Random self-similar trees:
A mathematical theory of Horton laws[*]

Yevgeniy Kovchegov[†]

Department of Mathematics, Oregon State University
2000 SW Campus Way, Corvallis, OR 97331-4605
e-mail: kovchegy@math.oregonstate.edu

and

Ilya Zaliapin[‡]

Department of Mathematics and Statistics, University of Nevada Reno,
1664 North Virginia st., Reno, NV 89557-0084
e-mail: zal@unr.edu

Abstract: The Horton laws originated in hydrology with a 1945 paper by Robert E. Horton, and for a long time remained a purely empirical finding. Ubiquitous in hierarchical branching systems, the Horton laws have been rediscovered in many disciplines ranging from geomorphology to genetics to computer science. Attempts to build a mathematical foundation behind the Horton laws during the 1990s revealed their close connection to the operation of pruning – erasing a tree from the leaves down to the root. This survey synthesizes recent results on invariances and self-similarities of tree measures under various forms of pruning. We argue that pruning is an indispensable instrument for describing branching structures and representing a variety of coalescent and annihilation dynamics. The Horton laws appear as a characteristic imprint of self-similarity, which settles some questions prompted by geophysical data.

MSC 2010 subject classifications: Primary 05C05, 05C80; secondary 05C63, 58-02.

Contents

1. Introduction

Invariance of the critical binary Galton-Watson tree measure with respect to
pruning (erasure) that begins at the leaves and progresses down to the tree root
has been recognized since the late 1980s. Both continuous [123] and discrete
[34] versions of prunings have been studied. The prune-invariance of the trees
naturally translates to the symmetries of the respective Harris paths [80]. The
richness of such a connection is supported by the well-studied embeddings of the
Galton-Watson trees in the excursions of random walks and Brownian motions
(e.g. [125, 105, 135]). These observations provide a point of departure for this
survey of recent results on *prune-invariance* and *self-similarity* of tree measures
and related stochastic processes on the real line. The prune-invariant measures
are abundant; they allow a concise parameterization via *Tokunaga sequences*
(Sect. 3) and reveal a variety of structures and symmetries (e.g. Thms. 1, 9, 14,
17, 25, 26). Looking at random trees through a prism of *self-similarity*, one rec-
ognizes the renowned critical binary Galton-Watson tree (Sect. 5) as a boundary
point of a general family of *invariant Galton-Watson measures* (Sect. 6), and a
member of the most symmetric class of prune-invariant binary tree measures –
critical Tokunaga trees (Sect. 7). The main attention is paid here to a combina-
torial *Horton pruning* for finite trees (Sects. 2–9), yet we also consider infinite
and real trees, and more general forms of pruning (Sects. 10–12). The surveyed
results suggest that particular forms of pruning may underline the evolution of
familiar dynamical systems, such as Kingman's coalescent (Sect. 9) or ballistic
annihilation (Sect. 11), allowing their efficient analytical treatment. The survey
also poses new questions related to random self-similar trees.

   We begin by summarizing the key empirical observations that provided an
impetus for the topic (Sect. 1.1) and discussing the structure and main results
of this survey (Sect. 1.2). Here, we keep references to a minimum, and indicate
survey sections where one can find future information.

1.1. Early empirical evidence: self-similar river networks

The theory of random self-similar trees originates from hydrogeomorphology,
which supplied the key empirical observations related to analysis of river net-
works. The river studies commonly represent a stream network that drains a
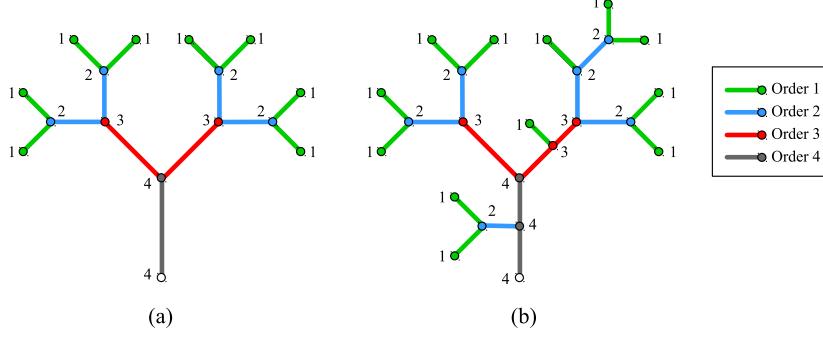
FIG 1. *Horton-Strahler orders in a binary tree. Different colors correspond to different orders of vertices and edges, as indicated in legend. (a) Perfect binary tree – orders are inversely proportional to vertex/edge depth. (b) General binary tree – orders are assigned according to the Horton-Strahler rule.*

single basin as a binary tree with planar embedding. The basin *outlet* (point furthest downstream) corresponds to the tree root, *sources* (points furthest upstream) to leaves, *junctions* (points where two streams meet) to internal vertices, and *links* (stream segments between two successive nodes) to edges. We use the graph-theoretic nomenclature below, although the original works commonly use the hydrologic terms italicized above.

Horton-Strahler orders (Sects. 2.4, 2.5). Informally, the aim of orders is to quantify the importance of vertices and edges in the tree hierarchy. It is natural to agree that the orders of a vertex and its parental edge are the same. Hence, we are only concerned with ordering vertices. In a perfect binary tree (where all leaves are located at the same depth, i.e., at the same distance from the root) one can assign orders inversely proportional to the vertex depth; see Fig. 1(a). In other words, we start with order 1 at the leaves and increase the order by unity with every step towards the root.

A celebrated ordering scheme that generalizes this idea to an arbitrary tree (not necessarily binary) was originally developed by Robert E. Horton [85], and later redesigned by Arthur N. Strahler [151] to its present form. It assigns integer orders to tree vertices and edges, beginning with order 1 at the leaves and increasing the order by unity every time a pair of edges of the same order meets at a vertex; see Fig. 1(b). A sequence of adjacent vertices/edges with the same order is called a *branch* (in hydrologic studies – *stream* or *channel*).

An example of Horton-Strahler ordering is shown in Fig. 2(a) for a small river network in the south-central US. Here, the orders serve as a good proxy for (a logarithm of) various physical characteristics of river streams: length, the area of the contributing basin, etc. The Horton-Strahler orders (a.k.a. Strahler numbers) provide an efficient ranking of the tree branches and have proven essential in numerous fields (see Sect. 4.4). As an example, the highest-order stream in a river basin commonly coincides with the basin's namesake river (e.g., Amazon river is the highest-order stream of the Amazon basin). One may find it quite
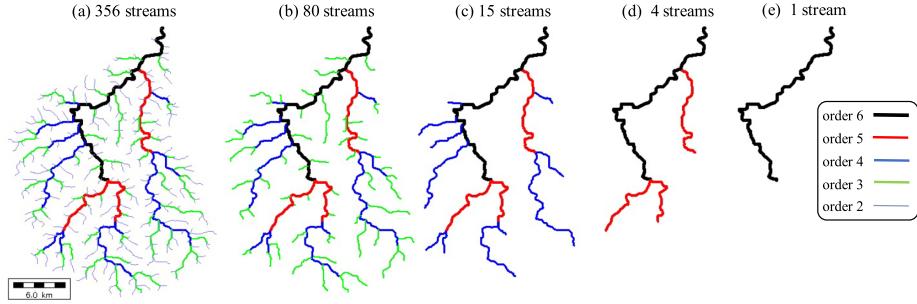
FIG 2. *Stream network of Beaver creek, Floyd County, KY. (a) Streams (branches) of orders* $K = 2, \ldots, 6$ *are shown by different colors (see legend on the right). Streams of order* 1 *are not shown for visual convenience. (b)–(e) Consecutive Horton prunings of the river network; uses the same color code for branch orders as panel (a). The channel extraction is done using* RiverTools *software (* http://rivix.com *).*

impressive that such an identification can be done using purely combinatorial properties of the river network. Further examples of Horton-Strahler ordering are shown in Figs. 10, 11, 12.

The Horton-Strahler orders are instrumental in describing geometry of river networks and connecting it to the river runoff and other surficial processes, as we discuss next.

Horton law and its implications (Sect. 4). In his pioneering study "of streams and their drainage basins" [85], Robert E. Horton described a geometric decay of the number N_K of branches (streams) of increasing Horton-Strahler order K. Specifically, the observed counts N_K in a large basin are closely approximated by

$$\frac{N_K}{N_{K+1}} = R_B \quad \Leftrightarrow \quad N_K \ \propto \ R_B^{-K} \tag{1}$$

for some *Horton exponent*[1] $R_B \geq 2$. The lower bound on R_B follows immediately from the definition of Horton-Strahler orders, since it takes at least two streams of order K to create a single stream of order $K + 1$. Figure 3(a) illustrates Horton law (1) for branch numbers in the Beaver creek network of Fig. 2(a); here $R_B \approx 4.55$.

Similar geometric relations, commonly referred to as *Horton laws* [131, 141], are found for multiple other stream statistics. This includes basin contributing area, basin magnitude (the number of sources), stream length, link slope, mean annual discharge, energy expenditure, etc. [141]. Figure 3(b) illustrates the Horton laws for the average magnitude M_K in a subbasin of order K, and the average number L_K of edges in a stream of order K in the Beaver creek network of Fig. 2(a). The respective Horton exponents here are $R_M \approx 4.55$ (for magnitude) and $R_L \approx 2.275$ (for edge number).

Horton laws play elemental role in statistical modeling of river basins, which

---

[1]Later in this survey, we only examine this Horton exponent and denote it simply by R.

rests upon empirical regularities that describe the frequencies of and relations among the key geometric and physical characteristics of individual streams. Remarkably, many such regularities heuristically follow from Horton laws and are parameterized by the respective Horton exponents. We give below several key examples, which are further discussed in Sect. 4:

- *Power-law frequencies of link statistics,* such as upstream contributing area, water discharge, or energy expenditure. Specifically, analysis [140] of river basins extracted from digital elevation models (DEM's) suggests

$$\#\{i : A_{(i)} \geq x\} \ \propto \ x^{-\beta} \quad \text{and} \quad \#\{i : E_{(i)} \geq x\} \ \propto \ x^{-2\beta},$$

 where $A_{(i)}$ is the area upstream of link i, $E_{(i)}$ is the energy expenditure at link i, and $\beta \approx 0.45$. These power law frequency distributions, as well as a deterministic relation between the respective exponents (β and 2β), are explained by Horton laws for the involved quantities (area, energy expenditure, link slope).
- *Power law relations among river statistics.* A well studied example is Hack's law that relates the length L of the longest stream in a basin to the basin area A via $L \propto A^h$ with $h \approx 0.6$ [78, 139]. The parameter h is expressed via the Horton exponents for length L and area A.
- *Modeling physical characteristics of a stream.* Classical hydrologic and geomorphologic studies of the mid 20-th century revealed that the key physical characteristics of streams – such as stream width, depth, slope, and flow velocity, can be modeled as power functions of the stream magnitude (number of upstream sources) [141, Chapter 1]. Hence, rather unexpectedly, essential physical characteristics of a river network can be calculated from purely combinatorial statistics of its tree representation. Horton laws for the stream magnitude and branch counts supply important constraints for the examined physical quantities.

It has been long recognized that Horton laws are related to self-similar (fractal) structure of river networks [154, 74]. Theorem 1 in Sect. 4 states a most general (to date) result on the topic; it shows that (a properly defined) Horton law for branch numbers necessarily holds in *mean self-similar trees* (Sect. 3). Section 7 introduces the Hierarchical Branching Process – a self-similar tree model that can reproduce arbitrary Horton exponents.

Horton pruning and its generalizations (Sects. 2.3, 10). The Horton-Strahler orders are naturally connected to the Horton pruning operation, which erases the leaves of a tree together with the adjacent edges, and removes the degree-2 vertices that might result from such erasure. Figure 2 illustrates a consecutive application of the Horton pruning to the Beaver creek network. The branches (streams) of order K are being erased at the K-th iteration of the Horton pruning. The mathematical theory of Horton laws concerns the tree measures that are invariant with respect to the Horton pruning. We also introduce a generalized dynamical pruning that allows one to erase a metric tree from the leaves down to the root in different ways, both continuous (metric) and discrete (combinatorial), and consider the respective prune-invariance.
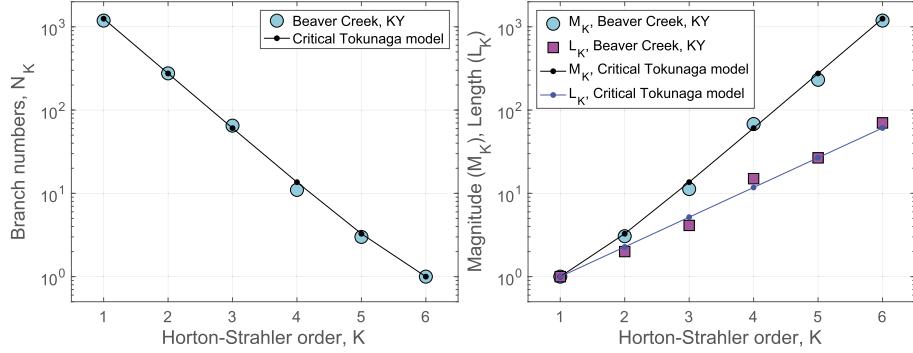
FIG 3. *Horton laws in the Beaver creek network of Fig. 2. (a) Number N_K of branches (streams) of order K. (b) Average magnitude (number of leaves) M_K in a subtree of order K. Average number L_K of edges in a branch (stream) of order K. Large circles and rectangles correspond to the Beaver creek data. Small dots and lines correspond to the critical Tokunaga process $S^{\mathrm{Tok}}(t; c, \gamma)$ of Sect. 7.5 with $c = 2.275$, $R = 2c = 4.55$ (γ is arbitrary, as it corresponds to the edge lengths that are not used in this analysis).*

Tokunaga model (Sects. 7.5, 7.6, 7.7). A notable observation inherited from the study of river networks is the Tokunaga constraint [155]. It complements the Horton law by describing the mergers of branches of distinct orders. Informally, the Tokunaga law suggests that the average number $\bar{N}_{i,j}$, $i < j$, of branches of order i that merge with a branch of order j in a given basin is an exponential function of the order difference, $\ln(\bar{N}_{i,j}) \propto j - i$. The Tokunaga model is surprisingly powerful in approximating the observed river networks [178] and predicting the values of multiple Horton exponents. Figure 3 shows how a one-parametric critical Tokunaga model S^{Tok} of Sect. 7.5 fits the average values of three branching statistics in the Beaver creek network.

In this work, we show the fundamental importance of the Toeplitz constraint $\bar{N}_{i,j} = f(j-i)$. We also provide a theoretical justification for the classical version of the Tokunaga law, which corresponds to a particular choice $\ln f(x) \propto x$.

1.2. Survey structure

Our primary goal is to survey the recent developments in the theory of random self-similar trees; yet a number of results, models, and approaches presented here are original. These novel results are motivated by the need to connect the dots and bridge the gaps when presenting a unified theory from the perspective of Horton pruning and its generalizations. We highlight some of these original contributions below in a list of survey topics.

The survey begins with the main definitions and notations in Sect. 2. This includes the definitions of finite rooted trees and tree spaces, and a brief overview of real trees. Next, *Horton pruning* and *Horton-Strahler orders* are introduced.

Section 3 defines the main types of invariances for tree measures sought-after in this survey. This includes a strong, distributional *Horton prune-invariance*

and a weaker *mean prune-invariance*. Importantly, we justify the requirement of *coordination*, which, together with mean prune-invariance, constitutes the mean self-similarity studied in this work. Every mean self-similar measure is associated with a sequence of nonnegative *Tokunaga coefficients* $\{T_k\}_{k\geq 1}$, which are theoretical analogs of the empirical averages $\bar{N}_{i,i+k} = f(k)$ discussed in Sect. 1.1. An important two-parameter family of the mean self-similar trees is *Tokunaga self-similar* trees specified by $T_k = ac^{k-1}$.

The *Horton law* for tree measures is formally defined in Sect. 4 in terms of the random counts $N_k[T]$ of branches of order k in a random tree T. We introduce two versions of the *strong Horton law*; one refers to convergence in probability (Def. 16) and the other to convergence of expectation ratios (Def. 17). The main result of the section (Thm. 1) establishes that the mean self-similarity implies the strong Horton law in expectation ratios, and expresses the Horton exponent R via the Tokunaga sequence $\{T_k\}$. Section 4.3 surveys computations of the entropy rate for trees that satisfy the strong Horton law, as a function of the Horton exponent $R \geq 2$ (Thm. 2), and for the Tokunaga self-similar trees, as a function of the Tokunaga parameters (a, c) (Thm. 3). These results emphasize a special role of the Horton exponent $R = 4$, which corresponds to the maximal possible entropy rate of unity (Fig. 21), and of the constraint $a = c - 1$, which corresponds to $R = 2c$ and achieves the maximal entropy rate among all trees with this value of Horton exponent (Fig. 22). These constraints correspond to the fundamental examples of self-similar trees; not surprisingly, they reappear in multiple later instances of our survey. Namely, $a = c - 1$, $R = 2c$ defines *critical Tokunaga trees* (Sect. 7), and a particular case of $a = 1$, $c = 2$, and $R = 4$ corresponds to the *critical binary Galton-Watson tree* (Sect. 5). The section concludes with a brief discussion of the applications of Horton-Strahler orders and Horton laws in natural and computer sciences.

Section 5 summarizes the results on the *critical binary Galton-Watson tree*. We start with a characterization theorem (Thm. 4), observing that this is the only mean self-similar tree among the combinatorial Galton-Watson trees with finite "$2 - \epsilon$" moment of the branching distribution $\{q_k\}$. Moreover, under consecutive application of Horton pruning, any such tree converges (conditioned on non-extinction) to the critical binary Galton-Watson tree (Thm. 5). The proofs of the strong Horton law for branch numbers (Cor. 2) and the Central Limit Theorem for branch numbers (Cor. 3) are novel, and emphasize the power of the pruning approach. We also find the length and height of the critical binary Galton-Watson tree with i.i.d. exponential edge lengths (Sect. 5.2), called the *exponential critical binary Galton-Watson tree.*

Section 6 includes new results. It extends the discussion of Sect. 5 by providing a complete description of invariants and attractors of the critical and subcritical Galton-Watson tree measures under the operation of Horton pruning. The prune-invariant measures, called *invariant Galton-Watson measures*, are critical; they are parameterized by the probability $q_0 \in [1/2, 1)$ of having no offspring. The case $q_0 = 1/2$ corresponds to the critical binary Galton-Watson tree of Sect. 5. Every other invariant measure has an unbounded branching, with distribution $q_k \sim k^{-(1+q_0)/q_0}$. The invariant Galton-Watson measures are the

only attractors of the critical Galton-Watson measures under consecutive Horton pruning; each of them has a non-empty domain of attraction (Thm. 10). The invariant Galton-Watson measures satisfy the strong Horton law with Horton exponent $R = (1 - q_0)^{-1/q_0}$ (Lem. 10).

Section 7 introduces a multi-type branching process, the *Hierarchical Branching Process (HBP)*, which is the main model of this work. The process trajectories are described by time oriented trees; this induces a probability measure on the space of planar binary trees with edge lengths. The HBP can generate trees with an arbitrary sequence of Tokunaga coefficients $\{T_k\}$. The combinatorial part of these trees is always mean self-similar; the measures are also (distributionally) Horton self-similar under additional constraints on the process parameters (Thm. 11). Section 7.3 defines the *criticality* for hierarchical branching processes as having a constant average progeny, in parallel with the definition of criticality for Galton-Watson processes. A hydrodynamic limit is established (Thm. 12) that describes the averaged branch dynamics as a deterministic system of ordinary differential equations (ODEs). This system of ODEs is used to detect a phase transition that separates fading and explosive behavior of the average process progeny (Thm. 13). A fundamental class of *critical Tokunaga processes* (Def. 25) is introduced that lies on the phase transition boundary and corresponds to $T_k = (c-1)c^{k-1}$ and $R = 2c$. The trees generated by the critical Tokunaga processes enjoy many of the symmetries seen in the exponential critical binary Galton-Watson tree; this includes Horton self-similarity, criticality, time-invariance, strong Horton law, Tokunaga self-similarity, and independence of edge lengths. The exponential critical binary Galton-Watson tree is a special case of the critical Tokunaga process with $c = 1$.

The results in Sect. 7.6 are original, applying martingale techniques in the analysis of self-similar trees. We introduce a Markov tree-valued process that generates the critical Tokunaga trees. We then construct a two-dimensional martingale with respect to the filtration of this process that describes the tree size and length. Doob's Martingale Convergence Theorem is applied to establish the strong Horton laws for the branch numbers (Thm. 16, Cor. 5) and the tree lengths (Cor. 7). The established Horton laws hold with the same Horton exponent $R = 2c$.

The *Geometric Branching Process* that describes the combinatorial part of a Horton self-similar HBP is examined in Sect. 7.7. We show, in particular, that invariance of this process with respect to the unit time shift is equivalent to the constraint $T_k = (c-1)c^{k-1}$ (i.e., $a = c-1$) on the Tokunaga coefficients (Thm. 17). This provides further justification for studying the critical Tokunaga processes. We show that the complete non-empty descendant subtrees in a combinatorial critical Tokunaga tree have the same distribution, and the two subtrees are independent if and only if the process is critical binary Galton-Watson (Cor. 8). Moreover, the empirical frequencies of edge/vertex orders in a large random critical Tokunaga tree approximate the order distribution in the respective space of trees (Props. 12, 13). This property is convenient for applied statistical analysis, where one might only be able to examine a handful of (large) trees.

Section 8 projects the Horton self-similarity results on time series via tree representation of continuous functions, a construction that goes back to Menger [117], Kronrod [92] and the celebrated Kolmogorov-Arnold representation theorem [9, 163]. The *level set tree* for a continuous function is defined following the well known pseudo-metric approach (161); see [4, 5, 105, 124, 52, 135]. We emphasize the connection of this construction with the Rising Sun Lemma (Lem. 19) of F. Riesz [138]. Proposition 15 reveals equivalence between Horton pruning and transition to the local extrema of a function. This allows us to interpret Horton self-similarity for the level set tree of a time series as invariance of this time series under transition to local extrema; see (171). An example of such an extreme-invariant process is given by the symmetric exponential random walk of Sect. 8.6.

The results in Sect. 8.5 are novel; they refer to the level set tree T of a positive excursion of a symmetric homogeneous random walk $\{X_k\}_{k \in \mathbb{Z}}$ on $\mathbb{R}$. The main result of this section (Thm. 18) shows that the combinatorial shape of T is distributed as the critical binary Galton-Watson tree, for any choice of the transition kernel for $\{X_k\}$. We also show (Lem. 21) that T has identically distributed edge lengths if and only if the transition kernel of $\{X_k\}$ is the probability density function of the Laplace distribution. The results of this section complement Thm. 20, a classical result on Galton-Watson representation of the level set tree of an exponential excursion, that can be found in [135, Lemma 7.3] and [105, 124].

Section 8.8 reveals a close connection between the level set tree of a sequence of i.i.d. random variables (discrete white noise) and the tree of the Kingman's coalescent process. Theorem 23 shows that the two trees are separated by a single Horton pruning.

Section 8.9 expands the level set tree construction to a Morse function defined on a multidimensional compact differentiable manifold. The key results from the Morse theory [121, 127, 36] are used to show (Cor. 19, Lem. 24) that the level set tree of a generic Morse function is binary. This result is new.

Section 9 establishes a weak form of Horton law for a tree representation of Kingman's coalescent process (Thm. 25). The proof is based on a Smoluchowski-type system of Smoluchowski-Horton ODEs (195) that describes evolution of the number of branches of a given Horton-Strahler order in a tree that represents Kingman's N-coalescent, in a hydrodynamic limit. Section 9.2 uses T. Kurtz's weak convergence results for density dependent population processes (Appendix A) to give a new, shorter than the original [97], derivation of the hydrodynamic limit. We present two alternative, more concise, versions of the Smoluchowski-Horton ODEs in (205) and (208), and use them to find a close numerical approximation to the Horton exponent in the Kingman's coalescent: $R = 3.0438279\ldots$. This exponent also applies to the level set tree of a discrete white noise, via the equivalence of Thm. 23 in Sect. 8.8.

Section 10 introduces the *generalized dynamical pruning* (218). This operation erases consecutively larger parts of a tree T, starting from the leaves and going down towards the root, according to a monotone nondecreasing pruning function φ along the tree. The generalized dynamical pruning encompasses a number

of discrete and continuous pruning operations, notably including the tree erasure of Jacques Neveu [123] (Sect. 10.1.1) and Horton pruning (Sect. 10.1.2). Important for our discussion, it generically includes erasures that do not satisfy the semigroup property (Sects. 10.1.3, 10.1.4). Theorem 26 establishes prune-invariance (Def. 34) of the exponential critical binary Galton-Watson tree with respect to a generalized dynamical pruning with an arbitrary admissible pruning function φ. The scaling exponents (Def. 34(ii)) for the pruning function φ equal to the tree length, tree height, and Horton-Starhler order are found in Thm. 27.

As an illuminating application of the generalized dynamical pruning, Sect. 11 examines the continuum 1-D ballistic annihilation model $A + A \to \emptyset$ for a constant initial particle density and initial velocity that alternates between the values of ± 1. The model dynamics creates coalescing shock waves, similar to those that appear in Hamilton-Jacobi equations [19], that have tree structure. We show (Cor. 21 of Thm. 28) that the shock tree is isometric to the level set tree of the initial potential (integral of velocity), and the model evolution is equivalent to a generalized dynamical pruning of the shock tree, with the pruning function equal to the total tree length (Thm. 30). This equivalence allows us to construct a complete probabilistic description of the annihilation dynamics for the initial velocity that alternates between the values of ± 1 at the epochs of a constant rate Poisson point process (Thms. 31, 32, 33). A real tree representation of the continuum ballistic annihilation is presented in Sect. 11.5.

Section 12 is novel. Here we construct an infinite level set tree, built from leaves down, for a time series $\{X_k\}_{k \in \mathbb{Z}}$. This gives a fresh perspective on multiple earlier results; e.g., those concerning the level set trees of random walks (Sect. 8.6), the generalized dynamical pruning (Sect. 10.5), or the evolution of an infinite exponential potential in the continuum annihilation model (Sect. 11.4). For instance, the infinite-tree version of prune-invariance for the exponential Galton-Watson tree (Thm. 34) can be established in a much simpler way than its finite counterpart (Thm. 26). Although this natural perspective has always influenced our research, this is the first time it is presented in an explicit form.

The survey concludes with a short list of open problems (Sect. 13).

Many concepts used in this survey are overlapping with the recent expositions on random trees, branching and coalescent processes by Aldous [4, 5, 6], Berestycki [25], Bertoin [29], Drmota [46], Duquesne and LeGall [52], Evans [63], Le Gall [108], Lyons and Peres [111], and Pitman [135]. We expect that the perspectives displayed in the present survey will with time connect and intertwine with better established topics in the theory of random trees.

2. Definitions and notations

2.1. Spaces of finite rooted trees

A connected acyclic graph is called a *tree* [32]. Consider the space $\mathcal{T}$ of finite unlabeled rooted reduced trees with no planar embedding. The (combinatorial) distance between a pair of tree vertices is the number of edges in the shortest
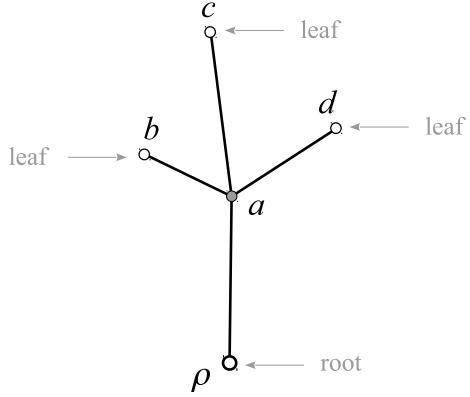
FIG 4. *Basic tree terminology: Illustration. Tree T consists of root ρ, one internal vertex a, and three leaves: b, c, d. Vertex ρ is the parent to a. Vertex a is the parent to b, c, d. Vertex a is the only offspring of ρ. Vertices b, c, d are offspring of a, and siblings to each other. Vertex a has degree 4. All other vertices have degree 1. Tree T is planted – its stem connects the root ρ and vertex a. The tree size is $\#T = 4$.*

path between them. A tree is called *rooted* if one of its vertices, denoted by ρ, is selected as the tree root (Fig. 4). The existence of root imposes a parent-offspring relation between each pair of adjacent vertices: the one closest to the root is called the *parent*, and the other the *offspring* (Fig. 4). The space $\mathcal{T}$ includes the *empty tree* ϕ comprised of a root vertex and no edges. The tree root is the only vertex that does not have parent. A *leaf* is a vertex with no offspring (Fig. 4). The absence of planar embedding in this context is the absence of order among the offspring of the same parent. We write $\#T$ for the number of non-root vertices, equal to the number of edges, in a tree T. Hence, a finite tree $T = \rho \cup \{v_i, \varepsilon_i\}_{1 \leq i \leq \#T}$ is comprised of the root ρ and a collection of non-root vertices v_i, each of which is connected to its unique parent $\mathsf{parent}(v_i)$ by the parental edge ε_i, $1 \leq i \leq \#T$. The degree of a vertex in a tree is the number of edges adjacent to the vertex (Fig. 4). Hence, the degree of a vertex in a rooted tree is the number of its offspring plus one. A tree is called *reduced* if it has no vertices of degree 2, with the root as the only possible exception.

The space of trees from $\mathcal{T}$ with positive edge lengths is denoted by $\mathcal{L}$. The trees in $\mathcal{L}$, also known as *weighted tree* [135, 111], can be considered as metric spaces with the metric denoted by $d(\cdot, \cdot)$. Specifically, the trees from $\mathcal{L}$ are isometric to one-dimensional connected sets comprised of a finite number of line segments that can share end points. The distance along tree paths is defined according to the Lebesgue measure on the edges. Each such tree can be embedded into $\mathbb{R}^2$ without creating additional edge intersections (see Fig. 5). Such a two-dimensional pictorial representation serves as the best intuitive model for the trees discussed in this work.

We write $\mathcal{T}_{\mathrm{plane}}$ and $\mathcal{L}_{\mathrm{plane}}$ for the spaces of trees from $\mathcal{T}$ and $\mathcal{L}$ with planar embedding, respectively. Any tree from $\mathcal{T}$ or $\mathcal{L}$ can be embedded in a plane

(a) $T^{(1)}$ (b) $T^{(2)}$ (c) $T^{(3)}$

(d) $T^{(4)}$ (e) $T^{(5)}$ (f) $T^{(6)}$
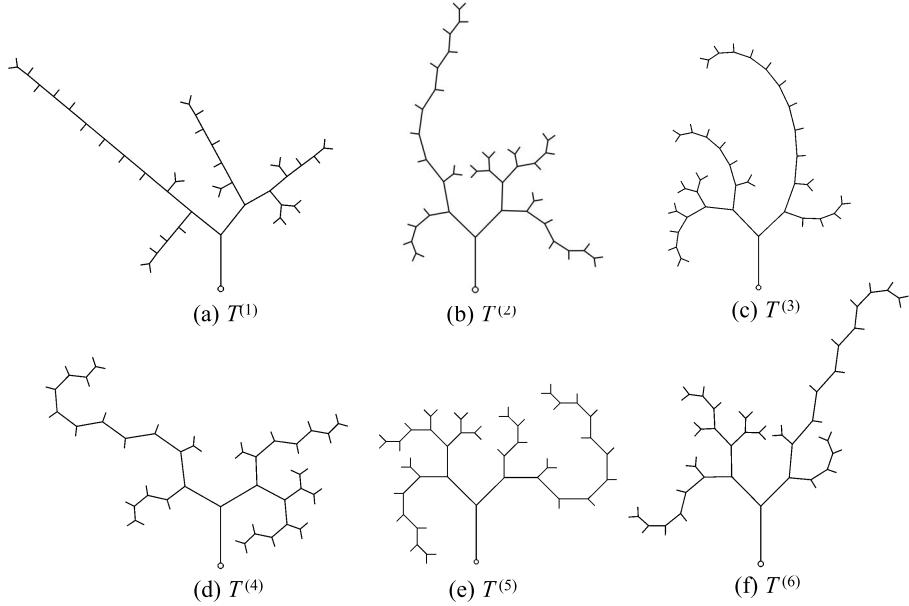
FIG 5. *Examples of alternative planar embeddings $T^{(i)} \in \mathcal{L}_{\mathrm{plane}}$, $i = 1, \ldots, 6$ of the same tree $T \in \mathcal{L}$, so that $\mathrm{SHAPE}(T^{(i)}) = T$. Different panels correspond to different (random) ordering of offspring of the same parent, and to different drawing styles.*

by selecting an order for the offsprings of the same parent. Unless indicated otherwise, the vertices of an embedded tree are indexed in order of depth-first search, starting from the root. Choosing different embeddings for the same tree $T \in \mathcal{T}$ (or $\mathcal{L}$) leads, in general, to different trees from $\mathcal{T}_{\mathrm{plane}}$ (or $\mathcal{L}_{\mathrm{plane}}$). Figure 5 illustrates alternative planar embeddings of a tree $T \in \mathcal{L}$. Planar embedding (offspring order) should not be confused with drawing style, related to how edges are represented in a plane. Each panel in Fig. 5 uses a separate drawing style.

Sometimes we focus on the combinatorial tree $\mathrm{SHAPE}(T)$, which retains the combinatorial structure of $T \in \mathcal{L}$ (or $\mathcal{L}_{\mathrm{plane}}$) while omitting its edge lengths and embedding. All tree shown in Fig. 5 have the same combinatorial structure. Similarly, the combinatorial tree $\mathrm{P\text{-}SHAPE}(T)$ retains the combinatorial structure of $T \in \mathcal{L}_{\mathrm{plane}}$ and planar embedding, and omits the edge length information. Here SHAPE is a projection from $\mathcal{L}$ or $\mathcal{L}_{\mathrm{plane}}$ to $\mathcal{T}$, and $\mathrm{P\text{-}SHAPE}$ is a projection from $\mathcal{L}_{\mathrm{plane}}$ to $\mathcal{T}_{\mathrm{plane}}$.

A non-empty rooted tree is called *planted* if its root has degree 1 (as in Figs. 4, 5); in this case the only edge connected to the root is called the *stem*. Otherwise the root has degree ≥ 2 and a tree is called *stemless*. We denote by $\mathcal{L}^{|}$ and $\mathcal{L}^{\vee}$ the subspaces of $\mathcal{L}$ consisting of planted and stemless trees, respectively. Hence $\mathcal{L} = \mathcal{L}^{|} \cup \mathcal{L}^{\vee}$. Also, we let the empty tree ϕ to be contained in each of the spaces. Therefore, $\mathcal{L}^{|} \cap \mathcal{L}^{\vee} = \{\phi\}$. Similarly, we write $\mathcal{L}^{|}_{\mathrm{plane}}$ and $\mathcal{L}^{\vee}_{\mathrm{plane}}$ for the

(a) Planted tree, $T \in \mathcal{L}^{|}_{\mathrm{plane}}$ (b) Stemless tree, $T \in \mathcal{L}^{\vee}_{\mathrm{plane}}$
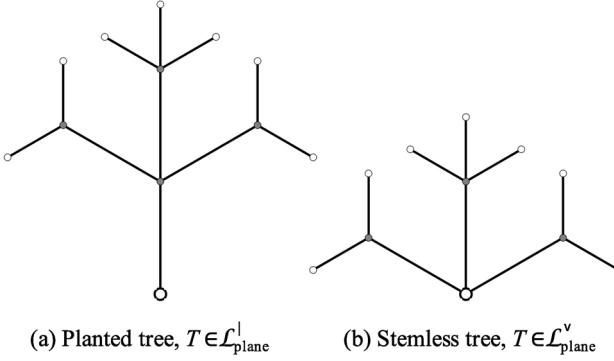
FIG 6. *Examples of planted (a) and stemless (b) trees. The combinatorial structure of both trees is the same, except the existence or absence of a stem. Internal vertices are marked by gray circles. Leaves are marked by small empty circles. Root is marked by large empty circle.*

subspaces of $\mathcal{L}_{\mathrm{plane}}$ consisting of planted and stemless trees, respectively. Clearly, $\mathcal{L}_{\mathrm{plane}} = \mathcal{L}^{|}_{\mathrm{plane}} \cup \mathcal{L}^{\vee}_{\mathrm{plane}}$ and $\mathcal{L}^{|}_{\mathrm{plane}} \cap \mathcal{L}^{\vee}_{\mathrm{plane}} = \{\phi\}$. Fig. 6 shows examples of a planted and a stemless tree.

For any space $\mathcal{S}$ from the list $\{\mathcal{T}, \mathcal{T}_{\mathrm{plane}}, \mathcal{L}, \mathcal{L}_{\mathrm{plane}}\}$ we write $\mathcal{BS}$ for the respective subspace of *binary* trees, $\mathcal{S}^{|}$ for the subspace of *planted* trees in $\mathcal{S}$ including ϕ, and $\mathcal{S}^{\vee}$ for the subspace of *stemless* trees in $\mathcal{S}$ including ϕ. We also consider subspaces $\mathcal{BS}^{|} = \mathcal{S}^{|} \cap \mathcal{BS}$ of *planted binary* trees and $\mathcal{BS}^{\vee} = \mathcal{S}^{\vee} \cap \mathcal{BS}$ of *stemless binary* trees. Most of the discussion and results in this work refer to planted trees.

Let $l_T = (l_1, \ldots, l_{\#T})$ with $l_i > 0$ be the vector of edge lengths of a tree $T \in \mathcal{L}$ (or $\mathcal{L}_{\mathrm{plane}}$). The length of a tree T is the sum of the lengths of its edges:

$$\mathrm{LENGTH}(T) = \sum_{i=1}^{\#T} l_i.$$

Recall that each weighted tree $T \in \mathcal{L}$ is itself a metric space with the metric denoted by $d(\cdot, \cdot)$. The height of a tree T is the maximal distance between the root and a vertex:

$$\mathrm{HEIGHT}(T) = \max_{1 \leq i \leq \#T} d(v_i, \rho).$$

2.2. Real trees

It is often natural to consider metric trees with structures more complicated than that allowed by finite spaces $\mathcal{L}$ and $\mathcal{L}_{\mathrm{plane}}$. In such cases, we use the following general definition.

Definition 1 (Metric tree [135, Sect. 7]). *A metric space (M, d) is called a tree if for each choice of $u, v \in M$ there is a unique continuous path $\sigma_{u,v}$:*

(a) Four point condition   (b) Three point condition
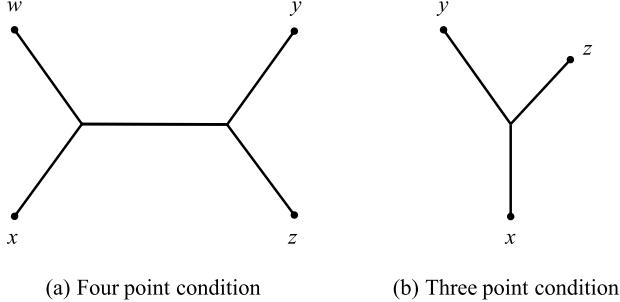
FIG 7. *Equivalent conditions for* 0-*hyperbolicity of a metric space* (M, d). *(a) Four point condition: any quadruple* $w, x, y, z \in M$ *is geodesically connected as shown in the figure. This configuration is algebraically expressed in Eq.* (2). *(b) Three point condition: any triplet* $x, y, z \in X$ *is geodesically connected as shown in the figure. There is no algebraic equivalent of the three point condition in terms of the lengths of the shown segments.*

$[0, d(u, v)] \to M$ *that travels from* u *to* v *at unit speed, and for any simple continuous path* $F : [0, L] \to M$ *with* $F(0) = u$ *and* $F(L) = v$, *the ranges of* F *and* $\sigma_{u,v}$ *coincide.*

As an example of a metric tree that does not belong to $\mathcal{L}_{\mathrm{plane}}$, consider a unit disk in the complex plane $M = \{z \in \mathbb{C} : |z| \leq 1\}$ and connect each point $z \in M$ to the origin $\mathbf{0}$ by a linear segment $[z, \mathbf{0}]$. Distances between points are computed in a usual way, but only along such segments. This is a tree whose (uncountable) set of leaves coincides with the unit circle $\{|z| = 1\}$. We refer to a book of Steve Evans [63] for a comprehensive discussion and further examples. Sects. 8, 11 of the present survey examine several natural constructions of a metric d on an n-dimensional manifold M with $n \geq 1$, such that (M, d) becomes a (one-dimensional) tree according to Def. 1.

Consider a metric tree $T = (M, d)$. As with the finite trees, we let $\rho \in M$ denote the root vertex of T. For any two points $x, y \in M$, we define a *segment* $[x, y] \subset M$ to be the image of the unique path $\sigma_{x,y}$ of the above definition. We call a point $y \in M$ a *descendant* of $x \in M$ if the path $[\rho, y]$ includes x. Equivalently, removing x from the tree separates its descendants from the root. To lighten the notations, we conventionally say $x \in T$ to indicate that point $x \in M$ belongs to tree T.

Metric trees benefit from an alternative characterization. Recall that a metric space (M, d) is called 0-hyperbolic, if any quadruple $w, x, y, z \in M$ satisfies the following *four point condition* [63, Lemma 3.12]:

$$d(w, x) + d(y, z) \leq \max\{d(w, y) + d(x, z), d(x, y) + d(w, z)\}. \qquad (2)$$

The four point condition is an algebraic description of an intuitive geometric constraint on geodesic connectivity of quadruples that is shown in Fig. 7(a). An equivalent way to define 0-hyperbolicity is the three point condition illustrated in Fig. 7(b). It is readily seen that the four point condition is satisfied by any finite

(a) Original tree   (b) Tree after series reduction
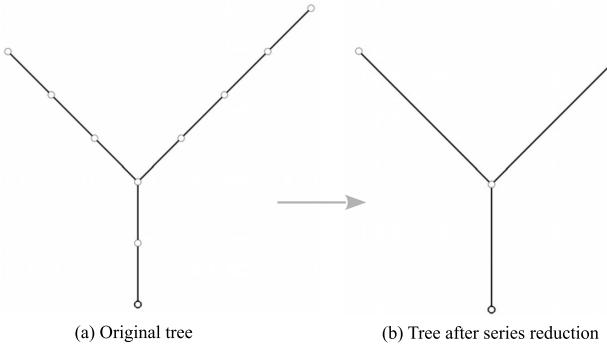
Fɪɢ 8. *Series reduction: Illustration.*

tree with edge lengths (considered as a metric space). In general, a connected and 0-hyperbolic metric space is called a *real tree*, or $\mathbb{R}$-tree [63, Theorem 3.40]. We denote by $\Delta_{p,T}$ the descendant tree at point p, that is the set of all descendants of point $p \in T$, including p as the tree root. The set of all descendant leaves of point p is denoted by $\Delta_{p,T}^{\circ}$. We use real trees in Sect. 11 to represent the dynamics of a continuum ballistic annihilation model.

2.3. Horton pruning

Here we give definition of Horton pruning for rooted trees [34, 131, 173, 96]; the related Horton-Strahler orders are defined in the next section. Horton pruning for a binary tree is illustrated in Fig. 9.

Definition 2 (**Series reduction**). *The operation of series reduction on a rooted tree (with or without edge lengths, plane or not) removes each degree-two non-root vertex by merging its adjacent edges into one (Fig. 8). For trees with edge lengths it adds the lengths of the two merging edges. The series reduction does not affect the left/right orientation in the planar trees.*

Thus, the series reduction is a mapping from the space of rooted trees (with or without edge lengths, plane or not) to the corresponding space of reduced rooted trees, which can be either $\mathcal{T}, \mathcal{T}_{\mathrm{plane}}, \mathcal{L},$ or $\mathcal{L}_{\mathrm{plane}}$.

Definition 3 (**Horton pruning**). *Horton pruning $\mathcal{R}$ on either of the spaces $\mathcal{T}, \mathcal{T}_{\mathrm{plane}}, \mathcal{L},$ or $\mathcal{L}_{\mathrm{plane}}$ is an onto function 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 $\mathcal{R}(\phi) = \phi$.*

Horton pruning induces a map on the underlying space of trees (Fig. 9). The trajectory of each tree T under $\mathcal{R}(\cdot)$ is uniquely determined and finite:

$$T \equiv \mathcal{R}^0(T) \to \mathcal{R}^1(T) \to \cdots \to \mathcal{R}^k(T) = \phi, \tag{3}$$

with the empty tree ϕ as the (only) fixed point. The preimage $\mathcal{R}^{-1}(T)$ of any non-empty tree T consists of an infinite collection of trees.

We observe that the space $T^|$ of planted trees is invariant under the operation of Horton pruning in Def. 3, while the space $T^\vee$ of stemless trees is not. The Horton analysis of stemless trees is simplified if one considers a *phantom edge* parental to the root. This phantom edge makes pruning in the space of stemless trees similar to pruning in the space of planted trees. In particular, this explains the difference in the treatment of planted and stemless trees in the definition of Horton-Strahler orders (Def. 4) that will be given in Sect. 2.4.

2.4. Horton-Strahler orders

It is natural to think of the distance to ϕ under the Horton pruning map and introduce the respective notion of tree order (see Fig. 9).

Definition 4 (**Horton-Strahler order**). *The Horton-Strahler order* $\mathsf{ord}(T) \in \mathbb{Z}_+$ *of a planted tree* T *in* $\mathcal{T}^|$, $\mathcal{T}^|_{\mathrm{plane}}$, $\mathcal{L}^|$, *or* $\mathcal{L}^|_{\mathrm{plane}}$ *is defined as the minimal number of Horton prunings necessary to eliminate the tree:*

$$\mathsf{ord}(T) = \min\left\{k \geq 0 \,:\, \mathcal{R}^k(T) = \phi\right\}.$$

The Horton-Strahler order for a stemless tree T *in* $\mathcal{T}^\vee \backslash \{\phi\}$, $\mathcal{T}^\vee_{\mathrm{plane}} \backslash \{\phi\}$, $\mathcal{L}^\vee \backslash \{\phi\}$ *or* $\mathcal{L}^\vee_{\mathrm{plane}} \backslash \{\phi\}$ *is defined similarly as follows:*

$$\mathsf{ord}(T) = 1 + \min\left\{k \geq 0 \,:\, \mathcal{R}^k(T) = \phi\right\}.$$

In particular, the order of the empty tree is $\mathsf{ord}(\phi) = 0$, because $\mathcal{R}^0(\phi) = \phi$. The distinction between the planted and the stemless trees in this definition is due to an extra Horton pruning needed for eliminating the stem of a planted tree (Fig. 6). Most of our discussion will be focused on non-empty planted trees with orders $\mathsf{ord}(T) > 0$. In addition, we will often consider measures on tree spaces that assign probability zero to the empty tree ϕ. Figures 1, 2, 11, 12 show examples of Horton-Strahler ordering in binary trees.

Horton pruning partitions the underlying space of planted trees (e.g. $\mathcal{T}^|$, $\mathcal{T}^|_{\mathrm{plane}}$, $\mathcal{L}^|$, $\mathcal{L}^|_{\mathrm{plane}}$, $\mathcal{BT}^|$, or $\mathcal{BT}^|_{\mathrm{plane}}$) into exhaustive and mutually exclusive collection of subspaces $\mathcal{H}_K$ of trees of Horton-Strahler 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 singe edge that connects the tree root and the only leaf descendant to the root, and all other subspaces $\mathcal{H}_K$, $K \geq 2$, consist of an infinite number of trees. In particular, the tree size in these subspaces is unbounded from above: for any $M > 0$ and any $K \geq 2$, there exists a tree $T \in \mathcal{H}_K$ such that $\#T > M$. At the same time, the definition of Horton-Strahler orders implies that for any $K \geq 2$ and $T \in \mathcal{H}_K$, the tree size $\#T \geq 2^{K-1}$. Specifically, the minimal size of a planted tree of order $K \geq 2$ is $2^{K-1} + 1$, and the minimal size of a stemless tree of order $K \geq 2$ is 2^{K-1}. This minimum is achieved for a perfect binary tree (see Ex. 1 in Sect. 3.4).
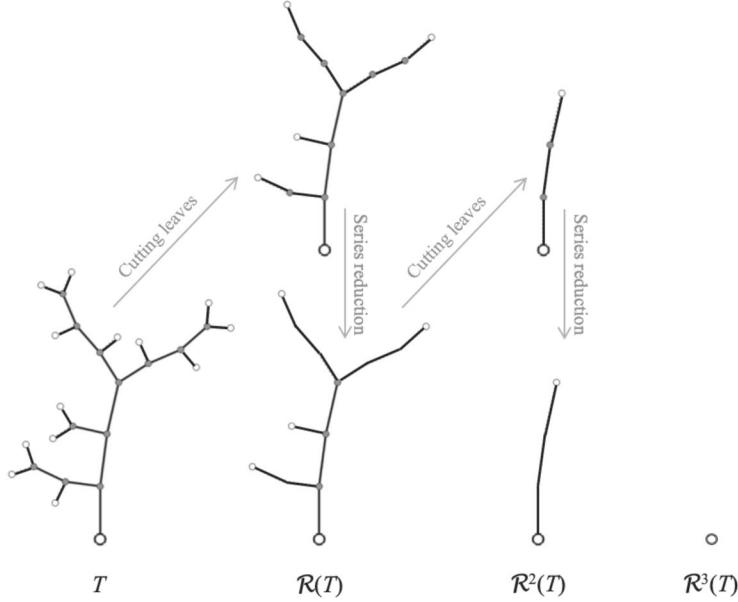
FIG 9. *Example of Horton pruning and Horton-Strahler ordering for a tree $T \in \mathcal{BL}^{|}_{\text{plane}}$. The figure shows the two stages of Horton pruning operation – cutting the leaves (top row), and consecutive series reduction (bottom row). The initial tree T is shown in the leftmost position of the bottom row. The order of the tree is $\mathsf{ord}(T) = 3$, since it is eliminated in three Horton prunings, $\mathcal{R}^3(T) = \phi$. Tree leaves are shown by small open circles, internal vertices – by filled circles, the tree root – by large open circle.*

Definition 5 (**Horton-Strahler terminology**)**.** *We introduce the following definitions related to the Horton-Strahler order of a tree (see Fig. 10):*

1. (**Descendant subtree at a vertex**) *For any non-root vertex v in $T \neq \phi$, a descendant subtree $T_v \subset T$ is the only planted subtree in T rooted at the parental vertex $\mathsf{parent}(v)$ of v, and comprised by v and all its descendant vertices together with their parental edges.*
2. (**Vertex order**) *For any vertex $v \in T \setminus \{\rho\}$ we set $\mathsf{ord}(v) = \mathsf{ord}(T_v)$ (Fig. 10a). We also set $\mathsf{ord}(\rho) = \mathsf{ord}(T)$.*
3. (**Edge order**) *The parental edge of a non-root vertex has the same order as the vertex.*
4. (**Branch**) *A maximal connected component consisting of vertices and edges of the same order is called a branch (Fig. 10a). Note that a tree T always has a single branch of the maximal order $\mathsf{ord}(T)$. In a stemless tree, the maximal order branch may consist of a single root vertex.*
5. (**Initial and terminal vertex of a branch**) *The branch vertex closest to the root is called the initial vertex of the branch. The branch vertex farthest from the root is called the terminal vertex of a branch. See Fig. 10a.*
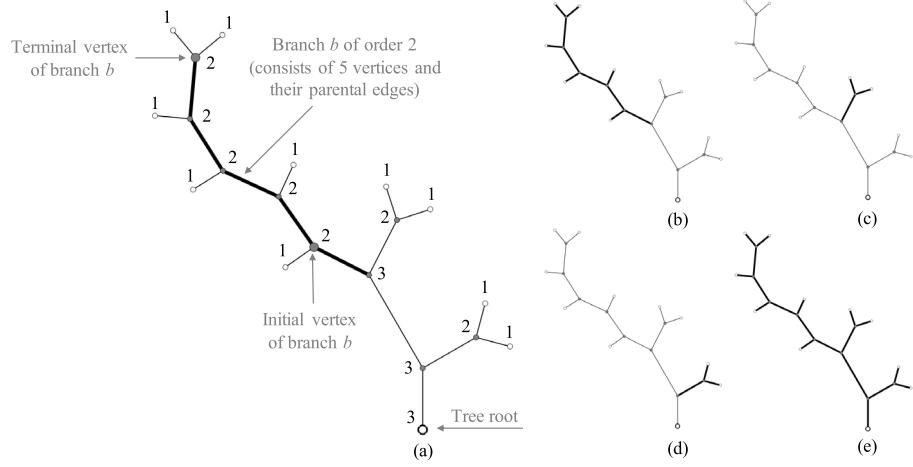
FIG 10. *Illustration of the Horton-Strahler terminology (Def. 5) in a tree $T \in \mathcal{BL}^|_{\mathrm{plane}}$ of order* $\mathsf{ord}(T) = 3$. *(a) Tree root, branch, initial and terminal vertex of a branch. The numbers indicate the Horton-Strahler orders of the vertices and their parental edges. The panel illustrates a branch of order* 2, *shown in bold. Here* $N_1 = 10$, $N_2 = 7$, $N_3 = 3$, $N_{1,2} = 4$, $N_{1,3} = 0$, *and* $N_{2,3} = 1$ *(see Sect. 3.3). (b), (c), (d) Complete subtrees of order* 2. *(e) Complete subtree of order* 3 *(coincides with the tree* T*).*

6. **(Complete subtree of a given order)** *Consider a connected component of tree T that has been completely removed in K pruning operations (but has not been completely removed in $K-1$ prunings). Each such component is a "planted tree without the root"; in other words it always starts with an open edge. This connected component together with the vertex that connects it to the rest of the tree is a subtree of T that will be called a complete subtree of order K.*

We observe that each subtree T_v at the initial vertex v of a branch of order $K \leq \mathsf{ord}(T)$ is a complete subtree of order K, and vice versa (Fig. 10b–d). A complete subtree of order $\mathsf{ord}(T)$ coincides with T (Fig. 10e). All subtrees of order $\mathsf{ord} = 1$ are complete (and consist of a single leaf, its parental edge, and the leaf's parent).

2.5. *Alternative definitions of Horton-Strahler orders*

Definition 4 connects the Horton-Strahler orders to the Horton pruning operation, which is the main theme of this survey. Here we give two alternative, equivalent, definitions of the Horton-Strahler orders. The proof of equivalence is straightforward and is left as an exercise.

Historically, the concept of tree order has been developed by Horton [85] and Strahler [151] for analysis of river streams; it uses hierarchical counting [43, 131, 126]. The first such definition for non-binary trees appeared in [34]. In
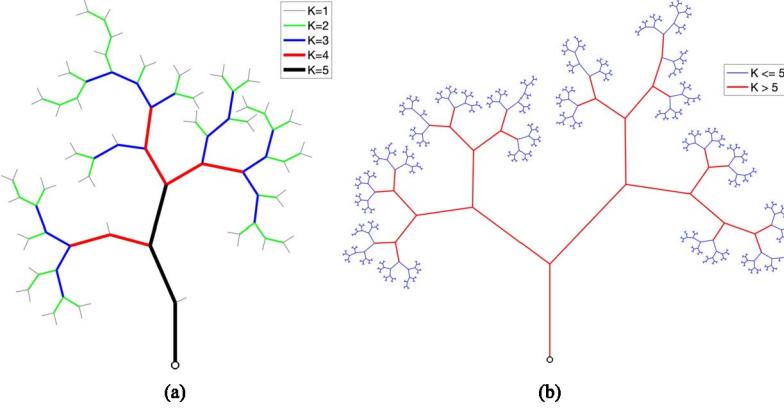
FIG 11. *Example of Horton-Strahler ordering of a binary tree $T \in \mathcal{BL}^|$. Different colors correspond to different orders of vertices and their parental edges, as indicated in legend. (a) $\#T = 121$, $\mathsf{ord}(T) = 5$. (b) $\#T = 1233$, $\mathsf{ord}(T) = 10$.*

this approach, each leaf is assigned order 1. If an internal vertex p has $m \geq 1$ offspring with orders $i_1, i_2, \ldots, i_m$ and $r = \max\{i_1, i_2, \ldots, i_m\}$, then

$$\mathsf{ord}(p) = \begin{cases} r & \text{if } \#\{s: i_s = r\} = 1, \\ r+1 & \text{otherwise.} \end{cases} \tag{4}$$

The parental edge of a non-root vertex has the same order as the vertex. The Horton-Strahler order of a tree $T \neq \phi$ is $\mathsf{ord}(T) = \max_{v \in T} \mathsf{ord}(v)$, where the maximum is taken over all vertices in T. This definition is most convenient for practical calculations, which explains its popularity in the literature.

In a special case of a reduced binary tree, an internal vertex p with two offspring of orders i and j has order

$$\mathsf{ord}(p) = \max(i, j) + \delta_{ij} = \lfloor \log_2(2^i + 2^j) \rfloor, \tag{5}$$

where δ_{ij} is the Kronecker's delta and $\lfloor x \rfloor$ denotes the maximal integer less than or equal to x. In words, the order increases by unity every time when two edges of the same order meet at a vertex (Figs. 1, 2, 11, 12).

Finally, we observe that $\mathsf{ord}(T)$ of a planted tree T equals the depth of the maximal planted perfect binary subtree of T with the same root (see Sect. 3.4, Ex. 1).

2.6. Tokunaga indices and side branching

The Tokunaga indices complement the Horton-Strahler orders (Sects. 2.4, 2.5) by cataloging the mergers of branches according to their orders. In this work, we define and use the Tokunaga indices in binary trees. It is straightforward to adopt these definitions for trees with general branching (see [34, 101]).
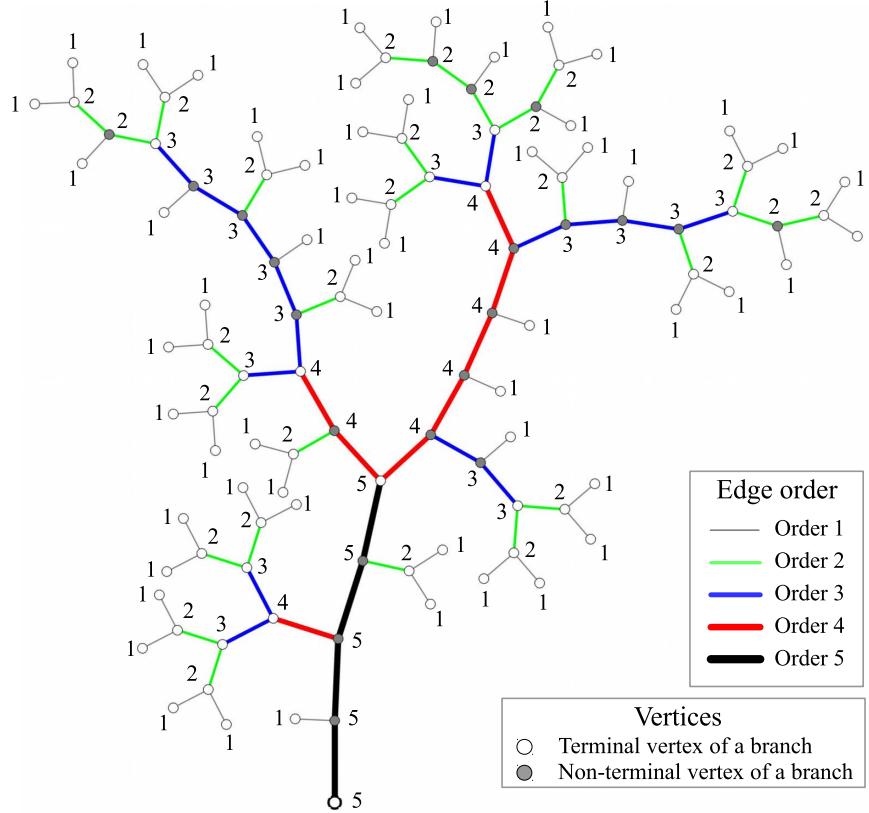
FIG 12. *Horton-Strahler orders of vertices in a binary tree: example. The order is shown next to every vertex. Edge orders are indicated by colors (see legend). Open circles mark terminal vertices of tree branches; they correspond either to leaves or mergers of principal branches. Shaded circles mark vertices that correspond to side branches. Here $N_1 = 56$, $N_2 = 22$, $N_3 = 8$, $N_4 = 3$, and $N_5 = 1$. Figure 13 shows the Tokunaga indexing for the same tree.*

Recall that a branch (Def. 5) is an uninterrupted sequence of vertices and edges of the same order (Fig. 10a). According to the Horton-Strahler ordering rules, every time when two branches of the same order i meet at a vertex, this vertex (and hence the branch for which this is the terminal vertex) is assigned order $i+1$. We refer to this as *principal branching*. A merger of two branches of distinct orders at a vertex, however, does not result in assigning this vertex (and the corresponding branch) a higher order; in this case a higher-order branch absorbs the lower-order branch. This phenomenon is known as *side branching* [126]. A branch of order i that merges with (and is being absorbed by) a branch of a higher order $j > i$ is referred to as a *side branch* of Tokunaga index $\{i, j\}$.

Formally, for a non-root vertex v in a reduced binary tree, we let sibling(v) denote the unique vertex of the tree that has the same parent as v, i.e., parent(v) = parent(sibling(v)).

Definition 6 (**Tokunaga indices**). *In a binary tree $T \in \mathcal{BT}$, consider a branch b of order $i \in \{1, \ldots, \mathsf{ord}(T) - 1\}$, and let v denote the initial vertex of the branch b, whence $\mathsf{ord}(v) = i$. The branch b is assigned the Tokunaga index $\{i, j\}$, where $j = \mathsf{ord}(\mathsf{sibling}(v))$. The Horton-Strahler ordering rules imply that $j \geq i$. A branch with Tokunaga index $\{i, i\}$ is called principal branch. A branch with Tokunaga index $\{i, j\}$ such that $i < j$ is called side branch.*

The definition of Tokunaga indices is illustrated in Fig. 13.

Remark 1. We emphasize that the Tokunaga indices refer to the tree branches, not to individual vertices and edges as is the case with the Horton-Strahler orders.

2.7. Labeling edges

The edges of a planar tree can be labeled by numbers $1, \ldots, \#T$ in order of depth-first search. 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 Horton pruning $\mathcal{R}$, as we describe in the following definition.

Definition 7 (**Proper embedding**). *An embedding function* EMBED $: \mathcal{T} \to \mathcal{T}_{\mathrm{plane}}$ *(respectively* $\mathcal{L} \to \mathcal{L}_{\mathrm{plane}}$*) is called proper if for any $T \in \mathcal{T}$ (resp. $T \in \mathcal{L}$),*

$$\mathcal{R}\left(\mathrm{EMBED}(T)\right) = \mathrm{EMBED}\left(\mathcal{R}(T)\right),$$

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

An example of proper embedding is given in [99].

2.8. Galton-Watson trees

The Galton-Watson distributions (aka Bienaymé-Galton-Watson distributions) over $\mathcal{T}^|$ are pivotal in the theory of random trees. Recall that a random Galton-Watson tree describes trajectory of a Galton-Watson branching process. This population process (and the corresponding tree) is constructed recursively as follows. The population starts with a single progenitor, represented by the tree root. For a given probability mass function $\{q_k\}$ over $\mathbb{Z}_+$, the population develops in discrete steps. At step $d = 0$, the progenitor produces a single offspring. It is represented in the tree by the vertex connected to the tree root by stem. At every discrete step $d > 0$ each existing population member (represented by a leaf of the currently constructed tree at the maximal depth $d - 1$) gives birth to $k \geq 0$ offsprings with probability q_k, where $k = 0$ represents the case when the population member leaves no offsprings, and terminates.

Assuming $q_1 < 1$, the resulting tree is finite with probability one if and only if $\sum_{k=0}^{\infty} k q_k \leq 1$ [81, 12]. Assuming in addition that $q_1 = 0$ in order to generate
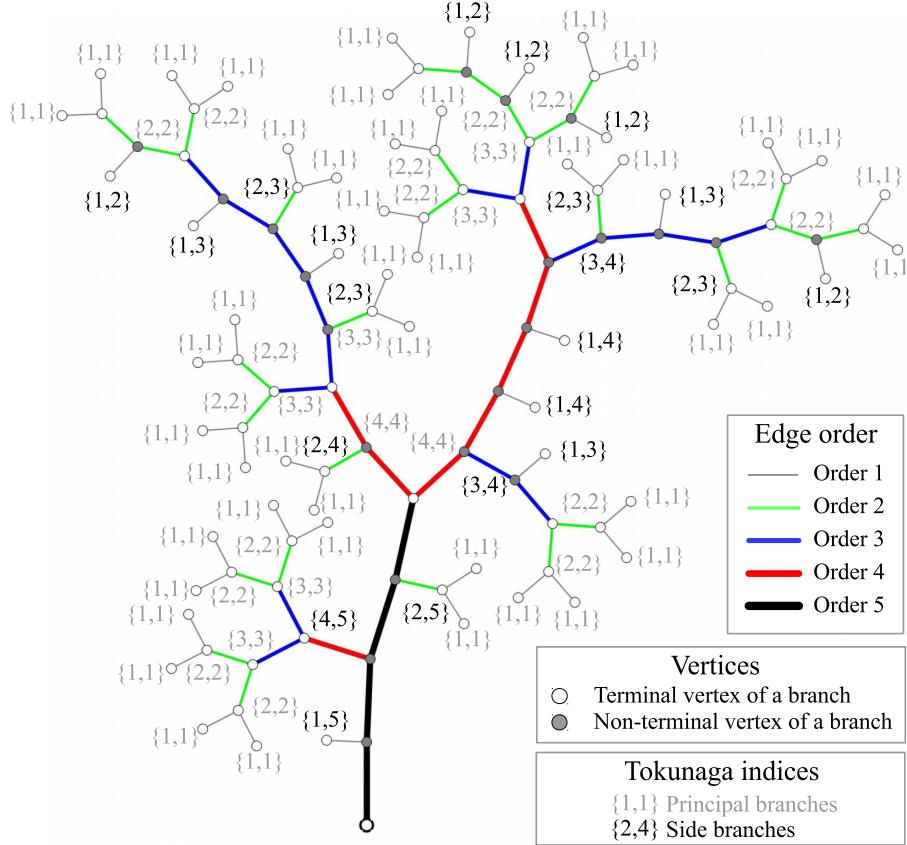
FIG 13. *Tokunaga indices for tree branches in a binary tree: example. The Tokunaga indices $\{i,j\}$ catalog mergers of tree branches, according to their Horton-Strahler orders. Edge orders are indicated by colors (see legend). Open circles mark terminal vertices of tree branches; they correspond either to leaves or mergers of principal branches. Shaded circles mark vertices that correspond to side branches. Here $N_{1,2} = 5$, $N_{1,3} = 4$, $N_{1,4} = 2$, $N_{1,5} = 1$, $N_{2,3} = 4$, $N_{2,4} = 1$, $N_{2,5} = 1$, $N_{3,4} = 2$, $N_{3,5} = 0$, and $N_{4,5} = 1$. Figure 12 shows the Horton-Strahler orders in the same tree.*

reduced trees, the Galton-Watson process induces a tree distribution on $\mathcal{T}^{|}$, denoted here by $\mathcal{GW}(\{q_k\})$. Finally, it is well known that in the critical case (i.e., for $\sum_{k=0}^{\infty} kq_k = 1$) the time to extinction (and hence the tree size) has infinite first moment.

We write $\mathcal{GW}(q_0, q_2)$ for the probability distribution of (combinatorial) binary Galton-Watson trees in $\mathcal{BT}^{|}$. The critical case (unit expected progeny) corresponds to $q_0 = q_2 = 1/2$. Finally, we let $\mathcal{GW}_{\text{plane}}(q_0, q_2)$ denote the probability distribution of (combinatorial) plane binary Galton-Watson trees in $\mathcal{BT}^{|}_{\text{plane}}$. A random tree sampled from $\mathcal{BT}^{|}_{\text{plane}}$ with distribution $\mathcal{GW}_{\text{plane}}(q_0, q_2)$ is obtained

from a random tree sampled from $\mathcal{BT}^{|}$ with distribution $\mathcal{GW}(q_0, q_2)$ by assigning the left-right orientation to each pair of offspring uniformly and independently for each parental vertex.

We conclude this section with a particular characterization of the critical binary Galton-Watson distribution $\mathcal{GW}\left(\frac{1}{2}, \frac{1}{2}\right)$; it follows directly from the process definition and will be used later.

Remark 2. A distribution μ on $\mathcal{BT}^{|}$ is $\mathcal{GW}\left(\frac{1}{2}, \frac{1}{2}\right)$ if and only if it can be constructed in the following way. Start with a stem (root edge). With probability $1/2$ this completes the tree generation process. With the complementary probability $1/2$, draw two trees independently from the distribution μ, and attach them (as descendant subtrees) to the non-root vertex of the stem. This completes the construction.

3. Self-similarity with respect to Horton pruning

This section introduces self-similarity for finite combinatorial and metric trees. We start in Sects. 3.1, 3.2 with a distributional self-similarity for measures on the spaces $\mathcal{T}^{|}$ and $\mathcal{L}^{|}$, respectively. A mean self-similarity that only considers the average values of selected branch statistics it discussed in Sect. 3.3; this is done for a narrower class of combinatorial binary trees from $\mathcal{BT}^{|}$. A general treatment of mean self-similarity in $\mathcal{T}^{|}$ is given in [101].

The term *self-similarity* is associated with invariance of a tree distribution with respect to the Horton pruning $\mathcal{R}$ introduced in Sect. 2.3. The prune-invariance alone, however, might not provide sufficient constraints for generating interesting families of trees suitable for applied statistical analysis (see Ex. 9 in Sect. 3.4). This calls for an additional property – *coordination* among trees of different orders introduced in Sect. 3.3.

3.1. Self-similarity of a combinatorial tree

Let $\mathcal{H}_K \subset \mathcal{T}^{|}$ be the subspace of planted trees of Horton-Strahler order $K \geq 0$. Naturally, $\mathcal{H}_K \bigcap \mathcal{H}_{K'} = \emptyset$ if $K \neq K'$, and $\bigcup_{K \geq 0} \mathcal{H}_K = \mathcal{T}^{|}$. Consider a set of conditional probability measures $\{\mu_K\}_{K \geq 0}$ each of which is defined on $\mathcal{H}_K$ by

$$\mu_K(T) = \mu(T \,|\, T \in \mathcal{H}_K) \tag{6}$$

and let $\pi_K = \mu(\mathcal{H}_K)$. Then μ can be represented as a mixture of the conditional measures:

$$\mu = \sum_{K=0}^{\infty} \pi_K \mu_K. \tag{7}$$

Definition 8 (**Horton prune-invariance/Horton self-similarity**). *Consider a probability measure μ on $\mathcal{T}^{|}$ such that $\mu(\phi) = 0$. Let ν be the pushforward measure, $\nu = \mathcal{R}_*(\mu)$, i.e.,*

$$\nu(T) = \mu \circ \mathcal{R}^{-1}(T) = \mu\left(\mathcal{R}^{-1}(T)\right).$$

Measure μ is called invariant with respect to the Horton pruning (Horton prune-invariant), or, equivalently, Horton self-similar, if for any tree $T \in \mathcal{T}^|$ we have

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

Remark 3 (**Prune-invariance vs. self-similarity**)**.** We follow Burd et al. [34] and a solid body of applied literature (see Sect. 4.4) when using the terms *prune-invariance* and *self-similarity* interchangeably. Section 3.3 below discusses a mean form of self-similarity, which is not equivalent to mean prune-invariance and requires additional property of *coordination* among trees of different orders (Def. 10). A distributional version of coordination is introduced in [99]. Informally, the distributional coordination property ensures that all complete subtrees of order K (see Def. 5) within a tree of a larger order H are identically distributed regardless of the value $H \geq K$. The coordination property holds by construction in Galton-Watson trees (Sect. 2.8), and in trees generated by our main model – the Hierarchical Branching Process of Sect. 7. Example 9 in Sect. 3.4 illustrates that coordination facilitates applied data-driven analysis of self-similar trees. It hence may be justified to define distributional Horton self-similarity as a combination of Horton prune-invariance and distributional coordination; this is done in [99].

Remark 4 (**Relation to invariant and quasi-stationary measures**)**.** Recall that the empty tree ϕ is the only fixed point for the map $\mathcal{R}$ from $\mathcal{T}^|$ to itself. Consequently, the above defined notion of Horton prune-invariant measure is a crossbreed between the invariant measures [164] and the quasi-stationary distributions [41]. The pushforward measure ν is induced by the original measure μ via the pruning operation: if $T' \overset{d}{\sim} \mu$ then $T = \mathcal{R}(T') \overset{d}{\sim} \nu$. Therefore, the probability $\nu(\phi) = \mu(\mathcal{H}_1)$ can be positive.

Proposition 1 (**Geometric order distribution**)**.** *Let μ be a Horton prune-invariant measure on $\mathcal{T}^|$. Then the distribution of orders, $\pi_K = \mu(\mathcal{H}_K)$, is geometric:*

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

where $p = \pi_1 = \mu(\mathcal{H}_1)$, and for any $T \in \mathcal{H}_K$

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

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

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

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

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

Combining (11) and (12) we find for any $K \geq 1$

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

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

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

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

Together with (12) this implies (10). □

Proposition 1 shows that a Horton prune-invariant measure μ is completely specified by its conditional measures μ_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 [34, Thm. 3.5].

3.2. Self-similarity of a tree with edge lengths

Consider a tree $T \in \mathcal{L}^|$ with edge lengths given by a positive vector $l_T = (l_1, \ldots, l_{\#T})$ and let $\text{LENGTH}(T) = \sum_i l_i$. We assume that the edges are labeled in a proper way as described in Sect. 2.7. 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 $x_T = (x_1, \ldots, x_{\#T})$ on the simplex $\sum_i x_i = 1$, $0 < x_i \le 1$, and conditional distribution $F(\cdot | x_T)$ of the tree length $\text{LENGTH}(T)$, where

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

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

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

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

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

The definition of self-similarity for a tree with edge lengths builds on its analog for combinatorial trees in Sect. 3.1. The combinatorial notion Horton prune-invariance (Def. 8), which we refer to as prune-invariance *in shapes*, is complemented with an analogous property *in edge lengths*. We consider the distribution of edge lengths after pruning. It is specified by distribution of the relative edge lengths after pruning

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

and distribution the total tree length after pruning

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

Definition 9 (Horton self-similarity of a tree with edge lengths). *We call a measure η on $\mathcal{L}^|$ self-similar with respect to Horton pruning $\mathcal{R}$ if the following conditions hold:*

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

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

(ii) *The measure is Horton 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

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

3.3. Mean self-similarity of a combinatorial tree

The distributional prune-invariance of previous sections (Defs. 8, 9) may be too restrictive and/or hard to verify in observations. This section introduces a weaker form of prune-invariance that only operates with the mean values of branch and side-branch counts, which makes it better suited for applied analyses.

We consider here the space $\mathcal{BT}^|$ of combinatorial planted binary trees; a generalization to $\mathcal{T}^|$ is given in [101]. Let $N_k = N_k[T]$ be the number of branches of order k in a tree T, and $N_{i,j} = N_{i,j}[T]$ be the number of side branches with Tokunaga index $\{i,j\}$ with $1 \leq i < j \leq \mathrm{ord}(T)$ in a tree T, i.e., the number of instances when an order-i branch merges with an order-j branch. Examples of counts $N_i[T]$ and $N_{i,j}[T]$ are given in Figs. 10, 12, 13. We do not consider the numbers $N_{i,i}[T]$ of principal branches in T – they are redundant with respect to the branch counts because of the relation $N_{i,i}[T] = 2N_{i+1}[T]$.

We write $\mathsf{E}_K[\cdot]$ for the mathematical expectation with respect to μ_K of Eq. (6). As before, we adopt the notation $\mathcal{H}_K$ for the subspace of trees of order K in $\mathcal{BT}^|$.

We define the *average Horton numbers* as

$$\mathcal{N}_k[K] = \mathsf{E}_K[N_k], \quad 1 \leq k \leq K, \quad K \geq 1,$$

and the *average side branch numbers* of index $\{i,j\}$ as

$$\mathcal{N}_{i,j}[K] := \mathsf{E}_K[N_{i,j}], \quad 1 \leq i < j \leq K, \quad K \geq 2.$$

We assume below that the average branch and side branch numbers are finite for any $K \geq 1$:

$$\mathcal{N}_j[K] < \infty \text{ and } \mathcal{N}_{i,j}[K] < \infty \text{ for all } 1 \leq i < j \leq K.$$

The *Tokunaga coefficient* $T_{i,j}[K]$ for subspace $\mathcal{H}_K$ is defined as the ratio of the average side branch number of index $\{i,j\}$ to the average Horton number of order j:

$$T_{i,j}[K] = \frac{\mathcal{N}_{i,j}[K]}{\mathcal{N}_j[K]}, \quad 1 \leq i < j \leq K. \tag{14}$$

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

Remark 5. Suppose that measure μ is such that all complete branches of order j within a random tree $T \in \mathcal{H}_K$ sampled with μ_K have the same distribution, independent of $N_j[T]$. (For instance, this is the case for a Galton-Watson tree.) In particular, the numbers $N_i[b_k]$ of branches of order i that merge with a particular branch b_k, $k = 1, \ldots, N_j[T]$ of order j in T has the same distribution for all b_k. Let $n_{i,j}$ be a random variable such that $N_i[b_k] \stackrel{d}{=} n_{i,j}$. Then we have, by Wald's equation,

$$
\begin{aligned}
\mathcal{N}_{i,j}[K] &= \mathsf{E}_K[N_{i,j}[T]] = \mathsf{E}_K \left[\sum_{k=1}^{N_j[T]} N_i[b_k] \right] \\
&= \mathsf{E}_K[N_j[T]]\mathsf{E}_K[n_{i,j}] = \mathcal{N}_j[K]\mathsf{E}_K[n_{i,j}],
\end{aligned}
$$

and, accordingly,

$$
T_{i,j}[K] = \frac{\mathcal{N}_j[K]\mathsf{E}_K[n_{i,j}]}{\mathcal{N}_j[K]} = \mathsf{E}_K[n_{i,j}].
$$

In other words, the Tokunaga coefficient in this case is the expected number of side branches of appropriate index in a randomly selected branch. This is how the Tokunaga coefficient is often defined (e.g., [34]). The definition (14) adopted here is more general, as it allows branches to depend on each other and on structure (shape, size, etc.) of T.

In applied analysis of data, it is convenient to assume that the values of Tokunaga coefficients $T_{i,j}[K]$ are independent of the tree order K. This motivates the following *coordination* property.

Definition 10 (Coordination). *A set of probability measures $\{\mu_K\}_{K \geq 1}$ on $\{\mathcal{H}_K\}_{K \geq 1}$ is called coordinated if the respective Tokunaga coefficients are independent of the tree order:*

$$
T_{i,j}[K] = T_{i,j}[K'] =: T_{i,j} \tag{15}
$$

for all $K, K' \geq 2$ and $1 \leq i < j \leq \min\{K, K'\}$. A measure μ on $\mathcal{BT}^|$ is called coordinated if the respective conditional measures $\{\mu_K\}$, as in Eq. (7), are coordinated.

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

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

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

Definition 11 (**Toeplitz property**). *A set of probability measures $\{\mu_K\}_{K \geq 1}$ on $\{\mathcal{H}_K\}_{K \geq 1}$ is said to satisfy the Toeplitz property if for every $K \geq 2$ there exists a sequence $T_k[K] \geq 0$, $k = 1, 2, \ldots$ such that*

$$T_{i,j}[K] = T_{j-i}[K] \quad \text{for each } K \geq 2. \tag{16}$$

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 μ on $\mathcal{BT}^|$ is said to satisfy the Toeplitz property if the respective conditional measures $\{\mu_K\}$, as in Eq. (7), satisfy the Toeplitz property.

Definition 12 (**Mean self-similarity**). *A set of probability measures $\{\mu_K\}_{K \geq 1}$ on $\{\mathcal{H}_K\}_{K \geq 1}$ is called mean self-similar if it is coordinated and satisfies the Toeplitz property. A measure μ on $\mathcal{BT}^|$ is called mean self-similar if the respective conditional measures $\{\mu_K\}$, as in Eq. (7), are mean self-similar.*

An alternative (yet equivalent) Def. 14 stated below will clarify the name selected for this property.

Combining Eqs. (15) and (16) we find that for a mean self-similar measure there exists a nonnegative *Tokunaga sequence* $\{T_k\}_{k=1,2,\ldots}$ such that

$$T_{i,j}[K] = T_{j-i} \quad \text{for all } 1 \leq i < j \leq K, \tag{17}$$

and the corresponding Tokunaga matrices $\mathbb{T}_K$ are Toeplitz:

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

Remark 6 (**Applied analysis of trees**). The Tokunaga matrix $\mathbb{T}_K$ is easily calculated for an observed collection of trees; this is the common way of examining self-similairty in applications [131, 126, 34, 178]. The Toeplitz property for a collection of branch and side-branch counts $N_i[T]$, $N_{i,j}[T]$ can be tested within the ANOVA framework.

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

$$N_k[T] = N_{k-1}[\mathcal{R}(T)], \quad k \geq 2, \tag{18}$$

$$N_{i,j}[T] = N_{i-1,j-1}[\mathcal{R}(T)], \quad 2 \leq i < j. \tag{19}$$

Consider the pushforward probability measure $\mathcal{R}_*(\mu)$ induced on $\mathcal{H}_K$ by the pruning operator:

$$\mathcal{R}_*(\mu)(A) = \mu_{K+1}\left(\mathcal{R}^{-1}(A)\right) \quad \forall A \subset \mathcal{H}_K.$$

The Tokunaga coefficients computed on $\mathcal{H}_K$ using the pushforward measure $\mathcal{R}_*(\mu)$ are denoted by $T_{i,j}^{\mathcal{R}}[K]$:

$$T_{i,j}^{\mathcal{R}}[K] = T_{i+1,j+1}[K+1] = \frac{\mathcal{N}_{i+1,j+1}[K+1]}{\mathcal{N}_{j+1}[K+1]}. \tag{20}$$

Definition 13 (**Mean prune-invariance**). *A set of probability measures* $\{\mu_K\}_{K \geq 1}$ *on* $\{\mathcal{H}_K\}_{K \geq 1}$ *is called mean prune-invariant if*

$$T_{i,j}[K] = T_{i,j}^{\mathcal{R}}[K] = T_{i+1,j+1}[K+1] \tag{21}$$

for any $K \geq 2$ *and all* $1 \leq i < j \leq K$. *A measure* μ *on* $\mathcal{BT}^{|}$ *is called mean prune-invariant if the respective conditional measures* $\{\mu_K\}$, *as in Eq.* (7), *are mean prune-invariant.*

Definition 14 (**Mean self-similarity**). *A set of probability measures* $\{\mu_K\}_{K \geq 1}$ *on* $\{\mathcal{H}_K\}_{K \geq 1}$ *is called mean self-similar with respect to Horton pruning, or mean self-similar, if it is coordinated and mean prune-invariant. A measure* μ *on* $\mathcal{BT}^{|}$ *is called mean self-similar with respect to Horton pruning if the respective conditional measures* $\{\mu_K\}$, *as in Eq.* (7), *are mean self-similar with respect to Horton pruning.*

Proposition 2. *Definitions* 12 *and* 14 *of mean self-similarity are equivalent.*

This equivalence was proven in [96]. Its validity is readily seen from the diagram of Fig. 14a, 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 (Def. 10), Toeplitz property (Def. 11), and mean prune-invariance (Def. 13). Moreover, we observe that if any two of these properties hold, the third also holds. The Venn diagram of Fig. 14b illustrates the relation among coordination, mean prune-invariance, Toeplitz property and mean self-similarity in the binary tree space $\mathcal{BT}^{|}$.

Consider a mean self-similar measure μ. Observe that since exactly two branches of order k are required to form a branch of order $(k+1)$, the average number of side branches of order $1 \leq k < K$ within $\mathcal{H}_K$ is $\mathcal{N}_k[K] - 2\mathcal{N}_{k+1}[K]$. This number can also be computed by counting the average number of side branches of order k for all higher-order branches:

$$\sum_{j=k+1}^{K} T_{k,j}\,\mathcal{N}_j[K] = \sum_{m=1}^{K-k} T_m\,\mathcal{N}_{k+m}[K].$$

Equalizing these two expressions we arrive at the main system of counting equations:

$$\mathcal{N}_k[K] = 2\,\mathcal{N}_{k+1}[K] + \sum_{j=1}^{K-k} T_j\,\mathcal{N}_{k+j}[K], \quad 1 \leq k \leq K-1, \quad K \geq 2. \tag{22}$$
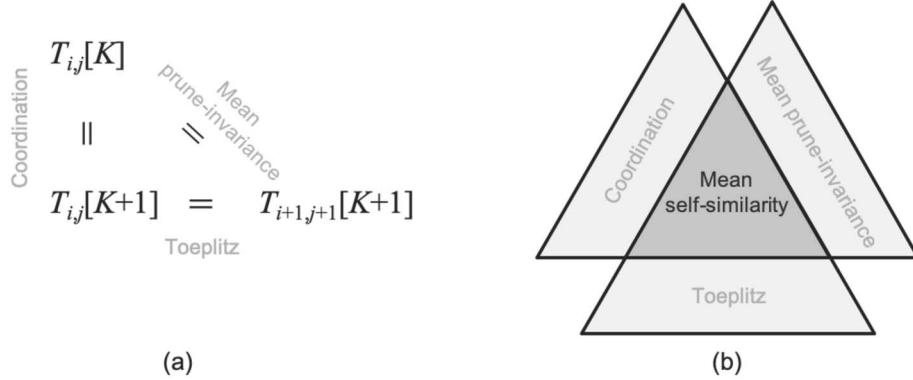
Fig 14. *Relations among coordination, mean prune-invariance, and Toeplitz property. (a) Pairwise equalities among the quantities $T_{i,j}[K]$, $T_{i,j}[K+1]$, and $T_{i+1,j+1}[K+1]$ involved in the definitions of coordination, mean prune-invariance, and Toeplitz property. (b) Venn diagram of the space $\mathcal{BT}^{|}$ illustrating the relation among coordination (left triangle), mean prune-invariance (right triangle), and Toeplitz property (bottom triangle). The mean self-similarity (the inner dark triangle) is formed by the intersection of the three properties (and, hence, by intersection of any two).*

Consider a $K \times K$ linear operator

$$
\mathbb{G}_K := \begin{bmatrix}
-1 & T_1 + 2 & T_2 & \dots & T_{K-1} \\
0 & -1 & T_1 + 2 & \dots & T_{K-2} \\
0 & 0 & \ddots & \ddots & \vdots \\
\vdots & \vdots & \ddots & -1 & T_1 + 2 \\
0 & 0 & 0 & 0 & -1
\end{bmatrix}. \tag{23}
$$

The counting equations (22) rewrite as

$$
\mathbb{G}_K \begin{pmatrix} \mathcal{N}_1[K] \\ \mathcal{N}_2[K] \\ \vdots \\ \mathcal{N}_K[K] \end{pmatrix} = -\mathbf{e}_K, \quad K \geq 1, \tag{24}
$$

where $\mathbf{e}_K$ is the K-th coordinate basis vector. Using this equation for $(K+1)$ and considering the last K components we obtain

$$
\mathbb{G}_K \begin{pmatrix} \mathcal{N}_2[K+1] \\ \mathcal{N}_3[K+1] \\ \vdots \\ \mathcal{N}_{K+1}[K+1] \end{pmatrix} = -\mathbf{e}_K, \quad K \geq 1.
$$

This proves the following statement.

Proposition 3. *Consider a mean Horton self-similar measure μ on $\mathcal{BT}^|$. Then for any $K \geq 1$ and $1 \leq k \leq K$ we have*

$$\mathcal{N}_{k+1}[K+1] = \mathcal{N}_k[K]$$

and

$$\mathcal{N}_{i+1,j+1}[K+1] = \mathcal{N}_{i,j}[K], \quad 1 \leq i < j \leq K, \quad K \geq 2.$$

Definition 15 (**Tokunaga self-similarity**)**.** *A mean self-similar measure μ on $\mathcal{BT}^|$ is called Tokunaga self-similar with parameters (a, c) if its Tokunaga sequence $\{T_j\}_{j=1,2,\dots}$ is expressed as*

$$T_j = a\,c^{j-1}, \quad k \geq 1 \tag{25}$$

for some constants $a \geq 0$ and $c > 0$.

Tokunaga self-similarity (25) specifies a combinatorial tree shape (up to a permutation of side branch attachment within a given branch) with only two parameters (a, c), hence suggesting a conventional modeling paradigm. The empirical validity of the Tokunaga self-similarity constraints (25) has been confirmed for a variety of river networks at different geographic locations [131, 153, 45, 112, 178], as well as in other types of data represented by trees, including botanical trees [126], the veins of botanical leaves [159, 133], clusters of dynamically limited aggregation [129, 126], percolation and forest-fire model clusters [175, 168], earthquake aftershock sequences [157, 84, 172], tree representation of symmetric random walks [173] (Sect. 8.6), and hierarchical clustering [69]. The conditions (25), however, lacks a theoretical justification. We make a step towards justifying this condition in Sect. 7.7.2.

Remark 7 (**Mean self-similarity is a property of conditional measures**)**.** The properties introduced in this section – coordination (Def. 10), Toeplitz (Def. 11), mean prune-invariance (Def. 13), and mean self-similarity (Defs. 12, 14) – are completely specified by a set of conditional measures $\{\mu_K\}$, and are independent of the randomization probabilities $\pi_K = \mu(\mathcal{H}_K)$, see Eq. (7).

Remark 8 (**Terminology**)**.** The self-similarity concepts studied in this work refer to a measure μ, or a collection of conditional measures $\{\mu_K\}$, on a suitable space of trees. For the sake of brevity, we sometimes use a common abuse of notations and discuss self-similarity of a random tree $T \overset{d}{\sim} \mu$ (e.g., claiming that a tree T is mean self-similar, etc.). Formally, such statements apply to the respective tree distribution μ.

3.4. Examples of self-similar trees

This section collects some examples (and non-examples) of self-similar trees and related properties.
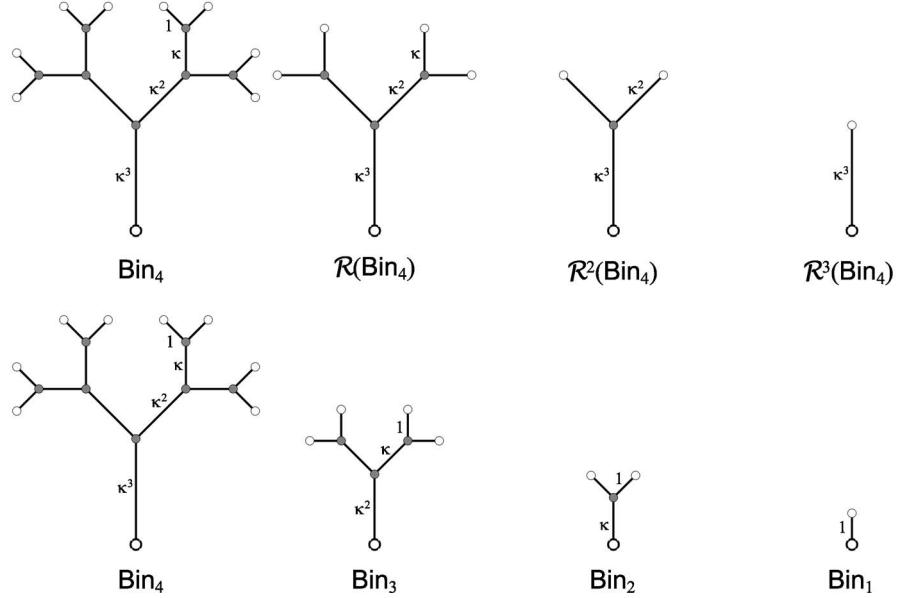
FIG 15. *Self-similarity of perfect binary trees* $\mathsf{Bin}(\kappa) \subset \mathcal{BL}^|$ *(Ex. 1). The length of edges of order i is κ^{i-1} for some $\kappa > 0$. The space $\mathsf{Bin}(\kappa)$ is Horton self-similar with $\zeta = \kappa$ and Tokunaga sequence $T_j = 0$, $k \geq 1$. In this figure, $\kappa = 1.5$. We write Bin_K for the tree of order K. Top row shows three consecutive Horton prunings of Bin_4. Bottom row shows trees $\mathsf{Bin}_{4,3,2,1}$. Here, for any $K \geq 1$ and $m \geq 0$, the tree Bin_K is obtained by scaling all edges of the tree $\mathcal{R}^m(\mathsf{Bin}_{K+m})$ by a multiplicative factor κ^{-m}. The four columns of the figure correspond to $m = 0, 1, 2, 3$ and $K + m = 4$. The lengths of selected edges are indicated in the figure.*

Example 1 (**Perfect binary trees**). Recall that a binary tree is called *perfect* if it is reduced and all its leaves have the same depth (combinatorial distance from the root). Consider space $\mathsf{Bin} \subset \mathcal{BT}^|$ of finite planted perfect binary trees; see Fig. 15. We write $D = D[T]$ for the depth of a tree T and $\mathsf{Bin}_D \subset \mathsf{Bin}$ for the subspace of trees of depth $D \geq 1$. The subspace Bin_D consists of a single tree with 2^{D-1} leaves; it has Horton-Strahler order D. Every conditional measure μ_K in this case is a point measure on Bin_K, $K \geq 1$. Moreover, the order of a vertex at depth $1 \leq d \leq D$ (and its parental edge) is $D - d + 1$, and for the tree Bin_K we have
$$N_k[\mathsf{Bin}_K] = 2^{K-k}, \quad K \geq 1, k \leq K.$$

We write $\mathsf{Bin}(\kappa) \subset \mathcal{BL}^|$ for the space of metric trees with combinatorial shapes from Bin and length κ^{i-1} assigned to edges of order $i \geq 1$. The bottom row of Fig. 15 shows the trees from space $\mathsf{Bin}(\kappa) \subset \mathcal{BL}^|$ with combinatorial shapes Bin_i, $i = 4, 3, 2, 1$, and $\kappa = 1.5$.

Let μ_K, $K \geq 1$, be a point-mass distribution on space $\mathcal{BT}^|$ that assigns probability one to the tree Bin_K. For a given $p \in (0, 1)$ and $\kappa > 0$, consider $\mu = \sum_{k=1}^{\infty} \pi_K \mu_K$ with $\pi_K = p(1-p)^{K-1}$ for all orders $K \geq 1$. Similarly, consider the probability measure η on $\mathcal{BL}^|$ that assigns the geometric probability π_K to

the tree in $\mathsf{Bin}(\kappa)$ with combinatorial shape Bin_K. Then, the measures μ and η satisfy the following properties:

(a) Coordination (Def. 10) and Toeplitz property (Def. 11). By construction, the trees in the space Bin have no side branching ($N_{i,j}[T] = 0$), and therefore,
$$T_{i,j}[K] = T_{j-i}[K] = T_{j-i} = 0, \quad i < j.$$
This implies coordination and Toeplitz property.

(b) Mean self-similarity (Defs. 12, 14) follow from (a).

(c) Combinatorial Horton self-similarity (Def. 8). Recall that subspace Bin_K consists of a single tree for any $K \geq 1$. Since
$$\mathsf{Bin}_K = \mathcal{R}(\mathsf{Bin}_{K+1}), \quad K \geq 1,$$
the space is Horton prune-invariant in shapes (Def. 8 or 9(i)). This establishes combinatorial Horton self-similarity.

(d) Tokunaga self-similarity with $a = 0$ (Def. 15) follows from (c).

(e) Horton prune-invariance in lengths (Def. 9(ii)). By construction, the leaves of a pruned tree have length κ; and the edge lengths change by a multiplicative factor κ with every combinatorial step toward the root. This implies Horton prune-invariance in lengths with $\zeta = \kappa$.

(f) Self-similarity (Def. 9) with $\zeta = \kappa$ follows from (c) and (e). It implies that for any $K \geq 1$ and $m \geq 0$, the tree Bin_K is obtained by scaling all edges of the tree $\mathcal{R}^m(\mathsf{Bin}_{K+m})$ by a multiplicative factor κ^{-m}. The four columns of Fig. 15 correspond to $m = 0, 1, 2, 3$ and $K + m = 4$.

Example 2 (**Combinatorial critical binary Galton-Watson trees**). The Galton-Watson distribution $\mu = \mathcal{GW}(\{q_k\})$ on $\mathcal{T}^|$ ($\sum_{k=0}^{\infty} kq_k \leq 1$ and $q_1 = 0$) has the coordination property for any distribution $\{q_k\}$ with $\pi_1 = \mu(\mathcal{H}_1) \neq 1$. Indeed, the Markovian branching mechanism (see Sect. 2.8) creates subtrees of the same structure, independently of the tree order. This implies coordination. However, mean and distributional prune-invariance (and hence mean and combinatorial Horton self-similarity) only hold in the critical binary case $\mathcal{GW}(\frac{1}{2}, \frac{1}{2})$ [34]. The corresponding Tokunaga sequence is $T_j = 2^{j-1}$, $j \geq 1$, which implies Tokunaga self-similarity with parameters $(a, c) = (1, 2)$. These results are summarized in Thm. 4, Sect. 5.1.

Example 3 (**Critical binary Galton-Watson trees with i.i.d. exponential edge lengths**). The space of critical binary Galton-Watson trees with independent exponential edge lengths is Horton self-similar with $\zeta = 2$; this is shown in Prop. 7 of Sect. 5.2.

Example 4 (**Hierarchical Branching Process**). Section 7 introduces a rich class of measures on $\mathcal{BL}^|$ induced by the Hierarchical Branching Process (HBP). Notably, one can select parameters so that the process is Horton self-similar (Def. 9) with an arbitrary Tokunaga sequence $\{T_j\}$ and arbitrary $\zeta > 0$. This class includes the critical binary Galton-Watson tree with independent exponential lengths as a special case; see Thm. 15.
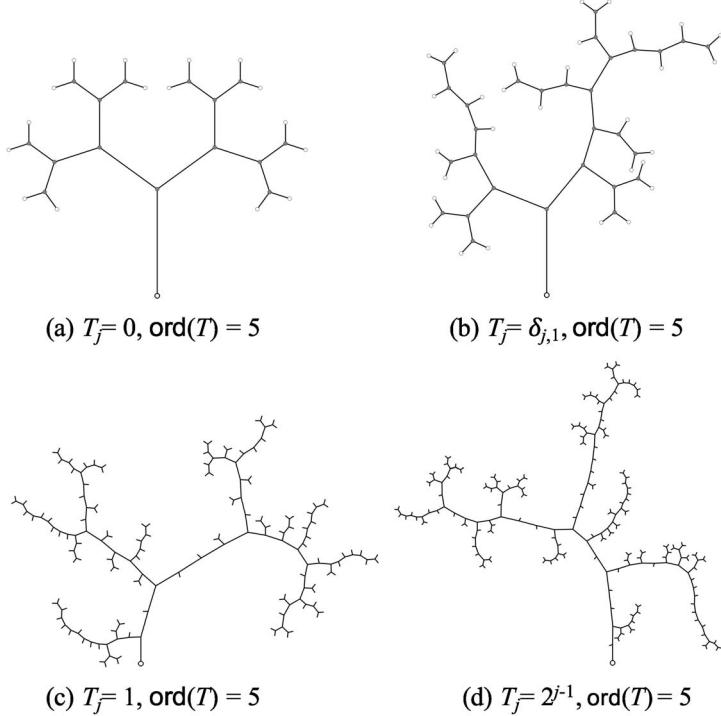
FIG 16. *Random Tokunaga trees: Examples. Each panel shows a random tree sampled with a particular Tokunaga self-similar tree measure (Def. 15), with the corresponding Tokunaga sequence $T_j = ac^{j-1}$. (a) $(a,c) = (0,0), T_j = 0$, (b) $(a,c) = (1,0), T_j = \delta_{j,1}$, (c) $(a,c) = (1,1), T_j = 1$, (d) $(a,c) = (1,2), T_j = 2^{j-1}$. The lengths of edges of order i equal κ^{i-1}, with $\kappa = 1.5$.*

Example 5 (**Combinatorial Tokunaga trees**). Tokunaga self-similar trees (Def. 15) are specified by a particular form of the Tokunaga sequence:

$$T_j = ac^{j-1}, \quad j \geq 1.$$

This is a very flexible model that can account for a variety of dendritic patterns. Figure 16 shows four selected examples:

$$
\begin{aligned}
\text{Fig. 16(a):} \quad & (a,c) = (0,0), \quad T_j = 0, \\
\text{Fig. 16(b):} \quad & (a,c) = (1,0), \quad T_j = \delta_{j,1}, \\
\text{Fig. 16(c):} \quad & (a,c) = (1,1), \quad T_j = 1, \\
\text{Fig. 16(d):} \quad & (a,c) = (1,2), \quad T_j = 2^{j-1}.
\end{aligned}
$$

The random trees in this example are generated using the Geometric Branching Process (see Sect. 7.7.1, Def. 26). This means, in particular, that the number of side branches with Tokunaga index $\{i,j\}$ (Def. 6) is a geometric random variable with mean T_{j-i}. The case $T_j = 0$ corresponds to perfect binary trees

with no side branching (see also Ex. 1). In this case, all branch mergers lead to increase of branch order by unity. This results in a most symmetric deterministic tree structure. Some (random) side branching appears for $T_j = \delta_{j,1}$ (hence $T_1 = 1, T_2 = 0, T_3 = 0, \dots$): every branch of order K has on average a single side branch of order $K-1$, and no side branches of lower orders. This destroys symmetry and introduce randomness in tree shape. The case $T_j = 1$ corresponds to an average of one side branch of any order $1 \leq k \leq K-1$ within a branch of order K, resulting in tentacle-shaped formations of varying length. The most complicated case illustrated here corresponds to $T_j = 2^{j-1}$, which is the Tokunaga sequence for critical binary Galton-Watson trees (but not necessarily vice versa); see Ex. 2. In this case the number of side branches increases geometrically with the difference of branch orders, hence producing branches with widely varying lengths and shapes.

Example 6 (**Tokunaga trees with i.i.d. exponential edge lengths**). Random edge lengths often appear as an element of applied modeling. Figure 17 illustrates the same four Tokunaga models as in Ex. 5, with i.i.d. exponential edge lengths. Clearly, this additional random element substantially affects the tree outlook. The edge length variability becomes a dominant element of the metric tree shape. We notice, in particular, that the four types of trees with exponential edge lengths in Fig. 17 look much more similar that the same four types with deterministic edge lengths related to branch order.

Example 7 (**Critical Tokunaga processes**). Section 7.5 introduces a subclass of HBP, called critical Tokunaga processes, with $T_j = (c-1)c^{j-1}$, $j \geq 1$ for an arbitrary $c \geq 1$. These processes generate tree distributions that are Horton self-similar with $\zeta = c$ and have i.i.d. exponential edge lengths.

Example 8 (**Independent random attachment**). A variety of mean self-similar measures on $\mathcal{T}^|$ can be constructed for an arbitrary sequence of Tokunaga coefficients $\{T_j\}_{j=1,2,\dots}$. Here we give a natural example [96].

Fix a sequence $\{T_j\}_{j=1,2,\dots}$ of Tokunaga coefficients. By Remark 7, it is sufficient to construct a set of Horton self-similar conditional measures μ_K, $K \geq 1$.

The subspace $\mathcal{H}_1$, which consists of a single-leaf tree τ_1, possesses a trivial unity mass conditional measure μ_1. To construct a random tree from $\mathcal{H}_2$, we select a probability mass function $\{P_{1,2}(n)\}_{n=0,1,\dots}$ with the mean value T_1. A random tree $T \in \mathcal{H}_2$ is obtained from the single-leaf tree τ_1 via the following two operations. First, we attach two offspring vertices to the leaf of τ_1. This creates a tree of order 2 with no side branches – one internal vertex of degree 3, two leaves, and the root (Y-shaped tree). Second, we draw a random number ℓ_2 from the probability distribution $P_{1,2}$, and attach ℓ_2 vertices to this tree so that they form side branches of index $\{1, 2\}$.

In general, we use a recursive construction procedure. Assume that a measure μ_{K-1}, $K \geq 2$, is constructed. To construct a random tree $T \in \mathcal{H}_K$ we select a set of probability mass functions $\{P_{k,K}(n)\}_{n=0,1,\dots}$, indexed by $k = 1, \dots, K-1$, 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

(a) $T_j = 0$, ord$(T) = 5$ (b) $T_j = \delta_{j,1}$, ord$(T) = 4$

(c) $T_j = 1$, ord$(T) = 4$ (d) $T_j = 2^{j-1}$, ord$(T) = 4$
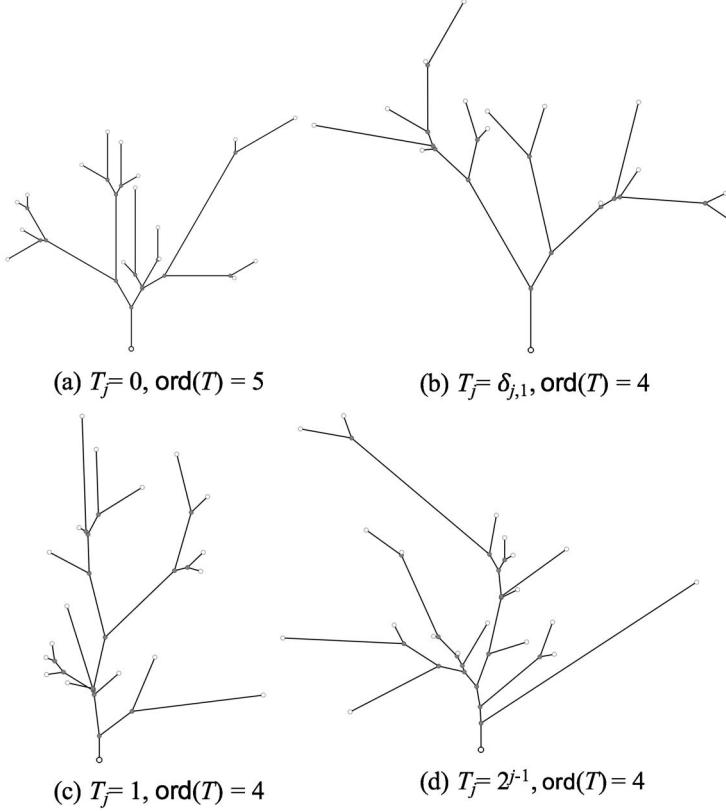
Fig 17. *Random Tokunaga trees with i.i.d. exponential edge lengths: Examples. Each panel shows a random tree drawn from a particular self-similar tree measure (Def. 15) with different Tokunaga sequences $T_j = ac^{j-1}$. (a) $(a,c) = (0,0), T_j = 0$, (b) $(a,c) = (1,0), T_j = \delta_{k,1}$, (c) $(a,c) = (1,1), T_j = 1$, (d) $(a,c) = (1,2), T_j = 2^{j-1}$.*

two new offspring vertices to every leaf of τ hence producing a tree $\tilde{T}$ of order K with no side branches of order 1. Second, for each order $2 \leq j \leq K$ and for all branches b_k of order j in $\tilde{T}$ ($k = 1, \ldots, N_j[\tilde{T}]$), we sample independent random variables $\ell_j(b_k)$ from the distribution $P_{j-1,K}$. Next, for each k, we attach $\ell_j(b_k)$ new leaves to the branch b_k so that they form side branches of index $\{1, j\}$. Each new leaf 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 $\ell_j(b_k)$ vertices among the $m + 1$ edges is given by the equiprobable multinomial distribution with $m + 1$ categories and $\ell_j(b_k)$ trials.

The procedure described above generates a set of coordinated measures $\{\mu_K\}$ on $\{\mathcal{H}_K\}$, since the mean values T_k of the distributions $P_{k,K}$ are unrelated to

K. Furthermore, observe that

$$N_{i,j} = \sum_{k=1}^{N_j[\tilde{T}]} \ell_{j-i+1}(b_k),$$

where each $\ell_{j-i+1}(b_k)$ was sampled independently from $P_{j-i,K-i+1}$, and therefore,

$$
\begin{aligned}
\mathcal{N}_{i,j}[K] &= \mathsf{E}_K[N_{i,j}] = \mathsf{E}_K\left[\mathsf{E}_K[N_{i,j}|N_j]\right] = \mathsf{E}_K[N_j\,T_{j-i}] \qquad (26)\\
&= T_{j-i}\,\mathsf{E}_K[N_j] = T_{j-i}\,\mathcal{N}_j[K].
\end{aligned}
$$

Hence, $T_{i,j}[K] = \mathcal{N}_{i,j}[K]/\mathcal{N}_j[K] = T_{j-i}$, and the sequence of tree measures $\{\mu_K\}$ is mean self-similar (Defs. 12, 14). Finally, in order to make this construction distributionally Horton self-similar (Def. 8), we select an arbitrary $p \in (0,1)$ and define

$$\mu = \sum_{K=1}^\infty \mu_K \pi_K \quad \text{with} \quad \pi_K = p(1-p)^{K-1}.$$

Example 9 (Why coordination?). Relating mean self-similarity (Def. 14) to mean prune-invariance (Def. 13) is quite intuitive (see also [34]). Much less so is the requirement of coordination of conditional measures (Def. 10), 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 (Def. 14) to a branch counting definition (Def. 12). 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 [131, 126, 45, 178]. The absence of coordination, at the same time, allows for a variety of prune-invariant measures with no Toeplitz constraint, which are hardly treatable in applications. To give an example of such a measure, let us select a tree τ_2 from the preimage of the only tree $\tau_1 \in \mathcal{H}_1$ of order $K = 1$ under the pruning operation: $\tau_2 \in \mathcal{R}^{-1}(\tau_1) = \mathcal{H}_2$. In a similar fashion, select any tree τ_{K+1} from the preimage of τ_K for $K \geq 2$. This gives us a collection of trees $\tau_K \in \mathcal{H}_K$, $K \geq 1$ such that $\mathcal{R}(\tau_{K+1}) = \tau_K$. Assign the full measure on $\mathcal{H}_K$ to τ_K: $\mu_K(\tau_K) = 1$. By construction, the measures $\{\mu_K\}$ are mean prune-invariant. They, however, may satisfy neither the coordination nor the Toeplitz property. This example illustrates how one can produce rather obscure collections of mean prune-invariant measures, providing a motivation for the coordination requirement.

4. Horton law in self-similar trees

In this section, we introduce the *strong Horton law* for the numbers of branches of different orders in a combinatorial tree on $\mathcal{T}$ (Def. 16) and for the respective averages (Def. 17). The main result of this section (Thm. 1) shows that the mean self-similarity (Defs. 12, 14) implies the strong Horton law for mean branch numbers (Def. 17).

Consider a measure μ on $\mathcal{T}$ and its conditional measures μ_K, each defined on subspace $\mathcal{H}_K \subset \mathcal{T}$ of trees of Horton-Strahler order $K \geq 1$. We write $T \overset{d}{\sim} \mu_K$ for a random tree T drawn from subspace $\mathcal{H}_K$ according to measure μ_K.

Definition 16 (**Strong Horton law for branch numbers**). *We say that a probability measure μ on $\mathcal{T}$ satisfies a strong Horton law for branch numbers if there exists such a positive (constant) Horton exponent $R \geq 2$ that for any $k \geq 1$,*

$$\left(\frac{N_k[T]}{N_1[T]}; \, T \overset{d}{\sim} \mu_K \right) \ \overset{p}{\longrightarrow} \ R^{1-k}, \quad as \quad K \to \infty, \tag{27}$$

that is, for any $\epsilon > 0$

$$\mu_K \left(\left| \frac{N_k[T]}{N_1[T]} - R^{1-k} \right| > \epsilon \right) \to 0 \quad as \quad K \to \infty. \tag{28}$$

Corollary 5 in Sect. 7.6.2 is an example of the strong Horton law for branch numbers. In the context of Horton laws, the adjective *strong* refers to the type of geometric decay, while the convergence of random variables is in probability. Section 4.2 discusses weaker types of geometric convergence. An alternative, weaker, definition of the Horton law is formulated in terms of expected branch counts.

Definition 17 (**Strong Horton law for mean branch numbers**). *We say that a probability measure μ on $\mathcal{T}$ satisfies a strong Horton law for mean branch numbers if there exists such a positive (constant) Horton exponent $R \geq 2$ that for any $k \geq 1$,*

$$\lim_{K \to \infty} \frac{\mathsf{E}_K \left[N_k[T] \right]}{\mathsf{E}_K \left[N_1[T] \right]} = \lim_{K \to \infty} \frac{\mathcal{N}_k[K]}{\mathcal{N}_1[K]} = R^{1-k}. \tag{29}$$

Lemma 1. *The strong Horton law for branch numbers (Def. 16) implies the strong Horton law for mean branch numbers (Def. 17).*

Proof. By construction, if $\mathsf{ord}(T) = K$, then $N_1[T] \geq 2^{K-1}$. Accordingly, for any $k \leq K$ we have $\frac{N_k[T]}{N_1[T]} \leq 2^{1-k}$. Assuming the strong Horton law (28) for branch numbers, for any given $\epsilon > 0$, we have

$$\mu_K \left(\left| \frac{N_k[T]}{N_1[T]} - R^{1-k} \right| > \epsilon \right) < \epsilon$$

for all sufficiently large K. Thus, for a given $k \in \mathbb{N}$ and for all sufficiently large K exceeding k, we have

$$\left| \frac{\mathcal{N}_k[K]}{\mathcal{N}_1[K]} - R^{1-k} \right| = \frac{\left| \mathsf{E}_K \left[N_1[T] \left(\frac{N_k[T]}{N_1[T]} - R^{1-k} \right) \right] \right|}{\mathsf{E}_K \left[N_1[T] \right]} \leq \frac{\mathsf{E}_K \left[N_1[T] \left| \frac{N_k[T]}{N_1[T]} - R^{1-k} \right| \right]}{\mathsf{E}_K \left[N_1[T] \right]}$$

$$\leq \frac{\epsilon \mathsf{E}_K \left[N_1[T] \right] + \epsilon 2^{1-k}}{\mathsf{E}_K \left[N_1[T] \right]} \ \leq \epsilon + \epsilon 2^{2-k-K} \ < 2\epsilon,$$

as $\left| \frac{N_k[T]}{N_1[T]} - R^{1-k} \right| \leq \max \left(2^{1-k}, R^{1-k} \right) \leq 2^{1-k}$. This establishes (29). $\qquad \square$

A similar calculation allows us to establish the following result.

Lemma 2. *Consider a probability measure μ on $\mathcal{T}$ and suppose the following properties hold:*

(i) μ *satisfies the strong Horton law for mean branch numbers (Def. 17), and*
(ii) $\forall k \geq 1$ $\exists L_k \in [0, \infty)$ *such that* $\left(\frac{N_k[T]}{N_1[T]}; T \overset{d}{\sim} \mu_K \right) \overset{p}{\to} L_k$ *as* $K \to \infty$.

Then, the measure μ satisfies the strong Horton law for branch numbers (Def. 16), i.e., $L_k = R^{1-k}$.

Sufficient conditions for the strong Horton law for mean branch numbers in binary trees were found in [96], hence providing rigorous foundations for the celebrated regularity that has escaped a formal explanation for a long time. These conditions are presented in Thm. 1 of this section. It has been shown in [97] that the tree that describes a trajectory of Kingman's coalescent process with N particles obeys a weaker version of Horton law as $N \to \infty$ (Sect. 9), and that the first pruning of this tree for any finite N is equivalent to a level set tree of a white noise (see Sect. 8 for definitions).

Consider a mean self-similar measure μ on $\mathcal{BT}^|$ with a Tokunaga sequence $\{T_j\}_{j=1,2,\dots}$. Define a sequence $t(j)$ as

$$t(0) = -1, \; t(1) = T_1 + 2, \; \text{and } t(j) = T_j \text{ for } j \geq 2, \tag{30}$$

and let $\hat{t}(z)$ denote the generating function of $\{t(j)\}_{j=0,1,\dots}$:

$$\hat{t}(z) = \sum_{j=0}^{\infty} z^j t(j) = -1 + 2z + \sum_{j=1}^{\infty} z^j T_j. \tag{31}$$

For a holomorphic function $f(z)$ represented by a power series $f(z) = \sum\limits_{j=0}^{\infty} a_j z^j$ in a nonempty disk $|z| \leq \rho$ we write

$$\check{f}(j) = \frac{1}{2\pi i} \oint\limits_{|z|=\rho} \frac{f(z)}{z^{j+1}} dz = a_j. \tag{32}$$

Theorem 1 (Strong Horton law in a mean self-similar tree). *Suppose μ is a mean self-similar measure on $\mathcal{BT}^|$ with a Tokunaga sequence $\{T_j\}_{j=1,2,\dots}$ such that*

$$\limsup_{j \to \infty} T_j^{1/j} < \infty. \tag{33}$$

Then the strong Horton law for mean branch numbers (Def. 17) holds with the Horton exponent $R = 1/w_0$, where w_0 is the only real zero of the generating function $\hat{t}(z)$ in the interval $\left(0, \frac{1}{2}\right]$. Moreover,

$$\mathcal{N}_1[K+1] = -\widetilde{\left(\frac{1}{\hat{t}}\right)}(K) \tag{34}$$

and

$$\lim_{K \to \infty} \left(\mathcal{N}_1[K] \, R^{-K} \right) = const. > 0. \tag{35}$$

Conversely, if $\limsup_{j \to \infty} T_j^{1/j} = \infty$, *then the limit* $\lim_{K \to \infty} \frac{\mathcal{N}_k[K]}{\mathcal{N}_1[K]}$ *does not exist at least for some* k.

Proof. The proof of Thm. 1 is given in Sect. 4.1.   □

That the Horton exponent R is reciprocal to the real root of $\hat{t}(z)$ was noticed by Peckham [131], under the assumption $\lim_{K \to \infty} \left(N_k R^{k-K} \right) = const. > 0$. We observe that formula (34) in Thm. 1 yields a closed form expression of $\mathcal{N}_k[K]$ via Prop. 3, and, therefore, is instrumental in evaluating the rate of convergence in (35). This is illustrated below in Ex. 10.

The following two examples illustrate application of Thm. 1.

Example 10 (Tokunaga self-similar trees). Consider a Tokunaga self-similar tree (Def. 15) with $T_j = a \, c^{j-1}$, where $a, c > 0$. (We exclude the case $a = 0 \Rightarrow T_j = 0$, which correspond to perfect binary trees with no side branching.) This model received considerable attention in the literature [116, 126, 131, 155], in part because of its ability to closely describe river networks [178]. Here we have

$$\limsup_{j \to \infty} T_j^{1/j} = c < \infty,$$

so the condition (33) of Thm. 1 is satisfied, and

$$
\begin{aligned}
\hat{t}(z) &= -1 + 2z + az \sum_{j=1}^{\infty} (cz)^{j-1} = -1 + 2z + \frac{az}{1 - cz} \\
&= \frac{-1 + (a + c + 2)z - 2cz^2}{1 - cz} \text{ for } |z| < 1/c. \tag{36}
\end{aligned}
$$

The discriminant of the quadratic polynomial in the numerator is positive,

$$(a + c + 2)^2 - 8c > (c + 2)^2 - 8c = (c - 2)^2 \geq 0.$$

Therefore, there exist two real roots, $z_1 < z_2$, of the numerator. It is easy to check that

$$z_1 z_2 = (2c)^{-1}, 0 < z_1 < \min\{2^{-1}, c^{-1}\}, \text{ and } z_2 > \max\{2^{-1}, c^{-1}\}.$$

Hence, there is a single root of $\hat{t}(z) = 0$ for $|z| < 1/c$ of algebraic multiplicity one:

$$z_1 \equiv w_0 = \frac{a + c + 2 - \sqrt{(a + c + 2)^2 - 8c}}{4c},$$

and the respective Horton exponent is

$$R = 1/w_0 = \frac{a + c + 2 + \sqrt{(a + c + 2)^2 - 8c}}{2} \tag{37}$$
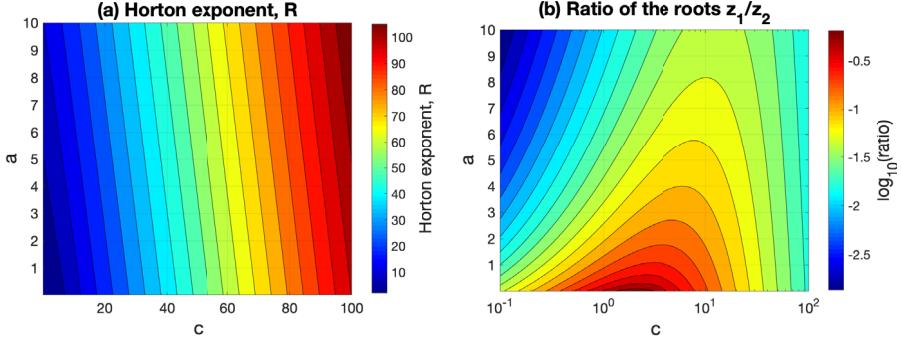
FIG 18. *Strong Horton law in a Tokunaga mean self-similar tree with $T_j = ac^{j-1}$, $j \geq 1$. (a) Horton exponent R as a function of the Tokunaga parameters (a, c). (b) The ratio $0 < z_1/z_2 < 1$ of the two roots of the equation $2cz^2 - (a + c + 2)z + 1 = 0$ as a function of the Tokunaga parameters (a, c). This ratio controls the rate of convergence in the strong Horton law – small values increase the rate.*

as was observed in [155, 131, 116]. A map of the values of the Horton exponent $R(a, c)$ is shown in Fig. 18a. As suggested by (37), the level sets of $R(a, c)$ are fairly approximated by $a + c = const$.

To examine the rate of convergence in the strong Horton law, we use Prop. 3 and formula (34). The reciprocal generating function for the examined Tokunaga tree is given by

$$
\begin{aligned}
-\frac{1}{\hat{t}(z)} &= \frac{1 - cz}{2c(z - z_1)(z - z_2)} \\
&= \frac{1 - cz}{2c(z_2 - z_1)} \left(\frac{1}{z - z_1} - \frac{1}{z - z_2} \right).
\end{aligned} \tag{38}
$$

Thus, since $\frac{1}{z-p} = -\sum\limits_{k=0}^{\infty} \frac{1}{p^{k+1}} z^k$ for $|z| < |p|$, formula (34) implies

$$
\begin{aligned}
\mathcal{N}_1[K + 1] &= \frac{1}{2c(z_2 - z_1)} \left(\frac{1 - cz_1}{z_1^{K+1}} - \frac{1 - cz_2}{z_2^{K+1}} \right) \\
&= \frac{1 - cz_1}{2c(z_2 - z_1)} \frac{1}{z_1^{K+1}} \left(1 - \left(\frac{z_1}{z_2} \right)^{K+1} \frac{1 - cz_2}{1 - cz_1} \right).
\end{aligned} \tag{39}
$$

We use the equality $R = z_1^{-1}$ of Thm. 1 and rewrite this as

$$
\mathcal{N}_1[K]R^{-K} = \frac{1 - cz_1}{2c(z_2 - z_1)} \left(1 - \left(\frac{z_1}{z_2} \right)^K \frac{1 - cz_2}{1 - cz_1} \right).
$$

This gives us an explicit limit in (35)

$$
\lim_{K \to \infty} \left(\mathcal{N}_1[K] R^{-K} \right) = \frac{1 - cz_1}{2c(z_2 - z_1)}
$$
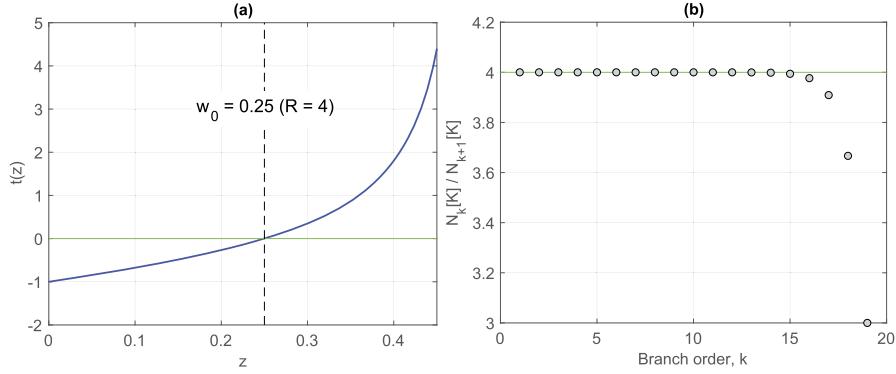
FIG 19. *The strong Horton law in a mean self-similar tree: an illustration. The figure refers to a Tokunaga mean self-similar measure μ with $T_j = 2^{j-1}$, $j \geq 1$. (a) Characteristic function $\hat{t}(z)$ (solid blue). The zero level is marked by a green horizontal line. The real solution $w_0 = 1/4$ is depicted by a vertical dashed line. (b) Ratio $\mathcal{N}_k[K]/\mathcal{N}_{k+1}[K]$ for tree order $K = 20$ and branch orders $k = 1, \ldots, 19$. The strong Horton law suggests $\mathcal{N}_k[K]/\mathcal{N}_{k+1}[K] \approx R = 4$ for large K and k not too close to K.*

and indicates that the respective rate of convergence is determined by the ratio $z_1/z_2 < 1$. Specifically, smaller values lead to faster convergence. Recall (Prop. 3) that

$$\mathcal{N}_1[m+1] = \mathcal{N}_{K-m}[K], \quad 0 \leq m \leq K-1, \quad K \geq 1.$$

Hence, the ratio z_1/z_2 also determines the rate of convergence in (29). Figure 18(b) shows the ratio z_1/z_2 as a function of (a, c). The only region when the ratio is approaching 1, hence slowing down the convergence rate in the strong Horton law, corresponds to $\{c \approx 2, a < 1\}$.

Figure 19 illustrates the strong Horton law in a Tokunaga mean self-similar tree with $a = 1, c = 2$, which corresponds to $T_j = 2^{j-1}$, $j \geq 1$. In this case (Figs. 19(a), 20)

$$\hat{t}(z) = \frac{-1 + 5z - 4z^2}{1 - 2z} = \frac{-4(z-1)(z-1/4)}{1 - 2z}, \quad |z| < 1/2. \tag{40}$$

The ratios $\mathcal{N}_k[K]/\mathcal{N}_{k+1}[K]$ for $K = 20$ are shown in Fig. 19(b). The ratios are very close to the theoretical value $R = 1/w_0 = 4$, except for the branch orders k close to the tree order K, $k > 15$. As suggested by Fig. 18(b), for most of the choices (a, c) the convergence rate is higher, so we expect to have a larger number of ratios in a close vicinity of the limit value R. As we discussed above, the convergence in (35) has the same rate, with first terms (small k) deviating from the limit value rather then the last ones, as was the case in (29) and Fig. 19(b).

We show below in Eq. 47 that, in general, the rate of convergence in the strong Horton law (29), (35) is controlled by

$$\min_{|z| < \gamma} |\hat{t}(z)|,$$
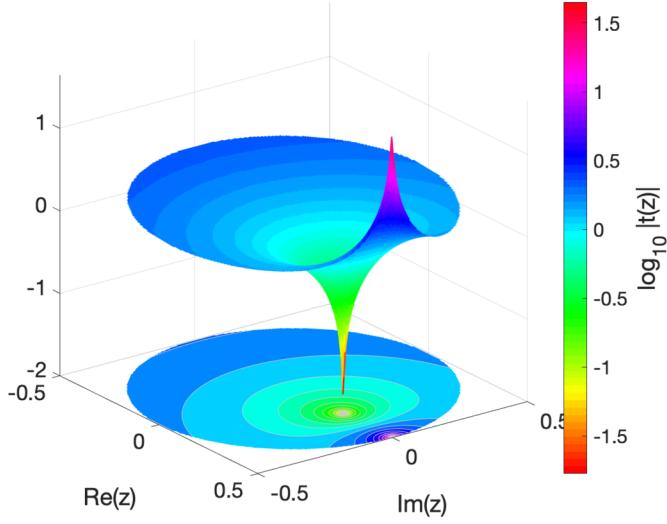
FIG 20. *Generating function $\hat{t}(z)$ for a mean self-similar tree with Tokunaga sequence $T_j = 2^{j-1}$, $j \geq 1$, see (40). The figure shows the value $\log_{10} |\hat{t}(z)|$ for visual convenience. The values of $\hat{t}(z)$ are well separated from its only zero at $z = 1/4$, ensuring a high convergence rate in the strong Horton law.*

where γ separates w_0 from other possible zeros of $\hat{t}(z)$ – higher values lead to faster convergence. Figure 20 shows the value $\log_{10} |\hat{t}(z)|$ on its disk on convergence for the Tokunaga tree of this example. Here, the only zero of $\hat{t}(z)$ at $z = 1/4$ (downward peak) is well isolated so that the surrounding values are separated from zero; this suggests a high rate of convergence that we already illustrated more directly in (39) and Figs. 18(b), 19(b).

Example 11 (Shallow side branching). Suppose $T_j = 0$ for $j \geq 3$, that is we only have "shallow" side branches of orders $\{k-2, k\}$ and $\{k-1, k\}$. Then

$$\hat{t}(z) = -1 + (T_1 + 2)\, z + T_2\, z^2.$$

The only root of this equation within $[0, 1/2]$ is

$$w_0 = \frac{\sqrt{(T_1 + 2)^2 + 4T_2} - (T_1 + 2)}{2\, T_2},$$

which leads to

$$R = 1/w_0 = \frac{\sqrt{(T_1 + 2)^2 + 4T_2} + (T_1 + 2)}{2}.$$

In particular, if $T_j = 0$ for $j \geq 2$, then $R = T_1 + 2$; such trees are called "cyclic" [131]. This shows that the entire range of Horton exponents $2 \leq R < \infty$ can be achieved by trees with only very shallow side branching.

We conclude this section with a linear algebra construction that clarifies the essence of Horton law in a mean self-similar tree. Define a vector $\zeta_K \in \mathbb{R}^K$ of average Horton numbers and a respective normalized vector $\xi_K \in \mathbb{R}^\infty$ as

$$\zeta_K = \begin{pmatrix} \mathcal{N}_1[K] \\ \mathcal{N}_2[K] \\ \vdots \\ \mathcal{N}_K[K] \end{pmatrix} \quad \text{and} \quad \xi_K := \frac{1}{\mathcal{N}_1[K]} \begin{pmatrix} \zeta_K \\ 0 \\ 0 \\ \vdots \end{pmatrix} = \begin{pmatrix} 1 \\ \mathcal{N}_2[K]/\mathcal{N}_1[K] \\ \vdots \\ \mathcal{N}_K[K]/\mathcal{N}_1[K] \\ 0 \\ 0 \\ \vdots \end{pmatrix}$$

and consider an infinite dimensional extension to operator $\mathbb{G}_K$ of (23):

$$\mathbb{G} := \begin{bmatrix} -1 & T_1+2 & T_2 & T_3 & \dots \\ 0 & -1 & T_1+2 & T_2 & \dots \\ 0 & 0 & -1 & T_1+2 & \ddots \\ 0 & 0 & 0 & -1 & \ddots \\ \vdots & \vdots & \ddots & \ddots & \ddots \end{bmatrix}. \tag{41}$$

Using these notations, the main counting equations (24) becomes $\mathbb{G}_K \zeta_K = -\mathbf{e}_K$, and therefore

$$\mathbb{G}\xi_K = -\frac{\mathbf{e}_K}{\mathcal{N}_1[K]}.$$

Here $\mathcal{N}_1[K] \geq (T_1+2)^{K-1} \to \infty$ as $K \to \infty$, and hence the strong Horton law for mean branch numbers (Def. 17) is equivalent to the existence of a limit solution $\lim_{K\to\infty} \xi_K = \xi$ to an infinite dimensional linear operator equation

$$\mathbb{G}\xi = 0$$

with coordinates $\xi(k) = R^{1-k}$.

4.1. Proof of Theorem 1

First, we establish (Prop. 4) necessary and sufficient conditions for the existence of the strong Horton law. Then we show that these conditions are satisfied and express the value of the Horton exponent R via the Tokunaga coefficients $\{T_j\}$.

Proposition 4. *Let μ be a mean self-similar measure on $\mathcal{BT}^|$. Suppose that the limit*

$$R = \lim_{K\to\infty} \frac{\mathcal{N}_1[K+1]}{\mathcal{N}_1[K]} \tag{42}$$

exists and is finite. Then, the strong Horton law for mean branch numbers holds; that is, for each positive integer k,

$$\lim_{K\to\infty} \frac{\mathcal{N}_k[K]}{\mathcal{N}_1[K]} = R^{1-k}. \tag{43}$$

Conversely, if the limit (42) *does not exist, then the limit in the left hand side of* (43) *also does not exist, at least for some k.*

Proof. Suppose the limit (42) exists and is finite. Proposition 3 implies that for any fixed integer $m \geq 1$

$$\frac{\mathcal{N}_{m+1}[K]}{\mathcal{N}_m[K]} = \frac{\mathcal{N}_1[K-m]}{\mathcal{N}_1[K-m+1]} \to R^{-1}, \text{ as } K \to \infty.$$

Thus, for any fixed integer $k \geq 2$,

$$\frac{\mathcal{N}_k[K]}{\mathcal{N}_1[K]} = \prod_{m=1}^{k-1} \frac{\mathcal{N}_{m+1}[K]}{\mathcal{N}_m[K]} \to R^{1-k}, \text{ as } K \to \infty.$$

Conversely, suppose the limit $\lim_{K \to \infty} \frac{\mathcal{N}_1[K+1]}{\mathcal{N}_1[K]}$ does not exist. Taking $k = 2$, we obtain by Prop. 3

$$\frac{\mathcal{N}_2[K]}{\mathcal{N}_1[K]} = \frac{\mathcal{N}_1[K-1]}{\mathcal{N}_1[K]}.$$

Thus $\lim_{K \to \infty} \frac{\mathcal{N}_2[K]}{\mathcal{N}_1[K]}$ diverges. $\qquad \square$

Next, we express $\mathcal{N}_1[K]$ via the elements of the Tokunaga sequence $\{T_j\}_{j=1,2,\dots}$ that satisfy condition (33). The quantity $\mathcal{N}_1[K+1]$ can be computed by counting, and expressed via convolution products as follows:

$$\begin{aligned}
\mathcal{N}_1[K+1] &= \sum_{r=1}^{K} \sum_{\substack{j_1,j_2,\dots,j_r \geq 1 \\ j_1+j_2+\dots+j_r=K}} t(j_1)t(j_2)\dots t(j_r) \\
&= \sum_{r=1}^{K} \underbrace{(t+\delta_0)*(t+\delta_0)*\dots*(t+\delta_0)}_{r \text{ times}}(K) \\
&= \sum_{r=1}^{\infty} \underbrace{(t+\delta_0)*(t+\delta_0)*\dots*(t+\delta_0)}_{r \text{ times}}(K),
\end{aligned}$$

where $\delta_0(j)$ is the Kronecker delta, and therefore, $(t+\delta_0)(0) = 0$. Hence, taking the z-transform of $\mathcal{N}_1[K]$, we obtain

$$\sum_{K=1}^{\infty} z^{K-1}\mathcal{N}_1[K] = 1 + \sum_{r=1}^{\infty} \left[\widehat{(t+\delta_0)}(z)\right]^r = 1 + \sum_{r=1}^{\infty} \left[\hat{t}(z)+1\right]^r = -\frac{1}{\hat{t}(z)} \quad (44)$$

for $|z|$ small enough. Recalling the definition (32) establishes (34):

$$\mathcal{N}_1[K+1] = -\widetilde{\left(\frac{1}{\hat{t}}\right)}(K).$$

Since $T_j \geq 0$ for any $j \geq 1$, the function $\hat{t}(z) = -1 + 2z + \sum_{j=1}^{\infty} z^j T_j$ has a single real root w_0 in the interval $(0, 1/2)$. Our goal is to show that the Horton exponent R is reciprocal to w_0. We begin by showing that w_0 is the root of $\hat{t}(z)$ closest to the origin.

Lemma 3. *Let w_0 be the only real root of $\hat{t}(z) = -1 + 2z + \sum_{j=1}^{\infty} z^j T_j$ in the interval $(0, 1/2]$. Then, for any other root w of $\hat{t}(z)$, we have $|w| > w_0$.*

Proof. Since $\{T_j\}$ are all nonnegative reals, we have $\overline{\hat{t}(\bar{z})} = \hat{t}(z)$. The radius of convergence of $\sum_{j=1}^{\infty} z^j T_j$ must be greater than w_0. Suppose $w = re^{i\theta}$ $(0 \leq \theta < 2\pi)$ is a root of magnitude at most w_0. That is $\hat{t}(w) = 0$ and $r := |w| \leq w_0$. Then $\hat{t}(\bar{w}) = 0$ and

$$0 = \frac{1}{2}\left[\hat{t}(w) + \hat{t}(\bar{w})\right] = -1 + 2r\cos(\theta) + \sum_{j=1}^{\infty} r^j T_j \cos(j\theta).$$

If $r < w_0$, then

$$0 = -1 + 2r\cos(\theta) + \sum_{j=1}^{\infty} r^j T_j \cos(j\theta) \leq -1 + 2r + \sum_{j=1}^{\infty} r^j T_j$$

$$< -1 + 2w_0 + \sum_{j=1}^{\infty} w_0^j T_j = 0$$

arriving to a contradiction. Thus $r = w_0$.

Next we show that $\theta = 0$. Suppose not. Then

$$0 = -1 + 2r\cos(\theta) + \sum_{j=1}^{\infty} r^j T_j \cos(j\theta) < -1 + 2r + \sum_{j=1}^{\infty} r^j T_j$$

$$= -1 + 2w_0 + \sum_{j=1}^{\infty} w_0^j T_j = 0$$

arriving to another contradiction. Hence $r = w_0$, $\theta = 0$, and $w = w_0$. $\square$

Let $L = \limsup_{j\to\infty} T_j^{1/j}$. Then L^{-1} is the radius of convergence of $\hat{t}(z)$ (we set $L^{-1} = \infty$ if $L = 0$), and $L^{-1} > w_0$. Lemma 3 asserts that there exists a positive real $\gamma \in (w_0, L^{-1})$ such that

$$\gamma < w \quad \text{for all} \quad w \neq w_0 \quad \text{such that} \quad \hat{t}(w) = 0. \tag{45}$$

Accordingly, for $0 < \rho < w_0$

$$\mathcal{N}_1[K] = \frac{-1}{2\pi i} \oint_{|z|=\rho} \frac{dz}{\hat{t}(z)z^K} = -Res\left(\frac{1}{\hat{t}(z)z^K}; w_0\right) - \frac{1}{2\pi i} \oint_{|z|=\gamma} \frac{dz}{\hat{t}(z)z^K}. \tag{46}$$

Observe that $Res\left(\frac{1}{\hat{t}(z)z^K}; w_0\right)$ is a constant multiple of w_0^{-K} since w_0 is a root of $\hat{t}(z)$ of algebraic multiplicity one. Thus, since $w_0 < \gamma$ and

$$\left| \frac{1}{2\pi i} \oint_{|z|=\gamma} \frac{dz}{\hat{t}(z)z^K} \right| \leq \frac{1}{\gamma^K \min_{|z|=\gamma} |\hat{t}(z)|} = o\left(w_0^{-K}\right), \quad K \to \infty, \qquad (47)$$

we have

$$\frac{\mathcal{N}_1[K+1]}{\mathcal{N}_1[K]} = \left| \frac{\mathcal{N}_1[K+1]}{\mathcal{N}_1[K]} \right| \to \frac{1}{w_0} \quad \text{as } K \to \infty.$$

Proposition 4 now implies the following lemma.

Lemma 4. *Suppose* $\limsup_{j\to\infty} T_j^{1/j} < \infty$. *Then, for each positive integer k*

$$\lim_{K\to\infty} \frac{\mathcal{N}_k[K]}{\mathcal{N}_1[K]} = w_0^{k-1}.$$

Moreover,

$$\lim_{K\to\infty} \left(\mathcal{N}_1[K]\, w_0^K \right) = const. > 0.$$

To establish the converse we need the following statement.

Proposition 5. *Suppose μ is a mean self-similar measure on $\mathcal{BT}^|$ with Tokunaga sequence $\{T_j\}_{j\geq 1}$. Then*

$$\mathcal{N}_1[K] \geq T_j^{(K-1)/j}$$

for all $j \in \mathbb{N}$ and $(K-1) \in j\mathbb{N}$.

Proof. Fix any $j \geq 1$. The main counting equations (22) show that for any integer $m \geq 0$

$$\mathcal{N}_{mj+1}[K] \geq T_j \mathcal{N}_{(m+1)j+1}[K].$$

Accordingly,

$$\mathcal{N}_1[K] \geq T_j^m \mathcal{N}_{mj+1}[K],$$

given $mj + 1 \leq K$. Choosing $m = (K-1)/j$ we obtain

$$\mathcal{N}_1[K] \geq T_j^{(K-1)/j} \mathcal{N}_K[K] = T_j^{(K-1)/j}. \qquad \square$$

Suppose the limit

$$R = \lim_{K\to\infty} \frac{\mathcal{N}_1[K+1]}{\mathcal{N}_1[K]}$$

exists and is finite. Proposition 5 asserts that $\mathcal{N}_1[K]^{1/(K-1)} \geq T_j^{1/j}$ for all $j \in \mathbb{N}$ and $(K-1) \in j\mathbb{N}$. Hence,

$$\limsup_{j\to\infty} T_j^{1/j} \leq \lim_{K\to\infty} \mathcal{N}_1[K]^{1/(K-1)} = R < \infty.$$

We summarize this in a lemma.

Lemma 5. *Suppose* $\limsup\limits_{j \to \infty} T_j^{1/j} = \infty$. *Then, the limit* $\lim\limits_{K \to \infty} \frac{\mathcal{N}_k[K]}{\mathcal{N}_1[K]}$ *does not exist at least for some* k.

Finally, Thm. 1 follows from Lem. 4 and Lem. 5.

4.2. Well-defined asymptotic Horton ratios

The setting for Horton law in (27) and (29) can be generalized beyond randomizing the tree measure with respect to Horton-Strahler orders as in (7). For instance, as it will be the case with the combinatorial critical binary Galton-Watson trees $\mathcal{GW}\left(\frac{1}{2}, \frac{1}{2}\right)$ in (66), the tree measure may be randomized with respect to the number of leaves in a tree. A general set up for the Horton laws is described below.

Let $\mathcal{Q}_n$, $n \in \mathbb{N}$, be a sequence of probability measures on $\mathcal{T}$. We write $N_j^{(\mathcal{Q}_n)}$ for the number of branches of Horton-Strahler order $j \geq 1$ in a tree generated according to $\mathcal{Q}_n$.

Definition 18 (**Well-defined asymptotic Horton ratios**). *We say that a sequence of probability measures* $\{\mathcal{Q}_n\}_{n \in \mathbb{N}}$ *has well-defined asymptotic Horton ratios if for each* $j \geq 1$

$$\frac{N_j^{(\mathcal{Q}_n)}}{N_1^{(\mathcal{Q}_n)}} \xrightarrow{p} \mathcal{N}_j \quad as \ \ n \to \infty, \tag{48}$$

where $\mathcal{N}_j$ *is a constant, called the asymptotic Horton ratio of the branches of order* j.

Sometimes it is possible to establish a stronger limit than in (48). One such example is the almost sure convergence in equation (135) of Sect. 7.6.2.

For a sequence of well-defined asymptotic Horton ratios $\mathcal{N}_j$, the *Horton law* states that $\mathcal{N}_j$ decreases in a geometric fashion as j goes to infinity. We consider three particular forms of geometric decay.

Definition 19 (**Root, ratio, and strong Horton laws**). *Consider a sequence* $\{\mathcal{Q}_n\}_{n \in \mathbb{N}}$ *of probability measures on* $\mathcal{T}$ *with well-defined asymptotic Horton ratios (Def. 18). Then, the sequence* $\{\mathcal{Q}_n\}$ *is said to obey*

- *a root-Horton law if the following limit exists:* $\lim\limits_{j \to \infty} \left(\mathcal{N}_j\right)^{-\frac{1}{j}} = R$;
- *a ratio-Horton law if the following limit exists:* $\lim\limits_{j \to \infty} \frac{\mathcal{N}_j}{\mathcal{N}_{j+1}} = R$;
- *a strong Horton law if the following limit exists:* $\lim\limits_{j \to \infty} \left(\mathcal{N}_j R^j\right) = const.$

The constant R *is called the Horton exponent. In each case, we require the Horton exponent* R *to be finite and positive.*

Observe that the Horton laws in Def. 19 above are listed in the order from weaker to stronger.

4.3. Entropy and information theory

The (information) *entropy* introduced by Claude Shannon in 1948 [144] is the main concept in information theory. Informally, entropy is a measure of noise, uncertainty, or surprise. It can be also interpreted as the average amount of information in a stochastic model. It has a broad range of applications, including data communication, cryptography, and combinatorics. The *entropy rate* can be thought as the asymptotic average information in a stochastic model.

The information theoretical aspects of self-similar trees were not addressed until very recently. This section reviews recent results by Chunikhina [38, 39], where the entropy rate is computed for trees that satisfy the strong Horton law for branch numbers (Def. 16) and for Tokunaga self-similar trees (Def. 15) as a function of the respective parameters, R and (a, c).

Consider a subspace $\mathcal{T}_{N_1,\ldots,N_K}$ of $\mathcal{BT}^|_{\text{plane}}$ of trees of a given order $\text{ord}(T) = K$ and given admissible ($N_K = 1$, $N_j \geq 2N_{j+1}$) branch counts $N_1, N_2, \ldots, N_K$:

$$\mathcal{T}_{N_1,\ldots,N_K} = \big\{ T \in \mathcal{BT}^|_{\text{plane}} : \text{ord}(T) = K, \ N_1[T] = N_1, \ldots, N_K[T] = N_K = 1 \big\}.$$

In [38], Chunikhina finds the size of $\mathcal{T}_{N_1,\ldots,N_K}$, providing an alternative form of expression that was first derived by Shreve [146].

Lemma 6 (**Branch counting lemma, [38]**).

$$\big| \mathcal{T}_{N_1,\ldots,N_K} \big| = 2^{N_1 - 1 - \sum_{j=2}^{K} N_j} \prod_{j=1}^{K-1} \binom{N_j - 2}{2N_{j+1} - 2}.$$

Subsequently, Lem. 6 is used to find the entropy rate for trees that satisfy the strong Horton law (Def. 16) with exponent $R > 2$.

Theorem 2 (**Entropy rate for Horton self-similar trees, [38]**). *For a given $R > 2$ and a given $\epsilon \in (0, R)$, consider a random tree T uniformly sampled from the space*

$$\mathcal{T}_{R,K} = \big\{ T \in \mathcal{BT}^|_{\text{plane}} : \text{ord}(T) = K, \ \big| N_j[T] - R^{K-j} \big| < (R - \epsilon)^{K-j} \ \ \forall j \leq K \big\}.$$

Let P and E denote the probability mass function and the expectation, respectively. Then, the entropy rate

$$\mathcal{H}_\infty(R) := \lim_{K \to \infty} \frac{-\mathsf{E}[\log_2 \mathsf{P}(T)]}{2R^{K-1}} = 1 - \frac{1 - H(2/R)}{2 - 2/R}, \tag{49}$$

where

$$H(z) = -z \log_2 z - (1 - z) \log_2 (1 - z), \quad 0 < z < 1 \tag{50}$$

is the binary entropy function illustrated in Fig. 21(a). The entropy rate $\mathcal{H}_\infty(R)$ is illustrated in Fig. 21(b).

Notice that the trees in $\mathcal{T}_{R,K}$ satisfy the strong Horton law (Def. 16) with the Horton exponent R, and $2R^{K-1}$ is the asymptotic number of nodes in a tree T from $\mathcal{T}_{R,K}$.
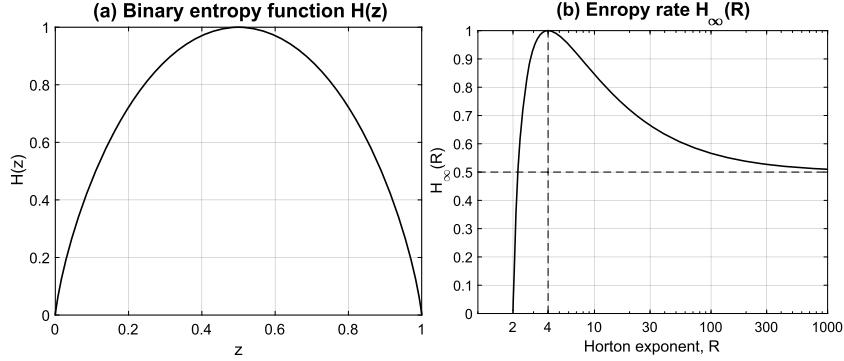
FIG 21. *Entropy rate in trees that satisfy the strong Horton law with exponent R. (a) Binary entropy function $H(z)$. (b) Entropy rate $\mathcal{H}_\infty(R)$.*

Remark 9. It is an easily verified fact that a random tree T selected uniformly from the subspace

$$\mathcal{BT}^|_{\text{plane}}(N) := \{T \in \mathcal{BT}^|_{\text{plane}} : \#T = 2N - 1\} \tag{51}$$

of $\mathcal{BT}^|_{\text{plane}}$ containing only the trees with N leaves ($2N$ nodes and $2N-1$ edges) is distributed as a random tree sampled from the critical plane Galton-Watson distribution $\mathcal{GW}_{\text{plane}}\left(\frac{1}{2}, \frac{1}{2}\right)$, conditioned on $\#T = 2N - 1$, i.e.,

$$\mathsf{Unif}\left(\mathcal{BT}^|_{\text{plane}}(N)\right) \;\overset{d}{=}\; \left(\mathcal{GW}_{\text{plane}}\left(\frac{1}{2}, \frac{1}{2}\right) \,\Big|\, \#T = 2N - 1\right). \tag{52}$$

Consequently, we have that

$$T \overset{d}{\sim} \mathsf{Unif}\left(\mathcal{BT}^|_{\text{plane}}(N)\right) \;\Rightarrow\; \text{SHAPE}(T) \overset{d}{\sim} \left(\mathcal{GW}\left(\frac{1}{2}, \frac{1}{2}\right) \,\Big|\, \#T = 2N - 1\right). \tag{53}$$

The number $\left|\mathcal{BT}^|_{\text{plane}}(N)\right|$ of different combinatorial shapes of rooted planted plane binary trees with N leaves and $2N-1$ edges, is given by C_{N-1}, where C_n denotes the *Catalan number* defined as

$$C_n = \frac{1}{n+1}\binom{2n}{n}. \tag{54}$$

Using $\left|\mathcal{BT}^|_{\text{plane}}(N)\right| = C_{N-1}$ and Stirling's formula, it is observed in [38] that the entropy rate for a tree T', selected uniformly from $\mathcal{BT}^|_{\text{plane}}(N)$ is

$$\mathcal{H}_\infty^{\text{GW}} := \lim_{N \to \infty} \frac{-\mathsf{E}[\log_2 \mathsf{P}(T')]}{2N} = 1. \tag{55}$$

Thus, scaling by the asymptotic number of nodes $2R^{K-1}$ in Thm. 2 implies

$$\mathcal{H}_\infty(R) \leq \mathcal{H}_\infty^{\text{GW}} = 1.$$

Indeed, by definition of the corresponding spaces,

$$\mathcal{T}_{R,K} \subseteq \bigcup_N \mathcal{BT}^{|}_{\text{plane}}(N),$$

where the union is taken over N ranging from

$$\lceil R^{K-1} - (R-\epsilon)^{K-1} \rceil \quad \text{to} \quad \lfloor R^{K-1} + (R-\epsilon)^{K-1} \rfloor,$$

and therefore

$$\left| \mathcal{T}_{R,K} \right| \leq 2(R-\epsilon)^{K-1} \left| \mathcal{BT}^{|}_{\text{plane}} \left(2R^{K-1} + 2(R-\epsilon)^{K-1} \right) \right|.$$

Hence, for the following limits known to converge, we have

$$\mathcal{H}_\infty(R) = \lim_{K \to \infty} \frac{\log_2 \left| \mathcal{T}_{R,K} \right|}{2R^{K-1}}$$

$$\leq \mathcal{H}_\infty^{\text{GW}} = \lim_{K \to \infty} \frac{\log_2 \left(2(R-\epsilon)^{K-1} \left| \mathcal{BT}^{|}_{\text{plane}} \left(2R^{K-1} + 2(R-\epsilon)^{K-1} \right) \right| \right)}{2R^{K-1}}.$$

Moreover, scaling by the asymptotic number of nodes $2R^{K-1}$ in Thm. 2 enables representing $\mathcal{H}_\infty(R)$ as the limit ratio of the entropy for Horton self-similar trees with parameter R to the entropy for uniformly selected binary trees. Specifically, let T be a random tree sampled uniformly from the space $\mathcal{T}_{R,K}$ and let T' be a random tree sampled uniformly from the space $\mathcal{BT}^{|}_{\text{plane}}(N)$ with $N = R^{K-1}$. Then, equations (49) and (55) imply that $\mathcal{H}_\infty(R)$ is the the limit ratio of entropies as the space sizes grow with $K \to \infty$:

$$\mathcal{H}_\infty(R) = \lim_{K \to \infty} \frac{-\mathsf{E}[\log_2 \mathsf{P}(T)]}{-\mathsf{E}[\log_2 \mathsf{P}(T')]} = 1 - \frac{1 - H(2/R)}{2 - 2/R}. \tag{56}$$

As an important consequence of Thm. 2, a special place of the parameter $R = 4$ is established amongst all Horton exponents $R \in [2, \infty)$ as

$$\operatorname{argmax}_R \mathcal{H}_\infty(R) = 4 \quad \text{and} \quad \max_R \mathcal{H}_\infty(R) = \mathcal{H}_\infty(4) = 1.$$

Not surprisingly, $R = 4$ is the parameter value for the strong Horton law results we will encounter in Sect. 5, primarily in the context of the critical binary Galton-Watson tree $\mathcal{GW}\left(\frac{1}{2}, \frac{1}{2}\right)$. Indeed, as stated in Rem. 9, the tree $T'' = \text{SHAPE}(T') \in \mathcal{BT}^{|}$ in (56) is a random tree sampled from the Galton-Watson distribution $\mathcal{GW}\left(\frac{1}{2}, \frac{1}{2}\right)$ conditioned on $\#T'' = 2N - 1$.

In [39], Chunikhina extended the results in [38] by counting the number of trees with the given merger numbers $N_{i,j}$ (see Sect. 3.3), and finding the entropy rates for the Tokunaga self-similar trees (Def. 15) represented as a function of the parameters (a, c). For a given integer $K > 1$, consider a finite sequence

of admissible branch counts $\{N_i\}_{i=1,\dots,K}$, and a finite sequence of admissible branch numbers $\{N_{i,j}\}_{1\leq i<j\leq K}$. Admissibility means that for all $i \leq K-1$,

$$N_i = 2N_{i+1} + \sum_{j=i+1}^{K} N_{i,j}$$

as all N_i branches of Horton-Strahler order i have to merge into a higher order branch (either two branches of order i merge and originate a branch of order $i+1$, or a branch of order i merges into a branch of order $j > i$). Consider the subspace

$$\mathcal{T}_{K,\{N_i\},\{N_{i,j}\}} = \big\{T \in \mathcal{T}_{N_1,\dots,N_K} : N_{1,2}[T] = N_{1,2}, \dots, N_{K-1,K}[T] = N_{K-1,K}\big\}.$$

Lemma 7 (Side branch counting lemma, [39]).

$$\Big|\mathcal{T}_{K,\{N_i\},\{N_{i,j}\}}\Big| = \prod_{j=2}^{K}\prod_{i=1}^{j-1} 2^{N_{i,j}}\binom{A_j}{N_{i,j}}, \qquad A_j = N_j - 1 + \sum_{k=i}^{j-1} N_{k,j}.$$

Lemma 7 is used to obtain the following asymptotic results. Consider Tokunaga self-similar tree with parameters (a, c). Such a tree satisfies the strong Horton law for mean branch numbers (Def. 17) with the Horton exponent (37)

$$R = R(a, c) = \frac{a + c + 2 + \sqrt{(a+c+2)^2 - 8c}}{2}.$$

Next, similarly to $\mathcal{T}_{R,K}$, one can define the space $\mathcal{T}_{a,c,K}$ of asymptotically Tokunaga self-similar trees of order K. Informally, this space includes the trees in $\mathcal{BT}^|_{\text{plane}}$ such that

$$\mathsf{ord}(T) = K, \quad N_j[T] \sim R^{K-j}, \quad \text{and} \quad \frac{N_{i,j}(T)}{N_j(T)} \sim ac^{j-i-1},$$

where $R = R(a, c)$, and the asymptotic equality $\sim$ is taken as $K \to \infty$.

Theorem 3 (Entropy rate for Tokunaga self-similar trees, [39]). *For given $a, c > 0$, let T be a random tree, uniformly sampled from the space $\mathcal{T}_{a,c,K}$. Let P and E denote the probability mass function and the expectation, respectively. Then, the entropy rate*

$$\begin{aligned}
\mathcal{H}_\infty(a,c) &:= \lim_{K\to\infty} \frac{-\mathsf{E}[\log_2 \mathsf{P}(T)]}{2R^{K-1}} \\
&= \frac{a}{2}\sum_{j=1}^{\infty} R^{-j}\left(\frac{1-c^j}{1-c} + \frac{1}{a}\right)\log_2\left(\frac{1-c^j}{1-c} + \frac{1}{a}\right) \\
&\quad + \frac{aR}{2(R-c)(R-1)} + \frac{\log_2 a}{2(R-1)} - \frac{aRc\log_2 c}{2(R-c)^2(R-1)}.
\end{aligned} \tag{57}$$
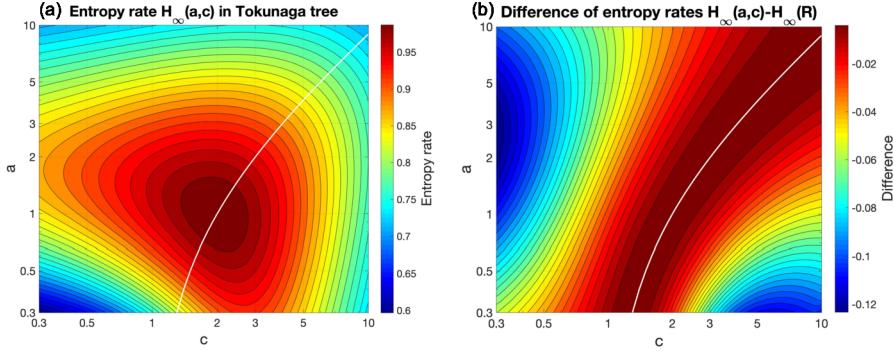
FIG 22. *Entropy rate in Tokunaga trees. (a) Entropy rate $\mathcal{H}_\infty(a,c)$ in a Tokunaga self-similar tree with parameters (a,c). (b) The difference $\mathcal{H}_\infty(a,c) - \mathcal{H}_\infty(R)$ of entropy rates in a Tokunaga tree with parameters (a,c) and in a tree satisfying Horton law with Horton exponent $R(a,c)$. A double-logarithmic scale is used to emphasize behavior of the plots at the boundary values. White line corresponds to $a = c - 1$.*

Figure 22(a) illustrates the entropy rate $\mathcal{H}_\infty(a,c)$.

If $a = c - 1$, then $R = 2c$ by (37), and the equation (57) simplifies, leading to the following corollary.

Corollary 1 ([**39**]). *Let T be a random tree, uniformly sampled from the space $\mathcal{T}_{a,c,K}$ with $c > 1$ and $a = c - 1$. Then T satisfies the strong Horton law (29) with $R = 2c$, and the entropy rate is given by*

$$\mathcal{H}_\infty(c - 1, c) = 1 - \frac{1 - H(1/c)}{2 - 1/c} = \mathcal{H}_\infty(R), \tag{58}$$

where $H(z)$ is the binary entropy function (50) and $\mathcal{H}_\infty(R)$ is defined by (49).

Figure 22(b) illustrates this result, by showing how the difference of entropy rates $\mathcal{H}_\infty(a,c) - \mathcal{H}_\infty(R)$ decreases away from the line $a = c - 1$. The special place for the line $a = c - 1$ within the parameter space of the Tokunaga self-similar random trees was observed earlier in [161, 98, 99]; see Rem. 14. The constraint $a = c - 1$ will reappear in many instances in Sect. 7 of the present work.

Finally, the maximum value $\max \mathcal{H}_\infty(a,c) = 1$ is achieved at the special point $(a, c) = (1, 2)$ of the special line $a = c - 1$. Once again, this is not surprising as $(a, c) = (1, 2)$ is the parameter value for the Tokunaga self-similarity results of Sect. 5, presented in the context of the critical binary Galton-Watson trees $\mathcal{GW}\left(\frac{1}{2}, \frac{1}{2}\right)$ and related processes. We recall that the combinatorial shape $T'' = \text{SHAPE}(T') \in \mathcal{BT}^{|}$ of the random binary tree T' in (55) is distributed according to $\mathcal{GW}\left(\frac{1}{2}, \frac{1}{2}\right)$ conditioned on $\#T'' = 2N - 1$.

4.4. Applications

A quantitative understanding of the branching patterns is instrumental in hydrology [141, 154, 114, 16, 74, 31, 91], geomorphology [45, 82], statistical seis-

mology [14, 157, 84, 177, 71, 174, 172], statistical physics of fracture [142], vascular analysis [87], brain studies [37], ecology [35], biology [159], and beyond, encouraging a rigorous treatment. Introduced in hydrology to describe the dendritic structure of river networks, which is among the most evident examples of natural branching, Horton-Strahler [85, 151] and Tokunaga [155] indexing schemes have been rediscovered and used in other fields. Subsequently, the Horton law (Def. 16) and Tokunaga self-similarities (Def. 15) have been empirically or rigorously established in numerous observed and modeled systems [126]. This includes hydrology (see Sect. 4.4.1), vein structure of botanical leaves [126, 159], diffusion limited aggregation [129, 115, 170], two dimensional site percolation [158, 168, 175, 176], a hierarchical coagulation model of Gabrielov et al. [69] introduced in the framework of self-organized criticality, and a random self-similar network model of Veitzer and Gupta [161] developed as an alternative to the Shreve's random topology model for river networks. The Horton exponent commonly reported in empirical studies is within the range $3 < R < 6$. Curiously, it has been observed in [98] that the critical Tokunaga model (Sect. 7.5) with this range of Horton exponents generates trees with fractal dimension in the range $\approx (1.6, 3)$, which includes all the trees that may exist in a three-dimensional world, excluding the range < 1.6 that corresponds to almost "linear", and probably less studied, structures.

4.4.1. Hydrology

An illuminating natural example of Horton laws and Tokunaga self-similarity is given by the combinatorial structure of river networks (Figs. 2, 3). The hydrological *Horton law* was first described by Robert E. Horton [85] who noticed that the empirical ratio N_K/N_{K+1} in river streams is close to $R_B = 4$; see (1). This observation has been strongly corroborated in numerous observational studies [90, 146, 109, 131, 153, 76, 141, 119, 156]. See Barndorff-Nielsen [18] for a 1993 survey for probabilists.

Geometric scaling is found for many other physical and combinatorial characteristics of river networks. This can be upstream area, magnitude (number of upstream sources), the lengths of the longest channel, the total channel lengths, etc. These quantities often increase with order (unlike branch counts N_K that decrease with order), which justifies a slightly different way the Horton law is commonly formulated. Specifically, consider the values Z_K of a selected statistic Z calculated using streams of order K. Horton law with exponent $R_Z \geq 1$ states that Z_K scale as

$$\frac{Z_{K+1}}{Z_K} = R_Z \quad \Leftrightarrow \quad Z_K \propto R_Z^K. \tag{59}$$

In both cases (1) and (59) the law is formulated in such a way that the Horton exponent is greater than unity. Informally, (59) suggests that the order K of a branch (stream) is proportional to $\ln(Z_K)$, where Z_K can be interpreted as the stream "size".

Section 1.1 mentioned three key regularities that are commonly observed in river stream networks and can be explained and parameterized via Horton laws. Below we discuss each of them in more detail.

Power-law empirical frequencies of river statistics. Consider empirical frequencies of a statistic Z calculated at every edge (link) in a large tree (basin). We write $Z_{(i)}$ for the value of Z calculated at i-th link. Assume that Horton law holds (i) for the examined statistic Z, with exponent R_Z and (for simplicity) proportionality constant 1; (ii) for the average number L_K of edges within a branch of order K, with exponent R_L; and (iii) for branch counts N_K with exponent R_B as in (1). The average number of edges of order K in such a tree is given by $N_K L_K$. One can now heuristically approximate the expected frequencies of $Z_{(i)}$ by using the same value Z_K for any edge of order K and considering a limit of an infinitely large tree:

$$\#\{i : Z_{(i)} \geq R_Z^K\} \ \propto \ \sum_{j=K}^{\infty} N_K L_K = \sum_{j=K}^{\infty} \left(\frac{R_L}{R_B}\right)^j \propto \left(\frac{R_L}{R_B}\right)^K.$$

This is a punctured (by discrete order) version of a general power law relation

$$\#\{i : Z_{(i)} \geq z\} \ \propto \ z^{-\beta}, \quad \beta = \frac{\log R_B - \log R_L}{\log R_Z}. \tag{60}$$

Power-law relations among river statistics. If stream statistic Z satisfies the Horton law (59) with exponent R_Z, and the branch counts N_K satisfy the Horton law (1) with Horton exponent R_B, then

$$Z_K \ \propto \ N_K^{-\alpha}, \quad \text{with} \quad \alpha = \frac{\ln R_Z}{\ln R_B}. \tag{61}$$

Similarly, suppose that Horton law (59) holds for selected river statistics Z and Y, with Horton exponents R_Z and R_Y, respectively. Then, Z_K and Y_K are connected via a power-law relation

$$Z_K \ \propto \ Y_K^{\alpha}, \quad \alpha = \frac{\log R_Z}{\log R_Y}. \tag{62}$$

As above, Eqs. (61), (62) are a punctured (by discrete orders) version of a general power-law relation $Z \propto Y^{\alpha}$ that is abound among hydrologic quantities. It is common to relate a statistic of interest to the basin area A. Section 1.1 mentions Hack's law that scales the length L of the longest stream as $L \propto A^h$ with $h \approx 0.6$ [139].

Modeling physical characteristics of streams. Data analysis suggests that many physical characteristics of streams, including stream width, depth, slope, and flow velocity, scale with the stream discharge (volume of water flowing through a river stream) Q [141, Chapter 1]. For example, the velocity v can be approximated by $v \propto Q^a$, etc. The discharge Q, in turn, is a power-law function of the basin area: $Q \propto A^{\beta}$. The value of exponent β depends on a precise definition of discharge (bankfull, mean annual, etc.) Finally, the basin area A is

closely approximated by the basin magnitude (number of outlets) M, since it is natural to think of a stream network as a space-filling tree. Combining these observations, we find that the stream velocity can be modeled as a power-law function of the stream magnitude:

$$v \propto M^{a\beta}.$$

Similar relations exist for other physical characteristics of a stream. A fundamental scaling is for the average link slope S [75]:

$$S \propto A^{-\theta} \propto M^{-\theta}, \quad \theta \approx 0.5.$$

A solid body of observational and modeling studies attributes the Horton laws and power-law relations among essential river statistics to self-similar structure of river stream networks [75, 131, 72, 153, 76, 40, 132, 133, 161, 141, 74, 119, 156]. It has been shown, in particular, that river networks are closely approximated by a two-parametric Tokunaga self-similar model (Def. 15) with parameters that are independent of river's geographic location [155, 131, 45, 178]. The Tokunaga model closely predicts values of the Horton exponents for multiple basin statistics with only two parameters (see Fig. 3).

One of the best explored models of a river network is the critical binary Galton-Watson tree (Sect. 5), also known in hydrology as Shreve's random topology model [146, 147]; it is equivalent to the uniform distribution on the space of planar binary trees with a fixed number of leaves [135]. This model has the Horton exponent $R = 4$ and Tokunaga parameters $(a, c) = (1, 2)$; see Thm. 4. For long time, the critical binary Galton-Watson tree has remained the only well-known probability model for which the Horton and Tokunaga self-similarity was rigorously established, and whose Horton-Strahler ordering has received attention in the literature [146, 147, 88, 18, 43, 131, 130, 165, 171, 34]. The model has been particularly popular in hydrology as an approximation to the combinatorial structure of the observed river networks [154]. Scott Peckham [131] has first explicitly noticed, by performing a high-precision extraction of river channels for Kentucky River, Kentucky and Powder River, Wyoming, that the Horton exponents and Tokunaga parameters for the observed rivers significantly deviate from that for the Galton-Watson model. He reported values $R \approx 4.6$ and $(a, c) \approx (1.2, 2.5)$ and emphasized the importance of studying a broad range of Horton exponents and Tokunaga parameters. The general interest to fractals and self-similar structures in natural sciences during the 1990s resulted in a quest, mainly inspired and led by Donald Turcotte, for Tokunaga self-similar trees of diverse origin [69, 126, 129, 133, 156, 158, 159, 168, 178]. As a result, the Horton and Tokunaga self-similarity, with a broad range of respective parameters, have been empirically or rigorously established in numerous observed and modeled systems, well beyond river networks.

4.4.2. Computer science

The Horton-Strahler orders are known in computer science as the *register function* or *register number*. They first appeared in the 1958 paper by Ershov [59]

as the minimal number of memory registers required for evaluating a binary arithmetic expression.

A study of Flajolet et al. [66] concerns calculating the average register function in a random plane planted binary tree with n leaves. That is, let the random tree T be uniformly sampled from all C_{n-1} trees in the subspace $\mathcal{BT}^{|}_{\text{plane}}(n)$ of $\mathcal{BT}^{|}_{\text{plane}}$ defined in (51), where C_n is the Catalan number (54). Following Rem. 9, we know that the combinatorial shape $\textsc{shape}(T) \in \mathcal{BT}^{|}$ of such binary tree T can also be obtained by sampling from the Galton-Watson distribution $\mathcal{GW}\left(\frac{1}{2}, \frac{1}{2}\right)$ conditioned on $\#T = 2n - 1$. The work [66] finds the average register function (Horton-Strahler order) in a random binary tree $T \stackrel{d}{\sim} \mathsf{Unif}\big(\mathcal{BT}^{|}_{\text{plane}}(n)\big)$,

$$\mathsf{E}\big[\mathsf{ord}(T)\big] = 1 + \frac{1}{C_n} \sum_{j=1}^{n-1} v_2(j) \left[\binom{2n}{n+j+1} - 2\binom{2n}{n+j} + \binom{2n}{n+j-1} \right],$$

where $v_2(n)$ is known as the *dyadic valuation* of n. Specifically, the dyadic valuation of n is the cardinality of the inverse image of

$$f(p, k) = k2^p \ : \ \mathbb{Z}_+ \times \mathbb{N} \to \mathbb{N},$$

i.e., $v_2(n) = \big|\{(p, k) \in \mathbb{Z}_+ \times \mathbb{N} \ : \ k2^p = n\}\big|$.

In addition, Flajolet et al. [66] proved that for $T \stackrel{d}{\sim} \mathsf{Unif}\big(\mathcal{BT}^{|}_{\text{plane}}(n)\big)$,

$$\mathsf{E}\big[\mathsf{ord}(T)\big] = \log_4 n + D\big(\log_4 n\big) + o(1), \quad \text{as} \quad n \to \infty, \tag{63}$$

where $D(\cdot)$ is a particular continuous periodic function of period one, explicitly derived in [66]. We illustrate Eq. (63) below in Fig. 52(a), which closely reproduces Fig. 6 from the original paper by Flajolet et al. [66]. Equation (63) is related to the tree size asymptotic (35) of Thm. 1, with the Horton exponent $R = 4$.

For more on register functions see [67, 136, 122, 48, 79] and references therein.

5. Critical binary Galton-Watson tree

The critical binary Galton-Watson tree is pivotal for the theory of random trees and for diverse applications because of its transparent generation process and multiple symmetries. This section summarizes some properties of this tree used in our further discussion.

5.1. Combinatorial case

Here we discuss combinatorial binary Galton-Watson trees. Observe that the Galton-Watson tree measures satisfy the coordination property (Def. 10) by construction, and hence the mean Horton self-similarity (Defs. 12, 14) for these measures is equivalent to mean prune-invariance (Def. 13).

5.1.1. Self-similarity and attraction properties

Burd, Waymire, and Winn [34] have first recognized a special position held by the critical binary tree with respect to the Horton pruning in the space of Galton-Watson distributions $\mathcal{GW}(\{q_k\})$ on $\mathcal{T}^|$. The following theorem provides a synopsis of the results in [34], revised and extended in [101].

Theorem 4 (**Self-similarity of critical binary Galton-Watson tree, [34, 101]**)**.** *Consider a Galton-Watson measure $\mathcal{GW}(\{q_k\})$ with $q_1 = 0$. Assume criticality or subcriticality, i.e., $\sum_{k=0}^{\infty} kq_k \leq 1$, and suppose*

$$\sum_{k=2}^{\infty} k^{2-\epsilon} q_k < \infty \quad \forall \ \epsilon > 0.$$

Then, the following statements are equivalent:

- *(a) A distribution is Horton self-similar (Def. 8);*
- *(b) A distribution is mean self-similar (Defs. 12, 14);*
- *(c) A distribution is critical binary: $q_0 = q_2 = 1/2$.*

Furthermore, the critical binary distribution has Tokunaga sequence $T_j = 2^{j-1}$, $j \geq 1$, which corresponds to Tokunaga self-similarity with $(a, c) = (1, 2)$ and strong Horton law (Def. 16) with exponent $R = 4$.

Burd et al. [34, Theorem 3.11] have shown, furthermore, that any critical Galton-Watson tree with bounded offspring number (i.e., $\exists b$ such that $q_k = 0$ for all $k \geq b$) converges to the critical binary Galton-Watson tree under iterative Horton pruning, conditioned on surviving the pruning. In [101], the bounded offspring condition was replaced with finite "$2-$" moment assumption.

Theorem 5 (**Attraction property of critical binary Galton-Watson tree, [34, 101]**)**.** *Consider a critical Galton-Watson process $\rho_0 \equiv \mathcal{GW}(\{q_k\})$ with $q_1 = 0$, such that the "$2-$" moment assumption is satisfied, i.e.,*

$$\sum_{k=2}^{\infty} k^{2-\epsilon} q_k < \infty \qquad \forall \epsilon > 0.$$

Starting with $k = 0$, and for each consecutive integer, let $\nu_k = \mathcal{R}_*(\rho_k)$ denote the pushforward probability measure induced by the pruning operator, i.e.,

$$\nu_k(T) = \rho_k \circ \mathcal{R}^{-1}(T) = \rho_k\big(\mathcal{R}^{-1}(T)\big),$$

and set $\rho_{k+1}(T) = \nu_k (T \mid T \neq \phi)$. Then, for any $T \in \mathcal{T}^|$,

$$\lim_{k \to \infty} \rho_k(T) = \rho^*(T),$$

where ρ^* is the critical binary Galton-Watson measure $\mathcal{GW}\left(\frac{1}{2}, \frac{1}{2}\right)$ on $\mathcal{T}^|$ with the support on $\mathcal{BT}^|$.

The following result of Burd et al. [34] describes the evolution of a binary Galton-Watson tree under iterative Horton pruning.

Theorem 6 (**Dynamics of branching** [34, Proposition 2.1]). *Consider a critical or subcritical combinatorial binary Galton-Watson probability measure $\rho_0 = \mathcal{GW}(q_0, q_2)$ on $\mathcal{BT}^|$, i.e., require $q_0 + q_2 = 1$ and $q_2 \leq 1/2$. Construct a recursion by repeatedly applying Horton pruning operation $\mathcal{R}$ as follows. Starting with $k = 0$, and for each consecutive integer, let $\nu_k = \mathcal{R}_*(\rho_k)$ be the pushforward probability measure induced by the pruning operator, i.e.,*

$$\nu_k(T) = \rho_k \circ \mathcal{R}^{-1}(T) = \rho_k\big(\mathcal{R}^{-1}(T)\big),$$

and set

$$\rho_{k+1}(T) = \nu_k\left(T \mid T \neq \phi\right).$$

Then for each $k \geq 0$, distribution $\rho_k(T)$ is a binary Galton-Watson distribution $\mathcal{GW}(q_0^{(k)}, q_2^{(k)})$ with $q_0^{(k)}$ and $q_2^{(k)}$ constructed recursively as follows: start with $q_0^{(0)} = q_0$ and $q_2^{(0)} = q_2$, and let

$$q_2^{(k+1)} = \frac{\left[q_2^{(k)}\right]^2}{\left[q_0^{(k)}\right]^2 + \left[q_2^{(k)}\right]^2}, \quad q_0^{(k+1)} = 1 - q_2^{(k+1)}. \tag{64}$$

Theorem 6 provides an alternative proof (besides Thm. 4) of why a combinatorial binary Galton-Watson probability distribution $\mathcal{GW}(q_0, q_2)$ is Horton prune-invariant (Def. 8) if and only if $q_0 = q_2 = \frac{1}{2}$. Remark 18 in Sec. 8.6 ties Thm. 6 to the analysis of an excursion of an exponential random walk.

The following statement (Prop. 6) that arises from the results in [34] is useful for applications as it provides a simple way for a computer simulation of a critical binary Galton-Watson tree $\mathcal{GW}\left(\frac{1}{2}, \frac{1}{2}\right)$ of an arbitrary Horton-Strahler order – see related Prop. 8 and Thm. 15 in Sect. 7 for further details.

In Prop. 6 below and throughout the manuscript, we let $\mathsf{Geom}_0(p)$ denote the geometric distribution with p.m.f. $p(m) = p\,(1-p)^m$ for $m \in \mathbb{Z}_+$, and $\mathsf{Geom}_1(p)$ denote the geometric distribution with p.m.f. $p(m) = p\,(1-p)^{m-1}$ for $m \in \mathbb{N}$; see Appendix D.

Proposition 6. *Suppose $T \overset{d}{\sim} \mathcal{GW}\left(\frac{1}{2}, \frac{1}{2}\right)$. Then, the tree order $\mathsf{ord}(T)$ has geometric distribution:*

$$\mathsf{ord}(T) \overset{d}{\sim} \mathsf{Geom}_1(1/2).$$

Fix $j \geq 2$ and $K \geq j$, and condition on the event $\mathsf{ord}(T) = K$. Let b be a branch of order j in T selected uniformly and randomly among all branches of order j in T, and let $N[b] \geq 0$ be the total number of side branches that merge with branch b. Then, $N[b]$ is geometrically distributed:

$$N[b] \overset{d}{\sim} \mathsf{Geom}_0(S_{j-1}^{-1}) = \mathsf{Geom}_0(2^{1-j}),$$

where $S_k = 1 + T_1 + \cdots + T_k = 2^{k-1}$, $k \geq 1$. *Furthermore, conditioned on the value of $N[b]$, each side branch that merges with branch b is assigned an order with probabilities*

$$\mathsf{P}(\textit{side branch has order } i) = \frac{T_{j-i}}{S_{j-1} - 1} = \frac{2^{j-i-1}}{2^{j-1} - 1}, \quad \textit{for } i = 1, \ldots, j-1,$$

independently of other side branches.

5.1.2. The Central Limit Theorem and strong Horton law

This section reviews limit results for the branch counts in a critical binary Galton-Watson tree.

For a given $N \in \mathbb{N}$, consider $T \overset{d}{\sim} \mathsf{Unif}\big(\mathcal{BT}^{|}_{\mathrm{plane}}(N)\big)$. Following Remark 9, we know that $\mathrm{SHAPE}(T) \overset{d}{\sim} \big(\mathcal{GW}\left(\tfrac{1}{2}, \tfrac{1}{2}\right) \big| \#T = 2N - 1\big)$. The branch counts

$$N_j^{(N)}[T] := \Big(N_j[T]; \; T \overset{d}{\sim} \mathsf{Unif}\big(\mathcal{BT}^{|}_{\mathrm{plane}}(N)\big)\Big)$$

are integer valued random variables induced by T. They are the same for T and $\mathrm{SHAPE}(T)$, i.e., $N_j^{(N)}[\mathrm{SHAPE}(T)] = N_j^{(N)}[T]$. The following Law of Large Numbers was proved in Wang and Waymire [165, Thm. 2.1].

Theorem 7 (**LLN for order two branches, [165]**). *For a random tree $T \overset{d}{\sim}$ $\mathsf{Unif}\big(\mathcal{BT}^{|}_{\mathrm{plane}}(N)\big)$,*

$$\frac{N_2^{(N)}[T]}{N} \overset{p}{\to} 4^{-1} \quad \textit{as } N \to \infty. \tag{65}$$

Recall that we know from Thm. 9 that the critical binary Galton-Watson tree is invariant under the Horton pruning operation $\mathcal{R}$. Thus, the *strong Horton law* for branch numbers is deduced from Thm. 7 as follows.

Corollary 2 (**The strong Horton law for branch counts**). *For a random tree $T \overset{d}{\sim} \mathsf{Unif}\big(\mathcal{BT}^{|}_{\mathrm{plane}}(N)\big)$ and for all $j \in \mathbb{N}$,*

$$\frac{N_j^{(N)}[T]}{N} \overset{p}{\to} 4^{-(j-1)} \quad \textit{as } N \to \infty. \tag{66}$$

Proof. For a fixed integer $k > 1$ and a tree $T^{\mathrm{GW}} \overset{d}{\sim} \mathcal{GW}\left(\tfrac{1}{2}, \tfrac{1}{2}\right)$, we have for any positive integers N and $M \leq 2^{-(k-1)}N$,

$$\Big(\mathcal{R}^{k-1}\big(T^{\mathrm{GW}}\big) \,\Big|\, N_1^{(N)}[T^{\mathrm{GW}}] = N, \; N_k^{(N)}[T^{\mathrm{GW}}] = M\Big) \tag{67}$$

$$\overset{d}{=} \Big(\mathcal{R}^{k-1}\big(T^{\mathrm{GW}}\big) \,\Big|\, N_k^{(N)}[T^{\mathrm{GW}}] = M\Big) \overset{d}{=} \Big(T^{\mathrm{GW}} \,\Big|\, N_1^{(N)}[T^{\mathrm{GW}}] = M\Big)$$

as $\mathcal{R}^{k-1}(T^{\mathrm{GW}}) \stackrel{d}{=} T^{\mathrm{GW}}$ by the Horton prune-invariance Thm. 9 (and a more general statement in Thm. 26 of Sect. 10.4). The first equality in (67) can be easily verified from permutability of attachments of smaller order branches to the larger order branches. Specifically, the event $N_k^{(N)}[T^{\mathrm{GW}}] = M$ is equivalent to the event that the pruned tree $\mathcal{R}^{k-1}(T^{\mathrm{GW}})$ will have $\#\mathcal{R}^{k-1}(T^{\mathrm{GW}}) = 2M - 1$ edges. Thus, conditioned of the combinatorial shape $\mathcal{R}^{k-1}(T^{\mathrm{GW}})$, all complete subtrees T_v (see Def. 5(6)) of T such that $\mathsf{ord}(T_v) = \mathsf{ord}(v) < k$ and $\mathsf{ord}(\mathsf{parent}(v)) \geq k$ will be attached to the edges and leaves of $\mathcal{R}^{k-1}(T^{\mathrm{GW}})$ in the same number of ways, for each $\mathcal{R}^{k-1}(T^{\mathrm{GW}})$ satisfying $\#\mathcal{R}^{k-1}(T^{\mathrm{GW}}) = 2M - 1$ edges.

Thus, for a fixed $k \in \mathbb{N}$ and a random tree

$$T \stackrel{d}{\sim} \mathsf{Unif}\big(\mathcal{BT}_{\mathrm{plane}}^{|}(N)\big),$$

we have by (67),

$$\left(\mathcal{R}^{k-1}(T)\Big| N_k^{(N)}[T] = M\right) \stackrel{d}{\sim} \mathsf{Unif}\big(\mathcal{BT}_{\mathrm{plane}}^{|}(M)\big)$$

for all $M \leq 2^{-(k-1)}N$. Hence, Thm. 7 implies

$$\left(\frac{N_k^{(N)}[T]}{N_{k-1}^{(N)}[T]} \,\Big|\, \mathsf{ord}(T) \geq k\right) = \left(\frac{N_2^{(N)}\big[\mathcal{R}^{k-1}(T)\big]}{N_1^{(N)}\big[\mathcal{R}^{k-1}(T)\big]} \,\Big|\, \mathsf{ord}(T) \geq k\right) \stackrel{p}{\to} 4^{-1} \text{ as } N \to \infty.$$

Next, we let $\frac{0}{0} = 0$ as here $N_k^{(N)}[T] \leq N_{k-1}^{(N)}[T]$, and

$$N_{k-1}^{(N)}[T] = 0 \text{ implies } N_k^{(N)}[T] = 0.$$

Then, as $\lim\limits_{N \to \infty,} \mathsf{P}\big(\mathsf{ord}(T) < k\big) = 0$ we have

$$\frac{N_k^{(N)}[T]}{N_{k-1}^{(N)}[T]} \stackrel{p}{\to} 4^{-1} \quad \text{as } N \to \infty. \tag{68}$$

Finally, iterating (68), we obtain

$$\frac{N_j^{(N)}[T]}{N} = \frac{N_j^{(N)}[T]}{N_{j-1}^{(N)}[T]} \frac{N_{j-1}^{(N)}[T]}{N_{j-2}^{(N)}[T]} \cdots \frac{N_2^{(N)}[T]}{N} \stackrel{p}{\to} 4^{-(j-1)} \quad \text{as } N \to \infty. \qquad \square$$

Following Thm. 7, the corresponding Central Limit Theorem was proved in Wang and Waymire [165, Thm. 2.4].

Theorem 8 (**CLT for order two branches, [165]**). *For a random tree $T \stackrel{d}{\sim}$* $\mathsf{Unif}\big(\mathcal{BT}_{\mathrm{plane}}^{|}(N)\big)$,

$$\sqrt{N}\left(\frac{N_2^{(N)}[T]}{N} - \frac{1}{4}\right) \stackrel{d}{\to} N(0, 4^{-2}) \quad \text{as } N \to \infty. \tag{69}$$

Next, using the pruning framework, the following Central Limit Theorem for $N_j^{(N)}[T]$ is readily obtained as a direct consequence of the original Thm. 8 of Wang and Waymire [165] and the Horton prune-invariance (Def. 8) of $\mathcal{GW}\left(\frac{1}{2}, \frac{1}{2}\right)$ as stated in Thm. 9 (and, alternatively, in Thm. 6), and a more general statement that will appear in Thm. 26 of Sect. 10.4.

Corollary 3 (**CLT for branch numbers, [169]**). *For a random tree $T \overset{d}{\sim}$* $\mathsf{Unif}\big(\mathcal{BT}_{\mathrm{plane}}^{|}(N)\big)$,

$$\sqrt{N}\left(\frac{N_{j+1}^{(N)}[T]}{N_j^{(N)}[T]} - \frac{1}{4}\right) \overset{d}{\to} N(0, 4^{r-3}) \quad \text{as } N \to \infty, \tag{70}$$

where we set $\frac{0}{0} = 0$.

Proof. Pruning $T \overset{d}{\sim} \mathsf{Unif}\big(\mathcal{BT}_{\mathrm{plane}}^{|}(N)\big)$ iteratively $j-1$ times, we obtain $T \overset{d}{\sim}$ $\mathsf{Unif}\Big(\mathcal{BT}_{\mathrm{plane}}^{|}\big(N_j^{(N)}[T]\big)\Big)$, where for the case when $j > \mathsf{ord}(T)$ and $N_j^{(N)}[T] = 0$, we set $\mathcal{BT}_{\mathrm{plane}}^{|}(0) := \{\phi\}$. Hence, Theorem 8 immediately implies

$$\sqrt{N_j^{(N)}[T]}\left(\frac{N_{j+1}^{(N)}[T]}{N_j^{(N)}[T]} - \frac{1}{4}\right) \overset{d}{\to} N(0, 4^{-2}) \quad \text{as } N \to \infty. \tag{71}$$

Thus, substituting (66) into (71), we obtain (70). □

The limit (70) was derived by Yamamoto [169] directly, after a series of technically involved calculations.

5.2. Metric case: exponential edge lengths

In this section we turn to the trees in $\mathcal{BL}_{\mathrm{plane}}^{|}$. Specifically, we assign i.i.d. exponential lengths to the edges of a critical plane binary Galton-Watson tree $\mathcal{GW}_{\mathrm{plane}}(\frac{1}{2}, \frac{1}{2})$ in $\mathcal{T}^{|}$ and introduce random ordering between the sibling offspring, thus obtaining the *exponential critical binary Galton-Watson tree.*

Definition 20 (**Exponential critical binary Galton-Watson tree**). *We say that a random tree $T \in \mathcal{BL}_{\mathrm{plane}}^{|}$ is an exponential critical binary Galton-Watson tree with (edge length) parameter $\lambda > 0$, and write $T \overset{d}{\sim} \mathsf{GW}(\lambda)$, if the following conditions are satisfied:*

(i) P-SHAPE(T) is a critical plane binary Galton-Watson tree $\mathcal{GW}_{\mathrm{plane}}(\frac{1}{2}, \frac{1}{2})$;
(ii) conditioned on a given P-SHAPE(T), the edges of T are sampled as independent $\mathsf{Exp}(\lambda)$ random variables, i.e., random variables with probability density function (p.d.f.)

$$\phi_\lambda(x) = \lambda e^{-\lambda x} \mathbf{1}_{\{x \geq 0\}}. \tag{72}$$

The branching process that generates an exponential critical binary Galton-Watson tree is known as the continuous time Galton-Watson process, and is sometimes simply called *Markov branching process* [81].

Remark 10. The distribution $\mathsf{GW}(\lambda)$ is defined here on the space of planar trees $\mathcal{BL}^|_{\text{plane}}$. The planar embedding (ordering of sibling offspring) is needed when working with tree representation of time series in Sect. 8. In other analyses of this work, one can drop the embedding and work with trees from $\mathcal{BL}^|$. All results presented below are valid for both versions of the definition (with and without embedding). We observe though that embedding is useful for counting distinct trees (e.g. Lem. 8).

5.2.1. Self-similarity

The self-similar properties of the exponential critical binary Galton-Watson tree $T \stackrel{d}{\sim} \mathsf{GW}(\lambda)$ are summarized in the following statement.

Proposition 7 (**Self-similarity of exponential critical binary Galton-Watson tree**). *Suppose $T \stackrel{d}{\sim} \mathsf{GW}(\lambda)$. Then T is distributionally Horton self-similar (Def. 9) with scaling exponent $\zeta = 2$. Furthermore, it has Tokunaga sequence $T_j = 2^{j-1}$, $j \geq 1$, which corresponds to Tokunaga self-similarity with $(a, c) = (1, 2)$ and strong Horton law (Def. 16) with exponent $R = 4$.*

Proof. We show here three alternative proofs to emphasize the pivotal role of the distribution $\mathsf{GW}(\lambda)$ in the developed theory.

Observe that the combinatorial Horton self-similarity and the values of a, c and R are found in Thm. 4. Hence, we only need to establish prune-invarience in lengths (Def. 9, part ii) and find the scaling exponent ζ.

First, one can use results on the hierarchical branching process in Sect. 7. Theorem 15 in Sect. 7.5 shows that the exponential critical binary Galton-Watson tree is a particular case of the critical Tokunaga process (Def. 25). Proposition 11 establishes distributional Horton self-similarity for the critical Tokunaga processes (and hence for the exponential critical binary Galton-Watson tree), and finds the respective parameters ζ, a, c, and R.

Second, one can use the results of Sect. 8 on tree representation of time series. The Horton prune-invariance of the exponential critical binary Galton-Watson tree with $\zeta = 2$ is established in Cor. 12 of Sect. 8.6 via tree representation of exponential random walks.

Third, one can use the framework of generalized dynamical pruning of Sect. 10. Theorem 26 establishes invariance of the exponential critical binary Galton-Watson tree with respect to any admissible generalized pruning. Horton pruning is a special case of the generalized dynamical pruning (Ex. 10.1.2), which ensures the invariance of the present theorem. Next, Thm. 27, part c, states the scaling exponent $\zeta = 2$ (which is found, as above, via Cor. 12). $\qquad\square$

5.2.2. Tree length

Recall the modified Bessel functions of the first kind

$$I_\nu(z) = \sum_{n=0}^{\infty} \frac{\left(\frac{z}{2}\right)^{2n+\nu}}{\Gamma(n+1+\nu)\,n!}.$$

Lemma 8. *Suppose* $T \overset{d}{\sim} \mathsf{GW}(\lambda)$ *is an exponential critical binary Galton-Watson tree with parameter* λ. *The total length* LENGTH(T) *of the tree* T *has the p.d.f.*

$$\ell(x) = \frac{1}{x}e^{-\lambda x}I_1\big(\lambda x\big), \quad x > 0. \tag{73}$$

Proof. Recall that the number of different combinatorial shapes of a planted plane binary tree with $n+1$ leaves, and therefore $2n+1$ edges, is given by the *Catalan number* (54), i.e.,

$$\big|\mathcal{BT}^|_{\text{plane}}(n+1)\big| = C_n = \frac{1}{n+1}\binom{2n}{n} = \frac{(2n)!}{(n+1)!n!}.$$

The total length of $2n+1$ edges is a gamma random variable with parameters λ and $2n+1$ and density function

$$\gamma_{\lambda,2n+1}(x) = \frac{\lambda^{2n+1}x^{2n}e^{-\lambda x}}{\Gamma(2n+1)}, \quad x > 0.$$

Hence, the total length of the tree T has the p.d.f.

$$\ell(x) = \sum_{n=0}^{\infty} \frac{C_n}{2^{2n+1}} \cdot \frac{\lambda^{2n+1}x^{2n}e^{-\lambda x}}{(2n)!} = \sum_{n=0}^{\infty} \frac{\lambda^{2n+1}x^{2n}e^{-\lambda x}}{2^{2n+1}(n+1)!n!}$$

$$= \frac{1}{x}e^{-\lambda x}\sum_{n=0}^{\infty} \frac{\left(\frac{\lambda x}{2}\right)^{2n+1}}{\Gamma(n+2)\,n!} = \frac{1}{x}e^{-\lambda x}I_1\big(\lambda x\big). \tag{74}$$

$\square$

Next, we compute the Laplace transform of $\ell(x)$. By the summation formula in (74),

$$\mathcal{L}\ell(s) = \int_0^{\infty} \sum_{n=0}^{\infty} \frac{C_n}{2^{2n+1}} \cdot \frac{\lambda^{2n+1}x^{2n}e^{-(\lambda+s)x}}{(2n)!}\,dx$$

$$= \sum_{n=0}^{\infty} \frac{C_n}{2^{2n+1}} \cdot \left(\frac{\lambda}{\lambda+s}\right)^{2n+1} \int_0^{\infty} \frac{(\lambda+s)^{2n+1}x^{2n}e^{-(\lambda+s)x}}{(2n)!}\,dx$$

$$= \sum_{n=0}^{\infty} \frac{C_n}{2^{2n+1}} \cdot \left(\frac{\lambda}{\lambda+s}\right)^{2n+1} = Z \cdot c(Z^2),$$

where we let $Z = \frac{\lambda}{2(\lambda+s)}$, and the characteristic function of Catalan numbers

$$c(z) = \sum_{n=0}^{\infty} C_n z^n = \frac{2}{1 + \sqrt{1 - 4z}} \tag{75}$$

is well known. Therefore

$$\mathcal{L}\ell(s) = Z \cdot c(Z^2) = \frac{\lambda}{\lambda + s + \sqrt{(\lambda+s)^2 - \lambda^2}}. \tag{76}$$

Note that the Laplace transform $\mathcal{L}\ell(s)$ could be derived from the total probability formula

$$\ell(x) = \frac{1}{2}\phi_\lambda(x) + \frac{1}{2}\phi_\lambda * \ell * \ell(x), \tag{77}$$

where $\phi_\lambda(x)$ is the exponential p.d.f. (72). Thus, $\mathcal{L}\ell(s)$ solves

$$\mathcal{L}\ell(s) = \frac{1}{2}\frac{\lambda}{\lambda + s}\Big(1 + \big(\mathcal{L}\ell(s)\big)^2\Big). \tag{78}$$

5.2.3. Tree height

Lemma 9 ([100]). *Suppose $T \overset{d}{\sim} \mathsf{GW}(\lambda)$ is an exponential critical binary Galton-Watson tree with parameter λ. Then, the height* HEIGHT(T) *of the tree T has the cumulative distribution function*

$$\mathsf{H}(x) = \frac{\lambda x}{\lambda x + 2}, \quad x > 0. \tag{79}$$

Proof. The proof is based on duality between trees and positive real excursions that we introduce in Sect. 8. In particular, Thm. 20 establishes equivalence between the level set tree (Sect. 8.2) of a positive excursion of an *exponential random walk* (Sect. 8.6) and an exponential critical binary Galton-Watson tree $\mathsf{GW}(\lambda)$. This implies, in particular, that for a tree $T \overset{d}{\sim} \mathsf{GW}(\lambda)$ the HEIGHT(T) has the same distribution as the height of a positive excursion of an exponential random walk Y_k with $Y_0 = 0$ and independent increments $Y_{k+1} - Y_k$ distributed according to the Laplace density function $\frac{\phi_\lambda(x)+\phi_\lambda(-x)}{2} = \frac{\lambda}{2}e^{-\lambda|x|}$, with $\phi_\lambda(x)$ defined in (72).

Notice that Y_k is a martingale. We condition on $Y_1 > 0$, and consider an excursion $Y_0, Y_1, \ldots, Y_{\tau_-}$ with $\tau_- = \min\{k > 1 \ : \ Y_k \leq 0\}$ denoting the termination step of the excursion. For $x > 0$, we write

$$p_x = 1 - \mathsf{H}(x) = \mathsf{P}\left(\max_{j \,:\, 0 < j < \tau_-} Y_j > x \;\Big|\; Y_1 > 0\right)$$

for the probability that the height of the excursion exceeds x. The problem of finding p_x is solved using the Optional Stopping Theorem. Let

$$\tau_x = \min\{k > 0 \ : \ Y_k \geq x\} \qquad \text{and} \qquad \tau := \tau_x \wedge \tau_-.$$

Observe that

$$p_x = \mathsf{P}(\tau = \tau_x \mid Y_1 > 0).$$

For a fixed $y \in (0, x)$, by the Optional Stopping Theorem, we have

$$
\begin{aligned}
y \;\; &= \;\; \mathsf{E}[Y_\tau \mid Y_1 = y] \\[2mm]
&= \;\; \mathsf{E}[Y_\tau \mid \tau = \tau_-, Y_1 = y]\, \mathsf{P}(\tau = \tau_- \mid Y_1 = y) \\
&\qquad + \mathsf{E}[Y_\tau \mid \tau = \tau_x, Y_1 = y]\, \mathsf{P}(\tau = \tau_x \mid Y_1 = y) \\[2mm]
&= \;\; \mathsf{E}[Y_\tau \mid Y_\tau \le 0, Y_1 = y]\mathsf{P}\,(\tau = \tau_- \mid Y_1 = y) \\
&\qquad + \mathsf{E}[Y_\tau \mid Y_\tau \ge x, Y_1 = y]\, \mathsf{P}(\tau = \tau_x \mid Y_1 = y) \\[2mm]
&= \;\; -\frac{1}{\lambda}\mathsf{P}(\tau = \tau_- \mid Y_1 = y) + \left(x + \frac{1}{\lambda}\right)\mathsf{P}(\tau = \tau_x \mid Y_1 = y) \\[2mm]
&= \;\; \left(x + \frac{2}{\lambda}\right)\mathsf{P}(\tau = \tau_x \mid Y_1 = y) - \frac{1}{\lambda}.
\end{aligned}
$$

Hence,

$$\mathsf{P}(\tau = \tau_x \mid Y_1 = y) = \frac{y + \frac{1}{\lambda}}{x + \frac{2}{\lambda}}.$$

Thus,

$$
\begin{aligned}
\mathsf{P}\!\left(\tau = \tau_x, 0 < Y_1 < x \mid Y_1 > 0\right) \;\; &= \;\; \int_0^x \mathsf{P}(\tau = \tau_x \mid Y_1 = y)\, \lambda e^{-\lambda y} dy \\[2mm]
&= \;\; \int_0^x \frac{y + \frac{1}{\lambda}}{x + \frac{2}{\lambda}}\, \lambda e^{-\lambda y} dy \\[2mm]
&= \;\; \frac{2}{\lambda x + 2} - e^{-\lambda x},
\end{aligned}
$$

and therefore,

$$
\begin{aligned}
p_x \;\; &= \;\; \mathsf{P}\!\left(\max_{j:\ 0 < j < K} Y_j > x \mid Y_1 > 0\right) \\[2mm]
&= \;\; \mathsf{P}\!\left(\tau = \tau_x, 0 < Y_1 < x \mid Y_1 > 0\right) + \mathsf{P}\!\left(\tau = \tau_x, Y_1 \ge x \mid Y_1 > 0\right) \\[2mm]
&= \;\; \frac{2}{\lambda x + 2} - e^{-\lambda x} + \mathsf{P}\!\left(Y_1 \ge x \mid Y_1 > 0\right) = \frac{2}{\lambda x + 2}.
\end{aligned}
$$

Hence,

$$\mathsf{H}(x) = 1 - p_x = \frac{\lambda x}{\lambda x + 2}. \qquad\qquad \square$$

We continue examining the height function $\textsc{height}(T)$ for $T \overset{d}{\sim} \mathsf{GW}(\lambda)$. This time, we condition on $\#T = 2n - 1$, i.e., the tree T has n leaves and $n - 1$ internal non-root vertices. We let $\mathsf{H}_n(x)$ denote the corresponding conditional cumulative distribution function,

$$\mathsf{H}_n(x) = \mathsf{P}\big(\textsc{height}(T) \leq x \mid \#T = 2n - 1\big). \tag{80}$$

There, for a one-leaf tree,

$$\mathsf{H}_1(x) = 1 - e^{-\lambda x}, \tag{81}$$

and for $n \geq 2$, the following recursion follows from conditioning on the length of the stem (root edge),

$$\mathsf{H}_n(a) = \sum_{k=1}^{n-1} \frac{C_{k-1}C_{n-k-1}}{C_{n-1}} \int_0^a \mathsf{H}_k(a - x)\mathsf{H}_{n-k}(a - x)\,\lambda e^{-\lambda x}\,dx, \tag{82}$$

where C_n is the Catalan number as defined in (54).

Next, we consider the following z-transform:

$$\mathsf{h}(a; z) = \sum_{n=1}^{\infty} \mathsf{H}_n(a)\,C_{n-1}\,z^n \quad \text{for} \ \ |z| < 1/4. \tag{83}$$

Then, (81) and (82) imply

$$\mathsf{h}(a; z) = (1 - e^{-\lambda a})z + \int_0^a \mathsf{h}^2(a - x; z)\,\lambda e^{-\lambda x}\,dx$$

which, if we let $y = a - x$, simplifies to

$$e^{\lambda a}\mathsf{h}(a; z) - e^{\lambda a}z = \int_0^a \mathsf{h}^2(y; z)\,\lambda e^{\lambda y}\,dy\ - z.$$

We differentiate the above equation, obtaining

$$\frac{\partial}{\partial a}\mathsf{h}(a; z) = \lambda\left(\mathsf{h}^2(a; z) - \mathsf{h}(a; z) + z\right). \tag{84}$$

Let

$$\mathsf{x}_1(z) = \frac{1 + \sqrt{1 - 4z}}{2} \quad \text{and} \quad \mathsf{x}_2(z) = \frac{1 - \sqrt{1 - 4z}}{2}$$

be the two roots of $\mathsf{x}^2 - \mathsf{x} + z = 0$. Here, $\mathsf{x}_2(z)/z = 1/\mathsf{x}_1(z) = c(z)$ is the z-transform of the Catalan sequence C_n, introduced in (75). Then, (84) solves as

$$\mathsf{h}(a; z) - \mathsf{x}_1(z) = \Phi(z)e^{\lambda a\sqrt{1-4z}}\big(\mathsf{h}(a; z) - \mathsf{x}_2(z)\big),$$

where due to the initial conditions $\mathsf{h}(0; z) = 0$, we have $\Phi(z) = \mathsf{x}_1(z)/\mathsf{x}_2(z)$, and

$$\mathsf{h}(a; z) - \mathsf{x}_1(z) = \frac{\mathsf{x}_1(z)}{\mathsf{x}_2(z)}e^{\lambda a\sqrt{1-4z}}\big(\mathsf{h}(a; z) - \mathsf{x}_2(z)\big). \tag{85}$$

Solution (85) implies

$$h(a; z) = \frac{2 \left(e^{\lambda a \sqrt{1-4z}} - 1\right) z}{e^{\lambda a \sqrt{1-4z}} - 1 + \left(e^{\lambda a \sqrt{1-4z}} + 1\right) \sqrt{1 - 4z}}. \qquad (86)$$

Here and throughout we use $-\pi < \arg(z) \leq \pi$ branch of the logarithm when defining $\sqrt{1 - 4z}$ for $|z| < 1/4$.

Now, since $\mathsf{P}\left(\#T = 2n - 1\right) = 2C_{n-1}4^{-n}$, the series expansion (83) implies

$$\mathsf{H}(a) = \lim_{z \uparrow \frac{1}{4}} 2\,\mathsf{h}(a; z), \qquad (87)$$

where $z \in (-1/4, \, 1/4)$ is real. We substitute (86) into the limit (87),

$$\begin{aligned}
\lim_{z \uparrow \frac{1}{4}} 2\,\mathsf{h}(a; z) &= \lim_{z \uparrow \frac{1}{4}} \frac{4 \left(e^{\lambda a \sqrt{1-4z}} - 1\right) z}{e^{\lambda a \sqrt{1-4z}} - 1 + \left(e^{\lambda a \sqrt{1-4z}} + 1\right) \sqrt{1 - 4z}} \\
&= \lim_{z \uparrow \frac{1}{4}} \frac{4z}{1 + \sqrt{1 - 4z} + \frac{2\sqrt{1-4z}}{e^{\lambda a \sqrt{1-4z}} - 1}} = \frac{1}{1 + \frac{2}{\lambda a}} \\
&= \frac{\lambda a}{\lambda a + 2}, \qquad (88)
\end{aligned}$$

thus obtaining an alternative proof of formula (79) in Lemma 9.

The asymptotic of the height distribution $\mathsf{H}_n(a)$ for a given number of leaves n was the object of analysis in [94, 166, 73, 51]. In particular, Gupta et al. [73] extended the results of Kolchin [94], by showing that

$$\lim_{n \to \infty} \mathsf{H}_n \left(\frac{a\sqrt{n}}{\lambda}\right) = \mathsf{H}_\infty(a) := 1 + 2 \sum_{k=1}^{\infty} (1 - 4k^2 a^2) \exp\left\{-2k^2 a^2\right\}. \qquad (89)$$

It was also observed in [73] that $\mathsf{H}_\infty \left(\frac{a}{2\sqrt{2}}\right)$ is the distribution function for the maximum of the Brownian excursion as shown in the work of Durrett and Iglehart [50]. The results of [73] were further developed in [51] for more general trees with edge lengths.

6. Invariant Galton-Watson trees

This section summarizes the results of [101]. It extends the invariance study of Burd et al. [34] reviewed in Sect. 5.1.1 and gives a complete description of invariants and attractors of the critical and subcritical Galton-Watson tree measures under the operation of Horton pruning (Def. 3).

Definition 21 (**Invariant Galton-Watson measures**)**.** *For a given* $q \in [1/2, 1)$*, a critical Galton-Watson measure* $\mathcal{GW}(\{q_k\})$ *is said to be the invariant*

Galton-Watson (IGW) measure with parameter q and denoted by $\mathcal{IGW}(q)$ if its generating function has the following expression:

$$\sum_{k=0}^{\infty} z^k q_k = z + q(1-z)^{1/q}. \tag{90}$$

The respective branching probabilities are $q_0 = q$, $q_1 = 0$, $q_2 = \frac{1-q}{2q}$, and

$$q_k = \frac{1-q}{k!\,q} \prod_{i=2}^{k-1}(i - 1/q) \quad (k \geq 3). \tag{91}$$

Here, if $q = \frac{1}{2}$, then the distribution is critical binary, i.e., $\mathcal{GW}\left(\frac{1}{2}, \frac{1}{2}\right)$. If $q \in (1/2, 1)$, the distribution is of Zipf type with

$$q_k = \frac{(1-q)\Gamma(k - 1/q)}{q\Gamma(2 - 1/q)\,k!} \sim k^{-(1+q)/q}. \tag{92}$$

The next theorem shows that the IGW measures are the only Galton-Watson measures invariant with respect to Horton pruning, hence justifying their name.

Theorem 9 (Self-similar Galton-Watson measures, [101]). *Consider a critical or subcritical Galton-Watson measure $\mathcal{GW}(\{q_k\})$ with $q_1 = 0$. It is Horton self-similar (Def. 8) if and only if it is the invariant Galton-Watson (IGW) measure $\mathcal{IGW}(q_0)$ with $q_0 \in [1/2, 1)$.*

Remark 11. The Horton prune-invariance is not the only invariance satisfied by the family of measures $\mathcal{IGW}(q)$. For example, Duquesne and Winkel [53] considered a random tree T' obtained by applying a Bernoulli leaf coloring scheme to a Galton-Watson tree T. Specifically, a subset of leaves of T is selected with independent Bernoulli trials and a subtree of T consisting of the union of all lineages of these selected leaves is considered. Next, a new tree T' is constructed by taking a series reduction of this subtree. Then, conditioned on $T' \neq \phi$, the tree T' is distributed as a Galton-Watson tree with the generating function given in [53, Sect. 2.2, Eq. (10)]. Substituting (90) into that equation implies the invariance of $\mathcal{IGW}(q)$ measures under the Bernoulli leaf coloring scheme.

The work [101] shows that the collection of $\mathcal{IGW}(q_0)$ measures for $q_0 \in [1/2, 1)$ and a point measure $\mathcal{GW}(q_0 = 1)$ are the only possible attractors of critical and subcritical Galton-Watson measures under iterative Horton pruning. Specifically, all subcritical measures converge to $\mathcal{GW}(q_0 = 1)$, and critical measures converge to $\mathcal{IGW}(q_0)$. The domain of attraction of $\mathcal{IGW}(q_0)$ for any $q_0 \in [1/2, 1)$ is non empty and is characterized by the tail behavior of the offspring distribution $\{q_k\}$ of the initial Galton-Watson measure. The next result illustrates the attraction property of the IGW measures for a particular class of initial Galton-Watson measures with power law (Zipf type) decay of the offspring distribution $\{q_k\}$.

Theorem 10 (**Attraction property of critical Galton-Watson trees,** [**101**]). *Consider a critical Galton-Watson process $\rho_0 \equiv \mathcal{GW}(\{q_k\})$ with $q_1 = 0$ and probability mass function q_k of Zipf type, i.e., $q_k \sim k^{-(\alpha+1)}$, with $\alpha \in (1,2]$. Starting with $k = 0$, and for each consecutive integer, let $\nu_k = \mathcal{R}_*(\rho_k)$ denote the pushforward probability measure induced by the pruning operator, i.e.,*

$$\nu_k(T) = \rho_k \circ \mathcal{R}^{-1}(T) = \rho_k\big(\mathcal{R}^{-1}(T)\big),$$

and set $\rho_{k+1}(T) = \nu_k\left(T \,|T \neq \phi\right)$. Then, for any $T \in \mathcal{T}^|$,

$$\lim_{k \to \infty} \rho_k(T) = \rho^*(T),$$

where ρ^* is the invariant Galton-Watson measure $\mathcal{IGW}\left(\frac{1}{\alpha}\right)$.

Finally, if the Galton-Watson measure is subcritical, then $\rho_k(T)$ converges to a point mass measure, $\mathcal{GW}(q_0 = 1)$.

Finally we show that the IGW measures satisfy the strong Horton law.

Lemma 10 (**Strong Horton law for IGW,** [**101**]). *Consider an invariant Galton-Watson measure $\mathcal{IGW}(q_0)$ for $q_0 \in [1/2, 1)$. Then, the strong Horton law* (29) *holds with Horton exponent $R = (1 - q_0)^{-1/q_0}$.*

7. Hierarchical branching process

The critical binary Galton-Watson tree discussed in Sect. 5 is an important although a very particular example of self-similarity. It corresponds to a specific Tokunaga sequence $T_j = 2^{j-1}$, Horton exponent $R = 4$, and self-similarity constant $\zeta = 2$ (Prop. 7).

In this section we construct a multi-type branching process [81, 12] that generates a self-similar tree for an arbitrary Tokunaga sequence $T_j \geq 0$ and any scaling constant $\zeta > 0$ (Thm. 11). This model, called *hierarchical branching process*, includes the exponential critical binary Galton-Watson tree as a special case (Thm. 15). We also introduce a class of critical self-similar *Tokunaga processes* (Sect. 7.5) that enjoy additional symmetries – their edge lengths are i.i.d. random variables (Prop. 11), and subtrees of large Tokunaga trees reproduce the probabilistic structure of the entire random tree space (Prop. 12). The results of this section are derived in [99].

7.1. Definition and main properties

Consider a probability mass function $\{\pi_K\}_{K \geq 1}$, a sequence $\{T_k\}_{k \geq 1}$ of non-negative Tokunaga coefficients, and a sequence $\{\lambda_j\}_{j \geq 1}$ of positive termination rates. We now define a hierarchical branching process $S(t)$.

Definition 22 (**Hierarchical Branching Process (HBP)**). *We say that $S(t)$ is a hierarchical branching process with a triplet of parameter sequences $\{T_k\}$, $\{\lambda_j\}$, and $\{\pi_K\}$, and write*

$$S(t) \overset{d}{\sim} \mathsf{HBP}\big(\{T_k\}, \{\lambda_j\}, \{\pi_K\}\big)$$

if $S(t)$ is a multi-type branching process that develops in continuous time $t > 0$ according to the following rules:

(i) *The process $S(t)$ starts at $t = 0$ with a single progenitor (root branch) whose Horton-Strahler order (type) is $K \geq 1$ with probability π_K.*

(ii) *Every branch of order $j \leq K$ produces offspring (side branches) of every order $i < j$ with rate $\lambda_j T_{j-i}$. Each offspring (side branch) is assigned a uniform random orientation (right or left).*

(iii) *A branch of order j terminates with rate λ_j.*

(iv) *At its termination time, a branch of order $j \geq 2$ splits into two independent branches of order $j - 1$. The two branches are assigned uniform random orientations, i.e., a uniformly randomly selected branch becomes right and the other becomes left.*

(v) *A branch of order $j = 1$ terminates without leaving offspring.*

(vi) *Generation of side branches and termination of distinct branches are independent.*

The definition implies that the process $S(t)$ terminates a.s. in finite time. Accordingly, the branching history of $S(t)$ creates a random binary tree $T[S]$ in the space $\mathcal{BL}^{|}_{\text{plane}}$ of planted binary trees with edge lengths and planar embedding. To avoid heavy notations, we sometimes use the process distribution name $\mathsf{HBP}(\cdot, \cdot, \cdot)$, as well as its various special cases introduced below, to also denote the measures induced by the process on suitable tree spaces ($\mathcal{T}^{|}_{\text{plane}}$, $\mathcal{L}^{|}_{\text{plane}}$ $\mathcal{BL}^{|}_{\text{plane}}$, etc.)

The next statement describes the branching structure of $T[S]$.

Proposition 8 (**Side branching in hierarchical branching process, [99]**).
Consider a hierarchical branching process $S(t) \overset{d}{\sim} \mathsf{HBP}\big(\{T_k\}, \{\lambda_j\}, \{\pi_K\}\big)$ and let $T[S]$ be the tree generated by $S(t)$ in $\mathcal{BL}^{|}_{\text{plane}}$. For a branch $b \subset T[S]$ of order $K \geq 1$, let $N_i[b] \geq 0$ be the number of its side branches of order $i = 1, \ldots, K-1$, and $N[b] = N_1[b] + \cdots + N_{K-1}[b]$ be the total number of the side branches. Conditioned on $N[b]$, let $l_i := l_i(b)$ be the lengths of $N[b] + 1$ edges within b, counted sequentially from the initial vertex, and $l := l(b) = l_1 + \cdots + l_{N[b]+1}$ be the total branch length. Define

$$S_K := 1 + T_1 + \cdots + T_K$$

for $K \geq 0$ by assuming $T_0 = 0$. Then the following statements hold:

1. *The tree order satisfies*

$$\mathsf{P}\left(\mathsf{ord}(T[S]) = K\right) = \pi_K, \quad K \geq 1. \tag{93}$$

2. *The total number $N[b]$ of side branches within a branch b of order K has geometric distribution:*

$$N[b] \overset{d}{\sim} \mathsf{Geom}_0\left(S_{K-1}^{-1}\right), \quad K \geq 1, \tag{94}$$

with $\mathsf{E}[N[b]] = S_{K-1} - 1 = T_1 + \cdots + T_{K-1}$.

3. *Conditioned on the total number m of side branches, the distribution of vector $(N_1[b], \ldots, N_{K-1}[b])$ is multinomial with $N[b]$ trials and success probabilities*

$$\mathsf{P}(side\ branch\ has\ order\ i) = \frac{T_{K-i}}{S_{K-1} - 1}. \tag{95}$$

The vector $(\mathsf{ord}_1, \ldots, \mathsf{ord}_{N[b]})$ of side branch orders, where the side branches are labeled sequentially starting from the initial vertex of b, is obtained from the sequence

$$\mathsf{orders} = \underbrace{(1, \ldots, 1}_{N_1[b]\ times}, \underbrace{2, \ldots, 2}_{N_2[b]\ times}, \ldots \underbrace{K-1, \ldots, K-1)}_{N_{K-1}[b]\ times}$$

by a uniform random permutation $\sigma_{N[b]}$ of indices $\{1, \ldots, N[b]\}$:

$$(\mathsf{ord}_1, \ldots, \mathsf{ord}_{N[b]}) = \mathsf{orders} \circ \sigma_{N[b]}.$$

4. *The total numbers of side branches and orders of side branches are independent in distinct branches.*
5. *The branch length l has exponential distribution with rate λ_K, 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*

$$\lambda_K S_{K-1}. \tag{96}$$

Proof. All the properties readily follow from Def. 22. $\qquad \square$

Combining properties 2 and 3 of Prop. 8 we find that the number $N_i[b]$ of side branches of order i within a branch b of order K has geometric distribution:

$$N_i[b] \stackrel{d}{\sim} \mathsf{Geom}_0\left([1 + T_{K-i}]^{-1}\right), \quad 1 \le i \le K-1, \tag{97}$$

with $\mathsf{E}[N_i[b]] = T_{K-i}$. We also notice that, because of property 3, the numbers $N_i[b]$ for $i = 1, \ldots, K-1$ within the same branch b are dependent.

Proposition 8 provides an alternative definition of the hierarchical branching process and suggests a recursive construction of $T[S]$ that does not require time-dependent simulations. This construction is useful for applied computer simulations. Specifically, a tree of order $K = 1$ consists of two vertices (root and leaf) connected by an edge of exponential length with rate λ_1. Assume now that we know how to construct a random tree of any order below $K \ge 2$. To construct a tree of order K, we start with a perfect (combinatorial) planted binary tree of depth K, which we call *skeleton*. The combinatorial shapes of such trees is illustrated in Fig. 15. All leaves in the skeleton have the same depth K, and all vertices at depth $1 \le \kappa \le K$ have the same Horton-Strahler order $K - \kappa + 1$. The root (at depth 0) has order K. Next, we assign lengths to the branches of the skeleton. Recall (Ex. 1) that each branch in a perfect tree consists of a single edge. To assign length to a branch b of order κ, with $1 \le \kappa \le K$, we generate a geometric number $N[b] \stackrel{d}{\sim} \mathsf{Geom}_0(S_{\kappa-1}^{-1})$ according to (94) and then $N[b] + 1$

i.i.d. exponential lengths l_i, $i = 1, \ldots, N[b] + 1$, with the common rate $\lambda_\kappa S_{\kappa-1}$ according to (96). The total length of the branch b is $l_1 + \cdots + l_{N[b]+1}$. Moreover, branch b has $N[b]$ side branches that are attached along b with spacings l_i, starting from the branch point closest to the root. The order assignment for the side branches is done according to (95). We generate side branches (each has order below K) independently and attach them to the branch b. This completes the construction of a random tree of order K. To construct a random HBP tree, one first generates a random order $K \geq 1$ according to (93) and then constructs a tree of order K using the above recursive process.

Now we establish various forms of self-similarity for the hierarchical branching process.

Theorem 11 (Self-similarity of hierarchical branching process, [99]).
Consider a hierarchical branching process $S(t) \overset{d}{\sim} \mathsf{HBP}\big(\{T_k\}, \{\lambda_j\}, \{\pi_K\}\big)$ and let $T := T[S]$ be the tree generated by $S(t)$ on $\mathcal{BL}^|_{\mathrm{plane}}$. The following statements hold.

1. *The combinatorial tree $\mathrm{SHAPE}(T)$ is mean self-similar (according to Defs. 12, 14) with Tokunaga coefficients $\{T_k\}$.*
2. *The combinatorial tree $\mathrm{SHAPE}(T)$ is Horton self-similar (according to Def. 8) with Tokunaga coefficients $\{T_k\}$ if and only if*

$$\pi_K = p(1-p)^{K-1} \text{ for all } K \geq 1 \text{ and some } 0 < p < 1.$$

3. *The tree T is Horton self-similar (according to Def. 9) with scaling exponent $\zeta > 0$ if and only if*

$$\pi_K = p(1-p)^{K-1}, \ K \geq 1, \quad and \quad \lambda_j = \gamma \zeta^{-j}, \ j \geq 1,$$

for some positive γ and $0 < p < 1$.

Proof. By process construction, the tree T is coordinated (Def. 10), with independent complete subtrees.

(1) According to (97), the expected value of the number $N_i[b]$ of side branches of order $i \geq 1$ within a branch b of order $j > i$ is given by $\mathsf{E}\,[N_i[b]] = T_{j-i}$. The mean self-similarity of Def. 12 with Tokunaga coefficients T_k immediately follows via conditioning argument as in (26).

(2) Assume that $\mathrm{SHAPE}\,(T)$ is self-similar. A geometric distribution of orders is then established in Prop. 1. 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 8, 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 $T_K/(S_K - 1)$ on the total side branch count m. Accordingly, the total side branch count after pruning has geometric distribution with success probability

$$q^{\mathcal{R}} = S_{K-1}^{-1}.$$

The order assignment among the remaining side branches (with possible orders $i = 1, \ldots, K - 1$) 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 ζ. 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 $\mathsf{ord} = 1$. Accordingly, the spacings between the remaining side branches are exponentially distributed with a decreased rate

$$\lambda_{K+1} S_{K-1} = \zeta \lambda_K S_{K-1}.$$

Comparing this with (96), and recalling the self-similarity of $\mathsf{SHAPE}\,(T)$, we conclude that Def. 9 is satisfied with scaling exponent ζ. □

7.2. 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 hierarchical branching process S. The values of $x_j^{(n)}(s)$ at branching and termination times are set equal to the right limit values. Thus, $x_j^{(n)}(s)$ are càdlàg[2] functions, and $x_0^{(n)}(s) \equiv 0$. 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,

$$x_j^{(n)}(s) \xrightarrow{\text{a.s.}} \mathsf{E}\,[N_j(s)] =: x_j(s).$$

Theorem 12 (Hydrodynamic limit for branch dynamics, [99]). *Suppose that the following conditions are satisfied:*

$$L := \limsup_{k \to \infty} T_k^{1/k} < \infty, \tag{98}$$

and

$$\sup_{j \geq 1} \lambda_j < \infty, \quad \limsup_{j \to \infty} \lambda_j^{1/j} \leq 1/L. \tag{99}$$

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

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

---

[2]The word "càdlàg" is an abbreviation for "continue à droite, limite à gauche" in French, which means "continuous on the right, limit on the left" in English. The space of càdlàg functions is called Skorokhod space.

converges almost surely, as $n \to \infty$, to the dynamical system

$$\mathbf{x}(s) = \Big(x_1(s), x_2(s), \dots\Big)^T, \qquad s \in [0, T],$$

that satisfies

$$\dot{\mathbf{x}} = \mathbb{G}\Lambda\mathbf{x} \quad \textit{with the initial conditions} \quad \mathbf{x}(0) = \pi := \sum_{K=1}^{\infty} \pi_K \mathbf{e}_K, \qquad (100)$$

where $\Lambda = diag\{\lambda_1, \lambda_2, \dots\}$ is a diagonal operator with the entries $\lambda_1, \lambda_2, \dots$, $\mathbf{e}_i$ are the standard basis vectors, and operator $\mathbb{G}$ defined in Eq. (41).

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

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

with

$$\beta_\ell(\mathbf{x}) = \begin{cases} \lambda_1 x_1 & \text{if } \ell = -\mathbf{e}_1, \\ \lambda_{i+1} x_{i+1} & \text{if } \ell = 2\mathbf{e}_i - \mathbf{e}_{i+1}, i \geq 1, \\ \sum\limits_{j=i+1}^{\infty} \lambda_j T_{j-i} x_j & \text{if } \ell = \mathbf{e}_i, i \geq 1. \end{cases}$$

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 $\mathbf{x}^{(n)}(s)$ is

$$\begin{aligned} L_n f(\mathbf{x}) &= n\lambda_1 x_1 \left[f\left(\mathbf{x} - \frac{1}{n}\mathbf{e}_1\right) - f(\mathbf{x}) \right] \\ &\quad + \sum_{i=1}^{\infty} n\lambda_{i+1} x_{i+1} \left[f\left(\mathbf{x} - \frac{1}{n}\mathbf{e}_{i+1} + \frac{2}{n}\mathbf{e}_i\right) - f(\mathbf{x}) \right] \\ &\quad + \sum_{i=1}^{\infty} \left(\sum_{j=i+1}^{\infty} n\lambda_j T_{j-i} x_j \right) \left[f\left(\mathbf{x} + \frac{1}{n}\mathbf{e}_i\right) - f(\mathbf{x}) \right]. \end{aligned} \quad (101)$$

Let

$$\begin{aligned} F(\mathbf{x}) &:= \sum_{\ell} \beta_\ell(\mathbf{x}) = -\lambda_1 x_1 \mathbf{e}_1 + \sum_{i=1}^{\infty} \lambda_{i+1} x_{i+1}(2\mathbf{e}_i - \mathbf{e}_{i+1}) \\ &\quad + \sum_{i=1}^{\infty} \left(\sum_{j=i+1}^{\infty} \lambda_j T_{j-i} x_j \right) \mathbf{e}_i. \end{aligned}$$

The convergence result of Kurtz ([60, Theorem 2.1, Chapter 11], [103, Theorem 8.1]) given here in Appendix A 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 Thm. 35. Specifically, we require that for a compact set $\mathcal{C}$ in $\ell^1(\mathbb{R})$,

$$\sum_\ell \|\ell\|_1 \sup_{\mathbf{x} \in \mathcal{C}} \beta_\ell(\mathbf{x}) < \infty, \tag{102}$$

and there exists $M_{\mathcal{C}} > 0$ such that

$$\|F(\mathbf{x}) - F(\mathbf{y})\|_1 \le M_{\mathcal{C}} \|\mathbf{x} - \mathbf{y}\|_1, \qquad \mathbf{x}, \mathbf{y} \in \mathcal{C}. \tag{103}$$

Here the condition (102) follows from

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

which in turn follow from conditions (99). Similarly, Lipschitz conditions (103) are satisfied in $\mathcal{C}$ due to conditions (99). Thus, by Thm. 35 extended for $\ell^1(\mathbb{R})$, the process $\mathbf{x}^{(n)}(s)$ converges almost surely to $\mathbf{x}(s)$ that satisfies $\dot{\mathbf{x}} = F(\mathbf{x})$, which expands as the following system of ordinary differential equations:

$$\begin{cases} x_1'(s) & = -\lambda_1 x_1 + \lambda_2(T_1 + 2)x_2 + \lambda_3 T_2 x_3 + \dots \\ x_2'(s) & = -\lambda_2 x_2 + \lambda_3(T_1 + 2)x_3 + \lambda_4 T_2 x_4 + \dots \\ \quad \vdots \\ x_k'(s) & = -\lambda_k x_k + \lambda_{k+1}(T_1 + 2)x_{k+1} + \lambda_{k+2} T_2 x_{k+2} + \dots \\ \quad \vdots \end{cases} \tag{104}$$

with the initial conditions $\mathbf{x}(0) = \lim_{n \to \infty} \mathbf{x}^{(n)}(0) = \pi := \sum_{K=1}^\infty \pi_K \mathbf{e}_K$ by the law of large numbers. Finally, we observe that $\|\pi\|_1 = 1$, and conditions (99) imply that $\mathbb{G}\Lambda$ is a bounded operator in $\ell^1(\mathbb{R})$. □

7.3. Criticality and time invariance

7.3.1. Definitions

Assume that the hydrodynamic limit $\mathbf{x}(s)$, and hence the averages $x_j(s)$, exist. Write $\pi = \sum_{K=1}^\infty \pi_K \mathbf{e}_K$ for the initial distribution of the process. Consider the average progeny of the process, that is the average number of branches of any order alive at instant $s \ge 0$:

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

Definition 23 (Criticality). *A hierarchical branching process $S(s)$ is said to be critical if its average progeny is constant: $C(s) = 1$ for all $s \ge 0$.*

Recall that a Galton-Watson process that begins with a single progenitor is critical if and only if its average progeny (the average number of branches) equals to unity at all times [12]. This provides an analogy with the above defined critical hierarchical branching process.

Definition 24 (**Time invariance**). *A hierarchical branching process $S(s)$ with the initial distribution $\pi = \sum_{K=1}^{\infty} \pi_K \mathbf{e}_K$ is said to be time-invariant if*

$$e^{\mathbb{G}\Lambda s}\pi = \pi \quad \text{for all} \quad s \geq 0. \tag{105}$$

Proposition 9 (**Time invariance $\rightarrow$ Criticality**). *Suppose the hydrodynamic limit $\mathbf{x}(s)$ exists, and the hierarchical branching process $S(s)$ is time-invariant. Then the process $S(s)$ is critical.*

Proof. $C(s) = \|\mathbf{x}(s)\|_1 = \|e^{\mathbb{G}\Lambda s}\pi\|_1 = \|\pi\|_1 = 1.$ □

Recall the function $\hat{t}(z) = -1 + 2z + \sum_j z^j T_j$ defined in Eq. (31) for all complex $|z| < 1/L$, where the inverse radius of convergence L is defined in Eq. (98). We also recall that there is a unique real root w_0 of $\hat{t}(z)$ within $(0, \frac{1}{2}]$. We formulate some of the results below in terms of $\hat{t}(z)$ and the *Horton exponent* $R := w_0^{-1}$; see Thm. 1.

Proposition 10. *Suppose $\Lambda\pi$ is a constant multiple of the geometric vector $v_0 = \sum_{K=1}^{\infty} R^{-K}\mathbf{e}_K$, i.e., there exists a coefficient of proportionality $\beta > 0$ such that*

$$\lambda_K \pi_K = \beta R^{-K} \quad \forall K \geq 1. \tag{106}$$

Then the process $S(s)$ is time-invariant.

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

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

and

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

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

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

7.3.2. 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 11, 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 $\pi_K = p(1-p)^{K-1}$ and $\lambda_j = \gamma \zeta^{-j}$. We refer to a self-similar process by its parameter triplet, and write $S(s) \overset{d}{\sim} S_{p,\gamma,\zeta}(s)$. We denote the respective average progeny by $C_{p,\gamma,\zeta}(s)$. Observe that in the self-similar case the first of the conditions (99) is equivalent to $\zeta \geq 1$, and the second is equivalent to $\zeta \geq L$. Hence, the conditions (99) are equivalent to $\zeta \geq 1 \vee L$.

Theorem 13 (Average progeny of a self-similar process, [99]). *Consider a self-similar process $S_{p,\gamma,\zeta}(s)$ with $0 < p < 1$ and $\gamma > 0$. Suppose that (98) is satisfied and $\zeta \geq 1 \vee L$. Then*

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

Proof. The choice of the limits for ζ ensures that the conditions (99) are satisfied and hence, by Thm. 12, the hydrodynamic limit $\mathbf{x}(s)$ exists and the function $C_{p,\gamma,\zeta}(s)$ is well defined. Now we have

$$\Lambda \pi = \frac{\gamma p}{1-p} \sum_{K=1}^{\infty} \left(\zeta^{-1}(1-p)\right)^K \mathbf{e}_K,$$

and therefore

$$\mathbb{G}\Lambda \pi = \hat{t}\left(\zeta^{-1}(1-p)\right)\Lambda \pi. \tag{107}$$

Iterating recursively, we obtain

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

and in general,

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

Thus, taking $\mathbf{x}(0) = \pi$,

$$\mathbf{x}(s) = e^{\mathbb{G}\Lambda s}\pi = \pi + \sum_{m=1}^{\infty} \frac{s^m}{m!} \left[\prod_{i=1}^{m} \hat{t}\left(\zeta^{-i}(1-p)\right)\right] \Lambda^m \pi. \tag{108}$$

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

$$C_{p,\gamma,\zeta}(s) = \sum_{j=1}^{\infty} x_j(s)$$

$$= 1 + \sum_{m=1}^{\infty} \frac{s^m}{m!} \left[\prod_{i=1}^{m} \hat{t}\left(\zeta^{-i}(1-p)\right)\right] \sum_{j=1}^{\infty} \left(\Lambda^m \pi\right)_j$$

$$= 1 + \sum_{m=1}^{\infty} \frac{(s\gamma/\zeta)^m}{m!} \left[\prod_{i=1}^{m} \hat{t}\big(\zeta^{-i}(1-p)\big) \right] \frac{p}{1 - \zeta^{-m}(1-p)}, \quad (109)$$

since

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

Next, notice that by letting $p' = 1 - \zeta^{-1}(1-p)$, we have from (109) and the uniform convergence of the corresponding series for any fixed $M > 0$ and $s \in [0, M]$, that

$$\frac{d}{ds} C_{p,\gamma,\zeta}(s) = \frac{\gamma}{\zeta} \hat{t}(1 - p') C_{p',\gamma,\zeta}(s) \quad \text{with} \ \ C_{p,\gamma,\zeta}(0) = C_{p',\gamma,\zeta}(0) = 1. \quad (110)$$

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. $\qquad \square$

Remark 12. If $\zeta = 1$, then $p' = p$ and equation (110) has an explicit solution

$$C_{p,\gamma,1}(s) = \exp\big\{ s\gamma \hat{t}(1-p) \big\}.$$

Accordingly,

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

This case is further examined in Sect. 7.4. In general, the average progeny $C_{p,\gamma,\zeta}(s)$ 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 $\hat{t}\big(\zeta^{-d-1}(1-p)\big) = 0$. Accordingly, (108) implies that $C_{p,\gamma,\zeta}(s)$ is a polynomial of degree d.

Theorem 14 (Criticality of a self-similar process, [99]). *Consider a self-similar process $S_{p,\gamma,\zeta}(s)$ with $0 < p < 1$, $\gamma > 0$. Suppose that (98) 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 Thm. 13. (ii)$\rightarrow$(i) is established in Prop 9. (iii)$\rightarrow$(ii): Observe that $\hat{t}\left(\zeta^{-1}(1-p)\right) = \hat{t}\left(R^{-1}\right) = 0$. Time invariance now follows from (108). □

Remark 13. By Thm. 11, the product $\lambda_K \pi_K$ in a self-similar process is given by

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

for some $0 < p < 1$, $\gamma > 0$, and $\zeta \geq 1 \vee L$. Hence, a time-invariant process can be constructed, according to Prop. 10 and (106), by selecting any sequence $\{T_k\}$ such that the unique real zero w_0 on $[0, 1/2]$ of the respective function $\hat{t}(z)$ is given by

$$w_0 = R^{-1} = \zeta^{-1}(1 - p).$$

Theorem 14 states that this is the only possible way to construct a time-invariant process, given that the process is self-similar.

7.4. Closed form solution for equally distributed branch lengths

Consider a self-similar hierarchical branching process with $\Lambda = I$ and $\pi_K = p(1-p)^{K-1}$ for all $K \geq 1$. Here we have $\lambda_j = 1$ for all $j \geq 1$, which implies $\gamma = \zeta = 1$.

For a given integer $K \geq 1$, condition on the event $\mathsf{ord}(T[S]) = K$. In this case, the system of equation (104) is finite dimensional,

$$
\begin{cases}
x_1'(s) & = -x_1 + (T_1 + 2)x_2 + T_2 x_3 + \ldots + T_{K-1} x_K \\
x_2'(s) & = -x_2 + (T_1 + 2)x_3 + T_2 x_4 + \ldots + T_{K-2} x_K \\
\quad \vdots \\
x_{K-1}'(s) & = -x_{K-1} + (T_1 + 2)x_K \\
x_K'(s) & = -x_K
\end{cases}
\tag{111}
$$

with the initial conditions $\mathbf{x}(0) = \mathbf{e}_K$.

Recall the sequence $t(j)$ defined in Eq. (30), and let $\mathbf{y}(s) = e^s \mathbf{x}(s)$. Then (111) rewrites in terms of the coordinates of $y(s)$ as follows

$$
\begin{cases}
y_1'(s) & = t(1)y_2 + t(2)y_3 + \ldots + t(K-1)y_K \\
y_2'(s) & = t(1)y_3 + t(2)y_4 + \ldots + t(K-2)y_K \\
\quad \vdots \\
y_{K-2}'(s) & = t(1)y_{K-1} + t(2)y_K \\
y_{K-1}'(s) & = t(1)y_K \\
y_K'(s) & = 0
\end{cases}
\tag{112}
$$

with the initial conditions $\mathbf{y}(0) = \mathbf{e}_K$. The ODEs (112) can be solved recursively in a reversed order of equations in the system obtaining for $m = 1, \ldots, K - 1$,

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

Let $\delta_0(j) = \mathbf{1}_{\{j=0\}}$ be the Kronecker delta function. Then we arrive with the closed form solution

$$
\begin{aligned}
x_{K-m}(s) &= e^{-s} y_{K-m}(s) \\
&= e^{-s} \sum_{n=1}^{\infty} \underbrace{(t + \delta_0) * (t + \delta_0) * \ldots * (t + \delta_0)}_{n \text{ times}}(m) \frac{s^n}{n!}.
\end{aligned}
\tag{113}
$$

Next, removing condition $\mathsf{ord}(T[S]) = K$, the tree orders are randomized by assigning geometric probability $\pi_K = p(1-p)^{K-1}$ to each order $K \geq 1$, and the above closed form expression (113) yields the expression for the average progeny observed earlier in Rem. 12 of this section:

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

7.5. Critical Tokunaga process

We introduce here a class of hierarchical branching processes that enjoy all of the symmetries discussed in this work – Horton self-similarity, criticality, time-invariance, strong Horton law, Tokunaga self-similarity, and also have independently distributed edge lengths. Despite these multiple constraints, the class is sufficiently broad, allowing the self-similarity constant ζ (Def. 9, part (ii)) to take any value $\zeta \geq 1$, and the Horton exponent to take any value $R \geq 2$. The critical binary Galton-Watson process is a special case of this class.

Definition 25 (Critical Tokunaga process). *We say that $S(t)$ is a critical Tokunaga process with parameters (γ, c), and write $S(t) \overset{d}{\sim} S^{\mathrm{Tok}}(t; c, \gamma)$, if it is a hierarchical branching process with the following parameter triplet:*

$$\lambda_j = \gamma \, c^{1-j}, \ \ \pi_K = 2^{-K}, \ \ \text{and} \ \ T_k = (c-1) \, c^{k-1} \tag{114}$$

for some $\gamma > 0, \ c \geq 1$.

Proposition 11 (Critical Tokunaga process). *Suppose $S(t) \overset{d}{\sim} S^{\mathrm{Tok}}(t; c, \gamma)$ and let $T[S]$ be the tree of $S(t)$. Then,*

1. *$S(t)$ is a Horton self-similar, critical, and time invariant process*

$$S(t) \overset{d}{\sim} S_{\frac{1}{2},\gamma,c}(t).$$

2. *Independently of the combinatorial shape of $T[S]$, its edge lengths are i.i.d. exponential random variables with rate γ.*
3. *We have*

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

Proof. 1. Self-similarity follows from Thm. 11, part (3). Specification of parameters (114) implies $p = 2^{-1}$ and $\zeta = c$. The Horton exponent $R = 2c$ is found from (37). Criticality and time-invariance now follow from Thm. 14, since here

$$2^{-1} = p = 1 - \frac{\zeta}{R} = 1 - \frac{c}{2c} = 2^{-1}.$$

2. To establish the edge lengths property, observe that

$$\{T_0 = 1, T_k = (c-1)c^{k-1}, k \geq 1\} \Rightarrow S_K = 1 + T_1 + \cdots + T_K = c^K, K \geq 0.$$

Recall from Prop. 8, part(4) that the edge lengths within a branch of order $K \geq 1$ are i.i.d. exponential r.v.s. with rate

$$\lambda_K S_{K-1} = \gamma\, c^{1-K} c^{K-1} = \gamma.$$

3. The values of R, p, and ζ are found in 1. The expression for $\hat{t}(z)$ and equality $L = c$ are readily obtained from the geometric form of the Tokunaga coefficients T_k. □

Criticality and i.i.d. edge length distribution property characterize the critical Tokunaga process, as we explain in the following statement.

Lemma 11. *Consider a self-similar hierarchical branching process $S(t) \overset{d}{\sim} S_{p,\gamma,\zeta}(t)$ with $p \in (0,1)$ and $\gamma > 0$. Suppose that (98) holds and $\zeta \geq 1 \vee L$. Let $T[S]$ be the tree of $S(t)$. Then, the following conditions are equivalent:*

1. *$S(t)$ is critical and the edges in T have i.i.d. exponential lengths with rate $\gamma > 0$.*
2. *$S(t)$ is a critical Tokunaga process: $S(t) \overset{d}{\sim} S_{\frac{1}{2},\gamma,c}(t)$.*

Proof. The implication $(2 \Rightarrow 1)$ was established in Prop. 11. To show $(1 \Rightarrow 2)$, recall from Prop. 8, Eq. (96), that the edge lengths within a branch of order K are i.i.d. with rate $\lambda_K S_{K-1}$. If the rate is independent of K, we have for any $K \geq 1$:

$$\lambda_K S_{K-1} = \lambda_{K+1} S_K$$

or

$$\zeta = \frac{\lambda_K}{\lambda_{K+1}} = \frac{S_K}{S_{K-1}}.$$

Given $S_0 = 1$, we find $S_K = \zeta^K$, and hence $T_K = (\zeta - 1)\zeta^{K-1}$. By (37), the Horton exponent is $R = 2\zeta$. Criticality implies (Prop. 14, part (iii)):

$$p_c = 1 - \frac{\zeta}{R} = 2^{-1},$$

which completes the proof. □

It follows from the proof of Lemma 11 that the i.i.d. edge length property alone (and no criticality) is equivalent to the following constraints on the process parameters:

$$\lambda_j = \gamma\,\zeta^{1-j}, \quad \text{and} \quad T_k = (\zeta - 1)\zeta^{k-1},$$

while allowing an arbitrary choice of $p \in (0,1)$. The tree of such process is Tokunaga self-similar, although not critical unless $p = 2^{-1}$.

The next results shows that the critical binary Galton-Watson tree $\mathsf{GW}(\lambda)$ with i.i.d. exponential edge lengths is a special case of the critical Tokunaga process.

Theorem 15 (Critical binary Galton-Watson tree, [99]). *Suppose $S(t)$ is a critical Tokunaga process with parameters*

$$\lambda_j = \gamma 2^{1-j}, \ \pi_K = 2^{-K}, \ and \ T_k = 2^{k-1} \ for \ some \ \gamma > 0, \qquad (115)$$

which means $S(t) \stackrel{d}{\sim} S^{\mathrm{Tok}}(t; 2, \gamma)$. Let $T[S]$ be the tree of $S(t)$. Then $T[S]$ has the same distribution on $\mathcal{BL}^|_{\mathrm{plane}}$ as the critical binary Galton-Watson tree with i.i.d. edge lengths: $T[S] \stackrel{d}{\sim} \mathsf{GW}(\gamma)$.

Proof. Consider a tree $T \stackrel{d}{\sim} \mathsf{GW}(\gamma)$ in $\mathcal{BL}^|_{\mathrm{plane}}$. We show below that this tree can be dynamically generated according to Def. 22 of the hierarchical branching process with parameters (115).

First, notice that by Prop. 6

$$\mathsf{P}(\mathrm{ord}(T) = K) = 2^{-K}.$$

We will establish later in Corollary 12 that the length of every branch of order j in T is exponentially distributed with parameter $\lambda_j = \gamma 2^{1-j}$, which matches the branch length distribution in the hierarchical branching process (115). Furthermore, by Corollary 12, conditioned on $\mathcal{R}^i(T) \neq \phi$ (which happens with a positive probability), we have $\mathcal{R}^i(T) \stackrel{d}{\sim} \mathsf{GW}(2^{-i}\gamma)$. This means that the distribution of Galton-Watson trees pruned i times is a linearly scaled version of the original distribution (the same combinatorial structure, linearly scaled edge lengths). Recall (Prop. 6) the total number m_j of side branches within a branch of order $j \geq 2$ in T is geometrically distributed with mean $T_1 + \cdots + T_{j-1} = 2^{j-1} - 1$, where $T_i = 2^{i-1}$, $i \geq 1$. Conditioned on m_j, the assignment of orders among the m_j side branches is done according to the multinomial distribution with m_j trials and success probability for order $i = 1, \ldots, j-1$ given by $T_{j-i}/(T_1 + \cdots + T_{j-1})$. This implies that 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, the branches of order i merge into any branch of order j in the pruned tree $\mathcal{R}^i(T)$ as a Poisson point process with intensity $\gamma\, 2^{-i} = \lambda_j T_{j-i}$ for every $j > i$.

Finally, the orientation of the two offspring of the same parent in $\mathsf{GW}(\gamma)$ is uniform random, by Def. 20. We conclude that tree $\mathsf{GW}(\gamma)$ has the same distribution on $\mathcal{BL}^|_{\text{plane}}$ as the critical Tokunaga process with parameters (115). $\square$

Remark 14. The condition $T_{i,i+k} = T_k = a\,c^{k-1}$ was first introduced in hydrology by Eiji Tokunaga [155] 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 λ_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 Prop. 8(4). Recall the special place case $a = c - 1$ plays for the entropy rate of Tokunaga self-similar trees as elaborated in Sect. 4.3. See Cor. 1. Interestingly, the constraint $a = c - 1$ appears in the *random self-similar network* (RSN) model introduced by Veitzer and Gupta [161], which uses a purely topological algorithm of recursive local replacement of the network generators to construct random self-similar trees. The importance of the constraint $a = c - 1$ in purely combinatorial context is revealed in Sect. 7.7.

7.6. Martingale approach

In this section, we propose a martingale representation for the size and length of a critical Tokunaga tree of a given order. This leads, via the martingale techniques, to the strong Horton laws for both these quantities, and allows us to find the asymptotic order of a tree of a given size. The proposed martingale representation is related to an alternative construction of a critical Tokunaga tree, via a Markov tree process on $\mathcal{BL}^|_{\text{plane}}$.

7.6.1. Markov tree process

Consider a critical Tokunaga process $S^{\text{Tok}}(t; c, \gamma)$ (Def. 25, Prop. 11) with $c > 1$ (hence excluding a trivial case $c = 1$ of perfect binary trees), and let μ be the measure induced by this process on $\mathcal{BL}^|_{\text{plane}}$. Following the notations introduced in Sect. 3.1, Eq. (6), we consider conditional measures

$$\mu_K(T) = \mu(T \,|\, \mathsf{ord}(T) = K).$$

Next, we construct a discrete time Markov tree process $\left\{\Upsilon_K\right\}_{K \in \mathbb{N}}$ on $\mathcal{BL}^|_{\text{plane}}$ such that for each $K \in \mathbb{N}$,

$$\mathsf{ord}(\Upsilon_K) = K, \quad \Upsilon_K \stackrel{d}{\sim} \mu_K, \quad \text{and} \quad \mathcal{R}(\Upsilon_{K+1}) = \Upsilon_K. \tag{116}$$

Let

$$X_K = N_1[\Upsilon_K] = \frac{1 + \#\Upsilon_K}{2} \in \mathbb{N}$$

be the number of leaves in Υ_K and $Y_K = \text{LENGTH}(\Upsilon_K) \in \mathbb{R}_+$ be the tree length. Consequently, $X_{K-k+1} = N_1[\Upsilon_{K-k+1}] = N_1[\mathcal{R}^{k-1}(\Upsilon_K)] = N_k[\Upsilon_K]$ for all $1 \le k \le K$.

We let Υ_1 be an I-shaped tree of Horton-Strahler order one, with the edge length $Y_1 \overset{d}{\sim} \text{Exp}(\gamma)$. This tree has one leaf, $X_1 = 1$.

Conditioned on Υ_K, the tree Υ_{K+1} is constructed according to the following transition rules. Denote by Υ'_K the tree Υ_K with edge length scaled by c. That is, the tree Υ'_K is obtained by multiplying the edge lengths in Υ_K by c, while preserving the combinatorial shape and planar embedding:

$$\text{P-SHAPE}(\Upsilon'_K) = \text{P-SHAPE}(\Upsilon_K).$$

Next, recall that $\Upsilon'_K \in \mathcal{BL}^|_{\text{plane}}$ is itself a metric space (Sec 2.1). We attach new leaf edges to Υ'_K at the points sampled by a homogeneous Poisson point process with intensity $\gamma(c-1)c^{-1}$ along the carrier space Υ'_K. The left-right orientation of the newly added edges is determined independently and uniformly. We also attach a pair of new leaf edges to each of the leaves in Υ'_K; there is exactly $2X_K$ such attachments (X_K pairs). The lengths of all the newly attached leaf edges are i.i.d. exponential random variables with parameter γ that are independent of the combinatorial shape and the edge lengths in Υ_K. Finally, we let the tree Υ_{K+1} consist of Υ'_K and all the attached leaves and leaf edges.

Lemma 12. *The process $\{\Upsilon_K\}_{K \in \mathbb{N}}$ is a Markov process that satisfies* (116).

Proof. The process construction readily implies the Markov property, and ensures that $\text{ord}(\Upsilon_K) = K$ and $\mathcal{R}(\Upsilon_{K+1}) = \Upsilon_K$. Next, we show that a random tree Υ_K satisfies Def. 22, conditioned on the tree order $K \ge 1$, with the critical Tokunaga parameters

$$\lambda_j = \gamma c^{1-j} \quad \text{and} \quad T_k = (c-1)c^{k-1}.$$

The tree Υ_1 has exponential edge length with parameter $\lambda_1 = \gamma$ and no side branching, hence $\Upsilon_1 \overset{d}{\sim} \mu_1$. Assume now that $\Upsilon_K \overset{d}{\sim} \mu_K$ for some $K \ge 1$. Then, for Υ_{K+1}, we establish each of the rules in Def. 22 characterizing the tree measure μ induced by $S^{\text{Tok}}(t; c, \gamma)$, except the tree order rule (i).

Fix any j such that $1 < j \le K$. Every branch of order j in Υ_{K+1} is formed by a branch of order $j - 1$ in Υ_K. In particular, the length of the branch is multiplied by c. Accordingly, every branch of order j within Υ_{K+1} produces offspring of every order i such that $1 < i < j$ with rate

$$c^{-1}\left(\lambda_{j-1}T_{(j-1)-(i-1)}\right) = c^{-1}\gamma c^{1-(j-1)}T_{j-i} = \gamma c^{1-j}T_{j-i} = \lambda_j T_{j-i}.$$

By construction, the side branches of order $i = 1$ are generated with rate

$$\gamma(c-1)c^{-1} = \lambda_j T_{j-1}.$$

This establishes rule (ii) in Def. 22.

Using the same argument as above, each branch of order $j > 1$ in Υ_{K+1} terminates with rate $c^{-1}\lambda_{j-1} = \lambda_j$. By construction, each branch of order $i = 1$ terminates with rate $\gamma = \lambda_1$. This establishes rule (iii) in Def. 22.

Rules (iv), (v), and (vi) in Def. 22 follow trivially from the process construction. This completes the proof.   □

Notice that sampling a random variable $\kappa \stackrel{d}{\sim} \mathsf{Geom}_1\left(\frac{1}{2}\right)$ independently of the process Υ_K, we have the stopped process $\Upsilon_\kappa \stackrel{d}{\sim} \mu$.

7.6.2. Martingale representation of tree size and length

By construction, the pairs (X_K, Y_K) and (X_{K+1}, Y_{K+1}) are related in an iterative way as follows. Conditioned on the values of (X_K, Y_K), we have

$$X_{K+1} = 2X_K + V_K, \tag{117}$$

where $V_K \stackrel{d}{\sim} \mathsf{Poi}\big(\gamma(c-1)Y_K\big)$ is the number of side branches of order one attached to Υ'_K. Next, conditioning on X_{K+1}, we have

$$Y_{K+1} = U_K + cY_K, \tag{118}$$

where $U_K \stackrel{d}{\sim} \mathsf{Gamma}\big(X_{K+1}, \gamma\big)$ is the sum of X_{K+1} i.i.d. edge lengths, each exponentially distributed with parameter γ.

Lemma 13 (Martingale representation). *The sequence*

$$M_K = R^{1-K}\left(X_K + \gamma(c-1)Y_K\right) \ \text{ with } K \in \mathbb{N} \tag{119}$$

is a martingale with respect to the Markov tree process $\big\{\Upsilon_K\big\}_{K\in\mathbb{N}}$.

Proof. Taking conditional expectations in (117) and (118) gives

$$
\begin{aligned}
\mathsf{E}[X_{K+1}\,|\,\Upsilon_K] &= 2X_K + \gamma(c-1)Y_K, \tag{120}\\
\mathsf{E}[Y_{K+1}\,|\,\Upsilon_K] &= \gamma^{-1}\mathsf{E}[X_{K+1}\,|\,\Upsilon_K] + cY_K \\
&= 2\gamma^{-1}X_K + (2c-1)Y_K. \tag{121}
\end{aligned}
$$

This can be summarized as

$$\mathsf{E}\left[\begin{pmatrix} X_{K+1} \\ Y_{K+1} \end{pmatrix}\,\Big|\,\Upsilon_K\right] = \mathbb{M}\begin{pmatrix} X_K \\ Y_K \end{pmatrix}, \tag{122}$$

where

$$\mathbb{M} = \begin{bmatrix} 2 & \gamma(c-1) \\ 2\gamma^{-1} & 2c-1 \end{bmatrix}.$$

The eigenvalues of the matrix $\mathbb{M}$ are $R = 2c$ and 1. The largest eigenvalue equals the Horton exponent R; the respective eigenspace is $y = 2\gamma^{-1}x$. Equation (122) implies that

$$\mathbb{M}^{1-K}\begin{pmatrix} X_K \\ Y_K \end{pmatrix} = \frac{1}{2c-1}\begin{pmatrix} [R^{1-K} + 2(c-1)]X_K + \gamma(c-1)[R^{1-K} - 1]Y_K \\ \\ 2\gamma^{-1}[R^{1-K} - 1]X_K + [2(c-1)R^{1-K} + 1]Y_K \end{pmatrix}$$

is a vector valued martingale with respect to the Markov tree process $\{\Upsilon_K\}_{K\in\mathbb{N}}$. Multiplying this martingale by the left eigenvector $\left(1, \gamma(c-1)\right)$ of $\mathbb{M}$ that corresponds to the largest eigenvalue R, we obtain a scalar martingale with respect to $\{\Upsilon_K\}_{K\in\mathbb{N}}$:

$$\left(1, \gamma(c-1)\right)\mathbb{M}^{1-K}\begin{pmatrix} X_K \\ Y_K \end{pmatrix} = R^{1-K}\left(X_K + \gamma(c-1)Y_K\right).$$

This completes the proof. $\qquad\square$

The next lemma establishes the limit of Y_K/X_K.

Lemma 14. *Suppose $S^{\mathrm{Tok}}(t; c, \gamma)$ is the distribution of a critical Tokunaga process and $\{\Upsilon_K\}_{K\in\mathbb{N}}$ is the corresponding Markov tree process. Then,*

$$Y_K/X_K \to 2\gamma^{-1} \quad a.s. \quad as \quad K \to \infty. \tag{123}$$

Proof. Recall that Y_K is a sum of $2X_K - 1$ independent edge lengths, each being exponentially distributed with parameter γ. Thus, since $X_K = N_1[\Upsilon_K] \geq 2^{K-1}$, the Chebyshev inequality implies for any $\epsilon > 0$,

$$\sum_{k=1}^{\infty} \mathsf{P}\left(\left|\frac{Y_K}{X_K} - 2\gamma^{-1}\right| \geq \epsilon\right) \leq \epsilon^{-2}\sum_{k=1}^{\infty} \mathsf{Var}\left(\frac{Y_K}{X_K} - 2\gamma^{-1}\right)$$

$$\leq \epsilon^{-2}\sum_{k=1}^{\infty} \mathsf{E}\left[\mathsf{E}\left[\left(\frac{Y_K}{X_K} - 2\gamma^{-1}\right)^2 \Big| X_K\right]\right]$$

$$= \epsilon^{-2}\gamma^{-2}\sum_{k=1}^{\infty} \mathsf{E}\left[X_K^{-1} + X_K^{-2}\right]$$

$$\leq \epsilon^{-2}\gamma^{-2}\sum_{k=1}^{\infty}\left(2^{1-K} + 2^{2(1-K)}\right) \quad < \infty$$

as $\mathsf{E}[Y_K | X_K] = 2\gamma^{-1}X_K - \gamma^{-1}$ and $\mathsf{E}[Y_K^2 | X_K] = 4\gamma^{-2}X_K^2 - 3\gamma^{-2}X_K + \gamma^{-1}$.

Hence, by the Borel-Cantelli lemma, we arrive with the almost sure convergence in (123). $\qquad\square$

The following lemma is needed for the proof of Thm. 16.

Lemma 15. *Suppose $S^{\mathrm{Tok}}(t; c, \gamma)$ is the distribution of a critical Tokunaga process and $\{\Upsilon_K\}_{K \in \mathbb{N}}$ is the corresponding Markov tree process. Then,*

$$\mathsf{P}\left(\lim_{K \to \infty} R^{1-K} X_K = 0\right) = 0.$$

Proof. For a given integer $x \geq 2^{K-1}$, we condition on the event $X_K = x$. Then, Y_K is a sum of $2X_K - 1 = 2x - 1$ i.i.d. exponential edge lengths. Hence, $Y_K \overset{d}{\sim} \mathsf{Gamma}(2x - 1, \gamma)$. Finally, recall that in the setup of (117), $V_K \overset{d}{\sim} \mathsf{Poi}(\gamma(c - 1)Y_K)$. Therefore, we can compute the moment generating function of V_K conditioned on the event $X_K = x$ as follows

$$
\begin{aligned}
\mathcal{M}_{\mathsf{v}}(s; x) &:= \mathsf{E}\left[e^{sV_K} \,\middle|\, X_K = x\right] \\
&= \int_0^\infty \sum_{k=0}^\infty e^{sk} e^{-\gamma(c-1)y} \frac{\left(\gamma(c-1)y\right)^k}{k!} \frac{\gamma^{2x-1} y^{2x-2} e^{-\gamma y}}{\Gamma(2x-1)} \, dy \\
&= \int_0^\infty e^{-\gamma\left(c - (c-1)e^s\right)} \frac{\gamma^{2x-1} y^{2x-2}}{\Gamma(2x-1)} \, dy \\
&= \frac{1}{\left(c - (c-1)e^s\right)^{2x-1}},
\end{aligned}
\tag{124}
$$

with the domain $s \in \left(-\infty, \log \frac{c}{c-1}\right)$.

Next, we use (124) in the exponential Markov inequality (a.k.a. Chernoff bound). For a given $\varepsilon \in \left(0, (c-1)c^{-1}\right)$ and $x \geq 2^{K-1}$, by (117) we have, for all $s \geq 0$,

$$
\begin{aligned}
\mathsf{P}\left(\frac{X_{K+1}}{RX_K} \leq 1 - \varepsilon \,\middle|\, X_K = x\right) &= \mathsf{P}\left(V_K \leq 2\left((1-\varepsilon)c - 1\right)x \,\middle|\, X_K = x\right) \\
&= \mathsf{P}\left(-sV_K \geq -2s\left((1-\varepsilon)c - 1\right)x \,\middle|\, X_K = x\right) \\
&\leq e^{2s\left((1-\varepsilon)c-1\right)x} \mathcal{M}_{\mathsf{v}}(-s; x) \\
&= \frac{e^{2s\left((1-\varepsilon)c-1\right)x}}{\left(c - (c-1)e^{-s}\right)^{2x-1}} \\
&= \left(c - (c-1)e^{-s}\right)\left(\frac{e^{(1-\varepsilon)cs}}{ce^s - (c-1)}\right)^{2x},
\end{aligned}
\tag{125}
$$

where we used (117) and $R = 2c$ to rewrite the event $\left\{\frac{X_{K+1}}{RX_K} \leq 1 - \varepsilon\right\}$ as $\left\{V_K \leq 2\left((1-\varepsilon)c - 1\right)X_K\right\}$.

We find the extreme value of $\frac{e^{(1-\varepsilon)cs}}{ce^s - (c-1)}$ in (125), and substitute

$$e^s = \frac{(1-\varepsilon)(c-1)}{(1-\varepsilon)c - 1} = \frac{1-\varepsilon}{1 - \frac{c}{c-1}\varepsilon}$$

into the right hand side of (125), obtaining

$$
\mathsf{P}\left(\frac{X_{K+1}}{RX_K} \leq 1 - \varepsilon \,\big|\, X_K = x\right)
$$

$$
\leq \quad (c - (c-1)e^{-s})\left(\frac{e^{(1-\varepsilon)cs}}{ce^s - (c-1)}\right)^{2x}
$$

$$
= \quad (1-\varepsilon)^{-1}\left(\left(1 - \frac{c}{c-1}\varepsilon\right)\left(\frac{1-\varepsilon}{1 - \frac{c}{c-1}\varepsilon}\right)^{(1-\varepsilon)c}\right)^{2x}
$$

$$
= \quad (1-\varepsilon)^{-1}\exp\left\{-x\left(\frac{c}{c-1}\varepsilon^2 + O(\varepsilon^3)\right)\right\}. \tag{126}
$$

Now, since $X_K \geq 2^{K-1}$, (126) implies

$$
\mathsf{P}\left(\frac{X_{K+1}}{RX_K} \leq 1 - \varepsilon\right) = \sum_{x=2^{K-1}}^{\infty} \mathsf{P}\left(\frac{X_{K+1}}{RX_K} \leq 1 - \varepsilon \,\big|\, X_K = x\right)\mathsf{P}(X_K = x)
$$

$$
\leq \exp\left\{-2^{K-1}\left(\frac{c}{c-1}\varepsilon^2 + O(\varepsilon^3)\right)\right\}. \tag{127}
$$

Next, plugging $\varepsilon = 1 - e^{-1/K^2}$ into (127), we find that

$$
\sum_{K=1}^{\infty} \mathsf{P}\left(\frac{X_{K+1}}{RX_K} \leq e^{-1/K^2}\right) < \infty, \tag{128}
$$

and equivalently,

$$
\sum_{K=1}^{\infty} \mathsf{P}\left(\log\left(\frac{R^{1-K}X_K}{R^{-K}X_{K+1}}\right) \geq \frac{1}{K^2}\right) < \infty. \tag{129}
$$

Therefore, by the Borel-Cantelli lemma,

$$
\mathsf{P}\left(\left|\left\{K \in \mathbb{N} : \log\left(\frac{R^{1-K}X_K}{R^{-K}X_{K+1}}\right) \geq \frac{1}{K^2}\right\}\right| < \infty\right) = 1, \tag{130}
$$

where $|\cdot|$ denotes the magnitude of sets. Hence, as $\sum_{K=1}^{\infty} \frac{1}{K^2} < \infty$,

$$
\mathsf{P}\left(\lim_{K\to\infty} R^{1-K}X_K = 0\right) = \mathsf{P}\left(\lim_{S\to\infty}\prod_{K=1}^{S}\frac{R^{1-K}X_K}{R^{-K}X_{K+1}} = \infty\right)
$$

$$
= \mathsf{P}\left(\prod_{K=1}^{\infty}\frac{R^{1-K}X_K}{R^{-K}X_{K+1}} = \infty\right) = \mathsf{P}\left(\sum_{K=1}^{\infty}\log\left(\frac{R^{1-K}X_K}{R^{-K}X_{K+1}}\right) = \infty\right) = 0. \tag{131}
$$

This completes the proof. $\qquad\square$

7.6.3. Strong Horton laws in a critical Tokunaga tree

The martingale representation of Lem. 13 has an immediate implication for the asymptotic behavior of the average size of a critical Tokunaga tree, stated below.

Corollary 4. *Suppose $S^{\mathrm{Tok}}(t; c, \gamma)$ is the distribution of a critical Tokunaga process and μ is a tree measure on $\mathcal{BL}^{\mid}_{\mathrm{plane}}$ induced by $S^{\mathrm{Tok}}(t; c, \gamma)$. Then, the following closed form expression holds for all $1 \leq k \leq K$:*

$$\frac{(2c-1)\mathcal{N}_k[K] - (c-1)}{(2c-1)\mathcal{N}_1[K] - (c-1)} = R^{1-k}, \ \ with \ R = 2c. \tag{132}$$

The equation (132) *implies, in particular,*

$$\mathcal{N}_1[K] = \frac{R^{K-1}c + c - 1}{2c - 1} = \frac{R^K + R - 2}{2(R-1)}. \tag{133}$$

Notice that (132) yields yet another proof (besides Thms. 1 and 11) of the strong Horton law for mean branch numbers (Def. 17) in the random trees induced by the critical Tokunaga processes.

Proof. Since Y_K is a sum of $2X_K - 1$ independent edge lengths, each exponentially distributed with parameter γ, we have $\mathsf{E}[Y_K] = \gamma^{-1}(2\mathsf{E}[X_K] - 1)$. Therefore,

$$\mathsf{E}[M_K] = R^{1-K}\mathsf{E}[X_K] + \gamma(c-1)R^{1-K}\mathsf{E}[Y_K]$$
$$= (2c-1)R^{1-K}\mathsf{E}[X_K] - (c-1)R^{1-K}.$$

Furthermore, for all $1 \leq k \leq K$, substituting $K - k + 1$ instead of K in the above equation, we obtain

$$\mathsf{E}[M_{K-k+1}] = (2c-1)R^{k-K}\mathsf{E}[X_{K-k+1}] - (c-1)R^{k-K}.$$

Since M_K is a martingale (see Lemma 13), we have $\mathsf{E}[M_{K-k+1}] = \mathsf{E}[M_K]$. Hence,

$$1 = \frac{\mathsf{E}[M_{K-k+1}]}{\mathsf{E}[M_K]} = R^{k-1}\frac{(2c-1)\mathsf{E}[X_{K-k+1}] - (c-1)}{(2c-1)\mathsf{E}[X_K] - (c-1)}$$
$$= R^{k-1}\frac{(2c-1)\mathsf{E}[N_k[\Upsilon_K]] - (c-1)}{(2c-1)\mathsf{E}[N_1[\Upsilon_K]] - (c-1)}$$

as $\mathsf{E}[X_{K-k+1}] = \mathsf{E}[N_k[\Upsilon_K]]$ and $\mathsf{E}[X_K] = \mathsf{E}[N_1[\Upsilon_K]]$. This establishes (132). The strong Horton law (29) for mean branch numbers follows from (132). The expression (133) is obtained by using $k = K$ in (132). This completes the proof. □

We also suggest an alternative proof that emphasizes the spectral property of the transition matrix $\mathbb{M}$ of (122).

Alternative proof of Corollary 4. Taking expectation in (122) we obtain, for any $K > 1$,

$$\begin{pmatrix} \mathsf{E}[X_K] \\ \mathsf{E}[Y_K] \end{pmatrix} = \mathbb{M} \begin{pmatrix} \mathsf{E}[X_{K-1}] \\ \mathsf{E}[Y_{K-1}] \end{pmatrix} = \mathbb{M}^{K-1} \begin{pmatrix} \mathsf{E}[X_1] \\ \mathsf{E}[Y_1] \end{pmatrix}. \tag{134}$$

Since Y_K is a sum of $2X_K - 1$ independent edge lengths, each exponentially distributed with parameter γ, we have $\mathsf{E}[Y_K] = \gamma^{-1}(2\mathsf{E}[X_K] - 1)$. Recall also that $\left(1, \gamma(c-1)\right)$ is the left eigenvector of $\mathbb{M}$ that corresponds to the eigenvalue R. Accordingly,

$$\left(1, \gamma(c - 1)\right)\mathbb{M}^{K-1} = R^{K-1}\left(1, \gamma(c - 1)\right).$$

Premultiplying (134) by the eigenvector $\left(1, \gamma(c - 1)\right)$ we hence obtain

$$(2c - 1)\mathsf{E}[X_K] - (c - 1) = R^{K-1}\left((2c - 1)\mathsf{E}[X_1] - (c - 1)\right),$$

which establishes (132), since $\mathsf{E}[X_1] = \mathsf{E}\left[N_K[\Upsilon_K]\right]$ and $\mathsf{E}[X_K] = \mathsf{E}\left[N_1[\Upsilon_K]\right]$. The strong Horton law (29) for mean branch numbers follows from (132). The expression (133) is obtained by using $k = K$ in (132). This completes the proof. $\square$

The sizes of trees of distinct orders have fixed asymptotic ratios in a much stronger (almost sure) sense, as we show below.

Theorem 16. *Suppose $S^{\mathrm{Tok}}(t; c, \gamma)$ is the distribution of a critical Tokunaga process, μ is a tree measure on $\mathcal{BL}^|_{\mathrm{plane}}$ induced by $S^{\mathrm{Tok}}(t; c, \gamma)$, and $\left\{\Upsilon_K\right\}_{K \in \mathbb{N}}$ is the corresponding Markov tree process. Then,*

$$\frac{N_k[\Upsilon_K]}{N_1[\Upsilon_K]} \overset{a.s.}{\to} R^{1-k} \quad as \ K \to \infty. \tag{135}$$

Proof. Recall that by Lem. 13, M_K defined in (119) is a martingale. Also, $M_K > 0$ and is in L^1 for all $K \in \mathbb{N}$. Thus, by the Doob's Martingale Convergence Theorem, M_K converges almost surely. Hence, by (123), $R^{1-K}X_K$ also converges almost surely, and

$$\lim_{K \to \infty} R^{1-K}X_K = \lim_{K \to \infty} \frac{M_K}{2c - 1}. \tag{136}$$

In other words, for almost every trajectory of the process $\left\{\Upsilon_K\right\}_{K \in \mathbb{N}}$, we have $R^{1-K}X_K = R^{1-K}N_1[\Upsilon_K]$ converging to a finite limit V_∞, where V_∞ is a random variable. Hence, for any $k \in \mathbb{N}$, the random sequences

$$R^{1-K}X_K = R^{1-K}N_1[\Upsilon_K] \quad \text{and} \quad R^{k-K}X_{K-k+1} = R^{k-K}N_k[\Upsilon_K]$$

converge almost surely to the same finite V_∞, where $V_\infty > 0$ *a.s.* by Lem. 15. The almost sure convergence (135) follows. $\square$

The almost sure convergence (135) in Thm. 16 implies the corresponding week convergence

$$\mathsf{P}\left(\left|\frac{N_k[\Upsilon_K]}{N_1[\Upsilon_K]} - R^{1-k}\right| > \epsilon\right) \to 0 \quad \text{as} \quad K \to \infty,$$

via the Bounded Convergence Theorem. We restate it as the following corollary.

Corollary 5 (**Strong Horton law for branch numbers**). *The tree measure* μ *on* $\mathcal{BL}^{|}_{\mathrm{plane}}$ *induced by a critical Tokunaga process* $S^{\mathrm{Tok}}(t; c, \gamma)$ *satisfies the strong Horton law for branch numbers (Def. 16). That is, for any* $\epsilon > 0$,

$$\mu_K\left(\left|\frac{N_k[T]}{N_1[T]} - R^{1-k}\right| > \epsilon\right) \to 0 \quad as \quad K \to \infty.$$

Corollary 6 (**Asymptotic tree order**). *Suppose* $S^{\mathrm{Tok}}(t; c, \gamma)$ *is the distribution of a critical Tokunaga process and* $\{\Upsilon_K\}_{K \in \mathbb{N}}$ *is the corresponding Markov tree process. Then,*

$$\frac{\log_R \#\Upsilon_K}{K} \overset{a.s.}{\to} 1, \quad as \ K \to \infty.$$

Proof. Recall from (136) that

$$R^{1-K} X_K \overset{a.s.}{\to} V_\infty,$$

where V_∞ is finite (by Doob's Martingale Convergence Theorem) and $V_\infty > 0$ *a.s.* by Lemma 15. Accordingly,

$$\log_R X_K - K \overset{a.s.}{\to} \log_R V_\infty - 1, \tag{137}$$

with $-\infty < \log_R V_\infty < \infty$ *a.s.* Recalling that $\#\Upsilon_K = 2X_K - 1$ completes the proof. $\qquad\square$

The almost sure convergence (123) allows to restate the limit results of this section in terms of the tree length Y_K.

Corollary 7 (**Strong Horton laws for tree lengths**). *Suppose* $S^{\mathrm{Tok}}(t; c, \gamma)$ *is the distribution of a critical Tokunaga process,* μ *is a tree measure on* $\mathcal{BL}^{|}_{\mathrm{plane}}$ *induced by* $S^{\mathrm{Tok}}(t; c, \gamma)$, *and* $\{\Upsilon_K\}_{K \in \mathbb{N}}$ *is the corresponding Markov tree process. Then, for a tree* $T \overset{d}{\sim} \mu$,

$$\mathsf{E}\big[\mathrm{LENGTH}(T) \,|\, \mathrm{ord}(T) = K\big] = \mathsf{E}[Y_K] = \frac{R^K - 1}{\gamma(R-1)}, \quad K \geq 1. \tag{138}$$

Furthermore, we have, for any $k \geq 1$,

$$\frac{Y_{K-k}}{Y_K} \overset{a.s.}{\to} R^{-k}, \ as \ K \to \infty, \tag{139}$$

which implies the strong Horton law for tree lengths: for any $\epsilon > 0$,

$$\mu_K\left(\left|\frac{\mathrm{LENGTH}\big(\mathcal{R}^k(T)\big)}{\mathrm{LENGTH}(T)} - R^{-k}\right| > \epsilon\right) \to 0 \quad as \quad K \to \infty. \tag{140}$$

TABLE 1
Mean size, $\mathsf{E}[X_K] = \mathcal{N}_1[K]$, *and length,* $\mathsf{E}[Y_K] = \mathsf{E}[\text{LENGTH}(T)]$, *of an exponential critical binary Galton-Watson tree* $T \overset{d}{\sim} \mathsf{GW}(1)$; *here* $c = 2$, $R = 4$.

$K = \text{ord}(T)$	$\mathcal{N}_1[K]$	$\mathsf{E}[\text{LENGTH}(T)]$	$2 - \frac{\mathsf{E}[Y_K]}{\mathsf{E}[X_K]}$	$4 - \frac{\mathsf{E}[X_K]}{\mathsf{E}[X_{K-1}]}$
1	1	1	1	–
2	3	5	$1/3$	1
3	11	21	9×10^{-2}	$1/3$
4	43	85	2×10^{-2}	9×10^{-2}
5	171	341	6×10^{-3}	2×10^{-2}
6	683	1365	1×10^{-3}	6×10^{-3}
7	2731	5461	4×10^{-4}	1×10^{-3}
8	10923	21845	9×10^{-5}	4×10^{-4}
9	43691	87381	2×10^{-5}	9×10^{-5}
10	174763	349525	6×10^{-6}	2×10^{-5}
11	699051	1398101	1×10^{-6}	6×10^{-6}
12	2796203	5592405	4×10^{-7}	1×10^{-6}
13	11184811	22369621	9×10^{-8}	4×10^{-7}
14	44739243	89478485	2×10^{-8}	9×10^{-8}
15	178956971	357913941	6×10^{-9}	2×10^{-8}
16	715827883	1431655765	1×10^{-9}	6×10^{-9}
17	2863311531	5726623061	3×10^{-10}	1×10^{-9}
18	11453246123	22906492245	9×10^{-11}	3×10^{-10}
19	45812984491	91625968981	2×10^{-11}	9×10^{-11}
20	183251937963	366503875925	5×10^{-12}	2×10^{-11}

Example 12 (Critical binary Galton Watson tree). Theorem 15 asserts that the exponential critical binary Galton-Watson tree, $T \overset{d}{\sim} \mathsf{GW}(\lambda)$, has the same distribution as a critical Tokunaga branching process with $c = 2$ and $\gamma = \lambda$. In this case $R = 2c = 4$ and the expressions (132), (133) give, for any $K \geq 1$,

$$\mathcal{N}_1[K] = \frac{4^K + 2}{6}.$$

Fixing $\lambda = 1$, by the expression (138) we have, for any $K \geq 1$,

$$\mathsf{E}\big[\text{LENGTH}(T) \,|\, \text{ord}(T) = K\big] = \frac{4^K - 1}{3}.$$

Table 1 shows the values of the mean size and mean length of a critical binary Galton-Watson tree $T \overset{d}{\sim} \mathsf{GW}(1)$, conditioned on selected values of tree order.

7.7. Combinatorial HBP: geometric branching process

This section focuses on combinatorial structure of a Horton self-similar hierarchical branching process [99]

$$S(t) \overset{d}{\sim} \mathsf{HBP}\big(\{T_k\}, \{\lambda_j\}, \{p(1-p)^{K-1}\}\big).$$

Let $T[S]$ be the tree generated by $S(t)$ in $\mathcal{BL}^{|}_{\text{plane}}$. Section 7.7.1 introduces a discrete time multi-type *geometric branching process* $\mathcal{G}(s) = \mathcal{G}(s; \{T_k\}, p)$ whose

trajectories induce a random tree $\mathcal{G}(\{T_k\}, p)$ on $\mathcal{BT}^{|}$ such that

$$\mathcal{G}(\{T_k\}, p) \stackrel{d}{=} \text{SHAPE}(T[S]) \in \mathcal{BT}^{|}. \tag{141}$$

We then show in Sect. 7.7.2 that geometric branching process is time invariant (in discrete time) if and only if it is Tokunaga self-similar with $T_k = (c-1)c^{k-1}$ and $p = 1/2$.

7.7.1. Definition and main properties

Our goal is to consider the combinatorial shape of a self-similar hierarchical branching process. The following definition suggests an explicit time dependent construction of such a process, which we denote $\mathcal{G}(s; \{T_k\}, p)$.

Definition 26 (**Geometric Branching Process**). *Consider a sequence of Tokunaga coefficients $\{T_k \geq 0\}_{k \geq 1}$ and $0 < p < 1$. Define*

$$S_K := 1 + T_1 + \cdots + T_K$$

for $K \geq 0$ by assuming $T_0 = 0$. The Geometric Branching Process (GBP) $\mathcal{G}(s) = \mathcal{G}(s; \{T_k\}, p)$ describes a discrete time multi-type population growth:

(i) The process starts at $s = 0$ with a progenitor of order $\text{ord}(\mathcal{G})$ such that $\text{ord}(\mathcal{G}) \stackrel{d}{\sim} \text{Geom}_1(p)$.

(ii) At every integer time instant $s > 0$, each population member of order $K \in \{1, \ldots, \text{ord}(\mathcal{G})\}$ terminates with probability $q_K = S_{K-1}^{-1}$, independently of other members. At termination, a member of order $K > 1$ produces two offspring of order $(K-1)$; and a member of order $K = 1$ terminates with leaving no offspring.

(iii) At every integer time instant $s > 0$, each population member of order $K \in \{1, \ldots, \text{ord}(\mathcal{G})\}$ survives (does not terminate) with probability

$$1 - q_K = 1 - S_{K-1}^{-1},$$

independently of other members. In this case, it produces a single offspring (side branch). The offspring order $i \in \{1, \ldots, K-1\}$, is drawn from the distribution

$$p_{K,i} = \frac{T_{K-i}}{T_1 + \cdots + T_{K-1}}. \tag{142}$$

The geometric tree $\mathcal{G}(\{T_k\}, p)$ is a combinatorial tree generated by the trajectories of $\mathcal{G}(s; \{T_k\}, p)$ in $\mathcal{BT}^{|}$.

By construction, the distribution of a geometric tree $\mathcal{G}(\{T_k\}, p)$ coincides with the combinatorial shape of the tree of a combinatorially Horton self-similar hierarchical branching process $S(t)$ with Tokunaga coefficients $\{T_k\}$, initial order distribution $\pi_K = p(1-p)^{K-1}$ and an arbitrary positive sequence of termination rates $\{\lambda_i\}$. Accordingly, the branching structure of a geometric tree is described
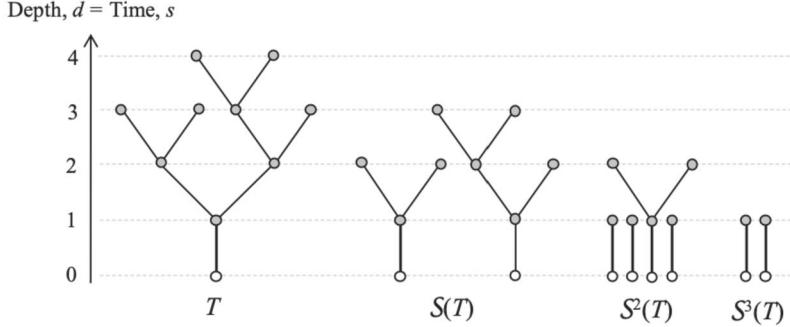
FIG 23. *Time shift $\mathcal{S}$: an illustration. The figure shows forests obtained by consecutive appli-cation of the time shift operator $\mathcal{S}$ to a tree T shown on the left. At every step, we remove the stem from each existing tree. This terminates the trees of order $\mathsf{ord} = 1$, and splits any other tree in two new trees. The operation $\mathcal{S}^d(T)$ removes all vertices at depth $\leq d$, together with their parental edges.*

by Prop. 8, items (1)–(4). The essential elements of the geometric trees (tree order, total number of side branches within a branch, numbers of side branches of a given order within a branch) are described by geometric laws, hence the model's name.

Similarly to the tree of an HBP, a geometric tree can be constructed without time-dependent simulations, following a suitable modification of the algorithm given after Prop. 8. Specifically, the step that involves generation and assignment of the edge lengths l_i should be skipped.

Consider a geometric tree $\mathcal{G} = \mathcal{G}(\{T_k\}, p)$ and its two subtrees, T^a and T^b, rooted at the internal vertex closest to the root, randomly and uniformly per-muted. We call T^a and T^b the *principal* subtrees of $\mathcal{G}$. Let K be the order of $\mathcal{G}$, and, conditioned on $K > 1$, let K_a, K_b be the orders of the principal subtrees T^a and T^b, respectively. Observe that the pair K_a, K_b uniquely defines the tree order K:

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

We write $K_1 \leq K_2$ for the order statistics of K_a, K_b.

Lemma 16 (Order of principal subtrees). *Conditioned on the tree order K, the joint distribution of the order statistics (K_1, K_2) is given by*

$$\mathsf{P}\left(K_1 = j, K_2 = m | K = k\right) = \begin{cases} S_{k-1}^{-1} & \text{if } j = m = k - 1, \\ T_{k-j} S_{k-1}^{-1} & \text{if } j < m = k, \end{cases} \tag{143}$$

where

$$\mathsf{P}(K = k | K > 1) = (1 - p)p^{k-2}, \quad k \geq 2.$$

Proof. Definition 26, part (ii) states that a branch of order K splits into two branches of order $K-1$ with probability S_{K-1}^{-1}, which establishes the first case of

(143). Definition 26, part (iii) states that, otherwise, with probability $1 - S_{K-1}^{-1}$, a side branch is created whose order equals j with probability $T_{K-j}(S_{K-1} - 1)^{-1}$. This gives

$$
\begin{aligned}
\mathsf{P}\,&(K_1 = j, K_2 = k | K = k) \\
&= \quad \mathsf{P}\,(K_1 = j | K = k, K_2 = k)\,\mathsf{P}\,(K_2 = k | K = k) \\
&= \quad \frac{T_{k-j}}{S_{k-1} - 1}\left(1 - \frac{1}{S_{k-1}}\right) = T_{k-j}S_{k-1}^{-1},
\end{aligned}
$$

which establishes the second case.   □

7.7.2. Tokunaga self-similarity of time invariant process

Let $\mathrm{x}_i(s)$, $i \geq 1$, denote the average number of vertices of order i at time s in the process $\mathcal{G}(s)$, and $\mathbf{x}(s) = (\mathrm{x}_1(s), \mathrm{x}_2(s), \dots)^T$ be the state vector. By definition we have

$$
\mathbf{x}(0) = \pi := \sum_{K=1}^{\infty} p(1-p)^{K-1}\mathbf{e}_K,
$$

where $\mathbf{e}_K$ are standard basis vectors. Furthermore, if $q_{a,b}$, $a \geq b$, denotes the probability that a vertex of order $\mathsf{ord} = a + \mathbf{1}_{\{a=b\}}$ that exists at time s splits into a pair of vertices of orders (a,b) at time $s+1$, then

$$
\mathrm{x}_K(s+1) = 2\,\mathrm{x}_{K+1}(s)q_{K,K} + \mathrm{x}_K(s)(1 - q_{K-1,K-1}) + \sum_{i=K+1}^{\infty}\mathrm{x}_i(s)\,q_{i,K}. \quad (144)
$$

The first term in the right-hand side of (144) corresponds to a split of an order-$(K+1)$ vertex into two vertices of order K, the second – to a split of an order-K vertex into a vertex of order K and a vertex of a smaller order, and the third – to a split of a vertex of order $i > K$ into a vertex of order K and a vertex of order i. The geometric branching implies (see Lemma 16, Eq. (143))

$$
q_{a,b} = \left\{ \begin{array}{ll} S_a^{-1} & \text{if } a = b, \\ T_{a-b}S_{a-1}^{-1} & \text{if } b < a. \end{array} \right. \quad (145)
$$

Accordingly, the system (144) rewrites as

$$
\mathbf{x}(s+1) - \mathbf{x}(s) \;=\; \mathbb{G}\mathbb{S}^{-1}\mathbf{x}(s), \quad (146)
$$

where $\mathbb{G}$ is defined in Eq. (41), and

$$
\mathbb{S} = \mathsf{diag}\{S_0, S_1, \dots\}.
$$

In this setup, the *unit time shift* operator $\mathcal{S}$, which advances the process time by unity, can be applied to individual trees and forests (collection of trees). For each tree $T \in \mathcal{T}^|$, the operator removes the root and stem, resulting in two principal subtrees T^a and T^b. A consecutive applications of d time shifts to a tree T is equivalent to removing the vertices at depth $\leq d$ from the root together with their parental edges (Fig. 23). Next we define time invariance with respect to the shift $\mathcal{S}$.

Definition 27 (**Combinatorial time invariance**). *Geometric branching process $\mathcal{G}(s)$, $s \in \mathbb{Z}_+$, is called time invariant if the state vector $\mathbf{x}(s)$ is invariant with respect to a unit time shift $\mathcal{S}$:*

$$\mathbf{x}(s) = \mathbf{x}(0) \equiv \pi \ \ \forall s \iff \mathbb{G}\mathbb{S}^{-1}\pi = \mathbf{0}. \tag{147}$$

Now we formulate the main result of this section.

Theorem 17 (**Combinatorial time invariance, [98]**). *A geometric branching process $\mathcal{G}(s; T_k, p)$ is time-invariant if and only if*

$$p = 1/2 \ \ and \ \ T_k = (c-1)c^{k-1} \ \ for \ any \ c \geq 1. \tag{148}$$

We call this family a (combinatorial) critical Tokunaga process, and the respective trees – (combinatorial) critical Tokunaga trees.

Theorem 17 is proven in Sect. 7.7.4 via solving a nonlinear system of equations that writes (147) in terms of ratios S_k/S_{k+1}.

Remark 15 (**Time invariance in combinatorial vs. metric trees**). Theorems 14 and 17 establish parameterization that is necessary and sufficient for time invariance of the hierarchical branching process in metric and combinatorial case, respectively. The combinatorial time invariance (Thm. 17) appears to be a much more restrictive property, as it only holds for $p = 1/2$ and a specific one-parametric class of Tokunaga coefficients (148). The metric time invariance (Thm. 14), at the same time, merely connects p to the Horton exponent R (via $p = 1 - \zeta/R$), hence allowing one to construct time invariant processes with a range of distinct Tokunaga sequences and values of p. The flexibility of the metric time invariance is explained by the possibility of the average edge lengths to scale with edge order (via parameter ζ), while each edge in the combinatorial case corresponds to a fixed time step of unit length. However, if one removes this additional degree of freedom in the metric case by requiring the edge lengths to be i.i.d. exponential random variables independently of the edge orders, then a mere criticality (which is weaker than time invariance, see Prop. 9) becomes equivalent to the parameterization of (148); this has been shown in Lem. 11.

Corollary 8. *Let $\mathcal{G}$ be a combinatorial critical Tokunaga tree. Then the distribution of the principal subtree T^a (and hence T^b) matches that of the initial tree $\mathcal{G}$. The distributions of T^a and T^b are independent if and only if $c = 2$.*

Proof. Let $\mathsf{ord}(\mathcal{G})$ denote the (random) order of a random geometric tree $\mathcal{G}$. Conditioned on $\mathsf{ord}(\mathcal{G}) > 1$, at instant $s = 1$ (equivalently, after a unit time shift $\mathcal{S}$) there exist exactly two vertices that are the roots of the principal subtrees T^a and T^b. Since the trees T^a and T^b have the same distribution, their roots have the same order distribution. Denote by y_k the probability that the tree T^a has order $k \geq 1$ and let $\mathbf{y} = (y_1, y_2, \dots)^T$. By Thm. 17, the process $\mathcal{G}(s)$ is time invariant. We have $p = \pi_1 = 1/2$, which, together with time invariance, implies

$$\mathbf{x}(0) = \mathbf{x}(1) = 2\mathbf{y}(1 - \pi_1) + \mathbf{0}\pi_1 = \mathbf{y}.$$

This establishes the first statement.

The second statement follows from examining the joint distribution $q_{a,b}$ of (145). Recall that we write K for the order of tree T, K_a, K_b for the orders of the principal subtrees T^a, T^b, and $K_1 < K_2$ for the order statistics of K_a, K_b. Observe that for $k > 1$,

$$P(K_a = m \mid K = k)$$

$$= \begin{cases} \frac{1}{2} \sum\limits_{j:j<k} P(K_1 = j, K_2 = k | K = k) & \text{if } m = k, \\ P(K_1 = K_2 = k - 1 | K = k) & \\ \quad + \frac{1}{2} P(K_1 = k - 1, K_2 = k | K = k) & \text{if } m = k - 1, \\ \frac{1}{2} P(K_1 = m, K_2 = k | K = k) & \text{if } m < k - 1, \end{cases}$$

$$= \begin{cases} \frac{1}{2}(S_{k-1} - 1)S_{k-1}^{-1} & = \frac{1}{2}(1 - c^{1-k}) & \text{if } m = k, \\ \left(1 + \frac{1}{2}T_1\right)S_{k-1}^{-1} & = \frac{1}{2}(c+1)c^{1-k} & \text{if } m = k - 1, \\ \frac{1}{2}T_{k-m}S_{k-1}^{-1} & = \frac{1}{2}(c-1)c^{-m} & \text{if } m < k - 1. \end{cases}$$

Furthermore,

$$P(K_a = m, \ K_b = j \mid K > 1)$$

$$= \sum_{k=m}^{\infty} P(K_a = m, \ K_b = j \mid K = k)P(K = k | K > 1)$$

$$= \begin{cases} (c - 1)c^{-j}\,2^{-m} & \text{if } j < m, \\ c^{-m}\,2^{-m} & \text{if } j = m. \end{cases}$$

Accordingly, the joint distribution of K_a, K_b equals the product of their marginals if and only if $c = 2$. This establishes the second statement.

We also notice that

$$P(K_a = m \mid K > 1) = \sum_{k=m}^{\infty} P(K_a = m \mid K = k)P(K = k | K > 1)$$

$$= \ (1 - c^{1-m})2^{-m} + c^{-m}2^{-m} + \frac{(c-1)c^{-m}}{2}\sum_{k=m+1}^{\infty} 2^{1-k} = 2^{-m}, \ \ (149)$$

which provides an alternative, direct proof of the first statement of the corollary that does not use the time invariance.   □

Remark 16. Corollary 8 asserts that the principal subtrees in a random critical Tokunaga tree are dependent, except the critical binary Galton-Watson case $c = 2$. 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 will be discussed in Sect. 8.6. The structure of H is

hence reminiscent of a self-similar random process with long-range dependence [118, 143]. Establishing the correlation structure of the Harris paths of critical Tokunaga processes is an interesting open problem (see Sect. 13).

7.7.3. Frequency of orders in a large critical Tokunaga tree

Combinatorial trees of the critical Tokunaga processes (Def. 25, Prop. 11), and hence the combinatorial critical Tokunaga trees of Thm. 17, have an additional important property: the frequencies of vertex orders in a large-order tree approximate the tree order distribution $\pi_K = 2^{-K}$ in the space $\mathcal{BT}^|$. To formalize this observation, let μ be a measure on $\mathcal{BT}^|$ induced by a combinatorial critical Tokunaga process (148), and let $T \overset{d}{\sim} \mu$ be the corresponding combinatorial critical Tokunaga tree. For a fixed $K \geq 1$, let $\mu_K(T) = \mu(T \,|\, \mathsf{ord}(T) = K)$. We write $V_k[T]$ for the number of non-root vertices of order k in a tree T, and let $\mathcal{V}_k[K] = \mathsf{E}_K\big[V_k[T]\big]$. Finally, the total number of non-root vertices in T is $\#T = \sum_{k=1}^{\mathsf{ord}(T)} V_k[T] = 2V_1[T] - 1$, where $V_1[T] = N_1[T]$. Thus, $\mathcal{V}[K] := \mathsf{E}_K\big[\#T\big] = 2\mathcal{V}_1[K] - 1 = 2\mathcal{N}_1[K] - 1$.

Proposition 12. *Let T be a combinatorial critical Tokunaga tree induced by a critical Tokunaga process* (148). *Then*

$$\lim_{K \to \infty} \frac{\mathcal{V}_k[K]}{\mathcal{V}_1[K]} = 2^{1-k}. \tag{150}$$

Let $v \in T$ be a vertex selected by uniform random drawing from the non-root vertices of T. Then, for any $k \geq 1$,

$$\lim_{K \to \infty} \mathsf{P}\big(\mathsf{ord}(v) = k \,\big|\, \mathsf{ord}(T) = K\big) = 2^{-k}. \tag{151}$$

Proof. Theorem 1 asserts that a critical Tokunaga tree T satisfies the strong Horton law (29) with Horton exponent $R = 2c$:

$$\lim_{K \to \infty} \frac{\mathcal{N}_k[K]}{\mathcal{N}_1[K]} = (2c)^{1-k}, \text{ for any } k \geq 1.$$

Conditioned on $\mathsf{ord}(T) = K$ we have, for any $k \in \{1, \ldots, K\}$,

$$V_k[K] = \sum_{i=1}^{N_k(T)} (1 + m_i(T)),$$

where $m_i(T)$ is the number of side branches that merge the i-th branch of order k in T, according to the proper branch labeling of Sect. 2.7. Proposition 8 gives

$$\mathcal{V}_k[K] = \mathcal{N}_k[K]S_{k-1}.$$

For a critical Tokunaga tree with $T_k = (1-c)c^{k-1}$ this implies

$$\lim_{K \to \infty} \frac{\mathcal{V}_k[K]}{\mathcal{V}_1[K]} = \lim_{K \to \infty} \frac{\mathcal{N}_k[K]S_{k-1}}{\mathcal{N}_1[K]} = (2c)^{1-k}c^{k-1} = 2^{1-k}.$$

To show (151), we write

$$V_k[T] = N_k[T] + \sum_{i=1}^{N_k[T]} m(i), \quad \mathsf{E}[m(i)] = S_{k-1} - 1,$$

where $m(i)$ is a random variable that represents the total number of side branches within i-th branch of order k within T. Since $N_k[T] \xrightarrow{p} \infty$ for any $k \geq 1$ as $\mathsf{ord}(T) \to \infty$, the Weak Law of Large Numbers gives

$$\frac{V_k[T]}{N_k[T]} \xrightarrow{p} S_{k-1} = c^{k-1} \text{ as } \mathsf{ord}(T) \to \infty.$$

Finally, the strong Horton law of Cor. 5 gives

$$\frac{V_k[T]}{\#T} = \frac{V_k[T]}{N_k[T]} \frac{N_k[T]}{2N_1[T] - 1} \xrightarrow{p} c^{k-1}\frac{1}{2}(2c)^{1-k} = 2^{-k}.$$

This implies (151) and completes the proof.   □

Proposition 12 has an immediate extension to trees with edge lengths, which we include here for completeness. Recall (Def. 1) that a tree $T \in \mathcal{BL}^|$ 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.

Proposition 13. *Let T be a tree generated by critical Tokunaga process (Def. 25). Let point $u \in T$ be sampled from a uniform density function on the metric space T, and let $\mathsf{ord}(u)$ denote the order of the edge to which the point u belongs. Then*

$$\lim_{K \to \infty} \mathsf{P}\Big(\mathsf{ord}(u) = k \ \Big| \ \mathsf{ord}(T) = K\Big) = 2^{-k}. \tag{152}$$

Proof. Proposition 11 establishes that the edge lengths in T are i.i.d. exponential random variables. Thus we can generate T by first sampling a combinatorial critical Tokunaga tree $\mathrm{SHAPE}(T)$, and then assigning i.i.d. exponential edge lengths. Provided that we already sampled $\mathrm{SHAPE}(T)$, 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 $\mathrm{SHAPE}(T)$, in order of proper labeling of Sect. 2.7. The order $\mathsf{ord}(u)$ is uniquely determined by the edge to which u belongs. The statement now follows from Prop. 12.   □

7.7.4. Proof of Theorem 17

Lemma 17 ([**98**]). *A geometric branching process $\mathcal{G}(s)$ is time invariant if and only if $p = 1/2$ and the sequence $\{T_k\}$ solves the following (nonlinear) system of equations:*

$$\frac{S_0}{S_k} = \sum_{i=1}^{\infty} 2^{-i} \frac{S_i}{S_{k+i}} \quad \text{for all } k \geq 1. \tag{153}$$

Proof. Assume that the process is time invariant. Then the process progeny is constant in time and equals unity:

$$\|\pi\|_1 = \sum_{k=1}^{\infty} p(1-p)^{k-1} = 1.$$

Observe that in one time step, every vertex of order $\mathsf{ord} = 1$ terminates, and any vertex of order $\mathsf{ord} > 1$ splits in two. Hence, the process progeny at $s = 1$ is

$$2\sum_{k=2}^{\infty} p(1-p)^{k-1} = 2(1-p) = 1,$$

which implies $p = 1/2$. Accordingly, $p(1-p)^{k-1} = 2^{-k}$ and the time invariance (147) takes the following coordinate form

$$-\frac{2^{-k}}{S_{k-1}} + 2^{-(k+1)}\frac{T_1 + 2}{S_k} + \sum_{i=k+2}^{\infty} 2^{-i}\frac{T_{i-k}}{S_{i-1}} = 0, \text{ for all } k \geq 1. \tag{154}$$

Multiplying (154) by 2^k and observing that $T_k = S_k - S_{k-1}$, we obtain

$$-\frac{1}{S_{k-1}} + \frac{1}{2}\frac{T_1 + 2}{S_k} + \sum_{i=2}^{\infty} 2^{-i}\frac{T_i}{S_{k+i-1}} = 0,$$

$$\frac{1}{S_{k-1}} - \sum_{i=1}^{\infty} 2^{-i}\frac{S_i}{S_{k+i-1}} = \frac{1}{S_k} - \frac{1}{2\,S_k} - \sum_{i=2}^{\infty} 2^{-i}\frac{S_{i-1}}{S_{k+i-1}},$$

and

$$\frac{1}{S_{k-1}} - \sum_{i=1}^{\infty} 2^{-i}\frac{S_i}{S_{k+i-1}} = \frac{1}{2}\left(\frac{1}{S_k} - \sum_{i=1}^{\infty} 2^{-i}\frac{S_i}{S_{k+i}}\right). \tag{155}$$

We prove (153) by induction. For $k = 1$ we have

$$\frac{1}{2} = \frac{1}{2\,S_1} + \sum_{i=1}^{\infty} 2^{-(i+1)}\frac{S_i - S_{i-1}}{S_i},$$

$$1 = \frac{1}{S_1} + \sum_{i=1}^{\infty} 2^{-i} - \sum_{i=1}^{\infty} 2^{-i}\frac{S_{i-1}}{S_i},$$

which establishes the base case

$$\frac{1}{S_1} = \sum_{i=1}^{\infty} 2^{-i} \frac{S_i}{S_{i+1}}.$$

Next, assuming that the statement is proven for $(k-1)$, the left-hand side of (155) vanishes, and the right-hand part rewrites as (153). This establishes necessity.

Conversely, we showed that the system (153) is equivalent to (147) in case $p = 1/2$. This establishes sufficiency. $\square$

Let $a_k = S_k/S_{k+1} \le 1$ for all $k \ge 0$. Then, for any $i \ge 0$ and any $k > 0$ we have $S_i/S_{k+i} = a_i\, a_{i+1} \ldots a_{i+k-1}$. The system (153) rewrites in terms of a_i as

$$\frac{1}{2}a_1 + \frac{1}{4}a_2 + \frac{1}{8}a_3 + \ldots = a_0,$$
$$\frac{1}{2}a_1 a_2 + \frac{1}{4}a_2 a_3 + \frac{1}{8}a_3 a_4 + \ldots = a_0 a_1,$$
$$\frac{1}{2}a_1 a_2 a_3 + \frac{1}{4}a_2 a_3 a_4 + \frac{1}{8}a_3 a_4 a_5 + \ldots = a_0 a_1 a_2,$$

and so on, which can be summarized as

$$\sum_{j=1}^{\infty} \frac{1}{2^j} \prod_{k=j}^{n+j-1} a_k = \prod_{k=0}^{n-1} a_k, \text{ for all } n \in \mathbb{N}. \tag{156}$$

Lemma 18 ([**98**])**.** *The system* (156) *with the initial value* $a_0 = 1/c > 0$ *has a unique solution*

$$a_0 = a_1 = a_2 = \ldots = 1/c.$$

Proof of Lemma 18. Suppose $\{a_0, a_1, a_2, \ldots\}$ is a solution to system (156). Then $\{1, a_1/a_0, a_2/a_0, \ldots\}$ is also a solution, since each equation only includes multinomial terms of the same degree. Thus, without loss of generality we assume $a_0 = 1$, and we need to prove that

$$a_1 = a_2 = \ldots = 1.$$

We consider two cases.

Case I. Suppose the sequence $\{a_j\}$ has a maximum: there exists an index $i \in \mathbb{N}$ such that $a_i = \max_{j \in \mathbb{N}} a_j$. Define

$$w_{j,\ell} := \frac{1}{2^j} \prod_{k=j}^{\ell+j-1} a_k \left[\prod_{k=0}^{\ell-1} a_k\right]^{-1}.$$

Using $n = \ell$ in (156) we obtain that for any $\ell \in \mathbb{N}$,

$$\sum_{j=1}^{\infty} w_{j,\ell} = 1, \tag{157}$$

and using $n = \ell + 1$ we find that an arbitrary a_ℓ is the weighted average of $\{a_{\ell+j}\}_{j=1,2,\ldots}$:

$$\sum_{j=1}^{\infty} w_{j,\ell}\, a_{\ell+j} = a_\ell. \tag{158}$$

Hence, since $a_i = \max_{j \in \mathbb{N}} a_j$,

$$a_i = a_{i+1} = a_{i+2} = a_{i+3} = \ldots = a.$$

Similarly, letting $\ell = i - 1$ in (157) and (158), we obtain $a_{i-1} = a$. Recursively, by plugging in $\ell = i - 2,\ i - 3, \ldots$, we show that

$$a_1 = a_2 = \ldots = a_{i-1} = a_i = a_{i+1} = \ldots = a.$$

Finally, $\frac{1}{2}a_1 + \frac{1}{4}a_2 + \frac{1}{8}a_3 + \ldots = 1$ implies $a = 1$.

Case II. Suppose there is no $\max_{j \in \mathbb{N}} a_j$. Let $U := \limsup_{j \to \infty} a_j$. From (156) we know via cancelation that

$$\frac{1}{2}a_n + \frac{1}{4}\frac{a_n a_{n+1}}{a_1} \quad + \quad \frac{1}{8}\frac{a_n a_{n+1} a_{n+2}}{a_1 a_2} + \ldots$$

$$+ \quad \frac{1}{2^{n-1}}\frac{\prod\limits_{k=n}^{2n-2} a_k}{\prod\limits_{k=0}^{n-2} a_k} + \sum_{j=n}^{\infty}\frac{1}{2^j}\frac{\prod\limits_{k=j}^{n+j-1} a_k}{\prod\limits_{k=0}^{n-1} a_k} = 1. \tag{159}$$

Thus, $2^{-1}\, a_n < 1$ and $U \le 2$. The absence of maximum implies $a_j < U \le 2$ for all $j \in \mathbb{N}$.

Plugging $n + 1$ in (156), we obtain

$$\left(\frac{1}{2}a_n\right) a_{n+1} + \left(\frac{1}{4}\frac{a_n a_{n+1}}{a_1}\right) a_{n+2} + \ldots$$

$$+ \quad \frac{1}{2^{n-1}}\frac{\prod\limits_{k=n}^{2n-2} a_k}{\prod\limits_{k=0}^{n-2} a_k} a_{n+j-1} + \sum_{j=n}^{\infty}\frac{1}{2^j}\frac{\prod\limits_{k=j}^{n+j-1} a_k}{\prod\limits_{k=0}^{n-1} a_k} a_{n+j} = a_n.$$

Thus, since $a_j < U$ for all $j \in \mathbb{N}$,

$$\left(\frac{1}{2}a_n\right) a_{n+1} + \left(\frac{1}{4}\frac{a_n a_{n+1}}{a_1}\right) U + \ldots$$

$$+ \quad \frac{1}{2^{n-1}}\frac{\prod\limits_{k=n}^{2n-2} a_k}{\prod\limits_{k=0}^{n-2} a_k} U + \sum_{j=n}^{\infty}\frac{1}{2^j}\frac{\prod\limits_{k=j}^{n+j-1} a_k}{\prod\limits_{k=0}^{n-1} a_k} U \quad > a_n$$

which simplifies via (159) to

$$\left(\frac{a_n}{2}\right) a_{n+1} + \left(1 - \frac{a_n}{2}\right) U > a_n. \tag{160}$$

For all $\varepsilon \in (0,1)$, there are infinitely many $n \in \mathbb{N}$ such that $a_n > (1-\varepsilon)U$. Then, for any such n, the above inequality (160) implies

$$a_{n+1} > 2 - \frac{2}{a_n}U + U > 2 - \frac{2\varepsilon}{1-\varepsilon} + U = \left(1 - \varphi(\varepsilon)\right)U,$$

where

$$\varphi(x) := \frac{2x}{(1-x)U}.$$

Let $\varphi^{(k)} = \varphi \circ \ldots \circ \varphi$. Repeating the argument for any given number of iterations $K \in \mathbb{N}$, we obtain

$$a_{n+2} > \left(1 - \varphi^{(2)}(\varepsilon)\right)U, \quad a_{n+3} > \left(1 - \varphi^{(3)}(\varepsilon)\right)U, \ldots, \quad a_{n+K} > \left(1 - \varphi^{(K)}(\varepsilon)\right)U.$$

Thus, given any $K \in \mathbb{N}$, fix $\varepsilon \in (0,1)$ small enough so that such that $\varphi^{(k)}(\varepsilon) \in (0,1)$ for all $k = 1, 2, \ldots, K$. Then, taking $n > K$ such that $a_n > (1-\varepsilon)U$, we obtain from (159) that

$$1 > \frac{1}{2}a_n + \frac{1}{4}\frac{a_n a_{n+1}}{a_1} + \frac{1}{8}\frac{a_n a_{n+1} a_{n+2}}{a_1 a_2} + \ldots + \frac{1}{2^{K+1}}\frac{\prod_{k=n}^{n+K} a_k}{\prod_{k=0}^{K} a_k}$$

$$> \frac{1}{2}(1-\varepsilon)U + \frac{1}{4}\frac{(1-\varepsilon)\left(1-\varphi(\varepsilon)\right)U^2}{U} + \ldots + \frac{1}{2^{K+1}}\frac{U^{K+1}\prod_{k=0}^{K}\left(1 - \varphi^{(k)}(\varepsilon)\right)}{U^K}.$$

Now, since ε can be chosen arbitrarily small,

$$1 \geq \left(1 - \frac{1}{2^{K+1}}\right) U.$$

Finally, since K can be selected arbitrarily large, we have proven that $1 \geq U$. However, this will contradict the assumption of *Case II*. Indeed, if $a_j < U \leq 1$ for all $j \in \mathbb{N}$, then

$$\frac{1}{2}a_1 + \frac{1}{4}a_2 + \frac{1}{8}a_3 + \ldots < 1,$$

contradicting the first equation in the statement of the theorem. Thus, the assumptions of *Case II* cannot be satisfied. We conclude that there exists a maximal element in the sequence $\{a_j\}_{j=1,2,\ldots}$ as assumed in *Case I*, implying the statement of the theorem. □

Proof of Theorem 17. Lemma 18 implies $a_k = S_k/S_{k+1} = 1/c$ for some $c \geq 1$. Hence $S_1 = 1 + T_1 = c$ and $T_1 = c - 1$. Furthermore,

$$S_{k+1} = c\, S_k = c^k$$

(a) Tree T (b) Harris path H_T
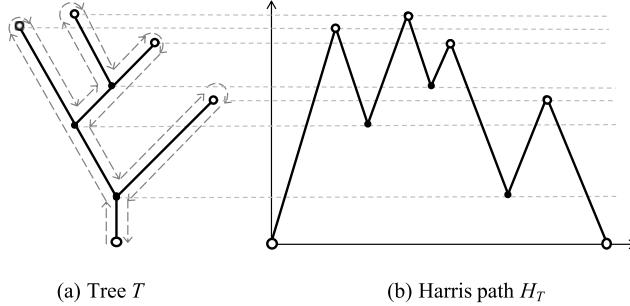
FIG 24. *(a) Tree T and its depth-first search illustrated by dashed arrows. (b) Harris path $H_T(t)$ for the tree T of panel (a). In this figure, the distances on a tree (edge lengths) are measured along the y-axis. Dashed horizontal lines illustrate correspondence between vertices of T and local extrema of $H_T(t)$.*

and, accordingly,

$$T_{k+1} = S_{k+1} - S_k = (c-1)c^{k-1},$$

which completes the proof. □

8. Tree representation of continuous functions

We review here the results of [105, 124, 135, 173] on tree representation of continuous functions. This representation allows us to apply the self-similarity concepts to time series.

8.1. Harris path

For any embedded tree $T \in \mathcal{L}_{\mathrm{plane}}$ with edge lengths, the *Harris path* (also known as the contour function, or Dyck path) is defined as a piece-wise linear function [80, 135]

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

that equals the distance from the root traveled along the tree T in the depth-first search, as illustrated in Fig. 24. 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 ± 1.

8.2. Level set tree

This section introduces a tree representation of continuous functions, which we call a *level set tree*. The following "great flood" analogy provides intuition behind our construction. One can think of a continuous function $f(x)$ as a landscape elevation profile. Imagine a flood that covers the entire region with water. When the water starts to recess, some isolated islands appear. With

(a) Function $f(x)$ (b) Tree LEVEL(f)
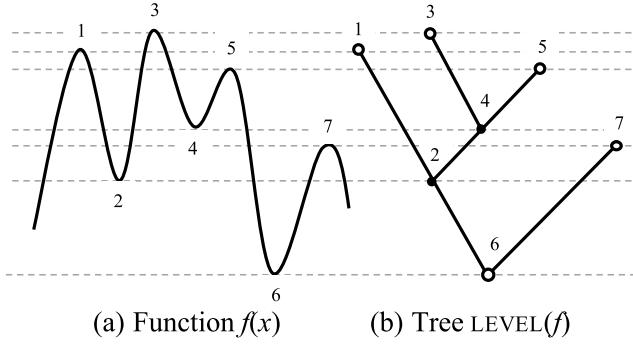
FIG 25. *Function $f(x)$ (panel a) with a finite number of local extrema and its level set tree* LEVEL(f) *(panel b). In this figure, the distances on a tree (edge lengths) are measured along the y-axis. Indices $1, \ldots, 7$ illustrate correspondence between the local extrema of $f(x)$ and vertices of* LEVEL(f). *Dashed horizontal lines show the thresholds α where the number of distinct intervals that comprise $\mathcal{L}_\alpha$ changes. Each such threshold corresponds to either a local maximum of $f(x)$ and leaf of* LEVEL(f) *(indices $1, 3, 5, 7$) or to a local minimum of $f(x)$ and an internal vertex of* LEVEL(f) *(indices $2, 4, 6$). All other threshold values affect the lengths of the intervals that comprise $\mathcal{L}_\alpha$ without affecting the number of distinct intervals.*

time, small islands merge to form bigger islands. Eventually, the water level goes low enough to result in a single connected dry continent. The level set tree describes the process of continent formation via island merging. It represents the appearance of an island as a leaf, growth of an island as an edge, and merging of two islands as an internal vertex.

Below we suggest a formal construction. We begin in Sect. 8.2.1 by assuming a finite number of local extrema; this case is more intuitive and is sufficient for analysis of finite trees from $\mathcal{L}_{\text{plane}}$. A general definition for continuous functions follows in Sect. 8.2.2.

8.2.1. Tamed functions: finite number of local extrema

Consider a closed interval $I \subset \mathbb{R}$ and function $f(x) \in C(I)$, where $C(I)$ is the space of continuous functions from I to $\mathbb{R}$. Suppose that $f(x)$ has a finite number of distinct local minima. The *level set* $\mathcal{L}_\alpha(f)$ is defined as the preimage of the function values equal to or above α:

$$\mathcal{L}_\alpha = \mathcal{L}_\alpha(f) = \{x \in I \,:\, f(x) \geq \alpha\}.$$

The level set $\mathcal{L}_\alpha$ for each α 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 $f(x)$ and $0 \leq |\mathcal{L}_\alpha| \leq n$, where n is the total number of the local maxima of $f(x)$ over I.

The *level set tree* LEVEL(f) $\in \mathcal{L}_{\text{plane}}$ is a tree that describes the structure of the level sets $\mathcal{L}_\alpha$ as a function of threshold α, as illustrated in Fig. 25. Specifically, there are bijections between
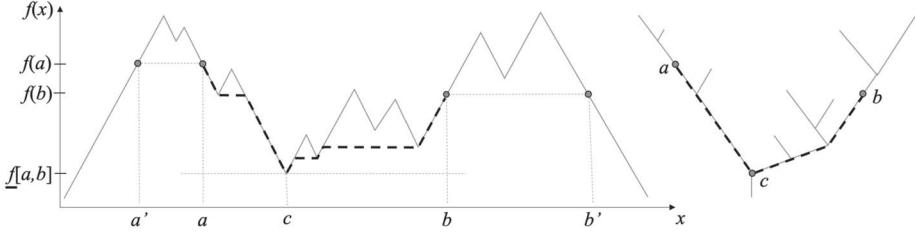
FIG 26. *Tree metric d_f on a real interval I defined by a function $f(x)$. (Left panel): The graph of a function $f(x)$, $x \in I$ is shown by solid gray line. The distance $d_f(a, b)$ between points $a, b \in I$ is given by (161); it equals the vertical distance along the path between $f(a)$ and $f(b)$ shown by black dashed line. This path coincides with graph of the function $\underline{f}(x)$ of (162). The panel also illustrates equivalence in metric d_f: here $a \sim_f a'$ and $b \sim_f b'$, since $d_f(a, a') = d_f(b, b') = 0$. (Right panel): The level set tree $\text{LEVEL}(f)$ of function $f(x)$ is shown by solid gray line. The distance $d_f(a, b)$ equals the length of the minimal tree path between points a and b, shown by dashed black line. Here, the tree distance is measured along the vertical axis.*

- **(i)** the leaves of $\text{LEVEL}(f)$ and the local maxima of $f(x)$;
- **(ii)** the internal (parental) vertices of $\text{LEVEL}(f)$ and the local minima of $f(x)$, excluding possible local minima achieved on the boundary ∂I;
- **(iii)** a pair of subtrees of $\text{LEVEL}(f)$ rooted in the parental vertex that corresponds to a local minima $f(x^*)$ and the adjacent positive excursions (or meanders bounded by ∂I) of $f(x) - f(x^*)$ to the right and left of x^*.

Furthermore, every edge in the tree is assigned a length equal to the difference of the values of $f(x)$ at the local extrema that correspond to the vertices adjacent to this edge according to the bijections (i) and (ii) above. The tree root corresponds to the global minimum of $f(x)$ on I. If the minimum is achieved at $x \in I \setminus \partial I$, then the level set tree is stemless, $\text{LEVEL}(f) \in \mathcal{L}_{\text{plane}}^{\vee}$; this case is shown in Fig. 25. Otherwise, if the minimum is on the boundary ∂I, then the level set tree is planted, $\text{LEVEL}(f) \in \mathcal{L}_{\text{plane}}^{|}$.

8.2.2. General case

For a function $f(x) \in C(I)$ on a closed interval $I \subset \mathbb{R}$, the level set tree is defined via the framework of Def. 1, following Aldous [4, 5] and Pitman [135]. Specifically, let $\underline{f}[a, b] := \inf_{x \in [a,b]} f(x)$ for any subinterval $[a, b] \subset I$. We define a *pseudo-metric* on I as [5, 135]

$$d_f(a, b) := \left(f(a) - \underline{f}[a, b]\right) + \left(f(b) - \underline{f}[a, b]\right), \quad a, b \in I. \tag{161}$$

We write $a \sim_f b$ if $d_f(a, b) = 0$. Here d_f is a metric on the quotient space $I_f \equiv I/\sim_f$. It can be shown [135] that (I_f, d_f) is a tree by Def. 1. Figure 26 illustrates this construction for a particular piece-wise function (left panel), and shows the respective tree (I_f, d_f) as an element of $\mathcal{L}_{\text{plane}}^{|}$ (right panel).

We describe now the unique path $\sigma_{a,b} \subset I_f$ between a pair of points a, b. Let $c \in [a, b]$ be the leftmost point where $f(x)$ achieves the minimum on $[a, b]$:

$$c = \min\{x \in [a, b] : f(x) = \underline{f}[a, b]\}.$$

We define a function $\underline{f}(x)$ on $[a, b]$ as

$$\underline{f}(x) = \begin{cases} \inf_{y \in [a,x]} f(y), & \text{if } x \in [a, c], \\ \inf_{y \in [x,b]} f(y), & \text{if } x \in [c, b]. \end{cases} \tag{162}$$

By construction, $\underline{f}(x)$ is a continuous function that is monotone non-increasing on $[a, c]$ and monotone nondecreasing on $[c, b]$. Furthermore, $\underline{f}(x) \le f(x)$ and, in particular, $\underline{f}(x) = f(x)$ for $x \in \{a, b, c\}$. An example of $\overline{f}(x)$ is shown in Fig. 26 (left panel, black dashed line).

Lemma 19 (Rising Sun Lemma, F. Riesz [138]). *Let*

$$S = \{x : \underline{f}(x) < f(x)\} \subset [a, b].$$

Then S is an open set that can be represented as a countable union of disjoint intervals

$$S = \bigcup_k (a_k, b_k)$$

such that $f(a_k) = f(b_k) = \underline{f}(a_k) = \underline{f}(b_k)$ and $f(x) > f(a_k)$ for any $x \in (a_k, b_k)$.

Proof. The statement is equivalent to that of the *Rising Sun Lemma* of Riesz [138, 152] applied to the functions $-f(x)$ on $[c, b]$ and $-f(-x)$ on $[a, c]$. We just notice that $f(c)$ is the global minimum of $f(x)$ on $[a, b]$ and so c cannot be a part of S. The union of two open sets, each represented as a countable union of disjoint intervals, is itself an open interval represented as a countable union of disjoint intervals. This completes the proof. □

Figure 27 illustrates the Rising Sun Lemma for the function $f(x)$ shown in Fig. 26 on the interval $[c, b]$. As the sun rises from east (right), it shades light (dashed gray arrows) on the those segments of the graph that satisfy $\underline{f}(x) = f(x)$, and leaves in shade the segments that satisfy $\underline{f}(x) < f(x)$. The preimage of the shadowed segments is the set S, while the preimage of the lighted segments is the path $\sigma_{c,b}$. The path, considered as a set in $[c, b]$, is making at most a countable number of jumps over the intervals (a_k, b_k) that comprise the set S of Lem. 19.

For a tamed function with a finite number of local extrema, the path $\sigma_{a,b}$ is the preimage of the graph of $\underline{f}(x)$ excluding the constant intervals. The Rising Sun Lemma generalizes this statement to any continuous function. Here, the path $\sigma_{a,b}$ is given by

$$\sigma_{a,b} = [a, b] \setminus S = \{x : \underline{f}(x) = f(x)\} \subset [a, b],$$

and is travelled at unit speed left to right. As a real set, the path $\sigma_{a,b}$ may have quite complicated structure. For instance, it can be the Cantor set. This, however, does not disturb the continuity of the map $[0, d_f(a, b)] \to I_f$ in Def. 1.
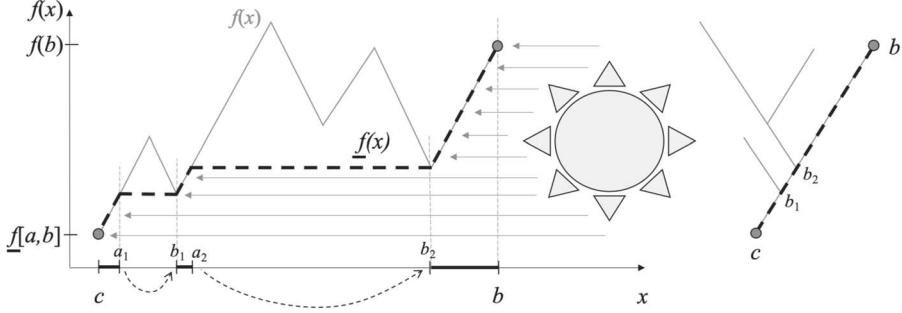
FIG 27. *Rising Sun Lemma: an illustration. The left panel shows the graph of the function $f(x)$ of Fig. 26 on the interval $[c, b]$ (solid gray line). Dashed black line shows the function $\underline{f}(x)$. The sun light falls from the right (gray arrows). The preimage of the lighted segments of $\overline{f}(x)$ (defined by $\underline{f}(x) = f(x)$) consists of three intervals, $[c, a_1] \cup [b_1, a_2] \cup [b_2, b]$. It comprises the path $\sigma_{c,b}$ (solid black segments on the x-axis). The preimage of the shadowed segments of $f(x)$ (defined by $\underline{f}(x) < f(x)$) consists of two intervals, $(a_1, b_1) \cup (a_2, b_2)$. It comprises the set S of Lem. 19. The path $\sigma_{c,b}$ jumps over the intervals (a_k, b_k) that form the set S, as shown by dashed arrows, using the equivalence of the interval ends, $a_k \sim_f b_k$. The right panel shows the tree that corresponds to $f(x)$ on $[c, b]$. The path $\sigma_{c,b}$ is shown by dashed black line. Every interval (a_k, b_k) corresponds to a tree attached to the path $\sigma_{c,b}$ at the vertex corresponding to b_k.*

The Rising Sun Lemma asserts that the function $\underline{f}(x)$ on $[a, b]$ has at most a countable set of constant disjoint intervals $I^{(k)} = (a_k, b_k)$, each of which corresponds to a positive excursion of $f(x)$. The end points of these intervals are equivalent in I_f, hence each interval generates a tree $(I_f^{(k)}, d_f)$ whose root corresponds to the equivalence class on I consisting of $\{a_k, b_k\}$. This observations leads to the following statement.

Corollary 9. *The level set tree* LEVEL(f) *of a continuous function $f(x)$ on a real closed interval $[a, b] \subset \mathbb{R}$ consists of a segment of length $d_f(a, b)$ and at most a countable number of trees attached to this segment with the same orientation. There is a one-to-one correspondence between these trees and the intervals (a_k, b_k) from the Rising Sun Lemma.*

It is straightforward to observe that the tree (I_f, d_f) is equivalent to the above defined level set tree LEVEL(f) for a function $f(x) \in C(I)$ with a finite number of distinct local minima. We just notice that for any subinterval $[a, b] \subset I$, the correspondence $a \sim_f b$ implies $\{f(x) : x \in [a, b]\}$ is a nonnegative excursion i.e.,

$$[a, b] \subset \mathcal{L}_\alpha(f) = \{x : f(x) \geq \alpha\} \quad \text{where} \quad \alpha = f(a) = f(b).$$

In other words, every point in (I_f, d_f) is an equivalence class of points on I with respect to $\sim_f$. There exist three types of equivalence classes, depending on the number of distinct points from I they include: (i) each single point class corresponds to a leaf vertex (local maximum), (ii) each two point class corresponds to an internal edge point (positive excursion), and (iii) each three

point class corresponds to an interval vertex (two adjacent positive excursions). For a general $f(x) \in C(I)$ there may exist equivalence classes that include an arbitrary number n of points from I, corresponding to $(n-1)$ adjacent positive excursions; and classes that consist of an infinite (countable or uncountable) number of points. Conversely, for every α, the level set $\mathcal{L}_\alpha(f)$ is a union of non-overlapping intervals $[a_j, b_j]$, i.e.,

$$\mathcal{L}_\alpha(f) = \bigcup_j [a_j, b_j],$$

where for each j, $a_j \sim_f b_j$.

Representing level sets of a continuous function as a tree goes back to works of Menger [117] and Kronrod [92]. A multivariate analog of level set tree is among the key tools in proving the celebrated Kolmogorov-Arnold representation theorem (every multivariate continuous function can be represented as a superposition of continuous functions of two variables) that gives a positive answer to a general version of the Hilbert's thirteenth problem [9, 163]. Such trees have also been discussed by Vladimir Arnold in connection to topological classification of Morse functions and generalizations of Hilbert's sixteenth problem [10, 11]. Level set trees for multivariate Morse functions (albeit slightly different from those considered by Arnold) are discussed in Sect. 8.9.

8.3. Reciprocity of Harris path and level set tree

Consider a function $f(x) \in C(I)$ with a finite number of distinct local minima. By construction, the level set tree LEVEL(f) is completely determined by the sequence of the values of local extrema of f, and is not affected by timing of those extrema, as soon as their order is preserved. This means, for instance, that if $g(x)$ is a continuous and monotone increasing function on I, then the trees LEVEL(f) and LEVEL$(f \circ g)$ are equivalent in $\mathcal{L}_{\text{plane}}$. Hence, without loss of generality we can focus on the level set trees of continuous functions with alternating slopes ± 1. We write $\mathcal{E}^{\text{ex}}$ for the space of all positive piece-wise linear continuous finite excursions with alternating slopes ± 1 and a finite number of segments (i.e., a finite number of local extrema).

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

This statement is readily verified by examining the excursions and trees in Figs. 24, 26.
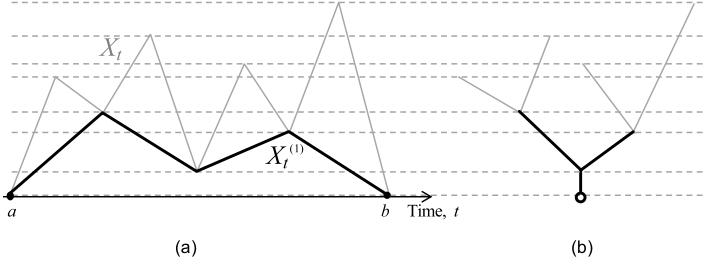
FIG 28. *Horton pruning of a positive excursion: transition to the local minima of an excursion X_t corresponds to pruning of the corresponding level set tree. (a) An original excursion X_t (gray line) and linearly interpolated sequence $X_t^{(1)}$ of the respective local minima (black line). (b) The level set tree* LEVEL$(X_t^{(1)})$ *of the local minima sequence (black lines) is obtained by pruning of the level set tree* 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.*

8.4. Horton pruning of positive excursions

This section examines the level set tree and its Horton pruning for a positive excursion on a finite real interval. We use these results for analysis of random walks X_k, $k \in \mathbb{Z}$, which motivates us to write here X_t, $t \in \mathbb{R}$, for a continuous function.

Consider a continuous positive excursion X_t, $t \in [a, b]$, with a finite number of distinct 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. 28a).

In the space of level set trees of tamed continuous functions, the Horton pruning $\mathcal{R}$ corresponds to coarsening the respective function by removing (smoothing) its local maxima, as illustrated in Fig. 28. An iterative pruning corresponds to iterative transition to the local minima, as we describe in the next statement.

Proposition 15 (Horton pruning of positive excursions, [173]). *The transition from a positive excursion X_t to the respective excursion $X_t^{(1)}$ of its local minima corresponds to the Horton pruning of the level set tree* LEVEL(X_t). *This is illustrated in a diagram of Fig. 29. In general,*

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

Proof. First,

$$\text{LEVEL}\left(X_t^{(1)}\right) = \mathcal{R}\left(\text{LEVEL}(X_t)\right) \tag{163}$$
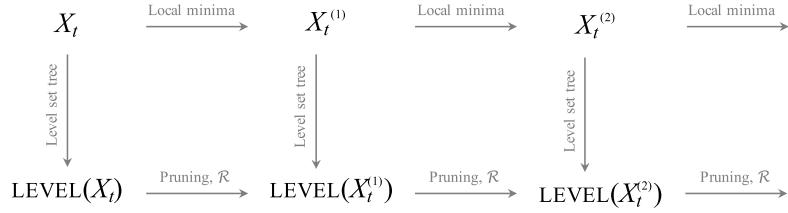
Fig 29. *Transition to the local minima of a function X_t corresponds to the Horton pruning $\mathcal{R}$ of the corresponding level set tree* LEVEL(X_t).

is established via the following observation. For a pair of consecutive local minima $s_1 < s_2$, the level set tree LEVEL$(\widetilde{X}_t)$ of the function

$$\widetilde{X}_t = X_t \mathbf{1}_{t \notin [s_1, s_2]} + \left(\frac{s_2 - t}{s_2 - s_1} X_{s_1} + \frac{t - s_1}{s_2 - s_1} X_{s_2} \right) \mathbf{1}_{t \in [s_1, s_2]}$$

is obtained from LEVEL (X_t) by removing the leaf that corresponds to the unique local maximum of X_t inside (s_1, s_2) together with its parental edge that connects it to the parental vertex, corresponding to $\max\{X_{s_1}, X_{s_2}\}$. Thus, substituting X_t with linear interpolation of local minima, $X_t^{(1)}$, will result in simultaneous removal of leaves together with the parental edges. The statement of the proposition follows via recursion of (163). □

It is straightforward to formulate an analog of Prop. 15 without the excursion assumption, for continuous functions with a finite number of distinct local minima within $[a, b]$.

8.5. *Excursion of a symmetric random walk*

We turn now to random walks $\{X_k\}_{k \in \mathbb{Z}}$. Linear interpolation of their trajectories corresponds to the tamed continuous functions. A random walk $\{X_k\}_{k \in \mathbb{Z}}$ with a transition kernel $p(x, y)$ is called *homogeneous* if $p(x, y) \equiv p(y - x)$ for any $x, y \in \mathbb{R}$. A homogeneous random walk is *symmetric* if $p(x) = p(-x)$ for all $x \in \mathbb{R}$. The transition kernel of a symmetric random walk can be represented as the even part of a p.d.f. $f(x)$ with support $\mathsf{supp}(f) \subseteq \mathbb{R}_+$:

$$p(x) = \frac{f(x) + f(-x)}{2}. \tag{164}$$

We assume that $p(x)$, and hence $f(x)$, is an atomless density function.

We write $\{X_k^{(1)}\}_{k \in \mathbb{Z}}$ for the sequence of local minima of $\{X_k\}_{k \in \mathbb{Z}}$, listed in the order of occurrence, from left to right. In particular, we set $X_0^{(1)}$ to be the value of the leftmost local minima of X_k for $k \geq 0$. Recursively, we let $\{X_k^{(j+1)}\}_{k \in \mathbb{Z}}$ denote the sequence of local minima of $\{X_k^{(j)}\}_{k \in \mathbb{Z}}$.

Lemma 20 (**Local minima of random walks**, [**173**]). *The following statements hold.*

(i) *The sequence of local minima $\{X_k^{(1)}\}_{k \in \mathbb{Z}}$ of a homogeneous random walk $\{X_k\}_{k \in \mathbb{Z}}$ is itself a homogeneous random walk.*

(ii) *The sequence of local minima $\{X_k^{(1)}\}_{k \in \mathbb{Z}}$ of a symmetric homogeneous random walk $\{X_k\}_{k \in \mathbb{Z}}$ is itself a symmetric homogeneous random walk.*

Proof. Let $d_j = X_{j+1}^{(1)} - X_j^{(1)}$. We have, for each $j \in \mathbb{Z}$,

$$d_j = \sum_{i=1}^{\xi_+} Y_i - \sum_{i=1}^{\xi_-} Z_i, \tag{165}$$

where the first sum corresponds to ξ_+ positive increments of X_k between a local minimum $X_j^{(1)}$ and the subsequent local maximum m_j, and the second sum corresponds to ξ_- negative increments between the local maximum m_j and the subsequent local minimum $X_{j+1}^{(1)}$.

Recall that $\mathsf{Geom}_1(p)$ denotes the geometric distribution with p.m.f. $p(m) = p(1-p)^{m-1}$, $m \in \mathbb{N}$. Accordingly, ξ_+ and ξ_- are independent geometric random variables

$$\xi_+ \stackrel{d}{\sim} \mathsf{Geom}_1(p^+), \quad \xi_- \stackrel{d}{\sim} \mathsf{Geom}_1(p^-)$$

with parameters, respectively,

$$p^+ = \int_0^\infty p(x)\,dx \quad \text{and} \quad p^- = \int_{-\infty}^0 p(x)\,dx,$$

and Y_i, Z_i are i.i.d. positive continuous random variables with p.d.f.s, respectively,

$$f_Y(x) = \frac{p(x)\mathbf{1}_{\{x \geq 0\}}}{p^+} \quad \text{and} \quad f_Z(x) = \frac{p(-x)\mathbf{1}_{\{x \leq 0\}}}{p^-}.$$

(i) By independence of increments of a random walk, the random jumps d_j have the same distribution for each j. This establishes the statement.

(ii) For the kernel of a symmetric random walk, we have representation (164). In this case, ξ_+ and ξ_- are independent geometric random variables with parameters $p^+ = p^- = 1/2$ and Y_i, Z_i are i.i.d. positive continuous random variables with p.d.f. $f(x)$. Hence, both sums in (165) have the same distribution, and their difference has a symmetric distribution. Thus $\{X_j^{(1)}\}_{j \in \mathbb{Z}}$ is a symmetric homogeneous random walk. $\qquad\square$

We notice that the symmetric kernel $p^{(1)}(x)$ for the chain of local minima $\{X_j^{(1)}\}_{j \in \mathbb{Z}}$ is necessarily different from $p(x)$ in both parts of Lemma 20. Hence, the random walk $\{X_j^{(1)}\}$ of local minima is always different from the initial random walk $\{X_k\}$. In a symmetric case, however, both the processes happen to be closely related in terms of the structure of their level set trees. Now we explore this relation.
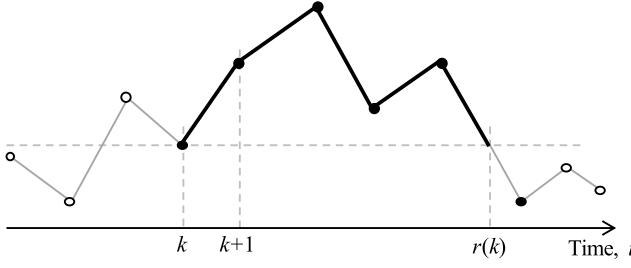
FIG 30. *Excursion of a symmetric homogeneous random walk: illustration. The values of time series* X_k, $k \in \mathbb{Z}$, *are shown by circles; the circles that form the excursion are filled. The linear interpolation function* X_t, $t \in \mathbb{R}$, *is shown by solid line; the excursion of* X_t *on* $[k, r(k)]$ *is shown in bold. This is the first positive excursion of* X_t *to the right of* k.

Consider linear interpolation $\{X_t\}_{t \in \mathbb{R}}$ of a symmetric homogeneous random walk $\{X_k\}_{k \in \mathbb{Z}}$ with an atomless transition kernel $p(x)$. For any $k \in \mathbb{Z}$, let

$$T^{\mathrm{ex}} = T^{\mathrm{ex}}(X_t, k) \in \mathcal{BL}^{|}_{\mathrm{plane}}$$

be the level set tree of the first positive excursion of $X_t - X_k$ to the right of k, with convention $T^{\mathrm{ex}} = \phi$ if $X_{k+1} - X_k < 0$. Formally, let $r = r(k) \in \mathbb{R}$ be the unique epoch such that (Fig. 30)

$$r \geq k, \quad X_t > X_k \text{ for all } t \in (k, r), \quad \text{and } X_r = X_k.$$

The epoch $r(k)$ is almost surely finite, as can be demonstrated by a renewal argument using the symmetry of the increments of X_k. We define

$$T^{\mathrm{ex}}(X_t, k) := \mathrm{LEVEL}(X_t, t \in [k, r(k)]).$$

It follows from this definition that

$$\{X_{k+1} - X_k > 0\} \Leftrightarrow \{T^{\mathrm{ex}}(X_t, k) \neq \phi\}.$$

The basic properties of symmetric homogeneous random walks imply that the distribution of $T^{\mathrm{ex}}(X_t, k)$ is the same for all points $k \in \mathbb{Z}$. This justifies the following definition.

Definition 28 (**Positive and nonnegative excursions**). *In the above setup, we call process* X_t^{ex} *a nonnegative excursion of the linearly interpolated symmetric homogeneous random walk* $\{X_t\}_{t \in \mathbb{R}}$ *if*

$$X_t^{\mathrm{ex}} \overset{d}{=} \{X_{s-k} - X_k, \ s \in [k, r(k)]\} \quad \text{for any } k \in \mathbb{Z}.$$

Furthermore, we call process X_t^{ex} *a positive excursion of the linearly interpolated symmetric homogeneous random walk* $\{X_t\}_{t \in \mathbb{R}}$ *if*

$$X_t^{\mathrm{ex}} \overset{d}{=} \{X_{s-k} - X_k, \ s \in [k, r(k)] \ \big| X_{k+1} - X_k > 0\} \quad \text{for any } k \in \mathbb{Z}.$$

A positive excursion defined above will also be called a *positive right excursion*. The corresponding positive excursion in the reversed time order, starting from k and going in the negative time direction, will be called *positive left excursion*. According to Def. 28, a nonnegative excursion may consist of a single point (if $r(k) = k$), in which case its level set tree is the empty tree. A positive excursion necessarily includes at least one positive value, and its level set tree is non-empty.

Consider a homogeneous random walk X_k with a symmetric atomless transition kernel $p(x)$, $x \in \mathbb{R}$, represented as in (164). Note that X_k is time reversible, with $p(x)$ also being the transition kernel of the reversed process. The increment between a pair of consecutive local extrema (a minimum and a maximum) of X_k is a sum of $\mathsf{Geom}_1(1/2)$-distributed number of i.i.d. $f(x)$-distributed random variables, and therefore has density

$$\mathsf{s}(x) := \sum_{k=1}^{\infty} 2^{-k} \underbrace{f * \ldots * f}_{k \text{ times}}(x). \tag{166}$$

We now examine a positive-time process $\{X_k\}_{k \geq 0}$, conditioned on $X_0 = 0$. Consider a sequence of local minima $\left\{X_j^{(1)}\right\}_{j \geq 1}$, where we set $X_0^{(1)} = 0$, and $X_1^{(1)}, X_2^{(1)}, \ldots$ are the local minima of the random walk X_k, listed from left to right. For a positive right excursion X^{ex} originating at $k = 0$, the number N of leaves in the level set tree $\textsc{level}(X^{\mathsf{ex}})$ is determined by the location of the first local minimum below zero:

$$N = \min\{j \geq 1 \,:\, X_j^{(1)} \leq 0\}.$$

The number of edges in the level set tree is $\#\textsc{level}(X^{\mathsf{ex}}) = 2N - 1$. Moreover, let $\kappa > 0$ be the time of the first local minimum below zero, $X_\kappa = X_N^{(1)}$. Next, we define the quantity by which the first nonpositive local minimum of X_k falls below the starting point at zero.

Definition 29 (Extended positive excursion and excess value). *In the above setup, the process $\check{X}^{\mathsf{ex}} = \{X_t\}_{t \in [0,\kappa]}$ is called the extended positive excursion or extended positive right excursion. That is, $\check{X}^{\mathsf{ex}}$ is obtained by extending the excursion X^{ex} until the first local minimum X_κ below the starting value. The quantity $\Lambda(\check{X}^{\mathsf{ex}}) := -X_N^{(1)}$ is called the excess value of the extended excursion. This definition is illustrated in Fig. 31(a).*

The notions of the extended positive excursion and the excess value $\Lambda(\check{X}^{\mathsf{ex}})$ can be expanded to the left and right excursions with arbitrary initial values.

Theorem 18 (Combinatorial excursion tree is critical Galton-Watson). *Suppose X^{ex} is a positive excursion of a homogeneous random walk on $\mathbb{R}$ with a symmetric atomless transition kernel and $T = \textsc{level}(X^{\mathsf{ex}})$. Then, the combinatorial shape of T has the same distribution on $\mathcal{BT}^|$ as the critical binary Galton-Watson tree:*

$$\textsc{shape}(T) \stackrel{d}{\sim} \mathcal{GW}\left(\frac{1}{2}, \frac{1}{2}\right).$$
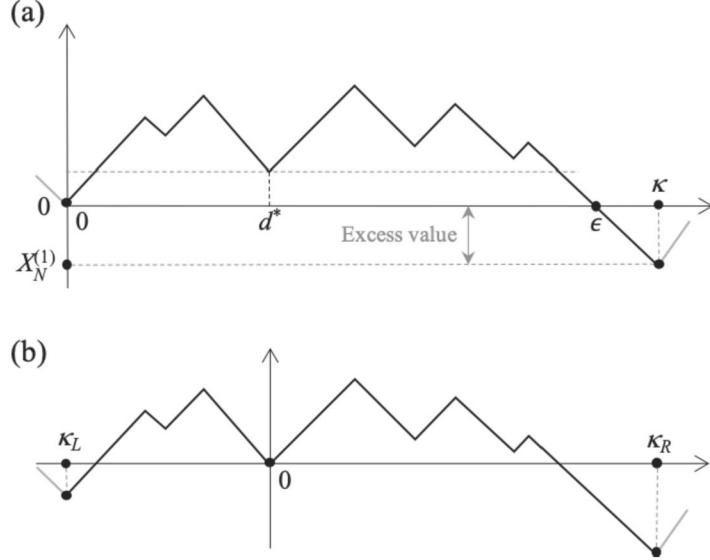
FIG 31. *Extended excursion: An illustration. (a) Extended positive right excursion $\check{X}^{\text{ex}}$ on the interval $[0, \kappa]$. It is obtained by extending the respective positive right excursion X^{ex} on the interval $[0, \varepsilon]$ until the first local minimum X_κ below zero. The panel also illustrates the excess value $-X_N^{(1)}$ and the lowest local minimum of the excursion at epoch d^*. (b) A trajectory from $\mathcal{X}_{LR}$ on the interval $[\kappa_L, \kappa_R]$ consists of a positive left excursion on $[\kappa_L, 0]$ and a positive right excursion on $[0, \kappa_R]$. Observe that the trajectory in panel (b) is obtained by a horizontal and vertical shift of the trajectory in panel (a). The proof of Thm. 18 uses the one-to-one correspondence between extended (left/right) positive excursions with $\text{ord}(T) \geq 2$ of panel (a) and trajectories of panel (b).*

Proof. Recall that $\text{SHAPE}(T)$ is almost surely in $\mathcal{BT}^{|}$. Without loss of generality we consider a positive right excursion X^{ex} originating at $k = 0$, where we set $X_0 = 0$. The tree $\text{SHAPE}(T)$ has exactly one leaf if and only if the first local minimum falls below zero. That is, if the jump from $X_0 = 0$ to the first local maximum is smaller than or equal to the size of the jump from the first local maximum to the consecutive local minimum. The probability of this event is:

$$\mathsf{P}\left(\text{ord}(T) = 1\right) = \int\limits_0^\infty \left(\int\limits_x^\infty \mathsf{s}(y)\, dy\right) \mathsf{s}(x) dx = \frac{1}{2}. \qquad (167)$$

According to the characterization of the critical Galton-Watson distribution $\mathcal{GW}\left(\frac{1}{2}, \frac{1}{2}\right)$ given in Remark 2 of Sect. 2.8, the proof will be complete if we show that conditioned on $\text{ord}(T) \geq 2$, the tree $\text{SHAPE}(T)$ splits into a pair of complete subtrees sampled independently from the same distribution as $\text{SHAPE}(T)$. This step is completed as follows.

Consider the space $\mathcal{X}_L$ of all the trajectories of all extended positive left excursions originating at $X_0 = 0$ and whose level set trees are of Horton-Strahler order ≥ 2. Similarly, consider the space $\mathcal{X}_R$ of all the trajectories of all extended

positive right excursions originating at $X_0 = 0$ and whose level set trees are of Horton-Strahler order ≥ 2. We know from (167) that the probability measure for each of the sets $\mathcal{X}_L$ and $\mathcal{X}_R$ totals $1/2$. Thus, we may consider the union set of left and right extended positive excursions $\mathcal{X}_L \cup \mathcal{X}_R$ and equip it with a new probability measure obtained by gluing together the two respective restrictions of probability measures for the left and the right positive excursions. That is the probability measure over the trajectories in $\mathcal{X}_L \cup \mathcal{X}_R$ when restricted to either $\mathcal{X}_L$ or $\mathcal{X}_R$, will coincide with the respective probability measures for the left and for the right positive excursions, with the total probability adding up to one. Now, since all the left and the right extended positive excursions in $\mathcal{X}_L \cup \mathcal{X}_R$ have Horton-Strahler order ≥ 2, for each $\breve{X}^{\mathsf{ex}} \in \mathcal{X}_L \cup \mathcal{X}_R$ there is almost surely a unique integer d^* such that $\breve{X}^{\mathsf{ex}}_{d^*} > 0$ is the smallest local minimum of the excursion $\breve{X}^{\mathsf{ex}}$.

Next, conditioning on $X_0 = 0$ being a local minimum of X_t, we consider a space $\mathcal{X}_{LR}$ of all possible trajectories such that each trajectory consists of the left and the right extended positive excursions originating from $X_0 = 0$ (with no restrictions on their orders). For a trajectory in $\mathcal{X}_{LR}$, let $\kappa_L < 0$ and $\kappa_R > 0$ denote the (random) endpoints of the left and the right extended positive excursions. Thus, a trajectory X_t, $t \in [\kappa_L, \kappa_R]$, in $\mathcal{X}_{LR}$ consists of a left extended positive excursion X_t ($\kappa_L \leq t \leq 0$) and a right extended positive excursion X_t ($0 \leq t \leq \kappa_R$). This construction is illustrated in Fig. 31(b). The probability measure over the space $\mathcal{X}_{LR}$ is a product measure of the left and the right positive excursions. We claim that there exists a bijective measure preserving shift map

$$\Psi : \mathcal{X}_{LR} \to \mathcal{X}_L \cup \mathcal{X}_R.$$

Indeed, if the excess value $\Lambda\big(\{X_t\}_{\kappa_L \leq t \leq 0}\big) = -X_{\kappa_L}$ for the left excursion is smaller than the excess value $\Lambda\big(\{X_t\}_{\kappa_L \leq t \leq 0}\big) = -X_{\kappa_R}$ for the right excursion, we set

$$\Psi\big(\{X_t\}_{\kappa_L \leq t \leq \kappa_R}\big) = \{X_{t+\kappa_L} - X_{\kappa_L}\}_{0 \leq t \leq -\kappa_L + \kappa_R} \ \in \mathcal{X}_R.$$

Otherwise, we set

$$\Psi\big(\{X_t\}_{\kappa_L \leq t \leq \kappa_R}\big) = \{X_{t+\kappa_R} - X_{\kappa_R}\}_{\kappa_L - \kappa_R \leq t \leq 0} \ \in \mathcal{X}_L.$$

The map Ψ is one-to-one onto as it consists of the vertical and the horizontal shifts. Also observe that under the mapping Ψ, the point $(0,0)$ of a trajectory in $\mathcal{X}_{LR}$ is sent to the point $(d^*, \breve{X}^{\mathsf{ex}}_{d^*})$ of the image trajectory in $\mathcal{X}_L \cup \mathcal{X}_R$. We can construct $\Psi^{-1} : \mathcal{X}_L \cup \mathcal{X}_R \to \mathcal{X}_{LR}$ accordingly as a map that shifts a trajectory $\breve{X}^{\mathsf{ex}}$ in $\mathcal{X}_L \cup \mathcal{X}_R$ by subtracting $(d^*, \breve{X}^{\mathsf{ex}}_{d^*})$. Finally, because we take the same product of the transition kernel values $\mathsf{s}(x)$ for the increments of a trajectory in $\mathcal{X}_{LR}$ as for its image in $\mathcal{X}_L \cup \mathcal{X}_R$ under the one-to-one mapping Ψ, the mapping Ψ is measure preserving.

Thus, since vertical and horizontal shifts of a function preserve its level set tree, we conclude that the distribution of the level set trees for the trajectories in $\mathcal{X}_{LR}$ and the trajectories in $\mathcal{X}_L \cup \mathcal{X}_R$ coincide. The level set tree for a trajectory in $\mathcal{X}_{LR}$ consists of a stem that branches into two level set trees of the left and

right positive excursions adjacent to $X_0 = 0$, sampled independently from the same distribution as $\textsc{shape}(T)$. This is so since for the trajectories in $\mathcal{X}_{LR}$, $X_0 = 0$ is the smallest local minimum. Finally, we observe that the distribution of $\textsc{shape}\big(\textsc{level}(\check{X}^{\mathsf{ex}})\big)$ is the same when $\check{X}^{\mathsf{ex}}$ is sampled from $\mathcal{X}_L$ as when it is sampled from $\mathcal{X}_R$. Thus, for $\check{X}^{\mathsf{ex}}$ sampled from $\mathcal{X}_R$, $\textsc{shape}\big(\textsc{level}(\check{X}^{\mathsf{ex}})\big)$ consists of a stem that branches into two level set trees. If X^{ex} is the right positive excursion corresponding to $\check{X}^{\mathsf{ex}}$ sampled from $\mathcal{X}_R$, then almost surely,

$$\textsc{shape}\big(\textsc{level}(X^{\mathsf{ex}})\big) = \textsc{shape}\big(\textsc{level}(\check{X}^{\mathsf{ex}})\big).$$

Thus, conditioned on $\mathsf{ord}(T) \geq 2$, the tree $\textsc{shape}(T)$ splits into a pair of complete subtrees sampled independently from the same distribution as $\textsc{shape}(T)$. This completes the proof.   □

Theorem 18 establishes that the level set trees of symmetric random walks have the same combinatorial structure (equivalent to that of a ciritical binary Galton-Watson tree), independently of the choice of the transition kernel $p(x)$. The planar embedding and metric structure of the level set trees, however, may depend on the kernel, as we illustrate in the following remark.

Remark 17. Consider an extended positive right excursion $\check{X}^{\mathsf{ex}}$ of a symmetric homogeneous random walk and let $T = \textsc{level}(\check{X}^{\mathsf{ex}})$ be its level set tree. Condition on the event $\mathsf{ord}(T) \geq 2$, which ensures that the left and right principal subtrees of T, which we denote T^{ℓ} and T^r, respectively, are non-empty.

It follows from the construction in the proof of Thm. 18 that the subtrees T^{ℓ} and T^r can be sampled as follows. Consider two independent excursions – an extended positive right excursion $\check{X}_t^{\mathsf{ex,r}}$, $t \in [0, \kappa_r]$, and an extended positive left excursion $\check{X}_t^{\mathsf{ex},\ell}$, $t \in [\kappa_\ell, 0]$. Next, condition on the event that the excess value of the left excursion is less than that of the right excursion:

$$\Lambda\big(\{\check{X}_t^{\mathsf{ex},\ell}\}_{t \in [\kappa_\ell, 0]}\big) < \Lambda\big(\{\check{X}_t^{\mathsf{ex,r}}\}_{t \in [0, \kappa_r]}\big).$$

Denote by $X^{\mathsf{ex},\ell}$ and $X^{\mathsf{ex,r}}$ the positive left and right excursions that correspond to the extended excursions $\check{X}^{\mathsf{ex},\ell}$ and $\check{X}^{\mathsf{ex,r}}$. Then,

$$T^{\ell} \stackrel{d}{=} \textsc{level}(X_t^{\mathsf{ex},\ell}) \ \ \text{and} \ \ T^r \stackrel{d}{=} \textsc{level}(X_t^{\mathsf{ex,r}}). \tag{168}$$

Write X^{ex} for the positive right excursion that corresponds to the extended excursion $\check{X}^{\mathsf{ex}}$. Then, the stem of the tree $\textsc{level}(X^{\mathsf{ex}}) \in \mathcal{BL}_{\mathrm{plane}}^{|}$ has length equal to $\Lambda\big(\{\check{X}_t^{\mathsf{ex},\ell}\}_{t \in [0, \kappa_\ell]}\big)$. This, in general, may introduce dependence between the planar embedding of T and its edge lengths. Such dependence is absent in the exponential critical binary Galton-Watson tree $\mathsf{GW}(\lambda)$.

Next, condition on the event that X^{ex} is an Λ-shaped excursion, which is equivalent to

$$\{\#\textsc{level}(X^{\mathsf{ex}}) = 1\} \ = \ \{\#\textsc{level}(\check{X}^{\mathsf{ex}}) = 1\}.$$
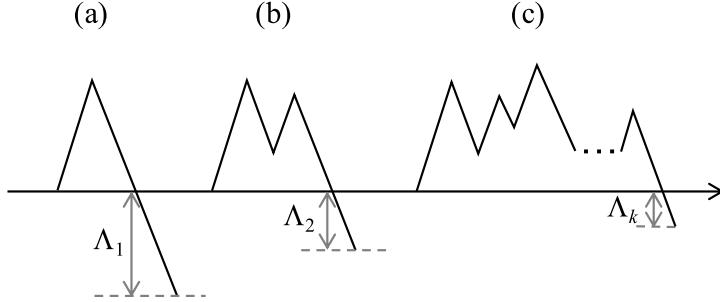
FIG 32. *Excess value* $\Lambda(\check{X}^{\mathsf{ex}})$ *may depend on the tree shape* SHAPE(LEVEL($\check{X}^{\mathsf{ex}}$)). *Equations* (169), (170) *demonstrate why the excess value for a Λ-shaped excursion of panel (a) may differ from the excess value of an M-shaped excursion of panel (b), and, hence, from the excess value of a general excursion of panel (c).*

Then, the density function of the excess value $\Lambda\big(\check{X}^{\mathsf{ex}}\big)$ that we denote by $\lambda_1(x)$ satisfies

$$\lambda_1(x) = 2 \int_0^\infty \mathsf{s}(x+y)\mathsf{s}(y)\,dy, \tag{169}$$

where $\mathsf{s}(x)$ was defined in (166). This is so because conditioned on

$$\{\#\text{LEVEL}(\check{X}^{\mathsf{ex}}) = 1\},$$

the extended excursion $\check{X}^{\mathsf{ex}}$ consists of an s-distributed jump upward, and a larger s-distributed jump downward. The excess value $\Lambda\big(\check{X}^{\mathsf{ex}}\big)$ is the difference between the jumps. The multiple of 2 in (169) is due to conditioning upon the event of probability $1/2$ that the jump up is smaller than the jump down.

Similarly, one can condition on the event that X^{ex} is an M-shaped excursion, which is equivalent to the event that the level set tree has 2 leaves and 3 edges, i.e.,

$$\{\#\text{LEVEL}(X^{\mathsf{ex}}) = 3\} \;=\; \{\#\text{LEVEL}(\check{X}^{\mathsf{ex}}) = 3\}.$$

Then, the density function of the excess value $\Lambda\big(\check{X}^{\mathsf{ex}}\big)$ that we denote by $\lambda_2(x)$ satisfies

$$\lambda_2(x) = 2 \int_0^\infty \lambda_1(x+y)\lambda_1(y)\,dy. \tag{170}$$

This is so because conditioned on

$$\{\#\text{LEVEL}(\check{X}^{\mathsf{ex}}) = 3\},$$

the extended excursion $\check{X}^{\mathsf{ex}}$ consists of two Λ-shaped (left and right) extended positive excursions originating from the only local minimum within the interior of the time domain $[0,\kappa]$ of $\check{X}^{\mathsf{ex}}$. The excess value $\Lambda\big(\check{X}^{\mathsf{ex}}\big)$ is the difference between the two λ_1-distributed excess values of the two Λ-shaped extended positive excursions.

Lemma 21. *Consider a homogeneous random walk X_k on $\mathbb{R}$ with a symmetric atomless transition kernel $p(x)$, $x \in \mathbb{R}$, i.e., there is a p.d.f. $f(x)$ with the support $\mathsf{supp}(f) \subseteq \mathbb{R}_+$ such that $p(x) = \frac{1}{2}(f(x) + f(-x))$. Consider an extended positive excursion $\check{X}^{\mathsf{ex}}$ of X_k, and the corresponding positive excursion X^{ex}. Let $T = \mathrm{LEVEL}(X^{\mathsf{ex}})$. Then, the following statements are equivalent:*

(a) T is independent of the excess value $\Lambda\big(\check{X}^{\mathsf{ex}}\big)$;
(b) conditioned on $\mathrm{P\text{-}SHAPE}(T)$, the edge lengths are identically distributed;
(c) $f(x)$ is an exponential p.d.f.

If any of these statements holds, then the edge lengths are i.i.d. exponential random variables.

Proof. $(c) \Rightarrow (a)$. It is easy to show via the characteristic functions that $\mathsf{s}(x)$ is an exponential p.d.f. if and only if $f(x)$ is an exponential p.d.f. The memoryless property of the exponential random variables implies that if $\mathsf{s}(x)$ is an exponential p.d.f., then $T = \mathrm{LEVEL}(X^{\mathsf{ex}})$ is independent of the excess value $\Lambda\big(\check{X}^{\mathsf{ex}}\big)$.

$(a) \Rightarrow (c)$. The excess value of a Λ-shaped extended positive excursion has the same distribution as the excess value of a M-shaped extended positive excursion if and only if $\lambda_1(x) \equiv \lambda_2(x)$. If this equality holds, then by equation (170) the p.d.f. $\lambda_1(x)$ satisfies equation (237) in Lemma 34, which implies that $\lambda_1(x) \equiv \lambda_2(x)$ is an exponential density function. Hence, from (169) and Lemma 35 we conclude that $\mathsf{s}(x) \equiv \lambda_1(x)$ is an exponential density function, which in turn implies that $f(x)$ is exponential.

$(b) \Rightarrow (c)$. The distribution of the leaf length is the minimum of two independent $\mathsf{s}(x)$-distributed random variables. Thus the cumulative distribution function of the leaf length equals

$$F_1(x) = 1 - \left(\int\limits_x^\infty \mathsf{s}(y)\, dy \right)^2.$$

The cumulative distribution function for the length of the non-leaf edge in a Y-shaped branch equals

$$F_2(x) = 1 - \left(\int\limits_x^\infty \lambda_1(y)\, dy \right)^2.$$

Here, $F_1(x) \equiv F_2(x)$ if and only if $\lambda_1(x) \equiv \mathsf{s}(x)$, which by Lemma 34 and equation (169) happens if and only if $\mathsf{s}(x)$ is exponential. This implies that $f(x)$ is an exponential p.d.f.

$(c) \Rightarrow (b)$. Suppose $f(x)$ is the exponential density with parameter λ, i.e., $f(x) = \phi_\lambda(x)$. According to the construction in the proof of Thm. 18, together with statement (a), and because any edge in T is a stem of a unique descendant subtree of T, it suffices to prove that conditioned on $\mathrm{P\text{-}SHAPE}(T)$, the tree stem (root edge) has exponential distribution with parameter λ.

According to (166), $\mathsf{s}(x)$ has the exponential density with parameter $\lambda/2$. Conditioned on $\mathsf{ord}(T) = 1$, the length of the stem (the only edge of the tree) equals the minimum of two independent exponentially distributed random variables with density $\mathsf{s}(x)$, and hence has the exponential density with parameter λ. Conditioned on $\mathsf{ord}(T) \geq 2$, the length of the stem is the minimum of the excess values of two independent extended positive excursions. By the memoryless property of the exponential distribution, each of these excess values has the exponential distribution with parameter $\lambda/2$. Hence, the stem length has the exponential distribution with parameter λ. This shows that the edge lengths in T have the same distribution.

Finally, suppose any and therefore all three of the statements (a)–(c) hold, then properties (b) and (c) insure that the edge lengths are identically and exponentially distributed, while property (a) insures the independence of edge lengths. This completes the proof. $\qquad\square$

8.6. Exponential random walks

Proposition 15 (and the subsequent comment) suggests that the problem of finding Horton self-similar trees with edge lengths is related to finding *extreme-invariant processes*

$$\left\{X_j^{(1)} - X_0^{(1)}\right\}_{j \in \mathbb{Z}} \overset{d}{=} \left\{\zeta(X_k - X_0)\right\}_{k \in \mathbb{Z}} \quad \text{for some } \zeta > 0, \qquad (171)$$

where $\{X_k\}_{k \in \mathbb{Z}}$, is a time series with an atomless distribution at every k and $X_j^{(1)}$ is the corresponding time series of local minima. The equality (171) is understood as the distributional equivalence of two time series.

In this section we establish a sufficient condition for a symmetric homogeneous random walk to solve (171), and show that in this case $\zeta = 2$. Moreover, we show that if a symmetric random walk X_k satisfies (171), the level set tree of its finite positive excursion, considered as elements of $\mathcal{L}_{\mathrm{plane}}$, is self-similar according to Def. 9. Symmetric random walks with exponential increments is an example of a process that solves (171).

The following result describes the solution of the problem (171) in terms of the characteristic function of $f(x)$.

Proposition 16 (**Extreme-invariance of a symmetric homogeneous random walk, [173]**). *Consider a symmetric homogeneous random walk $\{X_k\}_{k \in \mathbb{Z}}$ with a transition kernel $p(x) = \frac{f(x)+f(-x)}{2}$, where $f(x)$ is a p.d.f. with support $\mathsf{supp}(f) \subseteq \mathbb{R}_+$ and a finite second moment. Then, the local minima $\{X_j^{(1)}\}_{j \in \mathbb{Z}}$ of $\{X_k\}_{k \in \mathbb{Z}}$ form a symmetric homogeneous random walk with transition kernel*

$$p^{(1)}(x) = \zeta^{-1} p(x/\zeta), \quad \zeta > 0 \qquad (172)$$

if and only if $\zeta = 2$ and

$$\Re\left[\widehat{f}(2s)\right] = \left|\frac{\widehat{f}(s)}{2 - \widehat{f}(s)}\right|^2, \qquad (173)$$

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

Proof. Each increment between the consecutive local minima of X_k can be represented as d_j of (165) where $\{Y_i\}$ and $\{Z_i\}$ are i.i.d. with density $f(x)$, and ξ_+ and ξ_- are independent geometric random variables with parameter $1/2$, i.e., $\mathsf{Geom}_1(1/2)$.

The law of total variance readily implies that $\zeta = 2$. Indeed,

$$\mathsf{Var}\left(\sum_{i=1}^{\xi_+} Y_i\right) = \mathsf{E}\left[\mathsf{Var}\left(\sum_{i=1}^{\xi_+} Y_i \;\Big|\; \xi_+\right)\right] + \mathsf{Var}\left(\mathsf{E}\left[\sum_{i=1}^{\xi_+} Y_i \;\Big|\; \xi_+\right]\right)$$
$$= \sigma^2 \mathsf{E}[\xi_+] + \mu^2 \mathsf{Var}(\xi_+) = 2(\mu^2 + \sigma^2), \tag{174}$$

where μ and $\mu^2 + \sigma^2$ are the first and the second moments of $f(x)$ respectively. Thus, on one hand, the variance of the increments of X_k is

$$\mathsf{Var}(X_{k+1} - X_k) = \mu^2 + \sigma^2$$

since for a symmetric homogeneous random walk, $\mathsf{E}[X_{k+1} - X_k] = 0$. On the other hand, (165) and (174) imply that the variance of the increments in the sequence of local minima $X_j^{(1)}$ is

$$\mathsf{Var}(X_{j+1}^{(1)} - X_j^{(1)}) = \mathsf{Var}(d_j) = \mathsf{Var}\left(\sum_{i=1}^{\xi_+} Y_i\right) + \mathsf{Var}\left(\sum_{i=1}^{\xi_-} Z_i\right) = 4(\mu^2 + \sigma^2).$$

Hence, $\mathsf{Var}(X_{j+1}^{(1)} - X_j^{(1)}) = 4\,\mathsf{Var}(X_{k+1} - X_k)$, and therefore $\zeta = 2$ is the only value of ζ for which the scaling (172) may hold.

Taking the characteristic functions in (172), we obtain

$$\widehat{p}^{(1)}(s) = \widehat{p}(\zeta s) = \Re\left[\widehat{f}(\zeta s)\right],$$

while taking the characteristic function of d_j in (165) we have

$$\widehat{p}^{(1)}(s) = \mathsf{E}[e^{isd_j}] = \mathsf{E}\left[\left(\widehat{f}(s)\right)^{\xi_+}\right] \mathsf{E}\left[\left(\widehat{f}(-s)\right)^{\xi_-}\right] = \left|\frac{\widehat{f}(s)}{2 - \widehat{f}(s)}\right|^2.$$

Thus, (172) is satisfied if and only if

$$\Re\left[\widehat{f}(\zeta s)\right] = \left|\frac{\widehat{f}(s)}{2 - \widehat{f}(s)}\right|^2. \tag{175}$$

Substituting $\zeta = 2$ into (175) completes the proof. $\qquad\qquad\square$

Example 13. Exponential density $f(x) = \phi_\lambda(x)$ of (72) solves (173) with any $\lambda > 0$; see Thm. 19 below for more detail.

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 $l < k < r$ (see Fig. 30). 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 $X_{\tilde{r}} = X_l$. The level set tree of a positive excursion of X_k is that of the corresponding positive excursion of X_t.

Propositions 16 and 15 imply the following statement.

Corollary 10. *Consider a symmetric homogeneous random walk $\{X_k\}_{k\in\mathbb{Z}}$ with a transition kernel $p(x) = \frac{f(x)+f(-x)}{2}$, where $f(x)$ is a p.d.f. with support $\mathsf{supp}(f) \subseteq \mathbb{R}_+$ and a finite second moment. Let*

$$T = \mathrm{LEVEL}\big(\{X_t\}_{t\in[l,r]}\big)$$

be the level set tree for a positive excursion $\{X_t\}_{t\in[l,r]}$ generated by the random walk X_k as defined in Sect. 8.2. Then, the tree T has a Horton self-similar distribution (Def. 9) over $\mathcal{BL}^|_{\mathrm{plane}}$, if and only if the condition (173) holds for the characteristic function $\widehat{f}(s)$ of $f(x)$.

Proof. Horton prune-invariance is established in Props. 16, 15. □

A homogeneous random walk on $\mathbb{R}$ is called *exponential random walk* if its transition kernel is a mixture of exponential jumps:

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

where ϕ_λ is the exponential density with parameter $\lambda > 0$ as defined in Eq. (72). We refer to an exponential random walk by its parameter triplet $\{\rho, \lambda_u, \lambda_d\}$. Each interpolated exponential random walk with parameters $\{\rho, \lambda_u, \lambda_d\}$ is a piece-wise linear function whose positive (up) and negative (down) increments are independent exponential random variables with respective parameters λ_u and λ_d, and the probabilities of a positive or negative increment at every integer instant are ρ and $(1 - \rho)$, respectively. After a time change that makes all segments to have slopes ± 1, each interpolated exponential random walk with parameters $\{\rho, \lambda_u, \lambda_d\}$ corresponds to a piece-wise linear function with alternating rises and falls that 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 19 (**Self-similarity of exponential random walks, [173]**). *Let $\{X_k\}_{k\in\mathbb{Z}}$ be an exponential random walk with parameters $\{\rho, \lambda_u, \lambda_d\}$. Then*

(a) *The sequence $\{X_j^{(1)}\}_{j\in\mathbb{Z}}$ of the local minima of X_k is an exponential random walk with parameters $\{\rho^*, \lambda_u^*, \lambda_d^*\}$ such that*

$$\rho^* = \frac{\rho\,\lambda_d}{\rho\,\lambda_d + (1 - \rho)\,\lambda_u}, \quad \lambda_d^* = \rho\lambda_d, \quad and \quad \lambda_u^* = (1 - \rho)\lambda_u. \quad (176)$$

(b) *The exponential walk X_k satisfies the self-similarity condition* (171) *if and only if it is symmetric ($\rho = 1/2$ and $\lambda_u = \lambda_d$), i.e., when $p(x)$ is a mean zero Laplace p.d.f.*

(c) *The self-similarity* (171) *is achieved after the first Horton pruning, for the chain $\{X_j^{(1)}\}_{j \in \mathbb{Z}}$ of the local minima, if and only if the walk's increments have zero mean, $\rho \lambda_d = (1 - \rho) \lambda_u$.*

Proof. (a) By Lemma 20(i), the sequence of local minima $X_j^{(1)}$ of X_k is a homogeneous random walk with transition kernel $p^{(1)}(x)$. The latter is the probability distribution of the jumps d_j given by (165) with

$$\xi_+ \overset{d}{\sim} \mathsf{Geom}_1(1 - \rho), \quad \xi_- \overset{d}{\sim} \mathsf{Geom}_1(\rho), \quad Y_i \overset{d}{\sim} \phi_{\lambda_u}, \quad \text{and} \quad Z_i \overset{d}{\sim} \phi_{\lambda_d}.$$

The characteristic function $\widehat{p}^{(1)}(s)$ of the transition kernel $p^{(1)}(x)$ is found here as follows

$$\widehat{p}^{(1)}(s) = \mathsf{E}\left[\exp\left\{is\left(X_{j+1}^{(1)} - X_j^{(1)}\right)\right\}\right] = \frac{\rho(1 - \rho)\lambda_d \lambda_u}{((1 - \rho)\lambda_u - is)(\rho\lambda_d + is)}$$

$$= \rho^* \widehat{\phi}_{\lambda_u^*}(s) + (1 - \rho^*)\widehat{\phi}_{\lambda_d^*}(s),$$

where

$$\widehat{\phi}_\lambda(s) = \frac{\lambda}{\lambda - is} \tag{177}$$

is the characteristic function of an exponential random variable with parameter λ, and $\rho^*, \lambda_u^*, \lambda_d^*$ are given by (176). Thus,

$$p^{(1)}(x) = \rho^* \phi_{\lambda_u^*}(x) + (1 - \rho^*)\phi_{\lambda_d^*}(-x).$$

This means that the sequence of local minima $\{X_j^{(1)}\}$ also evolves according to a two-sided exponential transition kernel, only with different parameters, ρ^*, λ_d^*, and λ_u^*.

Part (b) of the theorem follows immediately from part (a). Alternatively, we observe that the exponential density $f(x) = \phi_\lambda(x)$ solves (173) with any $\lambda > 0$: by (177) we have

$$\Re\left[\widehat{\phi}_\lambda(2s)\right] = \Re\left[\frac{\lambda}{\lambda - 2is}\right] = \frac{\lambda^2}{\lambda^2 + 4s^2}$$

and

$$\left|\frac{\widehat{\phi}_\lambda(s)}{2 - \widehat{\phi}_\lambda(s)}\right|^2 = \left|\frac{\lambda}{\lambda - 2is}\right|^2 = \frac{\lambda^2}{\lambda^2 + 4s^2}.$$

Hence, Part (b) follows from Prop. 16.

(c) Observe that $\rho^* = 1/2$ and $\lambda_d^* = \lambda_u^*$ if and only if $\rho \lambda_d = (1 - \rho) \lambda_u$. $\square$

We now extend Def. 20 to non-critical Galton-Watson trees.

Definition 30 (**Exponential binary Galton-Watson tree, [135]**). *We say that a random planted embedded binary tree* $T \in \mathcal{BL}^|_{\mathrm{plane}}$ *is an exponential binary Galton-Watson tree and write* $T \stackrel{d}{\sim} \mathsf{GW}(\lambda', \lambda)$, *for* $0 \le \lambda' < \lambda$, *if*

(i) SHAPE(T) *is a binary Galton-Watson tree* $\mathcal{GW}(q_0, q_2)$ *with*

$$q_0 = \frac{\lambda + \lambda'}{2\lambda}, \quad q_2 = \frac{\lambda - \lambda'}{2\lambda};$$

(ii) *the orientation for every pair of siblings in* T *is random and symmetric; and*

(iii) *conditioned on a given* SHAPE(T), *the edges of* T *are sampled as independent exponential random variables with parameter* λ, *i.e., with density* (72).

In particular, we observe that $\mathsf{GW}(\lambda) = \mathsf{GW}(0, \lambda)$. A connection between exponential random walks and exponential Galton-Watson trees is provided by the following well known result.

Theorem 20 (**[135, Lemma 7.3], [105, 124]**). *Consider a random excursion* Y_t *in* $\mathcal{E}^{\mathrm{ex}}$. *The level set tree* LEVEL(Y_t) *is an exponential binary Galton-Watson tree* $\mathsf{GW}(\lambda', \lambda)$ *if and only if the alternating rises and falls of* Y_t, *excluding the last fall, are distributed as independent exponential random variables with parameters* $\frac{\lambda + \lambda'}{2}$ *and* $\frac{\lambda - \lambda'}{2}$, *respectively, for some* $0 \le \lambda' < \lambda$.

Equivalently, for a random excursion Y_t *of a homogeneous random walk in* $\mathcal{E}^{\mathrm{ex}}$, *the level set tree* LEVEL(Y_t) *is an exponential binary Galton-Watson tree* $\mathsf{GW}(\lambda', \lambda)$ *if and only if* Y_t, *as an element of* $\mathcal{E}^{\mathrm{ex}}$, *corresponds to an excursion of an exponential walk with parameters* $\{\rho, \lambda_u, \lambda_p\}$ *satisfying* $(1 - \rho)\lambda_u = \frac{\lambda + \lambda'}{2}$ *and* $\rho\lambda_d = \frac{\lambda - \lambda'}{2}$.

The following statement is a corollary of Lem. 8 in Sect. 5.2.2 and Thm. 20.

Corollary 11. *The p.d.f.* $f(x)$ *of the length of an excursion in an exponential symmetric random walk with parameter* λ *is given by*

$$f(x) = \frac{1}{2}\ell(x/2). \tag{178}$$

Proof. Observe that the excursion has twice the length of a tree $\mathsf{GW}(\lambda)$. $\square$

We also emphasize the following direct consequence of Thms. 19(a) and 20.

Corollary 12. *Suppose* $T \stackrel{d}{\sim} \mathsf{GW}(\gamma)$ *is an exponential critical binary Galton-Watson tree. Then, the following statements hold:*

(a) *The pruned exponential critical binary Galton-Watson tree is an exponential critical binary Galton-Watson tree:*

$$\left(\mathcal{R}^k(T) \mid \mathcal{R}^k(T) \neq \phi\right) \stackrel{d}{\sim} \mathsf{GW}\left(2^{-k}\gamma\right) \text{ for any } k \in \mathbb{N}.$$

(b) *The lengths of branches of Horton-Strahler order $j \geq 1$ in T (see Def. 5) has exponential distribution with parameter $2^{1-j}\gamma$. The lengths of branches (of all orders) are independent.*

Remark 18 (**A link between Thm. 19 and Thm. 6**). Consider an excursion of an exponential random walk X_t with parameters $\{\rho, \lambda_u, \lambda_d\}$. The geometric stability of the exponential distribution implies that the monotone rises and falls of X_t are exponentially distributed with parameters $(1 - \rho)\lambda_u$ and $\rho\lambda_d$, respectively. Thus, Thm. 20 implies that $\text{SHAPE}(\text{LEVEL}(X_t))$ is distributed as a binary Galton-Watson tree $\mathcal{GW}(q_0, q_2)$ with

$$q_2 = \frac{\rho\lambda_d}{(1-\rho)\lambda_u + \rho\lambda_d} = 1 - q_0.$$

The first pruning $X_t^{(1)}$ (see Sect. 8.4), according to (176), is an exponential random walk with parameters

$$\left\{\rho^* = \frac{\rho\lambda_d}{(1-\rho)\lambda_u + \rho\lambda_d}, \ \lambda_u^* = (1-\rho)\lambda_u, \ \lambda_d^* = \rho\lambda_d\right\}.$$

Its upward and downward increments are exponentially distributed with parameters, respectively,

$$(1-\rho^*)\lambda_u^* = \frac{[(1-\rho)\lambda_u]^2}{(1-\rho)\lambda_u + \rho\lambda_d} \quad \text{and} \quad \rho^*\lambda_d^* = \frac{[\rho\lambda_d]^2}{(1-\rho)\lambda_u + \rho\lambda_d}.$$

Accordingly, the level set tree for a positive excursion $X_t^{(1)}$ is a binary Galton-Watson tree $\mathcal{GW}(q_0^{(1)}, q_2^{(1)})$ with

$$q_2^{(1)} = \frac{[\rho\lambda_d]^2}{[(1-\rho)\lambda_u]^2 + [\rho\lambda_d]^2} = 1 - q_0^{(1)}.$$

Continuing this way, we find that n-th pruning $X_t^{(n)}$ of X_t is an exponential random walk such that the level set tree of its positive excursion has binary Galton-Watson distribution $\mathcal{GW}(q_0^{(n)}, q_2^{(n)})$ with

$$q_2^{(n)} = \frac{\left[q_2^{(n-1)}\right]^2}{\left[q_0^{(n-1)}\right]^2 + \left[q_2^{(n-1)}\right]^2} = \frac{[\rho\lambda_d]^{2^n}}{[(1-\rho)\lambda_u]^{2^n} + [\rho\lambda_d]^{2^n}} = 1 - q_0^{(n)} \quad (179)$$

The first equality in (179) defines the same iterative system as (64) in Thm. 6 of Burd et al. that describes iterative Horton pruning of Galton-Watson trees. Another noteworthy relation connecting the exponential random walk $X_t^{(n)}$ with parameters $\{\rho^{(n)}, \lambda_u^{(n)}, \lambda_d^{(n)}\}$ and the Galton-Watson tree $\mathcal{GW}(q_0^{(n-1)}, q_2^{(n-1)})$ is given by

$$\rho^{(n)} = q_2^{(n-1)} \text{ for any } n \geq 1 \text{ (where } q_2^{(0)} \equiv q_2).$$

8.7. Geometric random walks and critical non-binary Galton-Watson trees

A recent study by Barbosa et al. [17] examines the self-similar properties of the level-set trees corresponding to the excursions of the so-called *geometric random walk* on $\mathbb{Z}$, defined below (Def. 31). The results in [17] give a discrete-space version of the results discussed in Sect. 8.6.

For the given probabilities $\{p_1, p_2, r_1, r_2\}$ such that $p_1 + p_2 \leq 1$, consider a discrete-time random walk on $\mathbb{Z}$, where at each time step, p_1 is the probability of an upward jump, p_2 is the probability of a downward jump, and $1 - p_1 - p_2$ is the probability of remaining at the same location. Conditioned on jumping upward, the increment size is a $\mathsf{Geom}_1(r_1)$-distributed random variable, while conditioned on jumping downward, the increment size is a $\mathsf{Geom}_1(r_2)$-distributed random variable. Here is a formal definition.

Definition 31 (**Geometric random walk**). *A geometric random walk X_k, $k \in \mathbb{Z}$, with probability parameters*

$$\{p_1, p_2, r_1, r_2\}$$

is a discrete time space-homogeneous random walk on $\mathbb{Z}$ with transition probabilities $p(x, y) = p(y - x)$ such that its jump kernel $p(x)$ is a double-sided geometric probability mass function (discrete Laplace distribution) that can be expressed as

$$p(x) = p_1 g_1(x) + (1 - p_1 - p_2)\delta_0(x) + p_2 g_2(-x), \tag{180}$$

where $\delta_0(x)$ denotes the Kronecker delta function at 0, and $g_i(x)$ ($i = 1, 2$) is the probability mass function of a $\mathsf{Geom}_1(r_i)$-distributed random variable. The distribution for a geometric random walk is denoted by $\mathrm{GRW}(p_1, p_2, r_1, r_2)$.

Example 14. The most celebrated example of a geometric random walk is the simple random walk on $\mathbb{Z}$ with distribution $\mathrm{GRW}\left(\frac{1}{2}, \frac{1}{2}, 1, 1\right)$.

By (180), the characteristic function for the increments in a geometric walk is given by

$$\widehat{p}(s) = \frac{p_1 r_1 e^{is}}{1 - (1 - r_1)e^{is}} + \frac{p_2 r_2 e^{-is}}{1 - (1 - r_2)e^{-is}} + (1 - p_1 - p_2). \tag{181}$$

Equation (181) leads to the derivation of the following invariance result, analogous to Thm. 19(a) in a discrete space setting.

Theorem 21 ([17]). *Suppose X_k is a geometric random walk $\mathrm{GRW}(p_1, p_2, r_1, r_2)$, then the time series $X_k^{(1)}$ of its consecutive local minima (including flat plateaus) is also a geometric random walk $\mathrm{GRW}\left(p_1^{(1)}, p_2^{(1)}, r_1^{(1)}, r_2^{(1)}\right)$ with probability parameters*

$$r_1^{(1)} = \frac{p_2 r_1}{p_1 + p_2}, \qquad p_1^{(1)} = \frac{r_2^{(1)}(1 - r_1^{(1)})}{1 - (1 - r_1^{(1)})(1 - r_2^{(1)})},$$

$$r_2^{(1)} = \frac{p_1 r_2}{p_1 + p_2}, \qquad p_2^{(1)} = \frac{r_1^{(1)}(1 - r_2^{(1)})}{1 - (1 - r_1^{(1)})(1 - r_2^{(1)})}.$$

If $r_1 = r_2 = r$ and $p_1 = p_2 = p$, the geometric random walk $\mathrm{SGRW}(p, r) \equiv \mathrm{GRW}(p, p, r, r)$ is called *symmetric geometric random walk* (SGRW). In this case, Thm. 21 can be reinterpreted as the following statement, analogous to Thm. 19(b) adapted to the discrete space $\mathbb{Z}$.

Corollary 13 ([**17**]). *Suppose* $X_k \overset{d}{\sim} \mathrm{SGRW}(p, r)$ *is a symmetric geometric random walk on* $\mathbb{Z}$. *Then, the time series* $X_k^{(1)}$ *of its consecutive local minima is also a symmetric geometric random walk* $\mathrm{SGRW}(p^{(1)}, r^{(1)})$ *with probability parameters*

$$p^{(1)} = \frac{1 - r^{(1)}}{2 - r^{(1)}} \quad and \quad r^{(1)} = \frac{r}{2}.$$

Next, consider the case of a geometric random walk X_t with mean zero increments,

$$E[X_{k+1} - X_k] = \frac{p_1}{r_1} - \frac{p_2}{r_2} = 0.$$

In this case $p_1 r_2 = p_2 r_1$, and Thm. 21 and Cor. 13 imply the following result.

Corollary 14 ([**17**]). *Suppose* $X_k \overset{d}{\sim} \mathrm{GRW}(p_1, p_2, r_1, r_2)$ *is a geometric random walk with zero mean, i.e.,* $p_1 r_2 = p_2 r_1$. *Then, the time series* $X_k^{(1)}$ *of its consecutive local minima is a symmetric geometric random walk* $\mathrm{SGRW}(p^{(1)}, r^{(1)})$ *with probability parameters*

$$p^{(1)} = \frac{1 - r^{(1)}}{2 - r^{(1)}} \quad and \quad r^{(1)} = \frac{r}{2},$$

where $r = \frac{2p_1 r_2}{p_1 + p_2} = \frac{2p_2 r_1}{p_1 + p_2}$.

Furthermore, let $X_k^{(n+1)}$ for $n = 1, 2, \ldots$ be the time series of the consecutive local minima of $X_k^{(n)}$. Then, $X_k^{(n)}$ is also a symmetric geometric random walk $\mathrm{SGRW}(p^{(n)}, r^{(n)})$ with probability parameters

$$p^{(n)} = \frac{1 - r^{(n)}}{2 - r^{(n)}} \quad and \quad r^{(n)} = \frac{r}{2^n}. \tag{182}$$

For the remainder of this section, let $\{p^{(n)}, r^{(n)}\}$ denote the parameters of the symmetric geometric random walk $\mathrm{SGRW}(p^{(n)}, r^{(n)})$, obtained by taking k iterations of local minima of $X_k \overset{d}{\sim} \mathrm{GRW}(p_1, p_2, r_1, r_2)$, as in Cor. 14.

Corollary 15 ([**17**]). *Suppose* $X_k \overset{d}{\sim} \mathrm{GRW}(p_1, p_2, r_1, r_2)$ *is a mean zero geometric random walk, i.e.,* $p_1 r_2 = p_2 r_1$. *Then,*

$$\lim_{n \to \infty} r^{(n)} = 0 \quad and \quad \lim_{n \to \infty} p^{(n)} = \frac{1}{2}.$$

The following is a discrete analogue of Thm. 20, stated in Sect. 8.6.

Theorem 22 ([**17**]). *Suppose* $X_k \overset{d}{\sim} \mathrm{GRW}(p_1, p_2, r_1, r_2)$ *is a geometric random walk with a nonnegative drift, i.e.,* $p_1 r_2 \leq p_2 r_1$*. Let* T^{ex} *be the level set tree of a positive excursion of* X_k*. Then,*

$$\mathrm{SHAPE}(T^{\mathrm{ex}}) \overset{d}{\sim} \mathcal{GW}(\{q_m\}) \ on \ \mathcal{T}^{|}$$

with

$$q_0 = 1 - p_1^{(1)}, \quad and \quad q_m = p_1^{(1)} \big(r_2^{(1)} \big)^{m-2} (1 - r_2^{(1)}) \quad for \ m = 2, 3 \ldots,$$

where $r_1^{(1)}$*,* $r_2^{(1)}$*,* $p_1^{(1)}$*, and* $p_2^{(1)}$ *are as in Theorem 21 (recall that* $q_1 \equiv 0$ *since we work with reduced trees). Moreover, if* X_k *is a mean zero geometric random walk (i.e.,* $p_1 r_2 = p_2 r_1$*), then*

$$q_0 = \frac{1}{2 - r^{(1)}}, \quad and \quad q_m = \frac{\big(r^{(1)} \big)^{m-2} (1 - r^{(1)})^2}{2 - r^{(1)}} \quad for \ m = 2, 3 \ldots,$$

where $r_1^{(1)}$ *and* $r_2^{(1)}$ *are as in Corollary 14.*

Observe that, in the setting of Thm. 22, if we consider a mean zero GRW ($p_1 r_2 = p_2 r_1$ and, equivalently, $r_1^{(1)} = r_2^{(1)}$) then,

$$\sum_m m q_m = \left(\frac{1 - r_1^{(1)}}{1 - r_2^{(1)}} \right) \left(\frac{r_2^{(1)} + r_2^{(1)}(1 - r_2^{(1)})}{r_2^{(1)} + r_1^{(1)}(1 - r_2^{(1)})} \right) = 1.$$

In other words, the level set tree of its positive excursion is distributed as a *critical* Galton-Watson tree $\mathcal{GW}(\{q_m\})$. Combining Prop. 15 with Thm. 22 we have the following corollary.

Corollary 16 ([**17**]). *Suppose* $X_k \overset{d}{\sim} \mathrm{GRW}(p_1, p_2, r_1, r_2)$ *is a mean zero geometric random walk, i.e.,* $p_1 r_2 = p_2 r_1$*. Let* T^{ex} *be the level set tree of a positive excursion of* X_k*. Then,* $\mathrm{SHAPE}(T^{\mathrm{ex}}) \overset{d}{\sim} \mathcal{GW}(\{q_m\})$*, where* $\mathcal{GW}(\{q_m\})$ *is a critical Galton-Watson distribution on* $\mathcal{T}^{|}$*. Moreover, for any* $n \geq 1$*, the level set tree of a positive excursion of* $X_k^{(n)}$ *is distributed as*

$$\left(\mathcal{R}^n(T^{\mathrm{ex}}) \,\Big|\, \mathcal{R}^n(T^{\mathrm{ex}}) \neq \phi \right) \overset{d}{\sim} \mathcal{GW}(\{q_m^{(n)}\})$$

with

$$q_0^{(n)} = \frac{1}{2 - r^{(n+1)}} \quad and \quad q_m^{(n)} = \frac{(r^{(n+1)})^{m-2} (1 - r^{(n+1)})^2}{2 - r^{(n+1)}} \quad (\forall m \geq 2), \quad (183)$$

where $r^{(n)}$ *is given by equation (182) of Corollary 14.*

Finally, letting $n \to \infty$*, we have*

$$q_0^{(n)} \to \frac{1}{2}, \quad q_2^{(n)} \to \frac{1}{2}, \quad and \quad q_m^{(n)} \to 0 \quad (\forall m > 2). \quad (184)$$

The convergence in (184) follows from Cor. 15 as $r^{(n)} \to 0$. Writing $\nu = \mathcal{GW}(\{q_m\})$, we have by Cor. 16 that the pushforward measure $\nu_n := \mathcal{R}_*^n(\nu) = \nu \circ \mathcal{R}^{-n}$ satisfies

$$\nu_n(T \,|T \neq \phi) \stackrel{d}{=} \mathcal{GW}(\{q_m^{(n)}\})$$

while equation (184) additionally asserts that

$$\lim_{n \to \infty} \nu_n(T \,|T \neq \phi) = \rho^*(\tau), \tag{185}$$

where ρ^* denotes the critical binary Galton-Watson measure $\mathcal{GW}\left(\frac{1}{2}, \frac{1}{2}\right)$ on $\mathcal{T}^|$ with the support on $\mathcal{BT}^|$. Equation (185) provides a specific example of Thm. 5 showing that recursive pruning of a critical Galton-Watson tree satisfying

$$\sum_{m=2}^{\infty} m^{2-\epsilon} q_m \, < \infty \qquad \forall \epsilon > 0$$

will have the critical binary Galton-Watson tree measure as its limit.

8.8. White noise and Kingman's coalescent

This section establishes an interesting correspondence between the tree representations of a white noise (sequence of i.i.d. random variables) and celebrated Kingman's coalescent process [89]. We begin by an informal review of coalescent processes and their trees.

8.8.1. Coalescent processes, trees

Coalescent processes [135, 6, 29, 25, 62]. A general finite coalescent process begins with N singletons. The cluster formation is governed by a symmetric collision rate kernel $K(i, j) = K(j, i) > 0$. Specifically, a pair of clusters with masses (weights) i and j coalesces at the rate $K(i, j)/N$, independently of the other pairs, to form a new cluster of mass $i + j$. The process continues until there is a single cluster of mass N.

Formally, for a given $N \geq 1$ consider the space $\mathcal{P}_{[N]}$ of partitions of $[N] = \{1, 2, \dots, N\}$. Let $\Pi_0^{(N)}$ be the initial partition in singletons, and $\Pi_t^{(N)}$ $(t \geq 0)$ be a strong Markov process such that $\Pi_t^{(N)}$ transitions from partition $\pi \in \mathcal{P}_{[N]}$ to $\pi' \in \mathcal{P}_{[N]}$ with rate $K(i, j)/N$ provided that partition π' is obtained from partition π by merging two clusters of π of weights i and j. If $K(i, j) \equiv 1$ for all positive integer masses i and j, the process $\Pi_t^{(N)}$ is known as the N-particle *Kingman's coalescent process*. If $K(i, j) = i + j$ the process is called the N-particle *additive coalescent*. Finally, if $K(i, j) = ij$ the process is called the N-particle *multiplicative coalescent*.

Coalescent tree. A merger history of the N-particle coalescent process can be naturally described by a time oriented binary tree constructed as follows. Start with N leaves that represent the initial N particles and have time mark

$t = 0$. When two clusters coalesce (a transition occurs), merge the corresponding vertices to form an internal vertex with a time mark of the coalescent. The final coalescence forms the tree root. The resulting time oriented binary tree represents the history of the process. We notice that a given unlabeled tree corresponds to multiple coalescent trajectories obtained by relabeling of the initial particles.

Let $T_{\mathrm{K}}^{(N)}$ denote the coalescent tree for the N-particle Kingman's coalescent process. Let N_j denote the number of branches of Horton-Strahler order j in the tree $T_{\mathrm{K}}^{(N)}$. In Sect. 9 we will show that for each $j \geq 1$, the asymptotic Horton ratios $\mathcal{N}_j$ are well-defined (Def. 18), that is

$$\frac{N_j}{N} \xrightarrow{p} \mathcal{N}_j \quad \text{as} \quad N \to \infty. \tag{186}$$

Moreover, the Horton ratios $\mathcal{N}_j$ are finite and can be expressed as

$$\mathcal{N}_j = \frac{1}{2} \int_0^\infty g_j^2(x)\, dx,$$

where the sequence $g_j(x)$ solves the following system of ordinary differential equations (ODEs):

$$g_{j+1}'(x) - \frac{g_j^2(x)}{2} + g_j(x)g_{j+1}(x) = 0, \quad x \geq 0 \tag{187}$$

with $g_1(x) = 2/(x+2)$, $g_j(0) = 0$ for $j \geq 2$. Equivalently,

$$\mathcal{N}_j = \int_0^1 \left(1 - (1-x)\, h_{j-1}(x)\right)^2 dx,$$

where $h_0 \equiv 0$ and the sequence $h_j(x)$ satisfies the ODE system

$$h_{j+1}'(x) = 2h_j(x)h_{j+1}(x) - h_j^2(x), \quad 0 \leq x \leq 1 \tag{188}$$

with the initial conditions $h_k(0) = 1$ for $j \geq 1$.

The root-Horton law (Def. 19) for the well-defined Horton ratios $\mathcal{N}_j$ (186) of the Kingman's coalescent process is stated in Thm. 25, with the Horton exponent bounded by the interval $2 \leq R \leq 4$. Moreover, the Horton exponent is estimated to be $R = 3.0438279\ldots$ via the ODE representation in (187) and (188). The numerical computation (not shown here) affirms that the ratio-Horton and the strong Horton laws of Def. 19 are valid for the Kingman's coalescent as well.

8.8.2. White noise

In this section we will show that the combinatorial SHAPE function for the level set tree T_{wn} of white noise is closely connected to the SHAPE function of the Kingman's coalescent tree $T_{\mathrm{K}} = T_{\mathrm{K}}^{(N)}$ introduced in Sect. 8.8.1. Specifically, the

two are separated by a single Horton pruning $\mathcal{R}$. In other words, conditioning on the same number of leaves, $\text{SHAPE}\big(\mathcal{R}(T_K)\big) \overset{d}{=} \text{SHAPE}(T_{\text{wn}})$.

Let $W_j^{(N)}$ with $j = 1, \ldots, N-1$ be a *discrete white noise* that is a discrete time process comprised of $N-1$ i.i.d. random variables with a common atomless distribution. Next, we consider an auxiliary process $\tilde{W}_i^{(N)}$ with $i = 1, \ldots, 2N-1$, such that it has exactly N local maxima and $N-1$ internal local minima $\tilde{W}_{2j}^{(N)} = W_j^{(N)}$, $j = 1, \ldots, N-1$. We call $\tilde{W}_i^{(N)}$ an *extended white noise*. It can be constructed as in the following example.

Example 15 (Extended white noise).

$$\tilde{W}_i^{(N)} = \begin{cases} W_{i/2}^{(N)}, & \text{for even } i, \\ \max\left\{W_{i'}^{(N)}, W_{i''}^{(N)}\right\} + 1, & \text{for odd } i, \end{cases} \tag{189}$$

where $i' = \max\left(1, \frac{i-1}{2}\right)$ and $i'' = \min\left(N-1, \frac{i+1}{2}\right)$.

Let $T_{\text{wn}}^{(N)} = \text{LEVEL}\left(\tilde{W}_i^{(N)}\right)$ be the level set tree of $\tilde{W}_i^{(N)}$. By construction, $T_{\text{wn}}^{(N)}$ has exactly N leaves. Also observe that the level set trees $T_{\text{wn}}^{(N)}$ and $\text{LEVEL}\left(W_j^{(N)}\right)$ are separated by a single Horton pruning:

$$\mathcal{R}\left(T_{\text{wn}}^{(N)}\right) = \text{LEVEL}\left(W_j^{(N)}\right). \tag{190}$$

Lemma 22. *The distribution of* $\text{SHAPE}\left(T_{\text{wn}}^{(N)}\right)$ *on* $\mathcal{BT}^{|}$ *is the same for any atomless distribution F of the values of the associated white noise $W_j^{(N)}$.*

Proof. The condition of atomlessness of F is necessary to ensure that the level set tree is binary with probability one. By construction, the combinatorial level set tree is completely determined by the ordering of the local minima of the respective trajectory, independently of the particular values of its local maxima and minima. We complete the proof by noticing that the distribution for the ordering of $W_j^{(N)}$ is the same for any choice of atomless distribution F. $\qquad\square$

Let $T_K^{(N)}$ be the tree that corresponds to the Kingman's N-coalescent, and let $\text{SHAPE}\left(T_K^{(N)}\right)$ be its combinatorial version that drops the time marks of the vertices. Both the trees $\text{SHAPE}\left(T_{\text{wn}}^{(N)}\right)$ and $\text{SHAPE}\left(T_K^{(N)}\right)$, belong to the space $\mathcal{BT}^{|}$ (or, more specifically, to $\mathcal{BT}^{|}$ conditioned on N leaves).

Theorem 23. *The trees* $\text{SHAPE}\left(T_{\text{wn}}^{(N)}\right)$ *and* $\text{SHAPE}\left(T_K^{(N)}\right)$ *have the same distribution on* $\mathcal{BT}^{|}$.

Proof. The proof uses a construction similar in some respect to the celebrated Kingman paintbox process [89, 135, 29, 25]. For the Kingman's N-coalescent, let us enumerate the initial singletons from 1 to N. We will identify each cluster

with a collection of singletons listed from left to right, where the order in which they are listed is important as it contains a certain amount of information regarding the process's merger history. Specifically, consider a pair of clusters $\mathbf{i}$ and $\mathbf{j}$, identified with the corresponding collection of singletons as follows

$$\mathbf{i} = \{i_1, \ldots, i_k\} \quad \text{and} \quad \mathbf{j} = \{j_1, \ldots, j_m\}.$$

Next, we split the merger rate of $\frac{1}{N}$ into two. We let the clusters $\mathbf{i}$ and $\mathbf{j}$ merge into the new cluster

$$\{\mathbf{i}, \mathbf{j}\} = \{i_1, \ldots, i_k, j_1, \ldots, j_m\}$$

with rate $\frac{1}{2N}$, or into the new cluster

$$\{\mathbf{j}, \mathbf{i}\} = \{j_1, \ldots, j_m, i_1, \ldots, i_k\}$$

also with rate $\frac{1}{2N}$. The final merger results in a cluster consisting of all N singletons, listed as a permutation from S_N,

$$\sigma = \{\sigma_1, \ldots, \sigma_N\}.$$

Conditioning on the final permutation σ, the merger history is described by the random connection times,

$$t_1, t_2, \ldots, t_{N-1},$$

where t_j is the merger time when the singletons σ_j and σ_{j+1} meet in the same cluster. The following diagram helps visualize the connection times:

$$\sigma_1 \xrightarrow{t_1} \sigma_2 \xrightarrow{t_2} \sigma_3 \xrightarrow{t_3} \ldots \sigma_{N-1} \xrightarrow{t_{N-1}} \sigma_N.$$

Since all $(N-1)!$ orderings of the connection times $t_1, \ldots, t_{N-1}$ are equiprobable, the combinatorial shape of the resulting coalescent tree is distributed as the combinatorial tree $\mathrm{SHAPE}\left(T_{\mathsf{wn}}^{(N)}\right)$, where all $(N-1)!$ orderings of the analogous connection times $W_1^{(N)}, W_2^{(N)}, \ldots, W_{N-1}^{(N)}$ are also equiprobable. $\qquad \square$

The following result is a consequence of the above Thm. 23 and Thm. 25 that we state and prove in Sect. 9 establishing the root-Horton law (Def. 19) for Kingman's coalescent tree $\mathrm{SHAPE}\left(T_{\mathrm{K}}^{(N)}\right)$.

Corollary 17. *The combinatorial level set tree of a discrete white noise $W_j^{(N)}$ is root-Horton self similar with the same Horton exponent R as that for Kingman's N-coalescent.*

Proof. Together, Thms. 23 and 25 imply the root-Horton self-similarity for $\mathrm{SHAPE}\left(T_{\mathsf{wn}}^{(N)}\right)$, with the same Horton exponent R.

By definition, Horton pruning corresponds to an index shift in Horton statistics: $N_j\big[\mathcal{R}(T)\big] = N_{j+1}[T]$ $(j \geq 1)$. Thus, the root-Horton self-similarity for $\mathrm{SHAPE}\left(T_{\mathsf{wn}}^{(N)}\right)$ implies the root-Horton self-similarity for $\mathrm{SHAPE}\left(\mathrm{LEVEL}\big(W_j^{(N)}\big)\right)$. Finally, the Horton exponent is preserved under the extra Horton pruning as

$$\lim_{j \to \infty} \left(\mathcal{N}_{j+1}\right)^{-\frac{1}{j}} = \lim_{j \to \infty} \left(\mathcal{N}_j\right)^{-\frac{1}{j}} = R. \qquad \square$$
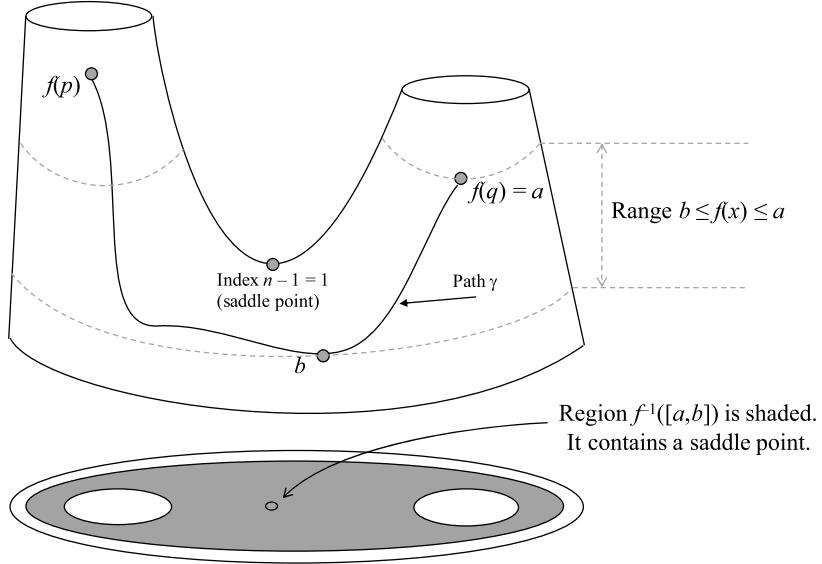
FIG 33. *Illustration to Lem. 23 (a counterexample). Here, a function $f : M \subset \mathbb{R}^2 \to \mathbb{R}$ is such that the region $f^{-1}([a,b])$, which is shaded in the bottom part, contains a saddle (critical point of index $n - 1 = 1$); hence the conditions of the lemma are violated. Observe, accordingly, that the image of any path $\gamma : p \to q$ must go below the point $a = f(q)$ by a finite amount, i.e., there exists $\delta > 0$ such that $\gamma \not\subset \mathcal{L}_{a-\delta}$.*

8.9. Level set trees on higher dimensional manifolds and Morse theory

Here we give a brief introduction to the theory of level set trees of Morse functions on multidimensional manifolds. This section is intended to outline a natural next step beyond the theory of level set trees for real functions on $\mathbb{R}$ discussed earlier. The reader, however, will not find here any results on self-similarity of trees under pruning, the main topic of this survey.

Consider an n-dimensional differentiable manifold $M = M^n$, and a differentiable function $f : M \to \mathbb{R}$. A point p is called a *critical point* of f if $df(p) = 0$, in which case, $f(p)$ is said to be a *critical value* of f. A point $x \in M$ is called a *regular point* of f if it is not a critical point.

If p is a critical point of f, then

$$f(x) = f(p) + \frac{1}{2}H_{f,p}(x, x) + O(3)$$

is the Taylor expansion of f around p, where

$$H_{f,p}(u, v) = \sum_{i,j} \frac{\partial^2 f}{\partial x_i \, \partial x_j}(p) \, u_i v_j \ : \ T_pM \times T_pM \to \mathbb{R}$$

(a) $f(x)$: $M \subset \mathbb{R}^2 \mapsto \mathbb{R}$ (b) LEVEL(f)
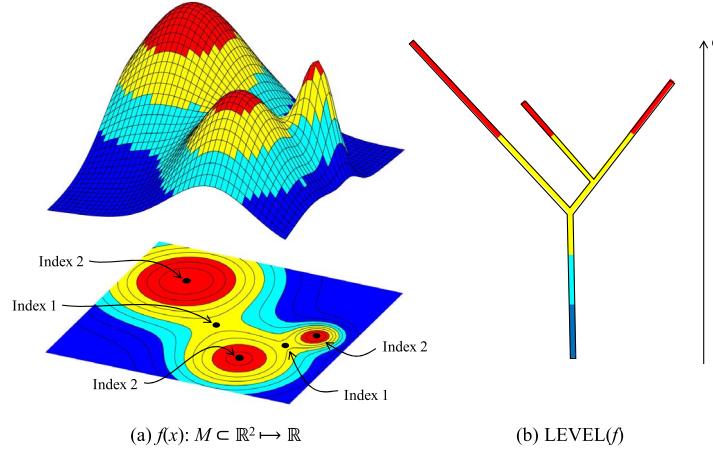
FIG 34. *Level set tree of a Morse function: An illustration. (a) A Morse function $f(x)$: $M \subset \mathbb{R}^2 \to \mathbb{R}$ (top) and its level sets $\mathcal{L}_\alpha$ (bottom). (b) The level set tree* LEVEL(f) *shows how distinct components of $\mathcal{L}_\alpha$ merge as threshold α decreases. The color code illustrates the value of $f(x)$ at different level lines. Each critical point of index 2 (local maximum) corresponds to a leaf. In this figure, each critical point of index 1 (saddle) corresponds to an internal vertex.*

is a *symmetric bilinear form* over the tangent space $T_p M$ generated by the *Hessian* matrix $\frac{\partial^2 f}{\partial x_i \, \partial x_j}(p)$, and $O(3)$ denotes the third and higher order terms.

Definition 32 (**Nondegenerate points and Morse functions** [**127**]). *Let M and f to be as above. A critical point $p \in M$ of f is said to be nondegenerate if the determinant of its Hessian matrix $\frac{\partial^2 f}{\partial x_i \, \partial x_j}(p)$ is not equal to zero. A differentiable function $f : M \to \mathbb{R}$ is said to be a Morse function if all of its critical points are nondegenerate.*

Theorem 24 (**Morse,** [**127**]). *Consider an n-dimensional differentiable manifold M, and a differentiable function $f : M \to \mathbb{R}$. If $p \in M$ is a nondegenerate critical point of f, then there exists an open neighborhood U of p and local coordinates $(x_1, \ldots, x_n)$ on U with*

$$\big(x_1(p), \ldots, x_n(p)\big) = (0, \ldots, 0)$$

such that in this coordinates $f(x)$ is a quadratic polynomial represented as

$$f(x) = f(p) + \frac{1}{2} H_{f,p}(x, x).$$

If $B(u, v) : V \times V \to \mathbb{R}$ is a nondegenerate (i.e., with non-zero determinant) symmetric bilinear form over an n-dimensional vector space V, then there exists a unique nonnegative integer $\lambda \leq n$ and at least one basis $\mathcal{B}$ of V such that, in basis $\mathcal{B}$,

$$B(x, x) = -x_1^2 - \ldots - x_\lambda^2 + x_{\lambda+1}^2 + \ldots + x_n^2.$$

This implies the following corollary to the Morse Theorem (Thm. 24), known as the Morse Lemma.

Corollary 18 (**Morse Lemma** [**127**]). *Consider an n-dimensional differentiable manifold M, and a differentiable function $f : M \to \mathbb{R}$. If $p \in M$ is a nondegenerate critical point of f, then there exists and open neighborhood U of p and local coordinates $(x_1, \ldots, x_n)$ on U with*

$$\big(x_1(p), \ldots, x_n(p)\big) = (0, \ldots, 0)$$

such that in this coordinates,

$$f(x) = f(p) - x_1^2 - \ldots - x_\lambda^2 + x_{\lambda+1}^2 + \ldots + x_n^2.$$

The integer λ in Cor. 18 is called the *index* of the nondegenerate critical point $p \in M$. The next lemma concerns directly the structure of the level set trees for $f : M \to \mathbb{R}$. Let M and f to be as above. Following the one-dimensional setup of Sect. 8.2.1, for $\alpha \in \mathbb{R}$ we consider the level set

$$\mathcal{L}_\alpha = \mathcal{L}_\alpha(f) = \{x \in M \,:\, f(x) \geq \alpha\}.$$

Lemma 23 ([**121**, **36**]). *Consider an n-dimensional differentiable manifold M, and a Morse function $f : M \to \mathbb{R}$. Given points $p, q \in M$ and a differentiable curve $\gamma : [0, 1] \to M$ such that $\gamma(0) = p$ and $\gamma(1) = q$. Let $a = \min\big\{f(p), f(q)\big\}$ be the minimal endpoint value, and let $b = \min\limits_{t \in [0,1]} \big(f \circ \gamma(t)\big)$.*

Suppose $f^{-1}\big([a, b]\big)$ is compact and does not contain any critical points of index n or $n - 1$. Then, for any $\delta > 0$, there exists a differentiable curve $\widetilde{\gamma} : [0, 1] \to M$ homotopic to γ such that $\widetilde{\gamma}(0) = p$ and $\widetilde{\gamma}(1) = q$, and

$$\widetilde{\gamma}\big([0, 1]\big) \subset \mathcal{L}_{a-\delta}.$$

Consider an n-dimensional compact differentiable manifold M, and a Morse function $f : M \to \mathbb{R}$. Recalling the definition of a level set tree in dimension one, for $p, q \in M$, let

$$\underline{f}(p, q) := \sup_{\gamma : p \to q} \inf_{x \in \gamma} f(x),$$

where the supremum is taken over all continuous curves $\gamma : [0, 1] \to M$ such that $\gamma(0) = p$ and $\gamma(1) = q$. Next, as it was the case when $\dim(M) = 1$, we define a *pseudo-metric* on M as

$$d_f(p, q) := \big(f(p) - \underline{f}(p, q)\big) + \big(f(q) - \underline{f}(p, q)\big), \quad p, q \in M. \tag{191}$$

We write $p \sim_f q$ if $d_f(p, q) = 0$, and observe that d_f is a metric over the quotient space $M/\sim_f$. Thus, $(M/\sim_f, d_f)$ is a metric space, satisfying Def. 1 of a tree. This tree will be called the *level set tree* of f, and denoted by LEVEL(f). Here, $d_f(p, q) \geq |f(p) - f(q)|$, with $d_f(p, q) = |f(p) - f(q)|$ if and only if points $(p/\sim_f)$ and $(q/\sim_f)$ of LEVEL(f) belong to the same lineage. In particular, if $d_f(p, q) = f(p) - f(q)$, then $(p/\sim_f)$ is the descendant point to $(q/\sim_f)$, and respectively, $(q/\sim_f)$ is the ancestral point to $(p/\sim_f)$. Figures 34, 35 show examples of level set trees for functions f on $\mathbb{R}^2$.

(a) $f(x)$: $M \subset \mathbb{R}^2 \mapsto \mathbb{R}$ (b) LEVEL(f)
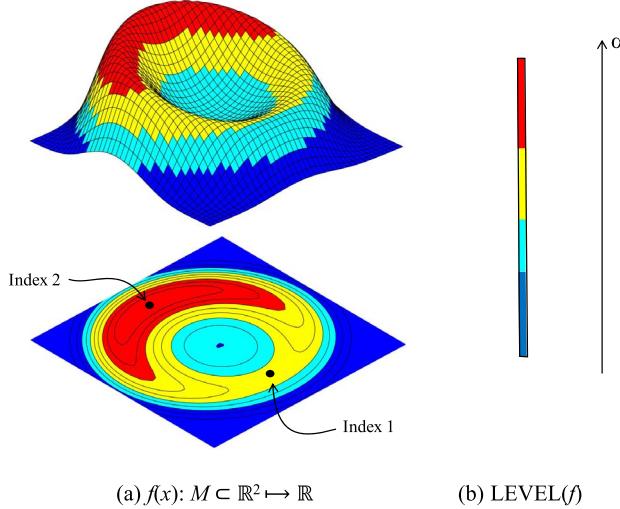
FIG 35. *Level set tree of a Morse function: An illustration. In this figure, the critical point of index* 1 *(saddle) does not correspond to an internal vertex. The rest of notations are the same as in Fig. 34.*

Example 16 (**Compactness requirement**). The requirement for the manifold M to be compact is necessary to ensure that there are no pairs of disjoint closed sets such that the distance between the two sets equals zero. As a counterexample, consider a function $f(x,y) = x^2 - e^y$ on $M = \mathbb{R}^2$ (Fig. 36). Here, the level set $\mathcal{L}_0$ consists of two nonintersecting closed regions, marked by gray shading in Fig. 36(b):

$$A = \{f(x,y) \geq 0, x > 0\} = \left\{x \geq e^{y/2}\right\}$$

and

$$B = \{f(x,y) \geq 0, x < 0\} = \left\{x \leq -e^{y/2}\right\}.$$

The distance between A and B is zero, as the two sets get arbitrary close along the line $x = 0$ as $y \to -\infty$. Consider points $p = (e, 2) \in A$ and $q = (-e, 2) \in B$ marked in Fig. 36. The points p and q are not connected by a continuous path inside $\mathcal{L}_0$, since each such a path must intersect the line $x = 0$ along which $f < 0$. Yet, if we were to extend the distance in (191) to $M = \mathbb{R}^2$, then $\underline{f}(p, q) = 0$ since for any $\delta > 0$ there exists a path similar to γ in Fig. 36(b), with the tip on the line $x = 0$ for large enough y, so that $\gamma \subset \mathcal{L}_{-\delta}$. Consequently, we have $d_f(p, q) = 0$ implying that the points p and q are equivalent on the level set tree of f, $p =_{\sim_f} q$, albeit they belong to two disconnected components of $\mathcal{L}_0$.

Naturally, if $f : M \to \mathbb{R}$ is a Morse function, the critical points of index n (local maxima) correspond to the leaves of the level set tree LEVEL(f). As we decrease α, new segments of $\mathcal{L}_\alpha$ appear at the critical points of index n, and disconnected components of $\mathcal{L}_\alpha$ merge at some critical points of index less than

(a) $f(x,y) = x^2 - e^y$ (b) Level set $\mathcal{L}_0$ (gray)
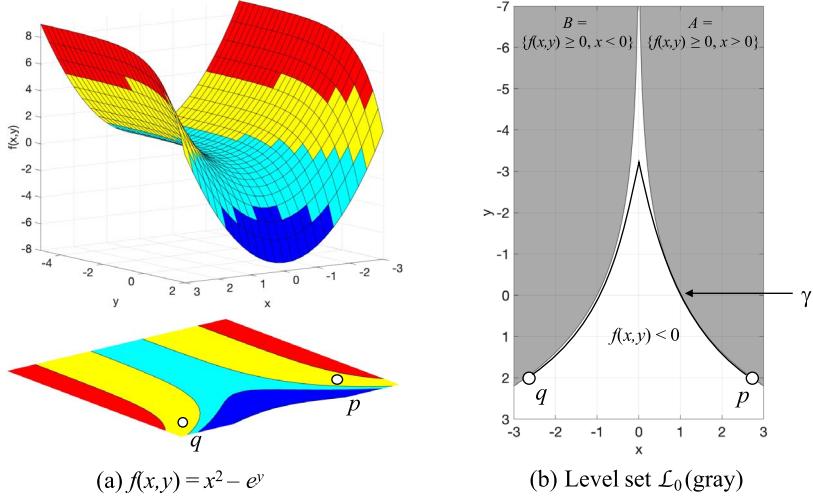
FIG 36. *Illustration to Example 16. The manifold M must be compact to properly define the level set tree of a function $f : M \to \mathbb{R}$. In this example, $M = \mathbb{R}^2$ is not compact. This allows for the existence of points p and q such that $d_f(p,q) = 0$, while they belong to disconnected components of $\mathcal{L}_0$.*

n. If M is a compact manifold and $f : M \to \mathbb{R}$ is a Morse function, then by Lem. 23 the critical points of index less than $n-1$ cannot be the merger points of separated pieces of $\mathcal{L}_\alpha$. Thus, we obtain the following corollary of Lem. 23.

Corollary 19. *Consider an n-dimensional compact differentiable manifold M, and a Morse function $f : M \to \mathbb{R}$. Then, there is a bijection between the leaves of* LEVEL(f) *and the critical points of f of index n, and a one-to-one (but not necessarily onto) correspondence between the internal (non-leaf) vertices of* LEVEL(f) *and the critical points of f of index $n-1$.*

Proof. Suppose $c \in M$ is a critical point of f of index less than $n-1$ such that $(c/\sim_f)$ is an internal (non-leaf) vertex of LEVEL(f). Then, $(c/\sim_f)$ is a parent vertex to at least one pair of points $(p/\sim_f)$ and $(q/\sim_f)$ of LEVEL(f) that do not belong to the same lineage, $\underline{f}(p,q) = f(c)$, and therefore

$$d_f(p,q) = f(p) + f(q) - 2f(c) = |f(p) - f(c)| + 2(a - f(c)) > |f(p) - f(q)|, \ (192)$$

where $a = \min\{f(p), f(q)\}$. Thus, since M is a differentiable manifold, there exists a differentiable curve $\gamma : [0,1] \to M$ such that $\gamma(0) = p$ and $\gamma(1) = q$, and $\min_{t \in [0,1]} (f \circ \gamma(t)) = f(c)$. Then, by Lemma 23, for any $\delta > 0$, there exists a differentiable curve $\widetilde{\gamma} : [0,1] \to M$ homotopic to γ such that $\widetilde{\gamma}(0) = p$ and $\widetilde{\gamma}(1) = q$, and

$$\widetilde{\gamma}([0,1]) \subset \mathcal{L}_{a-\delta}.$$

Hence,

$$d_f(p,q) \leq f(p) + f(q) - 2(a - \delta) = |f(p) - f(q)| + 2\delta$$

for any $\delta > 0$. Therefore, $d_f(p, q) = |f(p) - f(q)|$, contradicting (192), i.e., contradicting the assumption that $(p/\sim_f)$ and $(q/\sim_f)$ do not belong to the same lineage in LEVEL(f). □

Remark 19. Corollary 19 asserts that while every internal vertex of the level set tree corresponds to a critical point of index 1, not every critical point of index 1 may correspond to an internal vertex. Figure 34 shows an example of a function where every critical point of index 1 (saddle) corresponds to an internal vertex. Figure 35 shows an example of a function where the critical point of index 1 (saddle) does not corresponds to an internal vertex.

Finally, Cor. 19 together with Morse Lemma (Cor. 18) imply the following lemma.

Lemma 24. *Consider an n-dimensional compact differentiable manifold M, and a Morse function $f : M \to \mathbb{R}$. Suppose there is no two distinct critical points p and q of index $n - 1$ with the same value $f(p) = f(q)$. Then, the level set tree* LEVEL(f) *is binary.*

Note that an alternative commonly used definition of Morse function is the one where we require that all of its critical points are nondegenerate and have distinct function values, e.g. [56]. For so defined Morse function, the statement of Lem. 24 will only call for $f : M \to \mathbb{R}$ to be a Morse function.

Proof. Suppose p is a critical point of f corresponding to an internal (non-leaf) vertex in LEVEL(f). Then, by Corollary 19, p has index $\lambda = n - 1$. Corollary 18 asserts that there exists and open neighborhood U of p and local coordinates $(x_1, \ldots, x_n)$ on U with

$$\big(x_1(p), \ldots, x_n(p)\big) = (0, \ldots, 0)$$

such that in this coordinates,

$$f(x) = f(p) - x_1^2 - \ldots - x_{n-1}^2 + x_n^2.$$

Hence, as α decreases, the merger of distinct components of $\mathcal{L}_\alpha$ happens along the x_n-coordinate axis. This allows for the merger of at most two components. □

Vladimir Arnold studied an alternative (albeit similar in spirit) construction of level set trees that he called the *graph of Morse function* $f : M \to \mathbb{R}$, concentrating mainly on the spheres $M = S^2$; see [9, 10, 11] and references therein. Arnold has shown that these graphs are binary trees as well. These trees are constructed in such a way that both the local minima (index 0) and the local maxima (index 2) points of f correspond to the leaves, while the saddle points (index 1) correspond to the internal (non-leaf) vertices. The goal of Arnold's study was to shed light on the problem of classifying all possible configurations of the horizontal lines on the topographical maps formulated by A. Cayley in 1868. In [11], Arnold quotes a communication with Morse: *M. Morse has told me, in 1965, that the problem of the description of the possible*

combinations of several critical points of a smooth function on a manifold looks hopeless to him. L. S. Pontrjagin and H. Whitney were of the same opinion. Arnold's work of topological classification of level lines for Morse functions on S^2 enriched the collection of questions accompanying the Hilbert's sixteenth problem, which promoted the study of the topological structures of the level lines of real polynomials $p(x)$ over $x \in \mathbb{R}^n$, [83, 10, 11].

9. Kingman's coalescent process

We refer to a general definition of a coalescent process in Section 8.8.1. Recall that in an N-particle coalescent process, a pair of clusters with masses i and j coalesces at the rate $K(i,j)/N$. The mass-independent rate $K(i,j) = 1$ defines the Kingman's coalescent process [89]. The following result establishes a weak form of Horton law for Kingman's coalescent.

Theorem 25 (**Root-Horton law for Kingman's coalescent,** [**97**]). *Consider Kingman's N-coalescent process and its tree representation $T_{\mathrm{K}}^{(N)}$. Let $N_j = N_j^{(N)}$ denote the number of branches of Horton-Strahler order j in the tree $T_{\mathrm{K}}^{(N)}$.*

(i) The asymptotic Horton ratios $\mathcal{N}_j$ exist and are finite for all $j \in \mathbb{N}$, as in Def. 18. That is, for each j, the following limit exists and is finite:

$$N_j^{(N)}/N \xrightarrow{p} \mathcal{N}_j \quad as \ N \to \infty. \tag{193}$$

(ii) Furthermore, $\mathcal{N}_j$ satisfy the root-Horton law (Def. 19):

$$\lim_{j \to \infty} (\mathcal{N}_j)^{-\frac{1}{j}} = R$$

with Horton exponent $2 \leq R \leq 4$.

9.1. Smoluchowski-Horton ODEs for Kingman's coalescent

In this section we provide a heuristic derivation of Smoluchowski-type ODEs for the number of Horton-Strahler branches in the coalescent tree $T_{\mathrm{K}}^{(N)}$ and consider the asymptotic version of these equations as $N \to \infty$. Section 9.2 formally establishes the validity of the hydrodynamic limit.

Recall that $K(i,j) \equiv 1$. Let $|\Pi_t^{(N)}|$ denote the total number of clusters at time $t \geq 0$, and let $\eta_{(N)}(t) := |\Pi_t^{(N)}|/N$ be the total number of clusters relative to the system size N. Then $\eta_{(N)}(0) = N/N = 1$ and $\eta_{(N)}(t)$ decreases by $1/N$ with each coalescence of clusters; this happens with the rate

$$\frac{1}{N} \binom{N \eta_{(N)}(t)}{2} = \frac{\eta_{(N)}^2(t)}{2} \cdot N + o(N), \quad \text{as } N \to \infty,$$

since $1/N$ is the coalescence rate for any pair of clusters regardless of their masses. Informally, this implies that the large-system limit relative number of

clusters $\eta(t) = \lim\limits_{N\to\infty} \eta_{(N)}(t)$ satisfies the following ODE:[3]

$$\frac{d}{dt}\eta(t) = -\frac{\eta^2(t)}{2}. \tag{194}$$

The initial condition $\eta(0) = 1$ implies a unique solution $\eta(t) = 2/(2+t)$. The existence of the limit $\eta(t)$ is established in Lem. 25(a) of Sect. 9.2.

Next, for any $k \in \mathbb{N}$ we write $\eta_{k,N}(t)$ for the relative number of clusters (with respect to the system size N) that correspond to branches of Horton-Strahler order k in tree $T_{\mathrm{K}}^{(N)}$ at time t. Initially, each particle represents a leaf of Horton-Strahler order 1. Accordingly, the initial conditions are set to be, using Kronecker's delta notation,

$$\eta_{k,N}(0) = \delta_1(k).$$

Below we describe the evolution of $\eta_{k,N}(t)$ using the definition of Horton-Strahler orders.

Observe that $\eta_{k,N}(t)$ increases by $1/N$ with each coalescence of clusters of Horton-Strahler order $k-1$ that happens with the rate

$$\frac{1}{N}\binom{N\,\eta_{k-1,N}(t)}{2} = \frac{\eta_{k-1,N}^2(t)}{2}\cdot N + o(N).$$

Thus $\frac{\eta_{k-1,N}^2(t)}{2} + o(1)$ is the instantaneous rate of increase of $\eta_{k,N}(t)$.

Similarly, $\eta_{k,N}(t)$ decreases by $1/N$ when a cluster of order k coalesces with a cluster of order strictly higher than k that happens with the rate

$$\eta_{k,N}(t)\left(\eta_{(N)}(t) - \sum_{j=1}^{k}\eta_{j,N}(t)\right)\cdot N,$$

and it decreases by $2/N$ when a cluster of order k coalesces with another cluster of order k that happens with the rate

$$\frac{1}{N}\binom{N\,\eta_{k,N}(t)}{2} = \frac{\eta_{k,N}^2(t)}{2}\cdot N + o(N).$$

Thus the instantaneous rate of decrease of $\eta_{k,N}(t)$ is

$$\eta_{k,N}(t)\left(\eta_{(N)}(t) - \sum_{j=1}^{k}\eta_{j,N}(t)\right) + \eta_{k,N}^2(t) + o(1).$$

---

[3]A similar heuristic argument leading to the hydrodynamic limit $\eta(t)$ satisfying (194) was given in [25, proof of Thm. 2.1]. Furthermore, a rigorous argument employing similar hydrodynamical limits and martingale inequalities is used in [23] to obtain an asymptotic estimate for the total length of a Λ-coalescent tree.

We can informally write the limit rates-in and the rates-out for the clusters of Horton-Strahler order via the following *Smoluchowski-Horton system* of ODEs:

$$\frac{d}{dt}\eta_k(t) = \frac{\eta_{k-1}^2(t)}{2} - \eta_k(t)\left(\eta(t) - \sum_{j=1}^{k-1}\eta_j(t)\right), \tag{195}$$

with the initial conditions $\eta_k(0) = \delta_1(k)$. Here we interpret $\eta_k(t)$ as the hydrodynamic limit of $\eta_{k,N}(t)$ as $N \to \infty$, which will be rigorously established in Lem. 25(b) of Sect. 9.2. We also let $\eta_0 \equiv 0$.

Since $\eta_k(t)$ has the instantaneous rate of increase $\eta_{k-1}^2(t)/2$, the relative total number of clusters corresponding to branches of Horton-Strahler order k is then

$$\mathcal{N}_k = \delta_1(k) + \int_0^\infty \frac{\eta_{k-1}^2(t)}{2}dt. \tag{196}$$

This equation has a simple heuristic interpretation. Specifically, according to the Horton-Strahler rule (5), a branch of order $k > 1$ can only be created by merging two branches of order $k-1$. In Kingman's coalescent process these two branches are selected at random from all pairs of branches of order $k-1$ that exist at instant t. As N goes to infinity, the asymptotic density of a pair of branches of order $(k-1)$, and hence the instantaneous intensity of newly formed branches of order k, is $\eta_{k-1}^2(t)/2$. The integration over time gives the relative total number of order-k branches. The validity of equation (196) is established within the proof of Thm. 25(i) that follows Lem. 25.

It is not hard to compute the first three terms of the sequence $\mathcal{N}_k$ by solving equations (194) and (195) in the first three iterations:

$$\mathcal{N}_1 = 1, \quad \mathcal{N}_2 = \frac{1}{3}, \quad \text{and} \quad \mathcal{N}_3 = \frac{e^4}{128} - \frac{e^2}{8} + \frac{233}{384} = 0.109686868100941\ldots$$

Hence, we have $\mathcal{N}_1/\mathcal{N}_2 = 3$ and $\mathcal{N}_2/\mathcal{N}_3 = 3.038953879388\ldots$ Our numerical results yield, moreover,

$$\lim_{k\to\infty} (\mathcal{N}_k)^{-\frac{1}{k}} = \lim_{k\to\infty} \frac{\mathcal{N}_k}{\mathcal{N}_{k+1}} = 3.0438279\ldots$$

9.2. Hydrodynamic limit

This section establishes the existence of the asymptotic ratios $\mathcal{N}_k$ of (193) as well as the validity of the equations (194), (195) and (196) in a hydrodynamic limit. We refer to Darling and Norris [42] for a survey of techniques for establishing convergence of a Markov chain to the solution of a differential equation.

Notice that if the first $k-1$ functions $\eta_1(t), \ldots, \eta_{k-1}(t)$ are given, then (195) is a linear equation in $\eta_k(t)$. This *quasilinearity* implies the existence and uniqueness of a solution.

We now proceed with establishing a hydrodynamic limit for the Smoluchowski-Horton system of ODEs (195). Let

$$\eta_{k,N}(t) := \frac{N_k(t)}{N} \quad \text{and} \quad g_{k,N}(t) := \eta_{(N)}(t) - \sum_{j:j<k} \eta_{j,N}(t).$$

Lemma 25. *Let $\eta_{(N)}(t)$ be the relative total number of clusters and $\eta(t)$ be the solution to equation (194) with the initial condition $\eta(0) = 1$. Let $\eta_{k,N}(t)$ denote the relative number of clusters that correspond to branches of Horton-Strahler order k and let functions $\eta_k(t)$ solve the system of equations (195) with the initial conditions $\eta_k(0) = \delta_1(k)$. Then, as $N \to \infty$,*

(a) $\left\| \eta_{(N)}(t) - \eta(t) \right\|_{L^\infty[0,\infty)} \xrightarrow{p} 0$;

(b) $\| \eta_{k,N}(t) - \eta_k(t) \|_{L^\infty[0,\infty)} \xrightarrow{p} 0, \ \forall k \geq 1$.

Proof. We adopt here the approach of [95] that uses the weak limit law established in [60, Theorem 2.1, Chapter 11] and [103, Theorem 8.1]; it is briefly explained in Appendix A of this manuscript. This approach is different from the original proof given in [97], and also from the method developed in Norris [128] for the Smoluchowski equations.

For a fixed positive integer K, let

$$\hat{X}_N(t) = \Big(N_1(t), N_2(t), \ldots, N_K(t), N(t) \Big) \in \mathbb{Z}_+^{K+1}$$

with $\hat{X}_N(0) = N\mathbf{e}_1$. The process $\hat{X}_N(t)$ is a finite dimensional Markov process. Its transition rates can be found using the formalism (232) for density dependent population processes. Specifically, let $x = (x_1, x_2, \ldots, x_{K+1})$. Then, for any $1 \leq k \leq K$, the change vector $\ell = -\mathbf{e}_k - \mathbf{e}_{K+1}$ corresponding to a merger of a cluster of order k into a cluster of order higher than k has the rate

$$q^{(n)}(x, x + \ell) = \frac{1}{N} x_k \left(x_{K+1} - \sum_{j=1}^{k} x_j \right) = N\beta_\ell \left(\frac{x}{N} \right),$$

where $\beta_\ell(x) = x_k \left(x_{K+1} - \sum_{j=1}^{k} x_j \right)$. For a given k such that $1 \leq k \leq K$, the change vector

$$\ell = -2\mathbf{e}_k + \mathbf{e}_{k+1}\mathbf{1}_{k<K} - \mathbf{e}_{K+1}$$

corresponding to a merger of a pair of clusters of order k is assigned the rate

$$q^{(n)}(x, x + \ell) = \frac{1}{N} \left[\frac{x_k^2}{2} - \frac{x_k}{2} \right] = N \left[\beta_\ell \left(\frac{x}{N} \right) + O\left(\frac{1}{n} \right) \right], \qquad (197)$$

where $\beta_\ell(x) = \frac{x_k^2}{2}$. Finally, the change vector $\ell = -\mathbf{e}_{K+1}$ corresponding to a merger of two clusters, both of order greater than K, is assigned the rate

$$q^{(n)}(x, x + \ell) = \frac{1}{N} \left[\frac{x_{K+1}^2}{2} - \frac{x_{K+1}}{2} \right] = N \left[\beta_\ell \left(\frac{x}{N} \right) + O\left(\frac{1}{n} \right) \right],$$

where $\beta_\ell(x) = \frac{x_{K+1}^2}{2}$.

By Thm. 35, $X_N(t) = N^{-1}\hat{X}_N(t)$ converges to $X(t)$ as in (235), where $X(t)$ satisfies (234) with

$$F(x) := \sum_\ell \ell\beta_\ell(x) = \sum_{k=1}^K x_k \left(x_{K+1} - \sum_{j=1}^k x_j \right) [-\mathbf{e}_k - \mathbf{e}_{K+1}]$$

$$+ \frac{1}{2} \sum_{k=1}^{K+1} x_k^2 [-2\mathbf{e}_k \mathbf{1}_{k \leq K} + \mathbf{e}_{k+1} \mathbf{1}_{k < K} - \mathbf{e}_{K+1}]$$

$$= \sum_{k=1}^K \left(\frac{x_{k-1}^2}{2} - x_k \left(x_{K+1} - \sum_{j=1}^{k-1} x_j \right) \right) \mathbf{e}_k - \frac{x_{K+1}^2}{2} \mathbf{e}_{K+1}, \qquad (198)$$

where we let $x_{-1} = 0$ at all times. Here, $F(x)$ naturally satisfies the Lipschitz continuity conditions (233), and the initial conditions $X(0) = X_n(0) = \mathbf{e}_1$.

Therefore, for a given integer $K > 0$ and a fixed real $T > 0$, equation (234) in Thm. 35 with $F(x)$ as in (198) yields

$$\lim_{N \to \infty} \sup_{s \in [0,T]} \left| N^{-1}\eta_{(N)}(s) - \eta(s) \right| = 0 \qquad \text{a.s.} \qquad (199)$$

and

$$\lim_{N \to \infty} \sup_{s \in [0,T]} \left| N^{-1}\eta_{k,N}(s) - \eta_k(s) \right| = 0 \qquad \text{a.s.} \qquad (200)$$

for all $k = 1, 2, \ldots, K$, with $\eta_{(N)}$ satisfying (194) and $\eta_{k,N}$ satisfying the system of Smoluckowski-Horton system of ODEs (195).

Let T_m be the time when the first m clusters merge. The expectation for the time T_m is

$$E[T_m] = \frac{N}{\binom{N}{2}} + \frac{N}{\binom{N-1}{2}} + \cdots + \frac{N}{\binom{N-m+1}{2}} = \frac{2m}{N-m}. \qquad (201)$$

For given $\epsilon \in (0,1)$ and $\gamma > 1$ let $m = \lfloor (1-\epsilon)N \rfloor$. Taking $T > \frac{2(1-\epsilon)}{\epsilon}\gamma$, we have for all $t \geq T$,

$$0 < \eta(t) \leq \eta(T) < \eta\left(\frac{2(1-\epsilon)}{\epsilon}\gamma \right) < \eta\left(\frac{2(1-\epsilon)}{\epsilon} \right) = \epsilon.$$

Thus $\left| \eta_{(N)}(t) - \eta(t) \right| > \epsilon$ would imply $\eta_{(N)}(t) > \epsilon > \eta(t) > 0$, and by Markov's inequality, we obtain

$$P\left(\left\| \eta_{(N)}(t) - \eta(t) \right\|_{L^\infty[T,\infty)} > \epsilon \right) \leq P\left(\eta_{(N)}(T) > \epsilon \right) = P\left(T_m > T \right)$$

$$\leq \frac{2(1-\epsilon)}{\epsilon T} < 1/\gamma. \qquad (202)$$

Together (199) and the above equation (202) imply

$$\lim_{N \to \infty} P\left(\left\| \eta_{(N)}(t) - \eta(t) \right\|_{L^\infty[0,\infty)} < \epsilon \right) = 1.$$

Hence $\|\eta_{(N)}(t) - \eta(t)\|_{L^\infty[0,\infty)} \to 0$ in probability, establishing Lemma 25(a).

Finally, observe that for any $\epsilon > 0$ and for $T > 0$ large enough so that $\eta(T) < \epsilon$,

$$\eta_k(t) \leq \eta(t) \leq \eta(T) < \epsilon \text{ for all } t \geq T.$$

Thus,

$$
\begin{aligned}
P\Big(\big\|\eta_{k,N}(t) - \eta_k(t)\big\|_{L^\infty[T,\infty)} > \epsilon\Big) &\leq P\Big(\big\|\eta_{k,N}(t)\big\|_{L^\infty[T,\infty)} > \epsilon\Big) \\
&\leq P\Big(\big\|\eta_{(N)}(t)\big\|_{L^\infty[T,\infty)} > \epsilon\Big) \\
&= P\Big(\eta_{(N)}(T) > \epsilon\Big) \\
&\leq \frac{2(1-\epsilon)}{\epsilon T},
\end{aligned}
\tag{203}
$$

where the last bound is obtained from Markov inequality: for $m = \lfloor(1-\epsilon)N\rfloor$,

$$P\Big(\eta_{(N)}(T) > \epsilon\Big) = P(T_m > T) \leq \frac{E[T_m]}{T} = \frac{2m}{(N-m)T} \leq \frac{2(1-\epsilon)}{\epsilon T}$$

by (201). Together, equations (200) and (203) imply

$$\|\eta_{k,N} - \eta_k\|_{L^\infty[0,\infty)} \xrightarrow{p} 0 \qquad \forall k \geq 1. \qquad \square$$

Consequently, we establish a hydrodynamic limit for the Horton ratios (Thm. 25(i)) and validate formula (196).

Proof of Theorem 25(i). The existence of the limit $\mathcal{N}_j = \lim_{N\to\infty} N_j/N$ in probability and its expression (196) via the solution $\eta_{(N)}(t)$ of (194) follows from (197) in the context of Theorem 35 and the tail bound (202). $\qquad \square$

9.3. Some properties of the Smoluchowski-Horton system of ODEs

Here we restate the Smoluchowski-Horton system of ODEs (195) as a simpler quasilinear system of ODEs (205), which we later (Sect. 9.3.2) rescale to the interval $[0,1]$ (208). Some of the properties established in Prop. 17 and Lem. 26 of this section are used in the proof of Thm. 25(ii) in Sect. 9.4.

9.3.1. Simplifying the Smoluchowski-Horton system of ODEs

Let $g_1(t) = \eta(t)$ and $g_k(t) = \eta(t) - \sum_{j:\ j<k} \eta_j(t)$ be the asymptotic number of clusters of Horton-Strahler order k or higher at time t. We can rewrite (195) via g_k using $\eta_k(t) = g_k(t) - g_{k+1}(t)$:

$$\frac{d}{dt}g_k(t) - \frac{d}{dt}g_{k+1}(t) = \frac{\big(g_{k-1}(t) - g_k(t)\big)^2}{2} - (g_k(t) - g_{k+1}(t))g_k(t).$$

We now rearrange the terms, obtaining for all $k \geq 2$,

$$\frac{d}{dt}g_{k+1}(t) - \frac{g_k^2(t)}{2} + g_k(t)g_{k+1}(t) = \frac{d}{dt}g_k(t) - \frac{g_{k-1}^2(t)}{2} + g_{k-1}(t)g_k(t). \quad (204)$$

One can readily check that $\frac{d}{dt}g_2(t) - \frac{g_1^2(t)}{2} + g_1(t)\,g_2(t) = 0$; the above equations hence simplify as follows

$$g_{k+1}'(t) - \frac{g_k^2(t)}{2} + g_k(t)g_{k+1}(t) = 0 \quad\quad\quad (205)$$

$$\text{with } g_1(t) = \frac{2}{t+2}, \text{ and } g_k(0) = 0 \text{ for } k \geq 2.$$

Observe that the existence and uniqueness of the solution sequence g_k of (205) follows immediately from the quasilinear structure of the system (205): for a known $g_k(t)$, the next function $g_{k+1}(t)$ is obtained by solving a first-order linear equation.

From (205) one has $g_k(t) > 0$ for all $t > 0$, and similarly, from the equation (195) one has

$$\eta_k(t) = g_k(t) - g_{k+1}(t) > 0 \quad \text{for all } t > 0. \quad\quad (206)$$

Next, returning to the asymptotic ratios $\mathcal{N}_k$, we observe that (204) implies, for $k \geq 2$,

$$\mathcal{N}_k = \int_0^\infty \frac{\eta_{k-1}^2(t)}{2}dt = \int_0^\infty \frac{(g_{k-1}(t) - g_k(t))^2}{2}dt = \int_0^\infty \frac{g_k^2(t)}{2}dt$$

since

$$\frac{(g_{k-1}(t) - g_k(t))^2}{2} = \frac{d}{dt}g_k(t) + \frac{g_k^2(t)}{2},$$

where $0 \leq g_k(t) < g_1(t) \to 0$ as $t \to \infty$, and $\int_0^\infty \frac{d}{dt}g_k(t)dt = g_k(\infty) - g_k(0) = 0$ for $k \geq 2$. Let n_k represent the number of order-k branches relative to the number of order-$(k+1)$ branches:

$$n_k := \frac{\mathcal{N}_k}{\mathcal{N}_{k+1}} = \frac{\frac{1}{2}\int_0^\infty g_k^2(t)dt}{\frac{1}{2}\int_0^\infty g_{k+1}^2(t)dt} = \frac{\|g_k\|_{L^2[0,\infty)}^2}{\|g_{k+1}\|_{L^2[0,\infty)}^2}. \quad\quad (207)$$

Consider the following limits that represent, respectively, the root and the ratio asymptotic Horton laws:

$$\lim_{k\to\infty} (\mathcal{N}_k)^{-\frac{1}{k}} = \lim_{k\to\infty} \left(\prod_{j=1}^k n_j\right)^{\frac{1}{k}} \quad\text{and}\quad \lim_{k\to\infty} n_k = \lim_{k\to\infty} \frac{\|g_k\|_{L^2[0,\infty)}^2}{\|g_{k+1}\|_{L^2[0,\infty)}^2}.$$

Theorem 25(ii) establishes the existence of the first limit. We expect the second, stronger, limit also to exist and both of them to be equal to $3.043827\ldots$ according to our numerical results. We now establish some basic facts about g_k and n_k.

Proposition 17. *Let $g_k(x)$ solve the ODE system (205). Then*

(a) $\frac{1}{2}\int_0^\infty g_k^2(t)dt = \int_0^\infty g_k(t)g_{k+1}(t)dt,$

(b) $\int_0^\infty g_{k+1}^2(t)dt = \int_0^\infty (g_k(t) - g_{k+1}(t))^2 dt,$

(c) $\lim_{t\to\infty} tg_k(t) = 2,$

(d) $n_k = \dfrac{\|g_k\|_{L^2[0,\infty)}^2}{\|g_{k+1}\|_{L^2[0,\infty)}^2} \geq 2,$

(e) $n_k = \dfrac{\|g_k\|_{L^2[0,\infty)}^2}{\|g_{k+1}\|_{L^2[0,\infty)}^2} \leq 4.$

Proof. Part (a) follows from integrating (205), and part (b) follows from part (a). Part (c) is done by induction, using the L'Hôpital's rule as follows. It is obvious that $\lim_{x\to\infty} tg_1(t) = 2$. Hence, for any $k \geq 1$, (206) implies

$$tg_k(t) \leq tg_1(t) = \frac{2t}{t+2} < 2 \quad \forall t \geq 0.$$

Also,

$$[tg_{k+1}]' = \frac{tg_k^2(t)}{2} - tg_k(t)g_{k+1}(t) + g_{k+1}(t)$$
$$= \frac{\big(g_k(t) - g_{k+1}(t)\big)tg_k(t) + \big(2 - tg_k(t)\big)g_{k+1}(t)}{2}$$

implying $[tg_{k+1}]' \geq 0$ for all $t \geq 0$ as $g_k(t) - g_{k+1}(t) \geq 0$ and $2 - tg_k(t) > 0$. Hence, $tg_{k+1}(t)$ is bounded and nondecreasing. Thus, $\lim_{t\to\infty} tg_{k+1}(t)$ exists for all $k \geq 1$.

Next, suppose $\lim_{t\to\infty} tg_k(t) = 2$. Then by the Mean Value Theorem, for any $t > 0$ and for all $y > t$,

$$\frac{g_{k+1}(t) - g_{k+1}(y)}{t^{-1} - y^{-1}} \leq \sup_{z:\, z\geq t} \frac{g_{k+1}'(z)}{-z^{-2}}.$$

Taking $y \to \infty$, obtain

$$\frac{g_{k+1}(t)}{t^{-1}} \leq \sup_{z:\, z\geq t} \frac{g_{k+1}'(z)}{-z^{-2}}.$$

Therefore

$$\lim_{t\to\infty} tg_{k+1}(t) = \lim_{t\to\infty} \frac{g_{k+1}(t)}{t^{-1}} = \limsup_{z\to\infty} \frac{g_{k+1}'(z)}{-z^{-2}} = \limsup_{z\to\infty} \frac{\frac{g_k^2(z)}{2} - g_k(z)g_{k+1}(z)}{-z^{-2}}$$

$$= \limsup_{z \to \infty} \left[z^2 g_k(z) g_{k+1}(z) - \frac{z^2 g_k^2(z)}{2} \right] = 2 \lim_{t \to \infty} t g_{k+1}(t) - 2$$

implying $\lim_{t \to \infty} t g_{k+1}(t) = 2$.

Statement (d) follows from (207) as we have $\mathcal{N}_k \geq 2\mathcal{N}_{k+1}$ from the definition of the Horton-Strahler order. An alternative proof of (d) using the system of ODEs (208) is given in Sect. 9.3.2.

Part (e) follows from part (a) together with Hölder inequality

$$\frac{1}{2} \|g_k\|_{L^2[0,\infty)}^2 = \int_0^\infty g_k(t) g_{k+1}(t) dt \leq \|g_k\|_{L^2[0,\infty)} \cdot \|g_{k+1}\|_{L^2[0,\infty)},$$

which implies $n_k = \dfrac{\|g_k\|_{L^2[0,\infty)}^2}{\|g_{k+1}\|_{L^2[0,\infty)}^2} \leq 4.$ $\square$

Remark 20. The statements **(a)** and **(b)** of Proposition 17 have a straightforward heuristic interpretation, similar to that of equation (196) above. Specifically, **(a)** claims that the asymptotic relative total number of vertices of order $k+1$ and above in the Kingman's tree (left-hand side) equals twice the asymptotic relative total number of vertices of order $k+1$ and above except the vertices parental to two vertices of order k (right-hand side). This is nothing but the asymptotic property of a binary tree – the number of leaves equals twice the number of internal nodes. The item **(a)** hence merely claims that the Kingman's tree formed by clusters of order above k is binary for any $k \geq 1$. Similarly, item **(b)** claims that the asymptotic relative total number of vertices of order $(k+2)$ and above (left-hand side) equals the asymptotic relative total number of vertices of order $(k+1)$ (right-hand side). This is yet another way of saying that the Kingman's tree is binary.

9.3.2. Rescaling to $[0,1]$ interval

Define
$$h_k(x) = (1-x)^{-1} - (1-x)^{-2} g_{k+1}\left(\frac{2x}{1-x}\right)$$

for $x \in [0,1)$. Then $h_0 \equiv 0$, $h_1 \equiv 1$, and the system of ODEs (205) rewrites as

$$h'_{k+1}(x) = 2h_k(x) h_{k+1}(x) - h_k^2(x) \tag{208}$$

with the initial conditions $h_k(0) = 1$.

Observe that the above quasilinearized system of ODEs (208) has $h_k(x)$ converging to $h(x) = \frac{1}{1-x}$ as $k \to \infty$, where $h(x)$ is the solution to Riccati equation $h'(x) = h^2(x)$ over $[0,1)$, with the initial value $h(0) = 1$. Specifically, we have proven that $g_k(x) \to 0$ as $k \to \infty$. Thus

$$h_k(x) = (1-x)^{-1} - (1-x)^{-2} g_{k+1}\left(\frac{2x}{1-x}\right) \longrightarrow h(x) = \frac{1}{1-x}.$$

Observe that $h_2(x) = (1+e^{2x})/2$, but for $k \geq 3$ finding a closed form expression becomes increasingly hard.

We observe from (207) that the quantity n_k rewrites in terms of h_k as follows

$$n_k = \frac{\left\|1 - h_k/h\right\|_{L^2[0,1]}^2}{\left\|1 - h_{k+1}/h\right\|_{L^2[0,1]}^2}. \tag{209}$$

Consequently, equation (209) implies

$$\lim_{k\to\infty} (\mathcal{N}_k)^{-\frac{1}{k}} = \lim_{k\to\infty} \left(\prod_{j=1}^{k} n_j\right)^{\frac{1}{k}} = \lim_{k\to\infty} \left(\int_0^1 \left(1 - \frac{h_k(x)}{h(x)}\right)^2 dx\right)^{-\frac{1}{k}}. \tag{210}$$

Now, for a known $h_k(x)$, (208) is a first-order linear ODE in $h_{k+1}(x)$. Its solution is given by $h_{k+1}(x) = \mathcal{H}h_k(x)$, where $\mathcal{H}$ is a nonlinear operator defined as follows

$$\mathcal{H}f(x) = \left[1 - \int_0^x f^2(y)e^{-2\int_0^y f(s)ds} dy\right] \cdot e^{2\int_0^x f(s)ds}. \tag{211}$$

Hence, the problem of establishing the limit (210) for the root-Horton law concerns the asymptotic behavior of an iterated nonlinear functional.

The following lemma will be used in Sect. 9.4.

Lemma 26.

$$\left\|1 - h_{k+1}/h\right\|_{L^2[0,1]} = \left\|h_{k+1}/h - h_k/h\right\|_{L^2[0,1]}$$

Proof. Observing $h'_{k+1}(x) + (h_{k+1}(x) - h_k(x))^2 = h_{k+1}^2(x)$, we use integration by parts to obtain

$$\int_0^1 \frac{(h_{k+1}(x) - h_k(x))^2}{h^2(x)} dx = \int_0^1 \frac{h_{k+1}^2(x)}{h^2(x)} dx - \int_0^1 \frac{h'_{k+1}(x)}{h^2(x)} dx$$

$$= \int_0^1 \frac{h_{k+1}^2(x)}{h^2(x)} dx + 1 - 2\int_0^1 \frac{h_{k+1}(x)}{h(x)} dx = \int_0^1 \frac{(1 - h_{k+1}(x))^2}{h^2(x)} dx$$

as $1/h(x) = 1 - x$. $\qquad\square$

Next, we notice that (206) implies

$$h(x) > h_{k+1}(x) > h_k(x) \quad \text{for all} \ \ x \in (0,1) \tag{212}$$

for all $k \geq 1$.

Finally, an alternative proof to Proposition 17(d) using the system of ODEs (208) follows from Lemma 26 and (212).

Alternative proof of Proposition 17(d). Lemma 26 implies

$$\left\|1 - h_k/h\right\|_{L^2[0,1]}^2 = 2\int\limits_0^1 (1 - h_k/h)(1 - h_{k+1}/h)dx$$

$$= 2\left\|1 - h_{k+1}/h\right\|_{L^2[0,1]}^2 + 2\int\limits_0^1 (h_{k+1}/h - h_k/h)(1 - h_{k+1}/h)dx.$$

Hence, equation (212) yields $n_k = \dfrac{\left\|1 - h_k/h\right\|_{L^2[0,1]}^2}{\left\|1 - h_{k+1}/h\right\|_{L^2[0,1]}^2} \geq 2$. $\square$

9.4. Proof of the existence of the root-Horton limit

Here we present a proof of Thm. 25(ii). The proof is based on Lemmas 27 and 28 stated below that will be proven in the Sects. 9.4.1 and 9.4.2.

Lemma 27. *If the limit* $\lim\limits_{k \to \infty} \dfrac{h_{k+1}(1)}{h_k(1)}$ *exists, then* $\lim\limits_{k \to \infty} (\mathcal{N}_k)^{-\frac{1}{k}} = \lim\limits_{k \to \infty} \left(\prod\limits_{j=1}^k n_j\right)^{\frac{1}{k}}$ *also exists, and*

$$\lim\limits_{k \to \infty} (\mathcal{N}_k)^{-\frac{1}{k}} = \lim\limits_{k \to +\infty} \left(\frac{1}{h_k(1)}\right)^{-\frac{1}{k}} = \lim\limits_{k \to \infty} \frac{h_{k+1}(1)}{h_k(1)}.$$

Lemma 28. *The limit* $\lim\limits_{k \to \infty} \dfrac{h_{k+1}(1)}{h_k(1)} \geq 1$ *exists, and is finite.*

Once Lemmas 27 and 28 are established, the validity of root-Horton law Theorem 25(ii) is proved as follows.

Proof of Theorem 25(ii). The existence and finiteness of $\lim\limits_{k \to \infty} \dfrac{h_{k+1}(1)}{h_k(1)}$ established in Lemma 28 is the precondition for Lemma 27 that in turn implies the existence and finiteness of the limit

$$\lim\limits_{k \to \infty} (\mathcal{N}_k)^{-\frac{1}{k}} = \lim\limits_{k \to \infty} \left(\prod\limits_{j=1}^k n_j\right)^{\frac{1}{k}} = R$$

as needed for the root-Horton law. Furthermore,

$$R = \lim\limits_{k \to \infty} \frac{h_{k+1}(1)}{h_k(1)}, \tag{213}$$

and $2 \leq R \leq 4$ by Proposition 17. $\square$

9.4.1. Proof of Lemma 27 and related results

Proposition 18.

$$\left\|1 - h_{k+1}(x)/h(x)\right\|^2_{L^2[0,1]} \le \frac{1}{h_{k+1}(1)} \le \left\|1 - h_k(x)/h(x)\right\|^2_{L^2[0,1]}. \qquad (214)$$

Proof. Equation (208) implies

$$\frac{h'_{k+1}(x)}{h^2_{k+1}(x)} = 1 - \frac{(h_{k+1}(x) - h_k(x))^2}{h^2_{k+1}(x)} \quad \forall x \in (0,1]. \qquad (215)$$

Integrating both sides of the equation (215) from 0 to 1 we obtain

$$\frac{1}{h_{k+1}(1)} = \int_0^1 \frac{(h_{k+1}(x) - h_k(x))^2}{h^2_{k+1}(x)} dx = \left\|1 - h_k(x)/h_{k+1}(x)\right\|^2_{L^2[0,1]}$$

as $h_{k+1}(0) = 1$.

Hence, using Lemma 26, the first inequality in (214) is proved as follows

$$\frac{1}{h_{k+1}(1)} = \int_0^1 \frac{(h_{k+1}(x) - h_k(x))^2}{h^2_{k+1}(x)} dx \ge \int_0^1 \frac{(h_{k+1}(x) - h_k(x))^2}{h^2(x)} dx$$

$$= \left\|1 - h_{k+1}/h\right\|^2_{L^2[0,1]} = \left\|1 - h_{k+1}(x)/h(x)\right\|^2_{L^2[0,1]}.$$

Finally, equations (212) and (215) imply

$$\frac{1}{h_{k+1}(1)} = \left\|1 - h_k(x)/h_{k+1}(x)\right\|^2_{L^2[0,1]} \le \left\|1 - h_k(x)/h(x)\right\|^2_{L^2[0,1]}.$$

This completes the proof. □

Proof of Lemma 27. If the limit $\lim_{k\to\infty} \frac{h_{k+1}(1)}{h_k(1)}$ exists and is finite, then so is the limit $\lim_{k\to\infty} \left(\frac{1}{h_k(1)}\right)^{-\frac{1}{k}}$. Then, the existence and the finiteness of the limit $\lim_{k\to\infty} (\mathcal{N}_k)^{-\frac{1}{k}}$ follow from equation (210) and Proposition 18. □

9.4.2. Proof of Lemma 28 and related results

In this subsection we use the approach developed by Drmota [47] to prove the existence and the finiteness of $\lim_{k\to\infty} \frac{h_{k+1}(1)}{h_k(1)} \ge 1$. As we saw earlier, this result was used for proving existence, finiteness, and positivity of $\lim_{k\to\infty} (\mathcal{N}_k)^{-\frac{1}{k}} = \lim_{k\to\infty} \left(\prod_{j=1}^k n_j\right)^{\frac{1}{k}}$, the root-Horton law.

Definition 33. *Given $\gamma \in (0,1]$. Let*

$$V_{k,\gamma}(x) = \begin{cases} \frac{1}{1-x} & \text{for } 0 \le x \le 1-\gamma, \\ \gamma^{-1}h_k\left(\frac{x-(1-\gamma)}{\gamma}\right) & \text{for } 1-\gamma \le x \le 1. \end{cases}$$

Note that the sequences of functions $h_k(x)$ and $V_{k,\gamma}(x)$ can be extended beyond $x=1$.

Next, we make some observations about the above defined functions.

Observation 1. $V_{k,\gamma}(x)$ are positive continuous functions satisfying

$$V'_{k+1,\gamma}(x) = 2V_{k+1,\gamma}(x)V_{k,\gamma}(x) - V^2_{k,\gamma}(x)$$

for all $x \in [0,1] \setminus (1-\gamma)$, with initial conditions $V_{k,\gamma}(0) = 1$.

Observation 2. Let $\gamma_k = \frac{h_k(1)}{h_{k+1}(1)}$. Then

$$V_{k,\gamma_k}(1) = h_{k+1}(1) \tag{216}$$

and

$$V_{k,\gamma}(1) = \gamma^{-1}h_k(1) \ge h_{k+1}(1) \quad \text{whenever } \gamma \le \gamma_k. \tag{217}$$

Observation 3.
$$V_{k,\gamma}(x) \le V_{k+1,\gamma}(x)$$

for all $x \in [0,1]$ since $h_k(x) \le h_{k+1}(x)$.

Observation 4. Since $h_1(x) \equiv 1$ and $\gamma_1 = \frac{h_1(1)}{h_2(1)}$,

$$h_2(x) \le V_{1,\gamma_1}(x) = \begin{cases} \frac{1}{1-x} & \text{for } 0 \le x \le 1-\gamma_1, \\ \gamma_1^{-1} = h_2(1) & \text{for } 1-\gamma_1 \le x \le 1. \end{cases}$$

Observation 4 generalizes as follows.

Proposition 19.

$$h_{k+1}(x) \le V_{k,\gamma_k}(x) = \begin{cases} \frac{1}{1-x} & \text{for } 0 \le x \le 1-\gamma_k, \\ \gamma_k^{-1}h_k\left(\frac{x-(1-\gamma_k)}{\gamma_k}\right) & \text{for } 1-\gamma_k \le x \le 1. \end{cases}$$

In order to prove Proposition 19 we will need the following lemma.

Lemma 29. *For any $\gamma \in (0,1)$ and $k \ge 1$, function $V_{k,\gamma}(x) - h_{k+1}(x)$ changes its sign at most once as x increases from $1-\gamma$ to 1. Moreover, since $V_{k,\gamma}(1-\gamma) = h(1-\gamma) > h_{k+1}(1-\gamma)$, function $V_{k,\gamma}(x) - h_{k+1}(x)$ can only change sign from nonnegative to negative.*

Proof. This is a proof by induction with base at $k = 1$. Here $V_{1,\gamma}(x) = \frac{1}{\gamma}$ is constant on $[1-\gamma,1]$, while $h_2(x) = (1+e^{2x})/2$ is an increasing function, and

$$V_{1,\gamma}(1-\gamma) = h(1-\gamma) > h_2(1-\gamma).$$

For the induction step, we need to show that if $V_{k,\gamma}(x) - h_{k+1}(x)$ changes its sign at most once, then so does $V_{k+1,\gamma}(x) - h_{k+2}(x)$. Since both sequences of functions satisfy the same ODE relation (see Observation 1), we have

$$\frac{d}{dx}\left[(V_{k+1,\gamma}(x) - h_{k+2}(x)) \cdot e^{-2 \int\limits_{1-\gamma}^{x} h_{k+1}(y)dy} \right]$$

$$= (2V_{k+1,\gamma}(x) - V_{k,\gamma}(x) - h_{k+1}(x)) \cdot (V_{k,\gamma}(x) - h_{k+1}(x)) \cdot e^{-2 \int\limits_{1-\gamma}^{x} h_{k+1}(y)dy},$$

where $h_{k+1}(x) \leq V_{k+1,\gamma}(x)$ by definition of $V_{k+1,\gamma}(x)$, and $V_{k,\gamma}(x) \leq V_{k+1,\gamma}(x)$ as in Observation 3.

Now, let

$$I(x) := \int\limits_{1-\gamma}^{x} (2V_{k+1,\gamma}(s) - V_{k,\gamma}(s) - h_{k+1}(s)) \cdot (V_{k,\gamma}(s) - h_{k+1}(s)) \cdot e^{-2 \int\limits_{1-\gamma}^{s} h_{k+1}(y)dy} ds.$$

Then

$$(V_{k+1,\gamma}(x) - h_{k+2}(x)) \cdot e^{-2 \int\limits_{1-\gamma}^{x} h_{k+1}(y)dy} = V_{k+1,\gamma}(1-\gamma) - h_{k+2}(1-\gamma) + I(x).$$

The function $2V_{k+1,\gamma}(x) - V_{k,\gamma}(x) - h_{k+1}(x) \geq 0$, and since $V_{k,\gamma}(x) - h_{k+1}(x)$ changes its sign at most once, then $I(x)$ should change its sign from nonnegative to negative at most once as x increases from $1 - \gamma$ to 1. Hence

$$V_{k+1,\gamma}(x) - h_{k+2}(x) = (V_{k+1,\gamma}(1-\gamma) - h_{k+2}(1-\gamma) + I(x)) \cdot e^{2 \int\limits_{1-\gamma}^{x} h_{k+1}(y)dy}$$

should change its sign from nonnegative to negative at most once as

$$V_{k+1,\gamma}(1-\gamma) = h(1-\gamma) > h_{k+2}(1-\gamma)$$

by (212). $\qquad\square$

Proof of Proposition 19. Take $\gamma = \gamma_k$ in Lemma 29. Then function $h_{k+1}(x) - V_{k,\gamma_k}(x)$ should change its sign from nonnegative to negative at most once within the interval $[1-\gamma_k, 1]$. Hence, $V_{k,\gamma_k}(1-\gamma_k) > h_{k+1}(1-\gamma_k)$ and $h_{k+1}(1) = V_{k,\gamma_k}(1)$ imply $h_{k+1}(x) \leq V_{k,\gamma_k}(x)$ as in the statement of the proposition. $\qquad\square$

Now we are ready to prove the monotonicity result.

Lemma 30.
$$\gamma_k \leq \gamma_{k+1} \qquad \text{for all } k \in \mathbb{N}^+.$$

Proof. We prove it by contradiction. Suppose $\gamma_k \geq \gamma_{k+1}$ for some $k \in \mathbb{N}^+$. Then

$$V_{k,\gamma_k}(x) \leq V_{k,\gamma_{k+1}}(x) = \begin{cases} \frac{1}{1-x} & \text{for } 0 \leq x \leq 1 - \gamma_{k+1}, \\ \gamma_{k+1}^{-1} h_k\left(\frac{x-(1-\gamma_{k+1})}{\gamma_{k+1}} \right) & \text{for } 1 - \gamma_{k+1} \leq x \leq 1 \end{cases}$$

and therefore

$$h_{k+1}(x) \leq V_{k,\gamma_k}(x) \leq V_{k,\gamma_{k+1}}(x) \leq V_{k+1,\gamma_{k+1}}(x)$$

as $h_{k+1}(x) \leq V_{k,\gamma_k}(x)$ by Proposition 19.

Recall that for $x \in [1 - \gamma_{k+1}, 1]$,

$$V'_{k+1,\gamma_{k+1}}(x) = 2V_{k,\gamma_{k+1}}(x)V_{k+1,\gamma_{k+1}}(x) - V^2_{k,\gamma_{k+1}},$$

where at $1 - \gamma_{k+1}$ we consider only the right-hand derivative. Thus for $x \in [1 - \gamma_{k+1}, 1]$,

$$\frac{d}{dx}\Big(V_{k+1,\gamma_{k+1}}(x) - h_{k+2}(x)\Big) = A(x) + B(x)\Big(V_{k+1,\gamma_{k+1}}(x) - h_{k+2}(x)\Big),$$

where $A(x) = 2V_{k+1,\gamma_{k+1}}(x) - V_{k,\gamma_{k+1}}(x) - h_{k+1}(x) \geq 0$, $B(x) = 2h_{k+1}(x) > 0$, and $V_{k+1,\gamma_{k+1}}(1 - \gamma_{k+1}) - h_{k+2}(1 - \gamma_{k+1}) = h(1 - \gamma_{k+1}) - h_{k+2}(1 - \gamma_{k+1}) > 0$. Hence

$$V_{k+1,\gamma_{k+1}}(1) - h_{k+2}(1) \geq V_{k+1,\gamma_{k+1}}(1 - \gamma_{k+1}) - h_{k+2}(1 - \gamma_{k+1}) > 0$$

arriving to a contradiction since $V_{k+1,\gamma_{k+1}}(1) = h_{k+2}(1)$. $\qquad\square$

Corollary 20. *Limit* $\lim\limits_{k \to \infty} \gamma_k$ *exists.*

Proof. Lemma 30 implies γ_k is a monotone increasing sequence, bounded by 1. $\qquad\square$

Proof of Lemma 28. Lemma 28 follows immediately from Corollary 20 and an observation that $\frac{h_{k+1}(1)}{h_k(1)} = \frac{1}{\gamma_k}$. $\qquad\square$

10. Generalized dynamical pruning

The Horton pruning (Def. 3), which is the key element of the self-similarity theory developed in previous sections, is a very particular way of erasing a tree. Here we suggest a general approach to erasing a finite tree from leaves down to the root that include both combinatorial and metric prunings, and discuss the respective prune-invariance.

Given a tree $T \in \mathcal{L}$ and a point $x \in T$, let $\Delta_{x,T}$ be the *descendant tree* of x: it is comprised of all points of T descendant to x, including x; see Fig. 37a. Then $\Delta_{x,T}$ is itself a tree in $\mathcal{L}$ with root at x. Let $T_1 = (M_1, d_1)$ and $T_2 = (M_2, d_2)$ be two metric rooted trees (Def. 1), and let ρ_1 denote the root of T_1. A function $f : T_1 \to T_2$ is said to be an *isometry* if $\mathsf{Image}[f] \subseteq \Delta_{f(\rho_1),T_2}$ and for all pairs $x, y \in T_1$,

$$d_2\big(f(x), f(y)\big) = d_1(x, y).$$

The tree isometry is illustrated in Fig. 37b. We use the isometry to define a *partial order* in the space $\mathcal{L}$ as follows. We say that T_1 is *less than or equal to* T_2 and write $T_1 \preceq T_2$ if there is an isometry $f : T_1 \to T_2$. The relation $\preceq$

(a) Descendant tree   (b) Isometry
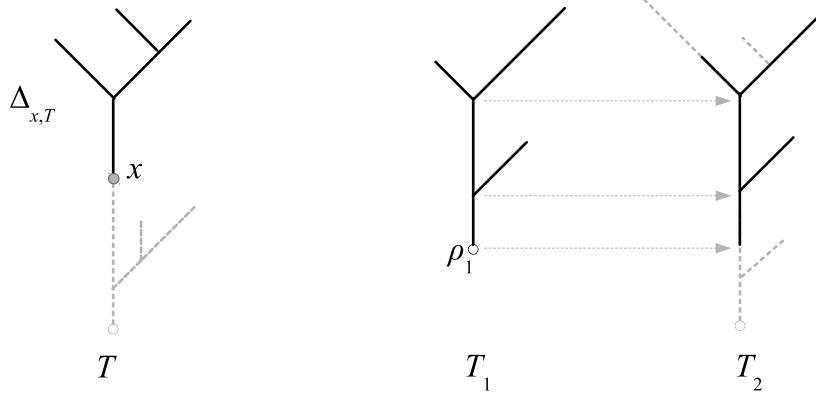
FIG 37. *Descendant subtree and tree isometry: an illustration. (a) Subtree $\Delta_{x,T}$ (solid black lines) descendant to a point x (gray circle) in a tree T (union of dashed gray and soling black lines). (b) Isometry of trees. Tree T_1 (left) is mapped to tree T_2 (right). The image of T_1 within T_2 is shown by black lines, the rest of T_2 is shown by dashed gray lines. Here, tree T_1 is less than tree T_2, $T_1 \preceq T_2$.*

is a partial order as it satisfies the reflexivity, antisymmetry, and transitivity conditions. Moreover, a variety of other properties of this partial order can be observed, including order denseness and semi-continuity.

We say that a function $\varphi : \mathcal{L} \to \mathbb{R}$ is *monotone nondecreasing* with respect to the partial order $\preceq$ if $\varphi(T_1) \leq \varphi(T_2)$ whenever $T_1 \preceq T_2$. Consider a monotone nondecreasing function $\varphi : \mathcal{L} \to \mathbb{R}_+$. We define the *generalized dynamical pruning* operator $\mathcal{S}_t(\varphi, T) : \mathcal{L} \to \mathcal{L}$ induced by φ for any $t \geq 0$ as

$$\mathcal{S}_t(\varphi, T) := \rho \cup \left\{ x \in T \setminus \rho \, : \, \varphi\big(\Delta_{x,T}\big) \geq t \right\}, \tag{218}$$

where ρ denotes the root of tree T. Informally, the operator $\mathcal{S}_t$ cuts all subtrees $\Delta_{x,T}$ for which the value of φ is below threshold t, and always keeps the tree root. Extending the partial order to $\mathcal{L}$ by assuming $\phi \preceq T$ for all $T \in \mathcal{L}$, we observe for any $T \in \mathcal{L}$ that $S_s(T) \preceq S_t(T)$ whenever $s \geq t$.

10.1. Examples of generalized dynamical pruning

The dynamical pruning operator $\mathcal{S}_t$ encompasses and unifies a range of problems, depending on a choice of φ, as we illustrate in the following examples.

10.1.1. Example: pruning via the tree height

Let the function $\varphi(T)$ equal the height of tree T:

$$\varphi(T) = \text{HEIGHT}(T). \tag{219}$$

In this case the operator $\mathcal{S}_t$ satisfies the **continuous semigroup property**:

$$\mathcal{S}_t \circ \mathcal{S}_s = \mathcal{S}_{t+s} \text{ for any } t, s \geq 0.$$

It coincides with the continuous pruning (a.k.a. tree erasure) studied by Jacques Neveu [123], who established invariance of a critical and sub-critical binary Galton-Watson trees with i.i.d. exponential edge lengths with respect to this operation.

It is readily seen that for a coalescent process (Sect. 8.8.1), the dynamical pruning $\mathcal{S}_t$ of the corresponding coalescent tree with $\varphi(T)$ as in (219) replicates the coalescent process. More specifically, the timing and order of particle mergers is reproduced by the dynamics of the leaves of $\mathcal{S}_t(\varphi, T)$. See Sect. 11.2.3, Thm. 29 for a concrete version of this statement for the coalescent dynamics of shocks in the continuum ballistic annihilation model.

10.1.2. Example: pruning via the Horton-Strahler order

Let the function $\varphi(T)$ be one unit less than the Horton-Strahler order $\text{ord}(T)$ of a tree T:

$$\varphi(T) = \text{ord}(T) - 1. \tag{220}$$

This function is also known as the *register number* [59, 66], as it equals the minimum number of memory registers necessary to evaluate an arithmetic expression described by a tree T, assuming that the result is stored in an additional register that also can be used for calculations.

With the choice (220), the dynamical pruning operator coincides with the Horton pruning (Def. 3): $\mathcal{S}_t = \mathcal{R}^{\lfloor t \rfloor}$, if we assume that all edge lengths equal to unity. It is readily seen that $\mathcal{S}_t$ satisfies the **discrete semigroup property**:

$$\mathcal{S}_t \circ \mathcal{S}_s = \mathcal{S}_{t+s} \text{ for any } t, s \in \mathbb{N}.$$

Most of the present survey is focused on invariance of a tree distribution with respect to this operation.

10.1.3. Example: pruning via the total tree length

Let the function $\varphi(T)$ equal the total lengths of T:

$$\varphi(T) = \text{LENGTH}(T). \tag{221}$$

The dynamical pruning by the tree length is illustrated in Fig. 38 for a Y-shaped tree that consists of three edges.
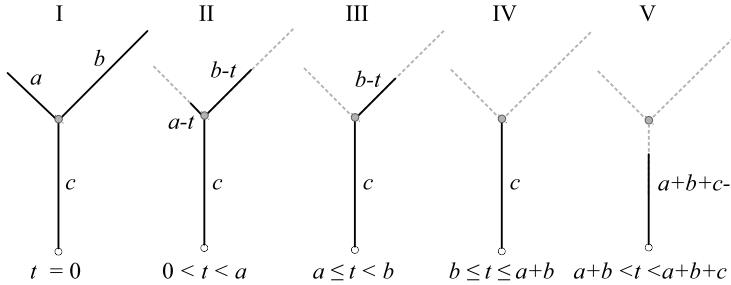
FIG 38. *Pruning by tree length: an illustration. Figure shows five generic stages in the dynamical pruning of a Y-shaped tree T, with pruning function $\varphi(T) = \text{LENGTH}(T)$. The pruned tree $\mathcal{S}_t$ is shown by solid black lines; the pruned parts of the initial tree are shown by dashed gray lines.*
Stage I: *Initial tree T consists of three edges, with lengths a, b, c indicated in the panel; without loss of generality we assume $a < b$.*
Stage II: *For any $t < a$ the pruned tree $\mathcal{S}_t$ has a Y-shaped form with leaf edges truncated by t.*
Stage III: *For any $a \leq t < b$ the pruned tree $\mathcal{S}_t$ consists of a single edge of length $c + b - t$.*
Stage IV: *For any $b \leq t \leq a + b$ the pruned tree $\mathcal{S}_t$ consists of a single edge of length c. Notice that during this stage the tree $\mathcal{S}_t$ does not change with t; this loss of memory causes the process to violate the semigroup property.*
Stage V: *For any $a + b < t < a + b + c$ the pruned tree $\mathcal{S}_t$ consists of a single edge of length $a + b + c - t$.*

Importantly, in this case $\mathcal{S}_t$ does not satisfy the semigroup property. To see this, consider an internal vertex point $x \in T$ (see Fig. 38, where the only internal vertex is marked by a gray ball). Then $\Delta_{x,T}$ consists of point x as its root, the left subtree of length a and the right subtree of length b. Observe that the whole left subtree is pruned away by time a, and the whole right subtree is pruned away by time b. However, since

$$\varphi(\Delta_{x,T}) = \text{LENGTH}(\Delta_{x,T}) = a + b,$$

the junction point x will not be pruned until time instant $a + b$. Thus, x will be a leaf of $\mathcal{S}_t(\varphi, T)$ for all t such that

$$\max\{a, b\} \leq t \leq a + b.$$

This situation corresponds to Stage IV in Fig. 38.

The semigroup property in this example can be introduced by considering **mass-equipped trees**. Informally, we replace each pruned subtree τ of T with a point of mass equal to the total length of τ. The massive points contain some of the information lost during the pruning process, which is enough to establish the semigroup property. Specifically, by time a, the pruned away left subtree (Fig. 38, Stage III) turns into a massive point of mass a attached to x on the left side. Similarly, by time b, the pruned away right subtree (Fig. 38, Stage IV) turns into a massive point of mass b attached to x on the right side. For $\max\{a, b\} \leq t \leq a + b$, this construction keeps track of the quantity

$a + b - t$ associated with point x, and when the quantity $a + b - t$ decreases to 0, the two massive points coalesce into one. If at instant t a single massive point seats at a leaf, its mass $m = t$, and the leaf's parental edge is being pruned. If at instant t two massive points (left and right) seat at a leaf, they total mass $m \geq t$, and further pruning of the leaf's parental edge is prevented until the instant $t = m$, when the two massive points coalesce. Keeping track of all such quantities makes $\mathcal{S}_t$ satisfy the continuous semigroup property. This construction is formally introduced in Sect. 11, which shows that the pruning operator $\mathcal{S}_t$ with (221) coincides with the potential dynamics of continuum mechanics formulation of the 1-D ballistic annihilation model, $A + A \to \emptyset$.

10.1.4. Example: pruning via the number of leaves

Let the function $\varphi(T)$ equal the number of leaves in a tree T. This choice is closely related to the mass-conditioned dynamics of an aggregation process. Specifically, consider N singletons (particles with unit mass) that appear in a system at instants $t_n \geq 0$, $1 \leq n \leq N$. The existing clusters merge into consecutively larger clusters by pair-wise mergers. The cluster mass is additive: a merger of two clusters of masses i and j results in a cluster of mass $i + j$. We consider a time-oriented tree T that describes this process. The tree T has N leaves and $(N-1)$ internal vertices. Each leaf corresponds to an initial particle, each internal vertex corresponds to a merger of two clusters, and the edge lengths represent times between the respective mergers. The action of $\mathcal{S}_t$ on such a tree coincides with a conditional state of the process that only considers clusters of mass $\geq t$. A well-studied special case is a coalescent process with a kernel $K(i, j)$ of Sect. 8.8.1.

10.2. Pruning for $\mathbb{R}$-trees

The generalized dynamical pruning is readily applied to *real trees* (Sect. 2.2), although this is not the focus of our work. We notice that the total tree length (Example 10.1.3) and number of leaves (Example 10.1.4) might be undefined (infinite) for an $\mathbb{R}$-tree. We introduce in Sect. 11.5.3 a *mass* function that can serve as a natural general analog of these and other functions on finite trees. We show (Sect. 11.2.3, Thm. 30) that pruning by mass is equivalent to the pruning by the total tree lengths in a particular situation of ballistic annihilation model with piece-wise continuous potential with a finite number of segments. Accordingly, our results should be straightforwardly extended to $\mathbb{R}$-trees that appear, for instance, as a description of the continuum ballistic annihilation dynamics for other initial potentials.

10.3. Relation to other generalizations of pruning

A pruning operation similar in spirit to the generalized dynamical pruning was considered in a work by Duquesne and Winkel [54] that extended a formalism by

Evans [63] and Evans et al. [64]. We notice that the two definitions of pruning, the generalized dynamical pruning of Sect. 10 and that in [54], are principally different, despite their similar appearance. In essence, the work [54] assumes the Borel measurability with respect to the Gromov-Hausdorff metric ([54], Section 2), which implies the semigroup property of the respective pruning ([54], Lemma 3.11). On the contrary, the generalized dynamical pruning defined here may have the semigroup property only under very particular choices of $\varphi(T)$ as in the examples in Sect. 10.1.1 and 10.1.2. The majority of natural choices of $\varphi(T)$, including the tree length $\varphi(T) = \text{LENGTH}(T)$ (Sect. 10.1.3) or the number of leaves in a tree (Sect. 10.1.4), do not satisfy the semigroup property, and hence are not covered by the pruning of [54]. The main results of our Sect. 11 refer to the pruning function $\varphi(T) = \text{LENGTH}(T)$ that does not satisfy the semigroup property, as shown in Sect. 10.1.3.

Curiously, for the above two examples with no semigroup property, i.e., when $\varphi(T) = \text{LENGTH}(T)$ and when $\varphi(T)$ equals the number of leaves in T, the following discontinuity property holds with respect to the Gromov-Hausdorff metric d_{GH} defined in [63, 64, 54]. For any $\epsilon > 0$ and any $M > 0$, there exist trees T and T' in $\mathcal{L}$ such that

$$|\varphi(T) - \varphi(T')| > M \quad \text{while} \quad d_{\mathsf{GH}}(T, T') < \epsilon.$$

Indeed, if $\varphi(T) = \text{LENGTH}(T)$, we consider a tree T with the number of leaves exceeding M/ϵ, and let T' be the tree obtained from T by elongating each of its leaves by ϵ. Similarly, if $\varphi(T)$ is the number of leaves in T, we construct T' from T by attaching at least M/ϵ new leaves, each of length ϵ.

10.4. Invariance with respect to the generalized dynamical pruning

Consider a tree $T \in \mathcal{L}_{\text{plane}}$ with edge lengths given by a vector $l_T = (l_1, \ldots, l_{\#T})$. The vector l_T can be specified by distribution $\chi(\cdot)$ of a point $x_T = (x_1, \ldots, x_{\#T})$ on the standard simplex

$$\Delta^{\#T} = \left\{ x_i : \sum_i^{\#T} x_i = 1, 0 < x_i \le 1 \right\},$$

and conditional distribution $F(\cdot | x_T)$ of the tree length $\text{LENGTH}(T)$, so that

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

Accordingly, a tree T can be completely specified by its planar shape, a vector of proportional edge lengths, and the total tree length:

$$T = \{\text{P-SHAPE}(T), x_T, \text{LENGTH}(T)\}.$$

A measure η on $\mathcal{L}_{\text{plane}}$ is a joint distribution of these three components:

$$\eta(T \in \{\tau, d\bar{x}, d\ell\}) = \mu(\tau) \cdot \chi_\tau(d\bar{x}) \cdot F_{\tau, \bar{x}}(d\ell),$$

where the tree planar shape is specified by

$$\mu(\tau) = \mathsf{Law}\left(\text{P-SHAPE}(T) = \tau\right), \quad \tau \in \mathcal{T}_{\text{plane}},$$

the relative edge lengths is specified by

$$\chi_\tau(\bar{x}) = \mathsf{Law}\left(x_T = \bar{x} \,|\, \text{P-SHAPE}(T) = \tau\right), \quad \bar{x} \in \Delta^{\#T},$$

and the total tree length is specified by

$$F_{\tau,\bar{x}}(\ell) = \mathsf{Law}\left(\text{LENGTH}(T) = \ell \,|\, x_T = \bar{x}, \ \text{P-SHAPE}(T) = \tau\right), \quad \ell \geq 0.$$

Let us fix $t \geq 0$ and a function $\varphi : \mathcal{L}_{\text{plane}} \to \mathbb{R}_+$ that is monotone nondecreasing with respect to the partial order $\preceq$. We denote by $\mathcal{S}_t^{-1}(\varphi, T)$ the preimage of a tree $T \in \mathcal{L}_{\text{plane}}$ under the generalized dynamical pruning:

$$\mathcal{S}_t^{-1}(\varphi, T) = \{\tau \in \mathcal{L}_{\text{plane}} : \mathcal{S}_t(\varphi, \tau) = T\}.$$

Consider the distribution of edge lengths induced by the pruning:

$$\Xi_\tau(\bar{x}) = \mathsf{Law}\left(x_{\tilde{T}} = \bar{x} \,|\, \text{P-SHAPE}(\tilde{T}) = \tau\right)$$

and

$$\Phi_{\tau,\bar{x}}(\ell) = \mathsf{Law}\left(\text{LENGTH}(\tilde{T}) = \ell \,|\, x_{\tilde{T}} = \bar{x}, \ \text{P-SHAPE}(\tilde{T}) = \tau\right),$$

where the notation $\tilde{T} := \mathcal{S}_t(\varphi, T)$ is used for brevity.

Definition 34 (Generalized prune-invariance). *Consider a function $\varphi : \mathcal{L}_{\text{plane}} \to \mathbb{R}_+$ that is monotone nondecreasing with respect to the partial order $\preceq$. A measure η on $\mathcal{L}_{\text{plane}}$ is called invariant with respect to the generalized dynamical pruning $\mathcal{S}_t(\cdot) = \mathcal{S}_t(\varphi, \cdot)$ (or simply prune-invariant) if the following conditions hold for all $t \geq 0$:*

(i) The measure is prune-invariant in shapes. This means that for the push-forward measure $\nu = (\mathcal{S}_t)_*(\mu) = \mu \circ \mathcal{S}_t^{-1}$ we have

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

(ii) The measure is prune-invariant in edge lengths. This means that for any combinatorial planar tree $\tau \in \mathcal{T}_{\text{plane}}$

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

and there exists a scaling exponent $\zeta \equiv \zeta(\varphi, t) > 0$ such that for any relative edge length vector $\bar{x} \in \Delta^{\#\tau}$ we have

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

Remark 21 (**Pruning trees with no embedding**). The generalized dynamical pruning (218) and the notion of prune-invariance (Def. 34) can be similarly defined on the space $\mathcal{L}$ of metric trees with no planar embedding. In this work we only apply the concept of prune-invariance to planar trees.

Remark 22 (**Relation to Horton prune-invariance**). Definition 34 is similar to Def. 9 of prune-invariance with respect to the Horton pruning, with combinatorial Horton pruning $\mathcal{R}$ being replaced with metric generalized dynamical pruning $\mathcal{S}_t$.

The prune-invariance of Def. 34 unifies multiple invariance properties examined in the literature. For example, the classical work by Jacques Neveu [123] establishes the prune-invariance of the exponential critical binary Galton-Watson trees $\mathsf{GW}(\lambda)$ with respect to the tree erasure from the leaves down to the root at a unit rate, which is equivalent to the generalized dynamical pruning with function $\varphi(T) = \textsc{height}(T)$ (Sect. 10.1.1). The prune-invariance with respect to the Horton pruning (Sect. 10.1.2) has been established by Burd et al. [34] for the combinatorial critical binary Galton-Watson $\mathcal{GW}\left(\frac{1}{2}, \frac{1}{2}\right)$ trees (Thm. 4 in Sect. 5.1.1). Duquesne and Winkel [54] established the prune-invariance of the exponential critical binary Galton-Watson $\mathsf{GW}(\lambda)$ trees with respect to the so-called *hereditary property*, which includes the tree erasure of Sect. 10.1.1 and Horton pruning of Sect. 10.1.2. The *critical Tokunaga trees* analyzed in Sect. 7.5 are prune-invariant with respect to the Horton pruning; this model includes $\mathsf{GW}(\lambda)$ trees as a special case. Section 10.5 below establishes the prune-invariance of the exponential critical binary Galton-Watson $\mathsf{GW}(\lambda)$ trees with respect to the generalized pruning with an arbitrary pruning function $\varphi(T)$.

10.5. *Prune-invariance of* $\mathsf{GW}(\lambda)$

This section establishes prune-invariance of exponential critical binary Galton-Watson trees with respect to arbitrary generalized pruning.

Theorem 26 ([100]). *Let $T \overset{d}{\sim} \mathsf{GW}(\lambda)$, $T \in \mathcal{BL}^{|}_{\mathrm{plane}}$, be an exponential critical binary Galton-Watson tree with parameter $\lambda > 0$. Then, for any monotone nondecreasing function $\varphi : \mathcal{BL}^{|}_{\mathrm{plane}} \to \mathbb{R}_+$ and any $\Delta > 0$ we have*

$$T^{\Delta} := \{\mathcal{S}_{\Delta}(\varphi, T) | \mathcal{S}_{\Delta}(\varphi, T) \neq \phi\} \overset{d}{\sim} \mathsf{GW}(\lambda p_{\Delta}(\lambda, \varphi)),$$

where $p_{\Delta}(\lambda, \varphi) = \mathsf{P}(\mathcal{S}_{\Delta}(\varphi, T) \neq \phi)$. That is, the pruned tree T^{Δ} conditioned on surviving, is an exponential critical binary Galton-Watson tree with parameter

$$\mathcal{E}_{\Delta}(\lambda, \varphi) = \lambda p_{\Delta}(\lambda, \varphi).$$

Proof. Let X denote the length of the stem (edge adjacent to the root) in T, and Y denote the length of the stem in T^{Δ}. Let x be the nearest descendent vertex (a junction or a leaf) to the root in T. Then X, which is an exponential random variable with parameter λ, represents the distance from the root of T
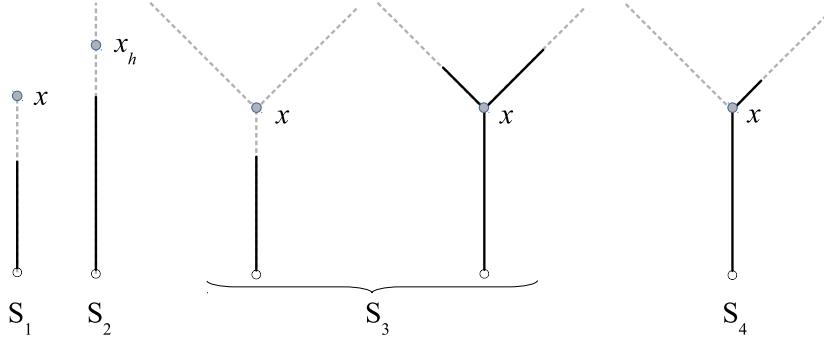
FIG 39. *Sub-events used in the proof of Thm. 26. Gray dashed line shows (a part of) initial tree T. Solid black line shows (a part of) pruned tree T^Δ. We denote by x_h a point in T located at distance h from the root, if it exists.*

to x. Let $\deg_T(x)$ denote the degree of x in tree T and $\deg_{T^\Delta}(x)$ denote the degree of x in tree T^Δ. If $T^\Delta = \phi$, then $Y = 0$. Let

$$F(h) = \mathsf{P}(Y \le h \mid \mathcal{S}_\Delta(\varphi, T) \ne \phi).$$

Let x_h denote a point in T located at distance h from the root, if such exists. If $X \ge h$, the choice of x_h is unique. The event $\{Y \le h\}$ is partitioned into the following non-overlapping sub-events $\mathsf{S}_1, \ldots, \mathsf{S}_4$ illustrated in Fig. 39:

(S$_1$) The event $\mathsf{S}_1 = \big\{\deg_T(x) = 1 \text{ and } X \le h\big\}$ has probability

$$\mathsf{P}(\mathsf{S}_1) = \frac{1}{2}(1 - e^{-\lambda h}).$$

(S$_2$) The event

$$\mathsf{S}_2 = \big\{X > h \text{ and all points of } T \text{ descendant to } x_h \text{ do not belong to } T^\Delta\big\}$$

has probability
$$\mathsf{P}(\mathsf{S}_2) = e^{-\lambda h}(1 - p_\Delta).$$

(S$_3$) The event $\mathsf{S}_3 = \Big\{X \le h \text{ and } \deg_T(x) = 3 \text{ and either both subtrees of } T$ descending from x are pruned away completely (not intersecting T^Δ) or $\big\{x \in T^\Delta,\ \deg_{T^\Delta}(x) = 3\big\}\Big\}$ has probability

$$\mathsf{P}(\mathsf{S}_3) = \frac{1}{2}(1 - e^{-\lambda h})\big((1 - p_\Delta)^2 + p_\Delta^2\big).$$

(S$_4$) The event

$$\mathsf{S}_4 = \{X \le h,\ \deg_T(x) = 3\} \cap \{x \in T^\Delta,\ \deg_{T^\Delta}(x) = 2\} \cap \{Y \le h\}$$

has probability[4]

$$\mathsf{P}(\mathsf{S}_4) = \frac{1}{2}\int\limits_0^h \lambda e^{-\lambda t}\cdot 2p_\Delta(1-p_\Delta)\cdot F(h-t)\, dt = p_\Delta(1-p_\Delta)\int\limits_0^\infty \lambda e^{-\lambda t}F(h-t)\, dt.$$

Using this we have two representations for the probability $\mathsf{P}(Y \le h)$:

$$\mathsf{P}(Y \le h) = (1 - p_\Delta) + p_\Delta F(h)$$

and

$$\begin{aligned}
\mathsf{P}(Y \le h) =& \mathsf{P}(\mathsf{S}_1) + \mathsf{P}(\mathsf{S}_2) + \mathsf{P}(\mathsf{S}_3) + \mathsf{P}(\mathsf{S}_4) \\
=& \frac{1}{2}(1 - e^{-\lambda h}) + e^{-\lambda h}(1 - p_\Delta) \\
& + \frac{1}{2}(1 - e^{-\lambda h})\big((1 - p_\Delta)^2 + p_\Delta^2\big) \\
& + p_\Delta(1 - p_\Delta)\int\limits_0^\infty \lambda e^{-\lambda t}F(h - t)\, dt.
\end{aligned}$$

Equating the two expressions of $\mathsf{P}(Y \le h)$ and simplifying, we obtain

$$(1 - p_\Delta) + p_\Delta F(h) = (1 - p_\Delta + p_\Delta^2) - e^{-\lambda h}p_\Delta + p_\Delta(1 - p_\Delta)\int\limits_0^\infty \lambda e^{-\lambda t}F(h - t)\, dt.$$

Differentiating the above equality we obtain the following equation for the p.d.f. $f(t) = \frac{d}{dt}F(t)$ of Y:

$$f(h) = p_\Delta\, \phi_\lambda(h) + (1 - p_\Delta)\, \phi_\lambda * f(h),$$

where as before ϕ_λ denotes the exponential density with parameter λ as in (72). Applying integral transformation on both sides of the equation, we obtain the characteristic function $\widehat{f}(s) = \mathsf{E}\big[e^{isY}\big]$ of Y,

$$\widehat{f}(s) = \frac{\lambda p_\Delta}{\lambda p_\Delta - is} = \widehat{\phi}_{\lambda p_\Delta}(s).$$

Thus, we conclude that Y is an exponential random variable with parameter λp_Δ.

Next, let y be the descendent vertex (a junction or a leaf) to the root in T^Δ. If $T^\Delta = \phi$, let y denote the root. Let

$$q = \mathsf{P}\big(\deg_{T^\Delta}(y) = 3 \mid S_\Delta(\varphi, T) \ne \phi\big).$$

Then,

$$q\, p_\Delta = \mathsf{P}(\deg_{T^\Delta}(y) = 3) = \mathsf{P}(\deg_T(x) = 3)\left[\mathsf{P}\big(\deg_{T^\Delta}(x) = 3 \mid \deg_T(x) = 3\big)\right.$$

---

[4]Here, $\deg_{T^\Delta}(x) = 2$ means x is neither a junction nor a leaf in T^Δ.

$$+ q\, \mathsf{P}\big(\deg_{T^\Delta}(x) = 2 \mid \deg_T(x) = 3\big)\Big]$$

$$=\frac{1}{2}\Big[p_\Delta^2 + 2p_\Delta(1 - p_\Delta)q\Big]$$

implying

$$q = \frac{1}{2}p_\Delta + (1 - p_\Delta)q,$$

which in turn yields $q = \frac{1}{2}$.

Finally, if $Y \geq h$, let x_h denote the unique point of T located at distance h from the root that survived the pruning. Letting $T_{x_h}^\Delta := \mathcal{S}_\Delta(\varphi, \Delta_{x_h, T})$, we have

$$\big(T_{x_h}^\Delta \mid Y > h\big) \overset{d}{=} \big(T^\Delta \mid S_\Delta(\varphi, T) \neq \phi\big).$$

Hence, conditioned on $S_\Delta(\varphi, T) \neq \phi$, the events $\big\{\deg_{T^\Delta}(y) = 3\big\}$ and $\big\{Y > h\big\}$ are independent.

We saw that conditioning on $\mathcal{S}_\Delta(\varphi, T) \neq \phi$, the pruned tree T^Δ has the stem length distributed exponentially with parameter λp_Δ. Next, independently of the stem length Y, we have the pruned tree T^Δ branching at y (the stem end point farthest from the root), with probability $q = \frac{1}{2}$, into two independent subtrees, each distributed[5] as $\{T^\Delta \mid T^\Delta \neq \phi\}$. Thus, we recursively obtain that T^Δ is a critical binary Galton-Watson tree with i.i.d. exponential edge length with parameter λp_Δ. □

Next, we find an exact form of the survival probability $p_\Delta(\lambda, \varphi)$ for three particular choices of φ, thus obtaining $\mathcal{E}_\Delta(\lambda, \varphi)$.

Theorem 27 ([100]). *In the settings of Theorem 26, we have*

(a) If $\varphi(T)$ equals the total length of T ($\varphi = \text{LENGTH}(T)$), then

$$\mathcal{E}_\Delta(\lambda, \varphi) = \lambda e^{-\lambda\Delta}\Big[I_0(\lambda\Delta) + I_1(\lambda\Delta)\Big].$$

(b) If $\varphi(T)$ equals the height of T ($\varphi = \text{HEIGHT}(T)$), then

$$\mathcal{E}_\Delta(\lambda, \varphi) = \frac{2\lambda}{\lambda\Delta + 2}.$$

(c) If $\varphi(T) + 1$ equals the Horton-Strahler order of the tree T, then

$$\mathcal{E}_\Delta(\lambda, \varphi) = \lambda 2^{-\lfloor\Delta\rfloor},$$

where $\lfloor\Delta\rfloor$ denotes the maximal integer $\leq \Delta$.

Proof. **Part (a).** Suppose $T \overset{d}{\sim} \mathsf{GW}(\lambda)$, and let $\ell(x)$ once again denote the p.d.f. of the total length $\text{LENGTH}(T)$. Then, by Lemma 8,

$$p_\Delta = 1 - \int_0^\Delta \ell(x)\, dx = 1 - \int_0^{\lambda\Delta} \frac{1}{x}e^{-x}I_1\big(x\big)\, dx$$

---

[5]Here, y is also a junction vertex in T of which it is only known that both of its descendent subtrees survived pruning (were not completely erased).

$$= e^{-\lambda\Delta}\Big[I_0(\lambda\Delta) + I_1(\lambda\Delta)\Big], \tag{222}$$

where for the last equality we used formula 11.3.14 in [3].

Part (b). Suppose $T \overset{d}{\sim} \mathsf{GW}(\lambda)$. Let $\mathsf{H}(x)$ once again denote the cumulative distribution function of the height $\textsc{height}(T)$. Then by Lemma 9, for any $\Delta > 0$,

$$p_\Delta = 1 - \mathsf{H}(\Delta) = \frac{2}{\lambda\Delta + 2}.$$

Part (c). Follows from Corollary 12(a). $\qquad\qquad\square$

Remark 23. Let $\mathcal{E}_\Delta(\lambda, \varphi) = \frac{2\lambda}{\lambda\Delta+2}$ as in Theorem 27(b). Here $\mathcal{E}_0(\lambda, \varphi) = \lambda$ and $\mathcal{E}_\Delta(\lambda, \varphi)$ is a linear-fractional transformation associated with matrix

$$\mathcal{A}_\Delta = \begin{pmatrix} 1 & 0 \\ \frac{\Delta}{2} & 1 \end{pmatrix}.$$

Since $\mathcal{A}_\Delta$ form a subgroup in $SL_2(\mathbb{R})$, the transformations $\{\mathcal{E}_\Delta\}_{\Delta \geq 0}$ satisfy the semigroup property

$$\mathcal{E}_{\Delta_1}\mathcal{E}_{\Delta_2} = \mathcal{E}_{\Delta_1 + \Delta_2}$$

for any pair $\Delta_1, \Delta_2 \geq 0$.

We notice also that the operator $\mathcal{E}_\Delta(\lambda, \varphi)$ in part (c) of Theorem 27 satisfies only the discrete semigroup property for nonnegative integer times. Finally, one can check that $\mathcal{E}_\Delta(\lambda, \varphi)$ in part (a) does not satisfy the semigroup property.

11. Continuum 1-D ballistic annihilation

As an illuminating application of the generalized dynamical pruning (Sect. 10) and its invariance properties (Sect. 10.4), we consider the dynamics of particles governed by 1-D ballistic annihilation model, traditionally denoted $A + A \to \emptyset$ [57]. This model describes the dynamics of particles on a real line: a particle with Lagrangian coordinate x moves with a constant velocity $v(x)$ until it collides with another particle, at which moment both particles annihilate, hence the model notation. The annihilation dynamics appears in chemical kinetics and bimolecular reactions and has received attention in physics and probability literature [57, 21, 20, 134, 49, 22, 58, 33, 102, 148].

In a continuum version of the ballistic annihilation model introduced in [100], the moving *shock waves* represent the *sinks* that aggregate the annihilated particles and hence accumulate the mass of the media. Dynamics of these sinks resembles a coalescent process that generates a tree structure for their trajectories, which explains the term *shock wave tree* that we use below. The dynamics of a ballistic annihilation model with two coalescing sinks is illustrated in Fig. 40.

Sect. 11.1 introduces the continuum annihilation model and describes the natural emergence of sinks (shocks). The model initial conditions are given by a particle velocity distribution and particle density on $\mathbb{R}$. Subsequently, we only consider a constant density and initial velocity distribution with alternating
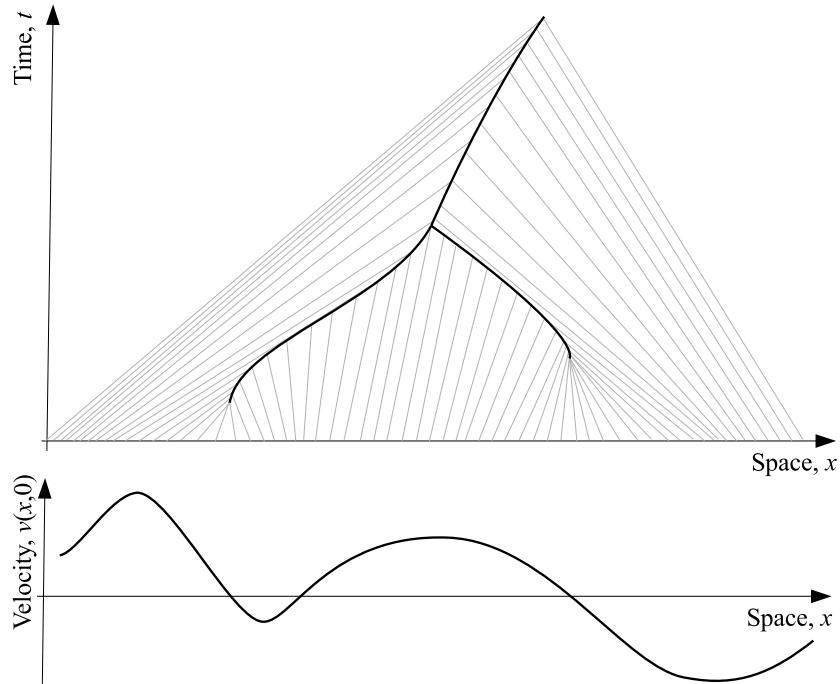
FIG 40. *Ballistic annihilation model: an illustration. A particle with Lagrangian coordinate x moves with velocity $v(x, 0)$ until it collides with another particle and annihilates. (Bottom panel): Initial velocity $v(x, 0)$. (Top panel): The space-time portrait of the system. The trajectories of selected particles are depicted by gray thin lines. The shock wave that describes the motion and coalescence of sinks is shown by solid black line. The sink trajectory forms an inverted Y-shaped tree.*

values ± 1, or, equivalently, initial piece-wise linear potential $\psi(x, 0)$ with alternating slopes ± 1 (Fig. 41). Section 11.2 discusses a construction of the graphical embedding of the shock wave tree into the phase space $(x, \psi(x, t))$ and space-time domain (x, t). Theorems 29, 30 in Sect. 11.2.3 establish equivalence of the ballistic annihilation dynamics to the generalized dynamical pruning of a (mass-equipped) shock wave tree. Sections 11.3, 11.4 illustrate how the pruning interpretation of annihilation dynamics facilitates analytical treatment of the model. Specifically, we give a complete description of the time-advanced potential function $\psi(x, t)$ at any instant $t > 0$ for the initial potential in a form of exponential excursion (Thm. 31), and describe the temporal dynamics of a random sink (Thms. 32, 33). A real tree representation of ballistic annihilation is discussed in Sect. 11.5.

11.1. Continuum model, sinks, and shock trees

Consider a Lebesgue measurable initial density $g(x) \geq 0$ of particles on an interval $[a, b] \subset \mathbb{R}$. The initial particle velocities are given by $v(x, 0) = v(x)$.

Prior to collision and subsequent annihilation, a particle located at x_0 at time $t = 0$ moves according to its initial velocity, so its coordinate $x(t)$ changes as

$$x(t) = x_0 + tv(x_0). \tag{223}$$

When the particle collides with another particle, it annihilates. Accordingly, two particles with initial coordinates and velocities (x_-, v_-) and (x_+, v_+) collide and annihilate at time t when they meet at the same new position,

$$x_- + tv_- = x_+ + tv_+,$$

given that neither of the particles annihilated prior to t. In this case, the annihilation time is given by

$$t = -\frac{x_+ - x_-}{v_+ - v_-}. \tag{224}$$

Let $g(x, t)$ and $v(x, t)$ be the Eulerian specification of the particle density and velocity field, respectively, at coordinate x and time instant t, with a convention that $g(x, t) = 0 \Rightarrow v(x, t) = 0$. We define the corresponding *potential function*

$$\psi(x, t) = -\int_a^x v(y, t)dy, \quad x \in [a, b], t \geq 0,$$

so that $v(x, t) = -\partial_x \psi(x, t)$. Let $\psi(x, 0) = \Psi_0(x)$ be the initial potential.

We call a point $\sigma(t)$ *sink* (or *shock*), if there exist two particles that annihilate at coordinate $\sigma(t)$ at time t. Suppose $v(x) \in C^1(\mathbb{R})$. The equation (224) implies that appearance of a sink is associated with a negative local minima of $v'(x^*)$; we call such points *sink sources*. Specifically, if x^* is a sink source, then a sink will appear at *breaking time* $t^* = -1/v'(x^*)$ at the location given by

$$\sigma(t^*) = x^* + t^*v(x^*) = x^* - \frac{v(x^*)}{v'(x^*)},$$

provided there exists a punctured neighborhood

$$N_\delta(x^*) = \{x : \ 0 < |x - x^*| < \delta\} \subseteq [a, b]$$

such that none of the particles with the initial coordinates in $N_\delta(x^*)$ is annihilated before time t^*.

Sinks, which originate at sink sources, can move and coalesce (see Fig. 40). We refer to a sink trajectory as a *shock wave*. We impose the conservation of mass condition by defining the *mass of a sink* at time t to be the total mass of particles annihilated in the sink between time zero and time t. When sinks coalesce, their masses add up. It will be convenient to assume that sinks do not disappear when they stop accumulating mass. Informally, we assume that the sinks are being pushed by the system particles. Formally, there exists three cases depending on the occupancy of a neighborhood of $\sigma(t)$. If there exists an empty neighborhood around the sink coordinate $\sigma(t)$, the sink is considered at
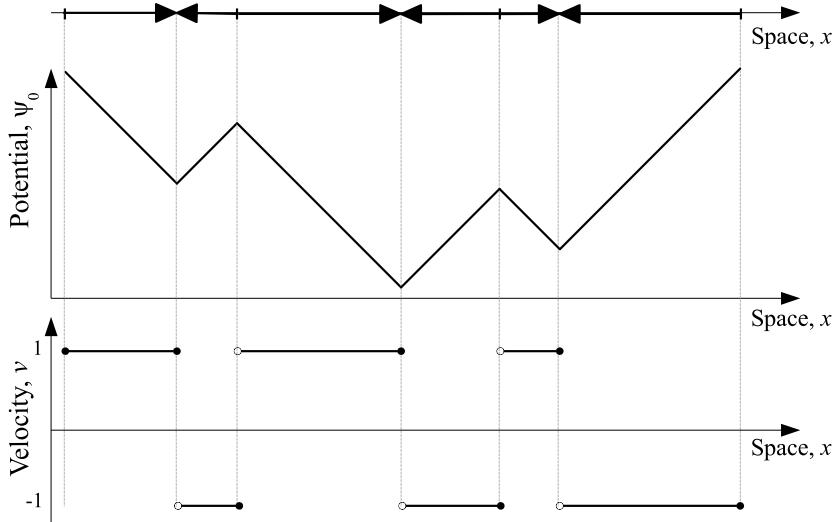
FIG 41. *Piece-wise linear unit slope potential: an illustration. (Top): Arrows indicate alternating directions of particle movement on an interval in* $\mathbb{R}$*. (Middle): Potential* $\Psi_0(x)$ *is a piece-wise linear unit slope function. (Bottom): Particle velocity alternates between values* ± 1 *within consecutive intervals.*

rest – its coordinate does not change. If only the left neighborhood of $\sigma(t)$ is empty, and the right adjacent velocity is negative:

$$v(\sigma_+, t) := \lim_{x \downarrow \sigma(t)} v(x, t) < 0,$$

the sink at $\sigma(t)$ moves with velocity $v(\sigma_+, t)$. A similar rule is applied to the case of right empty neighborhood. The appearance, motion, and subsequent coalescence of sinks can be described by a time oriented *shock tree*. In particular, the coalescence of sinks under initial conditions with a finite number of sink sources is described by a finite tree.

The dynamics of ballistic annihilation, either in discrete or continuum versions, can be quite intricate and is lacking a general description. The existing analyses focus on the evolution of selected statistics under particular initial conditions. In the following sections, we give a complete description of the dynamics in case of two-valued initial velocity and constant particle density.

11.2. Piece-wise linear potential with unit slopes

The discrete 1-D ballistic annihilation model with two possible velocities $\pm v$ was considered in [57, 20, 22, 58, 33]; the three velocity case (-1, 0, and $+1$) appeared in [49, 148]. Here, we explore a continuum version of the 1-D ballistic annihilation with two possible initial velocities and constant initial density, i.e.,

$v(x) = \pm v$ and $g(x,0) \equiv g(x) \equiv g_0$ for $x \in [a,b]$. Since we can scale both space and time, without loss of generality we let $v(x) = \pm 1$ and $g(x) \equiv 1$.

Recall (Sect. 8.3) the space $\mathcal{E}^{\text{ex}}$ of positive piece-wise linear continuous excursions with alternating slopes ± 1 and finite number of segments. We write $\mathcal{E}^{\text{ex}}([a,b])$ for the restriction of this space on the real interval $[a,b]$. We consider an initial potential $\psi(x,0) = \Psi_0(x)$ such that $-\psi(x,0) \in \mathcal{E}^{\text{ex}}([a,b])$; see Fig. 41. This space bears a lot of symmetries that facilitate our analysis.

The dynamics of a system with a simple unit slope potential is illustrated in Fig. 42. Prior to collision, the particles move at unit speed either to the left or to the right, so their trajectories in the (x,t) space are given by lines with slope ± 1 (Fig. 42, top panel, gray lines). The local minima of the potential $\Psi_0(x)$ correspond to the points whose right neighborhood moves to the left and left neighborhood moves to the right with unit speed, hence immediately creating a sink. Accordingly, the sinks appear at $t = 0$ at the local minima of the potential; and those are the only sinks of the system. The sinks move and merge to create a shock wave tree, shown in blue in Fig. 42.

Observe that the domain $[a,b]$ is partitioned into non-overlapping subintervals with boundaries x_j such that the initial particle velocity assumes alternating values of ± 1 within each interval, with boundary values $v(a,0) = v(a) = 1$ and $v(b,0) = v(b) = -1$. Because of the choice of potential $\Psi_0(x)$, we have

$$\int_a^b v(x)\,dx = \Psi_0(b) - \Psi_0(a) = 0,$$

i.e., the total length of the subintervals with the initial velocity -1 equals the total length of the subintervals with the initial velocity 1. For a finite interval $[a,b]$, there exists a finite time $t_{\max} = (b-a)/2$ at which all particles aggregate into a single sink of mass $m = (b-a) = 2\,t_{\max}$ [100]. We only consider the solution on the time interval $[0, t_{\max}]$, and assume that the density of particles vanishes outside of $[a,b]$.

11.2.1. *Graphical representation of the shock wave tree*

For our fixed choice of the initial particle density $g(x) \equiv 1$, the model dynamics is completely determined by the potential $\Psi_0(x)$. We will be particularly interested in the dynamics of sinks (shocks), which we refer to as *shock waves*. The trajectories of sinks can be described by a set (Fig. 42, top panel)

$$\mathcal{G}^{(x,t)}(\Psi_0) = \left\{ (x,t) \in \mathbb{R}^2 \,:\, \exists \text{ a sink satisfying } \sigma(t) = x \right\}$$

in the system space-time domain $(x,t) : x \in [a,b],\ t \in \left[0, (b-a)/2\right]$. These trajectories have a finite binary tree structure: the combinatorial planar shape of $\mathcal{G}^{(x,t)}(\Psi_0)$ is a finite tree in $\mathcal{BT}^|_{\text{plane}}$ [100]. For any two points $(x_i, t_i) \in \mathcal{G}^{(x,t)}(\Psi_0)$, $i = 1, 2$, connected by a unique self-avoiding path γ within $\mathcal{G}^{(x,t)}(\Psi_0)$,

we define the distance between them as

$$d^{(x,t)}\big((x_1, t_1), (x_2, t_2)\big) = \int_\gamma |dt| = 2t^* - t_1 - t_2,$$

where

$$t^* := \max\{t : (x, t) \in \gamma\}.$$

Equivalently, the distance between the points within a single edge is defined as their nonnegative time increment; this induces the distance $d^{(x,t)}$ on $\mathcal{G}^{(x,t)}(\Psi_0)$.

Similarly, the trajectories of the sinks can be described by a set (Fig. 42, bottom panel)

$$\mathcal{G}^{(x,\psi)}(\Psi_0) = \left\{ \big(x, \psi(x, t)\big) \in \mathbb{R}^2 \; : \; \exists \text{ a sink satisfying } \sigma(t) = x \right\}$$

in the system phase space $(x, \psi(x, t)) : \; x \in [a, b], \; t \in \big[0, (b - a)/2\big]$. For any two points $(x_i, \psi_i) \in \mathcal{G}^{(x,\psi)}(\Psi_0)$, $i = 1, 2$, connected by a unique self-avoiding path γ within $\mathcal{G}^{(x,\psi)}(\Psi_0)$, we define the distance between them as

$$d^{(x,\psi)}\big((x_1, \psi_1), (x_2, \psi_2)\big) = \int_\gamma \big(|dt| + |dx|\big).$$

Equivalently, one can consider the L^1 distance between the points within a single edge; this induces the distance $d^{(x,\psi)}$ on $\mathcal{G}^{(x,\psi)}(\Psi_0)$.

Lemma 31 ([100])**.** *The metric spaces $\big(\mathcal{G}^{(x,t)}(\Psi_0), d^{(x,t)}\big)$ and $\big(\mathcal{G}^{(x,\psi)}(\Psi_0), d^{(x,\psi)}\big)$ are trees (Def. 1). Furthermore, they have a finite number of edges and are isomeric to a unique binary tree from $\mathcal{BL}^|_{\mathrm{plane}}$ that we denote by $S(\Psi_0)$.*

We refer to the trees of Lem. 31 as the graphical trees $\mathcal{G}^{(x,t)}(\Psi_0)$ and $\mathcal{G}^{(x,\psi)}(\Psi_0)$ since they are two alternative graphical representations of the *shock wave tree* $S(\Psi_0)$.

11.2.2. Structure of the shock wave tree

Importantly, for our particular choice of the initial potential, the combinatorial structure and the planar embedding of the shock wave tree coincide with that of the level set tree $T = \mathrm{LEVEL}\big(-\Psi_0\big)$ of the initial potential, as we state in the following theorem.

Theorem 28 (Shock wave tree is a level set tree, [100]). *Suppose $g(x) \equiv 1$ and the initial potential $\Psi_0(x)$ is such that $-\Psi_0(x) \in \mathcal{E}^{\mathrm{ex}}$. Then*

$$\mathrm{P\text{-}SHAPE}\big(\mathrm{LEVEL}\,(-\Psi_0)\big) = \mathrm{P\text{-}SHAPE}\big(S(\Psi_0)\big).$$

Theorem 28 implies that there is one-to-one correspondence between internal local maxima of $\Psi_0(x)$ and internal non-root vertices of $S(\Psi_0)$. There is also a one-to-one correspondence between local minima and the leaves. We label the tree vertices with the indices j that correspond to the enumeration of the local extrema x_j of $\Psi_0(x)$; see Fig. 43. We write $\mathsf{parent}(i)$ for the index of the parent vertex to vertex i; $\mathsf{right}(i)$ and $\mathsf{left}(i)$ for the indices of the right and the left offsprings of an internal vertex i; and $\mathsf{sibling}(i)$ for the index of the sibling of vertex i.

For a local extremum x_j, we define its *basin* $\mathcal{B}_j$ as the shortest interval that contains x_j and supports a non-positive excursion of $\Psi_0(x)$. Formally, $\mathcal{B}_j = [x_j^{\mathrm{left}}, x_j^{\mathrm{right}}]$, where

$$x_j^{\mathrm{right}} = \inf \left\{ x : \ x > x_j \text{ and } \Psi_0(x) > \Psi(x_j) \right\},$$

$$x_j^{\mathrm{left}} = \sup \left\{ x : \ x < x_j \text{ and } \Psi_0(x) > \Psi(x_j) \right\}.$$

We observe that the basin $\mathcal{B}_j$ for a local minimum x_j coincides with its coordinate: $\mathcal{B}_j = \{x_j = x_j^{\mathrm{left}} = x_j^{\mathrm{right}}\}$.

The basin's length is $\left|\mathcal{B}_j\right| = x_j^{\mathrm{right}} - x_j^{\mathrm{left}}$. Point $c_j = (x_j^{\mathrm{right}} + x_j^{\mathrm{left}})/2$ denotes the center of the basin $\mathcal{B}_j$. Additionally, we let

$$\mathrm{v}_j = \Psi_0(x_{\mathsf{parent}(j)}) - \Psi_0(x_j) \quad \text{and} \quad \mathrm{h}_j = \left|\mathcal{B}_{\mathsf{sibling}(j)}\right|/2.$$

We are now ready to describe the metric structure of the shock tree $S(\Psi_0)$ and a constructive embedding $\mathcal{G}^{(x,\psi)}(\Psi_0)$ of the tree $S(\Psi_0)$ into the system's phase space.

Metric tree structure. The length l_j of the parental edge of a non-root vertex j within $S(\Psi_0)$ is given by $l_j = \mathrm{v}_j + \mathrm{h}_j$.

Graphical shock tree in the phase space. The tree $\mathcal{G}^{(x,\psi)}(\Psi_0)$ is the union of the following vertical and horizontal segments:

(v) For every local extremum x_j of $\Psi_0(x)$ there exists a vertical segment from $(c_j, \Psi_0(x_j))$ to $(c_j, \Psi_0(x_j) + \mathrm{v}_j)$.
(h) For every local maximum x_j of $\Psi_0(x)$ there exists a horizontal segment of length $\mathrm{h}_{\mathsf{left}(j)} + \mathrm{h}_{\mathsf{right}(j)}$ from $(c_{\mathsf{left}(j)}, \Psi_0(x_j))$ to $(c_{\mathsf{right}(j)}, \Psi_0(x_j))$.

Figure 43 shows the graphical shock trees $\mathcal{G}^{(x,\psi)}$ and $\mathcal{G}^{(x,t)}$ for an initial potential with two local maxima and three local minima, and illustrates the labeling of vertical (v_j) and horizontal (h_j) segments of the tree. Figure 44 shows an example of the graphical tree $\mathcal{G}^{(x,\psi)}$ for an initial potential with nine local minima (and, hence, with nine initial sinks).

Consider a tree $\mathcal{V}(\Psi_0) \in \mathcal{BL}_{\mathrm{plane}}^{|}$ that has the same planar combinatorial structure as $S(\Psi_0)$, and the length of the parental edge of vertex j is given by $l_j = \mathrm{v}_j$. Informally, this is a tree that consists of the vertical segments of the graphical tree $\mathcal{G}^{(x,\psi)}(\Psi_0)$ (Fig. 42, bottom). We have the following corollary of Thm. 28.
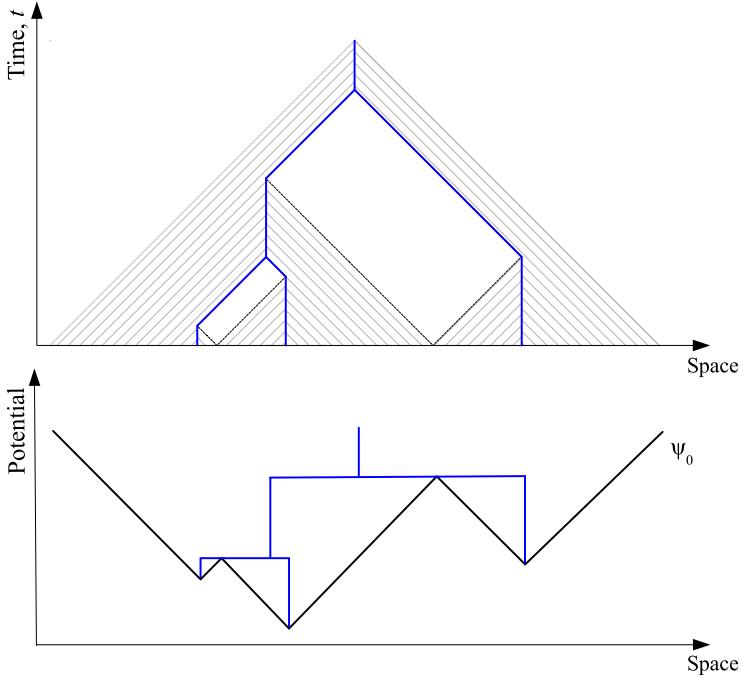
FIG 42. *Shock wave tree (sink tree) in a model with a unit slope potential: an illustration. (Top panel): Space-time dynamics of the system. Trajectories of particles are illustrated by gray lines. The trajectory of coalescing sinks is shown by blue line – this is the graphical representation $\mathcal{G}^{(x,t)}(\Psi_0)$ of the shock wave tree $S(\Psi_0)$. Notice the appearance of empty regions (zero particle density) in the space-time domain. (Bottom panel): Initial unit slope potential $\Psi_0(x)$ with three local minima (black line) and a graphical representation $\mathcal{G}^{(x,\psi)}(\Psi_0)$ of the shock wave tree (blue line) in the phase space $(x, \psi(x,t))$.*

Corollary 21 ([**100**])**.** *Suppose $g(x) \equiv 1$ and potential $\Psi_0(x)$ is such that $-\Psi_0(x) \in \mathcal{E}^{\mathsf{ex}}$. Then*

$$\mathcal{V}(\Psi_0) = \mathrm{LEVEL}\left(-\Psi_0\right).$$

11.2.3. Ballistic annihilation as generalized pruning

This section shows that the dynamics of continuum ballistic annihilation with constant initial density and unit-slope potential is equivalent to the generalized dynamical pruning of either the shock wave tree (Thm. 29) or the level set tree of the potential (Thm. 30).

Suppose a tree $T \in \mathcal{BL}^{|}_{\mathrm{plane}}$ has a particular graphical representation $\mathcal{G}_T \in \mathbb{R}^2$ implemented by a bijective isometry $f : T \to \mathcal{G}_T$ that maps the root of T into the root of $\mathcal{G}_T$. We extend the notion of the generalized dynamical pruning $\mathcal{S}_t(\varphi, \mathcal{G}_T)$ for the graphical tree $\mathcal{G}_T$ by considering the f-image of $\mathcal{S}_t(\varphi, T)$:

$$\mathcal{S}_t(\varphi, \mathcal{G}_T) = f\big(\mathcal{S}_t(\varphi, T)\big).$$
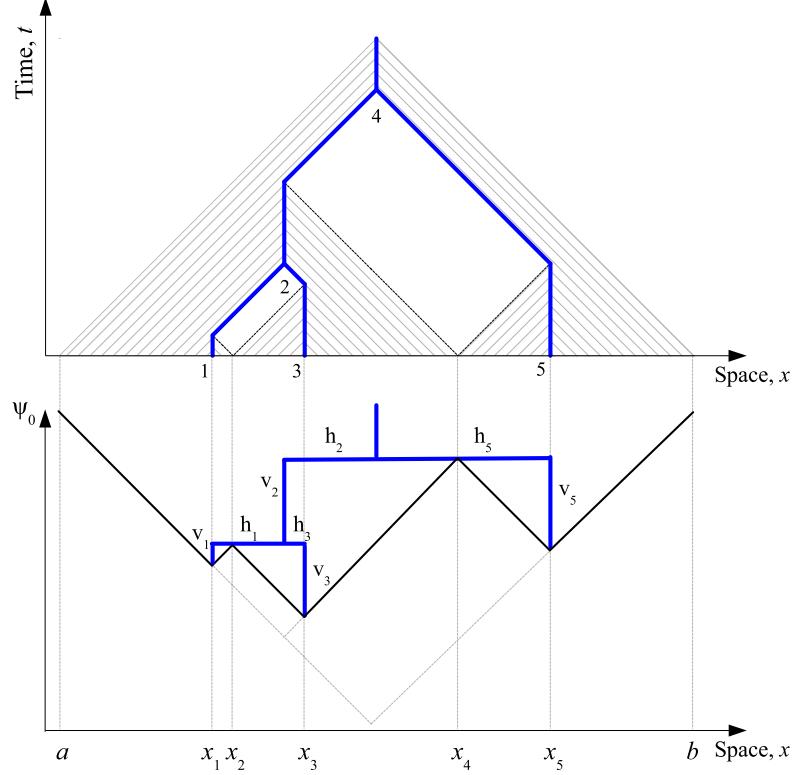
FIG 43. *Shock tree for a piece-wise linear potential with two local maxima. (Top): The shock tree in space-time domain (blue). Hatching illustrates motion of regular particles. There exist two empty rectangular areas, each corresponding to one of the local maxima. The panel illustrates indexing of the tree vertices. (Bottom): Potential $\Psi_0(x)$ (black) and the shock tree in the phase space (blue). The panel illustrates the labeling of vertical (v_j) and horizontal (h_j) segments of the tree.*

Consider a natural isometry (Lem. 31) between the shock wave tree $S(\Psi_0)$ and either of the graphical shock trees, $\mathcal{G}^{(x,t)}(\Psi_0)$ (in the space-time domain) or $\mathcal{G}^{(x,\psi)}(\Psi_0)$ (in the phase space). The next theorem formalizes an observation that the dynamics of sinks is described by the continuous pruning (Sect. 10.1.1) of the shock wave tree.

Theorem 29 (Annihilation pruning I, [100]). *Suppose $g(x) \equiv 1$, and the initial potential $\Psi_0(x)$ is such that $-\Psi_0(x) \in \mathcal{E}^{\text{ex}}$. Then, the dynamics of sinks is described by the generalized dynamical pruning $\mathcal{S}_t(\varphi, \mathcal{G})$ of either the graphical tree $\mathcal{G} = \mathcal{G}^{(x,\psi)}(\Psi_0)$ (in the phase space) or $\mathcal{G} = \mathcal{G}^{(x,t)}(\Psi_0)$ (in the space-time domain), with the pruning function $\varphi(T) = \text{HEIGHT}(T)$. Specifically, the locations of sinks at any instant $t \in [0, t_{\max})$ coincide with the location of the leaves of the pruned tree $\mathcal{S}_t(\varphi, \mathcal{G})$.*
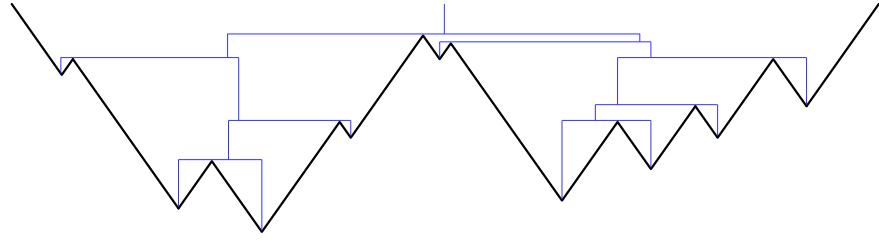
FIG 44. *Graphical representation* $\mathcal{G}^{(x,\psi)}(\Psi_0)$ *(blue) of the sink tree* $S(\Psi_0)$ *for initial potential* $\Psi_0(x)$ *with nine local minima (black). There are nine sinks that correspond to the leaves of the tree. The trajectory of each sink can be traced by going from the corresponding leaf to the root of the tree.*

Theorem 29 only refers to the dynamics of the sinks; it is, however, intuitively clear that the entire potential $\psi(x, t)$ at any given $t > 0$ can be uniquely reconstructed from either of the pruned graphical trees, $\mathcal{G}^{(x,t)}(\Psi_0)$ or $\mathcal{G}^{(x,\psi)}(\Psi_0)$. Because of the multiple symmetries [100], the graphical trees possess significant redundant information. It has been shown in [100] that the reduced tree $\mathcal{V}(\Psi_0)$ (Cor. 21) equipped with information about the sinks provides a minimal description sufficient for reconstructing the entire continuum annihilation dynamics.

Lemma 32 ([100]). *Suppose* $g(x) \equiv 1$, *and the initial potential* $\Psi_0(x)$ *is such that* $-\Psi_0(x) \in \mathcal{E}^{\mathsf{ex}}$. *Then,*

$$\text{LEVEL}(\psi(x, t)) = \mathcal{S}_t(\text{LENGTH}, \mathcal{V}(\Psi_0)).$$

Lemma 32 states that the level set tree (i.e., the sequence of the local extreme values) of $\psi(x, t)$ is uniquely reconstructed from the pruned tree $\mathcal{V}(\Psi_0)$. This, however, is not sufficient to reconstruct the entire time-advanced potential, which has plateaus corresponding to the intervals of zero density (recall the empty regions in the top panels of Fig. 42). The information about such plateaus is lost in the pruned tree. It happens that it suffices to remember "the size" of the pruned out parts of the tree in order to completely reconstruct the annihilation dynamics from $\mathcal{V}(\Psi_0)$. Specifically, we store the value $\varphi(\tau)$ for each subtree τ that has been pruned out. These values are stored in the *cuts* – the points where the pruned subtrees were attached to the initial tree; see Fig. 45(a). The cuts is a union of the leaves of the pruned tree and the vertices of the initial tree that became edge points in the pruned tree. A formal definition is given below.

Definition 35 (**Cuts**). *The set* $\mathcal{D}_t(\varphi, T)$ *of cuts in a pruned tree* $\mathcal{S}_t(\varphi, T)$ *is defined as the boundary of the pruned part of the tree*

$$\mathcal{D}_t(\varphi, T) = \partial\{x \in T \ : \ \varphi(\Delta_{x,T}) < t\}.$$

We now define an extension $\widetilde{\mathcal{S}}_t(\varphi, T)$ of the generalized dynamical pruning that preserves the sizes of pruned subtrees. Such pruning starts with a tree from

(a) Cuts   (b) Massive points
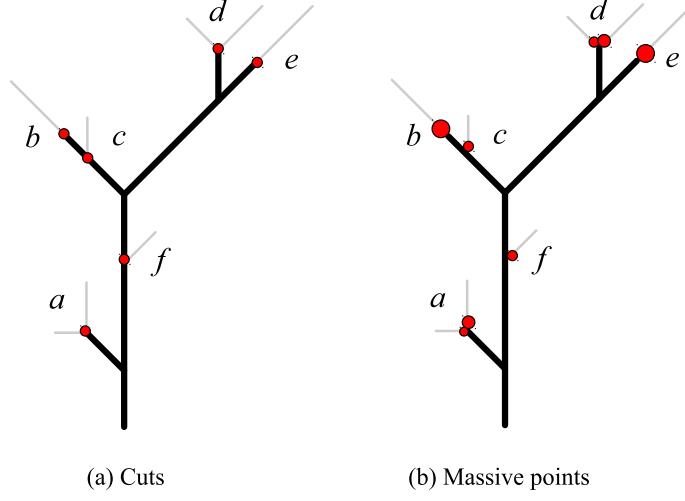
FIG 45. *Cuts and massive points: an illustration. (a) Pruned tree $\mathcal{S}_t(\text{LENGTH}, T)$ (solid black) with the set of cuts (red circles). The pruned parts of the initial tree T are shown in gray. Here, we prune by length; the cuts* a, d *correspond to Stage IV of Fig. 38. The cuts* a *and* d *are placed at vertices of T that became leaves within $\mathcal{S}_t(\text{LENGTH}, T)$. The cuts* b *and* e *are placed at the leaves of the pruned tree. The cuts* c *and* f *are placed at vertices of T that became non-vertex points within $\mathcal{S}_t(\text{LENGTH}, T)$. (b) Massive points (red circles) placed at the cuts. Each of the cuts* a *and* d *hosts two oriented massive points. Each of the cuts* b *and* e *hosts a single unoriented massive point. Each of the cuts* c *and* f *hosts a single oriented massive point. The circle size is proportional to the mass.*

$\mathcal{BL}^{|}_{\text{plane}}$ and results in a tree from the space of mass-equipped trees, denoted $\widetilde{\mathcal{BL}}^{|}_{\text{plane}}$. The pruning $\widetilde{\mathcal{S}}_t(\varphi, T)$ of a tree $T \in \mathcal{BL}^{|}_{\text{plane}}$ is a tree from $\widetilde{\mathcal{BL}}^{|}_{\text{plane}}$, whose projection to $\mathcal{BL}^{|}_{\text{plane}}$ coincides with $\mathcal{S}_t(\varphi, T)$. In addition, the tree is equipped with massive points placed at the cuts. Each massive point corresponds to a pruned out subtree τ of T, with mass equal $\varphi(\tau)$. If a cut is the boundary for two pruned subtrees (Fig. 45(a), cuts a, d), then it hosts two oriented masses. Such cuts are typical in prunings that do not have the semigroup property (see Fig. 38, Stage IV). Figure 45(b) illustrates mass-equipped pruning $\widetilde{\mathcal{S}}_t(\varphi, T)$ with pruning function $\varphi = \text{LENGTH}$.

Next, we describe how to construct a potential $\psi_{T,t}(x)$ for a given $t \in [0, t_{\max}]$ and all $x \in [a, b]$ from a pruned mass-equipped tree $T = \widetilde{\mathcal{S}}_t(\text{LENGTH}, \mathcal{V}(\Psi_0))$. Theorem 30 then shows that this reconstructed potential coincides with the time-advances potential of the annihilation dynamics.

Construction 1 (Tree $\rightarrow$ potential). *Suppose $T = \widetilde{\mathcal{S}}_t(\text{LENGTH}, \mathcal{V}(\Psi_0))$. The corresponding potential $\psi_{T,t}(x)$, with $-\psi_{T,t}(x) \in \mathcal{E}^{\text{ex}}$, is constructed in the following steps:*

(1) *Construct the Harris path $H_T(x)$ for the projection of T to $\mathcal{BL}^{|}_{\text{plane}}$ (i.e., disregarding masses), and consider the negative excursion $-H_T(x)$.*

(2) *At every local minimum of $-H_T(x)$ that corresponds to a double mass $(m_\mathrm{L}, m_\mathrm{R})$, insert a horizontal plateau of length*

$$\varepsilon = 2(m_\mathrm{L} + m_\mathrm{R} - t),$$

as illustrated in Fig. 46, Stage 3.

(3) *At every monotone point of $-H_T(x)$ that corresponds to an internal mass m, insert a horizontal plateau of length $2m$ (Fig. 46, Stage 2).*

(4) *At every internal local maxima of $-H_T(x)$, insert a horizontal plateau of length $2t$ (Fig. 46, Stage 1).*

The following theorem establishes the equivalence of the continuum annihilation dynamics and mass-equipped generalized dynamical pruning with respect to the tree length. In particular, it includes the statement of Lem. 32.

Theorem 30 (**Annihilation pruning II, [100]**). *Suppose $g(x) \equiv 1$ and the initial potential $\Psi_0(x)$ is such that $-\Psi_0(x) \in \mathcal{E}^{\mathsf{ex}}$. Then, for any $t \in [0, t_{\max}]$, the time-advances potential $\psi(x, t)$ is uniquely reconstructed (by Construction 1) from the pruned tree $T = \widetilde{\mathcal{S}}_t(\mathrm{LENGTH}, \mathcal{V}(\Psi_0))$. That is, $\psi(x, t) \equiv \psi_{T,t}$ for all $x \in [a, b]$.*

It is shown in [100] that, inversely, the mass-equipped tree $\widetilde{\mathcal{S}}_t(\mathrm{LENGTH}, \mathcal{V}(\Psi_0))$ can be uniquely reconstructed from the time-advanced potential $\psi(x, t)$. Hence, the continuum ballistic annihilation dynamics is equivalent to the mass-equipped generalized dynamical pruning of the level set tree of the initial potential. The next sections illustrates how this equivalence facilitates the analytical treatment of the model.

11.3. Ballistic annihilation of an exponential excursion

This section examines a special case of piece-wise linear potential with unit slopes: a negative exponential excursion. Consider potential

$$\psi(x, 0) = -H_{\mathsf{GW}(\lambda)}(x)$$

that is the negative Harris path (Sect. 8.1) of an exponential critical binary Galton-Watson tree with parameter λ (Def. 30). In words, the potential is a negative finite excursion with linear segments of alternating slopes ± 1, such that the lengths of all segments except the last one are i.i.d. exponential random variables with parameter $\lambda/2$. Accordingly, the initial particle velocity $v(x, 0)$ alternates between the values ± 1 at epochs of a stationary Poisson point process on $\mathbb{R}$ with rate $\lambda/2$, starting with $+1$ and until the respective potential crosses the zero level.

Corollary 22 (**Exponential excursion**). *Suppose $g(x) \equiv 1$ and initial potential $\Psi_0(x) = -H_{\mathsf{GW}(\lambda)}(x)$. Then the corresponding tree $\mathcal{V}(\Psi_0) \in \mathcal{BL}^{|}_{\mathrm{plane}}$ is an exponential binary critical Galton-Watson tree $\mathsf{GW}(\lambda)$.*
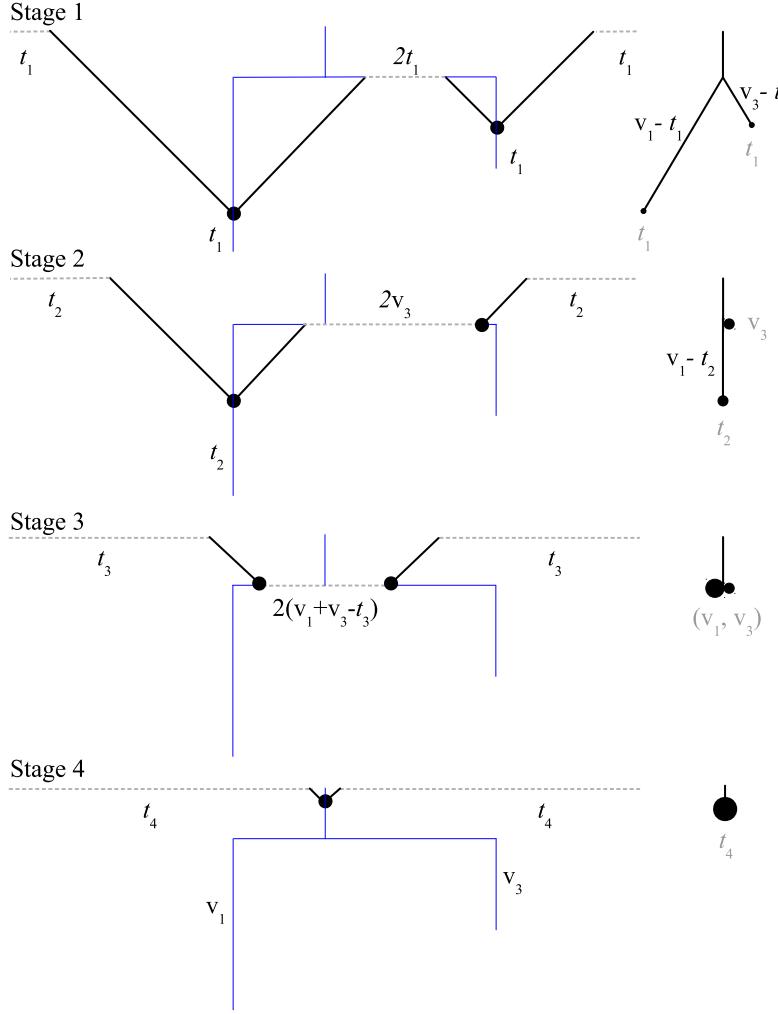
Fig 46. *Four generic stages in the ballistic annihilation dynamics of a W-shaped potential (left), and respective mass-equipped trees (right). The lengths v_1 and v_3 of the two vertical leaf segments are assigned as illustrated in the Stage 4 (see also Fig. 43). (Left): Potential $\psi(x,t)$ is shown in solid black. Each plateau (dashed gray) corresponds to an interval of zero density. The graphical shock tree $\mathcal{G}^{(x,\psi)}(\Psi_0)$ (blue) and sinks (black circles) are shown for visual convenience. (Right): Mass-equipped trees. Segment lengths are marked in black, point masses are indicated in gray. Notice progressive increase of the point masses from Stage 1 to 4. The Stages 1 to 4 refer to time instants $t_1 < t_2 < t_3 < t_4$. Here $v_3 < v_1$, $v_3 > t_1$, $v_3 < t_2 < v_1$, $v_1 < t_3$, and $t_3 < v_1 + v_3 < t_4$.*

Proof. By Cor. 21, the tree $\mathcal{V}(\Psi_0)$ is the level set tree of the negative potential $-\Psi_0(x)$. The statement now follows from Thm. 20. $\square$

To formulate the next result, recall that if $T \stackrel{d}{\sim} \mathsf{GW}(\lambda)$ and $\varphi(T) =$

LENGTH(T), then by (222),

$$p_t := \mathsf{P}(\varphi(T) > t) = e^{-\lambda t}\Big[I_0(\lambda t) + I_1(\lambda t)\Big].$$

Also, the p.d.f. of LENGTH(T) is given by $\ell(x)$ of (73).

Theorem 31 (**Ballistic annihilation dynamics of an exponential excursion, [100]**). *Suppose the initial particle density is constant, $g(x) \equiv 1$, and the initial potential $\psi(x,0)$ is the negative Harris path of an exponential critical binary Galton-Watson tree with parameter λ, i.e., $\mathcal{V}(\Psi_0) \stackrel{d}{\sim} \mathsf{GW}(\lambda)$. Then, at any instant $t > 0$ the mass-equipped shock tree $\mathcal{V}_t = \widetilde{\mathcal{S}}_t(\text{LENGTH}, \mathcal{V}(\Psi_0))$ conditioned on surviving, $\mathcal{V}_t \neq \phi$, is distributed according to the following rules.*

(i) *The planar shape of the tree, as an element of $\mathcal{BL}^{|}_{\text{plane}}$, is distributed as an exponential binary Galton-Watson tree $\mathsf{GW}(\lambda_t)$ with $\lambda_t := \lambda p_t$.*

(ii) *A single or double mass points are placed independently in each leaf with the probability of a single mass being*

$$\frac{2}{\lambda}\frac{\ell(t)}{p_t^2}.$$

(iii) *Each single mass at a leaf has mass $m = t$. For a double mass, the individual masses $(m_\mathrm{L}, m_\mathrm{R})$ have the following joint p.d.f.*

$$f(a,b) = \frac{\ell(a)\ell(b)}{p_t^2 - \frac{2}{\lambda}\ell(t)}$$

for $a, b > 0$, $a \vee b \leq t < a + b$.

(iv) *The number of mass points placed in the interior of any edge is distributed geometrically with the probability of placing k masses being*

$$p_t\big(1 - p_t\big)^k, \qquad k = 0, 1, 2, \ldots.$$

The locations of k mass points are independent uniform in the interior of the edge. The orientation of each mass is left or right independently with probability $1/2$.

(v) *The edge masses are i.i.d. random variables with the following common p.d.f.*

$$\frac{\ell(a)}{1 - p_t}, \qquad a \in (0, t).$$

11.4. Random sink in an infinite exponential potential

Here we focus on the dynamics of a random sink in the case of a negative exponential excursion potential. To avoid subtle conditioning related to a finite potential, we consider here an infinite exponential potential $\Psi_0^{\text{exp}}(x)$, $x \in \mathbb{R}$, constructed as follows. Let x_i, $i \in \mathbb{Z}$ be the epochs of a Poisson point process on
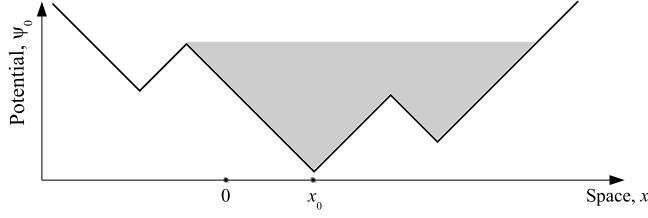
FIG 47. *Random sink $\mathcal{M}_0$ originates at point x_0 – the local minimum closest to the origin. Its dynamics during a finite time interval $[0,t]$ is completely specified by a finite negative excursion $\mathcal{B}_0^t$ similar to the one highlighted in the figure.*

$\mathbb{R}$ with rate $\lambda/2$, indexed so that x_0 is the epoch closest to the origin. The initial velocity $v(x,0)$ is a piece-wise constant continuous function that alternates between values ± 1 within the intervals $(x_i - 1, x_i]$ and with $v(x_0, 0) = 1$. Accordingly, the initial potential $\Psi_0^{\exp}(x)$ is a piece-wise linear continuous function with a local minimum at x_0 and alternating slopes ± 1 of independent exponential duration. The results in this section refer to the sink $\mathcal{M}_0$ with initial Lagrangian coordinate x_0. We refer to $\mathcal{M}_0$ as a *random sink*, using translation invariance of Poisson point process.

Observe that for any fixed $t > 0$, the dynamics of $\mathcal{M}_0$ is completely specified by a finite excursion within $\Psi_0^{\exp}(x)$. For instance, one can consider the shortest negative excursion of $\Psi_0^{\exp}(x)$ within interval $\mathcal{B}_0^t$ such that $x_0 \in \mathcal{B}_0^t$, $|\mathcal{B}_0^t| > 2t$, and one end of $\mathcal{B}_0^t$ is a local maximum of $\Psi_0^{\exp}(x)$ (see Fig. 47). The respective Harris path is an exponential Galton-Watson tree $\mathsf{GW}(\lambda)$. The dynamics of $\mathcal{M}_0$ consists of alternating intervals of mass accumulation (vertical segments of $\mathcal{G}^{(x,\psi)}$) and motion (horizontal segments of $\mathcal{G}^{(x,\psi)}$), starting with a mass accumulation interval. Label the lengths v_i of the vertical segments and the lengths h_i of the horizontal segments in the order of appearance in the examined trajectory. Corollary 22 implies that $\mathsf{v}_i, \mathsf{h}_i$ are independent; the lengths of v_i are i.i.d. exponential random variables with parameter λ; and the lengths of h_i equal the total lengths of independent Galton-Watson trees $\mathsf{GW}(\lambda)$. This description, illustrated in Fig. 48, allows us to find the mass dynamics of a random sink, which is described in the next two theorems.

Theorem 32 (**Growth probability of a random sink, [100]**). *The probability $\xi(t)$ that a random sink $\mathcal{M}_0$ is growing at a given instant $t > 0$ (that is, it is at rest and accumulates mass) is given by*

$$\xi(t) = e^{-\lambda t} I_0(\lambda t). \tag{225}$$

Theorem 33 (**Mass distribution of a random sink, [100]**). *The mass of a random sink $\mathcal{M}_0$ at instant $t > 0$ has probability distribution*

$$\begin{aligned}
\mu_t(da) = {} &\frac{\lambda}{2} e^{-\lambda t} \Big[I_0\big(\lambda(t - a/2)\big) + I_1\big(\lambda(t - a/2)\big) \Big] I_0(\lambda a/2) \cdot \mathbf{1}_{(0,2t)}(a) da \\
&+ e^{-\lambda t} I_0(\lambda t) \delta_{2t}(da),
\end{aligned}$$
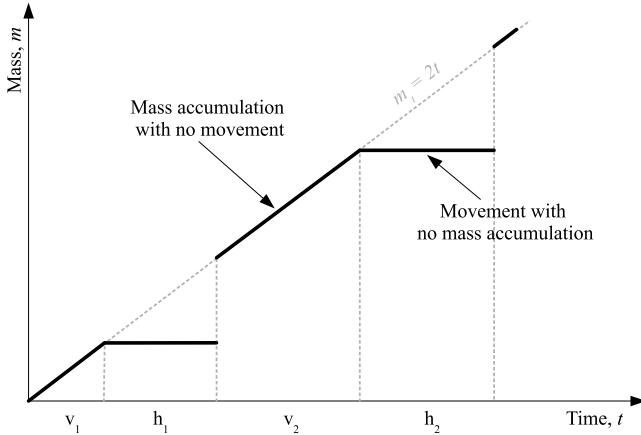
Fɪɢ 48. *Dynamics of a random sink: an illustration. The trajectory of a sink is partitioned into alternating intervals of mass accumulation of duration* v_j *and intervals of movement with no mass accumulation of duration* h_j. *Each* v_j *is an exponential random variable with parameter* λ. *Each* h_j *is distributed as the total length of a critical Galton-Watson tree with exponential edge lengths with parameter* λ.

where δ_{2t} *denotes Dirac delta function (point mass) at* $2t$.

Remark 24. The distribution of mass found in Thm. 33 is a mixture of two components. The first component (the first term on the right) is a distribution absolutely continuous with respect to the Lebesgue measure. It describes continuous mass accumulation during intervals of no movement; see Fig. 48. The second component (the second term on the right) is a delta function. It describes instantaneous mass increases associated with coalescences of the examined random sink with other sinks.

Remark 25. One can notice that the continuum annihilation dynamics of this section, with its shock waves, shock wave trees, and sink masses is reminiscent of that in the 1-D inviscid Burgers equation that describes the evolution of the velocity field $v(x,t)$:

$$\partial_t v(x,t) + v(x,t)\, \partial_x v(x,t) = 0, \quad x \in \mathbb{R}, t \in \mathbb{R}_+. \tag{226}$$

The Burgers dynamics appears in a surprising variety of problems, ranging from cosmology to fluid dynamics and vehicle traffic models; see [19, 68, 77] for comprehensive review. The solution of the Cauchy problem for the Burgers equation develops singularities (shocks) that correspond to intersection of individual particles. The shocks evolve via the shock waves that can be described as massive particles that aggregate the colliding regular particles and hence accumulate the mass of the media. The dynamics of these massive particles generates a tree structure for their world trajectories, the shock wave tree [28, 77].

The case of smooth random initial velocity can be treated explicitly via the Hopf-Cole solution. The case of non-smooth random initial velocities, e.g. a

white noise or a (fractional) Brownian motion, has been extensively studied, both numerically [145] and analytically [149, 27, 28, 70]. In this case, tracing the dynamics of the massive particles backward in time (from a point within a shock tree to the leaves) corresponds to fragmentation of the mass and describes the genealogy of the shocks, i.e., the sets of particles that merge with a given massive particle [26, 70]. In particular, it has been established in [28] that the shock wave tree for a Brownian motion initial velocity becomes the eternal additive coalescent after a proper time change; similar arguments apply for the Lévy type initial velocities [120]. However, despite general heuristic understanding of the structure of the Burgers shock wave tree, a complete analytical description is lacking (e.g., [145]).

11.5. Real tree description of ballistic annihilation

Recall that an $\mathbb{R}$-tree is a generalization of the concept of a finite tree with edge lengths to infinite spaces; see Sect. 2.2 for a formal setup. We construct here (Sect. 11.5.1) an $\mathbb{R}$-tree $\mathbb{T} = \mathbb{T}(\Psi_0)$ that describes the entire model dynamics as coalescence of particles and sinks; this tree is sketched by gray lines in the top panel of Figs. 42 and 49. Specifically, the tree consists of points (x, t) such that there exist either a particle or a sink with coordinate x at time t. There is one-to-one correspondence between the initial particles $(x, 0)$ and leaf vertices of $\mathbb{T}$. Each leaf edge of $\mathbb{T}$ corresponds (one-to-one) to the free (ballistic) run of a corresponding particle before annihilating in a sink. Four of such free runs are depicted by green arrows in Fig. 49. The shock wave tree (movement and coalescence of sinks) corresponds to the non-leaf part of the tree $\mathbb{T}$; it is shown by blue lines in Figs. 42, 49. We adopt a convention that the motion of a particle consists of two parts: an initial ballistic run at unit speed, and subsequent motion within a respective sink. For example, the within-sink motion of particles x and x' is shown by red line in Fig. 49. This interpretation extends motion of all particles to the same time interval $[0, t_{\max}]$, with $t_{\max}$ being the time when the last remaining sink accumulates the total mass on the initial interval. This final sink serves as the tree root. Section 11.5.1 introduces a proper metric on this space so that the model is represented by a time oriented rooted $\mathbb{R}$-tree. In particular, the metric induced by this tree on the initial particles $(x, 0)$ becomes an *ultrametric*, with the distance between any two particles equal to the time until their collision (as particles or as respective sinks).

Section 11.5.2 discusses two non-Lebesgue metrics of the system's domain $[a, b]$. Both describe the ballistic annihilation dynamics and are readily constructed from the initial potential $\Psi_0(x)$. One of these decsriptions is an $\mathbb{R}$-tree and the other is not. The $\mathbb{R}$-tree description establishes an equivalence between the pairs of points that collide with each other, like the pairs (x, x') and (y, y') in Fig. 49. This tree is isometric to the level set tree $\mathrm{LEVEL}(-\Psi_0)$ of the initial potential that is used in this work to describe the shock wave tree (Cor. 21); it is known in the literature as a *tree in continuous path* [135, Def. 7.6], [63, Ex. 3.14]. In Sect. 11.5.3 we briefly discuss a natural way of introducing prun-
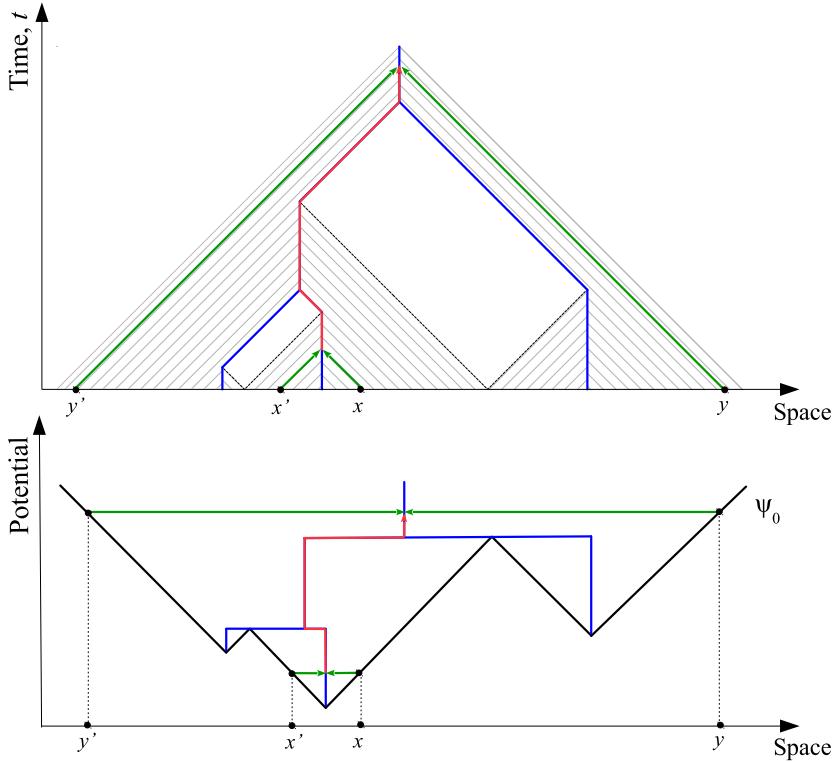
FIG 49. $\mathbb{R}$-tree representation of a ballistic annihilation model with a unit slope potential: an illustration. Figure illustrates dynamics of four points, x, x', y, and y', marked in the horizontal space axis. The pairs of points $\{x, x'\}$ and $\{y, y'\}$ collide and annihilate with each other. Green arrows correspond to ballistic runs of points x, x', y, y', and hence to leaves of tree $\mathbb{T}(\Psi_0)$. Red line corresponds to the trajectory of points x, x' after their collision, within a sink. The rest of notations are the same as in Fig. 42.

ings on $\mathbb{R}$-trees and show that a typical pruning does not have the semigroup property.

11.5.1. $\mathbb{R}$-tree representation of ballistic annihilation

We construct here a real tree representation of the continuum ballistic annihilation model of Sect. 11.2. Specifically, we assume a unit particle density $g(x) \equiv 1$ and initial potential $-\Psi_0(x) \equiv -\psi(x, 0) \in \mathcal{E}^{\mathrm{ex}}$, i.e., $\Psi_0(x)$ is a unit slope negative excursion with a finite number of segments on a finite interval $[a, b]$ (e.g., bottom panel of Fig. 42). Recall that the interval $[a, b]$ completely annihilates by time $t_{\max} = (b - a)/2$, producing a single sink at space-time location $((b + a)/2, t_{\max})$.

Consider the model's entire space-time domain $\mathbb{T} = \mathbb{T}(\Psi_0)$ that consists of all points of the form (x, t), $x \in [a, b]$, $0 \le t \le t_{\max}$, such that there exists either a

particle or a sink at location x at time instant t. The shaded (hatched) regions in the top panels of Figs. 42, 43 are examples of such sets of points. For any pair of points (x, t) and (y, s) in $\mathbb{T}$, we define their unique *earliest common ancestor* as a point

$$\mathsf{A}_\mathbb{T}((x, t), (y, s)) = (z, w) \in \mathbb{T}$$

such that w is the infimum over all w' such that

$$\exists\, z'\, :\, \{(x, t), (y, s)\} \in \Delta_{(z', w'), \mathbb{T}}.$$

The length of the unique segment between the points (x, t) and (y, s) is defined as

$$d\big((x, t), (y, s)\big) = \frac{1}{2}\big((w - t) + (w - s)\big) = \frac{1}{2}(2w - s - t), \qquad (227)$$

where w is the time component of $(z, w) = \mathsf{A}_\mathbb{T}((x, t), (y, s))$.

The tree $(\mathbb{T}, d)$ for a simple initial potential is illustrated in the top panel of Fig. 42 by gray lines. The tree has a relatively simple structure. There is a one-to-one correspondence between the initial particles $(x, 0)$, $x \in [a, b]$, and the leaf vertices of $\mathbb{T}$. There is a one-to-one correspondence between the ballistic runs of the initial particles (runs before collision and annihilation) and the leaf edges of $\mathbb{T}$. Four of such runs are shown by green arrows in Fig. 49. There is one-to-one correspondence between the sink points $(\sigma(t), t)$ and the non-leaf part of $\mathbb{T}$. In particular, the tree root corresponds to the final sink $((a + b)/2, t_{\max})$. The sink points are shown by blue line in Figs. 42, 43. It is now straightforward to check that the tree $(\mathbb{T}, d)$ satisfies the four point condition.

Consider again the sink subspace of $\mathbb{T}$, which consists of the points $\{\sigma(t), t)\}$ such that there exists a sink at location $\sigma(t)$ at time instant t, equipped with the distance (227). This metric subspace is also a tree, as a connected subspace of an $\mathbb{R}$-tree [63]. This tree is isometric to the shock wave tree $S(\Psi_0)$ and hence to either of its graphical representations $\mathcal{G}^{(x,t)}(\Psi_0)$ or $\mathcal{G}^{(x,\psi)}(\Psi_0)$ that are illustrated in Figs. 42, 43 (top and bottom panels, respectively).

From the above construction, it follows that all leaves $(x, 0)$ are located at the same depth (distance from the root) $t_{\max}$. To see this, consider the segment that connect a leaf and the root and apply (227). Moreover, each *time section* at a fixed instant t_0, $\mathsf{sec}(\mathbb{T}, t_0) = \{(x, t_0) \in \mathbb{T}\}$, is located at the same depth $t_{\max} - t_0$. This implies, in particular, that for any fixed $t_0 \geq 0$, the metric induced by $\mathbb{T}$ on $\mathsf{sec}(\mathbb{T}, t_0)$ is an *ultrametric*, which means that $d_1(p, q) \leq d_1(p, r) \vee d_1(r, q)$ for any triplet of points $p, q, r \in \mathsf{sec}(\mathbb{T}, t_0)$. Accordingly, each triangle $p, q, r \in \mathsf{sec}(\mathbb{T}, t_0)$ is an *isosceles*, meaning that at least two of the three pairwise distances between p, q and r are equal and not greater than the third [63, Def. 3.31]. The length definition (227) implies that the distance between any pair of points from any fixed section $\mathsf{sec}(\mathbb{T}, t_0)$ equals the time until the two points (each of which can be either a particle or a sink) collide.

We notice that the collection of leaf vertices $\Delta^\circ_{p, \mathbb{T}}$ descendant to a point $p \in \mathbb{T}$ can be either a single point $(x_p, 0)$, if p is within a leaf edge and represents the ballistic run of a particle, or an interval $\{(x, 0) : x_{\text{left}}(p) \leq x \leq x_{\text{right}}(p)\}$, if p

is a non-leaf point that represents a sink. We define the *mass* $m(p)$ of a point $p \in \mathbb{T}$ as

$$m(p) = \int\limits_{x:(x,0)\in\Delta^\circ_{p,\mathbb{T}}} g(z)dz = x_{\mathrm{right}}(p) - x_{\mathrm{left}}(p),$$

where the last equality reflects the assumption $g(z) \equiv 1$. The mass $m(p)$ generalizes the quantity "number of descendant leaves" (Sect. 10.1.4) to the $\mathbb{R}$-tree situation with an uncountable set of leaves. We observe that (i) a point $p \in \mathbb{T}$ represents a ballistic run if and only if $m(p) = 0$; (ii) a point $p \in \mathbb{T}$ represents a sink if and only if $m(p) > 0$. This means that the shock wave tree, which is isometric to the sink part of the tree $(\mathbb{T}, d)$, can be extracted from $(\mathbb{T}, d)$ by the condition $\{p : m(p) > 0\}$.

11.5.2. Metric spaces on the set of initial particles

In this section we discuss two metrics on the system's domain $[a, b]$, which is isometric to the set $\{(x, 0) : x \in [a, b]\}$ of initial particles. These spaces contain the key information about the system dynamics and, unlike the complete tree $(\mathbb{T}, d)$ of Sect. 11.5.1, can be readily constructed from the potential $\Psi_0(x)$.

Metric $h_1(x, y)$ reproduces the ultrametric induced by $(\mathbb{T}, d)$ on $[a, b]$. Below we explicitly connect this metric to $\Psi_0(x)$. For any pair of points $x, y \in [a, b]$ we define a basin $\mathsf{B}_{\Psi_0}(x, y)$ as the interval that supports the minimal negative excursion within $\Psi_0(x)$ that contains the points x, y. Formally, assuming without loss of generality that $x < y$ we find the maximum of Ψ_0 on $[x, y]$:

$$\mathsf{m}_{\Psi_0}(x, y) = \sup_{z \in [x,y]} \Psi_0(z)$$

and use it to define the basin $\mathsf{B}_{\Psi_0}(x, y) = [l, r]$, where

$$l = \sup\{z : z \le x, \Psi_0(z) \ge \mathsf{m}_{\Psi_0}(x, y)\},$$

$$r = \inf\{z : z \ge y, \Psi_0(z) \ge \mathsf{m}_{\Psi_0}(x, y)\}.$$

The metric is now defined as

$$h_1(x, y) = \frac{1}{2}|\mathsf{B}_{\Psi_0(x,y)}|.$$

It is straightforward to check that

$$h_1(x, y) = \text{the time until collision of the particles } (x, 0) \text{ and } (y, 0),$$

where the collision is understood as either collision of particles, collision of sinks that annihilated the particles, or collision between a sink that annihilated one of the particles and the other particle. For instance, the claim is readily verified, by examining the bottom panel of Fig. 49, for any pair of points from the set $\{x, x', y, y'\}$. The metric space $([a, b], h_1)$ is not a tree. Moreover, this space

is *totally disconnected*, since there only exists a finite number of points (local minima of $\Psi_0(x)$) that have a neighborhood of arbitrarily small size. Any other point at the Euclidean distance ϵ from the nearest local minimum is separated from other points by at least $\epsilon/2$.

Metric $h_2(x,y)$ describes the mass accumulation by sinks during the annihilation process. Specifically, we introduce an equivalence relation among the annihilating particles, by writing $x \sim_{\Psi_0} y$ if the particles with initial coordinates x and y collide and annihilate with each other. For example, in Fig. 49 we have $x \sim_{\Psi_0} x'$ and $y \sim_{\Psi_0} y'$. The following metric is now defined on the quotient space $[a,b]|_{\sim_{\Psi_0}}$:

$$h_2(x,y) = 2 \sup_{z \in [x,y]} [\Psi_0(z)] - \Psi_0(x) - \Psi_0(y).$$

In words, the distance $h_2(x,y)$ between particles x and y equals the total mass accumulated by the sinks to which the particles belong during the time intervals between the instants when the particles joined the respective sinks and the instant of particle (or respective sink) collision. Another interpretation is that $h_2(x,y)$ equals to the minimal Euclidean distance between points $x,y \in [a,b]|_{\sim_{\Psi_0}}$ in the quotient space; one can travel in this quotient space as along a regular real interval, with a possibility to jump (with no distance accumulation) between equivalent points. This $\mathbb{R}$-tree construction is know as the *tree in continuous path* [135, Def. 7.6], [63, Ex. 3.14].

The metric space $([a,b]|_{\sim_{\Psi_0}}, h_2)$ is a tree that is isometric to the level set tree of the potential $\Psi_0(x)$ on $[a,b]$ and hence to the (finite) shock wave tree $\mathcal{V}(\Psi_0)$ (by Cor. 21), with the convention that the root is placed in $a \sim_{\Psi_0} b$. This means, in particular, that prunings of these two trees, with the same pruning function and pruning time, coincide.

11.5.3. Other prunings on $\mathbb{T}$

One can introduce a large class of prunings on an $\mathbb{R}$-tree $(\mathbb{T}, d)$ following the approach used above to define the point mass $m(p)$. Specifically, consider a measure $\eta(\cdot)$ on $[a,b]$ and define $m_\eta(p) = \eta(\Delta_{p,\mathbb{T}}^\circ)$. The function $m_\eta(p)$ is non-decreasing along each segment that connect a leaf and the root $\rho_\mathbb{T}$ of $\mathbb{T}$. Hence, one can define a pruning with respect to m_η on $\mathbb{T}$ by cutting all points p with $m_\eta(p) < t$ for a given $t \geq 0$. It is readily seen that the function $m_\eta(p)$ typically has discontinuities along a path between a leaf and the root of $\mathbb{T}$. This means that pruning with respect to m_η typically does not have the semigroup property.

12. Infinite trees built from leaves down

Examples of *infinite trees built from the root up* are plentiful; they include the infinite trees induced by the Yule processes or any other birth processes; infinite trees generated by a supercritical branching process; the trees that represent depth-first search and breadth-first search algorithms on infinite networks. In

this section we explore the *infinite trees built from leaves down* that arise naturally in the context of infinitely many coalescing particles or the level set trees of continuous functions. Interestingly, many of the results about finite trees can be obtained from the characterizations of the corresponding infinite trees built from leaves down.

12.1. Infinite plane trees built from the leaves down

In the context of Sect. 8.2, set $I = \mathbb{R}$ and consider a function $f(x) \in C(\mathbb{R})$. Let $\mathcal{X}$ and $\mathcal{Y}$ be the sets containing all locations of local minima and local maxima of $f(x)$, respectively. Formally, $x_0 \in \mathcal{X}$ if $\exists \delta > 0$ s.t. $f(x) \geq f(x_0)$ $\forall x \in (x_0 - \delta, x_0 + \delta)$, and $\mathcal{Y}$ is defined analogously. Hence, the local extrema may include plateaus of constant values. We assume that $f(x)$ satisfies the following conditions:

(a) The set $\mathcal{X}$ of the locations of local minima has infinite image, i.e.,

$$|f(\mathcal{X})| = \infty.$$

This condition guarantees that the level set tree of $f(x)$ that we construct below has an infinite number of vertices.

(b) The intersection of $\mathcal{X}$ with any finite interval $[a, b]$ is either empty or consists of a finite number of closed intervals (possibly including separate points). This condition guarantees that every descendant subtree of the infinite level set tree of $f(x)$ is finite. The conditions **(a), (b)** guarantee that the level set tree has countably many vertices.

(c) $\forall a \in \mathbb{R}$, the sets

$$(a, \infty) \cap f^{-1}\left(\inf_{(a,\infty)} f(x) \right) \quad \text{and} \quad (-\infty, a) \cap f^{-1}\left(\inf_{(-\infty,a)} f(x) \right)$$

are empty or consist of finitely many closed intervals (including separate points). Here, $f^{-1}(-\infty)$ is an empty set. This condition, or equivalent, guarantees that the level set tree has finite branching (no vertices of infinite degree).

Recalling the construction in Sect. 8.2.2, the level set tree $T_\infty = \textsc{level}\big(f(x)\big)$ has infinitely many leaves. There, $T_\infty = \left(\mathbb{R}/\sim_f, d_f \right)$ is a metric quotient space obtained with respect to identification (denoted by $a_\ell \sim_f a_r$) of pairs of points a_ℓ and a_r in $\mathbb{R}$ as one point. Recall that we have $a_\ell \sim_f a_r$ whenever the following conditions are satisfied

1. $a_\ell < a_r$ and $f(a_\ell) = f(a_r)$;
2. $\forall x \in (a_\ell, a_r)$ we have $f(x) \geq f(a_\ell) = f(a_r)$.

The local maxima $\mathcal{Y}$ (including plateaus) constitute the leaves in T_∞, and the local minima $\mathcal{X}$ (including plateaus) constitute the internal vertices (junctions) in T_∞. Such T_∞ is also called an *infinite plane tree built from the leaves down*
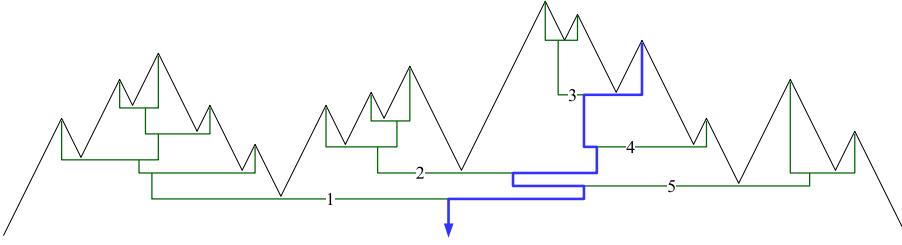
FIG 50. *Golden lineage representation of a level set tree: an illustration. The figure shows a part of the level set tree for a piece-wise linear function on $\mathbb{R}$. The initial part of the golden lineage is shown in blue. There exist five finite binary trees attached to the golden lineage – they are shown in green. The tree index is shown along the stem of each tree. The trees indexed 1, 2, 3 have left orientation, and the trees indexed 4, 5 have right orientation with respect to the golden lineage.*

induced by function $f(x)$. The reason for the name being that as we study $f(x)$ over larger and larger intervals (e.g. $[-a, a]$ as $a \to \infty$) we discover more and more leaves of T_∞ (local maxima) and their merger history (local minima) from leaves down, but never reaching the root.

To give a convenient description of an infinite tree T_∞ built from leaves down, we designate one leaf as the *golden leaf*, and its ancestral lineage is called the *golden lineage* (Fig. 50). In the above construction, we let the leaf that corresponds to the first local maximum in the nonnegative half-line,

$$\min\{x \in \mathcal{Y} : x \geq 0\},$$

to be designated as the golden leaf. Let $\mathcal{L}_{\mathrm{plane}}^\infty$ denote the space of infinite plane trees built from the leaves down, with edge lengths and designated golden leaf. For a tree $T_\infty \in \mathcal{L}_{\mathrm{plane}}^\infty$ with a designated golden leaf γ^*, we let $\ell = [\gamma^*, \phi]$ denote the unique ancestral path from the golden leaf γ^* to its parent, grandparent, great-grandparent and on towards the tree root ϕ, where ϕ is a *point at infinity*. Here, the ancestral path ℓ will be called the *golden lineage*. The golden lineage $\ell = \{\ell(i), e(i)\}$ consists of infinitely many vertices $\ell(i)$ that we enumerate by the index $i \geq 0$ along the path, starting from the golden leaf $\ell(0) = \gamma^*$ and increasing as we go down the golden lineage ℓ, and infinitely many edges $e(i) = [\ell(i), \ell(i+1)]$.

Each tree $T_\infty \in \mathcal{L}_{\mathrm{plane}}^\infty$ can be represented as a forest of finite trees attached to the golden lineage ℓ as follows

$$T_\infty = \left(\ell, \left\{ \mathcal{D}_i, \sigma_i \right\}_{i \geq 1} \right), \tag{228}$$

where for each $i \geq 1$, $\mathcal{D}_i = \Delta_{\ell(i)} \in \mathcal{L}_{\mathrm{plane}}^|$ denotes the complete subtree of T_∞ rooted at $\ell(i)$ that does not include the golden leaf, and $\sigma_i \in \{-1, +1\}$ denotes the left-right orientation of $\mathcal{D}_i$ with respect to the golden lineage ℓ. Figure 50 illustrates this construction.

The representation (228) of a tree $T_\infty \in \mathcal{L}^\infty_{\text{plane}}$ allows one to relate the space $\mathcal{L}^\infty_{\text{plane}}$ of infinite planar trees built from the leaves down with edge lengths and a designated golden leaf to the notion of a *forest of trees attached to the floor line* described in Sect. 7.4 of [135]. In addition, the golden lineage construct helps at meterizing the space $\mathcal{L}^\infty_{\text{plane}}$.

Importantly, for any point $x \in T_\infty$, the descendant tree $\Delta_{x,T}$ is a finite tree in $\mathcal{L}_{\text{plane}}$. Therefore, the definition of generalized dynamical pruning (218) extends naturally to the space $\mathcal{L}^\infty_{\text{plane}}$ of infinite plane trees built from the leaves down. Applying the generalized dynamical pruning $\mathcal{S}_t$ to an infinite tree built from the leaves down, the uppermost point of the golden lineage within $\mathcal{S}_t(\varphi, T)$ will become the golden leaf for the pruned tree $\mathcal{S}_t(\varphi, T)$.

Next, we extend the notion of prune-invariance in planar shapes from Def. 34(i) to a subspace S^∞ of the space $\mathcal{L}^\infty_{\text{plane}}$. Consider a subspace S^∞ of $\mathcal{L}^\infty_{\text{plane}}$. For a given monotone nondecreasing function $\varphi : \mathcal{L}_{\text{plane}} \to \mathbb{R}^+$, consider generalized pruning dynamics $\mathcal{S}_t(\varphi, T_\infty)$ ($T_\infty \in S^\infty$). We say that a probability measure μ on S_∞ is *prune-invariant in planar shapes* if

$$\mu(A) = \mu_t(A) \quad \forall t \geq 0, A \in \Sigma, \tag{229}$$

where $\mu_t = (\mathcal{S}_t)_*(\mu) = \mu \circ \mathcal{S}_t^{-1}$ is the pushforward measure, and Σ is the induced σ-algebra.

The above definition of prune-invariance (229) is significantly different from the original Def. 34(i) for finite trees as $\phi \notin S^\infty$ and we do not need to condition on the event $\mathcal{S}_t(\varphi, T) \neq \phi$ in the pushforward measure. Importantly, the prune-invariance in (229) coincides with the John Von Neumann [164] definition of the *invariant measure*, fundamental for ergodic theory and dynamical systems. At the same time, the definition of *prune-invariance in edge lengths* Def. 34(ii) does not need to be reformulated any differently for the infinite trees built from leaves down.

The renown Krylov-Bogolyubov theorem [93] states that for a compact metrizable topological space Ω with the induced Borel σ-algebra Σ, and a continuous function $\mathcal{S} : \Omega \to \Omega$, there exists an *invariant* probability measure μ on (Ω, Σ) satisfying

$$\mu(A) = \mu_*(A) \quad \forall A \in \Sigma, \tag{230}$$

where $\mu_* = (\mathcal{S})_*(\mu) = \mu \circ \mathcal{S}^{-1}$ is the pushforward measure.

Here we will not concentrate on constructing a suitable metric for the space $\mathcal{L}^\infty_{\text{plane}}$. However, in the spirit of the Krylov-Bogolyubov theorem, we will show in Thm. 34 that the infinite critical planar binary Galton-Watson tree $\mathsf{GW}_\infty(\lambda)$ built from the leaves down that we construct in Sect. 12.2 is prune-invariant under generalized dynamical pruning $\mathcal{S}_t$ induced by a monotone nondecreasing function $\varphi : \mathcal{L}_{\text{plane}} \to \mathbb{R}^+$. Additionally, it will be observed that Thm. 34 is a generalization of Thm. 26.

Remark 26. The reader would benefit from the following observation made by one of the anonymous reviewers of this paper. The golden lineage representation of an infinite planar tree built from the leaves down is reminiscent (but distinct

from) Bismut's decomposition of the Brownian excursion (see [30] and [137, Theorem 4.7 in Chapter 12]). Specifically, in [105], Bismut's decomposition of the Brownian excursion has been translated to a decomposition of real trees with respect to the ancestral lineage of a leaf chosen according to the mass measure in the setting of the continuum random tree (CRT), introduced by Aldous [4, 5]. The golden lineage idea is also similar to the Williams' decomposition [167] with respect to the ancestral lineage of the last individual alive in the genealogy of the total population given by a Lévy continuum random tree [106, 107, 52] which was established in Abraham and Delmas [2]. In the cases like ours, where the trees are infinite, the tree decompositions with respect to a particular lineage are also used in the context of superprocesses, where they are referred to as "skeleton decomposition", "backbone decomposition", or "spine decomposition". See [61, 24, 55] and references therein. Finally, [53] uses a similar approach when constructing the Galton-Watson forest.

12.2. Infinite exponential critical binary Galton-Watson tree built from the leaves down

Consider a Poisson point process $\{T_k\}_{k\in\mathbb{Z}}$ on $\mathbb{R}$ with parameter $\lambda/2$, enumerated from left to right (where T_0 is the epoch closest to zero). Let

$$
X_t = \begin{cases} \sum_{j=1}^{k}(-1)^{j+1}(T_j - T_{j-1}) + (-1)^k(T_k - t) & \text{if } t \in [T_{k-1}, T_k),\ k \geq 1, \\ \sum_{j=k+1}^{0}(-1)^{j+1}(T_j - T_{j-1}) + (-1)^k(T_k - t) & \text{if } t \in [T_k, T_{k+1}),\ k \leq -1. \end{cases}
$$

In other words, X_t is a continuous piecewise linear function with slopes alternating between ± 1 as it crosses the Poisson epochs $\{T_k\}_{k\in\mathbb{Z}}$, i.e., the slope

$$
\frac{d}{dt}X_t = \begin{cases} -1 & \text{if } t \in (T_{k-1}, T_k),\ k \text{ even,} \\ +1 & \text{if } t \in (T_{k-1}, T_k),\ k \text{ odd.} \end{cases}
$$

The level set tree $T_\infty = \text{LEVEL}(X_t)$ is invariant under shifting X_t vertically, or shifting and scaling X_t horizontally.

Fix a point $t^* \in \mathbb{R}$ and generate X_t with a Poisson point process $\{T_k\}_{k\in\mathbb{Z}}$. Then, with probability one, there will be a positive excursion of $X_t - X_{t^*}$ over an interval that begins or ends at t^*. By Thm. 20, the level set tree of this adjacent positive excursion is distributed as $\mathsf{GW}(\lambda)$. Therefore, the infinite binary level set tree $T_\infty = \text{LEVEL}(X_t)$ for X_t will be referred to as the *infinite planar exponential critical binary Galton-Watson tree built from the leaves down* with parameter λ, and denoted by $\mathsf{GW}_\infty(\lambda)$. We also refer to this tree as the infinite exponential critical binary Galton-Watson tree.

In the representation (228) of a tree $T_\infty \overset{d}{\sim} \mathsf{GW}_\infty(\lambda)$, the golden lineage ℓ is distributed as a one-dimensional Poisson process with parameter λ, the orientation variables σ_i are i.i.d. Bernoulli with parameter $1/2$, and the complete

(a) Tree   (b) Harris path
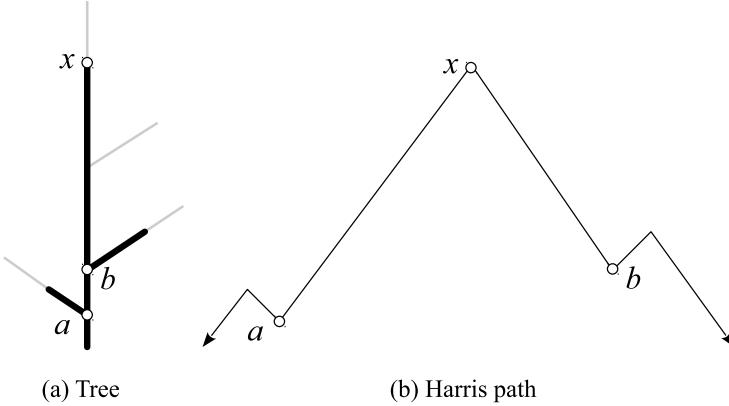
FIG 51. *Illustration to the proof of Thm. 34. (a) A fragment of the pruned tree T_∞^Δ (black). Parts of the pruned out subtrees of the initial tree T_∞ are shown in gray. The point x is a leaf of T_∞^Δ; it has the left parent a and right parent b in T_∞^Δ. The point x, considered as an edge point in T_∞, has the same left parent a in T_∞, but its right parent in T_∞ is different from b. (b) A part of the Harris path for the pruned tree T_∞^Δ. The proof of Thm. 34 finds that the increments of the segments $|xa|$ and $|xb|$ are i.i.d. exponential r.v.s. with parameter $\lambda p_\Delta/2$.*

subtrees $\mathcal{D}_i$ are i.i.d. $\mathsf{GW}(\lambda)$ trees. Finally, the golden lineage λ and the sequences, σ_i and $\mathcal{D}_i$, are all sampled independently of each other.

The following is a variation of Thm. 26 for the infinite critical exponential binary Galton-Watson tree.

Theorem 34. *Let $T_\infty \overset{d}{\sim} \mathsf{GW}_\infty(\lambda)$ with $\lambda > 0$. Then, for any monotone nondecreasing function $\varphi : \mathcal{BL}^|_{\text{plane}} \to \mathbb{R}_+$ and any $\Delta > 0$ we have*

$$T_\infty^\Delta := \mathcal{S}_\Delta(\varphi, T_\infty) \overset{d}{\sim} \mathsf{GW}_\infty\big(\lambda p_\Delta(\lambda, \varphi)\big),$$

where

$$p_\Delta := p_\Delta(\lambda, \varphi) = \mathsf{P}(\mathcal{S}_\Delta(\varphi, T) \neq \phi) \quad for \quad T \overset{d}{\sim} \mathsf{GW}(\lambda).$$

That is, the pruned tree T_∞^Δ is also an infinite exponential critical binary Galton-Watson tree with the scaled parameter

$$\mathcal{E}_\Delta(\lambda, \varphi) = \lambda p_\Delta(\lambda, \varphi).$$

Notice that since we are dealing with an infinite tree T_∞, we do not need to be concerned about it surviving under the pruning operation $\mathcal{S}_\Delta$. The survival probability p_Δ used in the statement of Thm. 34 is computed for finite trees, so the values of scaled parameter $\mathcal{E}_\Delta(\lambda, \varphi)$ for selected pruning functions are given by Thm. 27.

Proof. Let $\mathsf{par}(x)$ denote the *right parent* to a point x in T_∞. This means that the vertex $\mathsf{par}(x)$ is the parent of the first right subtree that one meets when

travels the tree T_∞ from x down to the root. In the Harris path of T_∞, there exist two points that correspond to x (they merge into a single point when x is a leaf). Consider the rightmost of these points, r_x, which belongs to a downward increment of the Harris path. The vertex $\mathsf{par}(x)$ corresponds to the nearest right local minima of r_x. Similarly, we let $\mathsf{par}_\Delta(\cdot)$ denote the right parent on T_∞^Δ.

Consider a leaf $x \in T_\infty^\Delta$, which is also a point in T_∞; see Fig. 51(a). We now find the distribution of the distance from x to $\mathsf{par}_\Delta(x)$, i.e., the length of the respective downward segment of the Harris path; see Fig. 51(b). Consider the descendant lineage of x in T_∞, which consists of vertices

$$x_1 = \mathsf{par}(x), \quad x_2 = \mathsf{par}(x_1), \quad x_3 = \mathsf{par}(x_2), \quad \dots \ .$$

Due to the memorylessness property of exponential distribution, and the symmetry of left-right orientation of subtrees in T_∞, the distance from x down to $\mathsf{par}(x)$ has exponential distribution with rate $\lambda/2$. The point x belongs to one (left) of the two complete subtrees rooted at $\mathsf{par}(x)$ in T_∞. Observe that $\mathsf{par}_\Delta(x) = \mathsf{par}(x)$ if and only if the subtree that does not contain x (we call it sibling subtree) has not been pruned out completely, i.e., the intersection of the sibling subtree with T_∞^Δ is not empty. (In the example of Fig. 51(a), we have $\mathsf{par}_\Delta(x) = x_2 \equiv b$.) The sibling subtree is known to be distributed as $\mathsf{GW}(\lambda)$. Therefore,

$$\mathsf{P}\big(\mathsf{par}_\Delta(x) = x_1\big) = p_\Delta.$$

Iterating this argument, we have for $k \geq 1$,

$$\mathsf{P}\big(\mathsf{par}_\Delta(x) = x_k \ \big| \ \mathsf{par}_\Delta(x) \neq x_1, \dots, \mathsf{par}_\Delta(x) \neq x_{k-1}\big) = p_\Delta(1 - p_\Delta)^{k-1}.$$

Therefore, the distance from a vertex x down to $\mathsf{par}_\Delta(x)$ is a geometric $\mathsf{Geom}_1(p_\Delta)$ sum of independent exponential random variables with parameter $\lambda/2$. Hence, it is itself an exponential random variable with parameter $\lambda p_\Delta/2$. In other words, the downward segment of the Harris path of the pruned tree T_∞^Δ adjacent to the local maximum that corresponds to the leaf x has exponential lengths with parameter $\lambda p_\Delta/2$; see Fig. 51(b).

The same argument (using left parents) shows that the upward segment of the Harris path of the pruned tree T_∞^Δ adjacent to the local maximum that corresponds to the leaf x has exponential lengths with parameter $\lambda p_\Delta/2$. The lengths of the upward and downward segments are independent; see Fig. 51(b).

Applying the above argument to all leaves in T_∞^Δ, we conclude that the Harris path of T_∞^Δ consists of alternating up/down increments with independent lengths, distributed exponentially with the parameter $\lambda p_\Delta/2$. Theorem 20 states that in this case T_∞^Δ is an exponential critical binary Galton-Watson tree with parameter λp_Δ. This completes the proof. $\qquad\square$

Observe that Thm. 26 can be obtained from Thm. 34 by considering finite excursions of X_t. Also notice that for the particular case of Horton pruning (Sect. 10.1.2), the statement of Thm. 34 follows from Thm. 19.

12.3. *Continuum annihilation*

One can observe that the continuum annihilation dynamics that begins with an infinite exponential potential $\Psi_0^{\mathrm{exp}}(x)$, $x \in \mathbb{R}$ (see Sect. 11.4), is nothing but the generalized dynamical pruning $\mathcal{S}_t(\varphi, T_\infty)$ of the infinite planar critical exponential binary Galton-Watson tree built from the leaves down

$$T_\infty := \mathrm{LEVEL}\big(- \Psi_0^{\mathrm{exp}}(x)\big) \overset{d}{\sim} \mathsf{GW}_\infty(\lambda),$$

where $\varphi(T) = \mathrm{LENGTH}(T)$ for $T \in \mathcal{BL}_{\mathrm{plane}}^|$. Moreover, the key results of Sect. 11.4, Thms. 32 and 33, that describe the growth dynamics of a sink in the continuum annihilation model are in fact describing the length distributions of pruned out sections of $T_\infty \overset{d}{\sim} \mathsf{GW}_\infty(\lambda)$ under the generalized dynamical pruning $\mathcal{S}_t(\varphi, T_\infty)$. The proofs of these results can be rewritten in the infinite tree style of Thm. 34.

13. Some open problems

1. Consider the cumulative distribution function $\mathsf{H}_n(x)$ for the height of an exponential critical binary Galton-Watson tree $\mathsf{GW}(\lambda)$ (Def. 20) conditioned on having n leaves; see (80) of Sect. 5.2.3. Can one derive the limit (89) from the equation (86)?
2. For a given sequence $\{T_k\}_{k \in \mathbb{Z}_+}$ of positive real numbers, construct a coalescent process whose symmetric kernel is a function of the clusters' Horton-Strahler orders, in such a way that the combinatorial part of the coalescent tree is mean self-similar with respect to Horton pruning (Defs. 12 and 14), with Tokunaga coefficients $\{T_k\}$. This would complement an analogous branching process construction of Sect. 7.
3. Generalize equation (63) of Flajolet et al. [66] for the critical Tokunaga processes (Sect. 7.5). Formally, consider a tree T that corresponds to a critical Tokunaga process $S^{\mathrm{Tok}}(t; c, \gamma)$ (Def. 25). Establish the following generalization of (63): for any given $c > 1$, there exists a periodic function $D_c(\cdot)$ of period one such that

$$\mathsf{E}\big[\mathsf{ord}(T) \,\big|\, \#T = 2n - 1\big] = \log_R n + D_c\big(\log_R n\big) + o(1) \qquad (231)$$

   as $n \to \infty$, where $R = 2c$. We confirmed the validity of (231) numerically; see Fig. 52.
4. For a hierarchical branching process $S(t)$ (Def. 22, Sect. 7.1), describe the correlation structure of its Harris path. A special case is given by Thm. 20; it shows that the Harris path of the exponential critical binary Galton-Watson tree $\mathsf{GW}(\lambda)$, which corresponds to the hierarchical branching process $S(t) \overset{d}{\sim} S^{\mathrm{Tok}}(t; c, \gamma)$ (Sect. 7.5), is an excursion of the exponential random walk (Sect. 8.6), with parameters $\big\{\frac{1}{2}, \lambda, \lambda\big\}$.
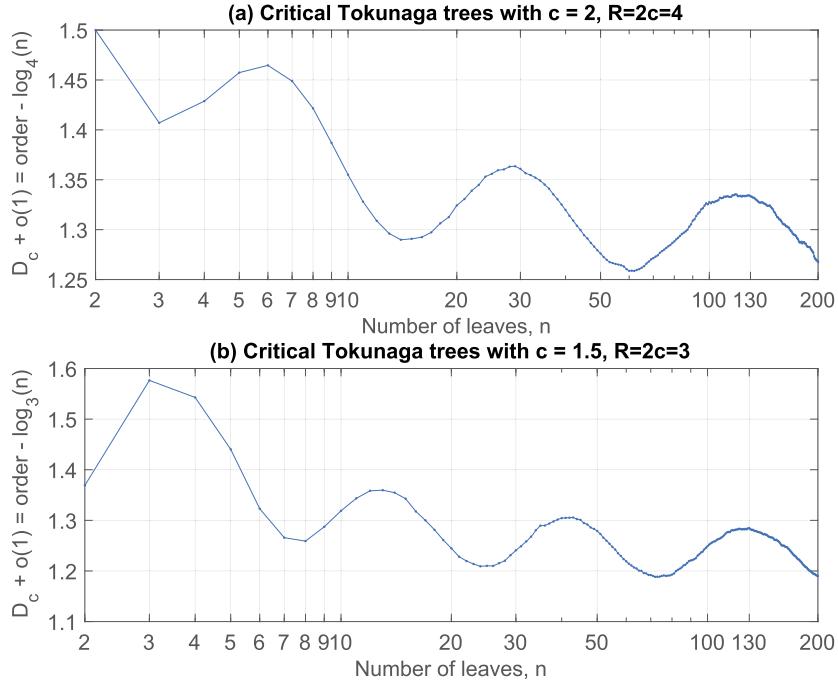
FIG 52. *Periodic fluctuations of the average order of a critical Tokunaga tree $S^{\mathrm{Tok}}(t; c, \gamma)$ as a function of the tree size (number n of leaves): numerical simulations. (a) Critical Tokunaga tree with $c = 2, R = 2c = 4$ (critical binary Galton-Watson tree); see* (63). *(b) Critical Tokunaga tree with $c = 1.5, R = 2c = 3$; see* (231). *The jitter at large values of n is due to stochastic variability in our numerical simulations.*

5. Recall that a rescaled Harris path of an exponential critical binary Galton-Watson tree $\mathsf{GW}(\lambda)$ converges to the excursion of a standard Brownian motion [105, 124]. For a hierarchical branching process $S(t)$ (Def. 22, Sect. 7.1), explore the existence of a proper infinite-tree limit and the respective limiting excursion process.

6. Prove the following extension of Lem. 21. In the setup of the Lemma, suppose that for any tree T, conditioned on P-SHAPE(T), the edge lengths in T are independent. Show that $f(x)$ is an exponential p.d.f.

7. Can the finite second moment assumption in Prop. 16 be removed? Also, does (173) characterize the exponential distribution (like the characterizations in Appendix B)?

8. In the context of Sect. 8.9, extend the one-dimensional result of Prop. 15 to higher dimensions. Specifically, consider an n-dimensional compact differentiable manifold $M = M^n$, and a Morse function $f : M \to \mathbb{R}$. Construct a natural Morse function $f^{(1)} : M \to \mathbb{R}$ such that

$$\text{LEVEL}\big(f^{(1)}\big) = \mathcal{R}\big(\text{LEVEL}(f)\big).$$

9. In the setting of Thm. 25 from Sect. 9, establish the asymptotic ratio-Horton law (Def. 19) for the Kingman's coalescent tree, and, if possible, prove the asymptotic strong Horton law (Def. 19). Specifically, prove $\lim_{j \to \infty} \frac{\mathcal{N}_j}{\mathcal{N}_{j+1}} = R$, and if possible, $\lim_{j \to \infty} (\mathcal{N}_j R^j) = const.$ Is it possible to derive a closed form expression for the Horton exponent R?

10. Find a suitable ramification of the generalized dynamical pruning sufficient for describing the evolution of the shock tree in the one-dimensional inviscid Burgers equation (226) and its multidimensional modification known as the *adhesion model* [19, 68, 77]. Use this to complement the framework developed in [149, 27, 28, 70].

Appendix A: Weak convergence results of Kurtz for density dependent population processes

We first formulate the framework for the convergence result of Kurtz as stated in [60, Chapter 11, Theorem 2.1], [103, Theorem 8.1]. There, the *density dependent population processes* are defined as continuous time Markov processes with state spaces in $\mathbb{Z}^d$, and transition intensities represented as follows

$$q^{(n)}(k, k + \ell) = n \left[\beta_\ell \left(\frac{k}{n} \right) + O \left(\frac{1}{n} \right) \right], \tag{232}$$

where $\ell, k \in \mathbb{Z}^d$, and β_ℓ is a given collection of rate functions.

In Section 5.1 of [6], Aldous observes that the results from Chapter 11 of Ethier and Kurtz [60] can be used to prove the weak convergence of a Marcus-Lushnikov process to the solutions of Smoluchowski system of equations in the case when the Marcus-Lushnikov process can be formulated as a finite dimensional density dependent population process. Specifically, the Marcus-Lushnikov processes corresponding to the multiplicative and Kingman's coalescent with the monodisperse initial conditions (n singletons) can be represented as finite dimensional density dependent population processes defined above.

Define $F(x) = \sum_\ell \ell \beta_\ell(x)$. Then, Theorem 2.1 in Chapter 11 of [60] (Theorem 8.1 in [103]) states the following law of large numbers. Let $\hat{X}_n(t)$ be the Markov process with the intensities $q^{(n)}(k, k + \ell)$ given in (232), and let $X_n(t) = n^{-1}\hat{X}_n(t)$. Finally, let $|x| = \sqrt{\sum x_i^2}$ denote the Euclidean norm in $\mathbb{R}^d$.

Theorem 35. *Suppose for all compact $\mathcal{K} \subset \mathbb{R}^d$,*

$$\sum_\ell |\ell| \sup_{x \in \mathcal{K}} \beta_\ell(\bar{x}) < \infty,$$

and there exists $M_{\mathcal{K}} > 0$ such that

$$|F(x) - F(y)| \leq M_{\mathcal{K}}|x - y|, \qquad \text{for all } x, y \in \mathcal{K}. \tag{233}$$

Suppose $\lim_{n \to \infty} X_n(0) = x_0$, and $X(t)$ satisfies

$$X(t) = X(0) + \int_0^t F(X(s))ds, \tag{234}$$

for all $T \geq 0$. Then

$$\lim_{n \to \infty} \sup_{s \in [0,T]} |X_n(s) - X(s)| = 0 \qquad a.s. \tag{235}$$

Appendix B: Characterization of exponential random variables

This section contains a number of characterization results for exponential random variables that we use in this manuscript. We refer the reader to [13, 8] for more on characterization of exponential random variables.

The following result of K. S. Lau and C. R. Rao [104] that implies a characterization of exponential random variables is used by us for establishing Lemma 21. See [15] for more on Integrated Cauchy Functional Equations.

Lemma 33 ([104]). *Consider an Integrated Cauchy Functional Equation*

$$\int_0^\infty \frac{G(x + y)}{G(y)} \, d\mu(y) = G(x) \quad \forall x \geq 0, \tag{236}$$

where $\mu(\cdot)$ is a p.d.f. on $[0, \infty)$ and $G(x) > 0$ for x in the support of μ. Then, $G(x) = e^{-\lambda x}$ for some $\lambda \geq 0$.

The following characterization of exponential random variables follows immediately from Lemma 33.

Lemma 34. *Consider a p.d.f. $g(x)$ defined on $[0, \infty)$, and satisfying*

$$g(x) = 2 \int_0^\infty g(x + y)g(y) \, dy \quad \forall x \geq 0. \tag{237}$$

Then, $g(x)$ is an exponential density function.

Proof. Let $G(a) = \int_a^\infty g(x) \, dx$. Then, integrating (237), we have for all $a \geq 0$,

$$G(a) = \int_a^\infty g(x) \, dx = 2 \int_0^\infty G(a + y)g(y) \, dy = \int_0^\infty \frac{G(a + y)}{G(y)} \, d\mu(y), \tag{238}$$

where $\mu(y) = 1 - G^2(y)$ is a p.d.f. on $[0, \infty)$. We notice that (238) produces equation (236). Hence, by Lem. 33, $G(x) = e^{-\lambda x}$, where $\lambda > 0$ as $g(x)$ is p.d.f. $\qquad \square$

Next, we recall the Parseval's identity, which we will use in the proof of characterization Lemma 35.

Theorem 36 (**Parseval's identity,** [160]). *For a pair of cumulative distribution functions $F(x)$ and $G(x)$ and their respective characteristic functions $\widehat{f}(s)$ and $\widehat{g}(s)$ the following identity holds for all $s \in \mathbb{R}$*

$$\int_{-\infty}^{\infty} e^{isx} \, \widehat{g}(x) \, dF(x) \;=\; \int_{-\infty}^{\infty} \widehat{f}(x+s) \, dG(x).$$

We give yet another characterization of the exponential p.d.f. $\phi_\lambda(x) = \lambda e^{-\lambda x} \mathbf{1}_{\{x \geq 0\}}$ as defined in (72).

Lemma 35. *Consider a p.d.f. $g(x)$ defined on $[0, \infty)$, and satisfying*

$$\phi_\lambda(x) = 2 \int_0^\infty g(x+y)g(y) \, dy \quad \forall x \geq 0. \tag{239}$$

Then, $g(x) = \phi_\lambda(x)$.

Proof. Observe that $\phi_\lambda(x)$ satisfies

$$\phi_\lambda(x) = 2 \int_0^\infty \phi_\lambda(x+y)\phi_\lambda(y) \, dy \quad \forall x \geq 0. \tag{240}$$

Thus,

$$\int_0^\infty \phi_\lambda(x+y)\phi_\lambda(y) \, dy = \int_0^\infty g(x+y)g(y) \, dy \quad \forall x \geq 0. \tag{241}$$

Hence, for the two pairs of independent random variables

$$X_1, X_2 \overset{d}{\sim} \text{ p.d.f. } \phi_\lambda(x) \quad \text{and} \quad Y_1, Y_2 \overset{d}{\sim} \text{ p.d.f. } g(x),$$

we have

$$X_1 - X_2 \overset{d}{=} Y_1 - Y_2.$$

Therefore, for the characteristic functions $\widehat{\phi}_\lambda$ and $\widehat{g}$, we have

$$\left| \widehat{\phi}_\lambda(s) \right|^2 \;=\; \mathsf{E}\left[e^{is(X_1 - X_2)} \right] \;=\; \mathsf{E}\left[e^{is(Y_1 - Y_2)} \right] \;=\; \left| \widehat{g}(s) \right|^2. \tag{242}$$

Observe that (242) can be also obtained from (241) via multiplying both sides by e^{isx} and integrating.

Next, from the Parseval's identity Theorem 36 and (241), we have $\forall s \geq 0$,

$$\int_0^\infty e^{isy} \widehat{g}(y) \, g(y) \, dy = \int_0^\infty g(s+y)g(y) \, dy$$

$$= \int\limits_0^\infty \phi_\lambda(s+y)\phi_\lambda(y)\,dy$$

$$= \int\limits_0^\infty e^{isy}\widehat{\phi}_\lambda(y)\,\phi_\lambda(y)\,dy. \tag{243}$$

Therefore,

$$\widehat{g}(x)\,g(x) \equiv \widehat{\phi}_\lambda(x)\,\phi_\lambda(x),$$

and (242) implies for any $x > 0$,

$$g(x) = \frac{\left|\widehat{g}(x)\,g(x)\right|}{\left|\widehat{g}(x)\right|} = \frac{\left|\widehat{\phi}_\lambda(x)\,\phi_\lambda(x)\right|}{\left|\widehat{\phi}_\lambda(x)\right|} = \phi_\lambda(x). \qquad \square$$

Appendix C: Notations

ρ	*root* vertex;
ϕ	the *empty tree* comprised of a root vertex and no edges;
$\mathcal{T}$	the space of finite unlabeled rooted reduced trees with no planar embedding;
$\mathcal{L}$	the space of trees from $\mathcal{T}$ with edge lengths;
$\mathcal{T}_{\mathrm{plane}}$	the space of finite unlabeled rooted reduced trees with planar embedding;
$\mathcal{L}_{\mathrm{plane}}$	the space of trees from $\mathcal{T}_{\mathrm{plane}}$ with edge lengths;
$\mathcal{BS}$	the subspace of *binary* trees in a given space of rooted trees $\mathcal{S}$, e.g., $S = \mathcal{T}, \mathcal{T}_{\mathrm{plane}}, \mathcal{L}, \mathcal{L}_{\mathrm{plane}}$;
$\mathcal{S}^{\vert}$	the subspace of *planted* trees in a given space of rooted trees $\mathcal{S}$;
$\mathcal{S}^{\vee}$	the subspace of *stemless* trees in a given space of rooted trees $\mathcal{S}$;
$\widetilde{\mathcal{BL}}_{\mathrm{plane}}$	the subspace of *mass-equipped* trees in $\mathcal{BL}_{\mathrm{plane}}$;
$\mathcal{GW}(\{q_k\})$	the probability distribution of (combinatorial) Galton-Watson trees on $\mathcal{T}^{\vert}$ with offspring p.m.f. $\{q_k\}$;
$\mathcal{GW}(q_0, q_2)$	the probability distribution of (combinatorial) binary Galton-Watson trees on $\mathcal{BT}^{\vert}$ with termination probability q_0 and split probability q_2;
$\mathcal{GW}_{\mathrm{plane}}(q_0, q_2)$	the planar embedding of trees in $\mathcal{GW}(q_0, q_2)$ that assigns the left-right orientation to each pair of offsprings uniformly and independently;
$\mathcal{IGW}(q)$	the invariant Galton-Watson (IGW) measure with parameter q and generating function $z + q(1-z)^{1/q}$ (see Def. 21);
$\mathsf{GW}(\lambda', \lambda)$	the probability distribution of exponential binary Galton-Watson trees (see Def. 30);
$\mathsf{GW}(\lambda)$	the probability distribution of exponential critical binary Galton-Watson trees (Def. 20);

$\mathsf{GW}_\infty(\lambda)$	the probability distribution of infinite exponential critical binary Galton-Watson tree built from the leaves down, with parameter $\lambda > 0$;
$\mathsf{parent}(v)$	the unique parent vertex of a non-root vertex v;
$X \stackrel{d}{\sim} D$	random element X has distribution D;
$X \stackrel{d}{=} Y$	random elements X and Y are equidistributed;
$\hat{t}(z)$	the generating function (z-transform) of a sequence $\{t(j)\}_{j=0,1,\ldots}$;
$\widehat{f}(s)$	the characteristic function of a random variable with p.d.f. $f(x)$;
$\mathcal{L}f(s)$	the Laplace transform of $f(x)$;
$\stackrel{a.s.}{\to}$	almost sure convergence;
$\stackrel{d}{\to}$	convergence in distribution;
$\stackrel{p}{\to}$	convergence in probability;
$x_n \sim y_n$	asymptotic equivalence: $\lim\limits_{n\to\infty} \dfrac{x_n}{y_n} = 1$;
$\mathcal{E}^{\mathrm{ex}}$	the space of all positive piece-wise linear continuous finite excursions with alternating slopes ± 1;
$\mathbb{N}$	the set of natural numbers $\{1, 2, \ldots\}$;
$\mathbb{Z}_+$	the set of nonnegative integer numbers $\{0, 1, 2, \ldots\}$;
$\mathbb{R}_+$	the set of nonnegative real numbers $[0, \infty)$.

Appendix D: Standard distributions

$\mathsf{Exp}(\lambda)$	the exponential distribution with rate λ; the respective p.d.f. is $\phi_\lambda(x) = \lambda e^{-\lambda x}$, $x \geq 0$;
$\mathsf{Gamma}(\alpha, \lambda)$	the gamma distribution with parameters $\alpha > 0$ and $\lambda > 0$; the respective p.d.f. is $f(x) = \lambda^\alpha x^{\alpha-1} e^{-\lambda x}/\Gamma(\alpha)$ for $x \geq 0$;
$\mathsf{Geom}_0(p)$	the geometric distribution with p.m.f. $p(m) = p\,(1-p)^m$ for $m = 0, 1, 2, \ldots$;
$\mathsf{Geom}_1(p)$	the geometric distribution with p.m.f. $p(m) = p\,(1-p)^{m-1}$ for $m = 1, 2, 3, \ldots$;
$\mathsf{Poi}(\lambda)$	the Poisson distribution with rate $\lambda > 0$; the respective p.m.f. is $p(m) = \lambda^m e^{-\lambda}/m!$ for $m = 0, 1, 2, \ldots$;
$\mathsf{Unif}(A)$	the uniform distribution over a set A.

Appendix E: Tree functions and mappings

$\textsc{length}(T)$	the length of a tree $T \in \mathcal{L}$ (or $\mathcal{L}_{\mathrm{plane}}$) defined as the sum of the lengths of its edges;
$\textsc{height}(T)$	the height of a tree $T \in \mathcal{L}$ (or $\mathcal{L}_{\mathrm{plane}}$) defined as the maximal distance between the root and a vertex;
$\textsc{shape}(T)$	the combinatorial shape of a tree $T \in \mathcal{L}$ (or $\mathcal{L}_{\mathrm{plane}}$); it is a mapping from $\mathcal{L}$ (or $\mathcal{L}_{\mathrm{plane}}$) to $\mathcal{T}$;

P-SHAPE(T)	the combinatorial shape of a tree $T \in \mathcal{L}_{\mathrm{plane}}$ together with the tree's planar embedding; it is a mapping from $\mathcal{L}_{\mathrm{plane}}$ to $\mathcal{T}_{\mathrm{plane}}$;
LEVEL(f)	the level set tree of a continuous function $f(x)$.

Acknowledgements

References

[1] R. Abraham, J.-F. Delmas, H. He, *Pruning Galton-Watson trees and tree-valued Markov processes* Ann. Inst. H. Poincaré Probab. Statist., **48**(3) (2012) 688–705. MR2976559

[2] R. Abraham and J.-F. Delmas, *Williams' decomposition of the Lévy continuum random tree and simultaneous extinction probability for populations with neutral mutations* Stochastic Processes and their Applications, **119**(4) (2009) 1124–1143. MR2508567

[3] M. Abramowitz and I. A. Stegun, *Handbook of mathematical functions: with formulas, graphs, and mathematical tables* Courier Corporation, **55** (1964). MR0208797

[4] D. J. Aldous, *The continuum random tree I.* The Annals of Probability, **19**(1) (1991) 1–28. MR1085326

[5] D. J. Aldous, *The continuum random tree III.* The Annals of Probability, **21**(1) (1993) 248–289. MR1207226

[6] 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. MR1673235

[7] D. J. Aldous and J. Pitman, *Tree-valued Markov chains derived from Galton-Watson processes* Ann. Inst. H. Poincaré Probab. Statist., **34**(5) (1998) 637–686. MR1641670

[8] B. C. Arnold and J. S. Huang, in *Exponential distribution: theory, methods and applications* (edited by K. Balakrishnan and A. P. Basu), CRC Press, Taylor & Francis Group (1996). MR1655093

[9] V. I. Arnold, *On the representation of continuous functions of three variables by superpositions of continuous functions of two variables* Matematicheskii Sbornik Vol. **48**(90), no. 1, (1959) 3–74. MR0121453

[10] V. I. Arnold, *Smooth functions statistics* Functional Analysis and Other Mathematics **1**(2), (2006) 111–118. MR2385493

[11] V. I. Arnold, *Topological classification of Morse functions and generalizations of Hilbert's 16-th problem* Mathematical Physics, Analysis and Geometry **10**(3), (2007) 227–236. MR2368960

[12] K. B. Athreya and P. E. Ney, *Branching processes* Courier Corporation (2004). MR2047480

[13] T. A. Azlarov and N. A. Volodin, *Characterization problems associated with the exponential distribution* Springer Verlag (1986). MR0841073

[14] M. Baiesi and M. Paczuski, *Scale-free networks of earthquakes and aftershocks* Physical Review E, (2004) **69**(6) 066106.

[15] R. Balasubrahmanyan and K. S. Lau, *Functional equations in probability theory* Elsevier (2014).

[16] J. R. Banavar, A. Maritan, and A. Rinaldo, *Size and form in efficient transportation networks* Nature, (1999) **399**(6732) 130.

[17] B. Barbosa, C. Campregher, J. Johnson, Y. Kovchegov, and I. Zaliapin, *Self-similarity in level set trees of geometric random walks* preprint.

[18] O. E. Barndorff-Nielsen, *River networks: a brief guide to the literature for statisticians and probabilists* Chapter 6 (pp.276–287) in Chaos and Networks. Statistical and Probabilistic Aspects. Chapman & Hall (1993).

[19] J. Bec and K. Khanin, *Burgers turbulence* Physics Reports, **447**(1) (2007) 1–66. MR2318582

[20] V. Belitsky and P. A. Ferrari, *Ballistic annihilation and deterministic surface growth* Journal of Statistical Physics, **80**(3-4) (1995) 517–543. MR1342240

[21] E. Ben-Naim, S. Redner, and F. Leyvraz, *Decay kinetics of ballistic annihilation* Physical Review Letters **70**(12), (1993) 1890–1893.

[22] E. Ben-Naim, S. Redner, and P. L. Krapivsky, *Two scales in asynchronous ballistic annihilation* J. Phys. A: Math. Gen., **29** L561 (1996).

[23] J. Berestycki, N. Berestycki, and V. Limic, *The Λ-coalescent speed of coming down from infinity* Ann. Probab. **38**(1), (2010) 207–233. MR2599198

[24] J. Berestycki, A. E. Kyprianou, and A. Murillo-Salas, *The prolific back-*

bone for supercritical superprocesses Stochastic Process. Appl. **121**, (2011) 1315–1331. MR2794978

[25] N. Berestycki, *Recent progress in coalescent theory* Ensaios Matemáticos, **16**, (2009) 1–193. MR2574323

[26] F. Bernardeau and P. Valageas, *The prolific backbone for supercritical superprocesses* Phys. Rev. E, **82**:016311 (2010).

[27] J. Bertoin, *The inviscid Burgers equation with Brownian initial velocity* Comm. Math. Phys. **193**(2) (1998) 397–406. MR1618139

[28] J. Bertoin, *Clustering statistics for sticky particles with Brownian initial velocity* Journal de Mathématiques Pures et Appliquées, **79**(2) (2000) 173–194. MR1749158

[29] J. Bertoin, *Random Fragmentation and Coagulation Processes* Cambridge University Press (2006). MR2253162

[30] J.-M. Bismut, *Last exit decompositions and regularity at the boundary of transition probabilities* Z. Wahrscheinlichkeitstheor. Verw. Geb. **69** (1985) 65–98. MR0775853

[31] B. Biswal and M. Marani, *Geomorphological origin of recession curves* Geophysical Research Letters, (2010) **37**(24).

[32] B. Bollobás, *Modern graph theory* (Vol. 184) Springer Science & Business Media (2013). MR1633290

[33] R. A. Blythe, M. R. Evans, and Y. Kafri, *Stochastic ballistic annihilation and coalescence* Physical Review Letters, **85**(18) (2000) 3750–3753.

[34] G. A. Burd, E. C. Waymire, and R. D. Winn, *A self-similar invariance of critical binary Galton-Watson trees* Bernoulli, **6** (2000) 1–21. MR1781179

[35] E. H. Campbell Grant, W. H. Lowe, and W. F. Fagan, *Living in the branches: population dynamics and ecological processes in dendritic networks* Ecology Letters, (2007) **10**(2) 165–175.

[36] M. P. do Carmo, *Riemannian geometry* Birkhäuser (1992). MR1138207

[37] F. Cassot, F. Lauwers, C. Fouard, S. Prohaska, and V. Lauwers-Cances, *A novel three-dimensional computer-assisted method for a quantitative study of microvascular networks of the human cerebral cortex* Microcirculation, (2006) **13**(1) 1–18.

[38] E. V. Chunikhina, *Entropy rates for Horton self-similar trees* Chaos, **28**(8), (2018) 081104. MR3846126

[39] E. V. Chunikhina, *Information Theoretical Analysis of Self-Similar Trees* Ph.D. thesis (2018). https://ir.library.oregonstate.edu/concern/graduate_thesis_or_dissertations/w66348893.

[40] M. Cieplak, A. Giacometti, A. Maritan, A. Rinaldo, I. Rodriguez-Iturbe and J. R. Banavar *Models of fractal river basins.* Journal of Statistical Physics, **91**(1–2) (1998) 1–15.

[41] P. Collet, S. Martínez, and J. San Martín, *Quasi-stationary distributions: Markov chains, diffusions and dynamical systems* Springer Science & Business Media (2012). MR2986807

[42] R. Darling and J. Norris, *Differential equation approximations for Markov chains* Probab. Surveys **5** (2008) 37–79. MR2395153

[43] L. Devroye and P. Kruszewski, *A note on the Horton-Strahler number for*

random trees, Inform. Processing Lett., **56** (1994) 95–99. MR1359172

[44] P. S. Dodds and D. H. Rothman, *Unified view of scaling laws for river networks* Phys. Rev. E, **59**(5) (1999) 4865.

[45] P. S. Dodds and D. H. Rothman, *Scaling, Universality, and Geomorphology*, Ann. Rev. Earth and Planet. Sci., **28** (2000) 571–610.

[46] M. Drmota, *Random trees: an interplay between combinatorics and probability* Springer Science & Business Media (2009). MR2484382

[47] M. Drmota, *The Height of Increasing Trees* Ann. Comb. **12** (2009) 373–402. MR2496124

[48] M. Drmota and H. Prodinger, *The register function for t-ary trees* ACM Transactions on Algorithms **2** (3) (2006) 318–334. MR2253784

[49] M. Droz, P.-A. Rey, L. Frachebourg, and J. Piasecki, *Ballistic-annihilation kinetics for a multivelocity one-dimensional ideal gas* Phys. Rev. E, **51**(6) (1995) 5541–5548.

[50] R. Durrett and D. L. Iglehart, *Functionals of Brownian meander and Brownian excursion* Ann. Prob. **5** (1977) 130–135. MR0436354

[51] R. Durrett, H. Kesten, and E. C. Waymire, *Functionals of Brownian meander and Brownian excursion* Journal of Theoretical Probability **4**(1) (1991) 223–237.

[52] T. Duquesne and J. F. Le Gall, *Random trees, Lévy processes and spatial branching processes* Société mathématique de France, Vol. **281**, (2002). MR1954248

[53] T. Duquesne and M. Winkel, *Growth of Lévy trees* Probab. Theory Related Fields **139**, (2007) 313–371. MR2322700

[54] T. Duquesne and M. Winkel, *Hereditary tree growth and Lévy forests* Stochastic Processes and their Applications, **129**(10), (2019) 3690–3747. MR3997659

[55] M. Eckhoff, A. E. Kyprianou, and M. Winkel, *Spines, skeletons and the strong law of large numbers for superdiffusions* Ann. Probab. **43**(5), (2015) 2545–2610. MR3395469

[56] H. Edelsbrunner and J. Harer, *Computational topology: an introduction* American Mathematical Soc. (2010). MR2572029

[57] Y. Elskens and H. L. Frisch, *Annihilation kinetics in the one-dimensional ideal gas* Physical Review A, **31**(6) (1985) 3812–3816.

[58] A. Ermakov, B. Toth, and W. Werner, *On some annihilating and coalescing systems* Journal of Statistical Physics, **91**(5–6) (1998) 845–870. MR1637254

[59] A. P. Ershov, *On programming of arithmetic operations* Communications of the ACM, **1** no. 8 (1958), 3–6. MR1045201

[60] S. N. Ethier and T. G. Kurtz, *Markov Processes. Characterization and convergence* Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York (1986). MR0838085

[61] S. N. Evans and N. O'Connell, *Weighted occupation time for branching particle systems and a representation for the supercritical superprocess*, Canad. Math. Bull. **37** (1994), 187–196. MR1275703

[62] S. N. Evans and J. Pitman, *Construction of Markovian coalescents*, Ann. Inst. H. Poincaré (B) Probab. Statist., vol. **34**, no. 3 (1998), 339–383. MR1625867

[63] S. N. Evans, *Probability and real trees: Ecole d'été de probabilités de Saint-Flour* Lectures on Probability Theory and Statistics. Springer (2008). MR2351587

[64] S. N. Evans, J. Pitman, and A. Winter, *Rayleigh processes, real trees, and root growth with re-grafting* Probability Theory and Related Fields **134**(1) (2006), 81–126. MR2221786

[65] W. Feller, *An introduction to probability theory and its applications* Volume 2, John Wiley & Sons (2008). MR0038583

[66] P. Flajolet, J.-C. Raoult, and J. Vuillemin, *The number of registers required for evaluating arithmetic expressions* Theoretical Computer Science **9**(1) (1979) 99–125. MR0535127

[67] P. Flajolet and H. Prodinger, *Register allocation for unary-binary trees* SIAM Journal on Computing **15** (3) (1986) 629–640. MR0850413

[68] U. Frisch, J. Bec, and B. Villone, *Singularities and the distribution of density in the Burgers/adhesion model* Physica D, **152/153** (2001) 620–635. MR1837931

[69] A. Gabrielov, W. I. Newman, and D. L. Turcotte, *Exactly soluble hierarchical clustering model: Inverse cascades, self-similarity, and scaling* Physical Review E, **60**(5) (1999) 5293. MR1719389

[70] C. Giraud, *Genealogy of shocks in Burgers turbulence with white noise initial velocity* Comm. Math.Phys., **223** (2001) 67–86. MR1860760

[71] C. Gu, A. Y. Schumann, M. Baiesi, and J. Davidsen, *Triggering cascades and statistical properties of aftershocks* Journal of Geophysical Research: Solid Earth, (2013) **118**(8) 4278–4295.

[72] V. K. Gupta, S. L. Castro, and T. M. Over, *On scaling exponents of spatial peak flows from rainfall and river network geometry* Journal of Hydrology, **187**(1–2) (1996) 81–104.

[73] V. K. Gupta, O. J. Mesa, and E. Waymire, *Tree-Dependent Extreme Values: The Exponential Case* Journal of Applied Probability, vol. **27** (1) (1990) 124–133. MR1039189

[74] V. K. Gupta, B. M. Troutman, and D. R. Dawdy *Towards a nonlinear geophysical theory of floods in river networks: an overview of 20 years of progress.* In Nonlinear Dynamics in Geosciences (pp. 121–151). Springer, New York, NY (2007).

[75] V. K. Gupta and E. D. Waymire *Statistical self-similarity in river networks parameterized by elevation.* Water Resources Research, **25**(3) (1989) 463–476.

[76] V. K. Gupta and E. Waymire, *Some mathematical aspects of rainfall, landforms and floods* In O. E. Barndorff-Nielsen, V. K. Gupta, V. Perez-Abreu, E. C. Waymire (eds) Rainfall, Landforms and Floods. Singapore: World Scientific (1998).

[77] S. Gurbatov, A. Malakhov, and A. Saichev, *Nonlinear random waves and turbulence in nondispersive media: waves, rays, particles* Manchester Uni-

versity Press, Manchester (1991). MR1255826

[78] J. T. Hack, *Studies of longitudinal stream profiles in Virginia and Maryland* US Government Printing Office, **294** (1957).

[79] H. Hackl, C. Heuberger, and H. Prodinger, *Reductions of binary trees and lattice paths induced by the register function* Theoretical Computer Science **705** (2018) 31–57. MR3721457

[80] T. E. Harris, *First passage and recurrence distribution* Trans. Amer. Math. Soc., **73** (1952) 471–486. MR0052057

[81] T. E. Harris *The theory of branching processes* Courier Corporation (2002). MR1991122

[82] T. Heckmann, W. Schwanghart, and J. D. Phillips, *Graph theory – Recent developments of its application in geomorphology* Geomorphology, **243** (2015) 130–146.

[83] D. Hilbert, *Mathematical problems* Göttinger Nachrichten, (1900) 253–297.

[84] J. R. Holliday, D. L. Turcotte, and J. B. Rundle, *Self-similar branching of aftershock sequences* Physica A: Statistical Mechanics and its Applications, (2008) **387**(4) 933–943. MR2359210

[85] R. E. Horton, *Erosional development of streams and their drainage basins: Hydrophysical approach to quantitative morphology* Geol. Soc. Am. Bull., **56** (1945) 275–370.

[86] A. Katok and B. Hasselblatt, *Introduction to the modern theory of dynamical systems* Vol. 54. Cambridge university press (1997). MR1326374

[87] G. S. Kassab, *The coronary vasculature and its reconstruction* Annals of Biomedical Engineering, (2000) **28**(8) 903–915.

[88] R. Kemp, *The average number of registers needed to evaluate a binary tree optimally* Acta Informatica, **11** (1979) 363–372. MR0533482

[89] J. F. C. Kingman, *The coalescent* Stoch. Process. Applic., **13**(3) (1982) 235–248. MR0671034

[90] J. W. Kirchner, *Statistical inevitability of Horton's laws and the apparent randomness of stream channel networks* Geology, **21**(7) (1993) 591–594.

[91] J. W. Kirchner, X. Feng, and C. Neal, *Fractal stream chemistry and its implications for contaminant transport in catchments* Nature, (2000) **403**(6769) 524.

[92] A. S. Kronrod, *On functions of two variables* Uspekhi matematicheskikh nauk, **5**(1) (1950) 24–134. MR0034826

[93] N. M. Krylov and N. N. Bogolyubov, *La theorie generalie de la mesure dans son application a l'etude de systemes dynamiques de la mecanique non-lineaire* Annals of Mathematics, **38** (1), (1937) 65–113. MR1503326

[94] V. F. Kolchin, *Moment of degeneration of a branching process and height of a random tree* Math. Notes, **6** (1978) 954–961. MR0458613

[95] Y. Kovchegov, P. T. Otto, and A. Yambartsev, *Cross-multiplicative coalescent processes and applications* submitted arXiv:1702.07764.

[96] Y. Kovchegov and I. Zaliapin, *Horton Law in Self-Similar Trees* Fractals, Vol. 24, No. 2 (2016) 1650017. MR3507614

[97] Y. Kovchegov and I. Zaliapin, *Horton self-similarity of Kingman's coa-*

lescent tree Ann. Inst. H. Poincaré (B) Probab. Statist., **53**(3) (2017) 1069–1107. MR3689961

[98] Y. Kovchegov and I. Zaliapin, *Tokunaga self-similarity arises naturally from time invariance* Chaos, **28**(4) (2018) 041102. MR3789870

[99] Y. Kovchegov and I. Zaliapin, *Random self-similar trees and a hierarchical branching process* Stochastic Processes and Their Applications, **129**(7), (2019) 2528–2560. MR3958441

[100] Y. Kovchegov and I. Zaliapin, *Dynamical pruning of rooted trees with applications to 1D ballistic annihilation* Journal of Statistical Physics, under revision arXiv:1707.01984.

[101] Y. Kovchegov and I. Zaliapin, *Invariance and attraction properties of Galton-Watson trees* submitted arXiv:1911.08095.

[102] P. L. Krapivsky, S. Redner, and E. Ben-Naim, *A kinetic view of statistical physics* Cambridge University Press (2010). MR2757286

[103] T. G. Kurtz, *Approximation of population processes* SIAM, **36** (1981). MR0610982

[104] K. S. Lau and C. R. Rao, *Integrated Cauchy functional equation and characterizations of the exponential law* Sankhyā: The Indian Journal of Statistics, Series **A** (1982) 1:72–90. MR0753078

[105] J. F. Le Gall, *The uniform random tree in a Brownian excursion* Probab. Theory Relat. Fields, **96** (1993) 369–383. MR1231930

[106] J. F. Le Gall and Y. Le Jan, *Branching processes in Lévy processes: The exploration process* Ann. Probab. **26** (1998) 213–252. MR1617047

[107] J. F. Le Gall and Y. Le Jan, *Branching processes in Lévy processes: Laplace functionals of snake and superprocesses* Ann. Probab. **26** (1998) 1407–1432. MR1675019

[108] J. F. Le Gall, *Spatial Branching Processes, Random Snakes and Partial Differential Equations* Springer Science & Business Media, (1999). MR1714707

[109] L. B. Leopold, M. G. Wolman, and J. P. Miller, *Fluvial Processes in Geomorphology* W. H. Freeman, San Francisco (1964).

[110] A. Lushnikov, *Coagulation in finite systems* J. Colloid and Interface Science, **65**, (1978) 276–285.

[111] R. Lyons and Y. Peres, *Probability on trees and networks* Vol. 42, Cambridge University Press (2017). MR3616205

[112] R. Mantilla, B. M. Troutman, and V. K. Gupta, *Testing statistical self-similarity in the topology of river networks* Journal of Geophysical Research: Earth Surface, (2010) **115**(F3).

[113] A. Marcus, *Stochastic coalescence* Technometrics, **10** (1968) 133–143. MR0223151

[114] A. Maritan, A. Rinaldo, R. Rigon, A. Giacometti, and I. Rodríguez-Iturbe, *Scaling laws for river networks* Physical Review E, (1996) **53**(2) 1510.

[115] J. G. Masek and D. L. Turcotte, *A Diffusion Limited Aggregation Model for the Evolution of Drainage Networks* Earth Planet. Sci. Let. **119** (1993) 379.

[116] M. McConnell and V. Gupta, *A proof of the Horton law of stream numbers*

for the Tokunaga model of river networks Fractals **16** (2008) 227–233.
MR2451617

[117] K. Menger, *Kurventheorie* Teubner, Leipzig, 1932. Google Scholar.

[118] M. M. Meerschaert and A. Sikorskii, *Stochastic models for fractional calculus* (Vol. 43). Walter de Gruyter Studies in Mathematics (2012). MR2884383

[119] O. J. Mesa, *Cuatro modelos de redes de drenaje* Revista de la Academia Colombiana de Ciencias Exactas, Físicas y Naturales, **42**(165), (2018) 379–391.

[120] G. Miermont, *Ordered additive coalescent and fragmentations associated to Lévy processes with no positive jumps* Electronic Journal of Probability, Vol. **6** (2001), paper no. 14, 1–33. MR1844511

[121] J. W. Milnor, M. Spivak, R. Wells, and R. Wells, *Morse theory* Princeton University Press (1963). MR0163331

[122] M. E. Nebel, *A unified approach to the analysis of Horton-Strahler parameters of binary tree structures* Random Structures & Algorithms, **21**(3–4) (2002) 252–277. MR1945370

[123] J. Neveu, *Erasing a branching tree* Advances in applied probability, **1** (1986) 101–108. MR0868511

[124] J. Neveu and J. Pitman, *Renewal property of the extrema and tree property of the excursion of a one-dimensional Brownian motion* Séminaire de Probabilités XXIII, **1372** of the series Lecture Notes in Mathematics, (1989) 239–247, Springer, Berlin. MR1022914

[125] J. Neveu and J. Pitman, *The branching process in a Brownian excursion* Séminaire de Probabilités XXIII, **1372** of the series Lecture Notes in Mathematics, (1989) 248–257 Springer, Berlin. MR1022915

[126] W. I. Newman, D. L. Turcotte, and A. M. Gabrielov, *Fractal trees with side branching* Fractals, **5** (1997) 603–614.

[127] L. Nicolaescu, *An invitation to Morse theory* (second edition). Springer Science & Business Media (2011). MR2883440

[128] J. R. Norris, *Smoluchowski's coagulation equation: uniqueness, nonuniqueness and a hydrodynamic limit for the stochastic coalescent* Ann. Appl. Probab. **9**, 1 (1999) 78–109. MR1682596

[129] P. Ossadnik, *Branch order and ramification analysis of large diffusion-limited-aggregation clusters* Physical Review A, (1992) **45**(2) 1058.

[130] M. Ossiander, E. Waymire, and Q. Zhang, *Some width function asymptotics for weighted trees* Ann. Appl. Probab. **7**, 4 (1997) 972–995. MR1484794

[131] S. D. Peckham, *New results for self-similar trees with applications to river networks* Water Resour. Res., **31** (1995) 1023–1029.

[132] S. D. Peckham and V. K. Gupta, *A reformulation of Horton's laws for large river networks in terms of statistical self-similarity.* Water Resources Research, **35**(9) (1999) 2763–2777.

[133] J. D. Pelletier and D. L. Turcotte, *Shapes of river networks and leaves: are they statistically similar?* Philosophical Transactions of the Royal Society of London B: Biological Sciences, **355**(1394) (2000) 307–311.

[134] J. Piasecki, *Ballistic annihilation in a one-dimensional fluid* Phys. Rev. E **51**(6) (1995) 5535–5540.

[135] J. Pitman, *Combinatorial Stochastic Processes: Ecole d'été de probabilités de Saint-Flour XXXII-2002* Lectures on Probability Theory and Statistics. Springer (2006).

[136] H. Prodinger, *Some recent results on the register function of a binary tree* Annals of Discrete Math **33** (1) (1987) 241–260. MR0930494

[137] D. Revuz and M. Yor, *Continuous martingales and Brownian motion* (3rd edition) Grundlehren der mathematischen Wissenschaften. Vol. **293**. Springer (2004). MR1083357

[138] F. Riesz, *Sur un Théoréme de Maximum de Mm. Hardy et Littlewood* Journal of the London Mathematical Society, **7**(1) (1932) 10–13. MR1574451

[139] R. Rigon, I. Rodriguez-Iturbe, A. Maritan, A. Giacometti, D. G. Tarboton, and A. Rinaldo, *On Hack's law* Water Resources Research, **32**(11) (1996) 3367–3374.

[140] I. Rodriguez-Iturbe, E. J. Ijjasz-Vasquez, R. L. Bras, and D. G. Tarboton, *Power law distributions of discharge mass and energy in river basins.* Water Resources Research, **28**(4) (1992) 1089–1093.

[141] I. Rodriguez-Iturbe and A. Rinaldo, *Fractal river basins: chance and self-organization* Cambridge University Press (2001).

[142] J. B. Rundle, D. L. Turcotte, R. Shcherbakov, W. Klein, and C. Sammis, *Statistical physics approach to understanding the multiscale dynamics of earthquake fault systems* Reviews of Geophysics, **41**(4), (2003).

[143] G. Samorodnitsky and M. S. Taqqu, *Stable non-Gaussian random processes: stochastic models with infinite variance* (Vol. 1). CRC press (1994). MR1280932

[144] C. E. Shannon, *A mathematical theory of communication* Bell System Technical Journal, **27**(3) (1948) 379–423. MR0026286

[145] Z.-S. She, E. Aurell, and U. Frisch, *The inviscid Burgers equation with initial data of Brownian type* Comm. Math. Phys., **148**(3) (1992), 623–641. MR1181072

[146] R. L. Shreve, *Statistical law of stream numbers* J. Geol., **74**(1) (1966) 17–37.

[147] R. L. Shreve, *Stream lengths and basin area in topologically random channel networks* J. Geol., **77**, (1969) 397–414.

[148] V. Sidoravicius and L. Tournier, *Note on a one-dimensional system of annihilating particles* Electron. Commun. Probab., **22**(59) (2017) 1–9. MR3718709

[149] Y. G. Sinai, *Statistics of shocks in solutions of inviscid Burgers equation* Comm. Math. Phys., **148**(3) (1992) 601–621. MR1181071

[150] M. Smoluchowski, *Drei Vorträge über Diffusion, Brownsche Molekularbewegung und Koagulation von Kolloidteilchen* Physik. Zeit., **17**, (1916) 557–571, 585–599.

[151] A. N. Strahler, *Quantitative analysis of watershed geomorphology* Trans. Am. Geophys. Un., **38** (1957) 913–920.

[152] T. Tao, *An introduction to measure theory* Providence, RI: American Mathematical Society (2011). MR2827917

[153] D. G. Tarboton, *Fractal river networks, Horton's laws and Tokunaga cyclicity* Journal of hydrology, **187**(1) (1996) 105–117.

[154] D. G. Tarboton, R. L. Bras, I. Rodriguez-Iturbe, *The fractal nature of river networks* Water Resour. Res., **24** (1988) 1317–1322.

[155] E. Tokunaga, *Consideration on the composition of drainage networks and their evolution* Geographical Rep. Tokyo Metro. Univ., **13** (1978) 1–27.

[156] D. L. Turcotte, *Fractals and chaos in geology and geophysics* Cambridge University Press (1997). MR1458893

[157] D. L. Turcotte, J. R. Holliday, and J. B. Rundle, *BASS, an alternative to ETAS* Geophysical Research Letters, **34**(12) (2007).

[158] D. L. Turcotte, B. D. Malamud, G. Morein, and W. I. Newman, *An inverse cascade model for self-organized critical behavior* Physica, A. **268** (1999) 629–643.

[159] D. L. Turcotte, J. D. Pelletier, and W. I. Newman, *Networks with side branching in biology* Journal of Theoretical Biology, **193**(4), (1998) 577–592.

[160] N. G. Ushakov, *Selected topics in characteristic functions* Walter de Gruyter (2011). MR1745554

[161] S. A. Veitzer, and 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) 1033–1048.

[162] X. G. Viennot, *Trees everywhere.* In CAAP'90 (pp. 18–41), Springer Berlin Heidelberg (1990).

[163] A. G. Vitushkin, *On Hilbert's thirteenth problem and related questions* Russian Mathematical Surveys, **59**(1) (2004) 11. MR2068840

[164] J. Von Neumann, *Invariant measures* American Mathematical Soc. (1998), originally published in 1941. MR1744399

[165] S. X. Wang and E. C. Waymire, *A large deviation rate and central limit theorem for Horton ratios* SIAM Journal on Discrete Mathematics, **4**(4), (1991) 575–588. MR1129394

[166] E. C. Waymire, *On the main channel length-magnitude formula for random networks: A solution to Moon's conjecture* Water Resour. Res., **25**, (1989) 1049–1050.

[167] D. Williams, *Path decomposition and continuity of local time for one-dimensional diffusions. I* Proc. London Math. Soc., **28**(3), (1974) 738–768. MR0350881

[168] G. Yakovlev, W. I. Newman, D. L. Turcotte, and A. Gabrielov, *An inverse cascade model for self-organized complexity and natural hazards* Geophysical Journal International, **163**(2), (2005) 433–442.

[169] K. Yamamoto, *Central limit theorem for the Horton-Strahler bifurcation ratio of general branch order* Journal of Applied Probability, **54**(4), (2017) 1111–1124. MR3731287

[170] I. Yekutieli and B. B. Mandelbrot, *Self-similarity of the branching structure in very large DLA clusters and other branching fractals* J. Phys. A:

Math. Gen., **27**(2) (1994) 275–284. MR1267412

[171] I. Yekutieli and B. B. Mandelbrot, *Horton-Strahler ordering of random binary trees* J. Phys. A: Math. Gen., **27**(2) (1994) 285–293. MR1267413

[172] M. R. Yoder, J. Van Aalsburg, D. L. Turcotte, S. G. Abaimov, and J. B. Rundle, *Statistical variability and Tokunaga branching of aftershock sequences utilizing BASS model simulations* Pure and Applied Geophysics, (2013) **170**(1–2) 155–171.

[173] I. Zaliapin and Y. Kovchegov, *Tokunaga and Horton self-similarity for level set trees of Markov chains* Chaos, Solitons & Fractals, **45**(3) (2012) 358–372. MR2881663

[174] I. Zaliapin and Y. Ben-Zion, *Earthquake clusters in southern California I: Identification and stability* Journal of Geophysical Research: Solid Earth, **118**(6), (2013) 2847–2864.

[175] I. Zaliapin, H. Wong, and A. Gabrielov, *Inverse cascade in a percolation model: Hierarchical description of time-dependent scaling* Physical Review E, **71**(6) (2005) 066118. MR2169012

[176] I. Zaliapin, H. Wong, and A. Gabrielov, *Hierarchical aggregation in percolation model* Tectonophysics **413** (2006) 93–107.

[177] I. Zaliapin, A. Gabrielov, V. Keilis-Borok, and H. Wong, *Clustering analysis of seismicity and aftershock identification* Physical Review Letters, (2008) **101**(1) 018501.

[178] S. Zanardo, I. Zaliapin, and 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.

Index