



# Random self-similar trees and a hierarchical branching process

Yevgeniy Kovchegov<sup>a,\*</sup>, Ilya Zaliapin<sup>b</sup>

<sup>a</sup> Department of Mathematics, Oregon State University, Corvallis, OR 97331, USA

<sup>b</sup> Department of Mathematics and Statistics, University of Nevada, Reno, NV, 89557-0084, USA

Received 5 May 2017; received in revised form 13 July 2018; accepted 18 July 2018

Available online 7 August 2018

---

## Abstract

We study self-similarity in random binary rooted trees. In a well-understood case of Galton–Watson trees, a distribution on a space of trees is said to be self-similar if it is invariant with respect to the operation of pruning, which cuts the tree leaves. This only happens for the critical Galton–Watson tree (a constant process progeny), which also exhibits other special symmetries. We extend the prune-invariance setup to arbitrary binary trees with edge lengths. In this general case the class of self-similar processes becomes much richer and covers a variety of practically important situations. The main result is construction of the *hierarchical branching processes* that satisfy various self-similarity definitions (including mean self-similarity and self-similarity in edge-lengths) depending on the process parameters. Taking the limit of averaged stochastic dynamics, as the number of trajectories increases, we obtain a deterministic system of differential equations that describes the process evolution. This system is used to establish a phase transition that separates fading and explosive behavior of the average process progeny. We describe a class of *critical Tokunaga* processes that happen at the phase transition boundary. They enjoy multiple additional symmetries and include the celebrated critical binary Galton–Watson tree with independent exponential edge length as a special case. Finally, we discuss a duality between trees and continuous functions, and introduce a class of *extreme-invariant* processes, constructed as the Harris paths of a self-similar hierarchical branching process, whose local minima has the same (linearly scaled) distribution as the original process.

© 2018 Elsevier B.V. All rights reserved.

MSC: primary 60C05; secondary 82B99

---

\* Corresponding author.

E-mail addresses: [kovchegy@math.oregonstate.edu](mailto:kovchegy@math.oregonstate.edu) (Y. Kovchegov), [zal@unr.edu](mailto:zal@unr.edu) (I. Zaliapin).

## 1. Introduction

Nature commonly exhibits dendritic structures, both static and dynamic, that can be represented by tree graphs [1,19,27]. Examples from diverse applications, together with a review of related coalescence and branching models can be found in Aldous [1], Berestycki [2], Bertoin [3], Evans [8], Le Gall [16], and Pitman [21]. Despite their apparent diversity, a number of rigorously studied dendritic structures possess structural self-similarity, which often allows a low-dimensional parameterization [11,19,20,26]. An illuminating example is the combinatorial structure of river networks, which is closely approximated by a two-parametric Tokunaga self-similar model with parameters that are independent of river's geographic location [6,20,25,29]. Tree self-similarity has been studied primarily in terms of the average values of selected branch statistics, and rigorous results have been obtained only for a very special classes of Markov trees (e.g., binary Galton–Watson trees with no edge lengths, as in [4]). At the same time, solid empirical evidence motivates a search for a flexible class of self-similar models that would encompass a variety of observed combinatorial and metric structures and rules of tree growth. We introduce here a general concept of self-similarity that accounts for both combinatorial and metric tree structure (Section 3.5, Definition 12) and describe a model (Section 5), called *hierarchical branching process*, that generates a broad range of self-similar trees (Theorem 4) and includes the critical binary Galton–Watson tree with exponential edge lengths as a special case (Theorem 8). We study *time-invariant* tree distributions, which is a convenient generalization of Markov growth (Theorem 7). We also introduce a class of critical self-similar *Tokunaga processes* (Section 5.7) that enjoy additional symmetries — their edge lengths are i.i.d. random variables (Proposition 9), and subtrees of large Tokunaga trees reproduce the probabilistic structure of the entire random tree space (Proposition 11). The duality between planar trees and continuous functions [9,21,28] allows us using the hierarchical branching process to construct a novel class of time series that satisfy the *extreme-invariance property*: the distribution of their local minima is the same as that of the original series (Section 4).

The paper is organized as follows. Section 2 introduces the main definitions, including the Horton–Strahler order of a tree, tree pruning, and a related concept of prune-invariance. Self-similarity for trees with edge lengths is defined in Section 3. The duality between trees and continuous functions is reviewed in Section 4. In particular, we define here *extreme-invariant* processes that are equivalent to self-similar trees. The main results are presented in Section 5. Section 5.1 introduces a *hierarchical branching process* that generates a rich collection of self-similar trees. The hydrodynamic limit for dynamics of the average numbers of Horton–Strahler branches is established in Section 5.2. The properties of *criticality* and *time-invariance* are defined in Section 5.3 and explored in a self-similar processes in Section 5.4. Critical Galton–Watson process and critical Tokunaga processes, which generate the most intriguing examples of self-similar trees, are discussed in Sections 5.6, 5.7. Section 6 discusses the combinatorial structure of the critical Tokunaga process. Section 7 concludes with two open problems.

## 2. Random trees

The focus of this paper is on finite unlabeled rooted reduced planted binary trees with no planar embedding. The space of such trees, which includes the *empty tree*  $\phi$  comprised of a single root vertex and no edges, is denoted by  $\mathcal{T}$ .

The existence of the root vertex imposes the parent–offspring relationship between each pair of the connected vertices in a tree  $T \in \mathcal{T}$ : the one closest to the root is called *parent*, and the

other — *offspring*. The absence of planar embedding in this context means the absence of order between the two offspring of the same parent. A tree is called *reduced* if it has no vertices of degree 2; such trees are also called *full binary trees*. A tree is called *planted* if its root has degree 1. Accordingly, there are three types of vertices in a tree from  $\mathcal{T} \setminus \{\phi\}$ : internal vertices of degree 3, leaves (degree 1) and the root (degree 1). The operation of *series reduction* removes each degree-two vertex of a binary tree by merging its adjacent edges into one. Series reduction turns a rooted binary tree into a reduced rooted binary tree. The edges of a tree from  $\mathcal{T}$  may be assigned positive lengths. The space of trees from  $\mathcal{T}$  with edge lengths is denoted by  $\mathcal{L}$ .

Any tree from  $\mathcal{T}$  or  $\mathcal{L}$  can be embedded (and represented graphically) in a plane by selecting an order for each pair of offspring of the same parent. The space of embedded trees from  $\mathcal{T}$  (and respectively  $\mathcal{L}$ ) is denoted  $\mathcal{T}_{\text{plane}}$  (and respectively  $\mathcal{L}_{\text{plane}}$ ). Examples of trees from  $\mathcal{L}_{\text{plane}}$  are found in the bottom row of Fig. 1. Choosing different embeddings for the same tree  $T \in \mathcal{T}$  (or  $T \in \mathcal{L}$ ) leads, in general, to different trees from  $\mathcal{T}_{\text{plane}}$  (or  $\mathcal{L}_{\text{plane}}$ ). Sometimes we focus on the combinatorial tree  $\text{SHAPE}(T) \in \mathcal{T}$ , which retains the branching structure of  $T$  while omitting its edge lengths and embedding.

2.1. Tree pruning and related concepts

The concept of self-similarity is related to the pruning operation [4,11,20]. *Pruning* (aka Horton pruning) of a tree is an onto function  $\mathcal{R} : \mathcal{T} \rightarrow \mathcal{T}$ , whose value  $\mathcal{R}(T)$  for a tree  $T \neq \phi$  is obtained by removing the leaves and their parental edges from  $T$ , followed by series reduction. We also set  $\mathcal{R}(\phi) = \phi$ .

The pruning is also well defined for trees with edge lengths ( $\mathcal{L}$ ), where series reduction adds the lengths of merging edges, and for planar trees ( $\mathcal{T}_{\text{plane}}, \mathcal{L}_{\text{plane}}$ ), where the embedding of the remaining part of a tree is unaffected by pruning. Pruning is illustrated in Fig. 1.

Pruning induces a contracting map on  $\mathcal{T}$ . The trajectory of each tree  $T$  under  $\mathcal{R}(\cdot)$  is uniquely determined and finite:

$$T \equiv \mathcal{R}^0(T) \rightarrow \mathcal{R}^1(T) \rightarrow \dots \rightarrow \mathcal{R}^k(T) = \phi, \tag{1}$$

with the empty tree  $\phi$  as the (only) fixed point. The pre-image  $\mathcal{R}^{-1}(T)$  of any non-empty tree  $T$  consists of an infinite collection of trees. It is natural to think of the distance to  $\phi$  under the pruning map and introduce the respective notion of tree order [10,24] (see Fig. 1).

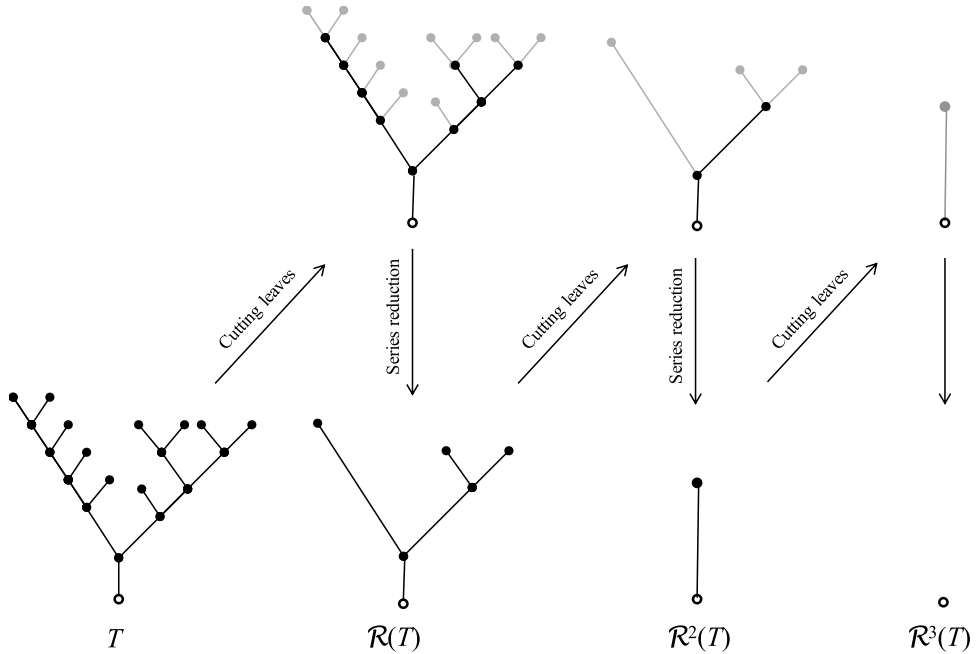
**Definition 1** (*Horton–Strahler Order of a Tree*). The Horton–Strahler order  $k(T) \in \mathbb{Z}_+$  of a tree  $T \in \mathcal{T}$  is defined as the minimal number of prunings necessary to eliminate the tree:

$$k(T) = \min_{k \geq 0} \{ \mathcal{R}^k(T) = \phi \}.$$

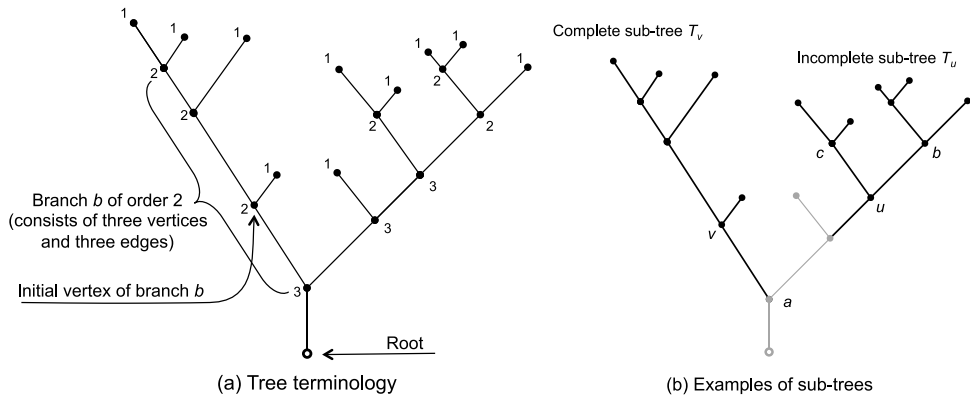
The definition of order is based on the combinatorial shape of a tree. Accordingly, the order of a tree  $T$  from either of spaces  $\mathcal{T}_{\text{plane}}, \mathcal{L}$ , or  $\mathcal{L}_{\text{plane}}$  is that of  $\text{SHAPE}(T)$ . The definition implies, in particular, that the order of the empty tree is  $k(\phi) = 0$ , because  $\mathcal{R}^0(\phi) = \phi$ . Most of our discussion will be focused on trees with orders  $k > 0$ , and we often assume that the empty tree  $\phi$  has zero probability.

Pruning partitions the tree space  $\mathcal{T}$  into exhaustive and mutually exclusive set of subspaces  $\mathcal{H}_K$  of trees of order  $K \geq 0$  such that  $\mathcal{R}(\mathcal{H}_{K+1}) = \mathcal{H}_K$ . Here  $\mathcal{H}_0 = \{\phi\}$ ,  $\mathcal{H}_1$  consists of a single tree comprised of a root and a leaf connected by an edge, and all other subspaces  $\mathcal{H}_K, K \geq 2$ , consist of an infinite number of trees.

**Definition 2** (*Horton–Strahler Terminology*). We introduce the following definitions related to the Horton–Strahler order of a tree (see Fig. 2):



**Fig. 1.** Example of pruning and Horton–Strahler ordering for a tree  $T \in \mathcal{L}_{\text{plane}}$ . The figure shows the two stages of pruning operation — cutting the leaves (top row), and consecutive series reduction (bottom row). The initial tree  $T$  is shown in the leftmost position of the bottom row. The gray color in the top row depicts vertices and edges being pruned. The order of the tree is  $k(T) = 3$ , since it is eliminated in three prunings,  $\mathcal{R}^3(T) = \phi$ .



**Fig. 2.** Illustration of the Horton–Strahler terminology (Definition 2) in a tree  $T$  of order  $k(T) = 3$ . (a) A tree root, a branch, and an initial vertex of a branch. The numbers indicate the Horton–Strahler orders of the vertices. In particular, we have  $N_1 = 10$ ,  $N_2 = 3$ ,  $N_3 = 1$ , and  $N_{1,2} = 3$ ,  $N_{1,3} = 1$ ,  $N_{2,3} = 1$ . (b) Examples of a complete ( $T_v$ ) and incomplete ( $T_u$ ) subtrees. The subtree  $T_u$  is incomplete since  $u$  is not the initial vertex of a branch. The tree shown in this panel has four complete subtrees of order  $k \geq 2$ :  $T_v$  of order 2,  $T_b$  of order 2,  $T_c$  of order 2, and  $T_a = T$  of order 3.

- (1) For any non-root vertex  $v$  in  $T \in \mathcal{T} \setminus \{\phi\}$ , a *subtree*  $T_v$  of  $T$  is defined as the only subtree in  $T$  rooted at the parental vertex  $p(v)$  of  $v$ , and comprised by  $v$  and all its descendant vertices together with their parental edges (Fig. 2b).
- (2) The Horton–Strahler order  $k(v)$  of a vertex  $v \in T$  coincides with the order of the subtree  $T_v \in T$  (Fig. 2a).
- (3) The parental edge of a vertex has the same order as the vertex.
- (4) A connected sequence of vertices of the same order together with their parental edges is called a *branch* (Fig. 2a).
- (5) The branch vertex closest to the root is called the *initial* vertex of a branch (Fig. 2a).
- (6) For the initial vertex  $v$  of a branch of order  $K \leq k(T)$ , a subtree of  $T_v$  is called a *complete subtree* of order  $K$  (Fig. 2b). The single complete subtree of order  $k(T)$  coincides with  $T$ . (All subtrees of order  $k = 1$  are complete.)

**Remark 1.** Equivalently, the Horton–Strahler ordering can be done by hierarchical counting [4, 5,10,19,20,24]. In this approach, each leaf is assigned order  $k(\text{leaf}) = 1$ . An internal vertex  $p$  whose children have orders  $i$  and  $j$  is assigned the order

$$k(p) = \max(i, j) + \delta_{ij} = \lfloor \log_2(2^i + 2^j) \rfloor,$$

where  $\delta_{ij}$  is the Kronecker’s delta and  $\lfloor x \rfloor$  denotes the maximal integer less than or equal to  $x$ . The Horton–Strahler order of the tree is  $k(T) = \max_v k(v)$ , where the maximum is taken over all non-root vertices of  $T \in \mathcal{T} \setminus \{\phi\}$ .

### 2.2. Labeling tree vertices

Sometimes we will need to label the vertices and edges of a tree (e.g., for selecting a branch or vertex uniformly). The vertices of a planar tree can be labeled by numbers  $1, \dots, \#T$  ( $\#T$  denoting the total number of vertices in  $T$ ) in order of depth-first search. We also assume that label of the parental edge for each vertex is taken from that vertex.

For a tree with no embedding, labeling is done by selecting a suitable embedding and then using the depth-first search labeling as above. Such embedding should be properly aligned with the pruning operation, as we describe in the following definition.

**Definition 3 (Proper Embedding).** An embedding function  $\text{EMBED} : \mathcal{T} \rightarrow \mathcal{T}_{\text{plane}} (\mathcal{L} \rightarrow \mathcal{L}_{\text{plane}})$  is called *proper* if for any  $T \in \mathcal{T} (T \in \mathcal{L})$

$$\mathcal{R}(\text{EMBED}(T)) = \text{EMBED}(\mathcal{R}(T)),$$

where the pruning on the left-hand side is in  $\mathcal{T}_{\text{plane}} (\mathcal{L}_{\text{plane}})$  and pruning on the right-hand side is in  $\mathcal{T} (\mathcal{L})$ .

A proper embedding for a tree with no edge lengths can be done using the following induction construction. A tree of order  $k = 1$  assumes a unique embedding. A tree of order  $k = 2$  is embedded by branching all its side-branches of order 1 to the right. Assuming there exists a proper embedding for trees of order  $k \leq K$ , we construct the labeling for a tree of order  $K + 1$ . All its side-branches (of any order) branch to the right. To embed the (only) two merging complete subtrees,  $\tau_1 \neq \tau_2$ , of order  $K$ , we consider their farthest non-identical pruning descendants: trees  $d_i = \mathcal{R}^k(\tau_i), i = 1, 2$  obtained by the maximal possible number  $k$  of pruning iterations such that  $d_1 \neq d_2$ . The number  $0 \leq k \leq K - 2$  is well defined since all trees of order 1,

which is the ultimate pruning limit, coincide. By construction, the trees  $d_i$  differ only by the number of side-branches of order 1 attached to the tree  $d_0 = \mathcal{R}^{k+1}(\tau_i)$ , which already has proper embedding. Consider the numbers of order-1 side-branches within each edge of  $d_0$ , in the order of its labeling:  $(n_1^{(i)}, \dots, n_{\#d_0}^{(i)})$ . The tree whose sequence has the smallest first non-coinciding number, will branch to the right.

A proper embedding for a tree  $T \in \mathcal{L}$  with edge length is constructed in the same fashion, with the only correction. From the two merging complete subtrees of order  $K$  with the same combinatorial structure, the one with the shortest root edge branches to the right. This definition covers the situation of atomless length distribution, which is of primary interest to us.

### 3. Tree self-similarity

This section defines self-similarity for combinatorial and metric trees. The term *self-similarity* is associated with invariance of a tree distribution with respect to the pruning operation  $\mathcal{R}$  introduced in Section 2.1. The prune-invariance alone, however, is often insufficient to generate interesting families of trees. This necessitates an additional property — *coordination* of conditional measures on subspaces of trees of a given order. Coordination together with prune-invariance constitutes the self-similarity studied in this work.

We start in Section 3.1 with a weak form of self-similarity that only considers the average values of selected branch statistics; it was introduced in [11]. Section 3.3 discusses a stronger version of self-similarity that operates with tree distributions. Self-similarity of metric trees is presented in Section 3.5.

#### 3.1. Mean self-similarity of a combinatorial tree

Let  $\mathcal{H}_K \subset \mathcal{T}$  be the subspace of trees of Horton–Strahler order  $K \geq 0$ . Naturally,  $\mathcal{H}_K \cap \mathcal{H}_{K'} = \emptyset$  if  $K \neq K'$ , and  $\bigcup_{K \geq 1} \mathcal{H}_K = \mathcal{T}$ . Consider a set of conditional probability measures  $\{\mu_K\}_{K \geq 0}$  each of which is defined on  $\mathcal{H}_K$  by  $\mu_K(T) = \mu(T|T \in \mathcal{H}_K)$  and let  $p_K = \mu(\mathcal{H}_K)$ . Then  $\mu$  is represented as a mixture of the conditional measures:

$$\mu = \sum_{K=1}^{\infty} p_K \mu_K. \tag{2}$$

We write  $E_K(\cdot)$  for the mathematical expectation with respect to  $\mu_K$ . Let  $N_k = N_k[T]$  denotes the number of branches of order  $k$  in a tree  $T \in \mathcal{T}$  (see Fig. 2a). We define the *average Horton numbers* for subspace  $\mathcal{H}_K$  as

$$\mathcal{N}_k[K] = E_K(N_k), \quad 1 \leq k \leq K, \quad K \geq 1.$$

Let  $N_{i,j} = N_{i,j}[T]$  denote the number of instances when an order- $i$  branch merges with an order- $j$  branch,  $1 \leq i < j$ , in a tree  $T$  (see Fig. 2a). Such branches are referred to as *side-branches* of order  $\{i, j\}$ . Consider the respective expectation  $\mathcal{N}_{i,j}[K] := E_K(N_{i,j})$ . The *Tokunaga coefficients*  $T_{i,j}[K]$  for subspace  $\mathcal{H}_K$  are defined as

$$T_{i,j}[K] = \frac{\mathcal{N}_{i,j}[K]}{\mathcal{N}_j[K]}, \quad 1 \leq i < j \leq K. \tag{3}$$

The Tokunaga coefficient  $T_{i,j}[K]$  hence reflects the average number of side-branches of order  $\{i, j\}$  per branch of order  $j$  in a tree of order  $K$ .

Next, we introduce a property that ensures independence of the side-branch structure of a tree order.

**Definition 4 (Mean Coordination).** A set of measures  $\{\mu_K\}_{K \geq 1}$  on  $\{\mathcal{H}_K\}_{K \geq 1}$  is called *mean coordinated* if  $T_{i,j} := T_{i,j}[K]$  for all  $K \geq 2$  and  $1 \leq i < j \leq K$ . A measure  $\mu$  on  $\mathcal{T}$  is called *mean coordinated* if the respective conditional measures  $\{\mu_K\}$ , as in Eq. (2), are mean coordinated.

For a mean coordinated measure  $\mu$ , the Tokunaga matrix  $\mathbb{T}_K$  is a  $K \times K$  matrix

$$\mathbb{T}_K = \begin{bmatrix} 0 & T_{1,2} & T_{1,3} & \dots & T_{1,K} \\ 0 & 0 & T_{2,3} & \dots & T_{2,K} \\ 0 & 0 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & 0 & T_{K-1,K} \\ 0 & 0 & \dots & 0 & 0 \end{bmatrix},$$

which coincides with the restriction of any larger-order Tokunaga matrix  $\mathbb{T}_M$ ,  $M > K$ , to the first  $K \times K$  entries.

**Definition 5 (Toeplitz Property).** A set of measures  $\{\mu_K\}_{K \geq 1}$  on  $\{\mathcal{H}_K\}_{K \geq 1}$  is said to satisfy the *Toeplitz property* if  $T_{i,j}[K] = T_{j-i}[K]$  for each  $K \geq 2$  and some sequence  $T_k[K] \geq 0$ ,  $k = 1, 2, \dots$ . The elements of the sequences  $T_k[K]$  are also referred to as Tokunaga coefficients, which does not create confusion with  $T_{i,j}[K]$ . A measure  $\mu$  on  $\mathcal{T}$  is said to satisfy the *Toeplitz property* if the respective conditional measures  $\{\mu_K\}$ , as in Eq. (2), satisfy the Toeplitz property.

**Definition 6 (Mean Self-similarity).** A measure  $\mu$  on  $\mathcal{T}$  is called *mean self-similar* if it is mean coordinated and satisfies the Toeplitz property.

For a mean self-similar measure the Tokunaga matrix  $\mathbb{T}_K$  becomes Toeplitz:

$$\mathbb{T}_K = \begin{bmatrix} 0 & T_1 & T_2 & \dots & T_{K-1} \\ 0 & 0 & T_1 & \dots & T_{K-2} \\ 0 & 0 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & 0 & T_1 \\ 0 & 0 & \dots & 0 & 0 \end{bmatrix}.$$

Pruning  $\mathcal{R}$  decreases the Horton–Strahler order of each vertex (and hence of each branch) by unity; in particular

$$N_k[T] = N_{k-1}[\mathcal{R}(T)], \quad k \geq 2, \tag{4}$$

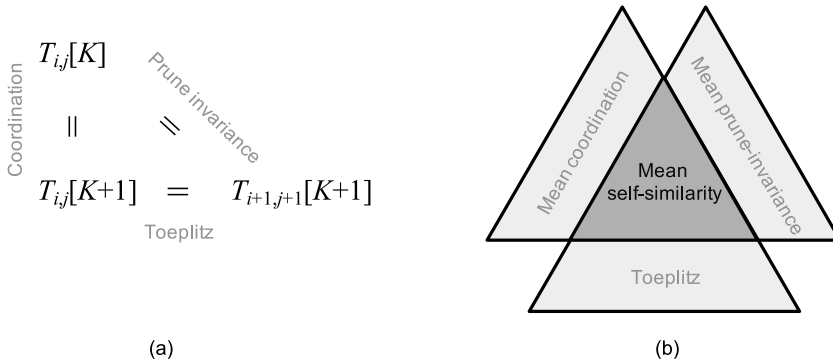
$$N_{i,j}[T] = N_{i-1,j-1}[\mathcal{R}(T)], \quad 2 \leq i < j. \tag{5}$$

Consider measure  $\mu_K^{\mathcal{R}}$  induced on  $\mathcal{H}_K$  by the pruning operator:

$$\mu_K^{\mathcal{R}}(A) = \mu_{K+1}(\mathcal{R}^{-1}(A)) \quad \forall A \subset \mathcal{H}_K.$$

The Tokunaga coefficients computed on  $\mathcal{H}_K$  using the induced measure  $\mu_K^{\mathcal{R}}$  are denoted by  $T_{i,j}^{\mathcal{R}}[K]$ . Formally,

$$T_{i,j}^{\mathcal{R}}[K] = T_{i+1,j+1}[K+1] = \frac{\mathcal{N}_{i+1,j+1}[K+1]}{\mathcal{N}_{j+1}[K+1]}. \tag{6}$$



**Fig. 3.** Relations among mean coordination, mean prune-invariance, and Toeplitz property. (a) Pairwise equalities among the quantities  $T_{i,j}[K]$ ,  $T_{i,j}[K + 1]$ , and  $T_{i+1,j+1}[K + 1]$  involved in the definitions of mean coordination, mean prune-invariance, and Toeplitz property. (b) Venn diagram of the space  $\mathcal{T}$  illustrating the relation among mean coordination (left triangle), mean prune-invariance (right triangle), and Toeplitz property (bottom triangle). The mean self-similarity (inner dark triangle) is formed by the intersection of any pair of the three properties.

**Definition 7 (Mean Prune-invariance).** A set of measures  $\{\mu_K\}_{K \geq 1}$  on  $\{\mathcal{H}_K\}_{K \geq 1}$  is called *mean prune-invariant* if  $T_{i,j}[K] = T_{i,j}^{\mathcal{R}}[K]$  (equivalently,  $T_{i,j}[K] = T_{i+1,j+1}[K + 1]$ ), for any  $K \geq 2$  and all  $1 \leq i < j \leq K$ . A measure  $\mu$  on  $\mathcal{T}$  is called *mean prune-invariant* if the respective conditional measures  $\{\mu_K\}$ , as in Eq. (2), are mean prune-invariant.

**Definition 8 (Mean Self-similarity).** A probability measures  $\mu$  on  $\mathcal{T}$  is called *mean self-similar with respect to pruning  $\mathcal{R}$*  if it is coordinated and mean prune-invariant.

**Proposition 1.** *Definitions 6 and 8 of mean self-similarity are equivalent.*

This equivalence was proven in [11]. Its validity is readily seen from the diagram of Fig. 3a, which shows relations among the quantities  $T_{i,j}[K]$ ,  $T_{i,j}[K + 1]$ , and  $T_{i+1,j+1}[K + 1]$  involved in the definitions of coordination, prune-invariance, and Toeplitz property. Moreover, we observe that if any two of these properties hold, the third also holds. The Venn diagram of Fig. 3b illustrates the relation among mean coordination, mean prune-invariance, Toeplitz property and mean self-similarity in the space  $\mathcal{T}$ . In this work, we refer to the mean self-similarity with respect to pruning  $\mathcal{R}$  simply as mean self-similarity.

A variety of mean self-similar measures can be constructed for an arbitrary sequence of Tokunaga coefficients  $T_k \geq 0, k \geq 1$ . Next, we give a natural example [11].

### 3.2. Example of a mean self-similar measure: Independent random attachment

The subspace  $\mathcal{H}_1$ , which consists of a single-leaf tree, possesses a trivial unity mass measure. To construct a random tree from  $\mathcal{H}_2$ , we select a discrete probability distribution  $P_{1,2}(n)$ ,  $n = 0, 1, \dots$ , with the mean value  $T_1$ . A random tree  $T \in \mathcal{H}_2$  is obtained from the single-leaf tree  $\tau_1$  of order 1 via the following two operations. First, we attach two offspring vertices to the leaf of  $\tau_1$ . This creates a tree of order 2 with no side-branches — one internal vertex of degree 3, two leaves, and the root. Second, we draw the number  $\tilde{N}_{1,2}$  from the distribution  $P_{1,2}$ , and attach  $\tilde{N}_{1,2}$  vertices to this tree so that they form side-branches of order  $\{1, 2\}$ .



In general, to construct a random tree  $T \in \mathcal{H}_K$  of order  $K \geq 2$  we select a set of discrete probability distributions  $P_{k,K}(n)$ ,  $k = 1, \dots, K - 1$ , on  $\mathbb{Z}_+$  with the respective mean values  $T_k$ . A random tree  $T \in \mathcal{H}_K$  is constructed by adding branches of order 1 (leaves) to a random tree  $\tau \in \mathcal{H}_{K-1}$ . First, we add two new child vertices to every leaf of  $\tau$  hence producing a tree  $\tilde{T}$  of order  $K$  with no side-branches of order 1. Second, for each branch  $b$  of order  $2 \leq j \leq K$  in  $\tilde{T}$  we draw a random number  $\tilde{N}_{1,j}(b)$  from the distribution  $P_{j-1,K}$  and attach  $\tilde{N}_{1,j}(b)$  new child vertices to this branch so that they form side-branches of order  $\{1, j\}$ . Each new vertex is attached in a random order with respect to the existing side-branches. Specifically, we notice that  $m \geq 0$  side-branches attached to a branch of order  $j$  are uniquely associated with  $m + 1$  edges within this branch. The attachment of the new  $\tilde{N}_{1,j}(b)$  vertices among the  $m + 1$  edges is given by the equiprobable multinomial distribution with  $m + 1$  categories and  $\tilde{N}_{1,j}(b)$  trials.

The procedure described above generates a set of measures  $\{\mu_K\}_{K \geq 1}$  on  $\{\mathcal{H}_K\}_{K \geq 1}$  that are mean coordinated by construction (recall that the mean values  $T_k$  of the distributions  $P_{k,K}$  are independent of  $K$ ).

Furthermore, observe that

$$N_{i,j} = \sum_{b_i=1}^{N_j} \tilde{N}_{1,j-i+1}(b_i),$$

$$\begin{aligned} \mathcal{N}_{i,j}[K] &= \mathbf{E}_K(N_{i,j}) = \mathbf{E}_K(\mathbf{E}_K(N_{i,j}|N_j)) = \mathbf{E}_K(N_j T_{j-i}) \\ &= T_{j-i} \mathbf{E}_K(N_j) = T_{j-i} \mathcal{N}_j[K], \end{aligned} \tag{7}$$

and hence  $T_{i,j}[K] = \mathcal{N}_{i,j}[K]/\mathcal{N}_j[K] = T_{j-i}$ , so the tree is mean self-similar, according to Definition 6.

**Remark 2.** The properties introduced in this section – mean coordination, mean prune-invariance, Toeplitz, and mean self-similarity – are completely specified by a set of conditional measures  $\{\mu_K\}$ , and are independent of the randomization probabilities  $p_K = \mu(\mathcal{H}_K)$ , see Eq. (2).

**Remark 3.** The idea of relating tree mean self-similarity (Definition 8) to mean prune-invariance (Definition 7) is quite intuitive (see also [4]). Much less so is the requirement of mean coordination of conditional measures (Definition 4), included in the definition of mean self-similarity. This requirement is motivated by our goal to bridge the measure-theoretic definition of self-similarity via the pruning operation (Definition 8) to a statistical definition via the branch counting (Definition 6). In applications, when a handful of trees of different orders is observed, the coordination assumption allows one to estimate the Tokunaga coefficients  $T_{i,j}$  and make inference regarding the Toeplitz property; see [6,19,20,29]. The absence of coordination, at the same time, opens a possibility of having a variety of prune-invariant measures with no Toeplitz constraint, which are hardly treatable in applications. To construct a simplest such measure, let select any tree  $T_2$  from the pre-image of the only tree  $T_1 \in \mathcal{H}_1$  of order  $K = 1$  under the pruning operation:  $T_2 \in \mathcal{R}^{-1}(T_1) \in \mathcal{H}_2$ . In a similar fashion, select any tree  $T_{K+1}$  from the pre-image of  $T_K$  for  $K \geq 2$ . This gives us a collection of trees  $T_K \in \mathcal{H}_K$ ,  $K \geq 1$  such that  $\mathcal{R}(T_{K+1}) = T_K$ . Assign the full measure on  $\mathcal{H}_K$  to  $T_K$ :  $\mu_K(T_K) = 1$ . By construction, the measures  $\{\mu_K\}$  are mean prune-invariant. They, however, may satisfy neither the mean coordination nor the Toeplitz property. This construction illustrates how one can produce rather obscure collections of trees that are mean prune-invariant, providing a motivation for the coordination requirement adopted in this work.

### 3.3. Self-similarity of a combinatorial tree

This section introduces a distribution-based approach to self-similarity.

**Definition 9 (Prune-invariance).** Consider a probability measure  $\mu$  on  $\mathcal{T}$  such that  $\mu(\phi) = 0$ . Let  $\nu(T) = \mu \circ \mathcal{R}^{-1}(T) = \mu(\mathcal{R}^{-1}(T))$ . (Note that  $\nu(\phi) > 0$ .) Measure  $\mu$  is called invariant with respect to the pruning operation (prune-invariant) if for any tree  $T \in \mathcal{T}$  we have

$$\nu(T | T \neq \phi) = \mu(T). \tag{8}$$

**Proposition 2.** Let  $\mu$  be a prune-invariant measure on  $\mathcal{T}$ . Then the distribution of orders,  $p_K = \mu(\mathcal{H}_K)$ , is geometric:

$$p_K = p(1 - p)^{K-1}, \quad K \geq 1, \tag{9}$$

where  $p = p_1 = \mu(\mathcal{H}_1)$ , and for any  $T \in \mathcal{H}_K$

$$\mu_{K+1}(\mathcal{R}^{-1}(T)) = \mu_K(T). \tag{10}$$

**Proof.** Pruning  $\mathcal{R}$  is a shift operator on the sequence of subspaces  $\{\mathcal{H}_k\}$ :

$$\mathcal{R}^{-1}(\mathcal{H}_{K-1}) = \mathcal{H}_K, \quad K \geq 2. \tag{11}$$

The only tree eliminated by pruning is the tree of order 1:  $\{\tau : \mathcal{R}(\tau) = \phi\} = \mathcal{H}_1$ . This allows to rewrite (8) for any  $T \neq \phi$  as

$$\mu(\mathcal{R}^{-1}(T)) = \mu(T)(1 - \mu(\mathcal{H}_1)). \tag{12}$$

Combining (11) and (12) we find for any  $K \geq 2$

$$\mu(\mathcal{H}_K) \stackrel{\text{by (11)}}{=} \mu(\mathcal{R}^{-1}(\mathcal{H}_{K-1})) \stackrel{\text{by (12)}}{=} (1 - \mu(\mathcal{H}_1))\mu(\mathcal{H}_{K-1}), \tag{13}$$

which establishes (9). Next, for any tree  $T \in \mathcal{H}_K$  we have

$$\mu(T) = \mu(\mathcal{H}_1)(1 - \mu(\mathcal{H}_1))^{K-1}\mu_K(T),$$

$$\mu(\mathcal{R}^{-1}(T)) = \mu(\mathcal{H}_1)(1 - \mu(\mathcal{H}_1))^K\mu_{K+1}(\mathcal{R}^{-1}(T)).$$

Together with (12) this implies (10).  $\square$

Proposition 2 shows that a prune-invariant measure  $\mu$  is completely specified by its conditional measures  $\mu_K$  and the mass  $p = \mu(\mathcal{H}_1)$  of the tree of order  $K = 1$ . The same result was obtained for Galton–Watson trees in [4, Thm. 3.5].

Next, we introduce a distributional analog of the mean coordination property; see Definition 4 and Remark 3. Specifically, we assume that a complete subtree  $T_K$  of a given order  $K$  randomly selected from a random tree  $T_H$  of order  $H \geq K$  has a common distribution independent of  $H$ . Since a tree  $T_K$  of order  $K$  has only one complete subtree of order  $K$ , which coincides with  $T_K$ , this common distribution must be  $\mu_K$ . Formally, consider the following process of selecting a uniform random complete subtree  $\text{subtree}_{K,H}$  of order  $K$  from a random tree  $T_H \in \mathcal{H}_H$ . First, select a random tree  $T_H$  according to the conditional measure  $\mu_H$ . Label all complete subtrees of order  $K$  in  $T_H$  in order of proper labeling of Section 2.2, and select a uniform random subtree, which we denote  $\text{subtree}_{K,H}$ . By construction,  $\text{subtree}_{K,H} \in \mathcal{H}_K$ ; we denote the corresponding sampling measure on  $\mathcal{H}_K$  by  $\mu_K^H$ .

**Definition 10 (Coordination).** A set of measures  $\{\mu_K\}_{K \geq 1}$  on  $\{\mathcal{H}_K\}_{K \geq 1}$  is called *coordinated* if  $\mu_K^H(T) = \mu_K(T)$  for any  $K \geq 1, H \geq K$ , and  $T \in \mathcal{H}_K$ . A measure  $\mu$  on  $\mathcal{T}$  is called *coordinated* if the respective conditional measures  $\{\mu_K\}$ , as in Eq. (2), are coordinated.

**Example 1.** The space of finite binary Galton–Watson trees has the coordination property. Recall that a random binary Galton–Watson tree starts with a single progenitor (root) and increases its depth in discrete steps: at every step each existing vertex can either split in two with probability  $p_2$  or become a leaf (disappear) with probability  $p_0 = 1 - p_2$ . We denote the resulting tree distribution on  $\mathcal{T}$  by  $\mathcal{GW}(p_0, p_2)$ . This Markovian generation mechanism creates complete subtrees of the same structure, independently of the tree order. This implies coordination.

**Definition 11 (Combinatorial Self-similarity).** A probability measure  $\mu$  on  $\mathcal{T}$  is called (*combinatorially*) *self-similar with respect to pruning  $\mathcal{R}$*  if it is coordinated and prune-invariant.

In this work, we refer to combinatorial self-similarity with respect to pruning  $\mathcal{R}$  simply as combinatorial self-similarity. It was established in [4] that critical binary Galton–Watson trees  $\mathcal{GW}(1/2, 1/2)$  are prune-invariant. Together with coordination (see Example 1), this implies combinatorial self-similarity. It also has been shown in [4] that non-critical binary Galton–Watson trees ( $p_0 \neq 1/2$ ) are not prune-invariant. This gives an example of coordinated measures that are not prune-invariant. Prune-invariant measures with no coordination can be easily constructed following the approach of Remark 3. To make that construction consistent with the definition of distributional prune-invariance (Definition 9), each tree  $T_K \in \mathcal{H}_K$  must be assigned the probability  $p_K = p(1 - p)^{K-1}$ .

### 3.4. Horton law in self-similar trees

We say that a random tree  $T$  satisfies a *strong Horton law* if the respective sequence  $\mathcal{N}_k[K]$  of branch numbers decays in geometric fashion as  $k$  increases. Formally, we require

$$\lim_{K \rightarrow \infty} \frac{\mathcal{N}_k[K]}{\mathcal{N}_1[K]} = R^{1-k}, \text{ for any } k \geq 1. \tag{14}$$

Horton law and its ramifications, which epitomize scale-invariance of dendritic hierarchical structures, are indispensable in hydrology (e.g., [6,20,23]) and have been reported in biology and other areas; see [13,19] and references therein. It has been shown in [12] that the tree that describes a trajectory Kingman’s coalescent process with  $N$  particles obeys a weaker version of Horton law as  $N \rightarrow \infty$ , and that the first pruning of this tree for any finite  $N$  is equivalent to a level set tree of a white noise (see Section 4 for definitions).

A necessary and sufficient condition for the strong Horton law in a mean self-similar tree has been established in [11]:

$$\limsup_{k \rightarrow \infty} T_k^{1/k} < \infty.$$

The Horton exponent  $R$  in this case is given by  $R = 1/w_0$ , where  $w_0$  is the only real root of

$$\hat{t}(z) = -1 + 2z + \sum_{k=1}^{\infty} z^k T_k$$

within the interval  $(0, 1/2]$ . Informally, this means that any mean self-similar tree with a “tamed” sequence of Tokunaga coefficients satisfies the strong Horton law.

### 3.5. Self-similarity of a tree with edge lengths

Consider a tree  $T \in \mathcal{L}$  with edge lengths given by a positive vector  $l_T = (l_1, \dots, l_{\#T})$  and let  $\text{LENGTH}(T) = \sum_i l_i$ . We assume that the edges are labeled in a proper way as described in Section 2.2. A tree is completely specified by its combinatorial shape  $\text{SHAPE}(T)$  and edge length vector  $l_T$ . The edge length vector  $l_T$  can be specified by distribution  $\chi(\cdot)$  of a point  $x_T = (x_1, \dots, x_{\#T})$  on the simplex  $\sum_i x_i = 1, 0 < x_i \leq 1$ , and conditional distribution  $F(\cdot|x_T)$  of the tree length  $\text{LENGTH}(T)$ , where

$$l_T = x_T \times \text{LENGTH}(T).$$

A measure  $\eta$  on  $\mathcal{L}$  is a joint distribution of tree’s combinatorial shape and its edge lengths; it has the following component measures.

Combinatorial shape :  $\mu(\tau) = \text{Law}(\text{SHAPE}(T) = \tau)$ ,

Relative edge lengths :  $\chi_\tau(\bar{x}) = \text{Law}(x_T = \bar{x} \mid \text{SHAPE}(T) = \tau)$ ,

Total tree length :  $F_{\tau, \bar{x}}(\ell) = \text{Law}(\text{LENGTH}(T) = \ell \mid x_T = \bar{x}, \text{SHAPE}(T) = \tau)$ .

The definition of self-similarity for a tree with edge lengths builds on its analog for combinatorial trees in Section 3.3. The combinatorial notions of coordination (Definition 10) and prune-invariance (Definition 9), which we refer to as coordination and prune-invariance *in shapes*, are complemented with analogous properties *in edge lengths*. Formally, we denote by  $\mu_K^H(\tau)$ ,  $\chi_\tau^H(\bar{x})$ , and  $F_{\tau, \bar{x}}^H(\ell)$  the component measures for a uniform complete subtree  $\text{subtree}_{K,H}$ . (Notice that the subtree order  $K$  is completely specified by the tree shape  $\tau$ , which explains the absence of subscript  $K$  in the component measures for subtree length). We also consider the distribution of edge lengths after pruning:

$$\Xi_\tau(\bar{x}) = \text{Law}(x_{\mathcal{R}(T)} = \bar{x} \mid \text{SHAPE}(\mathcal{R}(T)) = \tau)$$

and

$$\Phi_{\tau, \bar{x}}(\ell) = \text{Law}(\text{LENGTH}(\mathcal{R}(T)) = \ell \mid x_{\mathcal{R}(T)} = \bar{x}, \text{SHAPE}(\mathcal{R}(T)) = \tau).$$

Finally, we adopt here the notation  $\mathcal{H}_K$  for a subspace of trees of order  $K \geq 1$  from  $\mathcal{L}$ , and consider conditional measures  $\mu_K(\tau) = \mu(\tau \mid k(\tau) = K), K \geq 1$ , for a tree  $\tau \in \mathcal{L}$ .

**Definition 12 (Self-Similarity of a Tree with Edge Lengths).** We call a measure  $\eta$  on  $\mathcal{L}$  *self-similar with respect to pruning  $\mathcal{R}$*  if the following conditions hold

- (i) The measure is coordinated in shapes. This means that for every  $K \geq 1$  and every  $H \geq K$  we have

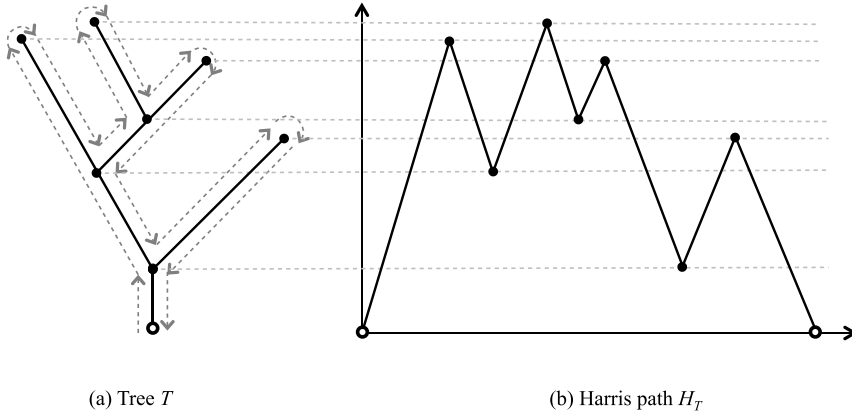
$$\mu_K^H(\tau) = \mu_K(\tau) \quad \forall \tau \in \mathcal{H}_K.$$

- (ii) The measure is coordinated in lengths. This means that for every  $K \geq 1, H \geq K$ , and  $\tau \in \mathcal{H}_K$  we have

$$\chi_\tau^H(\bar{x}) = \chi_\tau(\bar{x}) \quad \bar{x}\text{-a.s.},$$

and for every given  $\bar{x}$ ,

$$F_{\tau, \bar{x}}^H(\ell) = F_{\tau, \bar{x}}(\ell) \quad \ell\text{-a.s.}$$



**Fig. 4.** (a) Tree  $T$  and its depth-first search illustrated by dashed arrows. (b) Harris path  $H_T(t)$  for the tree  $T$  of panel (a).

(iii) The measure is prune-invariant in shapes. This means that for  $\nu = \mu \circ \mathcal{R}^{-1}$  we have

$$\mu(\tau) = \nu(\tau | \tau \neq \phi).$$

(iv) The measure is prune-invariant in lengths. This means that

$$\Xi_\tau(\bar{x}) = \chi_\tau(\bar{x})$$

and there exists a *scaling exponent*  $\zeta > 0$  such that for any combinatorial tree  $\tau \in \mathcal{T}$  we have

$$\Phi_{\tau, \bar{x}}(\ell) = \zeta^{-1} F_{\tau, \bar{x}}\left(\frac{\ell}{\zeta}\right).$$

In this work, we refer to self-similarity with respect to pruning  $\mathcal{R}$  simply as self-similarity. Section 5 introduces a rich class of measures that satisfy this definition.

#### 4. Tree representation of continuous functions

We review here the results of [15,18,21,28] on tree representation of continuous functions. This allows us to apply the self-similarity concepts to time series and motivates discussion in Section 5.6, 5.7.

##### 4.1. Harris path

For any embedded tree  $T \in \mathcal{L}_{\text{plane}}$  with edge lengths, the *Harris path* is defined as a piece-wise linear function [9,21]

$$H_T(t) : [0, 2 \cdot \text{LENGTH}(T)] \rightarrow \mathbb{R}$$

that equals the distance from the root traveled along the tree  $T$  in the depth-first search, as illustrated in Fig. 4. For a tree  $T$  with  $n$  leaves, the Harris path  $H_T(t)$  is a piece-wise linear positive excursion that consists of  $2n$  linear segments with alternating slopes  $\pm 1$  [21].

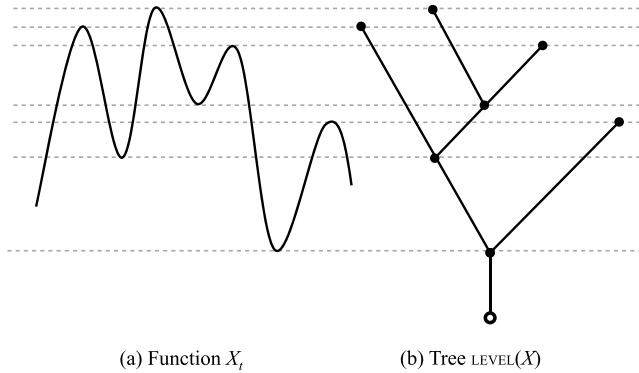


Fig. 5. Function  $X_t$  (panel a) with a finite number of local extrema and its level-set tree  $\text{LEVEL}(X)$  (panel b).

#### 4.2. Level set tree

Consider a continuous function  $X_t, t \in [a, b]$  with a finite number of distinct local minima. The level set  $\mathcal{L}_\alpha(X_t)$  is defined as the pre-image of the function values above  $\alpha$ :

$$\mathcal{L}_\alpha(X_t) = \{t : X_t \geq \alpha\}.$$

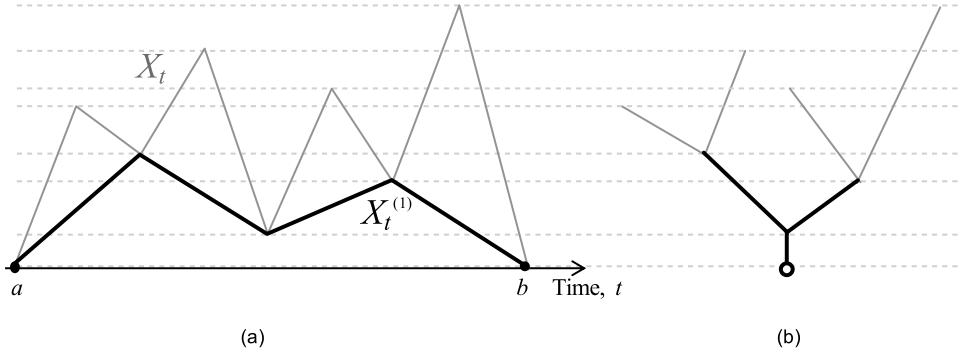
The level set  $\mathcal{L}_\alpha$  for each  $\alpha$  is a union of non-overlapping intervals; we write  $|\mathcal{L}_\alpha|$  for their number. Notice that  $|\mathcal{L}_\alpha| = |\mathcal{L}_\beta|$  as soon as the interval  $[\alpha, \beta]$  does not contain a value of local extrema of  $X_t$  and  $0 \leq |\mathcal{L}_\alpha| \leq n$ , where  $n$  is the number of the local maxima of  $X_t$ .

The *level set tree*  $\text{LEVEL}(X_t) \in \mathcal{L}_{\text{plane}}$  is a tree that describes the topology of the level sets  $\mathcal{L}_\alpha$  as a function of threshold  $\alpha$ , as illustrated in Fig. 5. Specifically, there are bijections between (i) the leaves of  $\text{LEVEL}(X_t)$  and the local maxima of  $X_t$ , (ii) the internal (parental) vertices of  $\text{LEVEL}(X_t)$  and the local minima of  $X_t$  (excluding possible local minima at the boundary points), and (iii) the pair of subtrees of  $\text{LEVEL}(X_t)$  rooted at the parental vertex that corresponds to a local minima  $X_{t^*}$  and the first positive excursions (or meanders bounded by  $t = a$  or  $t = b$ ) of  $X_t - X_{t^*}$  to right and left of  $t^*$ . Every edge in the tree is assigned a length equal the difference of the values of  $X_t$  at the local extrema that correspond to the vertices adjacent to this edge according to the bijections (i) and (ii) above. The lowest internal local minimum (achieved for  $a < t < b$ ) corresponds to the first descendant of the root, see Fig. 5. If the global minimum is achieved on the interval boundary ( $t = a$  or  $t = b$ ), then it corresponds to the tree root; otherwise, we artificially add a root to the tree, with an arbitrary root edge length; this situation is illustrated in Fig. 5.

By construction, the level-set tree  $\text{LEVEL}(X_t)$  is completely determined by the sequence of the values of local extrema of  $X_t$ , and is independent of timing of those extrema, as soon as their order is preserved. This means, for instance, that if  $g(t)$  is a continuous and monotone increasing function on  $[a, b]$ , then the following trees are equivalent:

$$\text{LEVEL}(X_t) = \text{LEVEL}(X_{g(t)}).$$

Hence, without loss of generality we can focus on the level set trees of continuous functions with alternating slopes  $\pm 1$ . By  $\mathcal{E}^{\text{ex}}$  we denote the space of all positive piece-wise linear continuous finite excursions with alternating slopes  $\pm 1$ .



**Fig. 6.** Pruning of a positive excursion: transition to the local minima of an excursion  $X_t$  corresponds to pruning of the corresponding level set tree. (a) An original excursion  $X_t$  (gray line) and linearly interpolated sequence  $X_t^{(1)}$  of the respective local minima (black line). (b) The level set tree  $\text{LEVEL}(X_t^{(1)})$  of the local minima sequence (black lines) is obtained by pruning of the level set tree  $\text{LEVEL}(X_t)$  of the original excursion (whole tree). The pruned edges are shown in gray — each of them corresponds to a local maximum of the original excursion.

By construction, the level set tree of an excursion from  $\mathcal{E}^{\text{ex}}$  and Harris path are reciprocal to each other as described in the following statement.

**Proposition 3 (Reciprocity of Harris Path and Level Set Tree).** *The Harris path  $H : \mathcal{L}_{\text{plane}} \rightarrow \mathcal{E}^{\text{ex}}$  and the level set tree  $\text{LEVEL} : \mathcal{E}^{\text{ex}} \rightarrow \mathcal{L}_{\text{plane}}$  are reciprocal to each other. This means that for any  $T \in \mathcal{L}_{\text{plane}}$  we have  $\text{LEVEL}(H_T(t)) \equiv T$ , and for any  $Y_t \in \mathcal{E}^{\text{ex}}$  we have  $H_{\text{LEVEL}(Y_t)}(t) \equiv Y_t$ .*

### 4.3. Pruning for positive excursions

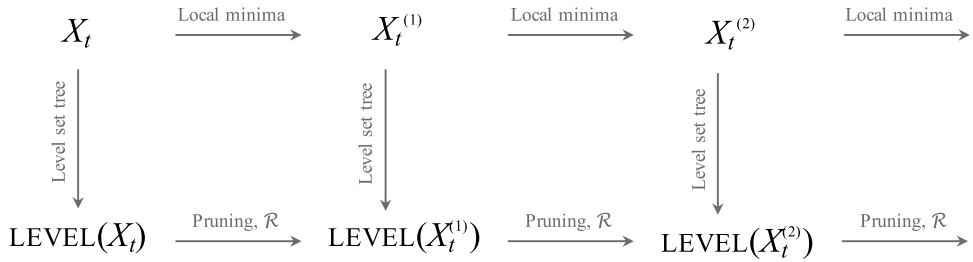
This section examines level set tree and the respective pruning for a positive continuous excursion  $X_t$  on  $[a, b]$ . Formally, consider a continuous function  $X_t$  with a finite number of local minima and such that  $X_a = X_b = 0$  and  $X_t > 0$  for  $a < t < b$ . Furthermore, consider excursion  $X_t^{(1)}, t \in [a, b]$ , obtained by a linear interpolation of the boundary values and the local minima of  $X_t$ ; as well as functions  $X_t^{(m)}, t \in [a, b]$ , for  $m \geq 1$ , obtained by taking the local minima of  $X_t$  iteratively  $m$  times, and linearly interpolating their values together with  $X_a = X_b = 0$  (see Fig. 6a).

In the space of level set trees of continuous functions, the pruning  $\mathcal{R}$  corresponds to coarsening the respective function by removing (smoothing) its local maxima, as illustrated in Fig. 6. An iterative pruning corresponds to iterative transition to the local minima.

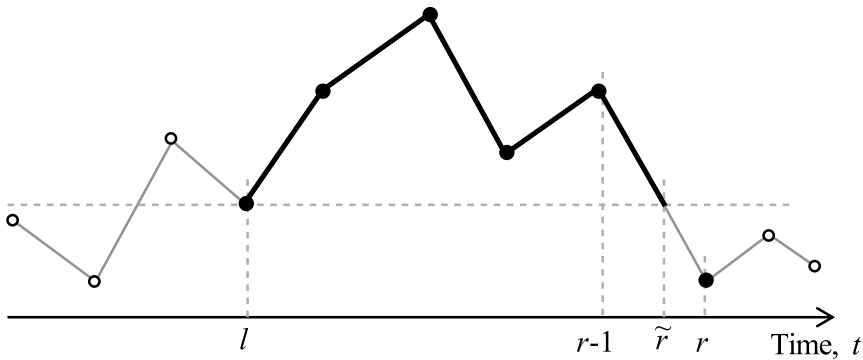
**Proposition 4 (Pruning for Positive Excursions, [28]).** *Using the definitions of this section, the transition from a positive excursion  $X_t$  to the respective excursion  $X_t^{(1)}$  of its local minima corresponds to pruning of the level-set tree  $\text{LEVEL}(X_t)$ . This is illustrated in a diagram of Fig. 7. Formally,*

$$\text{LEVEL}(X_t^{(m)}) = \mathcal{R}^m(\text{LEVEL}(X_t)), \forall m \geq 1.$$

It is straightforward to formulate an analog of Proposition 4 without the excursion assumption (for continuous functions with a finite number of local minima). This, however, involves technicalities that are tangential to the essence of this work and will be discussed elsewhere.



**Fig. 7.** Transition to the local minima of a function  $X_t$  corresponds to pruning  $\mathcal{R}$  of the corresponding level set tree  $\text{LEVEL}(X_t)$ .



**Fig. 8.** Excursion of a time series: illustration. The values of time series  $X_k$  are shown by circles; the circles that form the excursion on  $[l, r]$  are filled. The linear interpolation function  $X_t$  is shown by solid line; the excursion of  $X_t$  on  $[l, \tilde{r}]$  is shown in bold.

#### 4.4. Self-similarity for time series

Consider a time series  $X_k, k \in \mathbb{Z}$ , with an atomless distribution of values at each  $k$ . Let  $X_t, t \in \mathbb{R}$ , be a continuous function of linearly interpolated values of  $X_k$ . We define a *positive excursion* of  $X_k$  as a fragment of the time series on an interval  $[l, r], l, r \in \mathbb{Z}$  such that  $X_l \geq X_r$  and  $X_k > X_l$  for all  $l < k < r$ . To each positive excursion of  $X_k$  on  $[l, r]$  corresponds a positive excursion of  $X_t$  on  $[l, \tilde{r}]$ , where  $\tilde{r} \in (r - 1, r]$  is such that  $X_{\tilde{r}} = X_l$ . This construction is illustrated in Fig. 8. The level set tree of a positive excursion of  $X_k$  is that of the corresponding positive excursion of  $X_t$ .

Proposition 4 and a comment after it suggest that the problem of finding self-similar trees with edge lengths is equivalent to finding *extreme-invariant processes*

$$X_k \stackrel{d}{=} \zeta X_k^{(1)} \quad \text{for some } \zeta > 0, \tag{15}$$

where  $X_k, k \in \mathbb{Z}$ , is a time series with an atomless value distribution at every  $k$  and  $X_k^{(1)}$  is the corresponding time series of local minima. If  $X_k$  satisfies (15), the level set tree of an excursion from  $X_k$ , considered as an element of  $\mathcal{L}$ , is self-similar according to Definition 12. The next section describes a solution to (15) that corresponds to  $\zeta = 2$ .



4.5. Self-similarity for random walks on  $\mathbb{R}$

Consider a random walk  $\{X_k\}_{k \in \mathbb{Z}}$  with a homogeneous transition kernel  $p(x, y) \equiv p(x - y)$ , for any  $x, y \in \mathbb{R}$ , where  $p(x)$  is an atomless density function. A homogeneous random walk is called *symmetric* if  $p(x) = p(-x)$  for all  $x \in \mathbb{R}$ .

**Lemma 1** (Pruning for Random Walks, [28]). *The following statements hold.*

- a:** *The local minima of a homogeneous random walk  $\{X_k\}_{k \in \mathbb{Z}}$  form a homogeneous random walk (with a different transition kernel in general).*
- b:** *The local minima of a symmetric homogeneous random walk  $\{X_k\}_{k \in \mathbb{Z}}$  form a symmetric homogeneous random walk (with a different transition kernel in general).*

The transition kernel of a symmetric random walk can be represented as the even part of a probability density function  $f(x)$  with support in  $\mathbb{R}_+$ :

$$p(x) = \frac{f(x) + f(-x)}{2}.$$

The following result describes the solution of the problem (15) in terms of the characteristic function of  $f(x)$ .

**Proposition 5** (Self-similarity for a Symmetric Homogeneous Random Walk, [28]). *The local minima of a symmetric homogeneous random walk  $\{X_k\}_{k \in \mathbb{Z}}$  with a transition kernel  $p(x) = \frac{f(x)+f(-x)}{2}$  form a symmetric homogeneous random walk with a transition kernel*

$$p^{(1)}(x) = \zeta^{-1} p(x/\zeta), \quad c > 0$$

if and only if  $\zeta = 2$  and

$$\Re[\widehat{f}(2s)] = \left| \frac{\widehat{f}(s)}{2 - \widehat{f}(s)} \right|^2, \tag{16}$$

where  $\widehat{f}(s)$  is the characteristic function of  $f(x)$  and  $\Re[z]$  stays for the real part of  $z \in \mathbb{C}$ .

A solution to (16) is given for example by an exponential density  $f(x) = \phi_\lambda(x)$  of (18) for any  $\lambda > 0$ ; a detailed discussion of exponential kernels is given in Section 4.6. A weaker, mean self-similarity of Definitions 6, 8 is satisfied in any symmetric random walk, as discussed in the following statement.

**Theorem 1** (Mean Self-similarity of a Symmetric Homogeneous Random Walk, [28]). *The combinatorial level set tree  $T = \text{SHAPE}(\text{LEVEL}(X_t))$  of a finite symmetric homogeneous random walk  $X_k$  with  $k = 1, \dots, N$  is mean self-similar. Specifically, for a uniform random complete subtree  $\text{subtree}_{K, \mathfrak{k}(T)} \subset T$  of order  $K < \mathfrak{k}(T)$  the numbers  $\tau_{i,j}^{(r)}$  of side-branches of order  $i$  that merge the  $r$ th branch of order  $j$ , with  $2 \leq j \leq K$ , in  $\text{subtree}_{K, \mathfrak{k}(T)}$  are independent identically distributed random variables. If  $\tau_{i,j}$  is a random variable such that  $\tau_{i,j}^{(r)} \stackrel{d}{=} \tau_{i,j}$ , then*

$$T_{i,j} := \mathbb{E}[\tau_{i,j}] = 2^{j-i-1} =: T_{j-i}. \tag{17}$$

Moreover, by the strong law of large numbers  $\mathfrak{k}(T) \xrightarrow{a.s.} \infty$  as  $N \rightarrow \infty$ , and for any  $1 \leq i < j$  we have

$$T_{i,j} \xrightarrow{a.s.} T_{j-i} = 2^{j-i-1} \quad \text{as } N \rightarrow \infty,$$

where  $T_{i,j}$  can be computed over the entire  $T$ .

### 4.6. Exponential random walks

We call a homogeneous random walk *exponential* if its kernel is a mixture of exponential jumps constructed as follows

$$p(x) = \rho \phi_{\lambda_u}(x) + (1 - \rho) \phi_{\lambda_d}(-x), \quad 0 \leq \rho \leq 1, \quad \lambda_u, \lambda_d > 0,$$

where  $\phi_\lambda$  is the exponential density with parameter  $\lambda > 0$ ,

$$\phi_\lambda(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0, \\ 0 & \text{if } x < 0. \end{cases} \tag{18}$$

We refer to an exponential random walk by its parameter triplet  $\{\rho, \lambda_u, \lambda_d\}$ . Each exponential random walk with parameters  $\{\rho, \lambda_u, \lambda_d\}$  corresponds to a piece-wise linear function from  $\mathcal{E}(\mathbb{R})$  whose rises and falls have independent exponential lengths with parameters  $(1 - \rho)\lambda_u$  and  $\rho\lambda_d$ , respectively. An exponential random walk is symmetric if and only if  $\rho = 1/2$  and  $\lambda_u = \lambda_d$ .

**Theorem 2** (Self-similarity of Exponential Random Walks, [28]). *Let  $X_t$  be an exponential random walk with parameters  $\{\rho, \lambda_u, \lambda_d\}$ . Then*

**a:** *The local minima of  $X_t$  form a exponential random walk with parameters  $\{\rho^*, \lambda_u^*, \lambda_d^*\}$  such that*

$$\rho^* = \frac{\rho \lambda_d}{\rho \lambda_d + (1 - \rho) \lambda_u}, \quad \lambda_d^* = \rho \lambda_d, \quad \text{and} \quad \lambda_u^* = (1 - \rho) \lambda_u. \tag{19}$$

**b:** *The exponential walk  $X_t$  satisfies the self-similarity (15) if and only if it is symmetric, that is if  $\rho = 1/2$  and  $\lambda_u = \lambda_d$ .*

**c:** *The self-similarity (15) is achieved after the first pruning, for the chain  $X_t^{(1)}$  of the local minima, if and only if the walk’s increments have zero mean,  $\rho \lambda_d = (1 - \rho) \lambda_u$ .*

Recall that  $\mathcal{GW}(p_0, p_2)$  denotes the space of binary Galton–Watson trees with termination probability  $p_0$  and split probability  $p_2$  (see Example 1).

**Definition 13** (Exponential Binary Galton–Watson Tree, [21]). We say that a random embedded binary tree  $T \in \mathcal{L}_{\text{plane}}$  is an exponential binary Galton–Watson tree  $\mathbf{GW}(\lambda', \lambda)$ , for  $0 \leq \lambda' < \lambda$ , if  $\text{SHAPE}(T)$  is a binary Galton–Watson tree  $\mathcal{GW}(p_0, p_2)$  with

$$p_0 = \frac{\lambda + \lambda'}{2\lambda}, \quad p_2 = \frac{\lambda - \lambda'}{2\lambda},$$

and given  $\text{SHAPE}(T)$ , the edges of  $T$  are sampled as independent exponential random variables with parameter  $2\lambda$ , i.e., with density  $\phi_{2\lambda}(x)$ .

A connection between exponential random walks and Galton–Watson trees is given by the following well known result.

**Theorem 3** ([21, Lemma 7.3], [15, 18]). *Consider a random excursion  $Y_t$  in  $\mathcal{E}^{\text{ex}}$ . The level set tree  $\text{LEVEL}(Y_t)$  is an exponential binary Galton–Watson tree  $\mathbf{GW}(\lambda', \lambda)$  if and only if the rises and falls of  $Y_t$ , excluding the last fall, are distributed as independent exponential random variables with parameters  $(\lambda + \lambda')$  and  $(\lambda - \lambda')$ , respectively, for some  $0 \leq \lambda' < \lambda$ . Equivalently, the level set tree of a homogeneous random walk is a binary Galton–Watson tree  $\mathbf{GW}(\lambda', \lambda)$  if and only if  $Y_t$ , as an element of  $\mathcal{E}^{\text{ex}}$ , corresponds to an excursion of an exponential walk with parameters  $\{\rho, \lambda_u, \lambda_p\}$  such that  $(1 - \rho)\lambda_u = \lambda + \lambda'$  and  $\rho\lambda_d = \lambda - \lambda'$ .*

We emphasize the following direct consequence of [Theorems 2\(a\)](#) and [3](#).

**Corollary 1.** *Consider a critical binary Galton–Watson tree with independent exponential lengths,  $T = \text{GW}(0, \gamma)$ . The following statements hold:*

- a:** *The Harris path of  $\mathcal{R}^k(T)$  for any  $0 \leq k < k(T)$  corresponds to a positive excursion of a symmetric exponential random walk with parameters  $\{\frac{1}{2}, 2^{1-k}\gamma, 2^{1-k}\gamma\}$ , or, equivalently,  $\mathcal{R}^k(T) = \text{GW}(0, 2^{-k}\gamma)$ .*
- b:** *The length of any branch of order  $j \geq 1$  in  $T$  has exponential distribution with parameter  $2^{2-j}\gamma$ . The lengths of branches (of all orders) are independent.*

### 5. Hierarchical branching process

The results of previous section concern a very narrow class of mean self-similar trees — those with  $T_k = 2^{k-1}$ . Among such trees, the self-similarity is established only for the critical binary Galton–Watson tree  $\text{GW}(0, \gamma)$  with independent exponential edge lengths, i.e., continuous parameter Galton–Watson binary branching Markov processes; this case corresponds to the scaling exponent  $\zeta = 2$ . Here we construct a branching process that generates self-similar trees for an arbitrary sequence  $T_k \geq 0$  and for any  $\zeta > 0$ ; it includes the critical binary Galton–Watson tree as a special case.

#### 5.1. Definition and main properties

Consider a probability mass function  $\{p_K\}_{K \geq 1}$ , a sequence  $\{T_k\}_{k \geq 1}$  of nonnegative Tokunaga coefficients, and a sequence  $\{\lambda_j\}_{j \geq 1}$  of positive termination rates. A multi-type branching process  $S(t)$  starts with a root branch of Horton–Strahler order  $K \geq 1$  with probability  $p_K$ . Every branch of order  $j \leq K$  produces offspring of order  $i < j$  with rate  $\lambda_j T_{j-i}$ . A branch of order  $j$  terminates with rate  $\lambda_j$ . After termination, a branch of order  $j \geq 2$  splits into two branches of order  $j - 1$ . A branch of order  $j = 1$  terminates without leaving offspring. The branching history of  $S(t)$  creates a random binary tree  $T[S]$  in the space  $\mathcal{L}$  of binary trees with edge lengths and no planar embedding. The process is uniquely specified by the triplet

$$S(t) = (\{T_k\}, \{\lambda_j\}, \{p_K\}).$$

**Proposition 6** (*Side-branching in Hierarchical Branching Process*). *Consider a hierarchical branching process  $S(t) = (\{T_k\}, \{\lambda_j\}, \{p_K\})$ . For any branch  $b \subset T[S]$  of order  $K + 1 \geq 2$ , let  $m_i := m_i(b) \geq 0$  be the number of its side branches of order  $i = 1, \dots, K$ , and  $m := m(b) = m_1 + \dots + m_K$  be the total number of the side branches. Let  $l_i := l_i(b)$  be the lengths of  $m + 1$  edges within  $b$ , counted sequentially from the initial vertex, and  $l := l(b) = l_1 + \dots + l_{m+1}$  be the total branch length. Then the following statements hold:*

- (1) *The total numbers  $m(b)$  of side branches within different branches of order  $K + 1$  are i.i.d. random variables with a common geometric distribution:*

$$P(m = \kappa) = q(1 - q)^\kappa \quad \text{with} \quad q = \frac{1}{1 + T_1 + \dots + T_K}, \quad \kappa = 0, 1, \dots \tag{20}$$

- (2) *The number  $m_i$  of side branches of order  $i$  has geometric distribution:*

$$P(m_i = \kappa) = q_i(1 - q_i)^\kappa \quad \text{with} \quad q_i = \frac{1}{1 + T_{K-i+1}}, \quad \kappa = 0, 1, \dots \tag{21}$$

- (3) Conditioned on the total number  $m$  of side branches, the distribution of  $\{m_i\}$  is multinomial with  $m$  trials and success probabilities

$$P(\text{side branch has order } i) = \frac{T_{K-i+1}}{T_1 + \dots + T_K}. \tag{22}$$

The side branch order vector  $\mathbf{k} = (k_1, \dots, k_m)$ , where the side branches are labeled sequentially starting from the initial vertex of  $b$ , is obtained from the sequence

$$\text{orders} = \underbrace{(1, \dots, 1)}_{m_1 \text{ times}}, \underbrace{(2, \dots, 2)}_{m_2 \text{ times}}, \dots, \underbrace{(K, \dots, K)}_{m_K \text{ times}}$$

by a uniform random permutation  $\sigma_m$  of indices  $\{1, \dots, m\}$ :

$$\mathbf{k} = \text{orders} \circ \sigma_m.$$

- (4) The branch length  $l$  has exponential distribution with rate  $\lambda_{K+1}$ , independent of the lengths of any other branch (of any order). The corresponding edge lengths  $l_i$  are i.i.d. random variables; they have a common exponential distribution with rate

$$\lambda_{K+1}(1 + T_1 + \dots + T_K). \tag{23}$$

**Proof.** All the properties readily follow from process construction.  $\square$

Proposition 6 provides an alternative definition of the hierarchical branching process, and its construction – via parts (1), (3), and (4) – that does not require time-dependent simulations.

**Theorem 4 (Self-Similarity of Hierarchical Branching Process).** Consider a hierarchical branching process  $S(t) = (\{T_k\}, \{\lambda_j\}, \{p_K\})$  and let  $T := T[S]$  be the tree generated by  $S(t)$ . The following statements hold.

- (1) The combinatorial part of  $T$  is mean self-similar (according to Definitions 6,8) with Tokunaga coefficients  $\{T_k\}$ .
- (2) The combinatorial part of  $T$  is self-similar (according to Definition 11) with Tokunaga coefficients  $\{T_k\}$  if and only if

$$p_K = p(1 - p)^{K-1}$$

for some  $0 < p < 1$ .

- (3) The tree  $T$  is self-similar (according to Definition 12) with scaling exponent  $\zeta > 0$  if and only if

$$p_K = p(1 - p)^{K-1} \quad \text{and} \quad \lambda_j = \gamma \zeta^{-j}$$

for some positive  $\gamma$  and  $0 < p < 1$ .

**Proof.** By process construction, the tree  $T$  is coordinated in shapes and lengths (according to Definition 12), with independent complete subtrees.

(1) Proposition 6, part (3) implies that the expected value of the number  $\tilde{N}_{i,j}$  of side branches of order  $i \geq 1$  within a branch of order  $j > i$  is given by  $E(\tilde{N}_{i,j}) = T_{j-i}$ . The mean self-similarity of Definition 6 with coefficients  $T_k$  immediately follows, using a conditional argument as in (7).

(2) Assume that  $\text{SHAPE}(T)$  is self-similar. A geometric distribution of orders is then established in Proposition 2. Inversely, a geometric distribution of orders ensures that the total

mass  $\mu(\mathcal{H}_K)$ ,  $K \geq 1$ , is invariant with respect to pruning. The conditional distribution of trees of a given order is completely specified by the side branch distribution, described in Proposition 6, parts (1)–(3). Consider a branch of order  $K + 1$ ,  $K \geq 1$ . Pruning decreases the orders of this branch, and all its side branches, by unity. Pruning eliminates a random geometric number  $m_1$  of side-branches of order 1 from the branch. It acts as a thinning (with removal probability  $T_K/(T_1 + \dots + T_K)$ ) on the total side branch count  $m$ . Accordingly, the total side branch count after pruning has geometric distribution with success probability

$$q^{\mathcal{R}} = \frac{1}{1 + T_1 + \dots + T_{K-1}}.$$

The order assignment among the remaining side branches of orders  $i = 1, \dots, K - 1$  is done according to multinomial distribution with probabilities proportional to  $T_{K-i}$ . This coincides with the side branch structure in the original tree, hence completing the proof of (2).

(3) Having proven (2), it remains to prove the statement for the length structure of the tree. Assume that  $T$  is self-similar with scaling exponent  $\zeta$ . The branches of order  $j \geq 2$  become branches of order  $j - 1$  after pruning, which necessitates  $\lambda_j = \zeta \lambda_{j-1}$ . Inversely, pruning acts as a thinning on the side branches within a branch of order  $K + 1$ , eliminating the side branches of order  $k = 1$ . Accordingly, the spacings between the remaining side branches are exponentially distributed with a decreased rate

$$\lambda_{K+1}(1 + T_1 + \dots + T_{K-1}) = \zeta \lambda_K(1 + T_1 + \dots + T_{K-1}).$$

Comparing this with (23), and recalling the self-similarity of  $\text{SHAPE}(T)$ , we conclude that Definition 12 is satisfied with scaling exponent  $\zeta$ .  $\square$

### 5.2. Hydrodynamic limit

Here we analyze the average numbers of branches of different orders in a hierarchical branching process, using a hydrodynamic limit. Specifically, let  $n x_j^{(n)}(s)$  be the number of branches of order  $j$  at time  $s$  observed in  $n$  independent copies of the process  $S$ . Let  $N_j(s)$  be the number of branches of order  $j \geq 1$  in the process  $S$  at instant  $s \geq 0$ . We observe that, by the law of large numbers,

$$x_j^{(n)}(s) \xrightarrow{\text{a.s.}} \mathbf{E}(N_j(s)) =: x_j(s).$$

**Theorem 5 (Hydrodynamic Limit for Branch Dynamics).** *Suppose that the following conditions are satisfied:*

$$L := \limsup_{k \rightarrow \infty} T_k^{1/k} < \infty, \tag{24}$$

and

$$\sup_{j \geq 1} \lambda_j < \infty, \quad \limsup_{j \rightarrow \infty} \lambda_j^{1/j} \leq 1/L. \tag{25}$$

Then, for any given  $T > 0$ , the empirical process

$$x^{(n)}(s) = \left(x_1^{(n)}(s), x_2^{(n)}(s), \dots\right)^T, \quad s \in [0, T],$$

converges almost surely, as  $n \rightarrow \infty$ , to the process

$$x(s) = \left(x_1(s), x_2(s), \dots\right)^T, \quad s \in [0, T],$$

that satisfies

$$\dot{x} = \mathbb{G}Ax \quad \text{with the initial conditions} \quad x(0) = \pi := \sum_{K=1}^{\infty} p_K e_K, \tag{26}$$

where  $A = \text{diag}\{\lambda_1, \lambda_2, \dots\}$  is a diagonal operator with the entries  $\lambda_1, \lambda_2, \dots$ ,  $e_i$  are the standard basis vectors, and

$$\mathbb{G} := \begin{bmatrix} -1 & T_1 + 2 & T_2 & T_3 & \dots \\ 0 & -1 & T_1 + 2 & T_2 & \dots \\ 0 & 0 & -1 & T_1 + 2 & \ddots \\ 0 & 0 & 0 & -1 & \ddots \\ \vdots & \vdots & \ddots & \ddots & \ddots \end{bmatrix}. \tag{27}$$

**Proof.** The process  $x^{(n)}(s)$  evolves according to the transition rates

$$q^{(n)}(x, x + \ell) = n\beta_\ell \left( \frac{1}{n}x \right)$$

with

$$\beta_\ell(x) = \begin{cases} \lambda_1 x_1 & \text{if } \ell = -e_1, \\ \lambda_{i+1} x_{i+1} & \text{if } \ell = 2e_i - e_{i+1}, i \geq 1, \\ \sum_{j=i+1}^{\infty} \lambda_j T_{j-i} x_j & \text{if } \ell = e_i, i \geq 1. \end{cases}$$

Here the first term reflects termination of branches of order 1; the second term reflects termination of branches of orders  $i + 1 > 1$ , each of which results in creation of two branches of order  $i$ ; and the last term reflects side-branching. Thus, the infinitesimal generator of the stochastic process  $x^{(n)}(s)$  is

$$\begin{aligned} L_n f(x) &= n\lambda_1 x_1 \left[ f\left(x - \frac{1}{n}e_1\right) - f(x) \right] \\ &+ \sum_{i=1}^{\infty} n\lambda_{i+1} x_{i+1} \left[ f\left(x - \frac{1}{n}e_{i+1} + \frac{2}{n}e_i\right) - f(x) \right] \\ &+ \sum_{i=1}^{\infty} \left( \sum_{j=i+1}^{\infty} n\lambda_j T_{j-i} x_j \right) \left[ f\left(x + \frac{1}{n}e_i\right) - f(x) \right]. \end{aligned} \tag{28}$$

Let

$$F(x) := \sum_{\ell} \beta_\ell(x) = -\lambda_1 x_1 e_1 + \sum_{i=1}^{\infty} \lambda_{i+1} x_{i+1} (2e_i - e_{i+1}) + \sum_{i=1}^{\infty} \left( \sum_{j=i+1}^{\infty} \lambda_j T_{j-i} x_j \right) e_i.$$

The convergence result of Kurtz ([7, Theorem 2.1, Chapter 11], [14, Theorem 8.1]) extends (without changing the proof) to the Banach space  $\ell^1(\mathbb{R})$  provided the same conditions are satisfied for  $\ell^1(\mathbb{R})$  as for  $\mathbb{R}^d$  in the theorem of Kurtz. Specifically, we require that for a compact set  $\mathcal{C}$  in  $\ell^1(\mathbb{R})$ ,

$$\sum_{\ell} \|\ell\|_1 \sup_{x \in \mathcal{C}} \beta_\ell(x) < \infty, \tag{29}$$

and there exists  $M_C > 0$  such that

$$\|F(x) - F(y)\|_1 \leq M_C \|x - y\|_1, \quad x, y \in \mathcal{C}. \tag{30}$$

Here the condition (29) follows from

$$\sum_i \sup_{x \in \mathcal{C}} |\lambda_i x_i| < \infty \quad \text{and} \quad \sum_i \sup_{x \in \mathcal{C}} \sum_{j=i+1}^{\infty} \lambda_j T_{j-i} |x_j| < \infty,$$

which in turn follow from conditions (25). Similarly, Lipschitz conditions (30) are satisfied in  $\mathcal{C}$  due to conditions (25). Thus, by Kurtz ([7, Theorem 2.1, Chapter 11], [14, Theorem 8.1]), the process  $x^{(n)}(s)$  converges almost surely to  $x(s)$  that satisfies  $\dot{x} = F(x)$ , which expands as the following system of ordinary differential equations:

$$\begin{cases} x_1'(s) &= -\lambda_1 x_1 + \lambda_2(T_1 + 2)x_2 + \lambda_3 T_2 x_3 + \dots \\ x_2'(s) &= -\lambda_2 x_2 + \lambda_3(T_1 + 2)x_3 + \lambda_4 T_2 x_4 + \dots \\ &\vdots \\ x_k'(s) &= -\lambda_k x_k + \lambda_{k+1}(T_1 + 2)x_{k+1} + \lambda_{k+2} T_2 x_{k+2} + \dots \\ &\vdots \end{cases} \tag{31}$$

with the initial conditions  $x(0) = \lim_{n \rightarrow \infty} x^{(n)}(0) = \pi := \sum_{K=1}^{\infty} p_K e_K$  by the law of large numbers. Finally, we observe that  $\|\pi\|_1 = 1$ , and conditions (25) imply that  $\mathbb{G}A$  is a bounded operator in  $\ell^1(\mathbb{R})$ .  $\square$

### 5.3. Criticality and time invariance

Assume that the hydrodynamic limit  $x(s)$ , and hence the averages  $x_j(s)$ , exist. Let  $\pi = \sum_{K=1}^{\infty} p_K e_K$ . Then one can consider the average progeny of the process, that is the average number of branches of any order alive at instant  $s$ :

$$C(s) = \sum_{j=1}^{\infty} x_j(s) = \|e^{\mathbb{G}As} \pi\|_1.$$

In hydrological literature, an empirical version of the process  $C(s)$  is called the *width function* of a tree  $T[S]$ .

**Definition 14.** A hierarchical branching process  $S(s)$  is said to be *critical* if and only if the width function  $C(s) = 1$  for all  $s \geq 0$ .

**Definition 15.** A hierarchical branching process  $S(s)$  is said to be *time-invariant* if and only if

$$e^{\mathbb{G}As} \pi = \pi \quad \text{for all } s \geq 0. \tag{32}$$

**Proposition 7.** Suppose that the hydrodynamic limit  $x(s)$  exists, and  $\pi$  is time-invariant. Then the process  $S(s)$  is critical.

**Proof.**  $C(s) = \|x(s)\|_1 = \|e^{\mathbb{G}As} \pi\|_1 = \|\pi\|_1 = 1$ .  $\square$

Let  $\hat{t}(z) = -1 + 2z + \sum_j z^j T_j$  for  $|z| < 1/L$ , where  $L$  is defined in (24). Observe that there is a unique real root  $w_0$  of  $\hat{t}(z)$  within  $(0, \frac{1}{2}]$ . We formulate our results in terms of the Horton exponent  $R := w_0^{-1}$  (e.g., [11,20]).

**Proposition 8.** *Suppose  $\Lambda\pi$  is a constant multiple of the geometric vector  $v_0 = \sum_{K=1}^\infty R^{-K} e_K$ . Then the process  $S(s)$  is time-invariant.*

**Proof.** Observe that since  $\hat{t}(R^{-1}) = 0$  and  $\mathbb{G}$  is a Toeplitz operator,

$$\mathbb{G}v = \hat{t}(w)v \quad \text{for } v = \sum_{K=1}^\infty w^K e_K, \quad |w| < L.$$

and

$$\mathbb{G}v_0 = \hat{t}(R^{-1})v_0 = 0 \quad \text{for } v_0 := \sum_{K=1}^\infty R^{-K} e_K.$$

Hence  $\mathbb{G}\Lambda\pi = \hat{t}(R^{-1})\Lambda\pi = 0$  and

$$e^{\mathbb{G}\Lambda s}\pi = \pi + \sum_{m=1}^\infty \frac{s^m}{m!} (\mathbb{G}\Lambda)^m \pi = \pi. \quad \square$$

**Remark 4.** Proposition 8 states that the condition

$$\lambda_K p_K = b R^{-K}, \quad K \geq 1 \tag{33}$$

is sufficient for time-invariance, for any proportionality constant  $b > 0$ . This implies that a time-invariant process can be constructed for

- (i) an arbitrary sequence of Tokunaga coefficients  $\{T_k\}$  satisfying (24) — by selecting  $\lambda_K p_K = b R^{-K}$ ;
- (ii) arbitrary sequences  $\{T_k\}$  satisfying (24) and  $\{p_K\}$  — by selecting  $\lambda_K = b R^{-K} p_K^{-1}$ ;
- (iii) arbitrary sequences  $\{T_k\}$  satisfying (24) and  $\{\lambda_K\}$  — by selecting  $p_K = b R^{-K} \lambda_K^{-1}$ .

At the same time, arbitrary sequences  $\{\lambda_K\}, \{p_K\}$  will not, in general, satisfy (33) and hence will not correspond to a time-invariant process.

### 5.4. Criticality and time-invariance in a self-similar process

A convenient characterization of criticality can be established for self-similar hierarchical branching processes. Recall that by Theorem 4, part (3), a self-similar process  $S(s)$  is specified by parameters  $\gamma > 0, 0 < p < 1$  and length self-similarity constant  $\zeta > 0$  such that  $p_K = p(1 - p)^{K-1}$  and  $\lambda_j = \gamma \zeta^{-j}$ . We refer to a self-similar process by its parameter triplet,  $S(s) = S_{p,\gamma,\zeta}(s)$ , and denote the respective width function by  $C_{p,\gamma,\zeta}(s)$ . Observe that in the self-similar case the first of the conditions (25) is equivalent to  $\zeta \geq 1$ , and the second is equivalent to  $\zeta \geq L$ . Hence, the conditions (25) are equivalent to  $\zeta \geq 1 \vee L$ .



**Theorem 6** (Width Function of a Self-Similar Process). Consider a self-similar process  $S_{p,\gamma,\zeta}(s)$  with  $0 < p < 1, \gamma > 0$ . Suppose that (24) is satisfied and  $\zeta \geq 1 \vee L$ . Then

$$C_{p,\gamma,\zeta}(s) \begin{cases} \text{decreases} & \text{if } p > 1 - \frac{\zeta}{R}, \\ = 1 & \text{if } p = 1 - \frac{\zeta}{R}, \\ \text{increases} & \text{if } p < 1 - \frac{\zeta}{R}. \end{cases}$$

**Proof.** The choice of the limits for  $\zeta$  ensures that the conditions (25) are satisfied and hence, by Theorem 5, the hydrodynamic limit  $x(s)$  exists and the width function  $C_{p,\gamma,\zeta}(s)$  is well defined. Now we have

$$\Lambda\pi = \frac{\gamma p}{1-p} \sum_{K=1}^{\infty} (\zeta^{-1}(1-p))^K e_K,$$

and therefore

$$\mathbb{G}\Lambda\pi = \hat{t}(\zeta^{-1}(1-p))\Lambda\pi. \tag{34}$$

Iterating recursively, we obtain

$$(\mathbb{G}\Lambda)^2\pi = \hat{t}(\zeta^{-1}(1-p))\mathbb{G}\Lambda^2\pi = \hat{t}(\zeta^{-1}(1-p))\hat{t}(\zeta^{-2}(1-p))\Lambda^2\pi,$$

and in general,

$$(\mathbb{G}\Lambda)^m\pi = \hat{t}(\zeta^{-1}(1-p))\mathbb{G}\Lambda^m\pi = \left[ \prod_{i=1}^m \hat{t}(\zeta^{-i}(1-p)) \right] \Lambda^m\pi.$$

Thus, taking  $x(0) = \pi$ ,

$$x(s) = e^{\mathbb{G}\Lambda s}\pi = \pi + \sum_{m=1}^{\infty} \frac{s^m}{m!} \left[ \prod_{i=1}^m \hat{t}(\zeta^{-i}(1-p)) \right] \Lambda^m\pi. \tag{35}$$

The width function for the given values of  $p \in (0, 1), \gamma > 0$  and  $\zeta \geq 1$  can therefore be expressed as

$$\begin{aligned} C_{p,\gamma,\zeta}(s) &= \sum_{j=1}^{\infty} x_j(s) = 1 + \sum_{m=1}^{\infty} \frac{s^m}{m!} \left[ \prod_{i=1}^m \hat{t}(\zeta^{-i}(1-p)) \right] \sum_{j=1}^{\infty} (\Lambda^m\pi)_j \\ &= 1 + \sum_{m=1}^{\infty} \frac{(s\gamma/\zeta)^m}{m!} \left[ \prod_{i=1}^m \hat{t}(\zeta^{-i}(1-p)) \right] \frac{p}{1-\zeta^{-m}(1-p)} \end{aligned} \tag{36}$$

as  $\sum_{j=1}^{\infty} (\Lambda^m\pi)_j = \sum_{j=1}^{\infty} \lambda_j^m \pi_j = \sum_{j=1}^{\infty} \gamma^m \zeta^{-jm} p(1-p)^{j-1} = \gamma^m \zeta^{-m} \frac{p}{1-\zeta^{-m}(1-p)}$ .

Next, notice that by letting  $p' = 1 - \zeta^{-1}(1-p)$ , we have from (36) and the uniform convergence of the corresponding series for any fixed  $M > 0$  and  $s \in [0, M]$ , that

$$\frac{d}{ds} C_{p,\gamma,\zeta}(s) = \frac{\gamma}{\zeta} \hat{t}(1-p') C_{p',\gamma,\zeta}(s) \quad \text{with } C_{p,\gamma,\zeta}(0) = C_{p',\gamma,\zeta}(0) = 1. \tag{37}$$

Observe that  $\zeta \geq 1$  implies  $p' \geq p$  and  $C_{p',\gamma,\zeta}(s) \leq C_{p,\gamma,\zeta}(s)$ . Also, observe that

$$\hat{t}(1 - p') \begin{cases} < 0 & \text{if } p > 1 - \frac{\zeta}{R} \\ = 0 & \text{if } p = 1 - \frac{\zeta}{R} \\ > 0 & \text{if } p < 1 - \frac{\zeta}{R} \end{cases}$$

as  $\hat{t}$  is an increasing function on  $[0, \infty)$  and  $\hat{t}(1/R) = 0$ . This leads to the statement of the theorem.  $\square$

**Remark 5.** If  $\zeta = 1$ , Eq. (37) implies  $C_{p,\gamma,1}(s) = \exp\{s\gamma\hat{t}(1 - p)\}$  and hence

$$C_{p,\gamma,1}(s) \begin{cases} \text{exponentially decreases} & \text{if } p > 1 - R^{-1}, \\ = 1 \text{ for all } s \geq 0 & \text{if } p = 1 - R^{-1}, \\ \text{exponentially increases} & \text{if } p < 1 - R^{-1}. \end{cases}$$

This case is further examined in Section 5.5. In general, the width function  $C_{p,\gamma,\zeta}(s)$  may increase sub-exponentially for  $p < 1 - \frac{\zeta}{R}$ . For example, if there is a nonnegative integer  $d$  such that  $\zeta^{d+1} < R$ , then for  $p = 1 - \frac{\zeta^{d+1}}{R}$  we have  $\hat{t}(\zeta^{-d-1}(1 - p)) = 0$ . Hence, (35) implies that  $C_{p,\gamma,\zeta}(s)$  is a polynomial of degree  $d$ .

**Theorem 7 (Criticality of a Self-Similar Process).** Consider a self-similar process  $S_{p,\gamma,\zeta}(s)$  with  $0 < p < 1$ ,  $\gamma > 0$ . Suppose that (24) is satisfied and  $\zeta \geq 1 \vee L$ . Then the following conditions are equivalent:

- (i) The process is critical.
- (ii) The process is time-invariant.
- (iii) The following relations hold:  $\zeta < R$  and  $p = p_c := 1 - \frac{\zeta}{R}$ .

**Proof.** (i) $\leftrightarrow$ (iii) is established in Theorem 6. (ii) $\rightarrow$ (i) is established in Proposition 7. (iii) $\rightarrow$ (ii): Observe that  $\hat{t}(\zeta^{-1}(1 - p)) = \hat{t}(R^{-1}) = 0$ . Time invariance now follows from (35).  $\square$

**Remark 6.** In a self-similar process the sequences  $\lambda_K$  and  $p_K$  are geometric such that (Theorem 4)

$$\lambda_K p_K = \frac{\gamma p}{1 - p} \left( \frac{1 - p}{\zeta} \right)^K$$

for some  $0 < p < 1$ ,  $\gamma > 0$ , and  $\zeta \geq 1 \vee L$ . Hence, a time-invariant process can be constructed, according to Proposition 8 and (33), by selecting any sequence  $\{T_k\}$  that corresponds to

$$R = \zeta (1 - p)^{-1}.$$

Theorem 7 states that this is the only possible way to construct a time-invariant process, given that the process is self-similar.

### 5.5. A closed form solution for the case of equally distributed branch lengths

Observe that if  $\Lambda = \gamma I$ , then  $x(s) = e^{s\gamma\hat{t}(1-p)}\pi$  and  $C(s) = \|x(s)\|_1 = e^{s\gamma\hat{t}(1-p)s}$ .

Consider a hierarchical branching process with  $\Lambda = I$  and  $x(0) = e_K$  for a given integer  $K \geq 1$ . Here the system of Eq. (31) is finite dimensional,

$$\begin{cases} x'_1(s) &= -x_1 + (T_1 + 2)x_2 + T_2x_3 + \dots + T_{K-1}x_K \\ x'_2(s) &= -x_2 + (T_1 + 2)x_3 + T_2x_4 + \dots + T_{K-2}x_K \\ &\vdots \\ x'_{K-1}(s) &= -x_{K-1} + (T_1 + 2)x_K \\ x'_K(s) &= -x_K \end{cases} \tag{38}$$

with the initial conditions  $x(0) = e_K$ .

Define a sequence  $t(j)$  as

$$t(0) = -1, \quad t(1) = T_1 + 2, \quad \text{and } t(j) = T_j \text{ for } j \geq 2,$$

and let  $y(s) = e^s x(s)$ . Then (38) becomes

$$\begin{cases} y'_1(s) &= t(1)y_2 + t(2)y_3 + \dots + t(K - 1)y_K \\ y'_2(s) &= t(1)y_3 + t(2)y_4 + \dots + t(K - 2)y_K \\ &\vdots \\ y'_{K-2}(s) &= t(1)y_{K-1} + t(2)y_K \\ y'_{K-1}(s) &= t(1)y_K \\ y'_K(s) &= 0 \end{cases} \tag{39}$$

with the initial conditions  $y(0) = e_K$ . The ODEs (39) can be solved recursively in a reversed order of equations in the system obtaining for  $m = 1, \dots, K - 1$ ,

$$y_{K-m}(s) = \sum_{n=1}^m \left( \sum_{\substack{i_1, \dots, i_n \geq 1 \\ i_1 + \dots + i_n = m}} t(i_1) \cdot \dots \cdot t(i_n) \right) \frac{s^n}{n!}.$$

Let  $\delta_0(j) = I_{\{j=0\}}$  be the Kronecker delta function. Then we arrive with the closed form solution

$$x_{K-m}(s) = e^{-s} y_{K-m}(s) = e^{-s} \sum_{n=1}^{\infty} \underbrace{(t + \delta_0) * (t + \delta_0) * \dots * (t + \delta_0)}_{n \text{ times}}(m) \frac{s^n}{n!}. \tag{40}$$

Observe that if we randomize the orders of trees by assigning an order  $K$  to a tree with geometric probability  $p_K = p(1 - p)^{K-1}$ , then the above closed form expression (40) would yield an expression for the width function that was observed in Remark 5 of this section:

$$\begin{aligned} C(s) &= e^{-s} + e^{-s} \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} (1 - p)^m \underbrace{(t + \delta_0) * (t + \delta_0) * \dots * (t + \delta_0)}_{n \text{ times}}(m) \frac{s^n}{n!} \\ &= e^{-s} + e^{-s} \sum_{n=1}^{\infty} (\hat{t}(1 - p) + 1)^n \frac{s^n}{n!} = \exp \{s\hat{t}(1 - p)\}. \end{aligned}$$

### 5.6. Critical Galton–Watson process

The critical binary Galton–Watson process plays an important role in theory and applications because of its multiple symmetries. Burd, Waymire and Winn [4] have shown that the following three properties are equivalent for the binary Galton–Watson distributions  $\mathcal{GW}(p_0, p_2)$ : (i) A

distribution is prune-invariant; (ii) A distribution is mean self-similar with  $E(T_{i,j}) = T_{j-i} = 2^{j-i-1}$ ; and (iii) A distribution is critical:  $p_0 = p_2 = 1/2$ . The Markov structure of the critical Galton–Watson trees ensures the existence of two other special properties: (iv) Time-invariance (in discrete time): the forest of trees, obtained by removing the edges and the vertices below depth  $d$ , has the same frequency structure as the original space  $\mathcal{GW}(1/2, 1/2)$ ; and (v) The forest of trees obtained by considering subtrees rooted at every vertex of a random tree  $T$  approximates the frequency structure of the entire space of trees when the order of  $T$  increases.

The next result shows that the critical binary Galton–Watson tree is a special case of the hierarchical branching process.

**Theorem 8 (Critical Galton–Watson Tree).** *A hierarchical branching process with parameters*

$$\lambda_j = \gamma 2^{2-j}, \quad p_K = 2^{-K}, \quad \text{and } T_k = 2^{k-1} \text{ for any } \gamma > 0 \tag{41}$$

*is distributionally equivalent to the critical binary Galton–Watson tree  $\mathbf{GW}(0, \gamma)$  with i.i.d. edge lengths that have a common exponential distribution with rate  $2\gamma$ . This is a self-similar, critical, and time-invariant process with*

$$R = 4, \quad L = 2, \quad \text{and } \zeta = 2.$$

**Proof.** Consider a tree  $T = \mathbf{GW}(0, \gamma) \in \mathcal{L}$ . By Corollary 1, each branch of order  $j$  in  $T$  is exponentially distributed with parameter  $\lambda_j = \gamma 2^{2-j}$ , which matches the branch length distribution in the hierarchical branching process (41). Furthermore, conditioned on  $\mathcal{R}^i(T) \neq \emptyset$  (which happens with a positive probability), we have  $\mathcal{R}^i(T) = \mathbf{GW}(0, 2^{-i}\gamma)$ . This means that the space  $\mathcal{R}^i(\mathcal{L})$  of pruned trees is a linearly scaled version of the original space  $\mathcal{L}$  (the same combinatorial structure, linearly scaled edge lengths). Burd et al. [4] have shown that the total number of sub-branches within a branch of order  $j \geq 2$  in  $T$  is geometrically distributed over  $\mathbb{Z}_+$  with mean  $T_1 + \dots + T_{j-1} = 2^{j-1} - 1$  (that is  $P(m) = 2^{1-j}(1 - 2^{1-j})^m$  for  $m \in \mathbb{Z}_+$ ), where  $T_{i,j} = T_{j-i} = 2^{j-i-1}$ . The assignment of orders among the  $m$  side-branches is done according to the multinomial distribution with  $m$  trials and success probabilities  $T_i/(T_1 + \dots + T_{j-1})$ ,  $i = 1, \dots, j - 1$ . This implies that, conditioning on a particular implementation of the pruned tree  $\mathcal{R}(T)$ , the leaves of the original tree merge into every branch of the pruned tree as a Poisson point process with intensity  $\gamma = \lambda_j T_{j-1}$ . Iterating this pruning argument, conditioning on the particular implementation of  $\mathcal{R}^i(T) = \mathbf{GW}(0, 2^{-i}\gamma)$ , the branches of order  $i$  merge into any branch of the pruned tree  $\mathcal{R}^i(T)$  as a Poisson point process with intensity  $\gamma 2^{1-i} = \lambda_j T_{j-i}$  for every  $j > i$ . Finally, the critical binary Galton–Watson space has  $p_K = 2^{-K}$  [4]. We, hence, conclude that a  $\mathbf{GW}(0, \gamma)$  tree is distributionally identical to the hierarchical branching process with parameters (41).

By Theorem 4, the process (41) is self-similar with the scaling exponent  $\zeta = 2$ . Criticality and time-invariance follow from Theorem 7.  $\square$

### 5.7. Critical Tokunaga processes

We introduce here a class of processes that extends the symmetries observed in the critical binary Galton–Watson tree with exponential edge lengths (where  $\zeta = 2$ ) to the general case of  $\zeta \geq 1$ . Specifically, consider a hierarchical branching process  $S^{\text{Tok}}(t; c, \gamma)$ , which we call the *critical Tokunaga branching process*, with parameters

$$\lambda_j = \gamma c^{2-j}, \quad p_K = 2^{-K}, \quad \text{and } T_k = (c - 1)c^{k-1} \text{ for any } \gamma > 0, c \geq 1. \tag{42}$$

**Proposition 9.** *The process  $S^{\text{Tok}}(t; c, \gamma)$  is a self-similar critical time invariant process. Independently of the process combinatorial shape, its edge lengths are i.i.d. exponential random variables with rate  $\gamma c$ . In addition, we have*

$$\hat{t}(z) = \frac{(1 - 2cz)(z - 1)}{1 - cz}, \quad R = w_0^{-1} = 2c, \quad \zeta = L = c, \quad \text{and } p_c = 2^{-1}.$$

**Proof.** Self-similarity follows from [Theorem 4](#). Criticality and time-invariance are established in [Theorem 7](#). The edge lengths property follows from [Proposition 6](#), part (4).  $\square$

**Remark 7.** The condition  $T_{i,i+k} = T_k = ac^{k-1}$  was first introduced in hydrology by Eiji Tokunaga [[25](#)] in a study of river networks, hence the process name. The additional constraint  $a = c - 1$  is necessitated here by the self-similarity of tree lengths, which requires the sequence  $\lambda_j$  to be geometric. The sequence of the Tokunaga coefficients then also has to be geometric, and satisfy  $a = c - 1$ , to ensure identical distribution of the edge lengths, see [Proposition 6](#), part (4). Interestingly, the constraint  $a = c - 1$  appears in the Random Self-similar Network (RSN) model introduced by Veitzer and Gupta [[26](#)], which uses a purely topological algorithm of recursive local replacement of the network generators to generate self-similar random trees. The importance of the constraint  $a = c - 1$  in a combinatorial situation is discussed in the next section.

### 6. Combinatorial critical Tokunaga process

This section focuses on the combinatorial structure of a tree generated by the critical Tokunaga process (leaving aside the edge lengths). A general combinatorial reformulation of the hierarchical branching process can be found in [[13](#)].

Consider a combinatorially self-similar (according to [Definition 11](#)) tree  $T = \text{SHAPE}(T[S]) \in \mathcal{T}$  generated by a hierarchical branching process  $S$  with Tokunaga sequence  $\{T_k\}$  and initial distribution  $p_K = p(1 - p)^{K-1}$ . Let random variable  $K$  be the order of the tree  $T$ , and, conditioned on  $K > 1$ , let  $K_a, K_b$  be the orders of its two subtrees,  $T_a$  and  $T_b$ , rooted at the internal vertex closest to the root, randomly and uniformly permuted. We call  $T_a$  and  $T_b$  the *principal subtrees* of  $T$ . Observe that the pair  $K_a, K_b$  uniquely defines the tree order  $K$ :

$$K = \begin{cases} K_a \vee K_b, & \text{if } K_a \neq K_b, \\ K_a + 1, & \text{if } K_a = K_b. \end{cases}$$

Let  $K_1 \leq K_2$  be the order statistics of  $K_a, K_b$ . The joint distribution of  $(K_1, K_2)$  is given by

$$P(K_1 = j, K_2 = m | K = k) = \begin{cases} \frac{1}{1 + T_1 + \dots + T_{k-1}} & \text{if } j = m = k - 1 \\ \frac{T_{k-j}}{1 + T_1 + \dots + T_{k-1}} & \text{if } j < m = k, \end{cases} \tag{43}$$

where

$$P(K = k | K > 1) = (1 - p)p^{k-2}.$$

**Proposition 10.** *Consider a critical Tokunaga process  $S^{\text{Tok}}(t; c, \gamma)$ . Then, conditioned on  $K > 1$ , the marginal order distribution of  $K_a$  coincides with that of  $K$ :*

$$P(K_a = m | K > 1) = 2^{-m} = p_m \quad \text{for } m \geq 1. \tag{44}$$

At the same time, the joint distribution of  $(K_a, K_b)$  equals the product of the marginals,

$$\mathbb{P}(K_a = m, K_b = j \mid K > 1) = \mathbb{P}(K_a = m \mid K > 1)\mathbb{P}(K_b = j \mid K > 1), \tag{45}$$

if and only if  $c = 2$ .

**Proof.** Observe that for  $k > 1$ ,

$$\begin{aligned} &\mathbb{P}(K_a = m \mid K = k) \\ &= \begin{cases} \frac{1}{2} \sum_{j:j < k} \mathbb{P}(K_1 = j, K_2 = k \mid K = k) & \text{if } m = k, \\ \mathbb{P}(K_1 = K_2 = k - 1 \mid K = k) \\ \quad + \frac{1}{2} \mathbb{P}(K_1 = k - 1, K_2 = k \mid K = k) & \text{if } m = k - 1, \\ \frac{1}{2} \mathbb{P}(K_1 = m, K_2 = k \mid K = k) & \text{if } m < k - 1, \end{cases} \\ &= \begin{cases} \frac{1}{2} \frac{T_1 + \dots + T_{k-1}}{1 + T_1 + \dots + T_{k-1}} = \frac{1 - c^{1-k}}{2} & \text{if } m = k, \\ \frac{1 + \frac{1}{2}T_1}{1 + T_1 + \dots + T_{k-1}} = \frac{c + 1}{2} c^{1-k} & \text{if } m = k - 1, \\ \frac{1}{2} \frac{T_{k-m}}{1 + T_1 + \dots + T_{k-1}} = \frac{(c - 1)c^{-m}}{2} & \text{if } m < k - 1. \end{cases} \end{aligned}$$

Thus,

$$\mathbb{P}(K_a = m \mid K > 1) = (1 - c^{1-m})2^{-m} + c^{-m}2^{-m} + \frac{(c - 1)c^{-m}}{2} \sum_{k=m+1}^{\infty} 2^{1-k} = 2^{-m}.$$

At the same time,

$$\mathbb{P}(K_a = m, K_b = j \mid K > 1) = \begin{cases} (c - 1)c^{-j} 2^{-m} & \text{if } j < m, \\ c^{-m} 2^{-m} & \text{if } j = m. \end{cases}$$

Hence, (45) holds if and only if  $c = 2$ .  $\square$

**Remark 8.** Proposition 10 asserts that the principal subtrees in a random critical Tokunaga tree are dependent, except the critical Galton–Watson case. This implies that, in general, non-overlapping subtrees within a critical Tokunaga tree are dependent. Accordingly, the increments of the Harris path  $H$  of a critical Tokunaga process have (long-range) dependence. The only exception is the case  $c = 2$  that was discussed in Section 4.6. The structure of  $H$  is hence reminiscent of a self-similar random process [17,22]. Establishing the correlation structure of the Harris paths of critical Tokunaga processes is an interesting open problem (see Section 7).

The critical Tokunaga trees introduced in Proposition 9 have an additional important property: the frequencies of vertex orders in a large-order tree approximate the frequencies of orders in the entire space  $\mathcal{T}$ . To formalize this observation, let  $\mu$  be the measure on  $\mathcal{T}$  induced by  $S^{\text{Tok}}(t; c, \gamma)$ , i.e.  $\mu(T) = \mathbb{P}(\text{SHAPE}(T[S^{\text{Tok}}]) = T)$ . Next, for a fixed  $K \geq 1$ , let  $\mu_K(T) = \mu(T \mid T \in \mathcal{H}_K)$ . Let  $V_k[K]$  denote the number of vertices of order  $k \in \{1, \dots, K\}$  in a tree generated by  $\mu_K$ , and let  $\mathcal{V}_k[K] = \mathbb{E}(V_k[K])$ . Finally, we denote by  $V[K] = \sum_{k=1}^K V_k[K]$  the total number of non-root vertices, and notice that  $V[K] = 2\mathcal{V}_1[K] - 1$ . Thus,  $\mathcal{V}[K] := \mathbb{E}(V[K]) = 2\mathcal{V}_1[K] - 1$ .

**Proposition 11.** *Let  $S^{\text{Tok}}(t; c, \gamma)$  be a critical Tokunaga branching process, then*

$$\lim_{K \rightarrow \infty} \frac{\mathcal{V}_k[K]}{\mathcal{V}_1[K]} = 2^{1-k}. \tag{46}$$

*Let  $T = \text{SHAPE}(T[S^{\text{Tok}}]) \in \mathcal{H}_K$  be a tree generated by  $\mu_K$ , and let  $v$  be a vertex selected by uniform random drawing from the non-root vertices of  $T$ . Then*

$$\lim_{K \rightarrow \infty} \mathbf{P}(v \text{ has order } k) = p_k = 2^{-k}. \tag{47}$$

**Proof.** It has been shown in [11] that the mean self-similar trees satisfy the strong Horton law:

$$\lim_{K \rightarrow \infty} \frac{\mathcal{N}_k[K]}{\mathcal{N}_1[K]} = R^{1-k}, \text{ for any } k \geq 1.$$

Observe now that for any  $T \in \mathcal{H}_K$  we have

$$\mathcal{V}_k(T) = \sum_{i=1}^{N_k(T)} (1 + m_i(T)),$$

where  $m_i(T)$  is the number of sub-branches that merge the  $i$ th branch of order  $k \in \{1, \dots, K\}$  in  $T$ , according to the proper branch labeling of Section 2.2. Proposition 6 gives

$$\mathcal{V}_k[K] = \mathcal{N}_k[K](1 + T_1 + \dots + T_{k-1}).$$

For the process  $S^{\text{Tok}}(t; c, \gamma)$  this implies

$$\lim_{K \rightarrow \infty} \frac{\mathcal{V}_k[K]}{\mathcal{V}_1[K]} = \lim_{K \rightarrow \infty} \frac{\mathcal{N}_k[K](1 + T_1 + \dots + T_{k-1})}{\mathcal{N}_1[K]} = R^{1-k} c^{k-1} = 2^{1-k}.$$

The statement (47) is an immediate consequence of (46), since

$$\lim_{K \rightarrow \infty} \frac{\mathcal{V}_k[K]}{\mathcal{V}[K]} = \lim_{K \rightarrow \infty} \frac{\mathcal{V}_k[K]}{2\mathcal{V}_1[K] - 1} = 2^{-k}$$

as  $\mathcal{V}_1[K] \geq 2^{K-1}$ .  $\square$

Proposition 11 has an immediate extension to trees with edge lengths, which we include here for completeness. A tree  $T \in \mathcal{L}$  can be considered a metric space with distance  $d(a, b)$  between two points  $a, b \in T$  defined as the length of the shortest path within  $T$  connecting them; see [21, Sect. 7.3] for details.

**Proposition 12.** *Consider a random tree  $T = T[S^{\text{Tok}}] \in \mathcal{L}$  generated by a critical Tokunaga branching process  $S^{\text{Tok}}(t; c, \gamma)$  conditioned on the order  $k(T) = K$ . Let point  $u \in T$  be sampled from a uniform density function on the metric space  $T$ , and let  $r_u[K]$  denote the order of the edge to which the point  $u$  belongs. Then*

$$\lim_{K \rightarrow \infty} \mathbf{P}(r_u[K] = k) = p_k = 2^{-k}. \tag{48}$$

**Proof.** Proposition 9 establishes that the edge lengths in  $T$  are i.i.d. exponential random variables. Thus we can generate  $T$  by first sampling the combinatorial tree  $\text{SHAPE}(T)$  from  $\mathcal{H}_K$  according to conditional measure  $\mu_K(T) = \mu(T \in \mathcal{H}_K)$ , and then assigning i.i.d. exponential edge lengths. Provided that we already sampled  $\text{SHAPE}(T)$ , selecting the i.i.d. edge lengths and then selecting the point  $u \in T$  uniformly at random, and marking the edge that  $u$  belongs to, is equivalent to selecting a random edge uniformly from the edges of  $\text{SHAPE}(T)$ , in order of proper

labeling of Section 2.2. The order  $r_u[K]$  is uniquely determined by the edge to which  $u$  belongs. The statement now follows immediately from Proposition 11.  $\square$

### 7. Open problems

We conclude with two open problems, which refer to extending selected properties of the critical Galton–Watson tree with independent exponential edge lengths,  $\text{GW}(0, \gamma)$ , which is a special case of the hierarchical branding process (see Theorem 8), to a general case. Our formulations are intentionally informal, reflecting multiplicity of possible rigorous approaches to each of them. Here  $S(t) = (\{T_k\}, \{\lambda_j\}, \{p_K\})$  is a self-similar hierarchical branching process with

$$L = \limsup_{k \rightarrow \infty} T_k^{1/k} < \infty, \quad p_K = p(1 - p)^{K-1}, \quad \lambda_j = \gamma \zeta^{-j}$$

for some positive  $0 < p < 1, \gamma > 0$ , and  $\zeta \geq 1 \vee L$ .

**Open Problem 1.** Describe the correlation structure of the Harris path of  $S(t)$ . (The critical binary Galton–Watson tree with independent exponential edge lengths  $\text{GW}(0, \gamma)$  corresponds to a symmetric Markov chain with exponential jumps  $\{\frac{1}{2}, 2\gamma, 2\gamma\}$ , see Theorem 8.)

**Open Problem 2.** Establish a proper infinite-tree limit of  $S(t)$ , where the edge lengths go to zero and the tree length increases to infinity, that preserves a suitably defined limit version of the self-similarity property. Describe the respective limit Harris path processes. (The Harris path of the critical binary Galton–Watson tree  $\text{GW}(0, \gamma)$  can be rescaled to converge to excursion of the standard Brownian motion [15,18].)

### Acknowledgments

We are grateful to Ed Waymire for his continuing support and encouragement. YK wishes to thank Tom Kurtz for providing a feedback regarding infinite dimensional population processes during the workshop *Interplay of Stochastic and Deterministic Dynamics* at the Mathematical Biosciences Institute (MBI) held during February 22–26, 2016. We would like to thank Jim Pitman for suggesting relevant publications. Comments of an anonymous referee helped us to improve presentation of the results. We thank the participants of the conference *Random Trees and Maps : Probabilistic and Combinatorial Aspects*, June 6–10, 2016, Centre International de Rencontres Mathématiques (CIRM), Marseille, France, and the *31st Conference on Mathematical Geophysics* of the International Union of Geodesy and Geophysics, June 6–10, 2016, Paris, France to whom we presented an early version of this work. We acknowledge financial support from the National Science Foundation, awards NSF DMS-1412557 (Y.K.) and NSF EAR-1723033 (I.Z.)

### References

- [1] D.J. Aldous, Deterministic and stochastic models for coalescence (aggregation and coagulation): a review of the mean-field theory for probabilists, *Bernoulli* 5 (1999) 3–48. <http://dx.doi.org/10.2307/3318611>.
- [2] N. Berestycki, Recent progress in coalescent theory, *Ensaos Mat.* 16 (2009) 1–193 <https://arxiv.org/abs/0909.3985>.
- [3] J. Bertoin, *Random Fragmentation and Coagulation Processes*, Cambridge University Press, 2006. <http://dx.doi.org/10.1017/CBO9780511617768>.



- [4] G.A. Burd, E.C. Waymire, R.D. Winn, A self-similar invariance of critical binary Galton–Watson trees, *Bernoulli* 6 (2000) 1–21. <http://dx.doi.org/10.2307/3318630>.
- [5] L. Devroye, P. Kruszewski, A note on the Horton–Strahler number for random trees, *Inform. Process. Lett.* 56 (1994) 95–99. [http://dx.doi.org/10.1016/0020-0190\(95\)00114-R](http://dx.doi.org/10.1016/0020-0190(95)00114-R).
- [6] P.S. Dodds, D.H. Rothman, Scaling, universality, and geomorphology, *Annu. Rev. Earth Planet. Sci.* 28 (2000) 571–610. <http://dx.doi.org/10.1146/annurev.earth.28.1.571>.
- [7] S.N. Ethier, T.G. Kurtz, *Markov processes. characterization and convergence*, in: *Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics*, John Wiley & Sons, Inc., New York, 1986 MR0838085, ISBN:978-0-471-76986-6.
- [8] S.N. Evans, *Probability and Real Trees: Ecole d’Eté de Probabilités de Saint-Flour XXXV-2005*, Springer, 2007. <http://dx.doi.org/10.1007/978-3-540-74798-7>.
- [9] T.E. Harris, First passage and recurrence distribution, *Trans. Amer. Math. Soc.* 73 (1952) 471–486. <http://dx.doi.org/10.1090/S0002-9947-1952-0052057-2>.
- [10] R.E. Horton, Erosional development of streams and their drainage basins: Hydrophysical approach to quantitative morphology, *Geol. Soc. Am. Bull.* 56 (1945) 275–370. [http://dx.doi.org/10.1130/0016-7606\(1945\)56\[275:EDOSAT\]2.0.CO;2](http://dx.doi.org/10.1130/0016-7606(1945)56[275:EDOSAT]2.0.CO;2).
- [11] Y. Kovchegov, I. Zaliapin, Horton law in self-similar trees, *Fractals* 24 (2) (2016) 1650017. <http://dx.doi.org/10.1142/S0218348X16500171>.
- [12] Y. Kovchegov, I. Zaliapin, Horton self-similarity of Kingman’s coalescent tree, *Ann. Inst. H. Poincaré Probab. Statist.* 53 (3) (2017) 1069–1107. <http://dx.doi.org/10.1214/16-AIHP748>.
- [13] Y. Kovchegov, I. Zaliapin, Tokunaga self-similarity arises naturally from time invariance, *Chaos: Interdisciplinary J. Nonlinear Sci.* 28 (4) (2018) 041102. <http://dx.doi.org/10.1063/1.5029937>.
- [14] T.G. Kurtz, *Approximation of Population Processes*, Vol. 36, SIAM, 1981. <http://dx.doi.org/10.1137/1.9781611970333>.
- [15] J.F. Le Gall, The uniform random tree in a Brownian excursion, *Probab. Theory Related Fields* 96 (1993) 369–383. <http://dx.doi.org/10.1007/BF01292678>.
- [16] J.F. Le Gall, Random trees and applications, *Probab. Surv.* 2 (2005) 245–311. <http://dx.doi.org/10.1214/154957805100000140>.
- [17] M.M. Meerschaert, A. Sikorskii, *Stochastic Models for Fractional Calculus* (Vol. 43), Walter de Gruyter Studies in Mathematics, 2012. <http://dx.doi.org/10.1515/9783110258165>.
- [18] J. Neveu, J. Pitman, Renewal property of the extrema and tree property of the excursion of a one-dimensional Brownian motion, in: *Séminaire de Probabilités XXIII*, in: the series *Lecture Notes in Mathematics*, vol. 1372, 1989, pp. 239–247. <http://dx.doi.org/10.1007/BFb0083976>.
- [19] W.I. Newman, D.L. Turcotte, A.M. Gabriellov, Fractal trees with side branching, *Fractals* 5 (1997) 603–614. <http://dx.doi.org/10.1142/S0218348X97000486>.
- [20] S.D. Peckham, New results for self-similar trees with applications to river networks, *Water Resour. Res.* 31 (1995) 1023–1029. <http://dx.doi.org/10.1029/94WR03155>.
- [21] J. Pitman, *Combinatorial Stochastic Processes*, in: *Lecture Notes in Mathematics*, vol. 1875, Springer-Verlag, 2006. <http://dx.doi.org/10.1007/b11601500>.
- [22] G. Samorodnitsky, M.S. Taqqu, *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*, (Vol. 1), CRC press, 1994. ISBN0-412-05171-0.
- [23] R.L. Shreve, Statistical law of stream numbers, *J. Geol.* 74 (1966) 17–37. <http://dx.doi.org/10.1086/627137>.
- [24] A.N. Strahler, Quantitative analysis of watershed geomorphology, *Trans. Am. Geophys. Union* 38 (1957) 913–920. <http://dx.doi.org/10.1029/TR038i006p00913>.
- [25] E. Tokunaga, Consideration on the composition of drainage networks and their evolution, *Geogr. Rep. Tokyo Metrop. Univ.* 13 (1978) 1–27.
- [26] S.A. Veitzer, V.K. Gupta, Random self-similar river networks and derivations of generalized Horton Laws in terms of statistical simple scaling, *Water Resour. Res.* 36 (4) (2000) 10331048. <http://dx.doi.org/10.1029/1999WR900327>.
- [27] X.G. Viennot, Trees everywhere, in: *CAAP’90*, Springer Berlin Heidelberg, 1990, pp. 18–41. [http://dx.doi.org/10.1007/3-540-52590-4\\_38](http://dx.doi.org/10.1007/3-540-52590-4_38).
- [28] I. Zaliapin, Y. Kovchegov, Tokunaga and Horton self-similarity for level set trees of Markov chains, *Chaos Solitons Fractals* 45 (3) (2012) 358–372. <http://dx.doi.org/10.1016/j.chaos.2011.11.006>.
- [29] S. Zanardo, I. Zaliapin, E. Foufoula-Georgiou, Are American rivers Tokunaga self-similar? New results on fluvial network topology and its climatic dependence, *J. Geophys. Res.* 118 (2013) 166–183. <http://dx.doi.org/10.1029/2012JF002392>.