difference in the number of leaves between the two subtrees rooted at the two children of that node. Because of their wide applications, different mathematical issues of these two tree balance metrics and other indices have been extensively studied in the past decades (see the recent comprehensive survey \cite{10}).

\textsuperscript{∗}Corresponding author at: Department of Mathematics, National University of Singapore, 10 Lower Kent Ridge Road, Singapore 119076. Tel: (65) 65166579. Email: matzlx@nus.edu.sg
The Sackin and Colless indices of a random phylogenetic tree have been investigated under the Yule-Harding model and uniform models \[9, 11, 12, 13\]. Blum and co-authors proved that the expected Sackin and Colless indices of a phylogenetic tree are asymptotic to \(\sqrt{\pi n^{3/2}}\) under the uniform model and \(n \log n\) under the Yule-Harding model \[12, 13\]. Recently, Mir et al. \[14\] discovered surprisingly that the expected Sackin index of a phylogenetic tree is simply \(4n - 1\). An alternative proof of this closed formula was given by King and Rosenberg \[15\]. Both asymptotic and exact results on the variances of the Sackin and Colless indices have also been reported \[9, 12, 13, 16\].

The Sackin index of a tree is actually equal to the sum of the depths of all its leaves \[17\]. The Sackin index and the tree height have been studied for other types of tree in the combinatorics and theoretical computer science literature \[18, 19, 20, 21\].

In this short paper, we focus on the Sackin and Colless indices of a tree shape under the uniform model \[22\]. Here, tree shapes (also called Otter or Polya trees) are rooted full binary trees with unlabeled leaves where each internal node has two children. To the best of our knowledge, the statistic properties of these two indices and other tree balance indices have not been formally studied \[10\]. Here, we prove that the expected Sackin and Colless indices of a tree shape with \(n\) leaves are asymptotic to \(\sqrt{\pi \lambda^{-1} n^{3/2}}\) under the uniform model, where \(\lambda \approx 1.1300337163\).

We also present a short elementary proof of the close formula for the expected Sackin index of a phylogenetic tree that is mentioned above. A simple recurrence for the Sackin index is derived using the fact that all the phylogenetic trees on \(n\) taxa can be enumerated by inserting the \(n\)-th taxon into every edge of the phylogenetic trees on \(n - 1\) taxa \[23\]. This technique was first used by Zhang for counting the sum over all nodes of the number of the descendants of that node in \[22\].

2 Basic definitions and notation

2.1 Phylogenetic trees and shapes

A phylogenetic tree on \(n\) taxa is a full binary tree with a designed root \(\rho\) in which there are \(n\) leaves that are uniquely labeled with a taxon and each of the \(n - 1\) non-leaf nodes has exactly two children.

A tree shape is a rooted full binary tree in which all nodes are unlabeled.

Let \(T\) be a phylogenetic tree on \(n\) taxa or a tree shape. We use \(V_0(T)\) to denote the set of all non-leaf nodes of \(T\) and \(V(T)\) to denote the set of all nodes. We say that a leaf is below a non-leaf node \(u\) if the unique path from the root to the leaf passes through \(u\) in \(T\). We use \(\ell_T(u)\) to denote the number of leaves below \(u\) in \(T\). Also, we set \(\ell_T(u) = 1\) if \(u\) is a leaf.

Let \(u \in V_0(T)\). The balance number of \(u\) is defined to be \(|\ell_T(v) - \ell_T(w)|\), where \(v\) and \(w\) are the two children of \(u\). We use \(\delta_T(u)\) to denote the balance number of \(u\).
2.2 Sackin and Colless indices

**Definition 1.** The SI of a tree shape or a phylogenetic tree \( T \) is defined to be \( \sum_{u \in V_0(T)} \ell_T(u) \), and denoted by \( S(T) \).

**Definition 2.** The CI of a tree shape or a phylogenetic tree \( T \) is defined to be \( \sum_{u \in V_0(T)} \delta_T(u) \), and denoted by \( C(T) \).

In this short paper, we study the expected SI and CI of a tree shape in the uniform model where shapes are sampled with equal probability. More precisely, they are defined as:

\[
ES_{sh}(n) = \frac{1}{b_n} \sum_{T \in T(n)} S(T)
\]

and

\[
EC_{sh}(n) = \frac{1}{b_n} \sum_{T \in T(n)} C(T),
\]

where \( T(n) \) denotes the set of all tree shapes with \( n \) leaves and \( b_n = |T(n)| \). Note that there does not exist a closed formula for \( b_n \); but \( b_n \) can be computed using the following recurrence formulas for \( n > 1 \):

\[
b_n = \sum_{1 \leq k < n/2} b_k b_{n-k} + \begin{cases} 
0, & \text{if } n \text{ is odd}; \\
\frac{1}{2} b_{n/2}(b_{n/2} + 1), & \text{if } n \text{ is even}.
\end{cases}
\]

Equivalently, the generating function \( B(x) = \sum_i b_n x^n \) satisfies the following equation [17]:

\[
B(x) = x + \frac{1}{2} (B(x)^2 + B(x^2)).
\]

The expected SI of a phylogenetic tree in the uniform model is defined similarly, that is,

\[
ES_{p}(n) = \frac{1}{a_n} \sum_{P \in P(n)} S(P),
\]

where \( P(n) \) denote the set of all phylogenetic trees on \( n \) taxa and \( a_n = |P(n)| = \frac{(2n-2)!}{2^{n-1}(n-1)!} \).

3 Asymptotic analysis of \( ESI_{sh}(n) \) and \( ECI_{sh}(n) \)

Recall that \( T(n) \) denotes the set of all possible tree shapes with \( n \) leaves. Let \( S_n \) denote the cumulative Sackin index of all the tree shapes with \( n \) leaves. Obviously, \( S_1 = 0, S_2 = 2 \). For \( n > 2 \), \( T(n) \) can be obtained by combining every pair of tree shapes \( T \in T(k) \) and \( T' \in T(n-k) \), where \( k \) is from 1 to \( n/2 \). Using this fact, we can derive a recurrence formula for \( S_n \) as follows.
Since \( \ell(u) = n \) for the root of every tree shape,

\[
S_n = nb_n + \sum_{1 \leq k < n/2} \left( \sum_{T \in \mathcal{T}(k)} \sum_{T' \in \mathcal{T}(n-k)} (S(T) + S(T')) \right)
\]

\[
= nb_n + \sum_{1 \leq k < n/2} (b_{n-k}S_k + b_kS_{n-k})
\]

\[
= nb_n + \sum_{1 \leq k < n} S_kb_{n-k}, \text{ for odd } n
\]

(2)

and

\[
S_n = nb_n + \sum_{1 \leq k < n/2} \left( \sum_{T \in \mathcal{T}(k)} \sum_{T' \in \mathcal{T}(n-k)} (S(T) + S(T')) \right)
\]

\[
+ \sum_{T, T' \in \mathcal{T}(n/2); T \neq T'} (S(T) + S(T')) + \sum_{T \in \mathcal{T}(n/2)} 2S(T)
\]

\[
= nb_n + \sum_{1 \leq k < n/2} (b_{n-k}S_k + b_kS_{n-k}) + \sum_{T \in \mathcal{T}(n/2)} (b_{n/2} - 1)S(T) + 2S_{n/2}
\]

\[
= nb_n + \sum_{1 \leq k < n} S_kb_{n-k} + S_{n/2}, \text{ for even } n
\]

(3)

It is unknown whether one can derive closed formulas for \( S_n \) from the above recurrence formulas or not. However, the asymptotic analysis of \( b_n \) from Eqn. (1) is classical and summarized below.

**Lemma 1.** ([19]) The convergence radius \( \rho \) of the generation function \( B(z) \) of \( b_n \) satisfies \( 1/4 \leq \rho \leq 1/2 \), where \( \rho + B(\rho^2)/2 = 1/2 \). Moreover, \( B(z) \) is \( \Delta \)-analytic and satisfies as \( z \to \rho \) in the \( \Delta \)-domain:

\[
B(z) = 1 - \lambda \sqrt{1 - z/\rho} + O(1 - z/\rho), \quad \lambda = \sqrt{2\rho + 2\rho^2 B'(\rho^2)}.
\]

(4)

Thus,

\[
b_n \sim \frac{\lambda}{2\sqrt{\pi n^{3/2}} \rho^n}, \quad (n \to \infty).
\]

(5)

**Remark 1.** \( \rho \) and \( \lambda \) can be computed up to very high precision, e.g.,

\[
\rho = 0.40269750367 \cdots \quad \text{and} \quad \lambda = 1.1300337163 \cdots.
\]

We next will derive the asymptotic values of \( ESI_{sh}(n) \) and \( ECI_{sh}(n) \).

**Theorem 1.** Under the uniform model, the expected Sackin of a tree shape with \( n \) leaves, \( ESI_{sh}(n) \), is asymptotic to \( n^{1/2} \lambda^{-1} \lambda^{3/2} \).

**Proof.** The recurrence formulas (2) and (3) translate into the following equation for the generating function \( S(z) \) of \( S_n \):

\[
S(z) = zB'(z) + S(z)B(z) + S(z^2),
\]
or, equivalently,
\[ S(z) = \frac{zB'(z) + S(z^2)}{1 - B(z)} \]
from which we deduce that the radius of convergence of \( S(z) \) is equal to \( \rho \).
Moreover, from Eqn. [4], we see that \( S(z) \) is \( \Delta \)-analytic and satisfies as \( z \to \rho \) in the \( \Delta \)-domain:
\[ S(z) \sim \frac{1}{2} \cdot \frac{1}{1 - z/\rho}, \]
where we used that \( B'(z) \) satisfies as \( z \to \rho \) in the \( \Delta \)-domain:
\[ B'(z) \sim \frac{\lambda}{2\rho} \cdot \frac{1}{\sqrt{1 - z/\rho}}. \]
(This follows from the closure properties of singularity analysis; see Section VI.10 in [25].) Thus, by the transfer theorems of singularity analysis (see Section VI.3 in [25]),
\[ S_n \sim \frac{1}{2\rho^n}, \quad (n \to \infty). \]
By Eqn. [5], we obtain:
\[ \text{ESI}_{sh}(n) = \frac{S_n}{b_n} \sim \frac{1/(2\rho^n)}{\lambda/2\sqrt{\pi}n^{3/2}/\rho^n} = \sqrt{\pi} \lambda^{-1}n^{3/2}, \quad (n \to \infty). \]
This proves the claim.

For each internal node \( u \) of a tree, we use \( c_1(u) \) and \( c_2(u) \) to denotes the two children of \( u \). We have that \( \ell(u) = \ell(c_1(u)) + \ell(c_2(u)) \) and thus \( \delta(u) = |\ell(c_1(u)) - \ell(c_2(u))| = \ell(u) - 2\min(\ell(c_1(u)), \ell(c_2(u))). \)
Form this, it follows that for each tree shape \( T \), \( S(T) - C(T) = 2\sum_{u \in V_0(T)} \min(\ell(c_1(u)), \ell(c_2(u))). \)
Define \( D(T) = \frac{1}{2}(S(T) - C(T)) = \sum_{u \in V_0(T)} \min(\ell(c_1(u)), \ell(c_2(u))) \) and \( D_n = \sum_{T \in \tau(n)} D(T). \) Let \( C_n \) denote the cumulative Colless index of all tree shapes with \( n \) leaves. Clearly,
\[ C_n = S_n - 2D_n. \]
We have the following recurrence formula for \( D_n \).
\[
D_n = \sum_{1 \leq k < n/2} \left( \sum_{T \in \tau(k)} \sum_{T' \in \tau(n-k)} (D(T) + D(T') + k) \right)
= \sum_{1 \leq k < n/2} kbkb_{n-k} + \sum_{1 \leq k < n/2} (b_{n-k}D_k + b_kD_{n-k})
= \sum_{1 \leq k \leq n/2} kbkb_{n-k} + \sum_{1 \leq k < n} D_kb_{n-k}, \text{ for odd } n
\]
and

\[
D_n = \sum_{1 \leq k < n/2} \left( \sum_{T \in \mathcal{T}(k)} \sum_{T' \in \mathcal{T}(n-k)} (D(T) + D(T') + k) \right) \\
+ \sum_{T,T' \in \mathcal{T}(n/2): T \neq T'} (D(T) + D(T') + n/2) + \sum_{T \in \mathcal{S}(n/2)} (2D(T) + n/2)
\]

\[
= \sum_{1 \leq k < n/2} kb_k b_{n-k} + \sum_{1 \leq k < n/2} (b_{n-k} D_k + b_k D_{n-k}) \\
+ \sum_{T \in \mathcal{T}(n/2)} (b_{n/2} - 1)D(T) + \left( \frac{b_{n/2}}{2} \right) \frac{n}{2} + 2D_{n/2} + \frac{n}{2} b_{n/2}
\]

\[
= \sum_{1 \leq k \leq n/2} kb_k b_{n-k} + \sum_{1 \leq k < n} D_k b_{n-k} \\
- \frac{n}{2} \left( \frac{b_{n/2}}{2} \right) + D_{n/2}, \text{ for even } n.
\]

Next, set:

\[
F_n := \sum_{1 \leq k \leq n/2} kb_k b_{n-k} + \begin{cases} 
0, & \text{if } n \text{ is odd;} \\
- \frac{n}{2} \left( \frac{b_{n/2}}{2} \right), & \text{if } n \text{ is even.}
\end{cases}
\]

**Lemma 2.** We have \( F_n = O \left( n^{-1} \rho^{-n} \right) \).

**Proof.** By using Eqn. 5,

\[
F_n = O \left( \rho^{-n} \sum_{1 \leq k \leq n/2} k^{-1/2} (n-k)^{-3/2} + n^{-2} \rho^{-n} \right)
\]

\[
= O \left( n^{-1} \rho^{-n} \int_0^{1/2} x^{-1/2} (1-x)^{-3/2} dx \right) = O \left( n^{-1} \rho^{-n} \right), \text{ for large } n.
\]

Now, define:

\[
\tilde{D}_n = K n^{-1} \rho^{-n} + \sum_{1 \leq k < n} \tilde{D}_k b_{n-k} + \begin{cases} 
0, & \text{for } n \text{ is odd;} \\
\tilde{D}_{n/2}, & \text{for } n \text{ is even},
\end{cases}
\]

where \( K \) is the implied \( O \)-constant from the last lemma. Then, \( D_n \leq \tilde{D}_n \). Moreover, for the asymptotics of \( \tilde{D}_n \), we have the following.

**Lemma 3.** We have,

\[
\tilde{D}_n \sim K \lambda \sqrt{n}^{-1/2} n^{-1/2} (\log n) \rho^{-n}, \quad (n \to \infty).
\]

Consequently, \( D_n = O \left( n^{-1/2} (\log n) \rho^{-n} \right) \).
Proof. Denote by \( \tilde{D}(z) \) the generating function of \( \tilde{D}_n \). Then, the recurrence (8) translates into

\[
\tilde{D}(z) = K \log \frac{1}{1 - z/\rho} + \tilde{D}(z)B(z) + \tilde{D}(z^2).
\]

Solving this gives

\[
\tilde{D}(z) = \frac{K \log \frac{1}{1 - z/\rho} + \tilde{D}(z^2)}{1 - B(z)}.
\]

Thus, from Eqn. (4), \( \tilde{D}(z) \) satisfies as \( z \to \rho \) in a \( \Delta \)-domain:

\[
\tilde{D}(z) \sim \frac{K}{\lambda} \cdot \frac{1}{\sqrt{1 - z/\rho}}
\]

from which the claimed result follows by singularity analysis.

Now from Eqn. (6), Eqn. (7) and Lemma 3 we have the following result for the Colless index.

**Theorem 2.** Under the uniform model, the expected Colless index of a tree shape with \( n \) leaves, \( ECI_{sh}(n) \), is also asymptotic to \( \pi^{1/2}/n^{3/2} \).

### 4 An elementary proof of the formula for \( ECI_p(n) \)

We assume that there is an open edge entering the root of each tree of \( \mathcal{P}(n) \). Under such a convention, \( \mathcal{P}(n+1) \) can be obtained from \( \mathcal{P}(n) \) by attaching Leaf \( n + 1 \) on each of the \( 2n - 1 \) edges of every tree of \( \mathcal{P}(n) \) (see [23] for example), illustrated in Figure 1.

For each \( P \in \mathcal{P}(n) \), we let \( P_1, P_2, \ldots, P_{2n-1} \) be all the \( 2n - 1 \) phylogenetic trees on \( n + 1 \) taxa that are obtained from \( P \) by attaching Leaf \( n + 1 \) on each
of the $2n - 1$ tree edges of $P$. Note that Leaf $n + 1$ and its parent are the only nodes in $P_i$ that are not found in $P$.

Assume that $P_i$ is obtained by attaching Leaf $n + 1$ to the edge $(p, u)$ of $P$, where $u \in V(P)$. The number of leaves below the parent of Leaf $n + 1$ in $P_i$ is $1 + \ell_P(u)$. In total, therefore, the amount contributed by the parents of Leaf $n + 1$ to $\sum_i S(P_i)$ is

$$\sum_{u \in V(P)} \ell_P(u) + (2n - 1) = \sum_{u \in V_0(P)} \ell_P(u) + n + (2n - 1) = S(P) + 3n - 1.$$ 

For $u \in V(P)$, $\ell(u)$ changes in $P_i$ if and only if Leaf $n + 1$ is attached on an edge below $u$. There are $2\ell_P(u) - 2$ edges below $u$ in $P$, Hence, $\ell(u)$ increases by 1 in exactly $2\ell_P(u) - 2$ of $\{P_i\}_{i=1}^{2n-1}$. Thus,

$$\sum_{i=1}^{2n-1} S(P_i) = (2n - 1)S(P) + S(P) + (3n - 1) + \sum_{u \in V_0(P)} (2\ell_P(u) - 2)$$

$$= 2nS(P) + (3n - 1) + 2S(P) - 2|V_0(P)|$$

$$= 2(n + 1)S(P) + n + 1.$$ 

Adding $n+1$ to each term in the left-hand side of the above equality⁴ we further have:

$$\sum_{i=1}^{2n-1} (S(P_i) + (n + 1)) = 2(n + 1)S(P) + n + 1 + (2n - 1)(n + 1)$$

$$= 2(n + 1) (S(P) + n).$$ 

Let $S_{n+1}^{(p)} = \sum_{P \in \mathcal{P}(n+1)} S(P)$. Then,

$$S_{n+1}^{(p)} + (n + 1)a_{n+1} = \sum_{P \in \mathcal{P}(n)} \sum_{i=1}^{2n-1} (S(P_i) + (n + 1))$$

$$= \sum_{P \in \mathcal{P}(n)} 2(n + 1) (S(P) + n)$$

$$= 2(n + 1) \left( S_n^{(p)} + na_n \right). \tag{9}$$

Since $S_2^{(p)} = 2$ and $a_2 = 1$, Eqn. (9) implies that $S_n^{(p)} = 2^{n-1}n! - na_n$ and

$$\text{ESI}_p(n) = \frac{S_n^{(p)}}{a_n} = \frac{4^{n-1}n!(n - 1)!}{(2n - 2)!} - n$$

which is the formula first appearing in [14].

⁴We may consider $n + 1$ as the contribution of the $n + 1$ leaves.
CRediT authorship contribution statement

G. Goh: Recurrence formulas; L. Zhang: Recurrence formulas, writing; M. Fuchs: Asymptotic analysis, writing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

LZ was supported by MOE Tier 1 grant R-146-000-318-114; MF was supported by MOST-109-2115-M-004-003-MY2.

References

[1] M. Sackin, ‘Good’ and ‘bad’ phenograms, Syst. Zool. 21 (1972) 225–226.
[2] K.-T. Shao and R. R. Sokal, Tree balance, Syst. Zoology 39 (1990) 266–276.
[3] D. Colless, Review of “Phylogenetics: the theory and practice of phylogenetic systematics”, Syst. Zool. 31 (1982) 100–104.
[4] M. Avino, G. T. Ng, Y. He, M. S. Renaud, B. R. Jones, A. F. Poon, Tree shape-based approaches for the comparative study of cophylogeny, Ecol. Evol. 9 (2019) 6756–6771.
[5] C. Xue, Z. Liu, N. Goldenfeld, Scale-invariant topology and bursty branching of evolutionary trees emerge from niche construction, Proc. National Academy of Sciences 117 (2020) 7879–7887.
[6] A. O. Mooers and S. B. Heard. Inferring evolutionary process from phylogenetic tree shape, The Quarterly Review of Biology 72(1) 1997 31–54.
[7] J. G. Scott, P. K Maini, A. R. A. Anderson, A. G. Fletcher, Inferring tumor proliferative organization from phylogenetic tree measures in a computational model, Syst. Biol. 69(4) (2020) 623–637.
[8] M. G. B. Blum, E. Heyer, O. François, F. Austerlitz, Matrilineal fertility inheritance detected in hunter–gatherer populations using the imbalance of gene genealogies, PLoS Genetics 2 (2006) e122. doi: 10.1371/journal.pgen.0020122.
[9] M. Kirkpatrick and M. Slatkin, Searching for evolutionary patterns in the shape of a phylogenetic tree, Evolution 47 (1993) 1171–1181.
[10] M. Fischer, L. Herbst, S. Kersting, L. Kühn, K. Wicke, Tree balance indices: a comprehensive survey. arXiv preprint arXiv:2109.12281 2021.

[11] S. B. Heard, Patterns in tree balance among cladistic, phenetic, and randomly generated phylogenetic trees, Evolution 46 (1992) 1818–1826.

[12] M. G. B. Blum and O. François, On statistical tests of phylogenetic tree imbalance: the Sackin and other indices revisited, Math. Biosci. 195 (2005) 141–153.

[13] M. G. B. Blum, O. François, S. Janson, The mean, variance and limiting distribution of two statistics sensitive to phylogenetic tree balance, Ann. Appl. Probab. 16 (2006) 2195–2214.

[14] A. Mir, F. Rosselló, L. Rotger, A new balance index for phylogenetic trees, Math. Biosci. 241 (2013) 125–136.

[15] M. C. King and N. A. Rosenberg, A simple derivation of the mean of the Sackin index of tree balance under the uniform model on rooted binary labeled trees, Math. Biosci. 342 (2021) 108688.

[16] T. M. Coronado, A. Mir, F. Rosselló, L. Rotger, On Sackin’s original proposal: the variance of the leaves’ depths as a phylogenetic balance index, BMC Bioinform. 21 (2020) 154.

[17] M. Steel, Phylogeny: Discrete and Random Processes in Evolution, Society for Industrial and Applied Mathematics, Philadelphia, 2016.

[18] P. Flajolet and A. Odlyzko, The average height of binary trees and other simple trees, J. Comput. Syst. Sci. 25 (2) (1982) 171–213.

[19] N. Broutin and P. Flajolet, The distribution of height and diameter in random non-plane binary trees, Random Struct. Algor. 41 (2012) 215–252.

[20] J. A. Fill and N. Kapur, Limiting distributions for additive functionals on Catalan trees, Theoret. Comput. Sci. 326 (2004) 69–102.

[21] M. Fuchs and E. Y. Jin, Equality of Shapley value and fair proportion index in phylogenetic trees, J. Math. Biol. 71 (2015) 1133–1147.

[22] J. S. Rogers, Central moments and probability distributions of three measures of phylogenetic tree imbalance, Syst. Biol. 45 (1996) 99–110.

[23] J. Felsenstein, Inferring Phylogenies, Sunderland, MA. USA, 2004.

[24] L. Zhang, Generating normal networks via leaf insertion and nearest neighbor interchange, BMC Bioinforma. 20 (2019) 1–9.

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