Sparsification of Phylogenetic Covariance Matrices of k-Regular Trees

Sean Svihla
Department of Applied Mathematics, University of Colorado, Boulder, CO, USA

Manuel E. Lladser[1]
Department of Applied Mathematics, University of Colorado, Boulder, CO, USA

Abstract

Consider a tree $T = (V, E)$ with root $\circ$ and an edge length function $\ell : E \to \mathbb{R}_+$. The phylogenetic covariance matrix of T is the matrix C with rows and columns indexed by L, the leaf set of T, with entries $C(i, j) := \sum_{e \in [i \wedge j, o]} \ell(e)$, for each $i, j \in L$. Recent work [Gorman & Lladser 2023] has shown that the phylogenetic covariance matrix of a large but random binary tree T is significantly sparsified, with overwhelmingly high probability, under a change-of-basis to the so-called Haar-like wavelets of T. Notably, this finding enables manipulating the spectrum of covariance matrices of large binary trees without the necessity to store them in computer memory but instead performing two post-order traversals of the tree [Gorman & Lladser 2023]. Building on the methods of the aforesaid paper, this manuscript further advances their sparsification result to encompass the broader class of k-regular trees, for any given $k \geq 2$. This extension is achieved by refining existing asymptotic formulas for the mean and variance of the internal path length of random k-regular trees, utilizing hypergeometric function properties and identities.

2012 ACM Subject Classification Mathematics of computing → Trees; Mathematics of computing → Generating functions; Mathematics of computing → Random graphs

Keywords and phrases cophenetic matrix, Haar-like wavelets, hierarchical data, hypergeometric functions, metagenomics, phylogenetic covariance matrix, sparsification, ultrametric matrix

Digital Object Identifier 10.4230/LIPIcs.AofA.2024.4

Funding This research has been partially funded by the NSF grant No. 1836914.

Acknowledgements We are thankful to the reviewers for their comments and insightful suggestions.

1 Introduction

Hierarchical datasets are described, or presumed to be, by a rooted tree that recursively organizes data into clusters so that its leaves are in a one-to-one correspondence with the data points. Such datasets are common in various fields such as microbial ecology [5], where they arise as models of inter-species covariance [10, 21]. In this context, the associated covariance matrices are often large and dense, making their manipulation computationally challenging. Nevertheless, these matrices bear redundancies induced by their hierarchical structure, which may be exploited to sparsify them and make such manipulations tenable, if not trivial.

Ultrametric matrices, which arise in probability theory and statistical physics among other fields, are often also dense. A symmetric matrix $C \in \mathbb{R}_+^{n \times n}$ is called *ultrametric* when $C(i, j) \geq \min\{C(i, k), C(k, j)\}$, for all $i, j, k \in \{1, \ldots, n\}$. If, in addition, $C(i, i) > C(i, j)$ for all $j \neq i$ when $n > 1$, or $C(1, 1) > 0$ when $n = 1$, C is called *strictly ultrametric*. These matrices are fully dense, that is, all their entries are nonzero, but have a myriad of mathematical properties [6, 18].

---

[1] Corresponding author

35th International Conference on Probabilistic, Combinatorial and Asymptotic Methods for the Analysis of Algorithms (AofA 2024).
Editors: Cécile Mailler and Sebastian Wild; Article No. 4; pp. 4:1–4:17
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

A matrix C is ultrametric if and only if there is a rooted binary tree $T = (V, E)$ and edge length function $\ell : E \to \mathbb{R}_+$ such that [14, 20]:

$$C = \sum_{e \in E} \ell(e)\, \delta_e\, \delta_e',$$

where δ_e is the binary column vector, with entries indexed by the leaves of T, indicating the leaves that descend from e, and δ_e' is the transpose of δ_e. If L denotes the leaf set of T, the above identity is equivalent to having

$$C(i,j) = \sum_{e \in [i \wedge j, \circ]} \ell(e), \text{ for all } i, j \in L, \tag{1}$$

where $[i \wedge j, \circ]$ denotes the set of edges that connect $(i \wedge j)$, the least common ancestor of i and j, with the root of the tree, denoted as $\circ$. Ultrametric matrices have, therefore, a recursive structure, and their entries are redundant, suggesting that they may be amenable to some form of compression [14].

The formulation in (1) arises naturally as a model of "phylogenetic" covariance wherein the genetic drift of a particular trait follows a Brownian motion [15]. Under this model, each leaf represents a microbial species (or some notion thereof), and the trait variation among different species is a function of time since they diverged evolutionarily. Intuitively, since species sharing more of their evolutionary history should thrive or struggle in similar environments accordingly, a natural measure of trait covariance between two different species i and j is the length of their shared evolutionary history, namely the quantity in (1).

In the general setting of rooted trees – not necessarily binary – a matrix with entries such as in (1) is called the *phylogenetic covariance matrix* (or cophenetic matrix) of a tree. (The term "phylogenetic" is usually omitted from now on.) Tree covariance matrices arise naturally in the context of hierarchical datasets; in particular, the class of covariance matrices associated with datasets having a binary hierarchy is precisely the class of ultrametric matrices.

Recent work [14] has demonstrated that, in the case of large datasets with a binary hierarchy, or equivalently, weighted and rooted binary trees, the associated covariance matrices become asymptotically diagonal with overwhelmingly high probability after changing basis to the so-called Haar-like wavelets [11] of the tree. (By "asymptotically diagonal with overwhelmingly high probability," we mean that the fraction of non-zero off-diagonal entries of the covariance matrix, with respect to the wavelet basis, converges in probability to 0 as the tree size approaches infinity.) The sparsification of such covariance matrices facilitates manipulations that may be infeasible otherwise. For instance, the spectrum of ultrametric matrices can be derived, whether exactly or approximately depending on the matrix size, from just two post-order traversals of the associated tree without having to store the actual matrix in computer memory [14]. In addition, the subclass of ultrametric matrices diagonalized by Haar-like wavelets has been characterized, and their spectra shown to be in bijection with non-negative decreasing functions on the interior nodes of a binary tree [14].

Nevertheless, many hierarchical datasets, for example, in the context of phylogenomic studies [24], are non-binary and suffer the same downfall of having unmanageably large and dense covariance matrices. It is a natural question, then, whether the same technique may be used to sparsify covariance matrices belonging to a broader class of hierarchy. This manuscript extends some of the ideas in [14] to the broader context of k-regular trees, i.e., rooted trees for which each interior node has exactly k children. Specifically, Theorem 2, Theorem 4, Corollary 5, and Corollary 15 in this manuscript are generalizations of [14, Theorem 2.3, Theorem 3.4, Corollary 3.5, and Corollary 3.8], respectively, such that they are applicable to any $k \geq 2$. In addition, Theorem 10 supplies more precise asymptotic formulas for the mean and variance of the internal path length of random k-regular trees – beyond what is currently available in the literature.

This manuscript is based on the recent MS thesis [22].

Notation and Terminology

Depending on the context, we regard functions with finite domains as finite-dimensional column vectors, and vice versa. Throughout, $[\![\cdot]\!]$ is used to denote indicator functions.

We use standard terminology for trees unless otherwise stated. In particular, $T = (V, E)$ represents a rooted tree with vertex set V and edge set E. The size of T is the quantity $|T| := |V|$. The sets of leaves and interior nodes of T are denoted as L and I, respectively, and the internal path length of T is the quantity

$$\mathrm{IPL}(T) = \sum_{v \in I} \mathrm{depth}(v).$$

For $u, v \in V$, we denote by $[u, v]$ the set of edges in the shortest path from u to v and by $(u \wedge v)$ the least common ancestor of u and v.

$\mathring{T}$ denotes the *interior* of T, obtained by trimming the leaves of T.

In the subsequent discussion, T denotes a planted, ordered, and unlabelled k-regular tree endowed with an edge length function $\ell : E \to \mathbb{R}_+$. In particular, T is an ordered unlabelled k-regular tree whose root has been appended to a phantom one (denoted as $\circ$), which acts as the parent of the original root. (The term “out-rooted” instead of “planted” was introduced in [14].) See Figure 1(a). By definition, $\circ$ is an interior node of T (i.e., $\circ \in I$) and it has degree one. Furthermore, since each $v \in I \setminus \{\circ\}$ has k children, we denote these generically as $v_1, \ldots, v_k$ from left-to-right.

For a given $v \in V$, $T(v)$ denotes the sub-tree of T rooted at v and containing all of its descendants. We let $L(v)$ denote the sets of leaves of $T(v)$. In addition, we orient edges away from the root, that is, if $e = (u, v) \in E$, then u is understood to be the parent of v (and v is understood to be a child of u). We define $T(e) := T(v)$ and $L(e) := L(v)$.

The *trace length* of T is the function $\ell^* : E \to \mathbb{R}_+$ defined as [14]:

$$\ell^*(e) := |L(e)|\, \ell(e).$$

We also define for $u, v \in V$:

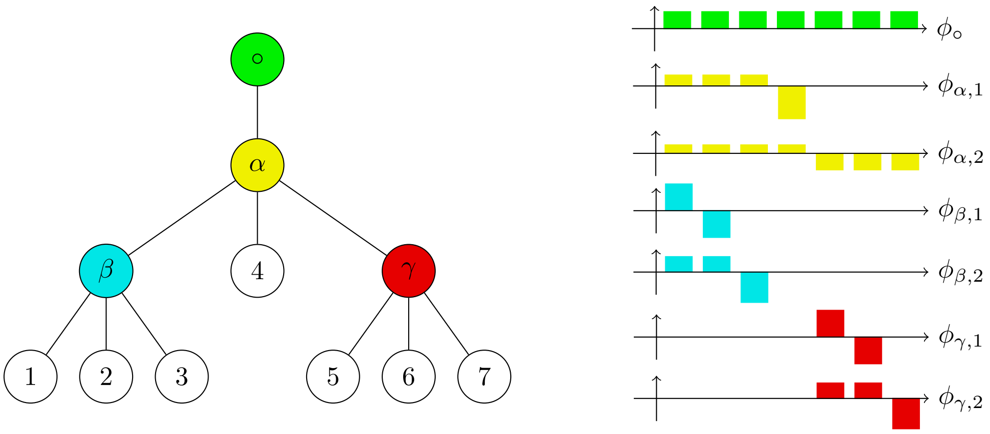
$$\ell(u, v) := \sum_{e \in [u,v]} \ell(e), \text{ and } \ell^*(u, v) := \sum_{e \in [u,v]} \ell^*(e),$$

and, given non-empty $J \subset V$, we denote by $\ell(J, v)$ the column vector with entries $\ell(j, v)$, for all $j \in J$, having dimension $|J|$. We give an analogous definition to $\ell^*(J, v)$.

For the remainder of this manuscript, the term “k-regular tree” encompasses planted, unlabelled, ordered, k-regular trees.

2 Haar-like Wavelets on k-Regular Trees

In this section, we specialize the Haar-like wavelet basis given in [11] to our setting of k-regular trees and present a useful interaction between the basis and the phylogenetic covariance matrix associated with any such tree (Theorem 2).



(a) 3-Regular Tree.

(b) Associated Haar-Like Wavelets.

Figure 1 Haar-like wavelets (right panel) associated with a 3-regular tree (left panel). As per our convention, the tree leaves are labeled $1, 2, \ldots$ following a depth-first search (DFS), while the non-root internal nodes are leaf-labeled. Thus $\alpha := \{1, 2, 3, 4, 5, 6, 7\}$, $\beta := \{1, 2, 3\}$, and $\gamma := \{5, 6, 7\}$. Except for the planted root (i.e., green node), each internal node has two wavelets associated with it. By construction, the wavelets are orthonormal piece-wise constant functions over the leave set and, except for $\phi_\circ$, each has mean zero.

While our manuscript focuses on ordered unlabeled trees, systematically naming their leaves is beneficial for constructing the wavelets. Henceforth, **we label the leaves of T as $1, \ldots, |L|$, following the order encountered in a depth-first search (DFS) of T.** Moreover, aside from the planted root, **we label interior nodes by the set of leaves descending from them.**

Wavelets are usually functions defined on a Euclidean space and derive their name from their commonly wave-like shape. They are prevalent in time series and image analysis to localize information across various scales.

In our context, the Haar-like wavelets associated with T are functions from L to $\mathbb{R}$ that are in a one-to-one correspondence with the elements in the set $\{\circ\} \cup \big(I \setminus \{\circ\} \times \{1, \ldots, k-1\}\big)$; in particular, there are $1 + (k-1)(|I| - 1)$ wavelets associated with a k-regular tree. The precise definition follows.

Given a $v \in I$ and integer $1 \leq n < k$, define

$$L_{v,n} := \bigcup_{j=1}^{n} L(v_j), \text{ and } L^+_{v,n} := L_{v,n} \cup L(v_{n+1}).$$

In particular, $L^+_{v,n} = L_{v,n+1}$. Nevertheless, we introduced the notation of $L^+_{v,n}$ because this is precisely the support of the wavelet with index (v, n) in the next definition.

▶ **Definition 1** (Haar-like wavelets of a k-regular tree). *The (mother) wavelet associated with the root of T is the constant function $\phi_\circ : L \to \mathbb{R}$ defined as*

$$\phi_\circ(i) := \frac{1}{\sqrt{|L|}}, \text{ for } i \in L.$$

Instead, for each $v \in I \setminus \{\circ\}$ and integer $1 \leq n < k$, the wavelet with index (v, n) is the function $\phi_{v,n} : L \to \mathbb{R}$ defined as

$$\phi_{v,n}(i) := \sqrt{\frac{|L(v_{n+1})|}{|L_{v,n}| \cdot |L_{v,n+1}|}} \, [\![i \in L_{v,n}]\!] - \sqrt{\frac{|L_{v,n}|}{|L(v_{n+1})| \cdot |L_{v,n+1}|}} \, [\![i \in L(v_{n+1})]\!] \, , \text{ for } i \in L.$$

It turns out that the Haar-like wavelets are mutually orthogonal [11]. Moreover, from well-known facts about graphs and trees,

$$|I| + |L| - 1 = \frac{1}{2}\sum_{v\in V}\deg(v) = \frac{1}{2}\Big[|I| + |L| + k\big(|I| - 1\big)\Big],$$

hence $|L| = 1 + (k-1)(|I|-1)$. That is, there are as many Haar-like wavelets as leaves on the tree, and the wavelets form therefore an orthonormal basis of the linear space of functions from L to $\mathbb{R}$.

We emphasize that the wavelet with index (v, n) assigns identical positive values to the leaves in $L_{v,n}$, and identical negative values to those in $L(v_{n+1})$. These values are chosen so that $\phi_{v,n}$ has mean zero and a unit ℓ^2-norm; namely,

$$\sum_{i\in L}\phi_{v,n}(i) = 0, \text{ and } \sum_{i\in L}\phi^2_{v,n}(i) = 1.$$

To fix ideas, Figure 1(b) provides an illustration of the Haar-like wavelets of a 3-regular tree.

Notice that $\phi_{v,n}$ is supported on $L^+_{v,n}$. As such, wavelets associated with nodes nearer the root capture coarser information about the leaves of the tree. As noted in [13, Theorem 1.1] for $k = 2$, given a function $\varphi : L \to \mathbb{R}$ and wavelet $\phi_{v,n}$, we have

$$\langle\varphi, \phi_{v,n}\rangle = c_{v,n}\cdot\left\{\frac{1}{|L_{v,n}|}\sum_{i\in L_{v,n}}\varphi(i) - \frac{1}{|L(v_{n+1})|}\sum_{i\in L(v_{n+1})}\varphi(i)\right\},$$

where

$$c_{v,n} := \sqrt{\frac{|L_{v,n}|\,|L(v_{n+1})|}{|L^+_{v,n}|}}.$$

Consequently, projecting a function onto $\phi_{v,n}$ is the same, up to a constant factor, as computing the difference between the average values of φ over $L_{v,n}$ and $L(v_{n+1})$. Hence, the projections of a real-valued function defined over L onto the Haar-like wavelets can be computed efficiently – which is relevant for applications involving large trees.

The following result highlights a remarkably simple action of the covariance matrix of a k-regular tree over its Haar-like wavelets. This property was first noticed in [14] for the case of 2-regular trees.

▶ **Theorem 2.** *If ψ is a Haar-like wavelet associated with $v \in I$, then $C\,\psi = \operatorname{diag}(\ell^*(L, v))\,\psi$.*

Proof. We first show the result for $\psi = \phi_\circ$. For this, note that for all $j \in L$

$$(C\psi)(j) = \frac{1}{\sqrt{L}}\sum_{i\in L}\ell(i\wedge j, \circ) = \psi(j)\sum_{i\in L}\sum_{e\in[i\wedge j,\circ]}\ell(e),$$

yet we have the logical equivalence:

$$\forall i, j \in L\;\forall v \in I, e \in [i\wedge j, v] \iff i \in L(e) \text{ and } e \in [j, v]. \tag{2}$$

Hence, $\ell(e)$ occurs $|L(e)|$ times in the previous double-sum, and

$$(C\psi)(j) = \psi(j)\sum_{e\in[j,v]}|L(e)|\,\ell(e) = \psi(j)\,\ell^*(j, v),$$

which is precisely the j-th entry of the vector $\operatorname{diag}(\ell^*(L, v))\,\psi$.

Next, we consider $\psi = \phi_{v,n}$, for a $v \in I \setminus \{\circ\}$ and $1 \le n < k$. Then, for each $j \in L$, we have the following cases:

(a) Assume $j \notin L^+_{v,n}$. Then, for all $i \in L^+_{v,n}$, $(i \wedge j) = (v \wedge j)$, so

$$(C\psi)(j) = \sum_{i \in L^+_{v,n}} \ell(i \wedge j, \circ)\, \phi_{v,n}(i) = \ell(v \wedge j, \circ) \sum_{i \in L^+_{v,n}} \phi_{v,n}(i) = 0,$$

because $\phi_{v,n}$ has mean zero. On the other hand, the entry associated with j in the vector $\operatorname{diag}(\ell^*(L,v))\,\psi$ also vanishes because ψ is supported on $L^+_{v,n}$.

(b) Instead, if $j \in L^+_{v,n}$ then

$$\begin{aligned}(C\psi)(j) &= \sum_{i \in L^+_{v,n}} \ell(i \wedge j, \circ)\, \phi_{v,n}(i) \\ &= \sum_{i \in L^+_{v,n}} \ell(i \wedge j, v)\, \phi_{v,n}(i) + \ell(v, \circ) \sum_{i \in L^+_{v,n}} \phi_{v,n}(i) \\ &= \sum_{i \in L^+_{v,n}} \phi_{v,n}(i) \sum_{e \in [i \wedge j, v]} \ell(e),\end{aligned}$$

where for the middle identity we have reused that $\phi_{v,n}$ has mean zero. But notice that $(i \wedge j) = v$ when $\ell(i \wedge j, v) = 0$. Otherwise, if $(i \wedge j) \neq v$ then $\phi_{v,n}(i) = \phi_{v,n}(j)$. Hence, reusing the logic equivalence in (2), we find that

$$(C\psi)(j) = \phi_{v,n}(j) \sum_{i \in L^+_{v,n} : (i \wedge j) \neq v} \; \sum_{e \in [i \wedge j, v]} \ell(e) = \psi(j)\, \ell^*(j, v).$$

which corresponds to the entry associated with j in the vector $\operatorname{diag}(\ell^*(L,v))\,\psi$.

The theorem now follows from parts (a)-(b). ◀

3 Sparsification of k-Regular Covariance Matrices

We now leverage the insights from the previous section to obtain a lower bound on the proportion of entries in the covariance matrix of a k-regular tree that vanish when switching to the Haar-type basis. To establish our main result in this section, we require the following definition, which is analogous to the one used in [14] for 2-regular trees.

▶ **Definition 3** (Haar-like matrix of a k-regular tree). *The Haar-like matrix of T is the square matrix Φ whose columns are the wavelets associated with the tree; that is, its rows are indexed by L and its columns by the wavelets.*

Let ψ_u, ψ_v be wavelets associated with interior nodes $u, v \in I$, respectively. We have as a direct consequence of Theorem 2 that

$$\big(\Phi' C \Phi\big)(\psi_u, \psi_v) = \psi_u \operatorname{diag}\big(\ell^*(L,v)\big)\, \psi_v. \tag{3}$$

But note that the support of $\operatorname{diag}\big(\ell^*(L,v)\big)\, \psi_v$ is contained in the support of ψ_v; therefore, the entry associated with row ψ_u and column ψ_v of $\Phi' C \Phi$ vanish when ψ_u and ψ_v have disjoint supports. Since $\Phi'\, C\, \Phi$ corresponds to C after changing basis to the Haar-like wavelets, (3) shows that the wavelets may be used to sparsify the covariance matrix of k-regular trees. Of course, there is no reason why such interactions should occur often enough to meaningfully sparsify the matrix. As detailed next, however, a minimum level of sparsification is guaranteed by the tree's size and internal path length.

We emphasize that the following result is a conservative bound on the number of vanished entries under the Haar-like basis. In practice, it is not uncommon to see better sparsification; however, we have found in all cases that the overwhelming majority of sparsification arises from the interactions considered in the following theorem. We discuss this topic further in Section 4.

▶ **Theorem 4.** *Let Φ and C denote the Haar-like matrix and (phylogenetic) covariance matrix of a k-regular tree T, respectively. If ζ is the fraction of vanishing entries of $\Phi' C \Phi$ then*

$$(1-\zeta) \leq (k-1)^2 \frac{1}{|I|} + 2(k-1)^2 \frac{\mathrm{IPL}(T)}{|I|^2}.$$

Proof. Note that C has dimensions $|L| \times |L|$ because T has as many Haar-like wavelets as leaves. Observe if $u, v \in I$ and ψ_u, ψ_v are any of their corresponding wavelets, identity (3) implies

$$\big(\Phi' \, C \, \Phi\big)(\psi_u, \psi_v) = \sum_{i \in L(u) \cap L(v)} \psi_u(i) \, \ell^*(i,v) \, \psi_v(i).$$

As $\big(\Phi' \, C \, \Phi\big)(\psi_u, \psi_v) = 0$ when $L(u) \cap L(v) = \emptyset$, then ζ obeys:

$$(1-\zeta) \leq \frac{(k-1)^2}{|L|^2} \Big| \Big\{ (u,v) \in I \times I \, ; \, L(u) \cap L(v) \neq \emptyset \Big\} \Big|.$$

Notice, however, that $L(u) \cap L(v) \neq \emptyset$ only if u descends from v (or vice versa) [14]. Hence, accounting for pairs of the form (u,u) and, when $u \neq v$, (u,v) and (v,u), we obtain that

$$(1-\zeta) \leq \frac{(k-1)^2}{|L|^2} \left(|I| + 2 \sum_{u \in I} \big(|\mathring{T}(u)| - 1 \big) \right) \leq \frac{(k-1)^2}{|L|^2} \left(2 \sum_{u \in I} |\mathring{T}(u)| - |I| \right).$$

Now, notice $\sum_{u \in I} |\mathring{T}(u)| = \sum_{u \in I} \sum_{v \in I} [\![v \in T(u)]\!]$; that is, each node is counted once for each of its ancestors, or

$$\sum_{u \in I} |\mathring{T}(u)| = \sum_{v \in I} \big(1 + \mathrm{depth}(v) \big) = |I| + \mathrm{IPL}(T).$$

Hence, we further obtain that

$$(1-\zeta) \leq \frac{(k-1)^2}{|L|^2} \Big(|I| + 2 \, \mathrm{IPL}(T) \Big).$$

As mentioned in Section 2, $|L| = 1 + (k-1)(|I|-1)$. Hence $|L| \geq |I|$; otherwise, $|I| < 1$, which is not possible because $k \geq 2$ and $\circ \in I$. The theorem is now a direct consequence of the above inequality. ◀

▶ **Corollary 5.** *If* $\mathrm{IPL}(T) \ll |I|^2$ *as* $|T| \to \infty$*, then* $\zeta = 1 - o(1)$*.*

▶ **Example 6** (Perfect k-Regular Trees.)**.** These are k-regular trees in which every leaf has the same depth. Let T be one such tree with height $(h+1)$. At each depth $j \geq 1$, there are k^{j-1} nodes. So

$$|I| = 1 + \sum_{j=1}^{h} k^{j-1} = \frac{k + k^h - 2}{k-1}, \text{ and } \mathrm{IPL}(T) = \sum_{j=1}^{h} j k^{j-1} = \frac{h k^{h+1} - (h+1) k^h + 1}{(k-1)^2},$$

and we find that

$$\frac{\mathrm{IPL}(T)}{|I|^2} = \frac{h k^{h+1} - (h+1) k^h + 1}{(k + k^h - 2)^2} \sim \frac{h}{k^{h-1}}.$$

Hence $\mathrm{IPL}(T) = o\big(|I|^2\big)$, as $h \to \infty$, and Corollary 5 implies that the Haar-like basis asymptotically diagonalizes the covariance matrix of perfect k-regular trees as their height tends to infinity.

▶ **Example 7** (k-Regular Caterpillar Trees)**.** These are k-regular trees in which the parent of every leaf is a node in a central path graph. Let T be a k-regular caterpillar tree of height $(h+1)$ so that its central path has length h. Since there is only one node at depth 1, and at each depth $2 \leq j \leq h$, there are $(k-1)$ leaves and one interior node, we have

$$|I| = 2 + h, \text{ and } \mathrm{IPL}(T) = \sum_{j=1}^{h} j = \frac{h(h+1)}{2}.$$

So

$$\frac{\mathrm{IPL}(T)}{|I|^2} = \frac{h(h+1)}{2(h+2)^2} \sim \frac{1}{2},$$

as $h \to \infty$, and the lower-bound for ζ in Theorem 4 is trivial (in fact, strictly negative), and we cannot guarantee that the covariance matrix associated with T is sparsified to a significant degree as $|T| \to \infty$.

The last example shows that covariance matrices of some k-regular trees do not meet the criterion for significant sparsification given by Theorem 4; moreover, the tree in Example 6 is exceptionally constrained. So, the question remains whether trees meeting the sparsification criterion are at all common. To this end, we note that the guarantee provided by Theorem 4 is solely based on the tree's topology; i.e., the edge lengths are irrelevant. Therefore, we may investigate our remaining question by considering random k-regular trees irrespective of their edge lengths.

3.1 Interlude on Hypergeometric Functions

For a concise introduction to hypergeometric functions, see [1,23].

A hypergeometric function is one whose power series is hypergeometric; that is, its ratio of consecutive coefficients indexed by n is a rational function of n. In particular, a power series $\sum_{n=0}^{\infty} f_n\, z^n$ is hypergeometric when there are constants $a_1, \ldots, a_p, b_1, \ldots, b_q$, for some integers $(q+1) \geq p \geq 1$, such that

$$\frac{f_{n+1}}{f_n} = \frac{1}{n+1} \cdot \frac{(n+a_1)\cdots(n+a_p)}{(n+b_1)\cdots(n+b_q)}.$$

(The constraints on p and q assume that the series coefficients are not eventually zero nor undefined due to division by zero.) Accordingly, the coefficients f_n may be written in terms of the Pochhammer symbol (defined such that $(c)_0 := 1$ and $(c)_n := \prod_{i=0}^{n-1}(c+i)$ for each integer $n \geq 1$) as

$$f_n = \frac{1}{n!} \cdot \frac{(a_1)_n \cdots (a_p)_n}{(b_1)_n \cdots (b_q)_n}, \text{ for all } n \geq 0.$$

The above hypergeometric function is denoted ${}_pF_q(a_1, a_2, \ldots, a_p; b_1, b_2, \ldots, b_q; z)$. A hypergeometric function of this form is said to be *s-balanced* if

$$\sum_{j=1}^{q} b_j - \sum_{j=1}^{p} a_j = s,$$

and if $s > 0$, the power series converges at $z = 1$ [8].

The following result will be crucial to identifying dominant singularities of the generating function enumerating internal path length of k-regular trees, as well as addressing their uniqueness in the closure of the disk of convergence. In stating this result, we use the following notation for the remainder of Section 3:

$$p(t) := t(1-t)^{k-1}, \text{ for all } t \in \mathbb{C}.$$

and

$$z_k := p\left(\frac{1}{k}\right) = \frac{(k-1)^{k-1}}{k^k}.$$

▶ **Proposition 8** (Reformulation of [23, Identity (25)]). *If* $k \geq 2$ *then the hypergeometric function* ${}_{k-1}F_{k-2}(\frac{1}{k}, \ldots, \frac{k-1}{k}; \frac{k-2}{k-1}, \frac{k}{k-1}; \frac{z}{z_k})$ *is analytic in the disc* $|z| < z_k$ *and continuous in* $|z| \leq z_k$*. Furthermore,*

$${}_{k-1}F_{k-2}\left[\begin{matrix} \frac{1}{k} & \cdots & \frac{k-1}{k} & \\ \frac{2}{k-1} & \cdots & \frac{k-2}{k-1} & \frac{k}{k-1} \end{matrix}; \frac{p(t)}{p\left(\frac{1}{k}\right)}\right] = \frac{1}{1-t}, \text{ for all } 0 \leq t \leq \frac{1}{k}.$$

Proof. We attribute some of the ideas in this proof to I. Gessel [12].

Consider the functional equation

$$F(z) = 1 + z\{F(z)\}^k, \tag{4}$$

with F analytic in an open neighborhood of $z = 0$. The Lagrange Inversion Theorem implies that (4) has a unique analytic solution in some open neighborhood of $z = 0$, with coefficients given by:

$$[z^n]\, F(z) = \frac{1}{(k-1)n+1}\binom{kn}{n}. \tag{5}$$

Therefore

$$\lim_{n\to\infty} \frac{[z^{n+1}]\, F(z)}{[z^n]\, F(z)} = \frac{k^k|z|}{(k-1)^{k-1}},$$

implying that $F(z)$ has radius of convergence z_k and, due to the Vivanti-Pringsheim Theorem [16, Theorem 5.7.1], z_k is a singular point of F.

On the other hand, the ratio of consecutive terms in $F(z)$ is

$$\frac{z^{n+1}\,[z^{n+1}]\, F(z)}{z^n\,[z^n]\, F(z)} = \frac{\left(n+\frac{1}{k}\right)\cdots\left(n+\frac{k-1}{k}\right)}{\left(n+\frac{2}{k-1}\right)\cdots\left(n+\frac{k-2}{k-1}\right)\cdot\left(n+\frac{k}{k-1}\right)} \cdot \frac{z/z_k}{n+1},$$

which reveals

$$F(z) = {}_{k-1}F_{k-2}\left[\begin{matrix} \frac{1}{k} & \cdots & \frac{k-1}{k} & \\ \frac{2}{k-1} & \cdots & \frac{k-2}{k-1} & \frac{k}{k-1} \end{matrix}; \frac{z}{z_k}\right], \tag{6}$$

is the only analytic solution of (4) in the disk $|z| < z_k$. Moreover, since the balance of this hypergeometric function is

$$s = \left(\sum_{j=2}^{k-2} \frac{j}{k-1}\right) + \frac{k}{k-1} - \sum_{j=1}^{k-1} \frac{j}{k} = \frac{1}{2} > 0,$$

it follows from [8] that the series of $F(z)$ converges at $z = z_k$. Since F has non-negative coefficients, the series is absolutely convergent for all $|z| \leq z_k$, implying that F is analytic for $|z| < z_k$ and continuous for $|z| \leq z_k$.

To complete the proof, notice that $F(z) > 0$ for all $0 \leq z \leq z_k$. From the functional equation (4) we have that

$$z = \left(1 - \frac{1}{F(z)}\right)\left(\frac{1}{F(z)}\right)^{k-1} = p\left(\frac{F(z)-1}{F(z)}\right), \text{ for all } 0 \leq z \leq z_k,$$

but $p'(t) \geq 0$ for all $0 \leq t \leq k^{-1}$, with equality only at $t = k^{-1}$, and $p(k^{-1}) = z_k$. Therefore,

$$p^{-1}(t) = \frac{F(t)-1}{F(t)}, \text{ i.e., } F(t) = \frac{1}{1-p^{-1}(t)}, \text{ for all } t \in \left[0, \frac{1}{k}\right],$$

or, equivalently,

$$F\big(p(t)\big) = \frac{1}{1-t}, \text{ for all } t \in \left[0, \frac{1}{k}\right],$$

which finalizes the proof. ◀

▶ Remark 9. The sequence in (5) belongs to the class of Fuss-Catalan numbers. They are cataloged in the OEIS as A000108 ($k = 2$), A001764 ($k = 3$), A002293 ($k = 4$), and so forth.

3.2 Expectation and Variance of Internal Path Length

Corollary 5 establishes that a sufficient condition for the Haar-like basis of a k-regular tree to diagonalize its (phylogenetic) covariance matrix asymptotically, as the tree grows, is that its internal path length becomes negligible compared to the square of the number of its interior points. To assess the prevalence and extent of trees meeting this criterion, we need asymptotic estimates for the mean and variance of large, uniformly at random, k-regular trees, which is precisely what our next result addresses. We emphasize that related results exist in the literature – see [9, Proposition VII.3], [19, Theorem 3.1], and [7, Theorem 2.19]. Our new contribution is the derivation of leading asymptotic estimates – with explicit multiplicative constants – in terms of k.

▶ **Theorem 10.** *For a uniformly at random k-regular tree $\mathbb{T}$ with $|I|$ internal nodes, the expectation and variance of its internal path length satisfy the asymptotic estimates*

$$\mathbb{E}\big[\mathrm{IPL}(\mathbb{T})\big] = \sqrt{\frac{\pi k |I|^3}{2(k-1)}}\,\big(1 + O(|I|^{-1/2})\big), \tag{7}$$

and

$$\mathbb{V}\big[\mathrm{IPL}(\mathbb{T})\big] = \frac{k}{2(k-1)}\Big(\frac{10}{3} - \pi\Big)|I|^3\,\big(1 + O(|I|^{-1/2})\big). \tag{8}$$

Proof (Sketch). Given that the asymptotic formulas in the Theorem agree with those in [14] for $k = 2$, we assume henceforth that $k \geq 3$.

Let $Q(z,u)$ be the generating function for the class of k-regular trees, where the variable z marks the number of internal nodes and u the internal path length of a tree. We find using the methods of [9, Section III] that $Q(z,u)$ obeys the functional equation

$$Q(z,u) = G\big(z, Q(zu,u)\big), \text{ where } G(z,w) := 1 + z\,w^k. \tag{9}$$

For brevity, define $Q(z) := Q(z,1)$, $Q_u(z) := \frac{\partial Q}{\partial u}(z,1)$, and $Q_{uu}(z) := \frac{\partial^2 Q}{\partial u^2}(z,1)$. In particular:

$$\mathbb{E}\big[\mathrm{IPL}(\mathbb{T})\big] = \frac{[z^n]Q_u(z)}{[z^n]Q(z)}; \tag{10}$$

$$\mathbb{V}\big[\mathrm{IPL}(\mathbb{T})\big] = \frac{[z^n]Q_{uu}(z)}{[z^n]Q(z)} + \frac{[z^n]Q_u(z)}{[z^n]Q(z)} - \left(\frac{[z^n]Q_u(z)}{[z^n]Q(z)}\right)^2. \tag{11}$$

The asymptotic formulas in (7)-(8) follow from a detailed asymptotic analysis of the numerators and denominators above. To this effect, we first note from (9) that

$$Q(z) = 1 + z\{Q(z)\}^k. \tag{12}$$

In particular, the proof of Proposition 8 implies that $Q(z)$ is hypergeometric. Namely:

$$Q(z) = {}_{k-1}F_{k-2}\left[\begin{matrix} \frac{1}{k} & \cdots & \frac{k-1}{k} \\ \frac{2}{k-1} & \cdots & \frac{k-2}{k-1} & \frac{k}{k-1} \end{matrix}\,;\,\frac{z}{z_k}\right], \text{ for all } |z| \leq z_k.$$

Moreover, since $z_k = p(k^{-1})$, the following result is immediate from Proposition 8.

▶ **Corollary 11.** *If $k \geq 3$ then $Q(z_k) = \frac{k}{k-1}$; in particular, $k\, z_k\, \{Q(z_k)\}^{k-1} = 1$.*

We next show that $Q(z)$ fits the "smooth implicit function schema" [9], and determine the asymptotic order of the denominator in (10).

▶ **Lemma 12.** *z_k is the only singularity of $Q(z)$ on the disk $|z| \leq z_k$ and, locally around z_k, $Q(z)$ admits the representation*

$$Q(z) = 1 + g(z) - h(z) \cdot \sqrt{1 - \frac{z}{z_k}},$$

for functions $g(z)$ and $h(z)$ analytic near z_k. Furthermore

$$[z^n]\, Q(z) = \sqrt{\frac{k}{2\pi n^3 (k-1)^3}} \cdot z_k^{-n}\big(1 + O(n^{-1})\big). \tag{13}$$

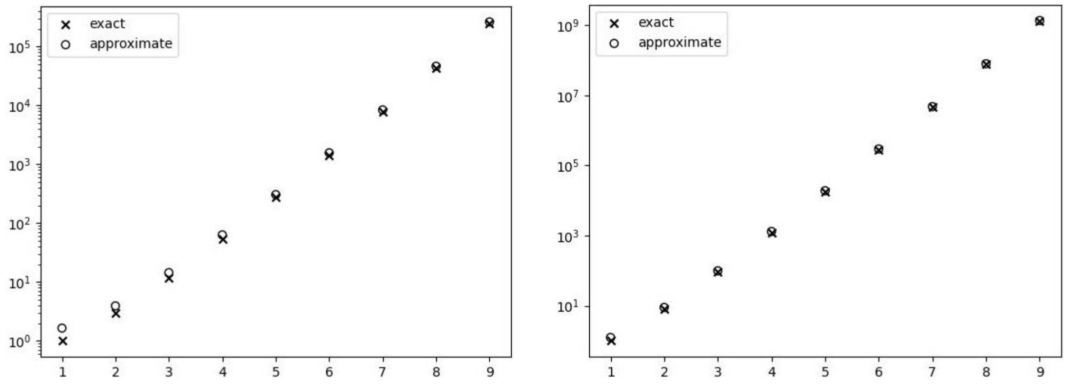
As seen in Figure 2, the asymptotic formula in (13) is highly accurate even for small values of n when $k = 3, 4$.

Proof. Define $P(z) := Q(z) - 1$. $P(0) = 0$ and $P(z) = F\big(z, P(z)\big)$, where $F(z,p) := z\{1+p\}^k$. Note that $F(0,p) = 0$ and F has only non-negative Taylor coefficients around $(0,0)$. Moreover, due to Corollary 11, $P(z_k) = \frac{1}{k-1}$, hence

$$\begin{aligned} F\big(z_k, P(z_k)\big) &= \tfrac{1}{k-1}; & F_z\big(z_k, P(z_k)\big) &= \tfrac{k^k}{(k-1)^k} \neq 0; \\ F_p\big(z_k, P(z_k)\big) &= 1; & F_{pp}\big(z_k, P(z_k)\big) &= \tfrac{(k-1)^2}{k} \neq 0. \end{aligned}$$

In particular, $F\big(z_k, P(z_k)\big) = P(z_k)$ and, since $[z^n]\, P(z) > 0$ for all $n \geq 1$, [7, Theorem 2.19] implies that that z_k is the only singularity of $P(z)$ and hence of $Q(z)$ in the disk $|z| \leq z_k$. Furthermore, it admits the singular expansion

$$P(z) = g(z) - h(z) \cdot \sqrt{1 - \frac{z}{z_k}}$$



(a) $k = 3$. **(b)** $k = 8$.

Figure 2 Plots of the first ten coefficients of the counting sequence associated with $Q(z)$, for $k = 3$ (left) and $k = 8$ (right), computed both by the exact formula and according to the asymptotic estimate in (13).

where $g(z)$ and $h(z)$ are analytic in an open neighborhood of z_k, and

$$[z^n]\,P(z) = \sqrt{\frac{z_k\,F_z\big(z_k, P(z_k)\big)}{2\pi\,F_{pp}\big(z_k, P(z_k)\big)}}\;z_k^{-n}\,n^{-3/2}\,\big(1 + \mathrm{O}(n^{-1})\big),$$

from which the lemma follows. ◀

We now address the asymptotic behavior of the numerators in (10)-(11). For this, note that by implicit differentiation in (12), we find that

$$Q_u(z) = \frac{k\,z^2\,\{Q(z)\}^{2k-1}}{\big(1 - kz\,\{Q(z)\}^{k-1}\big)^2}; \tag{14}$$

$$Q_{uu}(z) = \frac{k(7k-1)z^3\{Q(z)\}^{3k-2}}{(1 - kz\{Q(z)\}^{k-1})^3} + \frac{k^2(7k-1)z^4\{Q(z)\}^{4k-3}}{(1 - kz\{Q(z)\}^{k-1})^4} + \frac{5k^3(k-1)z^5\{Q(z)\}^{5k-4}}{(1 - kz\{Q(z)\}^{k-1})^5}. \tag{15}$$

(For details about these derivations refer to Appendix A of [22].) Our next result implies that z_k is the only singularity of the above partial derivatives in the disk $|z| \le z_k$.

▶ **Lemma 13.** *The equation* $kz\,\{Q(z)\}^{k-1} = 1$, *with* $|z| \le z_k$, *has only* z_k *as a solution.*

Proof. Per Corollary 11, we know that z_k solves the equation $kz\,\{Q(z)\}^{k-1} = 1$. On the other hand, because $Q(z)$ has non-negative coefficients, $|kz\,\{Q(z)\}^{k-1}| < k\,z_k\,\{Q(z_k)\}^{k-1} = 1$ for $|z| < z_k$; in particular, any solution to the equation must lie on the circle $|z| = z_k$. But, if $|z| = z_k$ and $z \ne z_k$ then $|Q(z)| < Q(|z|) = Q(z_k)$ because Q has all powers of z with strictly positive coefficients. Hence, $|kz\{Q(z)\}^{k-1}| = kz_k|Q(z)|^{k-1} < kz_k\{Q(z_k)\}^{k-1} = 1$, which completes the proof of the lemma. ◀

Notice that the generating functions in (14)-(15) share a common form, which as we see next may be exploited to yield a general asymptotic expansion of their coefficients.

▶ **Lemma 14.** *If $f : \mathbb{C} \to \mathbb{C}$ is an entire analytic function such that $f(z_k) \neq 0$, and $a \geq 0$ and $b \geq 1$ are integers, then the function*

$$[z^n]\,\frac{f(z)\{Q(z)\}^a}{(1-kz\{Q(z)\}^{k-1})^b} = \frac{f(z_k)}{2^{b/2}\,\Gamma(b/2)}\left(\frac{k}{k-1}\right)^{a+b/2} n^{(b-2)/2}\,z_k^{-n}\left(1+\mathrm{O}\left(n^{-1/2}\right)\right).$$

Proof (Sketch). For further details, refer to the proof of [22, Lemma 3.2.5].

Define

$$P(z) := \frac{A(z)}{\big(1-B(z)\big)^b},$$

where $A(z) := f(z)\{Q(z)\}^a$ and $B(z) := kz\{Q(z)\}^{k-1}$. By Lemma 12, we can write

$$Q(z) = g(z) - h(z)\sqrt{1-\frac{z}{z_k}},$$

where $g(z)$ and $h(z)$ are analytic in an open neighborhood of z_k. In particular, locally around z_k, we have that

$$f(z) = f_0 + \mathrm{O}\left(1-\frac{z}{z_k}\right);\quad g(z) = g_0 + \mathrm{O}\left(1-\frac{z}{z_k}\right);\ \text{and}\ h(z) = h_0 + \mathrm{O}\left(1-\frac{z}{z_k}\right);$$

where f_0, g_0, and h_0 are $f(z_k)$, $g(z_k)$, and $h(z_k)$, respectively.

To determine an asymptotic formula for the coefficients of $P(z)$, we first seek singular expansions for its numerator and denominator. For this, observe that locally around z_k:

$$A(z) = \alpha_0 + \mathrm{O}\left(\sqrt{1-\frac{z}{z_k}}\right);$$

$$B(z) = \beta_0 - \beta_1\sqrt{1-\frac{z}{z_k}} + \mathrm{O}\left(1-\frac{z}{z_k}\right);$$

where $\alpha_0 := f_0 g_0^a$, $\beta_0 := kz_k g_0^{k-1}$, and $\beta_1 := k(k-1)z_k g_0^{k-2}h_0$. But we know from Lemma 13 that $B(z_k) = 1$; in particular, $\beta_0 = 1$, and as a result

$$P(z) = \frac{f_0 g_0^a}{\{k(k-1)z_k g_0^{k-2}h_0\}^b}\left(1-\frac{z}{z_k}\right)^{-b/2}\left(1+\mathrm{O}\left(\sqrt{1-\frac{z}{z_k}}\right)\right).$$

Finally, note that $g_0 = Q(z_k)$, hence $g_0 = \frac{k}{k-1}$ due to Corollary 11. On the other hand, due to Lemma 12, $h_0 = \frac{1}{k-1}\sqrt{\frac{2k}{k-1}}$. Therefore

$$P(z) = \frac{f(z_k)}{2^{b/2}}\cdot\left(\frac{k}{k-1}\right)^{a+b/2}\left(1-\frac{z}{z_k}\right)^{-b/2}\left(1+\mathrm{O}\left(\sqrt{1-\frac{z}{z_k}}\right)\right),$$

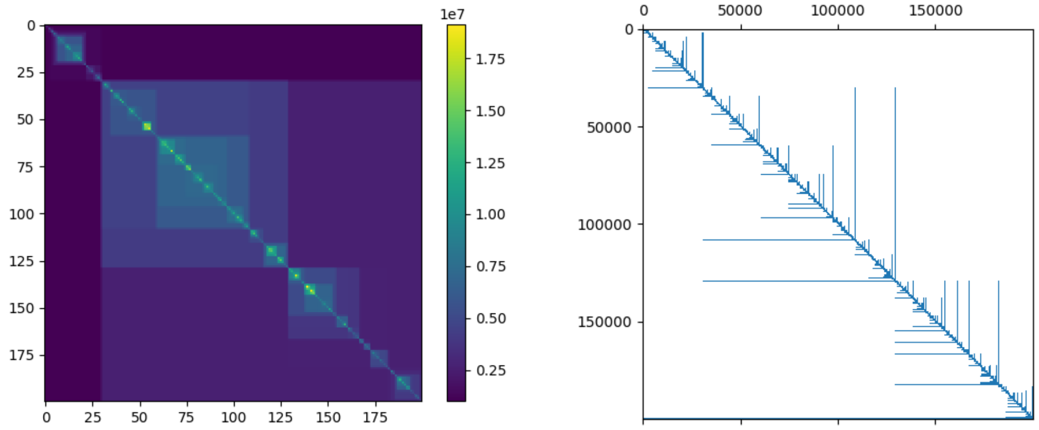
and the lemma follows from [9, Theorem VI.4]. ◀

Finally, by applying Lemma 13 to each of the terms in (14)-(15), we obtain that

$$[z^n]\,Q_u(z,1) = \frac{k}{2(k-1)^2}\cdot z_k^{-n}(1+\mathrm{O}(n^{-1/2}));$$

$$[z^n]Q_{uu}(z) = \frac{5k}{6(k-1)^2}\sqrt{\frac{2kn^3}{\pi(k-1)}}\cdot z_k^{-n}\left(1+\mathrm{O}\big(n^{-1/2}\big)\right);$$

from which the asymptotic formulas for the expected value and variance of the internal path length of a uniformly at random k-regular tree in Theorem 10 follow. ◀



(a) Dense covariance matrix.

(b) Sparsified covariance matrix.

Figure 3 (a) Heat-map visualization of a dense covariance matrix C, and (b) sparsity pattern of $\Phi' C \Phi$ for a 3-regular tree with 200,001 leaves. The dense matrix has over 40 billion non-zero entries. The sparse matrix, on the other hand, has only about 0.03% as many non-zero entries. The heatmap of the dense matrix was produced downsampling the sparse representation by a factor of 1000, prior to undoing the change-of-basis.

3.3 Sparsification of Large, Random, k-Regular Covariance Matrices

So far, we have seen that the Haar-like wavelets can partially sparsify the (phylogenetic) covariance matrix of a k-regular tree by changing the basis, obtaining a lower bound on the proportion of vanishing entries under the basis change. A remaining challenge is determining whether trees meeting the criteria for a high degree of sparsification are common. In this section, we find that such trees are, in fact, abundant and that large random k-regular trees are highly sparsified by the Haar-like wavelets with overwhelmingly high probability.

▶ **Corollary 15.** *Let $k \geq 2$ and $\mathbb{T}$ be a uniformly at random k-regular tree with $|I|$ internal nodes, and C its covariance matrix. If Φ is the Haar-like matrix associate with $\mathbb{T}$, and ζ the fraction of vanishing entries in $\Phi' C \Phi$, then $\zeta \to 1$ in probability, as $|I| \to \infty$.*

Proof. The argument mirrors the one for [14, Corollary 3.8].

Let μ and σ^2 denote the expectation and variance of the internal path length of $\mathbb{T}$, respectively. Cantelli's inequality states that $\mathbb{P}\big(\mathrm{IPL}(\mathbb{T}) > \mu + t\sigma\big) \leq (1+t^2)^{-1}$, for all $t > 0$. But, from Theorem 10, we know that $\mu + t\sigma = \Omega(t|I|^{3/2})$; that is, there is a constant $c > 0$ such that $\mathbb{P}\big(\mathrm{IPL}(T) > c\,t|I|^{3/2}\big) \leq (1+t^2)^{-1}$, or equivalently

$$\mathbb{P}\left(\frac{\mathrm{IPL}(\mathbb{T})}{|I|^2} \leq \frac{c\,t}{\sqrt{|I|}}\right) \geq \frac{t^2}{1+t^2}, \text{ for all } t > 0.$$

The result follows by choosing $t \to \infty$ so that $t = o(\sqrt{|I|})$. ◀

4 Discussion

Phylogenetic covariance matrices are often large and dense to the point of being computationally unmanageable; however, we have demonstrated that expressing them in the Haar-like basis can significantly sparsify them. In particular, we have shown that the covariance matrix of a random k-regular tree will be highly sparsified with overwhelmingly high probability as

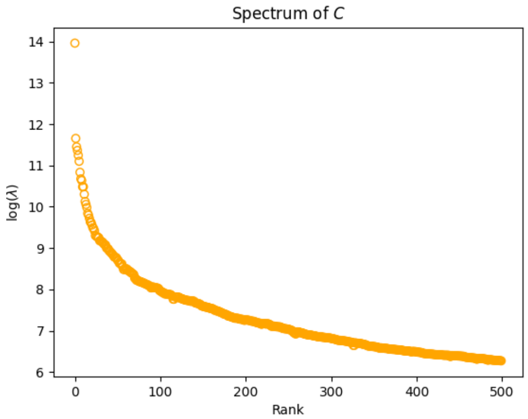


Figure 4 The largest 500 eigenvalues of a covariance matrix C associated with a 3-regular tree with 200,001 leaves. The spectrum was computed from its sparse representation, obtained after changing to the Haar-like basis.

the size of the tree tends to infinity. In this section, we illustrate the process of sparsifying a large phylogenetic covariance matrix and discuss some of the practical considerations of the method.

Consider Figure 3, which illustrates the application of the Haar-like wavelet transform on a random 3-regular tree with 200,001 leaves. The tree was generated by simulating a Galton-Watson process until a desired population size (i.e., number of leaves) was reached. We find that, although the original, dense matrix contains over 40 billion non-zero entries, the resulting sparse matrix contains a comparatively meager 11.3 million; that is, over 99.97% of the entries in the original matrix were zeroed by the transform. Importantly, because $\Phi' C \Phi$ and C are similar matrices, we can quickly calculate the spectrum of C from its sparse representation, illustrated in Figure 4. While this random tree model may not precisely induce actual phylogenetic trees [2–4], we emphasize that the sparsification observed results in an overwhelming part from the tree's topology, irrespective of its edge lengths (which effectively store covariances). Hence, the simulated data gives reasonable insight into the performance of this technique in general.

As already mentioned, the degree of sparsification guaranteed by Theorem 4 is conservative. For one, it does not account for the fact that the Haar-like wavelets are orthonormal, which would suggest that at least some entries associated with two wavelets ψ_u and ψ_v, where u is an ancestor of v, vanish. In fact, if a k-regular tree is *trace balanced* [14] (that is, for all $v \in I$ and $i, j \in L(v)$, $\ell^*(i, v) = \ell^*(j, v)$), Theorem 2 implies that the Haar-like wavelets fully diagonalize its covariance matrix. As a result, we often see a better degree of sparsification than what is guaranteed by Theorem 4.

Our analysis shows that the technique of sparsifying dense phylogenetic covariance matrices by a change-of-basis with the Haar-like wavelets extends to the broader class of k-regular trees. While further work is required to verify the exact performance of this method on random k-ary trees (i.e., ones for which each interior node contains *at most* k children), well-known properties of generating functions enumerating simple varieties of trees suggest that such a generalization is possible by the methods employed here. Comparison with k-regular trees indicates sparsification for what one may call "almost" k-regular trees (i.e., k-ary trees which are k-regular except at a relatively small number of internal nodes); however, initial investigation suggests that the worst-case k-ary tree might result in poor sparsification.

To conclude, it is worth noting that the k-regular and k-ary trees belong to the class of "simply generated trees" [17], a generalization which may prove helpful for further expanding our methodology. Exploring this avenue of research could grant access to new datasets characterized by more intricate yet richer hierarchical structures.

References

1 K. M. Abadir. An introduction to hypergeometric functions for economists. *Econometric Reviews*, 18(3):287–330, 1999.

2 D. Aldous. Probability distributions on cladograms. In D. Aldous and R. Pemantle, editors, *Random Discrete Structures*, pages 1–18, New York, NY, 1996. Springer New York.

3 D. Aldous and B. Pittel. The critical beta-splitting random tree: Heights and related results, 2023. `arXiv:2302.05066`.

4 D. J. Aldous. The critical beta-splitting random tree II: Overview and open problems, 2023. `arXiv:2303.02529`.

5 L. L. Cavalli-Sforza and A. W. Edwards. Phylogenetic analysis: models and estimation procedures. *Evolution*, 21(3):550, 1967. `doi:10.2307/2406616`.

6 C. Dellacherie, S. Martinez, and J. San Martín. *Inverse M-Matrices and Ultrametric Matrices*, volume 2118 of *Lecture Notes in Mathematics*. Springer, 2014.

7 M. Drmota. *Random Trees: An Interplay between Combinatorics and Probability*. Springer-Verlag/Wein, 2009.

8 R. J. Evans and D. Stanton. Asymptotic formulas for zero-balanced hypergeometric series. *SIAM J. Math. Anal.*, 1984. `doi:10.1137/0515078`.

9 P. Flajolet and R. Sedegwick. *Analytic Combinatorics*. Cambridge University Press, 2009. URL: `http://www.cambridge.org/uk/catalogue/catalogue.asp?isbn=9780521898065`.

10 J. Fukuyama, P. J. McMurdie, L. Dethlefsen, D. A. Relman, and S. Holmes. Comparisons of distance methods for combining covariates and abundances in microbiome studies. *Biocomputing*, pages 213–224, 2012. URL: `http://psb.stanford.edu/psb-online/proceedings/psb12/fukuyama.pdf`.

11 Matan Gavish, Boaz Nadler, and Ronald R. Coifman. Multiscale wavelets on trees, graphs and high dimensional data: theory and applications to semi supervised learning. In *Proceedings of the 27th International Conference on International Conference on Machine Learning*, ICML'10, pages 367–374, Madison, WI, USA, 2010. Omnipress.

12 I. Gessel. How can I verify this family of values for hypergeometric functions? MathOverflow, september 12 2023. URL: `https://mathoverflow.net/q/454420`.

13 E. Gorman and M. E. Lladser. Interpretable metric learning in comparative metagenomics: The adaptive Haar-like distance. PLoS Comput Biol, 2024 (to appear). URL: `https://www.biorxiv.org/content/10.1101/2023.09.27.559681v1`.

14 E. Gorman and M. E. Lladser. Sparsification of large ultrametric matrices: insights into the microbial Tree of Life. *Proc. R. Soc. A*, 479:20220847, 2023. `doi:10.1098/rspa.2022.0847`.

15 L. J. Harmon. *Phylogenetic Comparative Methods*. CreateSpace Independent Publishing Platform, 2019.

16 E. Hille. *Analytic function theory. Vol. 1*. Introduction to Higher Mathematics. Ginn and Company, 1959.

17 Svante Janson. Simply generated trees, conditioned Galton–Watson trees, random allocations and condensation. *Probability Surveys*, 9:103–252, 2012.

18 S. Martinez, G. Michon, and J. San Martín. Inverse of strictly ultrametric matrices are of Stieltjes type. *SIAM J. Matrix Anal. Appl.*, 15(1):98–106, 1994. `doi:10.1137/s0895479891217011`.

19 A. Meir and J. W. Moon. On the altitude of nodes in random trees. *Canadian Journal of Mathematics*, 30, 1978.

20 R. Nabben and R. S. Varga. A linear algebra proof that the inverse of a strictly ultrametric matrix is a strictly diagonally dominant stieltjes matrix. *SIAM J. Matrix Anal. Appl.*, 15(1):107–113, 1994. `doi:10.1137/s0895479892228237`.

21 S. Pavoine, A.-B. Dufour, and D. Chessel. From dissimilarities among species to dissimilarities among communities: a double principal coordinate analysis. *Journal of Theoretical Biology*, 228(4):523–537, 2004. `doi:10.1016/j.jtbi.2004.02.014`.

22 Sean Svihla. Sparsification of covariance matrices of k-regular trees. Master's thesis, The University of Colorado, 2024.

23 E. W. Weisstein. Hypergeometric function. `https://mathworld.wolfram.com/HypergeometricFunction.html`. Accessed: September 2023.

24 Q. Zhu, U. Mai, and W. Pfeiffer et al. Phylogenomics of 10,575 genomes reveals evolutionary proximity between domains bacteria and archaea. *Nature Communications*, 10:5477, 2019.