Submitted to Bernoulli

Invariance and attraction properties of Galton-Watson trees

YEVGENIY KOVCHEGOV[1]

[1] *Department of Mathematics, Oregon State University, Corvallis, OR 97331-4605*
E-mail: kovchegy@math.oregonstate.edu*;*
url: https://math.oregonstate.edu/people/view/kovchegy

ILYA ZALIAPIN[2]

[2] *Department of Mathematics and Statistics, University of Nevada, Reno, NV 89557-0084*
E-mail: zal@unr.edu*;*
url: https://wolfweb.unr.edu/homepage/zal/

We give a description of invariants and attractors of the critical and subcritical Galton-Watson tree measures under the operation of Horton pruning (cutting tree leaves with subsequent series reduction). Under a regularity condition, the class of invariant measures consists of the critical binary Galton-Watson tree and a one-parameter family of critical Galton-Watson trees with offspring distribution $\{q_k\}$ that has a power tail $q_k \sim Ck^{-(1+1/q_0)}$, where $q_0 \in (1/2, 1)$. Each invariant measure has a non-empty domain of attraction under consecutive Horton pruning, specified by the tail behavior of the initial Galton-Watson offspring distribution. The invariant measures satisfy the Toeplitz property for the Tokunaga coefficients and obey the Horton law with exponent $R = (1 - q_0)^{-1/q_0}$.

Keywords: Galton-Watson processes, self-similar trees, Horton-Strahler order, invariant measures, attractor.

MSC 2010: 60J80, 60C05

1. Introduction and motivation

The study of random trees invariant with respect to combinatorial pruning (erasure) from leaves down to the root emerges in attempts to understand symmetries of natural trees observed in fields as diverse as hydrology, phylogenetics, or computer science. In addition, it provides a unifying framework for analysis of coalescence and annihilation dynamical models, including the celebrated Kingman's coalescent, and self-similar stochastic processes on the real line; see a recent survey [13] for details. A special place in the invariance studies is occupied by the family of Galton-Watson trees, whose transparent generation mechanism makes it a convenient testbed for general theories and approaches. A Galton-Watson tree describes the trajectory of the Galton-Watson branching process [1] with a single progenitor and offspring distribution $\{q_k\}$, $k = 0, 1, \ldots$. We write $\mathcal{GW}(q_k)$ for

the probability measure that corresponds to this random tree. A tree is called *critical* if the expected progeny of a single member equals unity: $\sum_{k=1}^{\infty} kq_k = 1$. Similarly, a tree is *subcritical* if $\sum_{k=1}^{\infty} kq_k < 1$. In this paper we analyze the invariance and attraction properties of critical and subcritical Galton-Watson trees under the operation of *combinatorial Horton pruning* – cutting tree leaves and their parental edges followed by series reduction (removing vertices of degree 2). The Horton pruning (formally introduced in Sect. 2.2 and illustrated in Fig. 3) is a discrete, combinatorial analog of the continuous *erasure* or *trimming* studied by Neveu [16], Neveu and Pitman [17, 18], Le Jan [14], Evans [6], and Evans, Pitman and Winter [7].

1.1. Invariance

Combinatorial prune invariance of critical and subcritical Galton-Watson trees was first examined by Burd et al. [2], under the assumption of a finite second moment for the offspring distribution, $\sum_{k=1}^{\infty} k^2 q_k < \infty$. These authors have shown that the only invariant measure in this class corresponds to the critical binary Galton-Watson tree, $q_0 = q_2 = \frac{1}{2}$ [2, Thm. 3.9]. Here we substantially relax the regularity constraint on the offspring distribution; see Assumption 1 and Lem. 5 in Sect. 3.2. This reveals an abundance of prune-invariant measures with infinite second moment. Theorem 2 describes all such measures among the critical and subcritical Galton-Watson trees that satisfy Assumption 1. This infinite family of *Invariant Galton-Watson (IGW)* measures can be characterized by a single parameter – the probability $q_0 \in [1/2, 1)$ of having no offsprings. An individual distribution from this family is denoted by $\mathcal{IGW}(q_0)$; it is a critical distribution with the offspring generating function

$$Q(z) = \sum_{k=0}^{\infty} q_k z^k = z + q_0(1-z)^{1/q_0}.$$

The case $q_0 = 1/2$ with $Q(z) = (1+z^2)/2$ corresponds to the critical binary Galton-Watson tree $\mathcal{IGW}(1/2) = \mathcal{GW}(q_0 = q_2 = 1/2)$. Every invariant Galton-Watson measure $\mathcal{IGW}(q_0)$ with $q_0 \in (\frac{1}{2}, 1)$ corresponds to an unbounded offspring distribution of Zipf type with infinite second moment:

$$q_k \sim Ck^{-(1+1/q_0)} \quad \text{as} \quad k \to \infty.$$

1.2. Attraction

Burd et al. [2, Thm. 3.11] have shown that any critical Galton-Watson tree with a bounded offspring number (there 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 under the pruning. Our Thm. 3 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 that satisfy Assumption 1, with respect to the

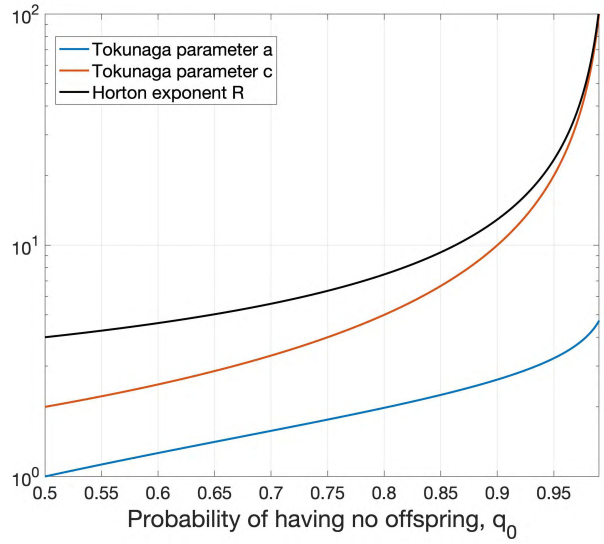

Figure 1: Tokunaga parameters a (blue), c (red) and Horton exponent R (black) in invariant Galton-Watson trees $\mathcal{IGW}(q_0)$ for $q_0 \in [0.5, 0.99]$.

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 characterized by the tail behavior of the offspring distribution $\{q_k\}$ of the initial Galton-Watson measure. In particular, Cor. 2 implies that every critical measure with Zipf tail $q_k \sim Ck^{-(1+1/q)}$ for $q \in [1/2, 1)$ and $C > 0$ converges to $\mathcal{IGW}(q)$. The subcritical attractor $\mathcal{GW}(q_0 = 1)$ is the limiting point of the IGW family for $q_0 = 1$ with generating function $Q(z) = z + (1 - z) = 1$. This distribution, however, is not prune-invariant.

Our results expand the attraction domain of the critical binary Galton-Watson tree $\mathcal{IGW}(1/2)$ initially described by Burd et al. [2]. Specifically, Lem. 2 shows that any critical offspring distribution that has an infinite second moment, satisfies Assumption 1, and has a finite $2-\epsilon$ moment for all $\epsilon > 0$ belongs to the attraction domain of $\mathcal{IGW}(1/2)$. We give an example of such a measure with $q_k \sim \frac{4}{3}k^{-3}$.

1.3. Toeplitz property

The results of Burd et al. [2] revealed an interesting characterization of the critical binary Galton-Watson distribution in terms of its *Tokunaga sequence*. Recall that the *Horton pruning* removes the leaf vertices and their parental edges from a finite tree T, with subsequent series reduction (removing degree-2 vertices). The *Horton order* of a tree T is the minimal number of Horton prunings sufficient to eliminate T. Informally, a branch of

Horton order k is a contiguous part of a tree (a collection of vertices and their parental edges in the initial tree) eliminated at k-th iteration of Horton pruning (see Figs. 3,4,5,6 for examples, and Sect. 2.2 for a formal definition). Each leaf (i.e., a leaf vertex with its parental edge) is a branch of order 1. Branches of higher orders may consist of lineages of vertices and their parental edges. The vertex farthest from the root is called the *terminal* vertex of a branch. Applied literature often examines the statistics of mergers of branches of distinct orders within a tree. Burd et al. [2] formalize this by considering the *Tokunaga coefficients* $T_{i,j}[T]$, for $i < j$, equal to the number of instances when a branch of order i joins a non-terminal vertex of the leftmost branch of order j closest to the root within T, given that the tree order is greater than j. This definition is suitable for describing a generic branch structure within a Galton-Watson tree, given its symmetric iterative generation mechanism. It has been shown [2, Thm. 3.16] that the critical binary Galton-Watson distribution $\mathcal{GW}(q_0 = q_2 = 1/2)$ is characterized, among the bounded offspring distributions, by the *Toeplitz property*:

$$\mathsf{E}\big[T_{i,j}[T]\big] = T_{j-i} \quad \text{for a positive Tokunaga sequence } \{T_k\}_{k=1,2,\dots}. \tag{1}$$

Specifically, the critical binary Galton-Watson distribution corresponds to $T_k = 2^{k-1}$. In Lem. 10, we show that all the invariant measures $\mathcal{IGW}(q_0)$ satisfy the Toeplitz property. In this analysis, we adopt an alternative, more general, definition of the Tokunaga coefficient $T_{i,j}$, which (i) accounts for branching at the terminal vertices, and (ii) can be applied to general (non Galton-Watson) trees. In our definition, the invariant measure $\mathcal{IGW}(q_0)$ corresponds to the Tokunaga sequence (Lem. 10)

$$T_1 = c^{c/(c-1)} - c - 1, \quad T_k = ac^{k-1}, \quad k = 2, 3, \dots$$

with (Fig. 1)

$$c = (1 - q_0)^{-1} \quad \text{and} \quad a = (c-1)(c^{1/(c-1)} - 1).$$

The critical binary Galton-Watson case with $q_0 = 1/2$ corresponds to $c = 2$ and $a = 1$, which reconstructs the Burd et al. [2] result $T_k = 2^{k-1}$. Moreover, using the Tokunaga sequence definition from Burd et al. [2], we obtain a particularly simple Tokunaga sequence $T_k = c^{k-1}$ for $k = 1, 2, \dots$.

1.4. Horton law

A ubiquitous empirical observation in the analysis of dendritic structures is the *Horton law* [9, 13, 20]. Informally, the law states that the numbers $N_k[T]$ of branches of order k in a large tree T decays geometrically:

$$\frac{N_k[T]}{N_{k+1}[T]} \approx R$$

for some *Horton exponent* $R \geq 2$. A formal definition of the Horton law for tree measures is given in Sect. 2.5.

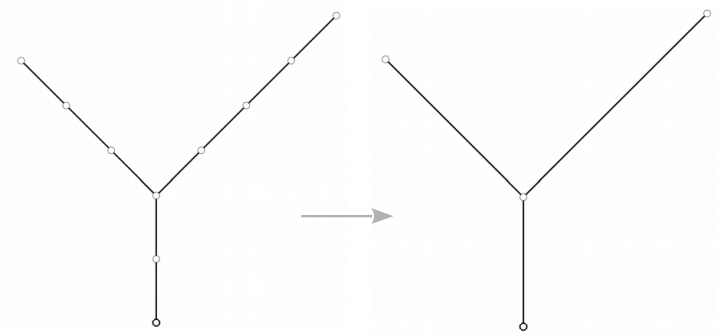

Figure 2: Series reduction: Example. Tree T before (a) and after (b) series reduction.

It has been shown by McConnell and Gupta [15] for a particular case of $T_k = ac^{k-1}$ with $a \geq 0$, $c > 0$, and generalized by the authors of this paper [11] to an arbitrary Tokunaga sequence $\{T_k\}$, that the Toeplitz property implies the Horton law. Lemma 10 shows that the invariant Galton-Watson measure $\mathcal{IGW}(q_0)$ for any $q_0 \in [1/2, 1)$ obeys the Horton law with the Horton exponent $R = (1 - q_0)^{-1/q_0}$ (Fig. 1).

2. Preliminaries

2.1. Galton-Watson tree measures

Consider the space $\mathcal{T}$ of finite unlabeled rooted reduced trees. A tree is called *rooted* if one of its vertices, denoted by ρ, is selected as the tree root. 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*. 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 a parent. Let $\mathcal{T}^{|}$ denote a subspace of *planted trees* in $\mathcal{T}$; it contains ϕ and all the trees in $\mathcal{T}$ with the root vertex having exactly one offspring (see Figs. 2,3). The degree of the root equals the number of its offsprings. The degree of a non-root vertex is the number of its offsprings plus one (to account for the parent). The number of the offsprings at a vertex is called the vertex *branching number*. A tree from $\mathcal{T}^{|}$ is called reduced is it has no vertices of degree 2.

For a given offspring distribution $\{q_k\}_{k=0,1,2,\ldots}$, we let $\mathcal{GW}(\{q_k\})$ denote the corresponding Galton-Watson tree measure. We assume that each tree begins with a single root vertex which produces a single offspring, so the resulting trees are in $\mathcal{T}^{|}$. In this renowned Markov chain construction, each non-root vertex produces k offsprings with probability q_k, independently of other vertices. We assume $\sum_{k=0}^{\infty} kq_k \leq 1$ and $q_1 = 0$ as we need $\mathcal{GW}(\{q_k\})$ to be a probability measure on $\mathcal{T}^{|}$ (the trees in $\mathcal{T}^{|}$ are required to be

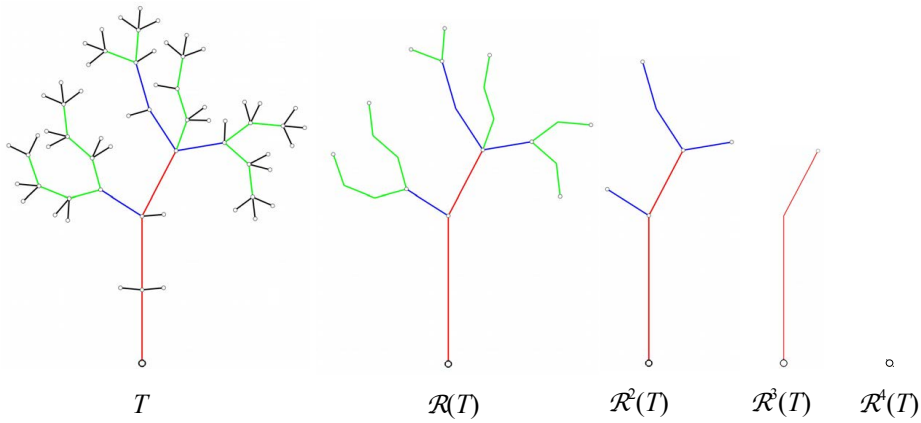

Figure 3: The Horton-Strahler orders: Example. Consecutive prunings $\mathcal{R}^k(T)$, $k = 0, 1, \ldots, 4$, of tree T. The order of tree is $\mathsf{ord}(T) = 4$ since $\mathcal{R}^4(T) = \phi$. Different colors depict branches of different orders: $\mathsf{ord} = 1$ (black), $\mathsf{ord} = 2$ (green), $\mathsf{ord} = 3$ (blue), and $\mathsf{ord} = 4$ (red).

finite and reduced). The assumption of subcriticality or criticality implies $q_0 \geq \frac{1}{2}$, since

$$1 \geq \sum_{k=2}^{\infty} k q_k \geq 2 \sum_{k=2}^{\infty} q_k \ = 2(1 - q_0).$$

2.2. Horton pruning, orders

Recall that *series reduction* on a tree T removes each vertex of degree 2 and merges its two adjacent edges into one (Fig. 2).

Definition 1 (Horton pruning). *Horton pruning $\mathcal{R} : \mathcal{T} \to \mathcal{T}$ 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 $\mathcal{R}(\phi) = \phi$.*

The trajectory of each tree T under $\mathcal{R}(\cdot)$ is uniquely determined and finite:

$$T \equiv \mathcal{R}^0(T) \to \mathcal{R}^1(T) \to \cdots \to \mathcal{R}^k(T) = \phi, \tag{2}$$

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

It is natural to think of the distance to ϕ under the Horton pruning map and introduce the respective notion of tree order [9, 13, 20, 21].

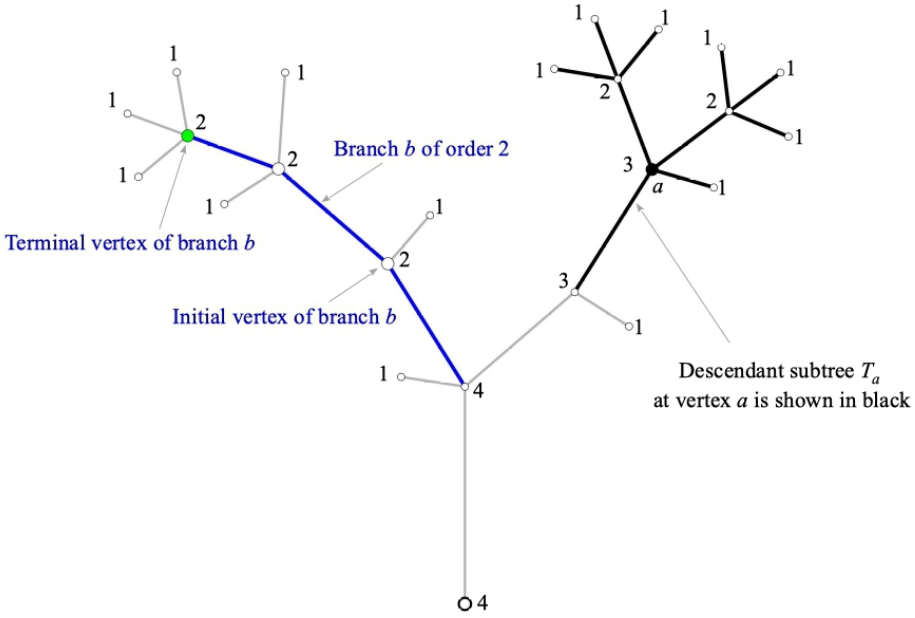

Figure 4: Illustration of the Horton-Strahler terminology (Def. 3). A branch b of oder 2 is shown in blue in the left part of the figure. The branch consists of three vertices of order 2 and their parental edges. The terminal vertex of branch b is shown by green circle. The descendant subtree T_a at vertex a is shown in black in the right part of the figure. The Horton-Strahler orders are shown next to the vertices.

Definition 2 (**Horton-Strahler order**). *The Horton-Strahler order* $\mathsf{ord}(T) \in \mathbb{Z}_+$ *of a tree* $T \in \mathcal{T}^{|}$ *is defined as the minimal number of Horton prunings necessary to eliminate the tree:*

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

In particular, the order of the empty tree is $\mathsf{ord}(\phi) = 0$, because $\mathcal{R}^0(\phi) = \phi$. This definition is illustrated in Fig. 3 for a tree T with $\mathsf{ord}(T) = 4$. In this paper we consider probability measures $\mathcal{GW}(\{q_k\})$ on $\mathcal{T}^{|}$ that satisfy $\sum_{k=0}^{\infty} kq_k \leq 1$ and $q_1 = 0$ and assign probability zero to the empty tree ϕ.

Definition 3 (**Horton-Strahler terminology**). *We introduce the following definitions related to the Horton-Strahler order of a tree (see Fig. 4):*

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. Figure 4 shows in black color the descendant subtree T_a at vertex a.

2. **(Vertex order)** *For any vertex $v \in T \setminus \{\rho\}$ we set $\mathsf{ord}(v) = \mathsf{ord}(T_v)$. 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. Figure 4 shows a branch b of order 2 (blue) that consists of three vertices and their parental edges. 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. Figure 4 shows the terminal vertex of branch b (blue) as a green circle.*

The Horton-Strahler orders can be equivalently defined via hierarchical counting [9, 21, 3, 20, 19]. The first such definition beyond the binary case appeared in [2]. In 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

$$\mathsf{ord}(p) = \begin{cases} r & \text{if } \#\{s:\ i_s = r\} = 1, \\ r+1 & \text{otherwise.} \end{cases} \tag{3}$$

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.

Figures 5,6 illustrate Horton-Strahler orders in trees with a constant branching number b ($q_0 + q_b = 1$) and with a bounded offspring distribution ($q_k = 0$ for $k > b$), repsectively.

2.3. Horton self-similarity

Here we define self-similarity of a Galton-Watson measure with respect to the Horton pruning $\mathcal{R}$, which is the main operation on trees discussed in this work.

Definition 4 (**Horton self-similarity**). *Consider a Galton-Watson measure μ on $\mathcal{T}$ (or $\mathcal{T}^{|}$) such that $\mu(\phi) = 0$. Let ν be the pushforward measure, $\nu = \mathcal{R}_*(\mu)$, i.e.,*

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

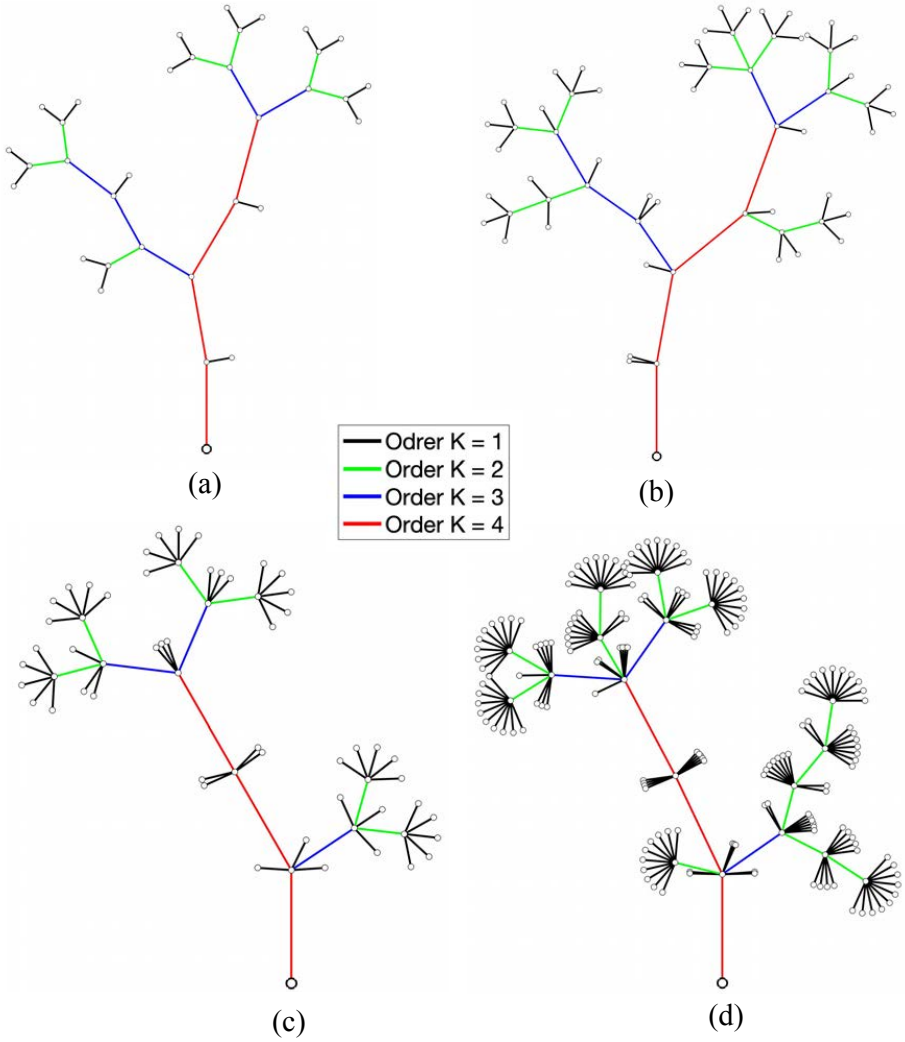

Figure 5: Examples of Horton-Strahler ordering in trees with constant branching number b ($q_0 + q_b = 1$). (a) $b = 2$, (b) $b = 3$, (c) $b = 5$, (d) $b = 10$. Each panel shows a tree of order $\mathsf{ord} = 4$. Edges of different orders are shown in different colors, as indicated in the legend.

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

$$\nu\left(T\,|T \neq \phi\right) = \mu(T). \tag{4}$$

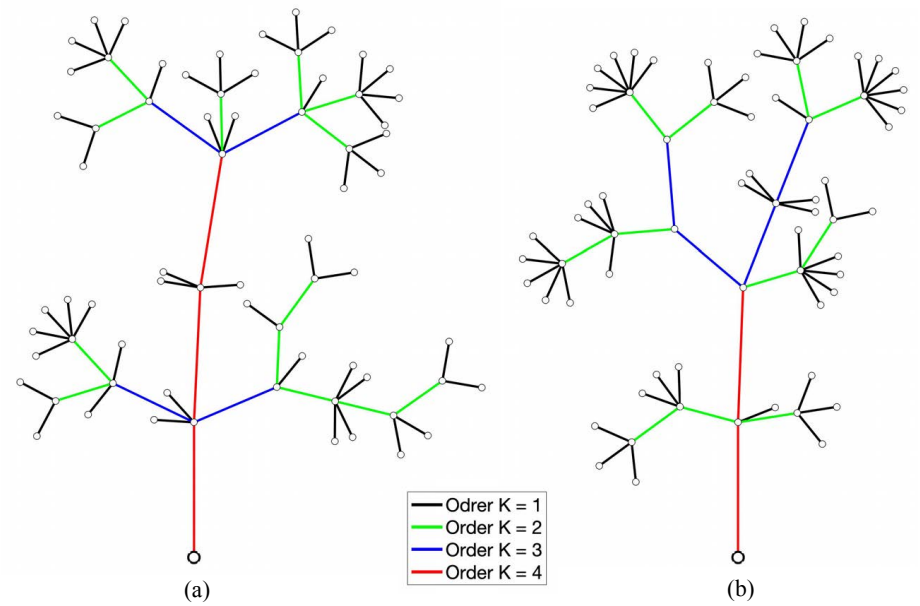

Figure 6: Examples of Horton-Strahler ordering in trees with bounded offspring distribution: $q_k = 0$ for $k > b$. (a) $b = 5$, (b) $b = 6$. Each panel shows a tree of order $\mathsf{ord} = 4$. Edges of different orders are shown in different colors, as indicated in the legend.

Definition 4 does not distinguish between prune-invariance and self-similarity. Such equivalence is a particular property of Galton-Watson measures connected to their Markov structure. In a general case, prune-invariance happens to be a weak property that allows a multitude of obscure measures. A general prune-invariant measure on $\mathcal{T}$ has to satisfy an additional property, called *coordination*, to be called self-similar. The Galton-Watson measures always satisfy the coordination property; see (10). We refer to [13] for a comprehensive discussion and examples.

2.4. Tokunaga coefficients and Toeplitz property

This section introduces *Tokunaga coefficients* that describe mergers of branches of different orders in a random tree. Empirically, a Tokunaga coefficient $T_{i,j}$ is the average number of branches of order i that merge a branch of order j within a tree T. The Markovian generation process ensures that all branches of a given order j in a Galton-Watson tree have the same probabilistic structure. Hence, one can follow Burd et al. [2] and define $T_{i,j}$ as the mean number of order i branches within a particular branch of order j, for instance – the leftmost branch closest to the root. We introduce below a more general definition, which is equivalent to that of Burd et al. [2] for Galton-Watson trees, and

can extend to non Markovian branching processes. This set up will also be needed to formulate the Horton law in Sect. 2.5.

Consider a measure μ on $\mathcal{T}$ (or $\mathcal{T}^{|}$) such that $\mu(\phi) = 0$. The Horton pruning partitions the underlying tree space 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 root and a leaf connected to the root by its parental edge, and all other subspaces $\mathcal{H}_k$, $k \geq 2$, consist of an infinite number of trees. Naturally, $\mathcal{H}_k \bigcap \mathcal{H}_{k'} = \emptyset$ if $k \neq k'$, and $\bigcup\limits_{k\geq 1} \mathcal{H}_k = \mathcal{T}$ (or $\mathcal{T}^{|}$). Consider a set of conditional probability measures $\{\mu_k\}_{k\geq 0}$ each of which is defined on $\mathcal{H}_k$. Specifically, we set $\mu_k(\cdot) \equiv 0$ for any k such that $\mu(\mathcal{H}_k) = 0$ and

$$\mu_k(T) = \mu(T\,|T \in \mathcal{H}_k) \tag{5}$$

otherwise. Letting $\pi_k = \mu(\mathcal{H}_k)$, the measure μ can be represented as a mixture of the conditional measures:

$$\mu = \sum_{k=1}^{\infty} \pi_k \mu_k. \tag{6}$$

Let $N_k = N_k[T]$ be the number of branches of order k in a tree T. For given integers $1 \leq i < j$, let $n_{i,j} = n_{i,j}[T]$ denote the total number of vertices of order i that have parent of order j in a tree $T \in \mathcal{T}$ (or $\mathcal{T}^{|}$). We write $\mathsf{E}_K[\cdot]$ for the expectation with respect to μ_K of Eq. (5).

We define the *average Horton numbers* for subspace $\mathcal{H}_K$ as

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

For subspace $\mathcal{H}_K$, let

$$t_{i,j}[K] = \frac{\mathsf{E}_K[n_{i,j}]}{\mathsf{E}_K[N_j]} = \frac{\mathsf{E}_K[n_{i,j}]}{\mathcal{N}_j[K]}, \quad 1 \leq i < j \leq K, \tag{7}$$

be the *total Tokunaga merger statistics* that is used to define the *Tokunaga coefficients*

$$T_{i,j}[K] = t_{i,j}[K] - 2\delta_{i,j-1} \qquad (1 \leq i < j). \tag{8}$$

Remark 1. *Recall that a branch of order j is formed by a merger of two or more branches of order $j-1$. We designate two arbitrarily selected branches of order $j-1$ that descend from the terminal vertex of a branch of order j as principle branches. The existence of such two branches follows from the definition of the Horton order (Def. 2). The other branches (if any) of order $i \leq j-1$ that descend from any vertex, including the terminal vertex, in a branch of order j are said to be side branches of Tokunaga index $\{i,j\}$. The Tokunaga coefficients $T_{i,j}[K]$ are intended to count the number of side branches of Tokunaga index $\{i,j\}$, which explains the need to subtract $2\delta_{i,j-1}$ in* (8).

Finally, let $n^o_{i,j}$ denote the total number of vertices of order i whose parent vertices are non-terminal vertices of order j. Then,

$$T^o_{i,j}[K] = \frac{\mathsf{E}_K[n^o_{i,j}]}{\mathsf{E}_K[N_j]} = \frac{\mathsf{E}_K[n^o_{i,j}]}{\mathcal{N}_j[K]} \qquad (1 \le i < j) \tag{9}$$

are called the *regular Tokunaga coefficients.*

Remark 2. *The quantities $\mathcal{N}_k[K]$, $t_{i,j}[K]$, $T_{i,j}[K]$, and $T^o_{i,j}[K]$ depend on the measure μ. We skip this dependence in our notations.*

We observe that for a subcritical or critical Galton-Watson measure μ we have the following *coordination* property [13]:

$$T_{i,j} := T_{i,j}[K] \quad \text{ for all } K \ge 2 \text{ and } 1 \le i < j \le K. \tag{10}$$

This is explained as follows. Consider all nodes in generation $d \in \mathbb{N}$ (which may be an empty set) in a critical or subcritical Galton-Watson tree T. The descendant subtrees T_v (see Def. 3) for v in generation d are independently distributed according to μ. Sampling of T_v can be split into two steps, first selecting its order with probability distribution π_j, next sampling the tree of order j according to the probability measure μ_j. The branching history $\mathcal{F}_d$ up to generation d together with the orders of the descendant subtrees T_v with v in generation d completely determines (i) the order of the tree T, and (ii) whether or not v is the initial vertex (Def. 3) of the corresponding branch of order $\mathsf{ord}(T_v)$. At the same time, conditioned on $\mathcal{F}_d$ and the orders $\mathsf{ord}(T_v)$ for v in generation d, each T_v is independently distributed according to μ_j, where $\mathsf{ord}(T_v) = j$.

The respective Tokunaga matrix $\mathbb{T}_K$ is a $K \times K$ matrix

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

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

Definition 5 (Toeplitz property). *A Galton-Watson measure μ is said to satisfy the Toeplitz property if there exists a sequence $T_k \ge 0$, $k = 1, 2, \dots$ such that*

$$T_{i,j} = T_{j-i}. \tag{11}$$

The elements of the sequences T_k are also referred to as Tokunaga coefficients, which does not create confusion with $T_{i,j}$.

For a Galton-Watson measure that satisfies the Toeplitz property, the corresponding Tokunaga matrices $\mathbb{T}_K$ are Toeplitz[1]:

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

The following statement has been proven in [13] for (not necessarily Galton-Watson) binary trees; the argument applies verbatim to general Galton-Watson trees.

Proposition 1 (**Prune-invariance implies Toeplitz**). *Suppose a Galton-Watson measure μ is Horton prune-invariant, then it satisfies the Toeplitz property (Def. 5).*

Definition 6 (**Tokunaga self-similarity**). *A Galton-Watson measure μ on $\mathcal{T}$ is called Tokunaga self-similar with parameters (a, c) if it satisfies the Toeplitz property and its Tokunaga sequence $\{T_j\}_{j=1,2,\dots}$ is expressed as*

$$T_j = a\, c^{j-1}, \quad j \geq 1 \tag{12}$$

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

2.5. Horton law

Consider a measure μ on $\mathcal{T}$ (or $\mathcal{T}^{|}$) and its conditional measures μ_K, each defined on subspace $\mathcal{H}_K \subset \mathcal{T}$ of trees of Horton-Strahler order $K \geq 1$ as discussed in Sect. 2.4. We write $T \overset{d}{\sim} \mu_K$ for a random tree T drawn from a subspace $\mathcal{H}_K$ ($\mu(\mathcal{H}_K) > 0$) according to measure μ_K.

Definition 7 (**Strong Horton law for mean branch numbers**). *We say that a probability measure μ on $\mathcal{T}$ (or $\mathcal{T}^{|}$) satisfies a strong Horton law for mean branch numbers if there exists such a positive (constant) Horton exponent $R \geq 2$ that for any $k \geq 1$,*

$$\lim_{K\to\infty} \frac{\mathcal{N}_k[K]}{\mathcal{N}_1[K]} = R^{1-k}. \tag{13}$$

Here, the adjective *strong* refers to the type of geometric convergence; see [13] for details.

The work [11] establishes the strong Horton law in a binary tree that satisfies the Toeplitz property (Def. 5). We observe that the results of [11] hold beyond the binary

[1]Note that in [2], the Tokunaga sequence was set to satisfy $T^o_{i,j} = T^o_{i,j}[K] = T_{j-i}$. That is, the offsprings adjacent to the terminal vertex of order j branch were not counted.

case, as the derivation steps are identical. Specifically, assume the Toeplitz property with a Tokunaga sequence $\{T_k\}$ and consider a sequence $t(k)$ defined by

$$t(0) = -1, \quad \text{and} \quad t(k) = T_k + 2\delta_{1,k} \quad \text{for } k \geq 1.$$

Observe that $t_{i,j} = t_{i,j}[K] = t(j-i)$. The generating function of $t(k)$ is

$$\hat{t}(z) = \sum_{k=0}^{\infty} z^k t(k) = -1 + 2z + \sum_{k=1}^{\infty} z^k T_k.$$

Theorem 1 (Strong Horton law in a mean self-similar tree, [11]). *Suppose μ is a Galton-Watson measure on $\mathcal{T}^{|}$ that satisfies the Toeplitz property with Tokunaga sequence $\{T_j\}_{j=1,2,...}$ such that*

$$\limsup_{j\to\infty} T_j^{1/j} < \infty. \tag{14}$$

Then the strong Horton law for mean branch numbers (Def. 7) 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,

$$\lim_{K\to\infty} \left(\mathcal{N}_1[K]\, R^{-K}\right) = const. > 0. \tag{15}$$

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.

3. Main results

3.1. Distribution of Horton orders and related functions

Consider a collection of critical or subcritical Galton-Watson measures $\mathcal{GW}(\{q_k\})$ with $q_1 = 0$ on $\mathcal{T}^{|}$. Let $Q(z) = \sum_{m=0}^{\infty} z^m q_m$ for $z \in [0,1]$ be the generating function of $\{q_k\}$. For $T \overset{d}{\sim} \mathcal{GW}(\{q_k\})$ we denote $\pi_j := P\big(\mathsf{ord}(T) = j\big)$. Finally, let $\sigma_0 = 0$ and $\sigma_j := \sum_{i=1}^{j} \pi_i$ $(j \geq 1)$.

Lemma 1 (Order distribution). *Consider a Galton-Watson measure $\mathcal{GW}(\{q_k\})$ with $q_1 = 0$. Assume criticality or subcriticality, i.e., $\sum_{k=0}^{\infty} k q_k \leq 1$. Then,*

$$\pi_1 = q_0 \quad \textit{and} \quad \pi_j = \frac{Q(\sigma_{j-1}) - Q(\sigma_{j-2}) - \pi_{j-1} Q'(\sigma_{j-2})}{1 - Q'(\sigma_{j-1})} \quad (j \geq 2). \tag{16}$$

Proof. The probability of tree T with a single leaf is $\pi_1 = P\big(\mathsf{ord}(T) = 1\big) = q_0$.

Next we find the Horton-Strahler order of the offspring of the root using the rule (3). The probability that the offspring of the root is a terminal vertex of a branch of order j, $j \geq 2$, is

$$\sum_{m=2}^{\infty} q_m \sum_{\ell=2}^{m} \binom{m}{\ell} \pi_{j-1}^{\ell} \sigma_{j-2}^{m-\ell} = Q(\sigma_{j-1}) - Q(\sigma_{j-2}) - \pi_{j-1} Q'(\sigma_{j-2}).$$

Here we take a sum over all possible numbers $m \geq 2$ of offsprings, and calculate the probability that $\ell \geq 2$ of the offsprings have order $j-1$, while the other $m-\ell$ offsprings have orders less than $j-1$.

Similarly, the probability of the offspring of the root to be a regular (non-terminal) vertex of order j equals

$$\sum_{m=2}^{\infty} q_m m \pi_j \sigma_{j-1}^{m-1} = \pi_j Q'(\sigma_{j-1}). \tag{17}$$

Therefore,

$$\pi_j = \pi_j Q'(\sigma_{j-1}) + \Big(Q(\sigma_{j-1}) - Q(\sigma_{j-2}) - \pi_{j-1} Q'(\sigma_{j-2})\Big),$$

which implies (16). □

Corollary 1. *Consider a Galton-Watson measure $\mathcal{GW}(\{q_k\})$ with $q_1 = 0$. Assume criticality or subcriticality, i.e., $\sum_{k=0}^{\infty} k q_k \leq 1$. Then, σ_j can be expressed via an iterated function (Fig. 7)*

$$\sigma_j = \underbrace{S \circ \ldots \circ S}_{j \text{ times}}(0) \qquad \text{for } \ j \geq 1, \tag{18}$$

where

$$S(z) = \frac{Q(z) - zQ'(z)}{1 - Q'(z)}. \tag{19}$$

Proof. Equation (16) implies

$$\pi_j = \big[Q(\sigma_{j-1}) + \pi_j Q'(\sigma_{j-1})\big] - \big[Q(\sigma_{j-2}) + \pi_{j-1} Q'(\sigma_{j-2})\big] \qquad \text{for } \ j \geq 2. \tag{20}$$

Hence, summing up the terms in (20), and substituting $\pi_1 = q_0$, we obtain

$$\sigma_j = \sum_{i=1}^{j} \pi_i = Q(\sigma_{j-1}) + \pi_j Q'(\sigma_{j-1}) = Q(\sigma_{j-1}) + (\sigma_j - \sigma_{j-1}) Q'(\sigma_{j-1})$$

for all $j \geq 1$. Thus, $\sigma_j = \frac{Q(\sigma_{j-1}) - \sigma_{j-1} Q'(\sigma_{j-1})}{1 - Q'(\sigma_{j-1})} = S(\sigma_{j-1})$. □

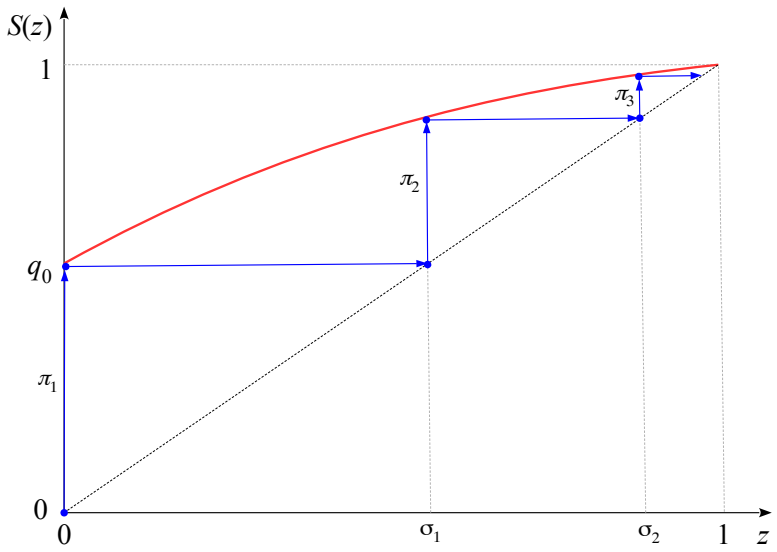

Figure 7: Illustration to Cor. 1. Function $S(z)$ is shown in red. Equation (18) implies that the values of σ_j are obtained by iterative application of $S(t)$, starting with $\sigma_0 = 0$. These iterations are illustrated by blue lines with arrows. Vertical increments correspond to the values of π_j.

Set $S(1) = \lim\limits_{x\to 1-} \frac{Q(x)-xQ'(x)}{1-Q'(x)}$. Then, by L'Hôpital's rule, $S(1) = \lim\limits_{x\to 1-} \frac{xQ''(x)}{Q''(x)} = 1$. Next, for the progeny variable $X \overset{d}{\sim} \{q_k\}$, consider the following important function

$$g(x) = \sum_{m=0}^{\infty} \mathsf{E}\big[(X-m-1)_+\big]\, x^m \;= \sum_{m=0}^{\infty} \sum_{k=m+1}^{\infty} (k-m-1) q_k\, x^m, \tag{21}$$

where $x_+ = \max\{x, 0\}$.

Proposition 2. *For a critical (i.e., $Q'(1) = 1$) Galton-Watson process $\mathcal{GW}(\{q_k\})$ with $q_1 = 0$, we have*

$$Q(x) - x = (1-x)^2 g(x)$$

for $g(x)$ as defined in (21).

Proof. Since $\sum_{k=2}^{\infty} kq_k = Q'(1) = 1$,

$$\begin{aligned}
Q(x) - x = q_0 + \sum_{k=2}^{\infty} q_k x^k \ - q_0 x - \sum_{k=2}^{\infty} q_k x \ &= (1-x)\left[q_0 - \sum_{k=2}^{\infty} q_k \frac{1-x^{k-1}}{1-x} x\right] \\
&= (1-x)\left[\sum_{k=2}^{\infty} kq_k - \sum_{k=2}^{\infty} q_k - \sum_{k=2}^{\infty} q_k \frac{1-x^{k-1}}{1-x} x\right] = (1-x)\sum_{k=2}^{\infty} q_k \left(k - 1 - \sum_{j=1}^{k-1} x^j\right) \\
&= (1-x)\sum_{k=2}^{\infty} q_k \left(\sum_{j=1}^{k-1}(1-x^j)\right) = (1-x)^2 \sum_{k=2}^{\infty} q_k \sum_{j=1}^{k-1}\sum_{m=0}^{j-1} x^m \\
&= (1-x)^2 \sum_{k=2}^{\infty} q_k \sum_{m=0}^{k-2}(k-m-1)x^m = (1-x)^2 \sum_{m=0}^{\infty}\sum_{k=m+2}^{\infty}(k-m-1)q_k x^m \\
&= (1-x)^2 g(x). \qquad (22)
\end{aligned}$$

□

Let L denote the limit $\lim_{x\to 1-}\left(\frac{\ln g(x)}{-\ln(1-x)}\right)$ whenever the limit exists.

Lemma 2. *For the progeny variable $X \overset{d}{\sim} \{q_k\}$ and $g(x)$ as defined in (21), if*

$$\mathsf{E}[X^{2-\epsilon}] = \sum_{k=0}^{\infty} k^{2-\epsilon} q_k \ < \infty \qquad \forall \epsilon > 0, \qquad (23)$$

then $L = \lim_{x\to 1-}\left(\frac{\ln g(x)}{-\ln(1-x)}\right) = 0$.

Proof. Suppose (23) holds, then by the Dominated Convergence Theorem, as $m \to \infty$,

$$(m+1)^{1-\epsilon}\,\mathsf{E}\big[(X-m-1)_+\big] \le \mathsf{E}\big[X^{1-\epsilon}(X-m-1)_+\big] \le \mathsf{E}\big[X^{2-\epsilon}\,\mathbf{1}_{\{X\ge m+1\}}\big] \ \to 0. \quad (24)$$

Accordingly, the m-th coefficient $\mathsf{E}\big[(X-m-1)_+\big]$ in the power series representation (21) of $g(x)$ is $o\left(m^{\epsilon-1}\right)$. Next, for $\epsilon > 0$, the m-th coefficient of the power series expansion of $(1-x)^{-\epsilon}$ is

$$\frac{\prod_{i=0}^{m-1}(\epsilon+i)}{m!} = \frac{\Gamma(\epsilon+m)}{\Gamma(\epsilon)m!} \sim \frac{m^{\epsilon-1}}{\Gamma(\epsilon)}, \quad m \to \infty. \qquad (25)$$

Together, (24) and (25) imply

$$\limsup_{x\to 1-} \frac{\ln g(x)}{\ln(1-x)^{-\epsilon}} \le 1 \quad \Leftrightarrow \quad \limsup_{x\to 1-} \frac{\ln g(x)}{-\ln(1-x)} \le \epsilon \qquad \forall \epsilon > 0.$$

Hence, $\limsup_{x\to 1-} \frac{\ln g(x)}{-\ln(1-x)} = 0$, while obviously $\liminf_{x\to 1-} \frac{\ln g(x)}{-\ln(1-x)} \ge 0$. □

3.2. Regularity condition

Many of the results of the paper are proved under the following assumption.

Assumption 1. *The following limit exists:*

$$S'(1) = \lim_{x\to 1-} \frac{1-S(x)}{1-x}. \tag{26}$$

Observe that since $S(x) - x = \frac{Q(x)-x}{1-Q'(x)}$, Assumption 1 is equivalent to the existence of the limit

$$\lim_{x\to 1-} \frac{Q(x)-x}{(1-x)\big(1-Q'(x)\big)} = 1 - S'(1). \tag{27}$$

Lemma 3. *Consider a critical Galton-Watson measure $\mathcal{GW}(\{q_k\})$ with $q_1 = 0$. If Assumption 1 is satisfied, then for $g(x)$ defined in (21) the following limit exists*

$$\lim_{x\to 1-} \left(\frac{\ln g(x)}{-\ln(1-x)} \right) = L, \tag{28}$$

and $S'(1) = \frac{1-L}{2-L}$.

Proof. By the L'Hôpital's rule,

$$\begin{aligned} L &= \lim_{x\to 1-} \left(\frac{\ln g(x)}{-\ln(1-x)} \right) = 2 - \lim_{x\to 1-} \frac{\ln g(x) + 2\ln(1-x)}{\ln(1-x)} = 2 - \lim_{x\to 1-} \frac{\ln\big(Q(x)-x\big)}{\ln(1-x)} \\ &= 2 - \lim_{x\to 1-} \frac{\frac{d}{dx}\ln\big(Q(x)-x\big)}{\frac{d}{dx}\ln(1-x)} = 2 - \lim_{x\to 1-} \frac{(1-x)\big(1-Q'(x)\big)}{Q(x)-x} = 2 - \lim_{x\to 1-} \frac{1-x}{S(x)-x} \\ &= 2 - \frac{1}{1-S'(1)}. \end{aligned}$$

□

Remark 3. *Notice that due to the conditions for the L'Hôpital's rule, there are cases when the limit in (28) exists while the limit in (26) does not exists. Indeed, the L'Hôpital's rule in (28) holds under the condition that the limit $\lim_{x\to 1-} \frac{\frac{d}{dx}\ln(Q(x)-x)}{\frac{d}{dx}\ln(1-x)}$ exists, or diverges to infinity. This gap will be apparent in the context of Lem. 9 and Thm. 2.*

***Remark* 4.** *Assumption 1 is satisfied with $S'(1) = 0$ for a subcritical Galton-Watson process $\mathcal{GW}(\{q_k\})$ with $q_1 = 0$. Indeed, we have*

$$\lim_{x\to 1-} \frac{Q(x)-x}{1-x} = \lim_{x\to 1-} \left(q_0 - \sum_{k=2}^{\infty} q_k \frac{1-x^{k-1}}{1-x} x \right) = \lim_{x\to 1-} \left(q_0 - \sum_{k=2}^{\infty} (k-1) q_k \right)$$
$$= 1 - \sum_{k=2}^{\infty} k q_k = 1 - Q'(1) > 0$$

and, therefore,

$$S'(1) = 1 - \lim_{x\to 1-} \frac{Q(x)-x}{(1-x)\big(1-Q'(x)\big)} = 1 - \frac{1-Q'(1)}{1-Q'(1)} = 0. \tag{29}$$

Lemma 4. *Consider a critical Galton-Watson measure $\mathcal{GW}(\{q_k\})$ with $q_1 = 0$. If the second moment of the offspring distribution is finite,*

$$\mathsf{E}[X^2] = \sum_{k=0}^{\infty} k^2 q_k \ < \infty,$$

then Assumption 1 is satisfied with $S'(1) = \frac{1}{2}$ and $L = \lim_{x\to 1-} \left(\frac{\ln g(x)}{-\ln(1-x)} \right) = 0$.

Proof. By L'Hôpital's rule,

$$\lim_{x\to 1-} \frac{Q(x)-x}{(1-x)^2} = \frac{1}{2} \lim_{x\to 1-} \frac{1-Q'(x)}{1-x} = \frac{Q''(1)}{2}.$$

Thus,

$$S'(1) = 1 - \lim_{x\to 1-} \left(\frac{Q(x)-x}{(1-x)^2} \frac{1-x}{1-Q'(x)} \right) = 1 - \frac{Q''(1)}{2Q''(1)} = \frac{1}{2},$$

and by Lem. 3, $L = 2 - \frac{1}{1-S'(1)} = 0$. □

The next statement suggests a sufficient condition for Assumption 1.

Lemma 5 (Regularity condition). *Consider a critical Galton-Watson process $\mathcal{GW}(\{q_k\})$ with $q_1 = 0$ and infinite second moment, i.e., $\sum_{k=0}^{\infty} k^2 q_k = \infty$. Suppose that for the progeny variable $X \overset{d}{\sim} \{q_k\}$ the following limit exists:*

$$\Lambda = \lim_{k\to\infty} \frac{k}{E[X \mid X \geq k]} = \lim_{k\to\infty} \frac{k \sum_{m=k}^{\infty} q_m}{\sum_{m=k}^{\infty} m q_m}. \tag{30}$$

Then Assumption 1 is satisfied with $S'(1) = \Lambda$.

Proof. For $x \in (0,1)$,

$$\frac{1-Q'(x)}{1-x} = \frac{Q'(1)-Q'(x)}{1-x} = \sum_{m=0}^{\infty} m q_m \frac{1-x^{m-1}}{1-x} = \sum_{m=0}^{\infty} \sum_{k=0}^{m-2} m q_m x^k = \sum_{k=0}^{\infty} b_k x^k,$$

where

$$b_k = \sum_{m=k+2}^{\infty} m q_m. \tag{31}$$

Recall that (22) shows that $g(x) = \sum\limits_{k=0}^{\infty} a_k x^k$, where

$$a_k = \sum_{m=k+2}^{\infty} (m-k-1) q_m = b_k - c_k \quad \text{with} \quad c_k = (k+1) \sum_{m=k+2}^{\infty} q_m \tag{32}$$

and b_k as defined in (31). Equations (31) and (32) yield

$$\frac{Q(x)-x}{(1-x)(1-Q'(x))} = \frac{(1-x)g(x)}{1-Q'(x)} = \frac{\sum\limits_{k=0}^{\infty} a_k x^k}{\sum\limits_{k=0}^{\infty} b_k x^k} = 1 - \frac{\sum\limits_{k=0}^{\infty} c_k x^k}{\sum\limits_{k=0}^{\infty} b_k x^k}. \tag{33}$$

The infinite second moment condition implies $\lim\limits_{x\to 1-} \sum\limits_{k=0}^{\infty} b_k x^k = \infty$ and $\lim\limits_{x\to 1-} \sum\limits_{k=0}^{\infty} c_k x^k = \infty$. As (30) postulates that $\lim\limits_{k\to\infty} (c_k/b_k) = \Lambda$, for a given small value of $\epsilon > 0$, there exists $K \in \mathbb{N}$ such that $|c_k/b_k - \Lambda| < \epsilon$, $\forall k \geq K$. Thus

$$\liminf_{x\to 1-} \frac{\sum\limits_{k=0}^{\infty} c_k x^k}{\sum\limits_{k=0}^{\infty} b_k x^k} = \liminf_{x\to 1-} \frac{\sum\limits_{k=K}^{\infty} c_k x^k}{\sum\limits_{k=K}^{\infty} b_k x^k} \geq \Lambda - \epsilon$$

and

$$\limsup_{x\to 1-} \frac{\sum\limits_{k=0}^{\infty} c_k x^k}{\sum\limits_{k=0}^{\infty} b_k x^k} = \limsup_{x\to 1-} \frac{\sum\limits_{k=K}^{\infty} c_k x^k}{\sum\limits_{k=K}^{\infty} b_k x^k} \leq \Lambda + \epsilon.$$

Consequently,

$$\lim_{x\to 1-} \frac{\sum\limits_{k=0}^{\infty} c_k x^k}{\sum\limits_{k=0}^{\infty} b_k x^k} = \Lambda.$$

Hence, the limit in (27) exists, and (33) implies

$$S'(1) = 1 - \lim_{x\to 1-} \frac{Q(x)-x}{(1-x)\big(1-Q'(x)\big)} = \lim_{x\to 1-} \frac{\sum\limits_{k=0}^{\infty} c_k x^k}{\sum\limits_{k=0}^{\infty} b_k x^k} = \Lambda.$$

□

Corollary 2 (**Zipf distribution**). *Consider a critical Galton-Watson process $\mathcal{GW}(\{q_k\})$ with $q_1 = 0$ and offspring distribution $\{q_k\}$ of Zipf type:*

$$q_k \sim Ck^{-(\alpha+1)} \quad \text{with} \quad \alpha \in (1,2] \quad \text{and} \quad C > 0. \tag{34}$$

Then Assumption 1 is satisfied,

$$S'(1) = \frac{\alpha-1}{\alpha} \quad \text{and} \quad L = \lim_{x\to 1-} \left(\frac{\ln g(x)}{-\ln(1-x)}\right) = 2-\alpha. \tag{35}$$

Proof. Suppose $q_k = Ck^{-(\alpha+1)}\big(1+o(1)\big)$. Then,

$$\sum_{m=k}^{\infty} mq_m = C\frac{k^{1-\alpha}}{\alpha-1}\big(1+o(1)\big) \quad \text{and} \quad \sum_{m=k}^{\infty} q_m = C\frac{k^{-\alpha}}{\alpha}\big(1+o(1)\big).$$

Hence, the limit Λ defined in (30) exists and is equal to

$$\Lambda = \lim_{k\to\infty} \frac{k\sum\limits_{m=k}^{\infty} q_m}{\sum\limits_{m=k}^{\infty} mq_m} = \frac{\alpha-1}{\alpha}$$

Consequently, Lem. 5 implies Assumption 1 and $S'(1) = \frac{\alpha-1}{\alpha}$. Finally, by Lem. 3 we have

$$L = 2 - \frac{1}{1-S'(1)} = 2-\alpha.$$

□

***Example* 1** (**Infinite second moment, $L = 0$**). *Consider a critical Galton-Watson process $\mathcal{GW}(\{q_k\})$ with $q_0 = \frac{2}{3}$, $q_1 = 0$, and*

$$q_k = \frac{4/3}{k(k^2-1)} \qquad (k \geq 2).$$

Observe that the offspring distribution q_k is of Zipf type (34) *with $\alpha = 2$. This offspring distribution has infinite second moment. Here,*

$$Q(x) - x = (1-x)^2\, g(x) \quad \text{with} \quad g(x) = -\frac{2/3}{x}\ln(1-x),$$

and therefore, the limit in (27) *exists and is equal to*

$$\lim_{x\to 1-} \frac{Q(x)-x}{(1-x)\big(1-Q'(x)\big)} = \lim_{x\to 1-} \frac{\ln(1-x)}{2\ln(1-x)+\frac{1}{x}} = \frac{1}{2}.$$

Hence, Assumption 1 is satisfied with $S'(1) = 1 - \frac{1}{2} = \frac{1}{2}$. On the other hand,

$$L = \lim_{x\to 1-}\left(\frac{\ln g(x)}{-\ln(1-x)}\right) = \lim_{x\to 1-}\left(\frac{\ln\left(-\ln(1-x)\right)}{-\ln(1-x)}\right) = 0,$$

which is consistent with Lem. 2. We also see that $S'(1) = \frac{1-L}{2-L} = \frac{\alpha-1}{\alpha}$, giving an example for statements in Lem. 3 and Cor. 2.

3.3. Tokunaga coefficients in recursive form

Here we derive a recursive expression for the Tokunaga coefficients of a Galton-Watson measure in the form $T_{i,j} = \pi_i f(\sigma_{j-2}, \pi_{j-1}, \pi_j)$. The recursive nature of this representation is connected to the recursive expression (16) for π_i of Lem. 1.

Lemma 6 (**Tokunaga coefficients**). *Consider a Galton-Watson measure $\mathcal{GW}(\{q_k\})$ with $q_1 = 0$. Assume criticality or subcriticality, i.e., $\sum_{k=0}^{\infty} kq_k \le 1$. Then, for all $1 \le i < j-1$, we have*

$$T_{i,j} = \pi_i \frac{Q'(\sigma_{j-1}) - Q'(\sigma_{j-2}) - \pi_{j-1}Q''(\sigma_{j-2})}{Q(\sigma_{j-1}) - Q(\sigma_{j-2}) - \pi_{j-1}Q'(\sigma_{j-2})} + T^o_{i,j}, \tag{36}$$

and for $1 \le i = j-1$,

$$T_{j-1,j} = \frac{\pi_{j-1}Q'(\sigma_{j-1}) + \pi_{j-1}Q'(\sigma_{j-2}) - 2Q(\sigma_{j-1}) + 2Q(\sigma_{j-2})}{Q(\sigma_{j-1}) - Q(\sigma_{j-2}) - \pi_{j-1}Q'(\sigma_{j-2})} + T^o_{j-1,j}, \tag{37}$$

where $T^o_{i,j} = \pi_i \frac{Q''(\sigma_{j-1})}{1-Q'(\sigma_{j-1})}$ is the expected number of offsprings of order i descendant to all regular (non-terminal) vertices of order j.

Note that (36) can be rewritten as

$$T_{i,j} = \pi_i \frac{d}{dx}\ln\left(\frac{Q(x+\pi_{j-1}) - Q(x) - \pi_{j-1}Q'(x)}{1 - Q'(x+\pi_{j-1})}\right)\Bigg|_{x=\sigma_{j-2}}.$$

Proof. For $i \le j-2$, let $M_{i,j}^{\text{term}}$ denote the expected number of descendants of order i of a terminal vertex of order j, i.e., the expected number of side branches of Tokunaga index $\{i,j\}$; see Remark 1. For $d \in \mathbb{N}$, consider all vertices in generation d. The probability that a vertex is a terminal vertex in a branch of order j is

$$\sum_{m=2}^{\infty} q_m \sum_{\ell=2}^{m} \binom{m}{\ell} \pi_{j-1}^{\ell} \sigma_{j-2}^{m-\ell} = Q(\sigma_{j-1}) - Q(\sigma_{j-2}) - \pi_{j-1} Q'(\sigma_{j-2}),$$

where $m \ge 2$ is its branching number (i.e., the number of descendants) and $\ell \ge 2$ is the number of descendants of order $j-1$.

Recall that $\sum_{m=0}^{k} m\binom{k}{m} a^m b^{k-m} = ka(a+b)^{k-1}$. The expected number of offsprings of order i descendant to a vertex conditioned on having a total of $m \ge 2$ offsprings, of which $\ell \ge 2$ are of order $j-1$ and $m-\ell$ are of order smaller than $j-1$, is

$$\frac{1}{\sigma_{j-2}^{m-\ell}} \sum_{k=0}^{m-\ell} k \binom{m-\ell}{k} \pi_i^k \big(\sigma_{j-2} - \pi_i\big)^{m-\ell-k} = \pi_i \frac{m-\ell}{\sigma_{j-2}}.$$

Thus, for $i \le j-2$,

$$\begin{aligned} M_{i,j}^{\text{term}} &= \frac{\sum_{m=2}^{\infty} q_m \sum_{\ell=2}^{m} \binom{m}{\ell} \pi_{j-1}^{\ell} \sum_{k=0}^{m-\ell} k \binom{m-\ell}{k} \pi_i^k \big(\sigma_{j-2} - \pi_i\big)^{m-\ell-k}}{\sum_{m=2}^{\infty} q_m \sum_{\ell=2}^{m} \binom{m}{\ell} \pi_{j-1}^{\ell} \sigma_{j-2}^{m-\ell}} \\ &= \frac{\pi_i \sum_{m=2}^{\infty} q_m \sum_{\ell=2}^{m} (m-\ell) \binom{m}{\ell} \pi_{j-1}^{\ell} \sigma_{j-2}^{m-\ell-1}}{Q(\sigma_{j-1}) - Q(\sigma_{j-2}) - \pi_{j-1} Q'(\sigma_{j-2})} \\ &= \frac{\pi_i \sum_{m=2}^{\infty} q_m \left(m\sigma_{j-1}^{m-1} - m(m-1)\pi_{j-1}\sigma_{j-2}^{m-2} - m\sigma_{j-2}^{m-1}\right)}{Q(\sigma_{j-1}) - Q(\sigma_{j-2}) - \pi_{j-1} Q'(\sigma_{j-2})} \\ &= \pi_i \frac{Q'(\sigma_{j-1}) - Q'(\sigma_{j-2}) - \pi_{j-1} Q''(\sigma_{j-2})}{Q(\sigma_{j-1}) - Q(\sigma_{j-2}) - \pi_{j-1} Q'(\sigma_{j-2})}. \end{aligned} \tag{38}$$

Next, for $i = j-1$, let $M_{j-1,j}^{\text{term}}$ denote the expected number of order $j-1$ side-branches adjacent to a terminal vertex of a branch of order j. The expected number of order $j-1$ offsprings of a vertex conditioned on being the terminal vertex in a branch of order j with a total of $m \ge 2$ offsprings is

$$\frac{1}{\sigma_{j-2}^{m-\ell}} \sum_{\ell=2}^{m} \ell \binom{m}{\ell} \pi_{j-1}^{\ell} \sigma_{j-2}^{m-\ell},$$

where $\ell \ge 2$ counts the offspring of order $j-1$, and the rest $m-\ell$ represent the offsprings of order smaller than $j-1$. Following Remark 1, we subtract two principal branches from

the number of order $j-1$ offsprings. Consequently, the expected number of order $j-1$ side branches adjacent to a vertex conditioned on being the terminal vertex in a branch of order j with a total of $m \geq 2$ offsprings is equal to

$$\frac{1}{\sigma_{j-2}^{m-\ell}} \sum_{\ell=2}^{m} (\ell-2) \binom{m}{\ell} \pi_{j-1}^{\ell} \sigma_{j-2}^{m-\ell}.$$

Here, of $\ell \geq 2$ offspring of order $j-1$, two are principle branches and $\ell-2$ are side branches. Hence, we have

$$\begin{aligned} M_{j-1,j}^{\text{term}} &= \frac{\sum\limits_{m=2}^{\infty} q_m \sum\limits_{\ell=2}^{m} (\ell-2)\binom{m}{\ell} \pi_{j-1}^{\ell} \sigma_{j-2}^{m-\ell}}{Q(\sigma_{j-1}) - Q(\sigma_{j-2}) - \pi_{j-1} Q'(\sigma_{j-2})} \\ &= \frac{\sum\limits_{m=2}^{\infty} q_m \left(m\pi_{j-1}\sigma_{j-1}^{m-1} + m\pi_{j-1}\sigma_{j-2}^{m-1} - 2\sigma_{j-1}^{m} + 2\sigma_{j-2}^{m} \right)}{Q(\sigma_{j-1}) - Q(\sigma_{j-2}) - \pi_{j-1} Q'(\sigma_{j-2})} \\ &= \frac{\pi_{j-1} Q'(\sigma_{j-1}) + \pi_{j-1} Q'(\sigma_{j-2}) - 2Q(\sigma_{j-1}) + 2Q(\sigma_{j-2})}{Q(\sigma_{j-1}) - Q(\sigma_{j-2}) - \pi_{j-1} Q'(\sigma_{j-2})}. \end{aligned} \tag{39}$$

The expected number V_j^o of regular (non-terminal) vertices in a branch of order j is computed as follows:

$$V_j^o = \frac{\sum\limits_{r=0}^{\infty} r \left(\sum\limits_{k=2}^{\infty} q_k k \sigma_{j-1}^{k-1} \right)^r}{\sum\limits_{r=0}^{\infty} \left(\sum\limits_{k=2}^{\infty} q_k k \sigma_{j-1}^{k-1} \right)^r} = \frac{Q'(\sigma_{j-1})}{1 - Q'(\sigma_{j-1})}, \tag{40}$$

where, following (17), the probability of a vertex being a regular vertex in a branch of order j, conditioned on it being of order j, equals

$$\sum_{k=2}^{\infty} q_k k \sigma_{j-1}^{k-1}.$$

Finally, the expected number $M_{i,j}^o$ of order i offsprings (and therefore, side branches of Tokunaga index $\{i,j\}$) in a regular (non-terminal) vertex on a branch of order j is

$$\begin{aligned} M_{i,j}^o &= \frac{1}{\sum\limits_{k=2}^{\infty} q_k k \sigma_{j-1}^{k-1}} \sum_{k=0}^{\infty} q_k k \sum_{s=0}^{k-1} s \binom{k-1}{s} \pi_i^s \left(\sigma_{j-1} - \pi_i\right)^{k-1-s} \\ &= \frac{1}{Q'(\sigma_{j-1})} \pi_i \sum_{k=2}^{\infty} q_k k(k-1) \sigma_{j-1}^{k-2} = \pi_i \frac{Q''(\sigma_{j-1})}{Q'(\sigma_{j-1})} \end{aligned} \tag{41}$$

for $1 \le i < j$. Here, k counts the total number of offsprings, of which we have k choices for the offspring of order j. Of the remaining $k-1$ ofsprings, we select s offsprings of order i and $k-1-s$ of order other than i, but less than j. There are $\binom{k-1}{s}$ such choices, with probability of $\pi_i^s \big(\sigma_{j-1} - \pi_i\big)^{k-1-s}$ for each such outcome.

The statement of the lemma follows from equations (38), (39), (40), (41) as $T_{i,j} = M_{i,j}^{term} + T_{i,j}^o$ with $T_{i,j}^o = V_j^o M_{i,j}^o$ by Wald's equation. □

***Example* 2 (Critical binary Galton-Watson tree).** *Consider the critical binary Galton-Watson distribution $\mathcal{GW}(q_0 = q_2 = 1/2)$. We have*

$$Q(z) = \frac{1+z^2}{2}, \quad S(z) = \frac{1+z}{2}, \quad \textit{and} \quad g(z) = 1/2.$$

Corollary 1 yields $\sigma_j = S(\sigma_{j-1})$ with $\sigma_0 = 0$, which implies by induction $\sigma_j = 1 - 2^{-j}$ and $\pi_j = 2^{-j}$ for $j \ge 1$. Equations (36) and (37) give

$$T_{i,j} = T_{i,j}^o = \frac{\pi_i}{1-\sigma_{j-1}} = 2^{j-i-1} \quad \textit{for all} \quad 1 \le i < j,$$

which implies the Toeplitz property (Def. 5) and Tokunaga self-similarity (Def. 6) with $(a,c) = (1,2)$ and $T_k = 2^{k-1}$.

Lemma 7 (Toeplitz implies criticality). *Consider a subcritical or critical Galton-Watson measure $\mathcal{GW}(\{q_k\})$ with $q_1 = 0$ that satisfies Assumption 1. If the Toeplitz property (Def. 5) is satisfied, then the measure is either critical or $q_0 = 1$, the order distribution is geometric with $\pi_k = q_0(1-q_0)^{k-1}$, and $q_0 = 1 - S'(1)$.*

Proof. The Toeplitz property implies the existence of the Tokunaga sequence $\{T_k\}_{k\in\mathbb{N}}$.

In the trivial case of $q_0 = 1$, we have $T_k = 0$ for any $k \ge 1$, $Q(z) = S(z) = 1$ so $S'(1) = 0 = 1 - q_0$, and $\pi_k = \delta_{1k}$. This establishes the statement.

Suppose that $q_0 < 1$. Equation (36) shows that there is a scalar $c > 0$ such that

$$\frac{T_{k+1}}{T_k} = \frac{\pi_i}{\pi_{i+1}} = c \qquad \forall k \ge 2,\, i \ge 1.$$

Thus, as $\pi_1 = q_0$, we have $\pi_j = q_0 c^{1-j}$ and since $\sum_j \pi_j = 1$ then $c = (1-q_0)^{-1}$.

Next, observe that since $S(x) = S(1) + S'(1)(x-1) + o(1-x)$ and $S(1) = 1$, we have

$$1 - q_0 = \frac{\pi_{i+1}}{\pi_i} = \frac{S(\sigma_i) - S(\sigma_{i-1})}{\pi_i} = \frac{S'(1)(\sigma_i - \sigma_{i-1}) + o(1-\sigma_{i-1})}{\pi_i} \to S'(1) \ \text{ as } i \to \infty$$

that leads to

$$q_0 = 1 - S'(1). \tag{42}$$

The criticality follows from the constraint $q_0 < 1$, since in the subcritical case we have $S'(1) = 0$ (see Rem. 4). □

The following statement gives an alternative proof to one of the main results of Burd et al. [2] using the framework of the present study.

Corollary 3. *Consider a subcritical or critical offspring distribution $\{q_k\}$ with $q_1 = 0$ and a finite second moment, $\sum_{k=1}^{\infty} k^2 q_k < \infty$. The measure $\mathcal{GW}(\{q_k\})$ satisfies the Toeplitz property (Def. 5) if and only if it is the critical binary Galton-Watson measure, $q_0 = q_2 = \frac{1}{2}$.*

Proof. By Lem. 4, the finite second moment implies Assumption 1 with $q_0 = 1 - S'(1) = \frac{1}{2}$. Assume the Toeplitz property holds. Then the criticality follows from Lem. 7. The criticality with $q_0 = \frac{1}{2}$ and $q_1 = 0$ yield $q_2 = \frac{1}{2}$ as

$$\sum_{k=2}^{\infty} \frac{k}{2} q_k = \frac{1}{2} = 1 - q_0 = \sum_{k=2}^{\infty} q_k.$$

Conversely, the Toeplitz property for the critical binary Galton-Watson tree is established in Ex. 2. □

3.4. Invariant Galton-Watson measures

The following result was originally proved in [2]. We state and prove it here since the expression (45) will be used in the proof of Thm. 2 below.

Lemma 8 (Pruning Galton-Watson tree, [2]). *Consider a critical or subcritical Galton-Watson measure $\mu \equiv \mathcal{GW}(\{q_k\})$ with $q_1 = 0$ on $\mathcal{T}^{|}$ with generating function $Q(z)$, and the corresponding pushforward probability measure induced by the Horton pruning operator $\mathcal{R}$,*

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

Then, $\nu(T \,|T \neq \phi)$ is a Galton-Watson measure $\mathcal{GW}(\{q_k^{(1)}\})$ on $\mathcal{T}^{|}$ with offspring probabilities

$$q_0^{(1)} = \frac{Q(q_0) - q_0}{(1 - q_0)\big(1 - Q'(q_0)\big)}, \tag{43}$$

$q_1^{(1)} = 0$, *and*

$$q_k^{(1)} = \frac{(1 - q_0)^{k-1} Q^{(k)}(q_0)}{k!\big(1 - Q'(q_0)\big)} \quad (k \geq 2), \tag{44}$$

and generating function

$$Q_1(z) = z + \frac{Q\big(q_0 + (1 - q_0)z\big) - q_0 - z(1 - q_0)}{(1 - q_0)\big(1 - Q'(q_0)\big)}. \tag{45}$$

Moreover, if $\mu(T)$ is critical, then so is $\nu(T \,|T \neq \phi)$. If $\mu(T)$ is subcritical, then the first moment is decreasing with pruning, i.e., $\sum\limits_{k=2}^{\infty} k q_k^{(1)} < \sum\limits_{k=2}^{\infty} k q_k < 1$.

Proof. The standard thinning argument (with $\pi_1 = q_0$ being the probability of eliminating an offspring) implies that $\mathcal{R}(T)$ is distributed as a Galton-Watson tree, i.e., $\mathcal{R}(T) \overset{d}{\sim} \mathcal{GW}(\{q_m^{(1)}\})$. Indeed, think of a random tree obtained as a result of the auxiliary branching process defined in the following way. We trace the branching process that starts with one generation zero progenitor vertex (the root) that produces exactly one offspring. From generation one on, the branching process evolves according to the offspring distribution $\left\{\frac{q_k}{1-q_0}\right\}_{k=2,3,\dots}$. Next, the process is thinned: once an offspring is produced (in each generation, including generation zero), it is either instantaneously eliminated with probability q_0 or is left untouched with probability $1-q_0$, where these Bernoulli trials are performed independently of each other and the branching history. Naturally, this generates a Galton-Watson branching process with branching probabilities $\{p_m\}$ calculated as follows

$$p_m = \sum_{k=m\vee 2}^{\infty} \binom{k}{m} q_0^{k-m}(1-q_0)^m \frac{q_k}{1-q_0}. \tag{46}$$

The above defined *thinned Galton-Watson process* can be equivalently formulated by tracking the original branching process with branching probabilities $\{q_k\}$. Here, for each offspring, it is instantaneously decided whether the offspring is a leaf or not via a Bernoulli trial with probabilities q_0 and $1-q_0$ for 'leaf' and 'no leaf' outcomes respectively. If the offspring is decided to be a leaf, it is pruned instantaneously. If not a leaf, it will branch according to the offspring distribution $\left\{\frac{q_k}{1-q_0}\right\}_{k=2,3,\dots}$. The thinned Galton-Watson process differs from the original one by pruning all the leaves. Hence, it implements the instantaneous Horton pruning, but not yet series reduction. Indeed, the above thinned Galton-Watson prices with branching probabilities $\{p_m\}$ can have single offspring nodes.

Next, we need to account for the series reduction by generating a Galton-Watson process with the branching probabilities $\{q_m^{(1)}\}$ by letting

$$q_0^{(1)} = \frac{p_0}{1-p_1} = \frac{(1-q_0)^{-1}\sum_{k=2}^{\infty} q_0^k q_k}{1-\sum_{k=2}^{\infty} k q_0^{k-1} q_k},$$

$q_1^{(1)} = 0$, and for $m \geq 2$,

$$q_m^{(1)} = \frac{p_m}{1-p_1} = \frac{(1-q_0)^{m-1}\sum_{k=m}^{\infty} \binom{k}{m} q_0^{k-m} q_k}{1-\sum_{k=2}^{\infty} k q_0^{k-1} q_k}.$$

This branding process induces the tree measure $\nu(T)$. Note that there is an alternative derivation of (43) as by Cor. 1, $q_0^{(1)} = \frac{\pi_2}{1-\sigma_1} = \frac{S(q_0)-q_0}{1-q_0} = \frac{Q(q_0)-q_0}{(1-q_0)(1-Q'(q_0))}$.

We notice that the corresponding generating function can be computed as follows

$$Q_1(z)=\sum_{m=0}^{\infty} z^m q_m^{(1)}=\frac{(1-q_0)^{-1}}{1-\sum\limits_{k=2}^{\infty} k q_0^{k-1} q_k}\left(\sum_{k=2}^{\infty} q_0^k q_k+\sum_{m=2}^{\infty}\sum_{k=m}^{\infty}\big(zq_0^{-1}(1-q_0)\big)^m\binom{k}{m}q_0^k q_k\right)$$
$$=\frac{(1-q_0)^{-1}}{1-Q'(q_0)}\left(\sum_{k=2}^{\infty} q_0^k q_k+\sum_{k=2}^{\infty}\sum_{m=2}^{k}\binom{k}{m}\big(zq_0^{-1}(1-q_0)\big)^m q_0^k q_k\right)$$
$$=\frac{(1-q_0)^{-1}}{1-Q'(q_0)}\big(Q\big(z+(1-z)q_0\big)-q_0-z(1-q_0)Q'(q_0)\big)$$

by the binomial theorem, implying (45). We proceed by differentiating $\frac{d}{dz}$ in (45), obtaining

$$Q_1'(z)=\frac{Q'(q_0+z(1-q_0))-Q'(q_0)}{1-Q'(q_0)}. \tag{47}$$

Next, we observe that if $\mu(T)$ is critical, (47) implies $\sum\limits_{k=2}^{\infty} k q_k^{(1)}=Q_1'(1)=\frac{Q'(1)-Q(q_0)}{1-Q(q_0)}=1$. That is, the critical process stays critical after a Horton pruning. Finally, in the subcritical case, $Q'(1)<1$, and by formula (47), $Q_1'(1)=\frac{Q'(1)-Q(q_0)}{1-Q(q_0)}<Q'(1)$. □

Formula (45) matches the evolution of the generator under tree erasure discussed by He and Winkel [8, Lemma 11]; see also Neveu [16] and Kesten [10]. Also, observe that expression (45) is of the same form as the generating function of a thinned Galton-Watson process in the work of Duquesne and Winkel [4, eqn. (10) of Sec. 2.2], where the thinning was done in the context of a Bernoulli leaf coloring scheme.

Lemma 9. *Consider a critical or subcritical Galton-Watson measure $\mathcal{GW}(\{q_k\})$ with $q_1=0$. If it is Horton prune-invariant (self-similar) (Def. 4), then the limit*

$$\lim_{x\to 1-}\left(\frac{\ln g(x)}{-\ln(1-x)}\right)=L$$

exists and is finite. Moreover,

$$L=1-\frac{\ln\big(1-Q'(q_0)\big)}{\ln(1-q_0)}.$$

Proof. The Horton prune-invariance implies $Q_1(z)=Q(z)$ in the recursion (45):

$$Q(z)=z+\frac{Q\big(q_0+(1-q_0)z\big)-q_0-z(1-q_0)}{(1-q_0)\big(1-Q'(q_0)\big)}, \tag{48}$$

which we rewrite as

$$Q\big(q_0+(1-q_0)z\big)-\big(q_0+z(1-q_0)\big) = M(q_0)\big(Q(z)-z\big), \text{ where } M(q_0) = (1-q_0)\big(1-Q'(q_0)\big). \tag{49}$$

Then, for any $k \in \mathbb{N}$,

$$Q\big(1-(1-q_0)^k + (1-q_0)^k z\big) - \big(1-(1-q_0)^k + (1-q_0)^k z\big) = \big(M(q_0)\big)^k \big(Q(z)-z\big)$$

and for $z \in [0,1)$,

$$\begin{aligned}\lim_{k\to\infty} \frac{\ln\big(Q\big(1-(1-q_0)^k + (1-q_0)^k z\big) - \big(1-(1-q_0)^k + (1-q_0)^k z\big)\big)}{\ln\Big(1-\big(1-(1-q_0)^k + (1-q_0)^k z\big)\Big)} \\ = \lim_{k\to\infty} \frac{k\ln M(q_0) + \ln\big(Q(z)-z\big)}{k\ln(1-q_0)+\ln(1-z)} = \frac{\ln M(q_0)}{\ln(1-q_0)}.\end{aligned} \tag{50}$$

Next, notice that for $z \in I_0 = [0, q_0)$,

$$\ln(q_0) \le \ln\big(Q(z)-z\big) \le \ln\big(Q(q_0)-q_0\big) \quad \text{and} \quad \ln(1-q_0) \le \ln(1-z) \le 0.$$

Hence, for any $x \in I_k = \big(1-(1-q_0)^k,\, 1-(1-q_0)^{k+1}\big)$, there is a $z \in I_0$ such that

$$x = 1-(1-q_0)^k + (1-q_0)^k z$$

and

$$\frac{k\ln M(q_0) + \ln\big(Q(q_0)-q_0\big)}{k\ln(1-q_0)+\ln(1-q_0)} \le \frac{\ln\big(Q(x)-x\big)}{\ln(1-x)} = \frac{k\ln M(q_0)+\ln\big(Q(z)-z\big)}{k\ln(1-q_0)+\ln(1-z)} \le \frac{k\ln M(q_0)+\ln(q_0)}{k\ln(1-q_0)}.$$

Hence, the following limit exists

$$\lim_{x\to 1-} \frac{\ln\big(Q(x)-x\big)}{\ln(1-x)} = \frac{\ln M(q_0)}{\ln(1-q_0)}.$$

Finally,

$$\begin{aligned}\lim_{x\to 1-}\left(\frac{\ln g(x)}{-\ln(1-x)}\right) &= \lim_{x\to 1-} \frac{2\ln(1-x) - \ln\big(Q(x)-x\big)}{\ln(1-x)} = 2 - \frac{\ln M(q_0)}{\ln(1-q_0)} \\ &= 1 - \frac{\ln\big(1-Q'(q_0)\big)}{\ln(1-q_0)}.\end{aligned} \tag{51}$$

□

Next, we define a single parameter family of critical Galton-Watson measures $\mathcal{GW}(\{q_k\})$ with $q_1 = 0$ on $\mathcal{T}^{|}$.

Definition 8 (**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 is given by*

$$Q(z) = z + q(1-z)^{1/q}. \tag{52}$$

The respective branching probabilities are $q_0 = q$, $q_1 = 0$, $q_2 = (1-q)/2q$, and

$$q_k = \frac{1-q}{k!\, q} \prod_{i=2}^{k-1} (i - 1/q) \quad (k \geq 3). \tag{53}$$

Here, if $q = 1/2$, then the distribution is critical binary, i.e., $\mathcal{GW}(q_0 = q_2 = 1/2)$. If $q \in (1/2, 1)$, the distribution is of Zipf type with

$$q_k = \frac{(1-q)\Gamma(k-1/q)}{q\Gamma(2-1/q)\, k!} \sim C k^{-(1+q)/q}, \quad \textit{where} \;\; C = \frac{1-q}{q\,\Gamma(2-1/q)}. \tag{54}$$

Theorem 2 (**Self-similar Galton-Watson measures**). *Consider a critical or sub-critical Galton-Watson measure $\mathcal{GW}(\{q_k\})$ with $q_1 = 0$ that satisfies Assumption 1. The measure is Horton prune-invariant (self-similar) (Def. 4) if and only if it is the invariant Galton-Watson (IGW) measure $\mathcal{IGW}(q_0)$ with $q_0 \in [1/2, 1)$.*

Proof. Combining equations (43) and (45), we have

$$Q_1(z) = z + q_0^{(1)} \frac{Q\big(q_0 + (1-q_0)z\big) - \big(q_0 + (1-q_0)z\big)}{Q(q_0) - q_0}. \tag{55}$$

If the Galton-Watson measure is Horton prune-invariant, then $Q_1(z) = Q(z)$, and (55) implies

$$R(z) = \frac{R\big(q_0 + (1-q_0)z\big)}{R(q_0)} \quad \text{for} \quad R(z) = \frac{Q(z) - z}{q_0}$$

for $z \in [0, 1)$. Hence, letting $\ell(z) = \ln R(1-z)$ for $z \in (0, 1]$, we have

$$\ell(z) + \ell(1-q_0) = \ell\big((1-q_0)z\big).$$

Finally, for $r(y) = \ell\big(e^{-y}\big) = \ln R\big(1 - e^{-y}\big)$ for $y \in [0, \infty)$ and $\kappa_0 = -\ln(1-q_0)$,

$$r(y + \kappa_0) = r(y) + r(\kappa_0) \qquad \forall y \in [0, \infty).$$

Therefore, $r'(y+\kappa_0) = r'(y)$ and $r(y) = -\int_0^y \alpha(w)\, dw$ for some κ_0-periodic function $\alpha(y)$. Thus,

$$Q(z) = z + q_0 R(z) = z + q_0 e^{\ell(1-z)} = z + q_0 e^{r\left(-\ln(1-z)\right)} = z + q_0 \exp\left\{ -\int_0^{-\ln(1-z)} \alpha(w)\, dw \right\}. \tag{56}$$

Next, $0 = q_1 = Q'(0) = 1 - \alpha(0)q_0$ implies $\alpha(0) = \frac{1}{q_0}$. Also, for $z \in (0,1)$, $R'(z) = \frac{Q'(z)-1}{q_0} < 0$ and $r(y)$ is a decreasing function. Hence, $\alpha(y) > 0$ for all $y \in (0,1)$.

Letting $w = -\ln(1-x)$ in (56), we have

$$\ln\big(Q(z)-z\big) = \ln(q_0) - \int_0^{-\ln(1-z)} \alpha(w)\,dw \;= \ln(q_0) - \int_0^z \frac{\alpha(-\ln(1-x))}{1-x}\,dx \qquad \forall z \in [0,1). \tag{57}$$

Recall that $\frac{d}{dz}\ln\big(Q(z)-z\big) = \frac{-1}{S(z)-z}$, and therefore,

$$\ln\big(Q(z)-z\big) = \ln(q_0) - \int_0^z \frac{dx}{S(x)-x} \qquad \forall z \in [0,1). \tag{58}$$

Equations (57) and (58) yield

$$S(z) = z + q_0(1-z)\varphi(z), \quad \text{where } \varphi(z) = \frac{1}{q_0\alpha(-\ln(1-z))}. \tag{59}$$

Here, $\alpha(0) = \frac{1}{q_0}$ implies $\varphi(0) = 1$. Since $\alpha(z)$ is κ_0-periodic function,

$$\alpha(-\ln(1-z)) = \alpha(-\ln(1-z)+\kappa_0) = \alpha\left(-\ln\big(1-(q_0+(1-q_0)z\big)\right)$$

and $\varphi(z)$ satisfies

$$\varphi(z) = \varphi\big(q_0 + (1-q_0)z\big) \qquad \forall z \in [0,1). \tag{60}$$

Equation (59) implies the existence of the limit

$$\varphi(1) = \lim_{x\to 1-} \varphi(x) = \frac{1}{q_0}\lim_{x\to 1-}\frac{S(x)-x}{1-x} = \frac{1-S'(1)}{q_0}.$$

Next, iterating (60), we have

$$\varphi(x) = \lim_{k\to\infty} \varphi\left(\big(1-(1-q_0)^k\big) + (1-q_0)^k x\right) = \varphi(1) \qquad \forall x \in [0,1).$$

Hence, $\varphi(x) \equiv 1$, and by (59),

$$S(z) = z + q_0(1-z).$$

Consequently, (58) implies $Q(z) = z + q_0(1-z)^{1/q_0}$.

Finally, observe that for an invariant Galton-Watson measure $\mathcal{IGW}(q_0)$ with any $q_0 \in [1/2, 1)$ satisfies (55). In particular, equation (52) implies

$$S(z) = q_0 + (1-q_0)z. \tag{61}$$

The statement of the theorem follows. □

***Remark* 5 (Heuristics for a linear $S(z)$).** *Consider a Horton prune-invariant measure (or at least a Toeplitz measure with $q_0 < 1$) that satisfies Assumption 1. Lemma 7 shows that in this case*

$$\frac{\pi_{k+1}}{\pi_k} = \frac{\sigma_{k+1} - \sigma_k}{\pi_k} = S'(1) = 1 - q_0 \quad \text{for all } k \geq 1.$$

Together with the recursion $\sigma_k = S(\sigma_{k-1})$ of Cor. 1 (see also Fig. 7), this implies that the points $(\sigma_k, S(\sigma_k))$ lie on the line

$$y(z) = q_0 + (1 - q_0)z.$$

This observation suggests $S(z) = q_0 + (1-q_0)z$ as a possible solution of the equation (45) with $Q_1(z) = Q(z)$, and the corresponding $Q(z) = z + q_0(1-z)^{1/q_0}$ is found by (58). Theorem 2 ensures that this is the only solution under the regularity Assumption 1.

***Remark* 6 (Intuition behind the regularity condition).** *The Horton pruning acts as a rescaling (vertical and horizontal) on the function $S(z) - z$ from the restricted domain $[q_0, 1]$ to $[0, 1]$, according to (45). After k consecutive prunings, function $S_k(z) - z$ with the domain $[0, 1]$ is obtained via scaling from a restriction of $S(z) - z$ to the interval $[1-(1-q_0)^k, 1]$. Thus, consecutive pruning rescales and maps the function $S(z) - z$ in the vicinity of $1-$ to the interval $[0, 1]$. Assumption 1 requires a smooth behavior of $S(z)$ at $z = 1-$. The rescaling translates this smooth behavior to the ultimate linearity of function $S(z)$ on the entire interval $[0, 1]$. The most general form of prune-invariant $Q(z)$ is given in (56), which allows a non-linear oscillatory behavior of $S(z)$ between the points $(\sigma_k, S(\sigma_k))$ discussed in Rem. 5. The rescaling argument shows that such oscillations necessarily lead to non-smooth behavior of $S(z)$ at $z = 1-$ and hence violate Assumption 1.*

***Remark* 7 (General prune-invariant measures).** *Recall that according to Lem. 9, the general Horton prune-invariant distributions adhere to the existence and finiteness of the limit $L = \lim_{x \to 1-} \left(\frac{\ln g(x)}{-\ln(1-x)} \right)$, which is weaker than $S'(1)$ required in Assumption 1 (see Lem. 3). The gap between the two conditions allows for the existence of Horton prune-invariant distributions that satisfy (48) and have a nonlinear function $S(z)$. An example of such a measure and further discussion is given in Sect. 4.*

3.5. Attractors and basins of attraction

Theorem 3 (Attraction property of critical Galton-Watson trees). *Consider a critical Galton-Watson measure $\rho_0 \equiv \mathcal{GW}(\{q_k\})$ with $q_1 = 0$ on $\mathcal{T}^{|}$. 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 $\rho_{k+1}(T) = \nu_k\left(T \mid T \neq \phi\right)$. Suppose Assumption 1 is satisfied. Then, for any $T \in \mathcal{T}^{|}$,*

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

where ρ^* denotes the invariant Galton-Watson measure $\mathcal{IGW}(q)$ with $q = 1 - S'(1)$.

Finally, if the Galton-Watson measure $\rho_0 \equiv \mathcal{GW}(\{q_k\})$ is subcritical, then $\rho_k(T)$ converges to a point mass measure, $\mathcal{GW}(q_0 = 1)$.

Proof. Let $q_m^{(k)}$ denote the offspring distribution corresponding to the critical Galton-Watson tree measure ρ_k, where $q_1^{(k)} = 0$ by series reduction. First, we observe that

$$\begin{aligned}\lim_{k\to\infty} q_0^{(k)} &= \lim_{k\to\infty} \frac{\pi_k}{1-\sigma_{k-1}} = \lim_{k\to\infty} \frac{S(\sigma_{k-1}) - \sigma_{k-1}}{1-\sigma_{k-1}} \\ &= \lim_{k\to\infty} \frac{1 + S'(1)\big(\sigma_{k-1} - 1\big) + o\big(1-\sigma_{k-1}\big) - \sigma_{k-1}}{1-\sigma_{k-1}} \ = 1 - S'(1). \end{aligned} \tag{62}$$

Let $Q_k(z) := \sum_{m=0}^{\infty} z^m q_m^{(k)}$ denote the generating function corresponding to the Galton-Watson measure ρ_k and $S_k(z) = \frac{Q_k(z) - zQ_k'(z)}{1 - Q_k'(z)}$. Equation (45) implies

$$S_1(z) = \frac{1}{1-q_0} S\big(q_0 + (1-q_0)z\big) - \frac{q_0}{1-q_0}. \tag{63}$$

For a given $z \in [0,1)$, we iterate (63), obtaining

$$S_k(z) = \prod_{i=0}^{k-1} \frac{1}{1-q_0^{(i)}}\, S\left(\left(1 - \prod_{i=0}^{k-1}(1-q_0^{(i)})\right) + z\prod_{i=0}^{k-1}(1-q_0^{(i)})\right) + \left(1 - \prod_{i=0}^{k-1}\frac{1}{1-q_0^{(i)}}\right), \tag{64}$$

where $\prod_{i=0}^{k-1}(1-q_0^{(i)}) \le 2^{-k} \to 0$ as $k \to \infty$. Next, we substitute

$$S\left(\left(1 - \prod_{i=0}^{k-1}(1-q_0^{(i)})\right) + z\prod_{i=0}^{k-1}(1-q_0^{(i)})\right) = 1 + (z-1)S'(1)\prod_{i=0}^{k-1}(1-q_0^{(i)}) + o\left(\prod_{i=0}^{k-1}(1-q_0^{(i)})\right)$$

into (64), getting

$$S_k(z) = 1 + (z-1)S'(1) + o(1).$$

Hence, for a given $z \in [0,1)$, we have

$$\frac{d}{dz}\ln\big(Q_k(z) - z\big) = \frac{1}{z - S_k(z)} \ \longrightarrow\ \frac{1}{(1-S'(1))(z-1)} \qquad \text{as} \ \ k \to \infty.$$

Also, we notice that $Q_k(x) - x$ is a decreasing function ($Q_k'(x) < Q_k'(1) = 1$) and

$$q_0^{(k)} \ge Q_k(x) - x \ge Q_k(z) - z > 0 \qquad \forall x \in [0, z].$$

Therefore, letting $k \to \infty$, we have

$$\ln\left(Q_k(z)-z\right) = \ln q_0^{(k)} + \int_0^z \frac{d}{dx} \ln\left(Q_k(x)-x\right)\, dx \;\longrightarrow\; \ln q + \frac{1}{q}\ln(1-z),$$

where $q = 1 - S'(1)$, as $\lim_{k\to\infty} q_0^{(k)} = q$ by (62). We conclude that

$$\lim_{k\to\infty} Q_k(z) = z + q\,(1-z)^{1/q}$$

where the right hand side is the generating function for $\mathcal{IGW}\,(q)$.

Finally, if $\rho_0 \equiv \mathcal{GW}(\{q_k\})$ is subcritical, (29) and (62) imply $\lim_{k\to\infty} q_0^{(k)} = 1 - S'(1) = 1$. □

Theorem 3 and Cor. 2 immediately imply the following result.

Corollary 4 (**Attraction property of critical Galton-Watson trees of Zipf type**). *Consider a critical Galton-Watson process $\rho_0 \equiv \mathcal{GW}(\{q_k\})$ with $q_1 = 0$, with offspring distribution q_k of Zipf type, i.e., $q_k \sim Ck^{-(\alpha+1)}$, with $\alpha \in (1,2]$ and $C > 0$. 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, and set $\rho_{k+1}(T) = \nu_k\,(T\,|T \neq \phi)$. Then, for any $T \in \mathcal{T}^{|}$,*

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

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

Next, Lem. 2 and 4 imply the following attraction result as a corollary of our Thm. 3. The same attraction property has been established in [2] under the assumption of a bounded offspring distribution.

Corollary 5 (**Attraction property of critical binary Galton-Watson tree, [2]**). *Consider a critical Galton-Watson process $\rho_0 \equiv \mathcal{GW}(\{q_k\})$ with $q_1 = 0$. Assume one of the following two conditions holds.*

(a) *The second moment assumption is satisfied:*

$$\sum_{k=2}^{\infty} k^2 q_k \;< \infty.$$

(b) *Assumption 1 is satisfied, and the "2−" moment assumption is satisfied, i.e.,*

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

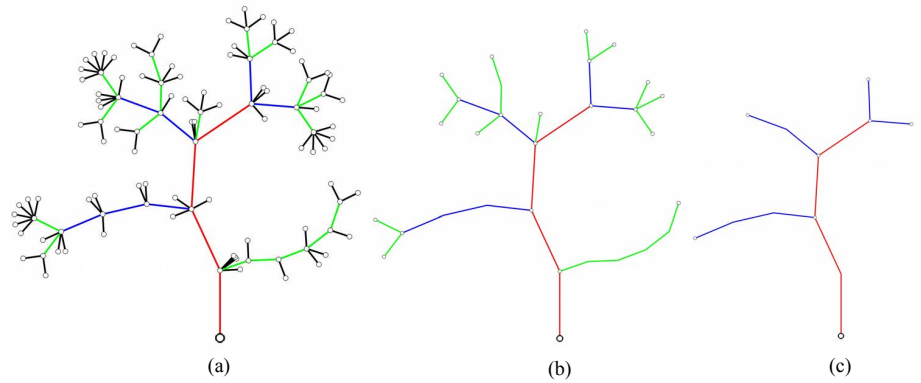

Figure 8: Binary attractor: Illustration. The tree T (panel a) has maximal branching number $b = 6$. Its first pruning (panel b) $\mathcal{R}(T)$ has maximal branching number $b = 3$. Its second pruning (panel c) $\mathcal{R}^2(T)$ has maximal branching number $b = 2$. This convergence to binary branching is generic in Galton-Watson trees that have offspring distribution with a finite $2 - \epsilon$ moment; see Cor. 5.

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, and set $\rho_{k+1}(T) = \nu_k(T \,|T \neq \phi)$. 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 $\mathcal{IGW}(1/2)$.

Figure 8 illustrates convergence of a tree with a large branching number to a binary tree.

3.6. Explicit Tokunaga coefficients and Horton law

In the next lemma we find the Tokunaga coefficients and the Horton exponent for an invariant Galton-Watson tree measure $\mathcal{IGW}(q_0)$.

Lemma 10 (Tokunaga coefficients). *Consider an invariant Galton-Watson measure $\mathcal{IGW}(q_0)$ for $q_0 \in [1/2, 1)$. Then,*

$$\pi_i = q_0\, c^{1-i} \quad \text{with} \quad c = \frac{1}{1-q_0}.$$

The measure satisfies Toeplitz property (Def. 5) with the Tokunaga coefficients

$$T^o_{i,j} = T^o_{j-i}, \quad \text{where} \quad T^o_k = c^{k-1} \quad (k = 1, 2, \ldots), \tag{65}$$

and

$$T_{i,j} = T_{j-i}, \quad \text{where} \quad T_1 = c^{c/(c-1)} - c - 1 \quad \text{and} \quad T_k = a\, c^{k-1} \quad (k = 2, 3, \ldots) \tag{66}$$

with $a = (c-1)\big(c^{1/(c-1)} - 1\big)$. *Finally, the strong Horton law* (13) *holds with Horton exponent* $R = c^{c/(c-1)} = (1-q_0)^{-1/q_0}$.

The functions $a(q_0), c(q_0)$ and $R(q_0)$ are illustrated in Fig. 1.

Proof. Equations (18) and (61) imply $\sigma_i = 1 - (1-q_0)^i + (1-q_0)^i\, z$. Hence, $\pi_i = \sigma_i - \sigma_{i-1} = q_0(1-q_0)^{i-1}$. Equations (65) and (66) are obtained via substituting π_i and σ_i into Lem. 6.

Finally, Thm. 1 implies the strong Horton law 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 $\big(0, \frac{1}{2}\big]$. We have

$$\hat{t}(z) = -1 + (T_1 + 2)z + \frac{acz^2}{1-cz},$$

which gives $w_0 = c^{-c/(c-1)}$ and $R = c^{c/(c-1)}$. □

4. Discussion

In this paper we described the invariance and attractor properties of combinatorial Galton-Watson trees with respect to the Horton pruning. The results hold under the regularity Assumption 1 that prohibits large tail oscillations of the offspring probabilities q_m that lead to a non-smooth behavior of $S(z)$ at $1-$. A sufficient condition under which the regularity assumption holds is suggested in Lem. 5. Theorem 2 introduces a one-parameter family of invariant Galton-Watson distributions $\mathcal{IGW}(q)$ and asserts that this family exhausts the Horton prune-invariant distributions within the examined regularity class. The invariant family has a power-law tail of the offspring distribution, $q_k \sim Ck^{-\alpha}$, with exactly one distribution for every $\alpha \in (2, 3]$, and also includes the critical binary Galton-Watson tree measure.

A similar approach can be applied to the search of invariant measures in a broader class of *generalized dynamical prunings* on trees with edge lengths introduced and analyzed in [12, 13] and to the pruning operation studied in Evans [6] and in Duquesne and Winkel [5]. Informally, a generalized dynamical pruning erases a tree from leaves down to the root at a rate that only depends on the descendant part of the tree. Most of such prunings, with a notable exception of the Horton pruning and the continuous erasure of Neveu [16], do not satisfy semigroup property. It has been shown in [12] that the critical binary Galton-Watson tree with i.i.d. exponential edge lengths is invariant with respect to any admissible generalized dynamical pruning. We conjecture that the Galton-Watson trees

that have i.i.d. exponential edge lengths and combinatorial shapes sampled from the invariant Galton-Watson measures $\mathcal{IGW}(q)$ introduced in this work (Def. 8) are the only Galton-Watson measures invariant with respect to all admissible generalized dynamical prunings, up to rescaling of the edge lengths. Heuristically, this is supported by the rescaling argument (Rem. 6) applied to the function $S(z)$. The Horton pruning of the present work only requires linearity of the function $S(z)$ on the grid σ_k, which is related to its discrete combinatorial action. This allows the existence of prune-invariant measures with oscillatory behavior, outside of the invariant Galton-Watson family. However, a continuous pruning, for instance the continuous erasure of Neveu [16], would constrain the function $S(z)$ on the entire interval $[0,1]$, hence leading to the family of invariant Galton-Watson trees. This will be explored in a follow-up paper.

We are grateful to the anonymous referee for finding a problem with the first version of this paper, caused by the gap consisting of all critical Galton-Watson measures for which the limit $L = \lim_{x\to 1-} \left(\frac{\ln g(x)}{-\ln(1-x)}\right)$ exists while the limit $S'(1) = \lim_{x\to 1-} \frac{1-S(x)}{1-x}$ does not; see Rem. 7. In particular, the referee suggested the following family of Horton prune-invariant critical Galton-Watson tree distributions different from the invariant distributions of Thm. 2. For a given probability $q_0 \in (1/2, 1)$, we let $q_1 = 0$, and

$$q_m = \frac{1}{m!A} \sum_{n\in\mathbb{Z}} B^n \rho^{nm} e^{-\rho^n} \qquad m = 2, 3, \ldots, \tag{67}$$

where $\rho = 1 - q_0$. Then, the second derivative of the generating function is equal to

$$Q''(z) = \sum_{m=2}^{\infty} m(m-1) q_m z^{m-2} = \frac{1}{A} \sum_{n\in\mathbb{Z}} B^n \rho^{2n} e^{-(1-z)\rho^n}, \qquad |z| < 1.$$

Observe that

$$Q''\big(q_0 + (1-q_0)z\big) = B^{-1}\rho^{-2} Q''(z). \tag{68}$$

Therefore, if $A > 0$ and $B \in \big((1-q_0)^{-1}, (1-q_0)^{-2}\big)$ are selected so that $Q(1) = Q'(1) = 1$, then $Q(z)$ will satisfy the invariance criterion (48). Such B is found by solving

$$\sum_{n\in\mathbb{Z}} B^n \left(1 - \rho^{n+1} - (1 + \rho^n - \rho^{n+1}) e^{-\rho^n}\right) = 0,$$

and $A = \sum_{n\in\mathbb{Z}} B^n \rho^n \left(1 - e^{-\rho^n}\right)$.

Hence,

$$Q(z) = q_0 + \frac{1}{A} \sum_{n\in\mathbb{Z}} B^n \left(e^{-(1-z)\rho^n} - (1 + \rho^n z) e^{-\rho^n}\right), \qquad |z| < 1,$$

satisfies (49) with

$$M(q_0) = (1-q_0)\big(1 - Q'(q_0)\big) = B^{-1}.$$

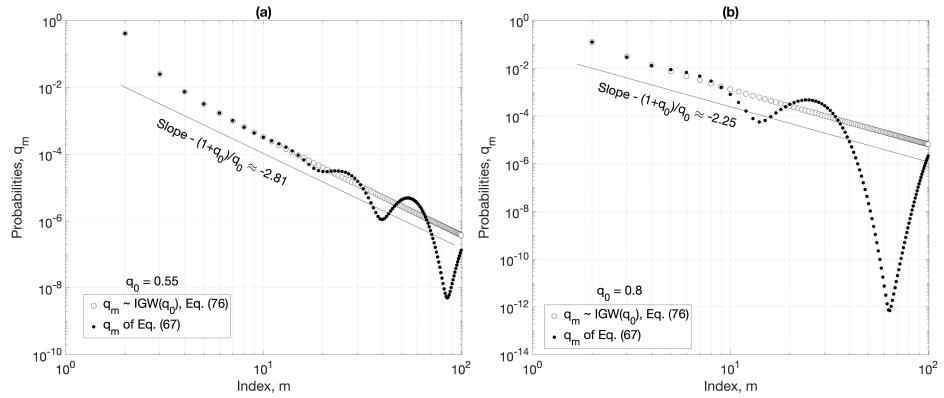

Figure 9: Horton prune-invariant measures: Illustration. Figure compares the offspring probabilities q_m, $m \geq 2$, of the invariant measure $\mathcal{IGW}(q_0)$ of Def. 8, also given in (76) (open circles), with those of the invariant measure of (67) that does not satisfy Assumption 1 (black circles). (a) $q_0 = 0.55$, (b) $q_0 = 0.8$.

Next, we show that this example belongs to the gap described in Rem. 7. Specifcally, we show that the limit L exists while the limit $S'(1)$ does not. First, Lem. 9 applies, yielding the existence of limit $L = \lim_{x\to 1-} \left(\frac{\ln g(x)}{-\ln(1-x)}\right)$. Moreover, equation (51) implies

$$L = 2 - \frac{\ln M(q_0)}{\ln(1-q_0)} = 2 + \frac{\ln B}{\ln(1-q_0)}.$$

Now, we show that the limit $S'(1) = \lim_{x\to 1-} \frac{1-S(x)}{1-x}$ does not exist, whence Assumption 1 is not satisfied. Since $\frac{1}{A}\sum_{n\in\mathbb{Z}} B^n\rho^n\left(1-e^{-\rho^n}\right) = 1$, we have for $x \in [0,1)$,

$$\frac{1-Q(x)}{1-x} = \frac{1}{A}\sum_{n\in\mathbb{Z}} B^n\rho^n\left(1-e^{-\rho^n}\right) - \frac{1-Q(x)}{1-x} = \frac{1}{A}\sum_{n\in\mathbb{Z}} B^n\rho^n\left(1-\frac{1-e^{-(1-x)\rho^n}}{(1-x)\rho^n}\right). \tag{69}$$

Also, for $x \in [0,1)$,

$$1-Q'(x) = \frac{1}{A}\sum_{n\in\mathbb{Z}} B^n\rho^n\left(1-e^{-(1-x)\rho^n}\right). \tag{70}$$

For a given $\alpha \in [0,1)$, consider a sequence $x_m = 1-\rho^{m+\alpha}$ for $m \in \mathbb{N}$, then, equation (69) implies

$$\frac{1-Q(x_m)}{1-x_m} = \frac{1}{A}\sum_{n\in\mathbb{Z}} B^n\rho^n\left(1-\frac{1-e^{-\rho^{n+m+\alpha}}}{\rho^{n+m+\alpha}}\right) = \frac{1}{A}B^{-(m+\alpha)}\rho^{-(m+\alpha)}C(\alpha),$$

where

$$C(\alpha) = \sum_{n\in\mathbb{Z}} B^{n+\alpha}\rho^{n+\alpha}\left(1 - \frac{1-e^{-\rho^{n+\alpha}}}{\rho^{n+\alpha}}\right). \tag{71}$$

Similarly, (70) implies

$$1 - Q'(x_m) = \frac{1}{A}\sum_{n\in\mathbb{Z}} B^n\rho^n\left(1-e^{-\rho^{n+m+\alpha}}\right) = \frac{1}{A}B^{-(m+\alpha)}\rho^{-(m+\alpha)}D(\alpha),$$

where

$$D(\alpha) = \sum_{n\in\mathbb{Z}} B^{n+\alpha}\rho^{n+\alpha}\left(1-e^{-\rho^{n+\alpha}}\right). \tag{72}$$

Hence,

$$\frac{1-Q(x_m)}{(1-x_m)\big(1-Q'(x_m)\big)} = \frac{C(\alpha)}{D(\alpha)} \qquad \forall m\in\mathbb{N}$$

with $C(\alpha)$ and $D(\alpha)$ as defined in (71) and (72). Thus, since $\frac{C(\alpha)}{D(\alpha)}$ is not constant for $\alpha\in[0,1)$, the limit in (26) does not exist, and the same is true about the limit $S'(1)$. Hence, in this example, Thms. 2 and 3 as currently stated do not apply since Assumption 1 does not hold.

The invariant measures (67) are in fact closely related to the $\mathcal{IGW}(q)$ measures of Def. 8. Specifically, consider a continuous integral version of (67), by selecting $q_0\in(1/2,1)$ and letting $q_1=0$ and

$$q_m = \frac{1}{m!A}\int_{-\infty}^{\infty} B^w\rho^{wm}e^{-\rho^w}\,dw \qquad m=2,3,\ldots, \tag{73}$$

where $\rho = 1-q_0$. Here too,

$$Q''(z) = \sum_{m=2}^{\infty} m(m-1)q_m x^{m-2} = \frac{1}{A}\int_{-\infty}^{\infty} B^w\rho^{2w}e^{-(1-x)\rho^w}\,dw$$

and so it satisfies (68), while $Q(1)=Q'(1)=1$ causes $Q(z)$ to satisfy (48). Analogously to (67), the constants $A>0$ and $B\in\big((1-q_0)^{-1},(1-q_0)^{-2}\big)$ are found by solving

$$\int_{-\infty}^{\infty} B^w\left(1-\rho^{w+1}-(1+\rho^w-\rho^{w+1})e^{-\rho^w}\right)dw = 0 \tag{74}$$

for B, and letting

$$A = \int_{-\infty}^{\infty} B^w\rho^w\left(1-e^{-\rho^w}\right)dw. \tag{75}$$

One can show that equation (74) has a unique solution using monotonicity of the integral on the left hand side in the equation (74) as a function of $B \in \big((1-q_0)^{-1}, (1-q_0)^{-2}\big)$. The uniqueness of B implies the uniqueness of A in (75).

In this situation, one easily shows that the limit $S'(1)$ exists. Remarkably, the unique offspring distribution q_m in (73) is the offspring distribution of $\mathcal{IGW}(q_0)$. Indeed, in equation (54) we have for $m \geq 2$,

$$q_m = \frac{(1-q_0)\Gamma(m-1/q_0)}{q_0\Gamma(2-1/q_0)\, m!} = \frac{1}{m!A} \int\limits_{-\infty}^{\infty} B^w \rho^{wm} e^{-\rho^w}\, dw \tag{76}$$

with $\rho = 1-q_0$, $A = \frac{q_0\Gamma(2-1/q_0)}{-(1-q_0)\ln(1-q_0)}$, and $B = (1-q_0)^{-1/q_0}$.

Figure 9 compares selected invariant measures $\mathcal{IGW}(q_0)$ of Def. 8, also given in (76), with the invariant measures of (67) that do not satisfy Assumption 1. Both types of measures decay in general as the power law $m^{-(1+q_0)/q_0}$, although the measures of (67) fluctuate around this general trend. The amplitude of the fluctuations (on logarithmic scale) increases with q_0. These fluctuations are related to the periodic function $\alpha(y)$ in the proof of Thm. 2; they are inevitable in the invariant measures that do not satisfy Assumption 1 (see Rem. 6).

Acknowledgements

The authors are grateful to the anonymous referee for finding a gap in the original version of the paper and suggesting an example of invariant offspring distribution with oscillatory tail (see Discussion and Fig. 9) and to Ed Waymire for multiple discussions about the topic of random self-similar trees. This research is supported by NSF awards DMS-1412557 (YK) and EAR-1723033 (IZ).

References

[1] K. B. Athreya and P. E. Ney, *Branching processes* Courier Corporation (2004).
[2] 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.
[3] L. Devroye and P. Kruszewski, *A note on the Horton-Strahler number for random trees*, Inform. Processing Lett., **56** (1994), 95–99.
[4] T. Duquesne and M. Winkel, *Growth of Lévy trees* Probab.Theory Related Fields, **139** (2007), 313–371.
[5] T. Duquesne and M. Winkel, *Hereditary tree growth and Lévy forests* Stochastic Processes and their Applications, Vol. **129**(10) (2019), 3690–3747.

[6] S. N. Evans, *Probability and real trees: Ecole d'été de probabilités de Saint-Flour* Lectures on Probability Theory and Statistics. Springer (2008).

[7] 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.

[8] H. He and M. Winkel, *Invariance principles for pruning processes of Galton-Watson trees* preprint *arXiv:1409.1014* https://arxiv.org/abs/1409.1014

[9] R. E. Horton, *Erosional development of streams and their drainage basins: Hydrophysical approach to quantitative morphology* Geol. Soc. Am. Bull., **56** (1945), 275–370.

[10] H. Kesten, *Subdiffusive behavior of random walk on a random cluster* Ann. Inst. H. Poincaré Probab. Statist., **22**(4) (1987), 425–487.

[11] Y. Kovchegov and I. Zaliapin, *Horton Law in Self-Similar Trees* Fractals, Vol. 24, No. 2 (2016) 1650017.

[12] Y. Kovchegov and I. Zaliapin, *Dynamical pruning of rooted trees with applications to 1D ballistic annihilation* Journal of Statistical Physics, **181**(2) (2020), 618–672.

[13] Y. Kovchegov and I. Zaliapin, *Random Self-Similar Trees: A mathematical theory of Horton laws* Probability Surveys, **17** (2020), 1–213.

[14] Y. Le Jan, *Superprocesses and projective limits of branching Markov process* Annales de l'IHP Probabilités et statistiques, **27**(1) (1991), 91–106.

[15] 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.

[16] J. Neveu, *Erasing a branching tree* Advances in Applied Probability, **1** (1986), 101–108.

[17] 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.

[18] 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.

[19] W. I. Newman, D. L. Turcotte, and A. M. Gabrielov, *Fractal trees with side branching* Fractals, **5** (1997), 603–614.

[20] S. D. Peckham, *New results for self-similar trees with applications to river networks* Water Resour. Res., **31** (1995), 1023–1029.

[21] A. N. Strahler, *Quantitative analysis of watershed geomorphology* Trans. Am. Geophys. Un., **38** (1957), 913–920.