AN ABSTRACT OF THE THESIS OF

Evgenia V. Chunikhina for the degree of Doctor of Philosophy in Computer Science presented on August 13, 2018.

Title: Information Theoretical Analysis of Self-Similar Trees

Abstract approved: ______________________________

Juan M. Restrepo

Tree-like patterns are ubiquitous in nature. Botanical trees, river networks, and blood systems are the most well-known examples of complex hierarchical systems met in observations. Interestingly, many of such systems exhibit statistical self-similarity. There are two main types of self-similarity: Horton self-similarity and Tokunaga self-similarity. Although there is an increased attention to the topic of trees' self-similarity, there is still a lack of theory that would allow to measure the information content embodied in vast variety of complex self-similar systems. In this work, we address the question of information quantification in tree-like self-similar structures. We start with combinatorial results that provide cardinality of several subspaces of binary trees with given structural characteristics. Then, using the notions of entropy we study the structural complexity of uniformly distributed binary trees. Furthermore, we consider classes of Horton self-similar and Tokunaga self-similar uniformly distributed trees and, using the notion of entropy rate, analyze how the structural complexity of such trees changes as the number of vertices of the tree grows to infinity.

Information Theoretical Analysis of Self-Similar Trees

by

Evgenia V. Chunikhina

A THESIS

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Presented August 13, 2018
Commencement June 2019

Doctor of Philosophy thesis of Evgenia V. Chunikhina presented on August 13, 2018.

APPROVED:

Major Professor, representing Computer Science

Head of the School of Electrical Engineering and Computer Science

Dean of the Graduate School

I understand that my thesis will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my thesis to any reader upon request.

Evgenia V. Chunikhina, Author

ACKNOWLEDGEMENTS

First, I would like to express my sincere gratitude to my advisor Prof. Juan M. Restrepo for his continuous support of my Ph.D research, motivation, and encouragement. I could not have imagined having a better advisor for my Ph.D study at Oregon State University. I would also like to thank the rest of my thesis committee: Prof. David Koslicki (my mathematics minor advisor), Prof. Bella Bose, Prof. Eric Walkingshaw, and Prof. Merrick C. Haller for their insightful comments and support. My special thanks goes to Prof. Ilya Zaliapin (University of Nevada, Reno), who provided invaluable feedback about the topic of my research. Our discussions helped me to expand my knowledge of the field. My gratitude extends to Prof. Edward C. Waymire (Oregon State University) and Dr. Scott D. Peckham (University of Colorado, Boulder) for their comments and suggestions. This work would not have been accomplished without the love and support of my family.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

LIST OF FIGURES

LIST OF FIGURES (Continued)

Figure | Page

LIST OF TABLES

Chapter 1: Introduction

Tree-like structures are among the most widely observed natural patterns, occurring in a wide variety of applied fields of study as diverse as river and drainage networks, branching structures of trees and vein structure of botanical leaves, respiratory and circulatory systems, crystals, snowflakes, lightening and register allocation for compilation of high level programming languages. In Figure 1.1 we depict several examples of tree-like structures drawn by celebrated scientists and artists. In addition, many processes like branching processes, percolation, nearest-neighbor clustering, binary search trees in computer science, tree representation of time series, evolution of an earthquake aftershock sequence, spread of a disease, spread of news or rumors on social platforms, or propagation of gene traits from parents to children can be represented as trees; see [1, 3, 8, 15, 18, 29, 30, 45, 47, 49, 50, 56, 58, 63–65, 68–73] and references therein.

Interestingly, a large number of different branching structures are statistically similar to each other and can be closely approximated by a low-dimensional statistical model [36, 45,47,65]. In other words, seemingly different dendritic structures, e.g., a human vascular tree and a Martial drainage network, have structural self-similarity and differ only in the values of the particular model parameters. There is two principal types of statistical self-similarity (Horton and Tokunaga) of tree-like structures that are associated with the Horton-Strahler ordering and Tokunaga indexing schemes [8], [71], [39]. These schemes were introduced in hydrology by Horton [29], Strahler [57, 58], and Tokunaga [62] to characterize the hierarchical structure of river networks and in computer science by Ershov [19] to address expression evaluation problems.

In particular, the Horton-Strahler number of a tree-like structure measures its branching complexity by assigning an order to each tree branch in accordance with its hierarchial importance. This measure found its practical application in many different areas, ranging from hydrology and biology to computer science, neuroscience, and financial mathematics [8, 14–16, 18, 19, 21, 22, 29, 30, 41, 44, 45, 47, 58, 64–68, 70]. In fact, Devroye and Kruszewski [15] wrote that “Horton-Strahler number occur in almost every field involving some kind of natural branching pattern”. For example, in hydrology, the Horton-Strahler

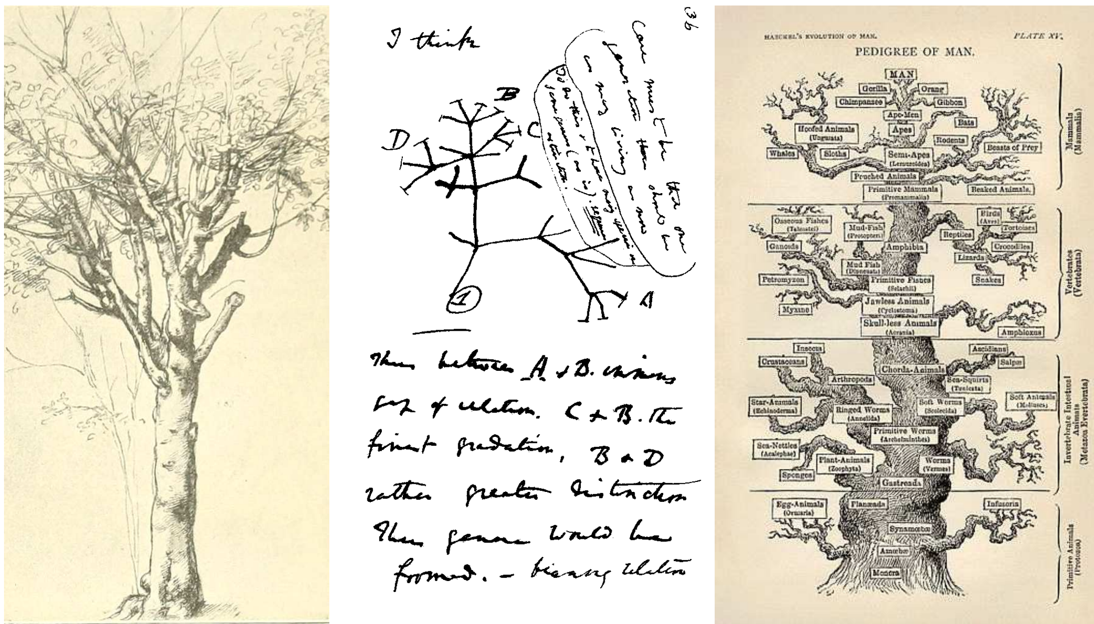

Figure 1.1: Examples of tree-like structures drawn by celebrated scientists and artists. Left: Leonardo da Vinci's drawing of a tree; Middle: Charles Darwin's sketch of an evolutionary tree from his First Notebook on Transmutation of Species (1837); Right: Ernst Haeckel's evolution tree from the The Evolution of Man (1879).

numbers were used to analyze the relationship between hydrological and geomorphical parameters and derive quantitative empirical laws for stream systems. In molecular biology, the Horton-Strahler numbers are used to analyze the secondary structures of single-stranded nucleic acids [68]. In computer science, the Horton-Strahler numbers (known as register numbers) are used in expression evaluation problems [16, 19, 44]. More precisely, recall that an arithmetic expression is stored as a binary tree. By traversing this tree, the expression is then evaluated by micro-operations, using registers. The operands are located in the external nodes of the binary tree and the operators in the internal nodes. Ershov [19] demonstrated that the minimal number of registers required to evaluate an arithmetic expression with binary operators is $1 + K$, where K is the Horton-Strahler order of the binary tree that is used to store the expression. Moreover, the minimum stack size required for a postorder traversal of a binary tree is also $1 + K$ [21, 22]. Furthermore, in the area of the computer graphics, the Horton-Strahler numbers were used to produce an image of a botanical tree; see [14, 41, 66, 67] and references therein.

In his study of river streams [28, 29], Horton observed the geometric decrease of Horton-Strahler numbers. This propriety is called Horton self-similarity and is common to a variety of hierarchical complex systems. It was verified in many different areas [25,32, 45, 47, 51, 54, 59]. Horton self-similarity implies a one-parameter Horton law with Horton exponent R. Empirical observations suggest that for a majority of natural dendritic structures the Horton exponent is in the range $(3, 6)$.

A stronger type of self-similarity, called Tokunaga self-similarity, addresses side branching statistics, i.e., the merging of branches of different orders [45, 47, 62]. The Tokunaga self-similarity implies that different levels of a hierarchical system have the same statistical structure. It is parameterized by the pair (a, c) of positive parameters. The Tokunaga self-similarity appears naturally in a wide variety of natural and simulated hierarchical systems such as river drainage networks, vein structure of botanical leaves, earthquake aftershock sequences, and nearest-neighbor clustering in Euclidean spaces. Interesting to note, that the combinatorial structure of river networks not only satisfies Horton law [45–49, 54], but can also satisfy Tokunaga self-similar model with parameters that are independent of river's geographic location [8, 18, 46, 47, 62, 74]. Tokunaga self-similarity was also established for several well-known processes that are essential for natural and computer sciences modeling: two dimensional site percolation, Shreve's random topology model, diffusion limited aggregation, level-set tree representation of white noise, level-set tree representation of random walk and Brownian motion, and Kingman's coalescent process; see [36, 39, 71] and reference therein.

In recent years there has been a growing interest in the area of trees self-similarity. The questions related to both Tokunaga and Horton self-similarity were addressed in a variety of scientific publications [8, 17, 18, 23, 36, 39, 43, 47, 60, 71]. However, there is one important aspects that has not yet been extensively addressed in scientific literature, namely, the quantification of information of different hierarchial, in particular self-similar hierarchial, structures. Although, the question of measuring the amount of information in an object is not a new one, there is still a fundamental need for a theory that would allow measuring the information content in a vast variety of structures, especially physical structures [7]. Brooks [7] considers "this missing metric to be the most fundamental gap in the theoretical underpinnings of information science and of computer science." Brooks also noted that this measure will be closely related to the theory of information, especially, to the concept of information entropy, introduced by Shannon in his famous work "A mathematical theory of communication" [53]. The notion of entropy originates from the concept of disorder in thermodynamics and statistical mechanics [20,52] and has been successfully connected to the concept of information [2,6,9,20,24,26,27,33,33–35].

In this work we provide information theoretical analysis of self-similar tree-like structures. We focus our attention on a space of finite unlabeled rooted planted binary plane trees with no edge lengths and study the statistical properties and structural complexity of several subspaces of this space. Kolmogorov [34] noted that the basis of the information theory has in its core finite combinatorial nature. Thus, we start with the combinatorial analysis of several subspaces of uniformly distributed planted trees with different structural features, such as the number of vertices in a tree, the Horton-Strahler order of a tree, the Horton-Strahler numbers, and the Tokunaga numbers. Next, we determine the number of trees in each subspace. We extend these results to subspaces of trees with a ghost edge and to the subspaces of stemless trees. Then we define entropy and entropy rate measures that provide useful descriptions of structural complexity of growing tree models. Specifically, we use the notion of entropy to study the structural complexity of uniformly distributed trees with N vertices. Note that the uniform distribution on the space of planted binary plane trees with N vertices is different from the uniform distribution over the space of planted binary non-plane trees, induced by the critical binary Galton-Watson process, conditioned on having N vertices [50]. Furthermore, we introduce the notion of entropy rate, that describes the growth of the entropy as the number of tree vertices grows to infinity, and use it to analyze structural complexity of sequences of Horton and Tokunaga self-similar trees. In particular, we consider subspaces of planted binary trees that satisfy Horton law with a given Horton exponent R and find a closed-form formula for its entropy rate. Also, we find closed-form formula for the entropy rate of a sequence of planted binary trees that satisfy Tokunaga law with parameters (a, c). Detailed examination of both entropy rates allows us to address the question of information quantification in self-similar tree-like structures. We extend the investigation of the behavior of entropy rates by rewriting the obtained formulae using the notion of I-divergence (or generalized Kullback-Leibler divergence). The information theoretical analysis and results of this work can be applied to a variety of questions related to the combinatorial structure of self-similar trees.

Chapter 2: Preliminaries

We begin with the statement of important definitions that will be used throughout this work.

2.1 Trees: definitions and set up

Definition 1. A *tree* is an acyclic connected graph.

Definition 2. A tree with one vertex labeled as the *root* is called a *rooted tree*. Presence of the root in a tree provides a natural child-parent relation between the neighboring vertices. More precisely, the *parent* of a vertex is the vertex connected to it on the path down, towards the root and the *child* of a vertex is the vertex connected to it on the path up, away from the root.

Note that a vertex can have more than one child and every vertex except the root has a unique parent and a unique *parental edge* that connects a vertex to its parent.

Definition 3. A *leaf* is a vertex with no children.

Definition 4. The *degree* of a vertex is the number of edges incident to a vertex.

Definition 5. A tree is called *binary* tree if each vertex has at most two children.

Definition 6. The operation of *series reduction* removes each degree-two vertex of a binary tree by merging its adjacent edges into one. Series reduction turns a rooted binary tree into a *reduced* rooted binary tree.

Definition 7. A *full* binary tree is a tree in which every vertex has either zero or two children.

Definition 8. A *perfect* binary tree is a binary tree in which all interior vertices have two children and all leaves have the same depth.

In Figure 2.1 (a) and (b) we depict full and perfect binary trees, respectively.

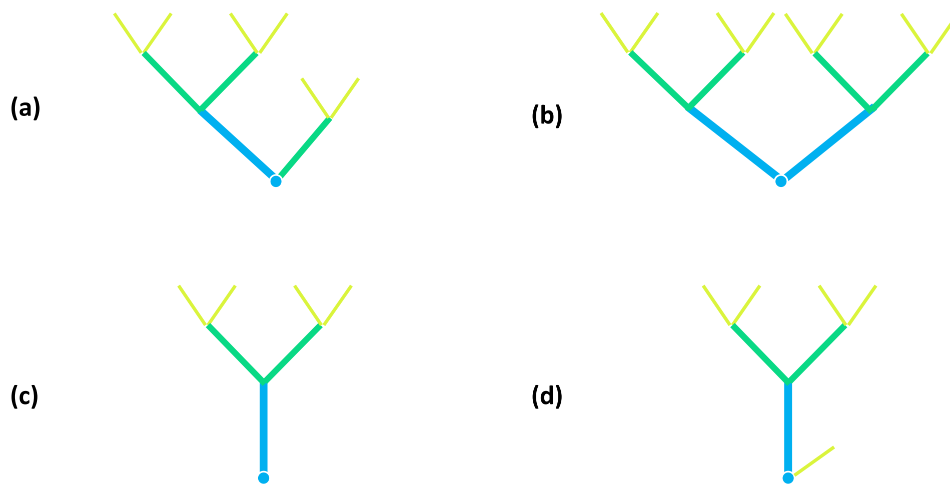

Figure 2.1: An example of a full binary tree (a) and a perfect binary tree (b). The tree (c) is a planted binary plane tree. The root node is depicted at the bottom of the tree and has degree one. The tree (c) has $n = 4$ leaves, $2n = 8$ vertices, and $2n-1 = 7$ edges. The tree (d) is not a planted binary plane tree since its root node has degree two.

Definition 9. A *plane* tree is a rooted tree with a specified ordering for the children of each vertex. This ordering is equivalent to an embedding of the tree in the plane and provides a natural left and right orientations for the children.

Definition 10. A *planted binary plane tree* is a rooted tree such that its root has degree one and every other vertex is either a leaf or an *internal vertex* of degree three (Please, see section 7.2 in [50]). Every planted tree is a reduced tree.

Definition 11. We denote a *stem* to be the unique edge that connects the root vertex with its only child. Assuming the tree grows from the root vertex upwards, the root vertex is located at the bottom of the stem.

Remark 1. *Every planted binary plane tree with n leaves has $2n-1$ edges and even number of vertices $2n$, such that $n-1$ of them are internal.*

In Figure 2.1 (c) we depict a planted binary plane tree. The tree has one root vertex, $n = 4$ leaves, $2n = 8$ vertices ($n - 1 = 3$ of them are internal), and $2n - 1 = 7$ edges. The root vertex of this tree is depicted at the bottom of the tree and has degree one. The tree depicted in Figure 2.1 (d) is not a planted binary plane tree since its root node has degree two.

Definition 12. Let $\mathcal{T}$ be a space of finite unlabeled rooted planted binary plane trees with no edge length.

In this work, we consider trees from the space $\mathcal{T}$. All the trees that we refer to are assumed to be from the space $\mathcal{T}$, unless stated otherwise.

Definition 13. Let $\mathscr{T}_N \subset \mathcal{T}$ be the subspace of all planted binary plane trees with n leaves and $N = 2n$ vertices.

Definition 14. Let the *cardinality* of a set be the measure of the number of elements in the set. We denote the cardinality of a set S to be $|S|$.

Remark 2. *The number of possible configurations of a planted binary plane tree with N vertices is given by the $(n-1)$th Catalan number $\mathcal{C}_{n-1}$ [50] as follows*

$$|\mathscr{T}_N| = \mathcal{C}_{n-1} = \frac{1}{n}\binom{2n-2}{n-1} = \frac{(2n-2)!}{n!\,(n-1)!},$$

where $n = \frac{N}{2}$ and $\binom{n}{k} = \frac{n!}{k!(n-k)!}$.

2.2 Horton-Strahler ordering scheme

In this section we introduce the Horton-Strahler ordering scheme.

Definition 15. Consider an arbitrary binary tree. The Horton-Strahler ordering of the vertices and branches in a binary tree is performed, from the leaves to the root node, by hierarchical counting [8, 29, 36, 45, 47, 58] as follows

- each leaf is assigned order 1;
- an internal vertex with children of orders i and j is assigned the order

$$k = \max(i, j) + \delta_{ij},$$

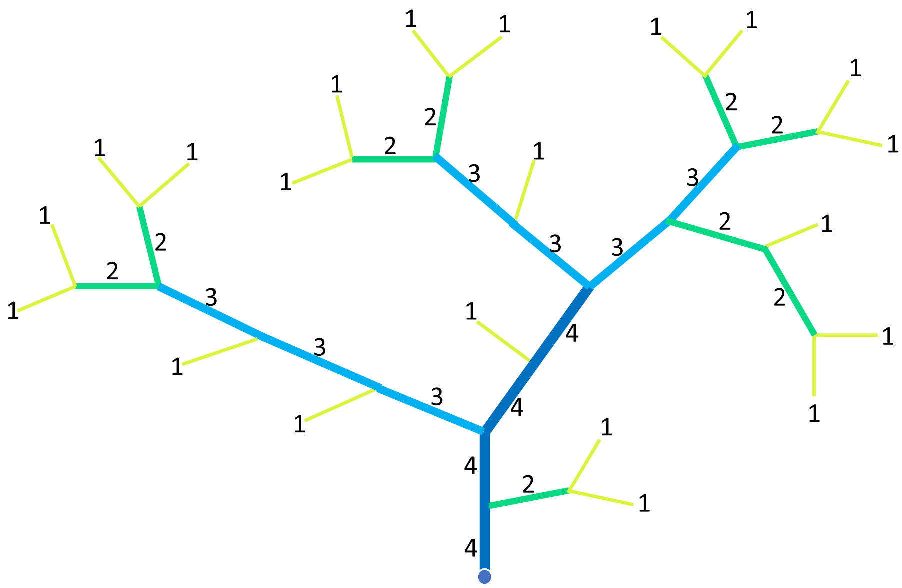

Figure 2.2: An example of Horton-Strahler ordering of a tree from the subspace $\mathcal{T}$. The tree has $N_4 = 1, N_3 = 3, N_2 = 8$, and $N_1 = 21$. The number of vertices is $N = 2N_1 = 42$. Branches of order 4 are depicted in indigo, branches of order 3 in blue, branches of order 2 in green, and branches of order 1 in pear. The order of the tree is $K = 4$. The tree has only one branch of order 4, although it consists of four edges. The stem of the tree is also of order 4 and is a part of the branch of order 4. The root node is depicted at the bottom of the tree and has order 4.

where δ_{ij} is the Kronecker's delta, defined as

$$\delta_{ij} = \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{if } i \neq j; \end{cases}$$

- the parental edge of a vertex has the same order as a vertex;
- a *branch* of order i is a sequence of neighboring vertices of order i together with their corresponding parental edges.

Definition 16. The order K of a non-empty binary tree is defined as the maximal order of its vertices. The order of an empty tree is $K = 0$.

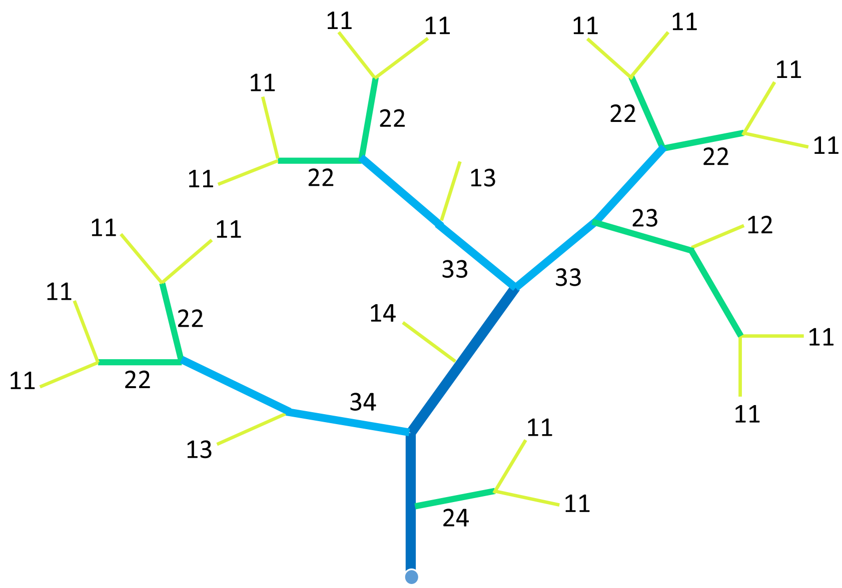

Figure 2.3: Example of Tokunaga ordering of a planted binary plane tree of order $K = 4$. The tree has $N_{1,2} = 1, N_{1,3} = 2, N_{1,4} = 1, N_{2,3} = 1, N_{2,4} = 1, N_{3,4} = 1$ and $N_4 = 1, N_3 = 3, N_2 = 8$, and $N_1 = 20$. The number of vertices is $N = 2N_1 = 40$. Branches of order 4 are depicted in indigo, branches of order 3 in blue, branches of order 2 in green, and branches of order 1 in pear.

Definition 17. The *Horton – Strahler ordering* of a non-empty binary tree is a set of numbers $N_i \geq 0$, $i = \overline{1, K}$, where each N_i represents the number of branches of order i.

Recall that in this work we consider trees from space $\mathcal{T}$, i.e., finite unlabeled rooted planted binary plane trees with no edge length. For such trees there are two important observations. First, notice that in order to have a branch of order $i + 1$ the binary tree needs to have at least two branches of order i. Therefore, $\forall i = \overline{1, K-1}$ the Horton-Strahler numbers $N_1, N_2, \cdots, N_K$ should satisfy inequality

$$N_i \geq 2N_{i+1}.$$

Moreover, if we assume a tree is of the order $K > 0$, then the tree has only one branch of order K, i.e., $N_K = 1$. The stem of the tree is also of order K and is a part of the branch of order K. Thus, when we consider trees from the space $\mathcal{T}$ we need to make sure that the Horton-Strahler numbers satisfy the two conditions described above.

Definition 18. The sequences $N_1, N_2, \cdots, N_K$ of Horton-Strahler numbers is called an *admissible sequence*, if $\forall i = \overline{1, K-1}$ N_i satisfies the following two conditions:

1. $N_i \geq 2N_{i+1}$;

2. $N_K = 1$.

Since we are working with finite unlabeled rooted planted binary plane trees with no edge length, we consider only *admissible sequences* of Horton-Strahler numbers. To simplify the notation, we call an admissible sequence $N_1, N_2, \cdots, N_K$ a set of Horton-Strahler numbers.

To illustrate the Horton-Strahler ordering of a tree from the space $\mathcal{T}$ consider an example in Figure 2.2.

2.3 Tokunaga indexing scheme

The Tokunaga indexing [45,47,62] is based on the Horton-Strahler ordering scheme and describes the merging between branches of different orders, called the $side - branching$.

Definition 19. Consider a finite tree of order K. Let N_j, $j \leq K$ be the number of branches of order j. We denote $\tau^l_{i,j}$ $1 \leq l \leq N_j$, $1 \leq i < j \leq K$ to be the number of branches of order i that join the non-terminal vertices of the lth branch of order j. Then $N_{i,j} = \sum\limits_l \tau^l_{i,j}$, where $i < j$ is the total number of branches of order i that merge with the branches of order j. The Tokunaga index $T_{i,j}$ is the average number of branches of order $i < j$ per branch of order j in a finite tree of order $K \geq j$, i.e.,

$$T_{i,j} = \frac{N_{i,j}}{N_j}.$$

We also denote $N_{i,i}$ be the total number of branches of order i that merge with other branches of order i. Note that for a planted binary tree $N_{i,i} = 2N_{i+1}$. We also define additional Tokunaga indices

$$T_{i,i} = \frac{N_{i,i}}{N_{i+1}} = 2.$$

The set of Tokunaga numbers $N_{i,j}$, $i \in [1, K-1]$, $j \in [1, K]$, $i \leq j$ provides a complete statistical description of the branching structure of a finite tree of order K.
In Figure 2.3 we depict an example of the Tokunaga ordering of a finite tree of order 4. Note that $\forall j \in [1, K]$, $N_j = \sum_{i=j}^{K} N_{j,i}$. In other words, the total number of branches of order j is the sum of the number of branches of order j that merge with branches of order j, the number of branches of order j that merge with branches of order $j+1$, and so on until $j = K$. Similarly, we conclude that the number of nodes in a planted binary tree of order K is $N = 2N_1 = 2\sum_{i=1}^{K} N_{1,i}$.

Chapter 3: Combinatorial results

In this chapter we consider several subspaces of binary trees and present combinatorial results on the cardinality of those subspaces.
In the next section we introduce the subspace of planted binary trees with a given set of Horton-Strahler numbers and explore its cardinality.

3.1 Cardinality of subspace $\mathscr{T}_{N_1,N_2,\cdots,N_K}$

Definition 20. Let $\mathscr{T}_{N_1,N_2,\cdots,N_K} \subset \mathcal{T}$ be the subspace of all finite unlabeled rooted planted binary plane trees with no edge length and with a particular set of Horton-Strahler numbers $N_1, N_2, \cdots, N_K$.

Note that $\mathscr{T}_{N_1,N_2,\cdots,N_K} \subset \mathscr{T}_N$, where $N = 2N_1$.
The following result evaluates the cardinality of the subspace $\mathscr{T}_{N_1,N_2,\cdots,N_K}$.

Theorem 1. *The number of trees of order K with a particular set of Horton-Strahler numbers $N_1, N_2, \cdots, N_K$ and with $N = 2N_1$ vertices is given by the following formula*

$$|\mathscr{T}_{N_1,N_2,\cdots,N_K}| = 2^{N_1-1-\sum_{i=1}^{K-1} N_{i+1}} \prod_{i=1}^{K-1} \binom{N_i - 2}{2N_{i+1} - 2}, \tag{3.1}$$

where $\binom{n}{k} = \frac{n!}{k!(n-k)!}$.

Proof. We prove this theorem by providing a method to construct and count trees with fixed Horton-Strahler numbers. We start by introducing a few helpful definitions.
For a given tree we define the *main frame* (also know as a skeleton in related publications) to be the minimal subtree of the same order with the same root. The main frame can be obtained by removing (with series reduction) the maximal number of branches, so that to preserve the order of the tree. Each branch of order $i+1$ is obtained by merging two *necessary frames* of order i, $\forall i = \overline{1, K-1}$. All other frames are *extra frames*. Note that each frame of order j is a perfect planted binary plane tree and has one branch of order j, two branches of order $j-1$ and so on and 2^{j-1} branches of order 1.

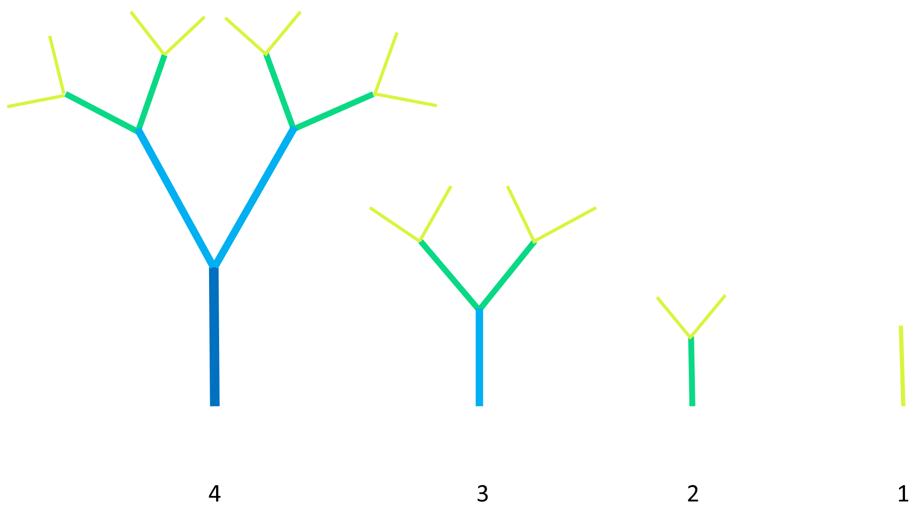

Figure 3.1: Frames of order 4, 3, 2, and 1, respectively. Branches of order 4 are depicted in indigo, branches of order 3 in blue, branches of order 2 in green, and branches of order 1 in pear.

Thus, given a set of Horton-Strahler numbers N_i such that $N_i \geq 2N_{i+1}$, $\forall i = \overline{1, K-1}$, the number of necessary frames of order i is $L_i = 2N_{i+1}$ and the number of extra frames of order i is $M_i = N_i - L_i$.

To illustrate the notion of necessary and extra frames, consider a planted binary plane tree of order $K = 3$ depicted in Figure 3.2.

The Horton-Strahler numbers of this tree are $N_3 = 1$, $N_2 = 3$, and $N_1 = 7$. The tree consists of one main frame of order 3 (that consists of two necessary frames of order 2 and four necessary frames of order 1), one extra frame of order 2 (that consists of two necessary frames of order 1), and one extra frame of order 1. Thus, this tree of order $S = 3$ is constructed by attaching all extra frames of orders 2 and 1 to the main frame of order 3. Note that each extra frame is attached to the branch of the order that is higher than the order of the extra frame. Both extra frames of orders 1 and 2 are attached to the branch of order 3. Generally, each extra frame of order i is attached to the branch of higher order $j > i$ because all other ways to attach extra frames will result in either non-binary trees or in binary trees with incorrect Horton-Strahler numbers and incorrect

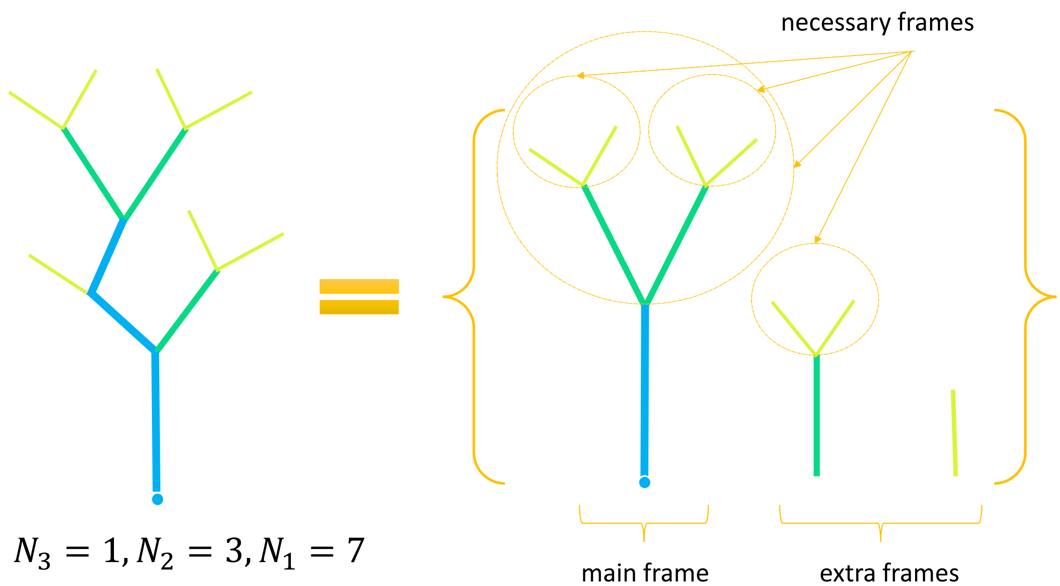

Figure 3.2: An example of a tree of order $K = 3$ with $N_3 = 1, N_2 = 3$, $N_1 = 7$. There are one main frame of order 3, two necessary frames of order 2 ($L_2 = 2$), six necessary frames of order 1 ($L_1 = 6$), one extra frame of order 2 ($M_2 = 1$), and one extra frame of order 1 ($M_1 = 1$). This tree is constructed by attaching extra frames of orders 2 and 1 to the main frame of order 3. Branches of order 3 are depicted in blue, branches of order 2 in green, and branches of order 1 in pear.

number of vertices.

We will illustrate this by considering all the cases separately.

First, notice that no extra frame can be attached to the root node of the main frame, since the root node should be of degree 1. Also, extra frames can not be attached to the internal vertices of the main frame, since it will result in a non-binary tree. Moreover, no extra frame can be attached to any of the leaf vertices, since it will result in a tree with incorrect number of branches of order i: instead of N_i the tree will have $N_i - 1$ branches of order i. In both examples in Figure 3.3 the resulting trees have incorrect Horton-Strahler numbers.

Moreover, attaching an extra frame of order i to the branch of lower order $j < i$ will result in a tree with incorrect number of branches of order i: instead of N_i the tree will

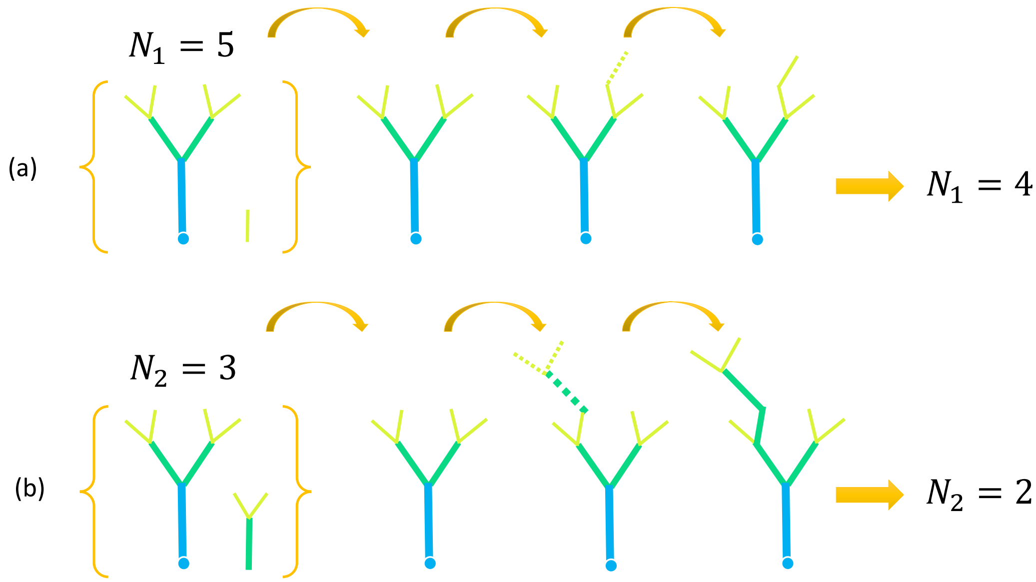

Figure 3.3: An example of how attaching extra frames to the leaf vertices results in the trees with incorrect Horton-Strahler numbers and incorrect number of vertices. In (a) the resulting tree has $N_1 = 4$ instead of $N_1 = 5$ and 9 vertices instead of 10. In (b) the resulting tree has $N_2 = 2$ instead of $N_2 = 3$ and 11 vertices instead of 12. Moreover, in both cases, the resulting trees are not planted binary trees, since one of the internal vertices has degree 2. Branches of order 3 are depicted in blue, branches of order 2 in green, and branches of order 1 in pear.

have $N_i - 1$ branches of order i. In Figure 3.4 (a) we depict an example of attaching one extra frame of order 2 to one of the branches of order 1 of the main frame. The resulting tree has incorrect Horton-Strahler number $N_2 = 2$, instead of $N_2 = 3$.

Suppose now we attach an extra frame of order i to the branch of higher order $j > i$. The resulting tree will have correct Horton-Strahler numbers and correct number of vertices, e.g., Figure 3.4 (b). Finally, suppose we attach an extra frame of order i to the branch of the same order $j = i$. In this case, all resulting trees will be redundant to the trees constructed by attaching an extra frame of order i to the branches of higher order. In Figure 3.5 (a) we depict an example of attaching an extra frame of order 2 to the branch of order 2 of the main frame. The resulting tree is identical to the tree depicted in

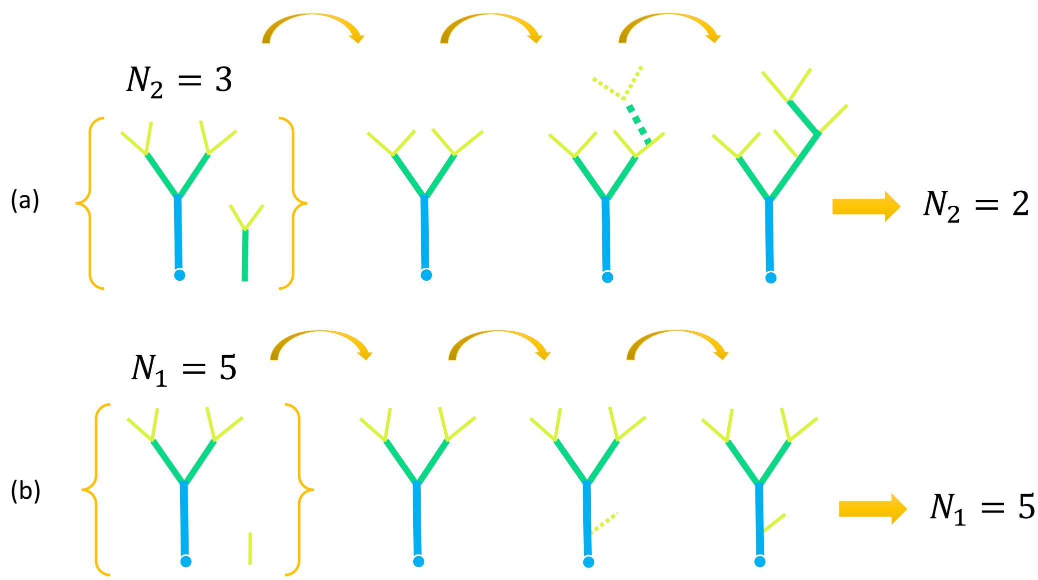

Figure 3.4: An example of attaching an extra frame to the branch of (a) a lower and (b) a higher order. In (a) we attach an extra frame of order 2 to the branch of order 1. The resulting tree has incorrect Horton-Strahler number $N_2 = 2$, instead of $N_2 = 3$. In (b) we attach an extra frame of order 1 to the branch of order 3. The resulting tree has correct Horton-Strahler numbers. Branches of order 3 are depicted in blue, branches of order 2 in green, and branches of order 1 in pear.

Figure 3.5 (b), which is obtained by attaching an extra frame of order 2 to the branch of order 3 of the main frame. Therefore, to find the total number of trees of order K with $N = 2N_1$ vertices and a given set of Horton-Strahler numbers $N_1, N_2, \cdots, N_K$, we should start with a main frame of order K and then count all possible ways we can attach all extra frames to the branches of higher orders, starting with the extra frames of order $K-1$, followed by the extra frames of order $K-2$, and so on. Extra frames of order 1 will be attached at the end.

We start with the main frame of order K. Denote $T_{K-1 \to K}$ to be the number of trees we obtain by attaching M_{K-1} extra frames of order $K-1$ to one branch of order K of the main frame. In general, the number of ways to place n identical objects into k different

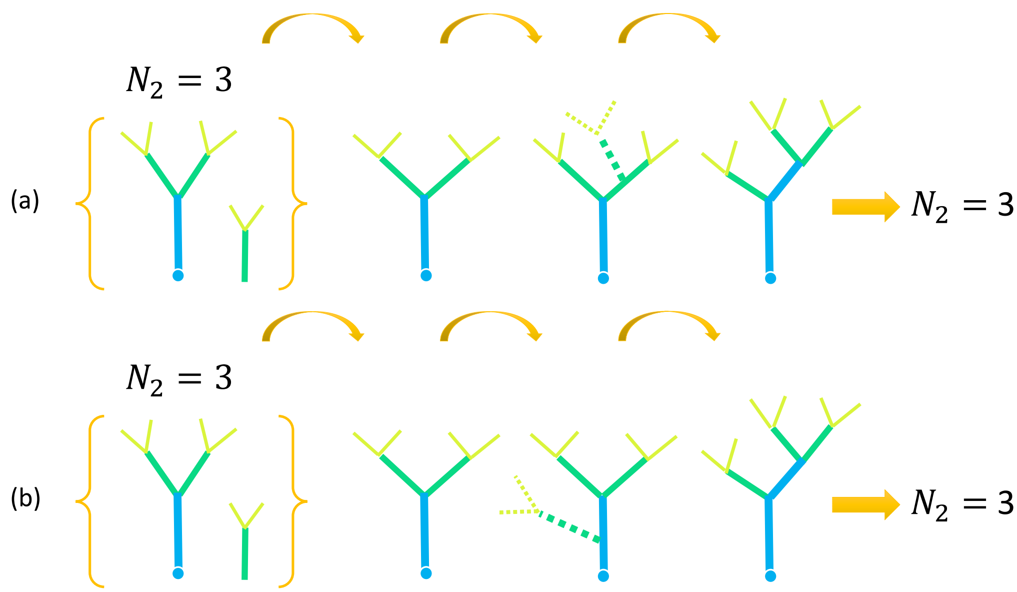

Figure 3.5: An example of attaching an extra frame to the branch of (a) the same and (b) a higher order. In (a) we attach an extra frame of order 2 to the branch of order 2. In (b) we attach an extra frame of order 2 to the branch of order 3. The resulting trees are identical and have correct Horton-Strahler numbers. Branches of order 3 are depicted in blue, branches of order 2 in green, and branches of order 1 in pear.

positions is given by the formula

$$\binom{n+k-1}{k-1} = \frac{(n+k-1)!}{(k-1)!n!}. \tag{3.2}$$

We also need to take into account that each extra frame can be attached to the middle point of a branch either form the left or from the right. In Figure 3.6 we depict an example of how attaching extra frames from the left or from the right result in different trees.

Therefore, $T_{K-1\to K}$ can be calculated as follows

$$T_{K-1\to K} = 2^{M_{K-1}}\binom{N_K + M_{K-1} - 1}{N_K - 1} = 2^{N_{K-1}-2N_K}\binom{N_{K-1}-2}{N_K - 1}. \tag{3.3}$$

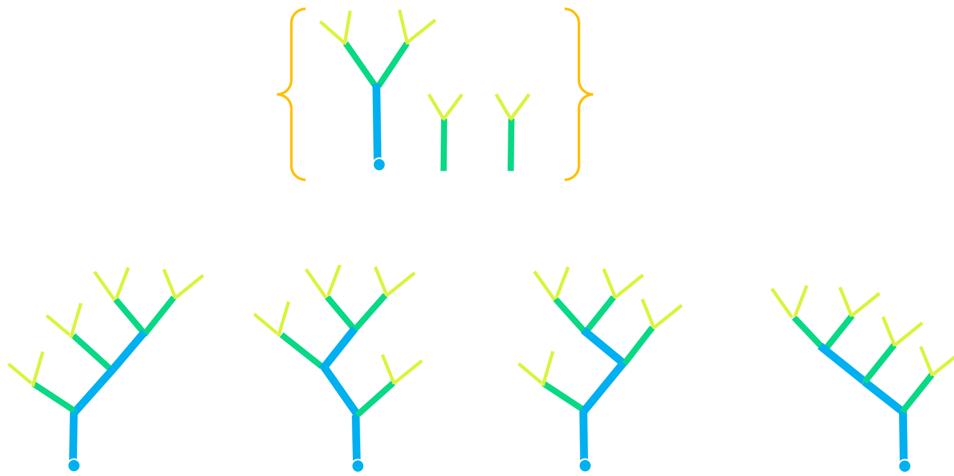

Figure 3.6: An example of placing $M_2 = 2$ extra frames of order 2 onto the main frame of order 3, in particular on the branch of order 3. Depending on the side of placement (right or left) we obtain $2^{M_2} = 4$ different trees. Branches of order 3 are depicted in blue, branches of order 2 in green, and branches of order 1 in pear.

Note that the term $T_{K-1\to K}$ can be also rewritten as

$$\begin{aligned} T_{K-1\to K} &= 2^{M_{K-1}}\binom{N_K + M_{K-1} - 1}{N_K - 1} = 2^{N_{K-1}-2N_K}\binom{N_{K-1} - N_K - 1}{N_K - 1} \\ &= 2^{N_{K-1}-2}\binom{N_{K-1} - 2}{0} = 2^{N_{K-1}-2}, \end{aligned}$$

where the last two equations follow from the fact that $N_K = 1$ and $\binom{N_{K-1}-2}{0} = 1$. However, for convenience, we leave the term $T_{K-1\to K}$ in more general form as in equation 3.3.

Note that when we attach M_{K-1} extra frames of order $K-1$ to one branch of order K, we brake the branch of order K into $N_K + M_{K-1}$ edges. The Horton-Strahler number does not change: there is still one branch of order K, i.e., $N_K = 1$, but it consists of $N_K + M_{K-1}$ edges of order K. For example in Figure 3.5 (b), attachment of an extra

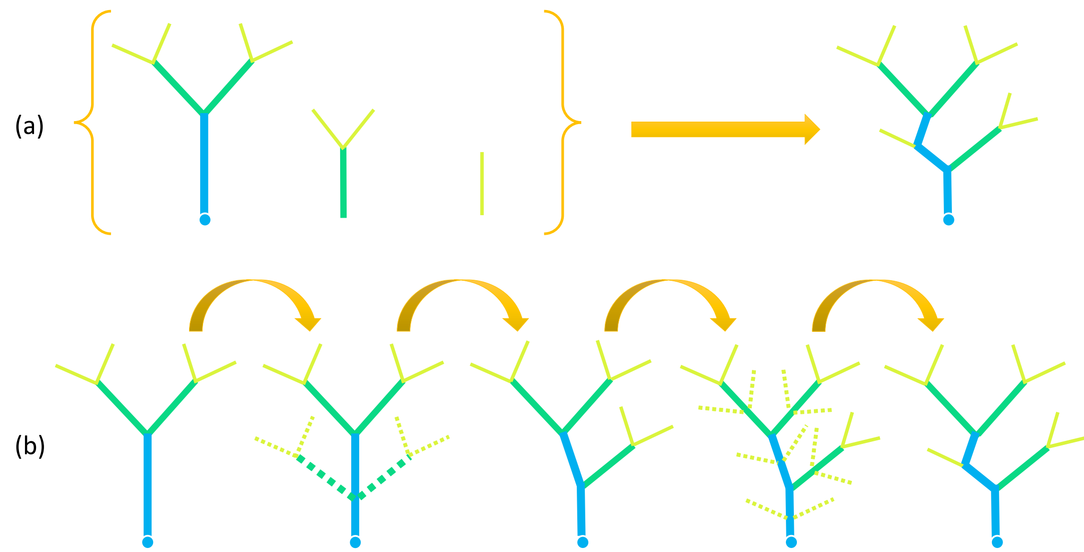

Figure 3.7: An example of constructing a tree of order 3 with $N_1 = 7$, $N_2 = 3$, and $N_3 = 1$. We start with a main frame of order 3 and attach all extra frames to the branches of higher orders, starting with the extra frames of order 2, followed by the extra frames of order 1. Extra frames depicted with dotted lines indicate possible placements of those frames. Branches of order 3 are depicted in blue, branches of order 2 in green, and branches of order 1 in pear.

frame of order 2 to the branch of order 3, broke the branch of order 3 into two edges.

Next, denote $T_{K-2\to K-1,K}$ to be the the number of different trees we obtain by attaching extra frames of order $K-2$ to branches of higher orders $K-1$ and K. There are $N_K + M_{K-1}$ edges of order K and N_{K-1} branches of order $K-1$. Thus, there are $k = N_K + M_{K-1} + N_{K-1} = 2N_{K-1} - 1$ edges of orders K and $K-1$ to which we can attach extra frames of order $K-2$. By using formula (3.2) and considering that each extra frame can be attached either from the left or from the right we obtain $T_{K-2\to K-1,K}$ as follows

$$T_{K-2\to K-1,K} \quad = \quad 2^{N_{K-2}-2N_{K-1}} \binom{N_{K-2}-2}{2N_{K-1}-2}. \tag{3.4}$$

Consider now an intermediate step. Let $T_{i\to i+1,\cdots,K}$ be the number of trees that we obtain by attaching M_i extra frames of order i to the branches of orders $i+1, i+2, \cdots, K$. Note that there are now

$$\begin{aligned} k &= N_K + M_{K-1} + N_{K-1} + M_{K-2} + N_{K-2} + \cdots + M_{i+1} + N_{i+1} \\ &= N_K + N_{K-1} - 2N_K + N_{K-1} + N_{K-2} - 2N_{K-1} + N_{K-2} \\ &+ \cdots + N_{i+1} - 2N_{i+2} + N_{i+1} \\ &= 2N_{i+1} - N_K = 2N_{i+1} - 1 \end{aligned} \tag{3.5}$$

edges of orders $i+1, i+2, \cdots, K$, to which we can attach extra frames of order i. Thus, using formula (3.2) and considering that each extra frame can be attached either from the left or from the right we obtain $T_{i\to i+1,\cdots,K}$ as follows

$$T_{i\to i+1,\cdots,K} = 2^{N_i - 2N_{i+1}} \binom{N_i - 2}{2N_{i+1} - 2}. \tag{3.6}$$

Equation (3.6) provides a general formula for the terms $T_{i\to i+1,\cdots,K}$, $\forall i = \overline{K-2, 1}$. Note now that for every possible attachment of extra frames of order $i+1$ there are $T_{i\to i+1,\cdots,K}$ possible attachments of extra frames of order i, $\forall i = \overline{K-1, 1}$. Thus, using the multiplication principle of combinatorics, we obtain the total number of trees of order K with a particular set of Horton-Strahler numbers $N_1, N_2, \cdots, N_K$ and $N = 2N_1$ vertices as follows

$$\begin{aligned} |\mathscr{T}_{N_1,N_2,\cdots N_K}| &= \prod_{i=K-1}^{1} T_{i\to i+1,\cdots,K} \\ &= \prod_{i=K-1}^{1} 2^{N_i - 2N_{i+1}} \binom{N_i - 2}{2N_{i+1} - 2} \\ &= 2^{\sum_{i=1}^{K-1}(N_i - 2N_{i+1})} \prod_{i=1}^{K-1} \binom{N_i - 2}{2N_{i+1} - 2} \\ &= 2^{N_1 - 1 - \sum_{i=1}^{K-1} N_{i+1}} \prod_{i=1}^{K-1} \binom{N_i - 2}{2N_{i+1} - 2}. \end{aligned}$$

□

Figure 3.8: An example of a subspace $\mathscr{T}_{7,3,1}$ of 20 planted binary plane trees with $N_3 = 1, N_2 = 3,\ N_1 = 7$ and $N = 2N_1 = 14$ vertices. Branches of order 3 are depicted in blue, branches of order 2 in green, and branches of order 1 in pear.

The above result was first published in a different form by Shreve [54].

3.1.1 Examples

Consider a subspace $\mathscr{T}_{7,3,1}$ of finite unlabeled rooted planted binary plane trees of order $K = 3$ with Horton-Strahler numbers $N_3 = 1, N_2 = 3,\ N_1 = 7$ and $N = 2N_1 = 14$ vertices. Using formula (3.1), we can find the cardinality of this subspace as follows

$$\begin{aligned} |\mathscr{T}_{7,3,1}| &= 2^{N_1-1-\sum_{i=1}^{K-1} N_{i+1}} \prod_{i=1}^{K-1} \binom{N_i - 2}{2N_{i+1} - 2} \\ &= 2^{7-1-3-1} \binom{7-2}{6-2} \binom{3-2}{2-2} \\ &= 2^2 \frac{5!}{4!1!} = 4 \times 5 = 20. \end{aligned}$$

Table 3.1: Each entry in this table represents the number of trees for different sets of Horton-Strahler numbers: for the first two columns - $|\mathscr{T}_{N_1,N_2,N_3}|$ and for the second two columns - $|\mathscr{T}_{N_1,N_2,N_3,N_4}|$. For the second two columns $N_4 = 1$. The last row has the number of trees, when $N_1 = 30$.

N_1	$N_2 = 2, N_3 = 1$	$N_2 = 3, N_3 = 1$	$N_2 = 4, N_3 = 2$	$N_2 = 5, N_3 = 2$
4	1			
5	6			
6	24	2		
7	80	20		
8	240	120	1	
9	672	560	14	
10	1792	2240	112	6
11	4608	8064	672	108
12	11520	26880	3360	1080
30	25,367,150,592	687,026,995,200	1,580,162,088,960	19,554,505,850,880

In Figure 3.8, we depict all 20 trees of order 3 from the subspace $\mathscr{T}_{7,3,1}$. Furthermore, in Table 3.1 we present the number of trees for different sets of Horton-Strahler numbers. Note that, although done for trees form space $\mathcal{T}$, the results of Theorem 1 can be applied to the stemless trees and to the trees with a *ghost edge* [8,71]. Both extensions are present below as corollaries.

3.1.2 Trees with a ghost edge

Definition 21. We define a *ghost edge* to be a parental edge of the root node. The tree with a ghost edge has N_1 leaves and $N = 2N_1 - 1$ vertices.

In Figure 3.9 (a) we depict an example of a tree with a ghost edge. The tree has order 4 and $N_1 = 10$, $N_2 = 4$, $N_3 = 2$, and $N_4 = 1$. Notice that the root node is located at the top of the ghost edge. There is no node at the bottom of the ghost edge.

Corollary 1. *Let $\tilde{\mathscr{T}}_{N_1,N_2,\cdots,N_K}$ be a subspace of finite unlabeled rooted binary plane trees with a ghost edge, a particular set of Horton-Strahler numbers $N_1, N_2, \cdots, N_K$, and $N = 2N_1 - 1$ vertices. Then the cardinality of this subspace is equal to the cardinality of* *the subspace* $\mathscr{T}_{N_1,N_2,\cdots,N_K}$*, i.e.,*

$$|\tilde{\mathscr{T}}_{N_1,N_2,\cdots,N_K}| = |\mathscr{T}_{N_1,N_2,\cdots,N_K}| = 2^{N_1-1-\sum_{i=1}^{K-1} N_{i+1}} \prod_{i=1}^{K-1} \binom{N_i-2}{2N_{i+1}-2}, \tag{3.7}$$

where $\binom{n}{k} = \frac{n!}{k!(n-k)!}$.

Proof. The trees from the subspace $\tilde{\mathscr{T}}_{N_1,N_2,\cdots,N_K}$ differ from the trees from the subspace $\mathscr{T}_{N_1,N_2,\cdots,N_K}$ in the number of vertices and in a position of the root vertex. The trees from the subspace $\tilde{\mathscr{T}}_{N_1,N_2,\cdots,N_K}$ have $2N_1-1$ vertices and trees from the subspace $\tilde{\mathscr{T}}_{N_1,N_2,\cdots,N_K}$ have $2N_1$ vertices. The root vertex in the the trees from the subspace $\tilde{\mathscr{T}}_{N_1,N_2,\cdots,N_K}$ is located at the top of the ghost edge and the root vertex in the trees from the subspace $\tilde{\mathscr{T}}_{N_1,N_2,\cdots,N_K}$ is located at the bottom of the stem. Therefore, one can obtain result 3.7 by following the proof of the Theorem 1 exactly, while treating a ghost edge as a stem and reassigning the root node to the top of the last link of the ghost edge. □

3.1.3 Stemless trees

In some applications it is convenient to consider binary trees without a stem or a ghost edge. In Figure 3.9 (b) we depict an example of a binary stemless tree. The tree is of order 3 and has Horton-Strahler numbers $N_1 = 11$, $N_2 = 5$, $N_3 = 2$ and $2N_1 - 1$ vertices. Note that the root node has degree two.

Corollary 2. *Let* $\hat{\mathscr{T}}_{N_1,N_2,\cdots,N_K}$ *be a subspace of finite unlabeled rooted binary plane trees with no edge length, such that the trees have a particular set of Horton-Strahler numbers* $N_1, N_2, \cdots, N_K$*,* $N = 2N_1 - 1$ *vertices, and a root node of degree two. Then the cardinality of this subspace is*

$$|\hat{\mathscr{T}}_{N_1,N_2,\cdots,N_K}| = 2^{N_1-2-\sum_{i=1}^{K-1} N_{i+1}} \prod_{i=1}^{K-1} \binom{N_i-3}{2N_{i+1}-3}, \tag{3.8}$$

where $\binom{n}{k} = \frac{n!}{k!(n-k)!}$.

Proof. As in the proof of Theorem 1 we should start with main frame of order K and then count all possible ways to attach extra frames of orders $K-1, K-2, \cdots 1$ to the

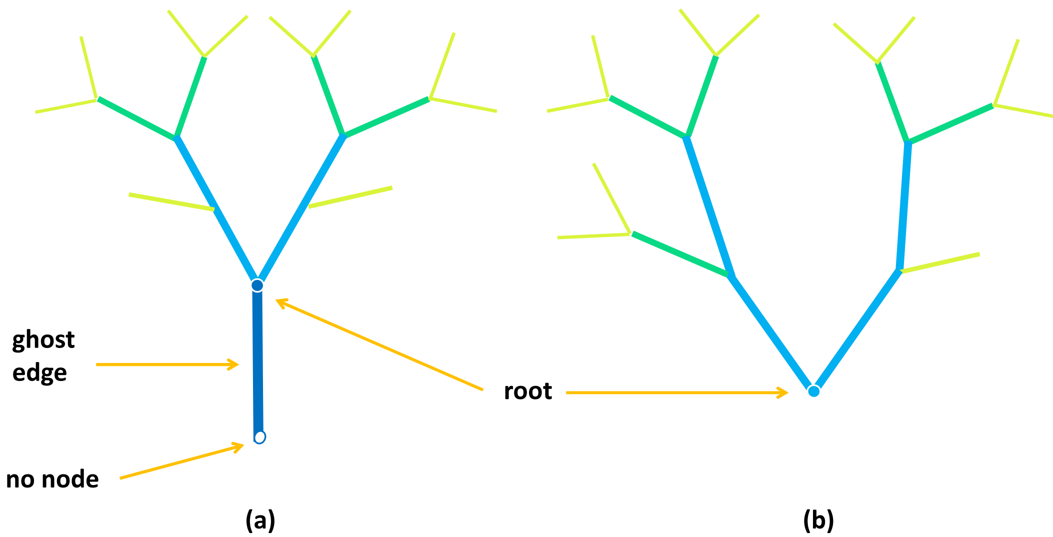

Figure 3.9: Examples of (a) a tree with a ghost edge and (b) a stemless tree. The tree depicted in (a) has order 4 and $N_1 = 10$, $N_2 = 4$, $N_3 = 2$, and $N_4 = 1$. The ghost edge is the parental edge of the root node. The tree has $N = 2N_1 - 1 = 19$ vertices. The tree depicted in (b) has order 3 and $N_1 = 11$, $N_2 = 5$, and $N_3 = 2$. The tree has $N = 2N_1 - 1 = 21$ vertices. Branches of order 4 are depicted in indigo, branches of order 3 in blue, branches of order 2 in green, and branches of order 1 in pear.

main frame. However, since the trees of interest do not have a stem or a ghost edge and have not one but two branches or the highest order K, the main frame for those trees will consist of two frames of order K connected with the root node. Now, starting with $M_{K-1} = N_{K-1} - 2N_K$ extra frames of order $K-1$ we attach them to the main frame, taking into account the fact that we can attach frames either from the right or from the left. Using formula (3.2) with $n = M_{K-1} = N_{K-1} - 2N_K$ and $k = N_K$ we obtain the following

$$
\begin{aligned}
T_{K-1\to K} &= 2^{M_{K-1}} \binom{N_K + M_{K-1} - 1}{N_K - 1} = 2^{N_{K-1}-2N_K} \binom{N_{K-1} - N_K - 1}{N_K - 1} \\
&= 2^{N_{K-1}-2N_K} \binom{N_{K-1} - 3}{1} = 2^{N_{K-1}-2N_K} (N_{K-1} - 3),
\end{aligned}
$$

where the last two equations were obtained using the fact that $N_K = 2$. However, for convenience, we leave the term $T_{K-1\to K}$ in more general form

$$\begin{aligned} T_{K-1\to K} &= 2^{N_{K-1}-2N_K}\binom{N_{K-1}-N_K-1}{N_K-1} \\ &= 2^{N_{K-1}-2N_K}\binom{N_{K-1}-3}{2N_K-3}. \end{aligned}$$

Now we consider all the extra frames of order $K-2$. Then, using formula (3.2) with $n = M_{K-2} = N_{K-2} - 2N_{K-1}$ and $k = N_K + N_{K-1} + M_{K-1} = 2N_{K-1} - N_K$ we get the following value of $T_{K-2\to K-1,K}$

$$\begin{aligned} T_{K-2\to K-1,K} &= 2^{M_{K-2}}\binom{M_{K-2}+2N_{K-1}-N_K-1}{2N_{K-1}-N_K-1} \\ &= 2^{N_{K-2}-2N_{K-1}}\binom{N_{K-2}-3}{2N_{K-1}-3}. \end{aligned}$$

In a similar fashion we obtain

$$T_{i\to i+1,\cdots K} = 2^{N_i-2N_{i+1}}\binom{N_i-3}{2N_{i+1}-3}.$$

Note now that for every possible attachment of extra frames of order $i+1$ there are $T_{i\to i+1,\cdots,K}$ possible attachments of extra frames of order i, $\forall i = \overline{K-1,1}$. Thus, using the multiplication principle of combinatorics, we obtain the total number of trees without a stem and of order K with a particular set of Horton-Strahler numbers $N_1, N_2, \cdots, N_K$

$$\begin{aligned} |\hat{\mathscr{T}}_{N_1,N_2,\cdots,N_K}| &= \prod_{i=K-1}^{1} T_{i\to i+1,\cdots,K} = \prod_{i=K-1}^{1} 2^{N_i-2N_{i+1}}\binom{N_i-3}{2N_{i+1}-3} \\ &= 2^{\sum_{i=1}^{K-1}(N_i-2N_{i+1})}\prod_{i=1}^{K-1}\binom{N_i-3}{2N_{i+1}-3} \\ &= 2^{N_1-2-\sum_{i=1}^{K-1}N_{i+1}}\prod_{i=1}^{K-1}\binom{N_i-3}{2N_{i+1}-3}. \end{aligned}$$

□

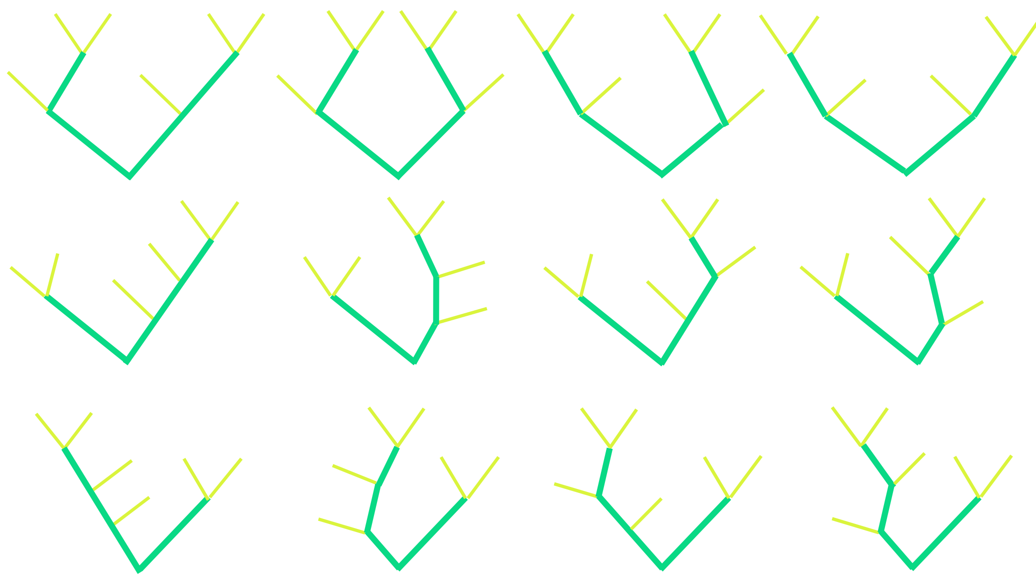

Figure 3.10: An example of a subspace $\hat{\mathscr{T}}_{6,2}$ of 12 binary stemless trees with Horton-Strahler numbers $N_2 = 2$, $N_1 = 6$, and $N = 2N_1 - 1 = 11$ vertices. Branches of order 2 are depicted in green and branches of order 1 in pear.

3.1.3.1 Example

Consider a subspace $\hat{\mathscr{T}}_{6,2}$ of finite unlabeled rooted binary plane stemless trees of order $K = 2$, with Horton-Strahler numbers $N_2 = 2$, $N_1 = 6$, and $N = 2N_1 - 1 = 11$ vertices. Using formula (3.8), we can find the cardinality of this subspace as follows

$$\begin{aligned} |\hat{\mathscr{T}}_{6,2}| &= 2^{N_1-2-\sum_{i=1}^{K-1} N_{i+1}} \prod_{i=1}^{K-1} \binom{N_i - 3}{2N_{i+1} - 3} \\ &= 2^{6-2-2} \binom{6-3}{4-3} \\ &= 2^2 \frac{3!}{2!1!} = 4 \times 3 = 12. \end{aligned}$$

In Figure 3.10, we depict 12 trees of order 2 from the subspace $\hat{\mathscr{T}}_{6,2}$.

3.2 Cardinality of subspace $\mathscr{T}_{K,N_{1,2},N_{1,3},\cdots,N_{K-1,K}}$

In this section we consider a subspace of binary trees with given Tokunaga numbers and examine the cardinality of this subspace. We start with the definition of the subspace of interest.

Definition 22. Let $\mathscr{T}_{K,N_{1,2},N_{1,3},\cdots,N_{K-1,K}} \subset \mathcal{T}$ be the subspace of all finite unlabeled rooted planted binary plane trees of order K and with Tokunaga numbers $N_{1,2}, N_{1,3}, \cdots, N_{K-1,K}$.

Note that $\mathscr{T}_{K,N_{1,2},N_{1,3},\cdots,N_{K-1,K}} \subset \mathscr{T}_N$, where $N = 2N_1 = 2\sum_{i=1}^{K} N_{1,i}$.
The next result evaluates the cardinality of the subspace $\mathscr{T}_{K,N_{1,2},N_{1,3},\cdots,N_{K-1,K}}$.

Theorem 2. *The number of finite unlabeled rooted planted binary plane trees of order K with a particular set of Tokunaga numbers $N_{i,j}$, $i,j = \overline{1,K}$ and with $N = 2N_1 = 2\sum_{i=1}^{K} N_{1,i}$ vertices is given by the following formula*

$$|\mathscr{T}_{K,N_{1,2},N_{1,3},\cdots,N_{K-1,K}}| = \prod_{j=2}^{K}\prod_{i=1}^{j-1} 2^{N_{i,j}} \binom{N_j - 1 + \sum_{l=i}^{j-1} N_{l,j}}{N_{i,j}}, \tag{3.9}$$

where $\binom{n}{k} = \frac{n!}{k!(n-k)!}$.

Proof. We prove this theorem in a similar fashion to that of the Theorem 1. In this proof, we also consider main, necessary, and extra frames, defined in the proof of the Theorem 1. However, using the definition of Tokunaga numbers, provided in the Section 2.3, we notice that the number of extra frames of order i is a sum of the Tokunaga numbers $N_{i,j}$, $j = \overline{i+1,K}$, i.e., $M_i = N_i - 2N_{i+1} = \sum_{j=i+1}^{K} N_{i,j}$. Thus, for every i, we do not attach all extra frames of order i at the same time, but attach them in groups: we first attach $N_{i,K}$ frames of order i to the branches of order K, then we attach $N_{i,K-1}$ frames of order i to the branches of order $K-1$, and so on. Lastly, we attach $N_{i,i+1}$ frames of order i to the branches of order $i+1$. More precisely, given the Tokunaga numbers $N_{i,j}$ we start with the main frame of order K and attach to this main frame $N_{K-1,K}$ frames of order $K-1$, where the number $N_{K-1,K}$ describes the number of branches of order $K-1$ that merge with the branches of order K. Denote $T_{K-1\to K}$ to be the number of trees we obtain by attaching $N_{K-1,K}$ frames of order $K-1$ to N_K branches of order K

of the main frame. We use formula (3.10),

$$\binom{n+k-1}{k-1} = \frac{(n+k-1)!}{(n)!(k-1)!} = \binom{n+k-1}{n} \tag{3.10}$$

which is a modified version of the formula (3.2), provided in the proof of the Theorem 1, and takeing into account that each branch of order $K-1$ can be attached to the middle point of a branch or order K either form the left or from the right, we obtain $T_{K-1\to K}$ as follows

$$T_{K-1\to K} = 2^{N_{K-1,K}} \binom{N_{K-1,K} + N_K - 1}{N_{K-1,K}}. \tag{3.11}$$

Note that when we attach $N_{K-1,K}$ branches of order $K-1$ to N_K branches of order K, we brake the branches of order K into $N_K + N_{K-1,K}$ edges.
Next, denote $T_{K-2\to K}$ to be the the number of different trees we obtain by attaching branches of order $K-2$ to branches of order K. There are $N_{K-2,K}$ branches of order $K-2$ that should be attached to $N_K + N_{K-1,K}$ edges of order K. By using formula (3.2) and considering that each branch of order $K-2$ can be attached either from the left or from the right we obtain $T_{K-2\to K}$ as follows

$$T_{K-2\to K} \;=\; 2^{N_{K-2,K}} \binom{N_{K-2,K} + N_{K-1,K} + N_K - 1}{N_{K-2,K}}. \tag{3.12}$$

Consider now an intermediate step. Let $T_{i\to K}$ be the number of trees that we obtain by attaching $N_{i,K}$ branches of order i to the edges of orders K. Note that there are now

$$\begin{aligned} k &= N_K + N_{K-1,K} + N_{K-2,K} + \cdots + N_{i+1,K} \\ &= N_K + \sum_{l=i+1}^{K-1} N_{l,K} \end{aligned}$$

edges of order K, to which we can attach $N_{i,K}$ branches of order i. Thus, using formula (3.2) and considering that each branch can be attached either from the left or from the right we obtain $T_{i\to K}$ as follows

$$\begin{aligned} T_{i\to K} &= 2^{N_{i,K}} \binom{N_{i,K} + N_K + \sum_{l=i+1}^{K-1} N_{l,K} - 1}{N_{i,K}} \\ &= 2^{N_{i,K}} \binom{N_K + \sum_{l=i}^{K-1} N_{l,K} - 1}{N_{i,K}}. \end{aligned} \tag{3.13}$$

Formula (3.13) provides a general description of the terms $T_{i\to K}$, $\forall i = \overline{K-1, 1}$. Similarly, we obtain a general form for the terms $T_{i\to j}$

$$T_{i\to j} = 2^{N_{i,j}} \binom{N_j + \sum_{l=i}^{j-1} N_{l,j} - 1}{N_{i,j}}. \tag{3.14}$$

Note now, that $i \in [1, j-1]$ for every term $T_{i\to j}$, since only branches of order $i \in [1, j-1]$ can be attached to the branches of order j. Moreover, $j \in [2, K]$. Thus, using the multiplication principle of combinatorics, we obtain the total number of trees from the subspace $\mathscr{T}_{K,N_{1,2},N_{1,3},\cdots,N_{K-1,K}}$ as follows

$$\begin{aligned} |\mathscr{T}_{K,N_{1,2},N_{1,3},\cdots,N_{K-1,K}}| &= \prod_{j=K}^{2} \prod_{i=j-1}^{1} T_{i\to j} \\ &= \prod_{j=K}^{2} \prod_{i=j-1}^{1} 2^{N_{i,j}} \binom{N_j - 1 + \sum_{l=i}^{j-1} N_{l,j}}{N_{i,j}} \\ &= \prod_{j=2}^{K} \prod_{i=1}^{j-1} 2^{N_{i,j}} \binom{N_j - 1 + \sum_{l=i}^{j-1} N_{l,j}}{N_{i,j}}. \end{aligned}$$

□

3.2.1 Examples

Consider now a subspace $\mathscr{T}_{3,0,1,1}$, i.e., the subspace of finite unlabeled rooted planted binary plane trees of order $K = 3$ with the Tokunaga numbers $N_{1,2} = 0, N_{1,3} = 1$, $N_{2,3} = 1$. Using formula (3.9), we find that the number of such trees can be found as follows

$$|\mathscr{T}_{3,0,1,1}| = 2^0 \binom{0+3-1}{0} 2^1 \binom{1+1+1-1}{1} 2^1 \binom{1+1-1}{1} = 8.$$

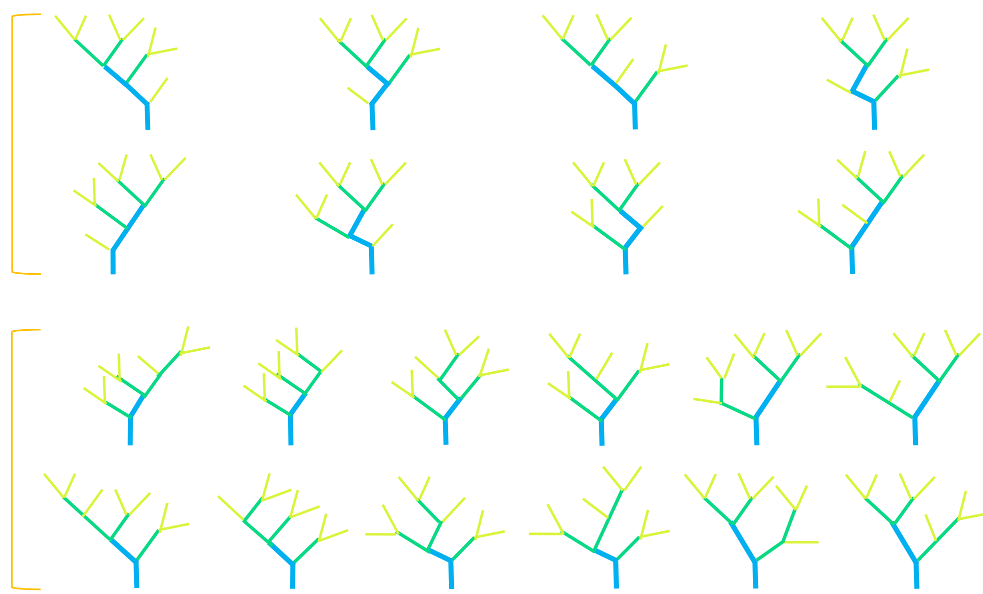

Figure 3.11: An example of (a) a subspace $\mathscr{T}_{3,0,1,1}$ of 8 trees of order 3 and with Tokunaga numbers $N_{1,2} = 0, N_{1,3} = 1$, $N_{2,3} = 1$ and (b) a subspace $\mathscr{T}_{3,1,0,1}$ of 12 trees of order 3 and with Tokunaga numbers $N_{1,2} = 1, N_{1,3} = 0$, $N_{2,3} = 1$. Branches of order 3 are depicted in blue, branches of order 2 in green, and branches of order 1 in pear.

In Figure 3.11 (a), we depict all 8 trees of order 3 with given Tokunaga numbers $N_{1,2} = 0, N_{1,3} = 1$, and $N_{2,3} = 1$.

Similarly, using formula (3.9), we can find the cardinality of the subspace $\mathscr{T}_{3,1,0,1}$, the subspace of finite unlabeled rooted planted binary plane trees of order $K = 3$ with the Tokunaga numbers $N_{1,2} = 1, N_{1,3} = 0$, $N_{2,3} = 1$

$$|\mathscr{T}_{3,1,0,1}| = 2^1 \binom{1+3-1}{1} 2^0 \binom{0+1+1-1}{0} 2^1 \binom{1+1-1}{1} = 12.$$

In Figure 3.11 (b), we depict all 12 trees of order 3 with given Tokunaga numbers $N_{1,2} = 1, N_{1,3} = 0$, and $N_{2,3} = 1$.

By observing Figures 3.8 and 3.11, we conclude that

$$|\mathscr{T}_{3,0,1,1}| + |\mathscr{T}_{3,1,0,1}| = |\mathscr{T}_{7,3,1}|.$$

Moreover, $\mathscr{T}_{3,0,1,1} \subset \mathscr{T}_{7,3,1}$, $\mathscr{T}_{3,1,0,1} \subset \mathscr{T}_{7,3,1}$ and $\mathscr{T}_{7,3,1} = \mathscr{T}_{3,0,1,1} \bigcup \mathscr{T}_{3,1,0,1}$. This is due to the fact that trees from the subspace $\mathscr{T}_{7,3,1}$ have 6 necessary leaves and one extra, which can be attached either to the edges of order 3 or to the branches of order 2. The subspace $\mathscr{T}_{3,0,1,1}$ contains all the trees which have that extra leaf attached to the edges of order 3. On the other hand, the subspace $\mathscr{T}_{3,1,0,1}$ contains all the trees which have that extra leaf attached to the edges of order 2. The subspace $\mathscr{T}_{7,3,1}$ contains all the trees which have that extra leaf attached to the edges of order 2 or 3.

In Table 3.2 we present more examples of the numbers of trees for different sets of Tokunaga numbers. Each column contains the cardinality of the subspace $\mathscr{T}_{3,N_{1,2},N_{1,3},N_{2,3}}$ for different values of the parameters $N_{1,j}$, $j = \{1,2\}$, which depend on the number of leaves N_1. The first column contains the cardinality of the subspace $\mathscr{T}_{3,N_{1,2},0,0}$, where parameter $N_{1,2}$ represents the number of branches of order 1 that merge with the branches of order 2, i.e., $N_{1,2} = N1 - 2N_2 = N_1 - 4$. In other words, the subspace $\mathscr{T}_{3,N_{1,2},0,0}$ contains all the trees for which extra leaves can only be placed on the branches of order 2. Similarly, the second column contains the cardinality of the subspace $\mathscr{T}_{3,0,N_{1,3},0}$, where $N_{1,3} = N_1 - 2N_2 = N_1 - 4$, i.e., the subspace of all trees for which extra leaves can only be placed on the branches of order 3. And the last two columns contain the cardinality of the subspaces $\mathscr{T}_{3,N_{1,2},0,1}$ with $N_{1,2} = N1 - 2N_2 = N_1 - 6$ and $\mathscr{T}_{3,0,N_{1,3},1}$ with $N_{1,3} = N1 - 2N_2 = N_1 - 6$, accordingly.

Although done for trees form space $\mathcal{T}$, the results of Theorem 2 can be extended for trees without a stem and for the trees with a ghost edge [2, 46]. We present both extensions below.

3.2.2 Stemless trees and trees with a ghost edge

Corollary 3. *Let $\tilde{\mathscr{T}}_{K,N_{1,2},N_{1,3},\cdots N_{K-1,K}}$ be a subspace of finite unlabeled rooted binary plane trees with a ghost edge, $N = 2N_1 - 1$ vertices, and a particular set of Tokunaga numbers $N_{1,2}, N_{1,3}, \cdots N_{K-1,K}$. Also, let $\hat{\mathscr{T}}_{K,N_{1,2},N_{1,3},\cdots N_{K-1,K}}$ be a subspace of finite unlabeled rooted binary plane stemless trees with $N = 2N_1 - 1$ vertices and a particular set of Tokunaga numbers $N_{1,2}, N_{1,3}, \cdots N_{K-1,K}$. Then the cardinality of both of those*

Table 3.2: Each entry in this table represents the cardinality of the subspace $\mathscr{T}_{3,N_{1,2},N_{1,3},N_{2,3}}$ for different values of the parameters $N_{1,j},\ j = \{1,2\}$, which depend on the number of leaves N_1. The first column contains the cardinality of the subspace $\mathscr{T}_{3,N_{1,2},0,0}$, where $N_{1,2} = N1 - 2N_2 = N_1 - 4$. The second column contains the cardinality of the subspace $\mathscr{T}_{3,0,N_{1,3},0}$ with $N_{1,3} = N_1 - 2N_2 = N_1 - 4$. The last two columns contain the cardinality of the subspaces $\mathscr{T}_{3,N_{1,2},0,1}$ with $N_{1,2} = N1 - 2N_2 = N_1 - 6$ and $\mathscr{T}_{3,0,N_{1,3},1}$ with $N_{1,3} = N1 - 2N_2 = N_1 - 6$, accordingly.

N_1	$\lvert\mathscr{T}_{3,N_1-4,0,0}\rvert$	$\lvert\mathscr{T}_{3,0,N_1-4,0}\rvert$	$\lvert\mathscr{T}_{3,N_1-6,0,1}\rvert$	$\lvert\mathscr{T}_{3,0,N_1-6,1}\rvert$
4	1	1		
5	4	2		
6	12	4	2	2
7	32	8	12	8
8	80	16	48	24
9	192	32	160	64
10	448	64	480	160
11	1024	128	1344	384
12	2304	256	3584	896
26	96,468,992	4,194,304	484,442,112	44,040,192

subspaces is equal to the cardinality of the subspace $\mathscr{T}_{K,N_{1,2},N_{1,3},\cdots N_{K-1,K}}$*, i.e.,*

$$\begin{aligned} |\tilde{\mathscr{T}}_{K,N_{1,2},N_{1,3},\cdots N_{K-1,K}}| &= |\hat{\mathscr{T}}_{K,N_{1,2},N_{1,3},\cdots N_{K-1,K}}| \\ &= |\mathscr{T}_{K,N_{1,2},N_{1,3},\cdots N_{K-1,K}}| \\ &= \prod_{j=2}^{K}\prod_{i=1}^{j-1} 2^{N_{i,j}} \binom{N_j - 1 + \sum_{l=i}^{j-1} N_{l,j}}{N_{i,j}}, \end{aligned} \quad (3.15)$$

where $\binom{n}{k} = \frac{n!}{k!(n-k)!}$.

Proof. Note that in the proof of Theorem 2 all the terms $T_{i\to j}$ were provided in a general form (3.14) and the fact that $N_K = 1$ for all trees from the subspace $\mathscr{T}_{K,N_{1,2},N_{1,3},\cdots N_{K-1,K}}$ was never used. Therefore, formula (3.15) can be used to find cardinalities of subspaces $\tilde{\mathscr{T}}_{K,N_{1,2},N_{1,3},\cdots N_{K-1,K}}$ and $\hat{\mathscr{T}}_{K,N_{1,2},N_{1,3},\cdots N_{K-1,K}}$, by setting the value of N_K to be either 1 for the subspace $\tilde{\mathscr{T}}_{K,N_{1,2},N_{1,3},\cdots N_{K-1,K}}$ or 2 for the subspace $\hat{\mathscr{T}}_{K,N_{1,2},N_{1,3},\cdots N_{K-1,K}}$. ∎

Chapter 4: Information theoretical analysis

In this section, we use Shannon entropy to quantify the structural complexity of a tree. We propose an entropy based measure, namely the entropy rate of a tree, to examine how the structural complexity of a tree changes as a tree is allowed to grow in size. We find entropy rates for several types of trees: the planted binary trees with N vertices, the planted binary trees that satisfy Horton law with Horton exponent R, and for the planted binary trees that satisfy Tokunaga law with Tokunaga parameters (a, c).

4.1 Entropy

Recall that entropy is a measure of the average uncertainty in the random variable [11,53]. For a discrete random variable X with possible values $x_1, x_2, \cdots, x_n$ and probability mass function $P(x)$ the entropy of X is defined by

$$H(X) = -\sum_{i=1}^{n} P(x_i) \log_2 P(x_i) = -\mathbb{E}[\log_2 P(X)], \tag{4.1}$$

where the quantity $0 \log_2 0$ is set to be 0. We can think of $\log_2 P(x_i)$ as the uncertainty of the outcome x_i (or the "surprise" of observing x_i). Thus, the entropy can be thought of as the average "surprise". Note, if possible values of X are uniformly distributed, i.e., $P(x_i) = 1/n$, then there is a maximal uncertainty about the outcome, maximum "surprise". In this case, the entropy achieves its maximal value $H(X) = \log_2 n$. For example, the entropy of a fair coin toss is 1 bit. In general, random variable with higher entropy are more "unpredictable" than the random variables with lower entropy. For example, the occurrence of a certain event ($P(x_i) = 1$, i.e., no "surprise") has minimal uncertainty, which corresponds to the minimal value of entropy $H(X) = 0$. Thus, the entropy of a random variable X that has n possible outcomes is bounded as follows

$$0 \leq H(X) \leq \log_2 n.$$

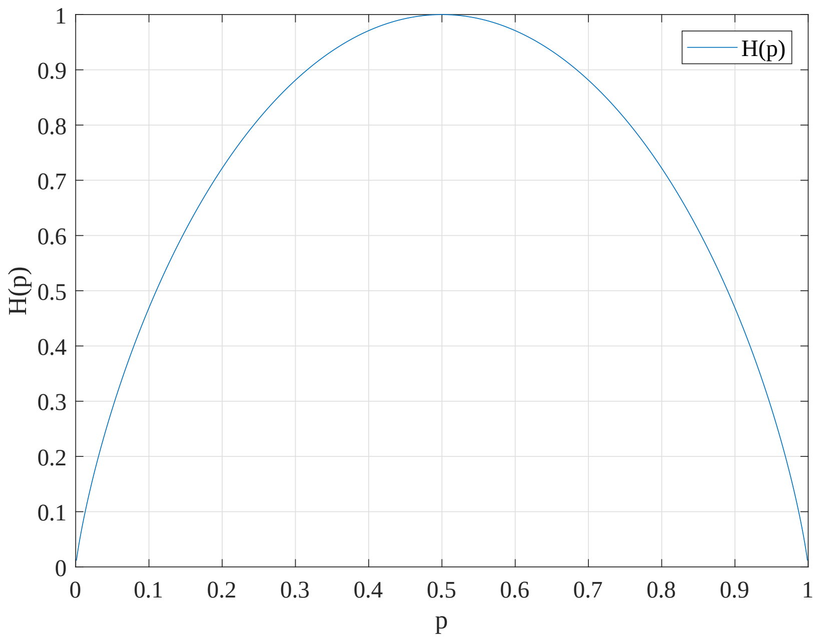

Figure 4.1: The binary entropy $H(p) = -p\log_2 p - (1-p)\log_2(1-p)$.

Note that the log in formula (4.1) is usually taken to base 2. In this case the entropy is measured in bits. If the log is taken to base e, then the entropy is measured in *nuts*.
In many applications random variable X is assumed to be binary, i.e., $n = 2$ and $P(X = 1) = p$. In this case, a binary entropy (or binary entropy function in related literature [5])

$$H(p) = -p\log_2 p - (1-p)\log_2(1-p)$$

is used. In Figure 4.1 we depict $H(p)$ for $p \in [0,1]$. The binary entropy function satisfies the following properties:

- $H(0) = H(1) = 0$,
- $H\left(\frac{1}{2}\right) = 1$,
- $H(p) = H(1-p)$, $\forall p \in [0,1]$.

The notion of entropy is closely related to the question of efficiently encoding data for storage or transmission. From the information theory point of view, the entropy of a random variable can be thought of as an average number of bits required to describe the random variable [11]. Consider a random variable X that has a uniform distribution over 8 outcomes, e.g., an eight-sided dice. The entropy of X is

$$H(X) = -\sum_{i=1}^{8} \frac{1}{8} \log_2 \frac{1}{8} = \log_2 8 = 3$$

bits. A 3-bit string takes on 8 different values and is sufficient to describe 8 outcomes of X. Note that all outcomes of X have representations of the same 3-bit length. Consider now a random variable Y with a nonuniform distribution. Assume Y can take 5 possible values $\{y_1, y_2, y_3, y_4, y_5\}$ with corresponding probabilities $\left(\frac{1}{2}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}\right)$. The entropy of Y is

$$H(Y) = -\frac{1}{2} \log_2 \frac{1}{2} - 4\frac{1}{8} \log_2 \frac{1}{8} = 2$$

bits. Note that 2-bit string will not be enough to encode all five values of Y. The 3-bit string takes on 8 different values and is sufficient to describe 5 outcomes of Y. However, we can get a better encoding by using Huffman coding technique, that produces an optimal code. It assigns a vector of probabilities to a set of leaf nodes and builds a code tree by repeatedly combining the two least probable nodes. The possible outcome y_1, y_2, y_3, y_4, y_5 are encoded as strings $1, 011, 010, 001, 000,$ respectively.

Definition 23. Let L be a random variable that represents the length of a codeword. The average coding length is given by the expected values of L, i.e., $\mathbb{E}[L]$.

For our example, the average coding length is

$$\mathbb{E}[L] = 3 \times \frac{1}{8} + 3 \times \frac{1}{8} + 3 \times \frac{1}{8} + 3 \times \frac{1}{8} + 1 \times \frac{1}{2} = 2.$$

Since we use shorter description for the more probable outcome y_1 and longer descriptions for the less probable outcomes y_2, y_3, y_4, y_5, the average description length is equal to the value of the entropy and is exactly 2 bits. Huffman coding technique provides savings in bits: instead of using 3 bits per symbol, this coding technique requires on average only 2 bits per symbol. This is an example of a *variable length* code. A wide variety of variable length codes can be described by binary trees. In Figure 4.2 we depict a binary tree that corresponds to the Huffman coding of random variable Y.

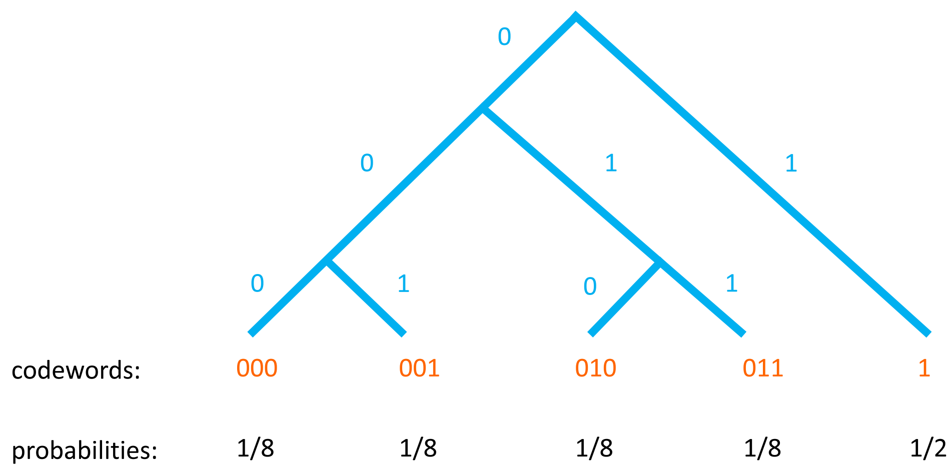

Figure 4.2: The binary tree that corresponds to the Huffman coding of a random variable Y, which takes values y_1, y_2, y_3, y_4, y_5 with corresponding probabilities $\left(\frac{1}{2}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}\right)$. The codewords are depicted in orange.

The central result in information theory is a well-known theorem by Claude Shannon [53], presented below, that ties together the notions of entropy and coding efficiency.

Theorem 3. *(Shannon, 1948)*
For any decodable (tree) code, the expected coding length is lower bounded by the entropy as follows

$$\mathbb{E}[L] \geq H(X).$$

It is easy to see that the the expected coding length is equal to the entropy when

$$P(x_i) = 2^{-l(x_i)},$$

where $l(x_i)$ is the length of the codeword for x_i.

Remark 3. *Note that in related literature quite often the entropy is viewed as a measure of average amount of information received when the value of a random variable X is* *observed. Thus, the average surprise yielded by X, the uncertainty of X, or the average information produces by X, are the same concepts observed from slightly different points of view.*

In the next section we introduce the notion of entropy and entropy rate for tree-like structures.

4.1.1 Entropy and entropy rate for spaces of trees

Definition 24. Consider the subspace $\mathscr{T}_N$, and let P be a probability measure over $\mathscr{T}_N$. We define the *entropy* of a random planted binary plane tree $T_N \in \mathscr{T}_N$ as follows

$$H(T_N) = -\mathbb{E}[\log_2 P(T_N)].$$

We consider it to be the measure of the structural complexity of a tree.
Informally, the larger the entropy, the more complex is the tree's dendritic structure. From the information theoretical point of view, the entropy of the tree gives the average number of bits needed to encode the tree.

Definition 25. We define *entropy rate* $\mathscr{H}_\infty$ to be the limit of normalized entropies $\frac{H(T_N)}{N}$, for T_N sampled from the corresponding subspace $\mathscr{T}_N$ with probability measure P, as $N \to \infty$

$$\mathscr{H}_\infty = \lim_{N\to\infty} \frac{H(T_N)}{N},$$

provided that the limit exists. The entropy rate describes the entropy's growth rate as $N \to \infty$.

The entropy rate quantifies per vertex entropy. In other words, for large N the entropy rate gives the average number of bits per vertex required to encode the tree. In fact, for large N there exists an arithmetic coding scheme that encodes a tree with N vertices using about $N\mathscr{H}_\infty$ bits [31]. Arithmetic coding can get arbitrarily close to the entropy, because it does not convert each vertex separately, but assigns one codeword to the entire tree. The tree can be recreated from this codeword.
In the next section we consider the subspace $\mathscr{T}_N$ and explore entropy and entropy rate for the trees from $\mathscr{T}_N$.

4.2 Entropy and entropy rate for subspaces $\mathscr{T}_N$

Theorem 4. *Consider a sequence of random trees T_N, each sampled uniformly from the corresponding subspace $\mathscr{T}_N \subset \mathcal{T}$. Then the entropy of a tree T_N is given by*

$$H(T_N) = N + \mathcal{O}(\log_2 N),$$

as $N \to \infty$.

Proof. We start the proof by noticing that under the assumption of the uniform distribution of trees in $\mathscr{T}_N$, the probability of a random tree $T_N \in \mathscr{T}_N$ is given as

$$P(T_N) = \frac{1}{\mathcal{C}_{n-1}},$$

where $\mathcal{C}_{n-1}$ is the $(n-1)$th Catalan number

$$\mathcal{C}_{n-1} = \frac{1}{n}\binom{2n-2}{n-1},$$

and $n = \frac{N}{2}$ is the number of leaves in a tree T_N. Thus, by the definition of the entropy we conclude that

$$\begin{aligned} H(T_N) &= -\mathbb{E}[\log_2 P(T_N)] \\ &= -\sum_{i=1}^{\mathcal{C}_{n-1}} \frac{1}{\mathcal{C}_{n-1}} \log_2 \frac{1}{\mathcal{C}_{n-1}} \\ &= \log_2 \mathcal{C}_{n-1}. \end{aligned} \tag{4.2}$$

We can now rewrite the term $\log_2 \mathcal{C}_{n-1}$ in the following way

$$\begin{aligned} \mathcal{C}_{n-1} &= \frac{(2n-2)!}{n!(n-1)!} = \frac{\left(2\left(\frac{N}{2}-1\right)\right)!}{\left(\frac{N}{2}\right)!\left(\frac{N}{2}-1\right)!} \\ &= \frac{\left(2\left(\frac{N}{2}-1\right)\right)!}{\left(\frac{N}{2}\right)\left(\left(\frac{N}{2}-1\right)!\right)^2} = \frac{2}{N}\binom{N-2}{\frac{N}{2}-1}, \end{aligned}$$

where we use the fact that $n = \frac{N}{2}$. Using the results of Lemma 2, given in Section 6.2, we obtain the entropy of $T_N \in \mathscr{T}_N$ as follows

$$\begin{aligned} H(T_N) &= \log_2 \left[\frac{2}{N} \binom{N-2}{\frac{N}{2}-1} \right] \\ &= 1 - \log_2 N + 2\left(\frac{N}{2}-1\right) H\left(\frac{1}{2}\right) + \mathcal{O}(\log_2 N) \\ &= 1 - \log_2 N + N - 2 + \mathcal{O}(\log_2 N) \\ &= N + \mathcal{O}(\log_2 N), \end{aligned} \tag{4.3}$$

as $N \to \infty$. □

Corollary 4. *Consider a sequence of random trees T_N, each sampled uniformly from the corresponding subspace $\mathscr{T}_N \subset \mathcal{T}$. Then the entropy rate of the sequence T_N is given by*

$$\mathscr{H}_\infty = \lim_{N\to\infty} \frac{H(T_N)}{N} = 1.$$

Proof. Dividing $H(T_N)$ by N and taking the limit as $N \to \infty$ we obtain the entropy rate as follows

$$\begin{aligned} \mathscr{H}_\infty(\mathscr{T}_N) &= \lim_{N\to\infty} \frac{H(T_N)}{N} \\ &= \lim_{N\to\infty} \frac{N + \mathcal{O}(\log_2 N)}{N} = 1. \end{aligned} \tag{4.4}$$

□

Theorem 4 and Corollary 4 demonstrate that for large enough N, we need about N bits per tree or about one bit per vertex to encode any tree $T_N \in \mathscr{T}_N$. While presented in a different context, theorem 4 reaffirms the entropy rate of the maximum entropy model in [31].

4.3 Entropy rates for subspaces of Horton self-similar trees

In this section we explore the structural complexity of the trees form the subspace of planted binary trees that satisfy Horton law with Horton exponent R. This section is based on the results published by the author in [10].

4.3.1 Horton self-similarity

In Section 3.1 we introduced a subspace $\mathscr{T}_{N_1,N_2,\cdots,N_K}$ of planted binary trees with an arbitrary (but admissible) set Horton-Strahler numbers $N_1, N_2, \cdots, N_K$. Quite often, however, observed tree-like structures display geometric decrease of the numbers N_i of elements of Horton-Strahler order $i \geq 1$. This property is known as Horton self-similarity, also referred to as the Horton Law. Formally, the strong Horton Law states the existence of the limit

$$\lim_{K\to\infty} \frac{N_i[K]}{N_{i+1}[K]} = R,$$

where N_i are different for different values of K, i.e., $N_i = N_i[K]$. The quantity R is called the Horton exponent.

Informally, as tree size grows, i.e., $K \to \infty$, the quantity $\frac{N_i}{N_{i+1}}$ approaches R. There are multiple models with broad range of Horton exponents that appear in different scientific areas and have practical importance in a variety of applications [8, 43, 45, 47, 74]. For example, a perfect binary tree satisfies the Horton law with $R = 2$ while the critical binary Galton-Watson tree [8, 47, 50, 55] satisfies the Horton law with $R = 4$. Despite their practical significance [8,45,47,74], models with Horton exponent other than 2 and 4 were not extensively addressed in the scientific literature, until two well-known processes, namely Kingman's coalescent and discrete white noise, were studied in [39]. Interestingly, for many natural tree-like structures $R \in (3, 6)$. For example, the real river networks have Horton exponent R in a range $(3, 6)$ [28, 29], e.g., for Amazon river $R = 4.51$ and for Mississippi river $R = 4.69$. This phenomenon was confirmed in hydrology, biology, and other areas; see [25, 32, 45, 47, 51, 54, 59] and reference therein.

In the next section, we introduce a subspace of trees that satisfy Horton law with Horton exponent R and examine the entropy rate of a sequence of trees, each sampled uniformly form a corresponding subspace.

4.3.2 Entropy rates for subspaces $\mathscr{T}_{K,R}$

Definition 26. We define $\mathscr{T}_{K,R} \subset \mathcal{T}$ to be a subspace of trees of order K with Horton-Strahler numbers N_k, $\forall k = \overline{1, K}$ that are defined in a special form as follows

$$N_k \in \left(R^{K-k} - \alpha^{K-k}, R^{K-k} + \alpha^{K-k}\right),$$

where $R, \alpha \in \mathbb{R}$ such that $R \geq 2$ and $\alpha \in (1, R)$.

Note that $N_k \approx R^{K-k}$ with an error $N_k - R^{K-k}$ dominated by the power of an exponent smaller than R. Moreover, the number of nodes N grows asymptotically as $2R^{K-1}$, i.e.,

$$\left|1 - \frac{N}{2R^{K-1}}\right| < \left(\frac{\alpha}{R}\right)^{K-1} \to 0$$

as $K \to \infty$. Therefore, we use $2R^{K-1}$ in the denominator in formula (4.5). It is easy to see that this model satisfies the Horton law with Horton exponent R.

Theorem 5. *Let $R \geq 2$. Consider a sequence of random trees T_K, each sampled uniformly from the corresponding subspace $\mathscr{T}_{K,R} \subset \mathcal{T}$. Then the entropy rate of a sequence T_K is given by*

$$\mathscr{H}_\infty(R) = \lim_{K\to\infty} \frac{H(T_K)}{2R^{K-1}} = 1 - \frac{1 - H\left(\frac{2}{R}\right)}{2 - \frac{2}{R}}, \tag{4.5}$$

where $H(z) = -z\log_2 z - (1-z)\log_2(1-z)$ is a binary entropy of z.

Proof. We begin the proof by noticing that states that $\forall k = \overline{1, K}$

$$N_k \in (R^{K-k} - \alpha^{K-k}, R^{K-k} + \alpha^{K-k}).$$

Thus $\forall k = \overline{1, K}$ there are no more than $2\alpha^{K-k}$ possible integer values for N_k. We denote $C(K, \alpha)$ to be the total number of possible collections of Horton-Strahler numbers $N_1, N_2, \cdots, N_K$, such that $N_k = R^{K-k} \pm \alpha^{K-k}$. Notice that for every particular collection of Horton-Strahler numbers $N_1, N_2, \cdots, N_K$ there are

$$|\mathscr{T}_{N_1,N_2,\cdots,N_K,R}| \quad = \quad 2^{N_1 - 1 - \sum_{k=1}^{K-1} N_{k+1}} \prod_{k=1}^{K-1} \binom{N_k - 2}{2N_{k+1} - 2}$$

trees. Thus, for a given set of parameters K and R the number of all trees with Horton-Strahler numbers $N_1, N_2, \cdots, N_K$ that satisfy $N_k = R^{K-k} \pm \alpha^{K-k}$, $\forall k = \overline{1, K}$ is given by

$$|\mathscr{T}_{K,R}| = \sum_{(N_1,N_2,\cdots,N_K)\in C(K,\alpha)} |\mathscr{T}_{N_1,N_2,\cdots,N_K,R}|,$$

where $\mathscr{T}_{N_1,N_2,\cdots,N_K,R} \subset \mathscr{T}_{K,R}$. Assuming uniform distribution of such trees, the probability of one tree $T_K \in \mathscr{T}_{K,R}$ is given by

$$P(T_K) = \frac{1}{|\mathscr{T}_{K,R}|}.$$

Therefore, the entropy rate is given as

$$\begin{aligned} \mathscr{H}_\infty(R) &= \lim_{K\to\infty} \frac{H(T_{K,R})}{2R^{K-1}} \\ &= \lim_{K\to\infty} \frac{1}{2R^{K-1}} \log_2 |\mathscr{T}_{K,R}|. \end{aligned} \tag{4.6}$$

To find the entropy rate in (4.6), first note that $|\mathscr{T}_{K,R}|$ can be bounded as follows

$$|\mathscr{T}_{N_1^*,N_2^*,\cdots,N_K^*,R}| \leq |\mathscr{T}_{K,R}| \leq |\mathscr{T}_{N_1^*,N_2^*,\cdots,N_K^*,R}| \times C(K,\alpha), \tag{4.7}$$

where

$$(N_1^*, N_2^*, \cdots, N_K^*) = \arg \max_{N_1,N_2,\cdots,N_K \in C(K,\alpha)} |\mathscr{T}_{N_1,N_2,\cdots,N_K,R}|.$$

Since

$$C(K,\alpha) \quad \leq \quad \prod_{k=1}^{K} 2\alpha^{K-k} = 2^K \alpha^{\sum_{k=1}^{K}(K-k)} = 2^K \alpha^{\frac{K(K-1)}{2}},$$

we can rewrite formula (4.7) in the following way

$$|\mathscr{T}_{N_1^*,N_2^*,\cdots,N_K^*,R}| \leq |\mathscr{T}_{K,R}| \leq |\mathscr{T}_{N_1^*,N_2^*,\cdots,N_K^*,R}| \times 2^K \alpha^{\frac{K(K-1)}{2}}. \tag{4.8}$$

Next we apply the logarithm and divide by $2R^{K-1}$ all sides of the inequality in (4.8). Taking the limit as $K \to \infty$, we conclude that

$$\mathscr{H}_\infty(R) \quad = \quad \lim_{K\to\infty} \frac{\log_2 |\mathscr{T}_{K,R}|}{2R^{K-1}} = \lim_{K\to\infty} \frac{\log_2 |\mathscr{T}_{N_1^*,N_2^*,\cdots,N_K^*,R}|}{2R^{K-1}},$$

where the last equality was obtained using the fact that

$$\begin{aligned} \lim_{K\to\infty} \frac{1}{2R^{K-1}} \log_2 \left(2^K \alpha^{\frac{K(K-1)}{2}}\right) &= \lim_{K\to\infty} \frac{K + \frac{K(K-1)}{2} \log_2 \alpha}{2R^{K-1}} \\ &= 0. \end{aligned}$$

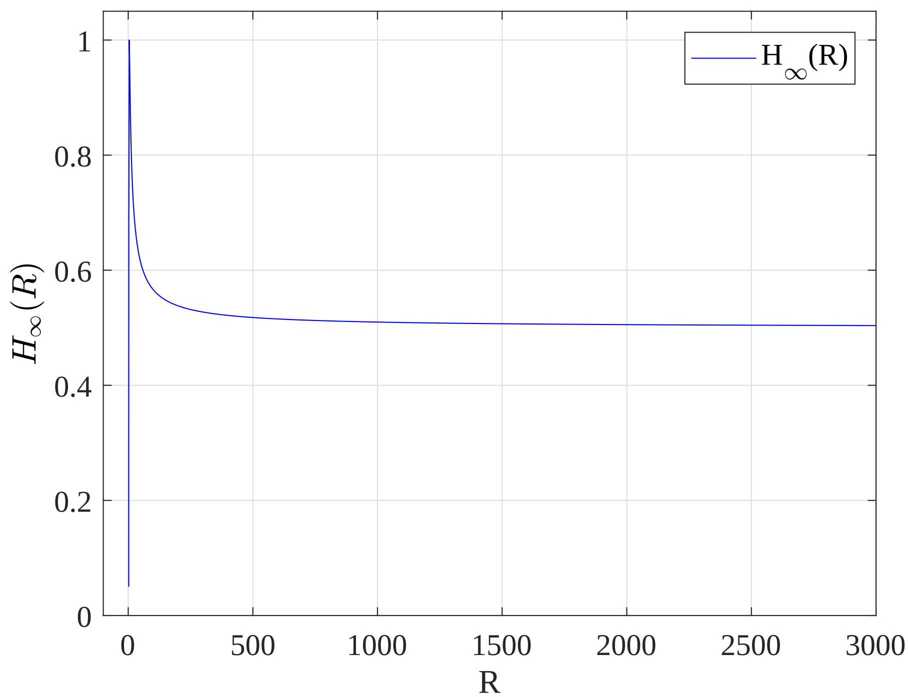

Figure 4.3: Entropy rate $\mathscr{H}_\infty(R)$ for $R \in (0, 3000]$. Note, $\lim_{R\to\infty} \mathscr{H}_\infty(R) = \frac{1}{2}$.

Finally, using results of Lemma 3, provided in Section 6.3, we conclude that

$$\mathscr{H}_\infty(R) = 1 - \frac{1 - H\left(\frac{2}{R}\right)}{2 - \frac{2}{R}}.$$

□

4.3.3 Discussion

In Figures 4.3 and 4.4 we depict the entropy rate $\mathscr{H}_\infty(R)$ for $R \in (0, 3000]$ and $R \in (0, 20]$, respectively. We observe that the entropy rate is equal to zero when $R = 2$ because the dendritic structure of a perfect planted binary plane tree is predetermined for any K. We also observe that for $R = 4$ the entropy rate attains its maximal value 1. Recall that the critical binary Galton-Watson process has parameter $R = 4$. This process was often used to model river networks. However, by performing a high-precision extraction of river

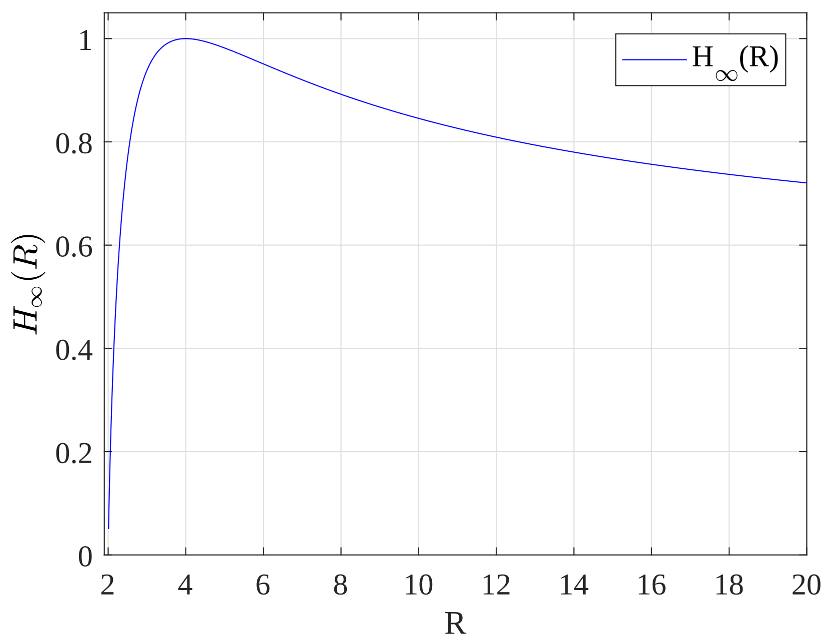

Figure 4.4: Entropy rate $\mathscr{H}_\infty(R)$ for $R \in (0, 20]$. The maximum of $\mathscr{H}_\infty(R)$ is attained at $R = 4$.

channels for Kentucky River (Kentucky) and Powder River (Wyoming), Peckham [47] noticed that the Horton exponents for real rivers are different from the theoretical parameter $R = 4$. For example, for Amazon river $R = 4.51$. Consequently, entropy rate for Amazon river is 0.9941. A natural question to ask would be: What physical phenomenon causes the nonoptimality of entropy rate of the rivers? A possible explanation of this phenomenon is given in [61]: although rivers adjust their configurations to maximize the entropy, this maximization happens within local feasibility constraints, thus global maximum is not achieved.

Note also, when R is allowed to grow, the entropy rate converges to $1/2$. More precisely,

$$\lim_{R\to\infty} \mathscr{H}_\infty(R) = \lim_{R\to\infty} \left(1 - \frac{1 - H\,(2/R)}{2 - 2/R}\right) = \frac{1}{2}.$$

This implies that for large R and K one would need about $N/2 \asymp R^{K-1}$ bits to decode the entire tree. It would be interesting to explain why trees with large enough R require

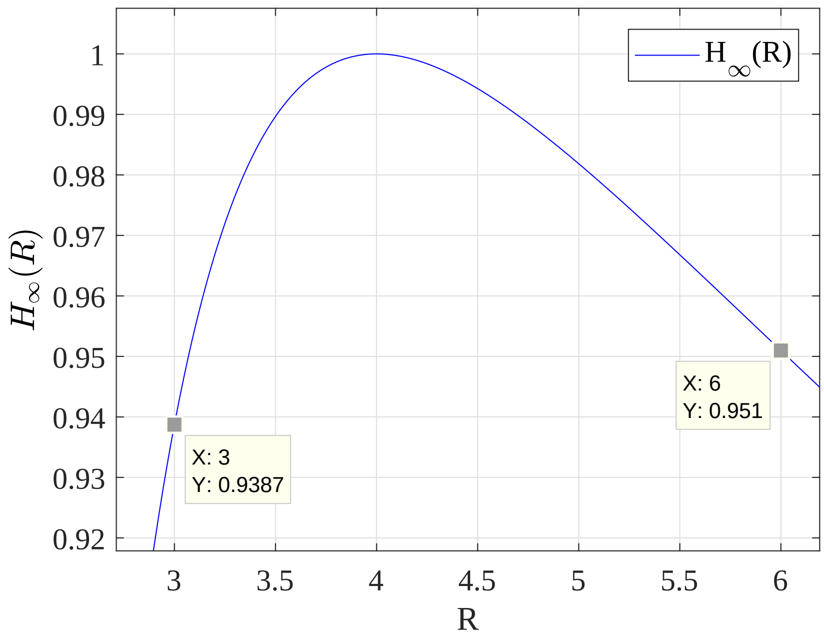

Figure 4.5: Entropy rate $\mathscr{H}_\infty(R)$ for $R \in [3, 6]$. Note $\mathscr{H}_\infty(3) = 0.9387$ and $\mathscr{H}_\infty(6) = 0.951$.

relatively fewer bits to encode them than the trees with $R = 4$. In other words, why the dendritic structure of planted binary trees with $R = 4$ is less predictable when compared with the the dendritic structure of planted binary trees with large R.

4.4 Entropy rates for subspaces $\mathscr{T}_{K,N_{1,2},N_{1,3},\cdots,N_{K-1,K}}$

In this section we explore the structural complexity of the trees form the subspace of planted binary trees that satisfy Tokunaga law with Tokunaga parameters (a, c).

4.4.1 Self-similarity and Tokunaga self-similarity

In this section we introduce the concepts of tree self-similarity and Tokunaga self-similarity. Both of those concepts focus on side-branching, which is the merging between branches of different order, and are based on the Horton-Strahler numbers.

Definition 27. We call a random tree T_K of order K a *self-similar* tree if the matrix $(\mathbf{T}_{i,j})$ of Tokunaga indices, defined as

$$\mathbf{T}_{i,j} = \mathbf{T}_{i,j}[K] = \frac{\mathbb{E}[N_{i,j}[K]]}{\mathbb{E}[N_j[K]]},$$

is Toeplitz. That is, there is a (Tokunaga) sequence $\{\mathbf{T}_k\}$, such that

$$\mathbf{T}_{i,j} = \mathbf{T}_{j-i}.$$

Definition 28. We call a random tree T_K of order K a *Tokunaga self-similar* tree if it is self-similar and

$$\frac{\mathbf{T}_{k+1}}{\mathbf{T}_k} = c,$$

which is equivalent to

$$\mathbf{T}_k = ac^{k-1},$$

where $a, c > 0$ and $k \in [1, K-1]$.

It has been shown [36] that self-similarity implies Horton self-similarity with $\frac{1}{R}$ being a root of

$$\sum_{j=1}^{\infty} \mathbf{T}_j Z^j = 1 - 2Z.$$

In particular, for a Tokunaga self-similar tree with parameters (a, c), the Horton exponent R can be expressed via Tokunaga parameters a and c as follows [36, 47, 62]

$$R = R(a, c) = \frac{2 + a + c + \sqrt{(2 + a + c)^2 - 8c}}{2}. \tag{4.9}$$

4.4.2 Entropy rates for subspaces $\mathscr{T}_{a,c}$

Definition 29. Let $\mathscr{T}_{a,c} \subset \mathscr{T}_{K,R} \subset \mathcal{T}$ with $R = R(a, c)$ as in equation (4.9) be a subspace of planted binary trees of order K with Tokunaga numbers $N_{i,j}$, $i \leq j$, $i \in [1, K-1]$, $j \in [1, K]$ defines in the following way

$$\frac{N_{i,j}}{N_j} \in \left(ac^{j-i-1} - \beta^{j-i}, ac^{j-i-1} + \beta^{j-i}\right),$$

where N_i, $i \in [1, K]$ are the Horton-Strahler numbers, $a, c \in \mathbb{R}$, and $\beta \in (0, c)$.

Note that the number of nodes N grows asymptotically as $2R^{K-1}$, i.e.,

$$\left|1 - \frac{N}{2R^{K-1}}\right| < \left(\frac{\alpha}{R}\right)^{K-1} \to 0$$

as $K \to \infty$. Moreover, the number of leaves N_1 grows asymptotically as R^{K-1}, i.e.,

$$\left|1 - \frac{N_1}{R^{K-1}}\right| < \left(\frac{\alpha}{R}\right)^{K-1} \to 0$$

as $K \to \infty$.
The trees form the subspace $\mathscr{T}_{a,c}$ satisfy the Tokunaga law with parameters (a, c).
We consider relatively small neighborhood of values around ac^{j-i-1} since $N_{i,j}$ may not be exactly equal to ac^{j-i-1} (and a,c may not be rational). Note that Theorem 5 was proved under similar assumptions. Thus, without loss of generality, $\forall i, j$ such that $i \in [1, K-1]$, $j \in [1, K]$, and $i \leq j$ we assume the following

- $N_j = R^{K-j}$,
- $N_{i,j} = N_j ac^{j-i-1}$,
- $T_{i,j} = \frac{N_{i,j}}{N_j} = ac^{j-i-1}$.

Notice that these assumptions will not affect the final limit in the following result.

Theorem 6. *Consider a sequence of random trees T_K, each sampled uniformly from the corresponding subspace $\mathscr{T}_{a,c} \subset \mathcal{T}$. Then the entropy rate of a sequence T_K is given by*

$$\begin{aligned} \mathscr{H}_\infty(a, c) &= \frac{a}{2}\sum_{j=1}^{\infty} R^{-j}\left(\frac{1-c^j}{1-c} + a^{-1}\right)\log_2\left(\frac{1-c^j}{1-c} + a^{-1}\right) \\ &+ \frac{aR}{2(R-c)(R-1)} + \frac{\log_2 a}{2(R-1)} + \frac{-aRc\log_2 c}{2(c-R)^2(R-1)}, \end{aligned}$$

where $H(z) = -z\log_2 z - (1-z)\log_2(1-z)$ is a binary entropy of z and

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

Proof. We start with the definition of the entropy rate and obtain the following

$$\begin{aligned}
\mathscr{H}_\infty(a,c) &= \lim_{K\to\infty} \frac{\log_2 |\mathscr{T}_{K,N_{1,2},N_{1,3},\cdots,N_{K-1,K}}|}{2R^{K-1}} \\
&= \lim_{K\to\infty} \frac{1}{2R^{K-1}} \log_2 \left[\prod_{j=2}^{K} \prod_{i=1}^{j-1} 2^{N_{i,j}} \binom{N_j - 1 + \sum_{l=i}^{j-1} N_{l,j}}{N_{i,j}} \right] \\
&= \lim_{K\to\infty} \frac{1}{2R^{K-1}} \sum_{j=2}^{K} \sum_{i=1}^{j-1} \left[\log_2 2^{N_{i,j}} + \log_2 \binom{N_j - 1 + \sum_{l=i}^{j-1} N_{l,j}}{N_{i,j}} \right] \\
&= \lim_{K\to\infty} \frac{1}{2R^{K-1}} \left[\sum_{j=2}^{K} \sum_{i=1}^{j-1} N_{i,j} + \sum_{j=2}^{K} \sum_{i=1}^{j-1} \log_2 \binom{N_j - 1 + \sum_{l=i}^{j-1} N_{l,j}}{N_{i,j}} \right] \\
&= \lim_{K\to\infty} \frac{1}{2R^{K-1}} \left[\sum_{j=2}^{K} \sum_{i=1}^{j-1} N_{i,j} \right] \\
&+ \lim_{K\to\infty} \frac{1}{2R^{K-1}} \left[\sum_{j=2}^{K} \sum_{i=1}^{j-1} \log_2 \binom{N_j - 1 + \sum_{l=i}^{j-1} N_{l,j}}{N_{i,j}} \right]. \qquad (4.10)
\end{aligned}$$

Thus, the entropy rate $\mathscr{H}_\infty(a,c)$ is a sum of two terms presented in equation (4.10). Consider each term in (4.10) separately. We start with the first term.

$$\begin{aligned}
\lim_{K\to\infty} \frac{1}{2R^{K-1}} \left[\sum_{j=2}^{K} \sum_{i=1}^{j-1} N_{i,j} \right] &= \lim_{K\to\infty} \frac{1}{2R^{K-1}} \left[\sum_{j=2}^{K} \sum_{i=1}^{j-1} R^{K-j} a c^{j-i-1} \right] \\
&= \lim_{K\to\infty} \frac{1}{2R^{K-1}} \left[R^K a c^{-1} \sum_{j=2}^{K} R^{-j} c^j \sum_{i=1}^{j-1} c^{-i} \right] \\
&= \lim_{K\to\infty} \frac{1}{2R^{K-1}} \left[R^K a c^{-1} \sum_{j=2}^{K} R^{-j} c^j \frac{1 - c^{1-j}}{c-1} \right] \\
&= \lim_{K\to\infty} \frac{1}{2R^{K-1}} \left[R^K a c^{-1} \sum_{j=2}^{K} R^{-j} \frac{c^j - c}{c-1} \right] \\
&= \lim_{K\to\infty} \frac{1}{2R^{K-1}} \left[R^K a c^{-1} \sum_{j=2}^{K} \left(R^{-j} \frac{c^j}{c-1} - R^{-j} \frac{c}{c-1} \right) \right]
\end{aligned}$$

$$
\begin{aligned}
&= \lim_{K\to\infty} \frac{1}{2R^{K-1}} \left[\frac{R^K a c^{-1}}{c-1} \sum_{j=2}^{K} \left(\frac{c}{R}\right)^j - \frac{R^K a}{c-1} \sum_{j=2}^{K} \left(\frac{1}{R}\right)^j \right] \\
&= \lim_{K\to\infty} \frac{1}{2R^{K-1}} \left[\frac{ac(R^{K-1}-c^{K-1})}{(R-c)(c-1)} - \frac{(R^{K-1}-1)a}{(c-1)(R-1)} \right] \\
&= \lim_{K\to\infty} \left[\frac{ac}{2(R-c)(c-1)} \right] - \lim_{K\to\infty} \left[\frac{ac(c/R)^{K-1}}{2(R-c)(c-1)} \right] \\
&- \lim_{K\to\infty} \left[\frac{a}{2(R-1)(c-1)} \right] + \lim_{K\to\infty} \left[\frac{a(1/R)^{K-1}}{2(R-1)(c-1)} \right].
\end{aligned}
$$

Recall that parameters R, c, and a are connected by the relationship given in formula (11) in [71], i.e.,

$$
R = \frac{2+c+a+\sqrt{(2+c+a)^2-8c}}{2}. \tag{4.11}
$$

Thus,

$$
c = \frac{R(R-a-2)}{R-2} = R - R\frac{a}{R-2} < R, \tag{4.12}
$$

and therefore

$$
\frac{c}{R} < 1. \tag{4.13}
$$

Notice also that since $R \geq 2$, then

$$
\frac{1}{R} < 1. \tag{4.14}
$$

Thus, using inequalities (4.13) and (4.14), we find the first term in equation (4.10) as follows

$$
\begin{aligned}
\lim_{K\to\infty} \frac{1}{2R^{K-1}} \left[\sum_{j=2}^{K} \sum_{i=1}^{j-1} N_{i,j} \right] &= \lim_{K\to\infty} \frac{1}{2R^{K-1}} \left[\sum_{j=2}^{K} \sum_{i=1}^{j-1} N_{ij} \right] \\
&= \frac{ac}{2(R-c)(c-1)} - \frac{a}{2(R-1)(c-1)} \\
&= \frac{aR}{2(R-c)(R-1)}.
\end{aligned} \tag{4.15}
$$

Now, we consider the second term in the equation (4.10). We first notice that

$$
\sum_{l=i}^{j-1} N_{l,j} = N_j a \sum_{l=i}^{j-1} c^{j-l-1} = N_j a \sum_{k=0}^{j-i-1} c^k. \tag{4.16}
$$

Thus, using formula (4.16) and the fact that $N_{i,j} = N_j a c^{j-i-1}$, we can rewrite term

$$\binom{N_j - 1 + \sum_{l=i}^{j-1} N_{l,j}}{N_{i,j}}$$

as follows

$$\begin{aligned}\binom{N_j - 1 + \sum_{l=i}^{j-1} N_{l,j}}{N_{i,j}} &= \binom{N_j a \sum_{k=0}^{j-i-1} c^k + N_j - 1}{N_j a c^{j-i-1}} \\ &= \frac{\left(N_j a \sum_{k=0}^{j-i-1} c^k + N_j - 1\right)!}{(N_j a c^{j-i-1})! \left(N_j a \sum_{k=0}^{j-i-2} c^k + N_j - 1\right)!}. \end{aligned} \tag{4.17}$$

Multiplying both the numerator and the denominator in formula (4.17) by

$$\left(N_j a \sum_{k=0}^{j-i-2} c^k + N_j\right) \left(N_j a \sum_{k=0}^{j-i-1} c^k + N_j\right)$$

and regrouping the terms such that to obtain the following

$$\left(N_j a \sum_{k=0}^{j-i-1} c^k + N_j\right)! \quad \text{and} \quad \left(N_j a \sum_{k=0}^{j-i-2} c^k + N_j\right)!,$$

$$\begin{aligned}\binom{N_j - 1 + \sum_{l=i}^{j-1} N_{l,j}}{N_{i,j}} &= \binom{N_j a \sum_{k=0}^{j-i-1} c^k + N_j}{N_j a c^{j-i-1}} \frac{\left(N_j a \sum_{k=0}^{j-i-2} c^k + N_j\right)}{\left(N_j a \sum_{k=0}^{j-i-1} c^k + N_j\right)} \\ &= \binom{N_j a \sum_{k=0}^{j-i-1} c^k + N_j}{N_j a c^{j-i-1}} \frac{\left(a \sum_{k=0}^{j-i-2} c^k + 1\right)}{\left(a \sum_{k=0}^{j-i-1} c^k + 1\right)} \end{aligned} \tag{4.18}$$

Then, taking the logarithm, we obtain

$$\log_2 \binom{N_j - 1 + \sum_{l=i}^{j-1} N_{l,j}}{N_{i,j}} = \log_2 \binom{N_j a \sum_{k=0}^{j-i-1} c^k + N_j}{N_j a c^{j-i-1}} + \log_2 \left(\frac{\sum_{k=0}^{j-i-2} c^k + a^{-1}}{\sum_{k=0}^{j-i-1} c^k + a^{-1}} \right) \tag{4.19}$$

Thus, the second term in the equation (4.10) can be rewritten as a sun of two terms as follows

$$\lim_{K\to\infty} \frac{1}{2R^{K-1}} \left[\sum_{j=2}^{K} \sum_{i=1}^{j-1} \log_2 \binom{N_j - 1 + \sum_{l=i}^{j-1} N_{l,j}}{N_{i,j}} \right]$$

$$= \lim_{K\to\infty} \frac{1}{2R^{K-1}} \left(\sum_{j=2}^{K} \sum_{i=1}^{j-1} \log_2 \binom{N_j a \sum_{k=0}^{j-i-1} c^k + N_j}{N_j a c^{j-i-1}} \right) + \lim_{K\to\infty} \frac{1}{2R^{K-1}} \left(\sum_{j=2}^{K} \sum_{i=1}^{j-1} \log_2 \left(\frac{\sum_{k=0}^{j-i-2} c^k + a^{-1}}{\sum_{k=0}^{j-i-1} c^k + a^{-1}} \right) \right). \tag{4.20}$$

We start with the first term in equation (4.20). Using auxiliary Lemma 2, provided in the Section 6.2, we obtain

$$\log_2 \binom{N_j a \sum_{k=0}^{j-i-1} c^k + N_j}{N_j a c^{j-i-1}} = \left[N_j a \sum_{k=0}^{j-i-1} c^k + N_j \right] H \left[\frac{N_j a c^{j-i-1}}{N_j a \sum_{k=0}^{j-i-1} c^k + N_j} \right] + \mathcal{E}$$

$$= N_j a \left[\sum_{k=0}^{j-i-1} c^k + a^{-1} \right] H \left[\frac{c^{j-i-1}}{\sum_{k=0}^{j-i-1} c^k + a^{-1}} \right] + \mathcal{E},$$

where

$$\mathcal{E} = \mathcal{O}\left(\log_2\left(N_j a \sum_{k=0}^{j-i-1} c^k + N_j\right)\right)$$

and

$$\lim_{K\to\infty} \frac{1}{2R^{K-1}} \sum_{j=2}^{K}\sum_{i=1}^{j-1} \log_2\left(N_j a \sum_{k=0}^{j-i-1} c^k + N_j\right) = 0.$$

Therefore, we obtain

$$\sum_{j=2}^{K}\sum_{i=1}^{j-1} \log_2 \binom{N_j a \sum_{k=0}^{j-i-1} c^k + N_j}{N_j a c^{j-i-1}}$$

$$= \sum_{j=2}^{K}\sum_{i=1}^{j-1} N_j a \left(\sum_{k=0}^{j-i-1} c^k + a^{-1}\right) H\left(\frac{c^{j-i-1}}{\sum_{k=0}^{j-i-1} c^k + a^{-1}}\right)$$

$$= \sum_{j=2}^{K} N_j a \left(\sum_{i=1}^{j-1}\left(\sum_{k=0}^{j-i-1} c^k + a^{-1}\right) H\left(\frac{c^{j-i-1}}{\sum_{k=0}^{j-i-1} c^k + a^{-1}}\right)\right). \qquad (4.21)$$

Note now that the terms in the internal sum in equation (4.21) can be written in the following form

$$b_i H\left(\frac{d_i}{b_i}\right) = -d_i \log_2(d_i) - (b_i - d_i)\log_2(b_i - d_i) + b_i \log_2(b_i), \qquad (4.22)$$

where

$$b_i = \sum_{k=0}^{j-i-1} c^k + a^{-1} = \sum_{k=i}^{j-1} d_k + a^{-1},$$

$$d_i = c^{j-i-1},$$

and

$$b_i - d_i = \sum_{k=0}^{j-i-1} c^k + a^{-1} - c^{j-i-1} = \sum_{k=0}^{j-i-2} c^k + a^{-1} = b_{i+1}.$$

Thus,

$$\sum_{i=1}^{j-1}\left(\sum_{k=0}^{j-i-1} c^k + a^{-1}\right) H\left(\frac{c^{j-i-1}}{\sum_{k=0}^{j-i-1} c^k + a^{-1}}\right)$$

$$\begin{aligned}
&= \sum_{i=1}^{j-1} b_i H\left(\frac{d_i}{b_i}\right) = b_1 \log_2 b_1 - d_1 \log_2 d_1 - b_2 \log_2 b_2 \\
&+ b_2 \log_2 b_2 - d_2 \log_2 d_2 - b_3 \log_2 b_3 \\
&+ \cdots \\
&+ b_{j-1} \log_2 b_{j-1} - d_{j-1} \log_2 d_{j-1} - (b_{j-1} - d_{j-1}) \log_2 (b_{j-1} - d_{j-1}) \\
&= b_1 \log_2 b_1 - (b_{j-1} - d_{j-1}) \log_2 (b_{j-1} - d_{j-1}) - \sum_{i=1}^{j-1} (d_i \log_2 d_i) \\
&= \left(\sum_{k=0}^{j-2} c^k + a^{-1}\right) \log_2 \left(\sum_{k=0}^{j-2} c^k + a^{-1}\right) - a^{-1} \log_2 a^{-1} \\
&- \sum_{i=1}^{j-1} \left(c^{j-i-1} \log_2 c^{j-i-1}\right). \qquad (4.23)
\end{aligned}$$

Therefore, the first term in equation (4.20) can be expressed a sum of three terms as follows

$$\lim_{K\to\infty} \frac{1}{2R^{K-1}} \left(\sum_{j=2}^{K} \sum_{i=1}^{j-1} \log_2 \left(\frac{N_j a \sum_{k=0}^{j-i-1} c^k + N_j}{N_j a c^{j-i-1}}\right)\right)$$

$$\begin{aligned}
&= \lim_{K\to\infty} \frac{1}{2R^{K-1}} \sum_{j=2}^{K} a N_j \left(\sum_{k=0}^{j-2} c^k + a^{-1}\right) \log_2 \left(\sum_{k=0}^{j-2} c^k + a^{-1}\right) \\
&- \lim_{K\to\infty} \frac{1}{2R^{K-1}} \sum_{j=2}^{K} a N_j \left(a^{-1} \log_2 a^{-1}\right) \\
&- \lim_{K\to\infty} \frac{1}{2R^{K-1}} \sum_{j=2}^{K} a N_j \sum_{i=1}^{j-1} \left(c^{j-i-1} \log_2 c^{j-i-1}\right). \qquad (4.24)
\end{aligned}$$

Next, we find all three terms in (4.24) separately. We start by rewriting the first term as follows

$$\lim_{K\to\infty}\frac{1}{2R^{K-1}}\sum_{j=2}^{K}aN_j\left(\sum_{k=0}^{j-2}c^k+a^{-1}\right)\log_2\left(\sum_{k=0}^{j-2}c^k+a^{-1}\right)$$

$$\begin{aligned}
&= \lim_{K\to\infty}\frac{1}{2R^{K-1}}\sum_{j=2}^{K}aR^{K-j}\left(\sum_{k=0}^{j-2}c^k+a^{-1}\right)\log_2\left(\sum_{k=0}^{j-2}c^k+a^{-1}\right)\\
&= \lim_{K\to\infty}\frac{1}{2R^{K-1}}\sum_{j=2}^{K}aR^{K-j}\left(\frac{1-c^{j-1}}{1-c}+a^{-1}\right)\log_2\left(\frac{1-c^{j-1}}{1-c}+a^{-1}\right)\\
&= \frac{aR}{2}\sum_{j=2}^{\infty}R^{-j}\left(\frac{1-c^{j-1}}{1-c}+a^{-1}\right)\log_2\left(\frac{1-c^{j-1}}{1-c}+a^{-1}\right) \qquad (4.25)
\end{aligned}$$

Examine now the second term in (4.24).

$$\begin{aligned}
-\lim_{K\to\infty}\frac{1}{2R^{K-1}}\sum_{j=2}^{K}aN_j\left(a^{-1}\log_2 a^{-1}\right) &= -\lim_{K\to\infty}\frac{1}{2R^{K-1}}\sum_{j=2}^{K}N_j a a^{-1}\log_2 a^{-1}\\
&= -\lim_{K\to\infty}\frac{1}{2R^{K-1}}\log_2 a^{-1}\left(\frac{R^{K-1}-1}{R-1}\right)\\
&= \frac{\log_2 a}{2(R-1)}. \qquad (4.26)
\end{aligned}$$

Finally, consider the third term in (4.24). Note first, that

$$\begin{aligned}
\sum_{i=1}^{j-1}\left(c^{j-i-1}\log_2 c^{j-i-1}\right) &= c^{j-1}\sum_{i=1}^{j-1}\left(c^{-i}\log_2 c^{j-i-1}\right)\\
&= c^{j-1}\sum_{i=1}^{j-1}\left(c^{-i}(j-i-1)\log_2 c\right)\\
&= (-1)c^{j-1}(\log_2 c)\sum_{i=1}^{j-1}ic^{-i}+c^{j-1}(\log_2 c)(j-1)\sum_{i=1}^{j-1}c^{-i}\\
&= (\log_2 c)\frac{jc-c^j+1-j}{(c-1)^2}+(\log_2 c)(j-1)\frac{c^{j-1}-1}{c-1}
\end{aligned}$$

$$= \quad (\log_2 c) \times \frac{(j-2)c^j - (j-1)c^{j-1} + c}{(c-1)^2},$$

where the last equation was obtained using formulas (4.27) and (4.28)

$$\sum_{i=1}^{j-1} c^{-i} = \frac{1 - c^{1-j}}{c-1}, \tag{4.27}$$

and

$$\sum_{i=1}^{j-1} i c^{-i} = \frac{(j-1)c^{1-j} - jc^{2-j} + c}{(c-1)^2}. \tag{4.28}$$

The proof of formula (4.28) is provided in Lemma 1 in the Section 6.1. Thus,

$$\sum_{j=2}^{K} aN_j (\log_2 c) \frac{(j-2)c^j - (j-1)c^{j-1} + c}{(c-1)^2}$$

$$\begin{aligned}
&= \sum_{j=2}^{K} aR^{K-j} (\log_2 c) \frac{(1-c^{-1})jc^j + (c^{-1}-2)c^j + c}{(c-1)^2} \\
&= aR^K (\log_2 c) \frac{(1-c^{-1})}{(c-1)^2} \sum_{j=2}^{K} j \left(\frac{c}{R}\right)^j + aR^K (\log_2 c) \frac{(c^{-1}-2)}{(c-1)^2} \sum_{j=2}^{K} \left(\frac{c}{R}\right)^j \\
&+ aR^K (\log_2 c) \frac{c}{(c-1)^2} \sum_{j=2}^{K} \left(\frac{1}{R}\right) j \\
&= a(\log_2 c) \times \frac{R^{K-1}(2cR - c^2) + c^K (cK - KR - R)}{(c-1)(c-R)^2} \\
&+ a(\log_2 c) \times \frac{R^{K-1}(c - 2c^2) + c^K (2c-1)}{(c-1)^2 (R-c)} + a(\log_2 c) \times \frac{R^{K-1}c - c}{(c-1)^2 (R-1)},
\end{aligned}$$

where the last equation was obtained using formulas (4.29) and (4.30)

$$\sum_{j=2}^{K} r^j = \frac{r^2 - r^{K+1}}{1-r}, \tag{4.29}$$

and

$$\sum_{j=2}^{K} jr^j = r \times \left(\sum_{j=2}^{K} r^j\right)' = r \times \left(\frac{r^2 - r^{K+1}}{1-r}\right)'$$
$$= r \times \frac{2r - r^2 + r^K(rK - K - 1)}{(r-1)^2}. \quad (4.30)$$

Therefore, the third term in equation (4.24) can be found as follows

$$-\lim_{K\to\infty} \frac{1}{2R^{K-1}} \sum_{j=2}^{K} aN_j \sum_{i=1}^{j-1} \left(c^{j-i-1} \log_2 c^{j-i-1}\right)$$

$$\begin{aligned}
&= -\lim_{K\to\infty} \frac{1}{2R^{K-1}} \left(a(\log_2 c) \times \frac{R^{K-1}(2cR - c^2) + c^K(cK - KR - R)}{(c-1)(c-R)^2}\right) \\
&- \lim_{K\to\infty} \frac{1}{2R^{K-1}} \left(a(\log_2 c) \times \frac{R^{K-1}(c - 2c^2) + c^K(2c-1)}{(c-1)^2(R-c)}\right) \\
&- \lim_{K\to\infty} \frac{1}{2R^{K-1}} \left(a(\log_2 c) \times \frac{R^{K-1}c - c}{(c-1)^2(R-1)}\right) \\
&= -\lim_{K\to\infty} \frac{1}{2R^{K-1}} \left(\frac{a \log_2 c}{(c-R)^2(R-1)} R^K c\right) \\
&= \frac{-aRc \log_2 c}{2(c-R)^2(R-1)}. \quad (4.31)
\end{aligned}$$

Hence, the first term in equation (4.20) can be obtained by combining formulas (4.25), (4.26), and (4.31) as follows

$$\lim_{K\to\infty} \frac{1}{2R^{K-1}} \left(\sum_{j=2}^{K}\sum_{i=1}^{j-1} \log_2 \left(\frac{N_j a \sum_{k=0}^{j-i-1} c^k + N_j}{N_j a c^{j-i-1}}\right)\right) = \frac{\log_2 a}{2(R-1)} + \frac{-aRc \log_2 c}{2(c-R)^2(R-1)}$$

$$+ \frac{aR}{2} \sum_{j=2}^{\infty} R^{-j} \left(\frac{1-c^{j-1}}{1-c} + a^{-1}\right) \log_2 \left(\frac{1-c^{j-1}}{1-c} + a^{-1}\right).$$

Now consider the second term in the equation (4.20). Note that

$$\sum_{i=1}^{j-1} \log_2 \left(\frac{\sum_{k=0}^{j-i-2} c^k + a^{-1}}{\sum_{k=0}^{j-i-1} c^k + a^{-1}} \right) = \sum_{i=1}^{j-1} \left(\log_2 b_{i+1} - \log_2 b_i \right),$$

where

$$b_i = \sum_{k=0}^{j-i-1} c^k + a^{-1} = \sum_{k=0}^{j-i} c^k - c^{j-i} + a^{-1}.$$

Thus, equation (4.32) can be expanded as follows

$$\begin{aligned}
\sum_{i=1}^{j-1} \log_2 \left(\frac{\sum_{k=0}^{j-i-2} c^k + a^{-1}}{\sum_{k=0}^{j-i-1} c^k + a^{-1}} \right) &= \sum_{i=1}^{j-1} \left(\log_2 b_{i+1} - \log_2 b_i \right) \\
&= \log_2 b_2 - \log_2 b_1 \\
&+ \log_2 b_3 - \log_2 b_2 \\
&+ \log_2 b_4 - \log_2 b_3 \\
&+ \cdots \\
&+ \log_2 b_{j-1} - \log_2 b_{j-2} \\
&+ \log_2 b_j - \log_2 b_{j-1} \\
&= \log_2 b_j - \log_2 b_1 \\
&= \log_2 a^{-1} - \log_2 \left(\sum_{k=0}^{j-2} c^k + a^{-1} \right) \\
&= \log_2 a^{-1} - \log_2 \left(\frac{1 - c^{j-1}}{1 - c} + a^{-1} \right). \qquad (4.32)
\end{aligned}$$

Therefore,

$$\sum_{j=2}^{K} \left(\log_2 a^{-1} - \log_2 \left(\frac{1 - c^{j-1}}{1 - c} + a^{-1} \right) \right) = (K-1) \log_2 a^{-1} - \sum_{j=2}^{K} \log_2 \left(\frac{1 - c^{j-1}}{1 - c} + a^{-1} \right)$$

and, hence, the second term in equation (4.20) is equal to zero

$$\lim_{K\to\infty} \frac{1}{2R^{K-1}} \left(\sum_{j=2}^{K} \sum_{i=1}^{j-1} \log_2 \left(\frac{\sum\limits_{k=0}^{j-i-2} c^k + a^{-1}}{\sum\limits_{k=0}^{j-i-1} c^k + a^{-1}} \right) \right)$$

$$= \lim_{K\to\infty} \frac{1}{2R^{K-1}} \left((K-1)\log_2 a^{-1} - \sum_{j=2}^{K} \log_2 \left(\frac{1-c^{j-1}}{1-c} + a^{-1} \right) \right) = 0.$$

Therefore, the entropy rate can be obtained as follows

$$\begin{aligned}
\mathscr{H}_\infty(a,c) &= \frac{aR}{2} \sum_{j=2}^{\infty} R^{-j} \left(\frac{1-c^{j-1}}{1-c} + a^{-1} \right) \log_2 \left(\frac{1-c^{j-1}}{1-c} + a^{-1} \right) \\
&+ \frac{aR}{2(R-c)(R-1)} + \frac{\log_2 a}{2(R-1)} + \frac{-aRc\log_2 c}{2(c-R)^2(R-1)} \\
&= \frac{a}{2} \sum_{j=1}^{\infty} R^{-j} \left(\frac{1-c^{j}}{1-c} + a^{-1} \right) \log_2 \left(\frac{1-c^{j}}{1-c} + a^{-1} \right) \\
&+ \frac{aR}{2(R-c)(R-1)} + \frac{\log_2 a}{2(R-1)} + \frac{-aRc\log_2 c}{2(c-R)^2(R-1)},
\end{aligned} \tag{4.33}$$

where the last equation was obtained by factoring out $\frac{1}{R}$ from the infinite sum and changing the sum limits. □

4.4.3 Entropy rates for subspaces $\mathscr{T}_{c-1,c}$.

In this section we consider a special subspace of trees $\mathscr{T}_{c-1,c} \subset \mathscr{T}_{a,c}$, when parameter a satisfies a constraint $a = c - 1$. This important condition appears in several well-known models such as Random Self-similar Network (RSN) model and critical Tokunaga processes; see [37,42,65] and references therein. The next result adds additional evidence that the condition $a = c - 1$ in the parameter domain is special and needs to be further explored.

Theorem 7. *Consider a sequence of random trees T_K, each sampled uniformly from the* *corresponding subspace* $\mathscr{T}_{c-1,c} \subset \mathcal{T}$. *Then the entropy rate of a sequence* T_K *is given by*

$$\mathscr{H}_\infty(c-1,c) = 1 - \frac{1 - H\left(\frac{1}{c}\right)}{2 - \frac{1}{c}} = \mathscr{H}_\infty(R),$$

where $R = 2c$ *and* $H(z) = -z\log_2 z - (1-z)\log_2(1-z)$ *is a binary entropy of* z.

Proof. We start the proof by noticing that when $a = c-1$ the Horton exponent R can be found as follows

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

Thus, to find the entropy rate for a special case when $a = c-1$ we consider each term in formula (4.33) separately and substitute $a = c-1$ and $R = 2c$. We start with the first term and, by substituting $a = c-1$ and $R = 2c$, we obtain the following expression

$$\frac{a}{2}\sum_{j=1}^{\infty} R^{-j}\left(\frac{1-c^j}{1-c} + a^{-1}\right)\log_2\left(\frac{1-c^j}{1-c} + a^{-1}\right)$$

$$\begin{aligned}
&= \frac{c-1}{2}\sum_{j=1}^{\infty}(2c)^{-j}\left(\frac{c^j}{c-1}\right)\log_2\left(\frac{c^j}{c-1}\right) \\
&= \frac{1}{2}\sum_{j=1}^{\infty}\frac{1}{2^j}\left(j\log_2(c) - \log_2(c-1)\right) \\
&= \frac{\log_2(c)}{2}\sum_{j=1}^{\infty} j\frac{1}{2^j} - \frac{\log_2(c-1)}{2}\sum_{j=1}^{\infty}\frac{1}{2^j} \\
&= \frac{\log_2(c)}{2}(2) - \frac{\log_2(c-1)}{2}(1) \\
&= \log_2(c) - \frac{1}{2}\log_2(c-1). \qquad (4.34)
\end{aligned}$$

Now we consider the second term in formula (4.33). For $a = c-1$ and $R = 2c$ we get

$$\frac{aR}{2(R-c)(R-1)} = \frac{(c-1)2c}{2(2c-c)(2c-1)} = \frac{(c-1)}{(2c-1)}. \qquad (4.35)$$

Plugging in $a = c - 1$ and $R = 2c$ into the third term of formula (4.33), we obtain

$$\frac{\log_2 a}{2(R-1)} = \frac{\log_2(c-1)}{2(2c-1)}. \tag{4.36}$$

Finally, we obtain the fourth term as follows

$$\begin{aligned}
\frac{-aRc\log_2 c}{2(c-R)^2(R-1)} &= \frac{-(c-1)(2c)c\log_2 c}{2(2c-c)^2(2c-1)} \\
&= \frac{-(c-1)\log_2 c}{(2c-1)}.
\end{aligned} \tag{4.37}$$

Combining formulas (4.34), (4.35), (4.36), and (4.37) together we obtain the final expression for the entropy rate when $a = c - 1$ and $R = 2c$

$$\begin{aligned}
\mathscr{H}_\infty(c-1, c) &= \log_2(c) - \frac{1}{2}\log_2(c-1) + \frac{(c-1)}{(2c-1)} \\
&+ \frac{\log_2(c-1)}{2(2c-1)} + \frac{-(c-1)\log_2 c}{(2c-1)} \\
&= \frac{c-1}{2c-1} + \frac{c\log_2 c}{2c-1} - \frac{(c-1)\log_2(c-1)}{2c-1} \\
&= \frac{c-1}{2c-1} + \frac{cH\left(\frac{1}{c}\right)}{2c-1},
\end{aligned} \tag{4.38}$$

where the last equation was obtained using the fact that

$$H\left(\frac{1}{c}\right) = \frac{1}{c}\left(c\log_2 c - (c-1)\log_2(c-1)\right).$$

Finally,

$$\begin{aligned}
\mathscr{H}_\infty(c-1, c) &= \frac{c-1}{2c-1} + \frac{cH\left(\frac{1}{c}\right)}{2c-1} \\
&= 1 - \frac{1 - H\left(\frac{1}{c}\right)}{2 - \frac{1}{c}} \\
&= \mathscr{H}_\infty(R),
\end{aligned}$$

if $R = 2c$. $\square$

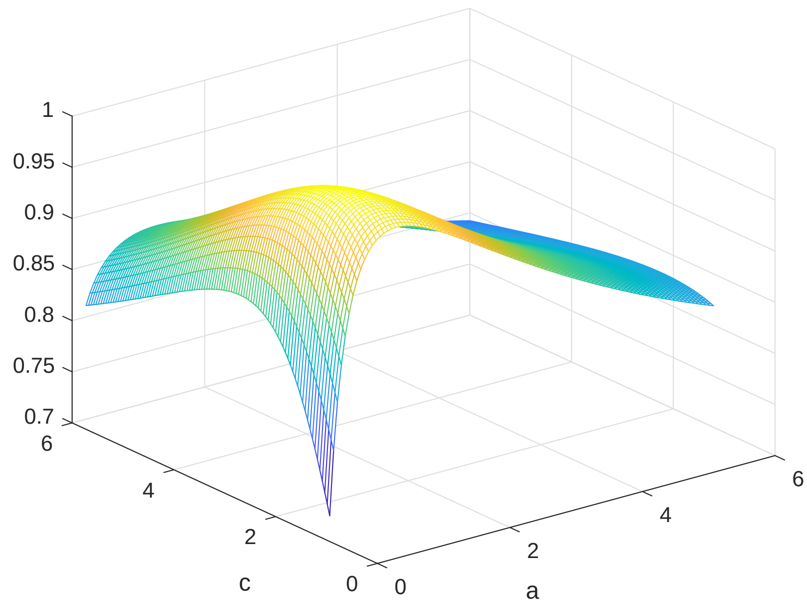

Figure 4.6: Entropy rate $\mathscr{H}_\infty(a, c)$ for $a \leq 6$ and $c \leq 6$. Note that the maximum is 1 and it is attained at $a = 1$ and $c = 2$.

Corollary 5. *The entropy rate $\mathscr{H}_\infty(a, c)$ is bounded from above by its maximal value 1, i.e.,*

$$\mathscr{H}_\infty(a, c) \leq 1.$$

Moreover, the maximum is attained at $a = 1$ and $c = 2$, i.e., $\mathscr{H}_\infty(1, 2) = 1$.

Proof. First, we note that

$$\begin{aligned} \mathscr{H}_\infty(a, c) &\leq \mathscr{H}_\infty(R(a, c)) \\ &\leq \mathscr{H}_\infty(4) = 1. \end{aligned}$$

Moreover, plugging in $a = 1$ and $c = 2$ into formula (4.33), we obtain $\mathscr{H}_\infty(1, 2) = 1$. $\square$

4.4.4 Discussion

In Figure 4.6 we depict the entropy rate $\mathscr{H}_\infty(a,c)$ for $a \leq 6$ and $c \leq 6$. The maximum of entropy rate is 1 and it is attained at $a = 1$ and $c = 2$, which corresponds to the $R = 4$ for entropy rate $\mathscr{H}_\infty(R)$. This fact provides the information-theoretic justification of the importance of the parameters $a = 1$ and $c = 2$. Observe also that,

$$\lim_{||(a,c)||\to\infty} \mathscr{H}_\infty(a,c) = \frac{1}{2}.$$

Similarly to the case of $\mathscr{H}_\infty(R)$, it would be interesting to investigate why for large values of the parameters a and c one requires on average $\frac{1}{2}$ bits per node to describe the tree.

In Figure 4.7 we demonstrate a map of Tokunaga parameters for several natural and synthetic processes. Note that natural hierarchical structures like river and drainage networks, botanical trees and and vein structure of botanical leaves have estimated Tokunaga parameters (a,c) that are close to the values $a = 1$ and $c = 2$. For example $a \approx 1.1$ and $c \approx 2.6$ for river basins [47, 74] and, consequently, the entropy rate for observed river basins is $\mathscr{H}_\infty(a,c) \approx 0.9916$. It was mentioned earlier, that the critical binary Galton-Watson model has $a = 1$, $c = 2$ and is often used to model real river networks. However, Tokunaga parameters for the real rivers deviate form that of critical binary Galton-Watson model. As in the case of Horton exponents, this nonoptimality of entropy rates (for example of the river basins) prompts questions about physical phenomena that need to be explained. Moreover, empirically observed values of Tokunaga parameters [47] (for example $a \approx 1.1$ and $c \approx 2.6$ for river basins) do not exactly satisfy the $a = c - 1$ condition. Explaining this phenomenon is an open problem.

There is also an interesting connection between the values of the parameter c and the fractal dimension of the trees. In particular, under the additional assumptions as in [38, 45], the fractal dimension is

$$d_c = \frac{\ln 2c}{\ln c} = 1 + \frac{\ln 2}{\ln c},$$

and therefore, depending on the value of the parameter c, the fractal dimension satisfies

$$1 \leq d_c < +\infty.$$

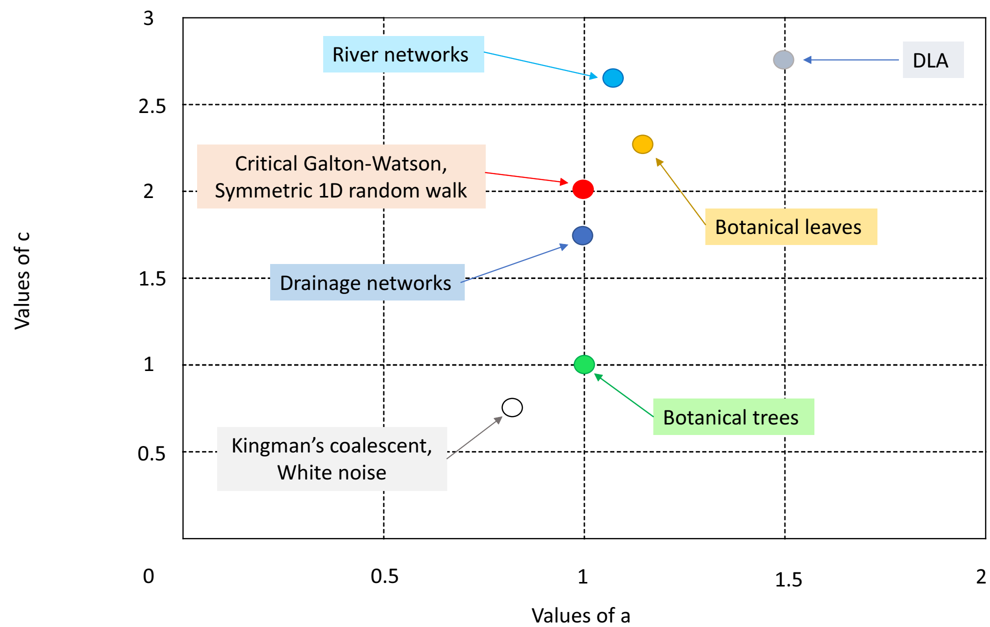

Figure 4.7: A map of Tokunaga parameters (a, c) for several natural and synthetic processes.

In general, as it was mentioned in [38], $\forall k \geq 1$, $d_{2\frac{1}{n}} = 1 + n$, which implies a volume-filling tree in $(1 + n)$ dimensional world. Recall, that the observed empirical range of the Horton exponent is $R \in (3, 6)$ and of the Tokunaga parameter is $c \in (1.4, 3)$. This corresponds to the fractal dimensions $1.6 < d_c < 3$. This range is enough to describe all possible tree dimensions that may exist in a real 3-dimensional world.

4.5 I-divergence analysis of entropy rates

In this section we examine both entropy rates $\mathscr{H}_\infty(R)$ and $\mathscr{H}_\infty(a, c)$ using the notion of I-divergence. We start with a few important definitions.

Definition 30. Function $D_{KL}(\cdot||\cdot) : \mathbb{R}_+^D \times \mathbb{R}_+^D \to \mathbb{R}_+$ is called the Kullback-Leibler divergence (also known as information gain or relative entropy) and is defined for vectors

$p, q \in \mathbb{R}_+^D$ such that $||p||_1 = 1$ and $||q||_1 = 1$ as follows

$$D_{KL}(p||q) = \begin{cases} \sum_{i=1}^{D} p(i) \log\left(\frac{p(i)}{q(i)}\right), & supp(p) \subseteq supp(q); \\ +\infty, & \text{otherwise}, \end{cases} \tag{4.39}$$

where the support of the distribution is defines as $supp(p) = \{i \in [1, D] : p(i) \neq 0\}$. Moreover, we use the following set of assumptions

- $0 \times (\pm\infty) = 0$
- $\log \frac{a}{0} = \infty$, $a \neq 0$
- $\log(0) = -\infty$.

The Kullback-Leibler divergence measures how one probability distribution diverges from a second probability distribution. Usually it is used to measure how the assumed probability distribution deviates from the true one. It was noted that Kullback-Leibler divergence plays a role of a (nonsymmetric) analogue of squared Euclidean distance for probability distributions [12, 13].

Definition 31. Function $I(\cdot||\cdot) : \mathbb{R}_+^D \times \mathbb{R}_+^D \to \mathbb{R}_+$ is the I-divergence function (also called the generalized Kullback-Leibler or Csiszar's divergence) and $\forall x, y \in \mathbb{R}_+^D$

$$I(x||y) = \begin{cases} \sum_{i=1}^{D} \left(x(i) \log\left(\frac{x(i)}{y(i)}\right) + y(i) - x(i) \right), & supp(x) \subseteq supp(y); \\ +\infty, & \text{otherwise}, \end{cases} \tag{4.40}$$

where $supp(x) = \{i \in [1, D] : x(i) \neq 0\}$.

Note that for formula (4.40) we use the same assumptions as for the formula (4.39).
I-divergence is the extension of the Kullback-Leibler divergence such that to allow quantification of the difference between functions, matrices, and sets [12, 13]. Note, I-divergence is also a special case of a Bregman divergence with a generating function $x \log_2 x$. Mutual information, Hamming distance, and precision and recall are just a few well-known examples of Bregman divergences. These and many other discrete Bregman distance measures naturally occur in a variety of computer science applications such as game theory, machine learning, and computer vision [4, 40].

Note, however, that I-divergence is not a metric in a classical sense, since it is not symmetric and does not satisfy a triangle inequality.
In the next Proposition, we explore entropy rate $\mathscr{H}_\infty(R)$ using I-divergence.

Proposition 1. *Consider a sequence of random trees T_K, each sampled uniformly from the corresponding subspace $\mathscr{T}_{K,R} \subset \mathcal{T}$. Then the entropy rate of a sequence T_K can be expressed in the following form*

$$\mathscr{H}_\infty(R) = \frac{1 + I\left(1||\frac{2}{R}\right)}{2 - \frac{2}{R}} - \frac{I\left(1 - \frac{2}{R}||\frac{2}{R}\right)}{2 - \frac{2}{R}}, \tag{4.41}$$

where $I(\alpha||\beta) = -\alpha \log_2\left(\frac{\alpha}{\beta}\right) + \beta - \alpha$ is I-divergence of the scalars α and β.

Proof. To obtain formula (4.41) we rewrite entropy rate $\mathscr{H}_\infty(R)$ as follows

$$\begin{aligned}
\mathscr{H}_\infty(R) &= 1 - \frac{1 - H(\frac{2}{R})}{2 - \frac{2}{R}} \\
&= \frac{2 - \frac{2}{R}}{2 - \frac{2}{R}} - \frac{1 - H\left(\frac{2}{R}\right)}{2 - \frac{2}{R}} \\
&= \frac{1 - \frac{2}{R} - \frac{2}{R}\log_2\left(\frac{2}{R}\right) - \left(1 - \frac{2}{R}\right)\log_2\left(1 - \frac{2}{R}\right)}{2 - \frac{2}{R}} \\
&= \frac{1}{2 - \frac{2}{R}} + \frac{-1 + \frac{2}{R} - \log_2\left(\frac{2}{R}\right)}{2 - \frac{2}{R}} \\
&- \frac{-1 + \frac{4}{R} + \left(1 - \frac{2}{R}\right)\log_2\left(1 - \frac{2}{R}\right) - \left(1 - \frac{2}{R}\right)\log_2\left(\frac{2}{R}\right)}{2 - \frac{2}{R}} \\
&= \frac{1 + I\left(1||\frac{2}{R}\right) - I\left(1 - \frac{2}{R}||\frac{2}{R}\right)}{2 - \frac{2}{R}},
\end{aligned} \tag{4.42}$$

where we used the fact that

$$I\left(1||\frac{2}{R}\right) = -\log_2\left(\frac{2}{R}\right) + \frac{2}{R} - 1$$

and

$$I\left(1 - \frac{2}{R}||\frac{2}{R}\right) = -1 + \frac{4}{R} + \left(1 - \frac{2}{R}\right)\log_2\left(1 - \frac{2}{R}\right) - \left(1 - \frac{2}{R}\right)\log_2\left(\frac{2}{R}\right).$$

□

Proposition 2. *Consider a sequence of random trees T_K, each sampled uniformly from the corresponding subspace $\mathscr{T}_{a,c} \subset \mathcal{T}$. Then the entropy rate of a sequence T_K can be expressed in the following form*

$$\begin{aligned}\mathscr{H}_\infty(a,c) &= \frac{a}{2}\left(I(x||y) + \frac{2}{(R-c)(2-\frac{2}{R})}\right)\\ &- \frac{a}{2}\left(\frac{2RI(\frac{1}{a}||1)}{(2-\frac{2}{R})} + 2\frac{(2c-R)+(c-2)+I(c||2)}{(R-c)^2\left(2-\frac{2}{R}\right)}\right),\end{aligned}$$

where $I(\alpha||\beta) = -\alpha\log_2\left(\frac{\alpha}{\beta}\right) + \beta - \alpha$ is I-divergence of scalars α and β and

$$I(x||y) = \sum_{i=1}^{\infty} -x_i \log_2\left(\frac{x_i}{y_i}\right) + y_i - x_i$$

is I-divergence of sequences $x_i = \frac{1}{R^i}\left(\frac{1-c^i}{1-c} + \frac{1}{a}\right)$ and $y_i = \frac{1}{R^i}$.

Proof. We start the proof by noticing that in the formula for the entropy rate

$$\begin{aligned}\mathscr{H}_\infty(a,c) &= \frac{a}{2}\sum_{j=1}^{\infty} R^{-j}\left(\frac{1-c^j}{1-c} + a^{-1}\right)\log_2\left(\frac{1-c^j}{1-c} + a^{-1}\right)\\ &+ \frac{aR}{2(R-c)(R-1)} + \frac{\log_2 a}{2(R-1)} + \frac{-aRc\log_2 c}{2(c-R)^2(R-1)}\end{aligned}$$

we can denote

$$x_j = \frac{1}{R^j}\left(\frac{1-c^j}{1-c} + \frac{1}{a}\right)$$

and

$$y_j = \frac{1}{R^j}.$$

Thus, the infinite sum can be rewritten as follows

$$\sum_{j=1}^{\infty} R^{-j}\left(\frac{1-c^j}{1-c} + a^{-1}\right)\log_2\left(\frac{1-c^j}{1-c} + a^{-1}\right) = \sum_{j=1}^{\infty} x_j \log_2\left(\frac{x_j}{y_j}\right).$$

Using the fact that I-divergence between two sequences is of the following form

$$I(x||y) = \sum_{j=1}^{\infty} x_j \log_2 \left(\frac{x_j}{y_j}\right) + \sum_{j=1}^{\infty} y_j - \sum_{j=1}^{\infty} x_j$$

we can express the infinite sum in the following way

$$\sum_{j=1}^{\infty} R^{-j} \left(\frac{1-c^j}{1-c} + a^{-1}\right) \log_2 \left(\frac{1-c^j}{1-c} + a^{-1}\right)$$

$$\begin{aligned}
&= I(x||y) - \sum_{j=1}^{\infty} \frac{1}{R^j} + \sum_{j=1}^{\infty} \frac{1}{R^j}\left(\frac{1-c^j}{1-c} + \frac{1}{a}\right) \\
&= I(x||y) - \frac{1}{R-1} + \frac{R}{(R-c)(R-1)} + \frac{1}{a(R-1)} \\
&= I(x||y) + \frac{R}{(R-c)(R-1)} + \frac{1-a}{a(R-1).}
\end{aligned}$$

Thus, the entropy rate is

$$\begin{aligned}
\mathscr{H}_{\infty}(a,c) &= \frac{a}{2} I(x||y) + \frac{aR}{2(R-c)(R-1)} + \frac{1-a}{2(R-1)} \\
&+ \frac{aR}{2(R-c)(R-1)} + \frac{\log_2 a}{2(R-1)} + \frac{-aRc\log_2 c}{2(c-R)^2(R-1)} \\
&= \frac{a}{2}\left(I(x||y) + \frac{2}{(R-c)(2-\frac{2}{R})}\right) \\
&+ \frac{a\left(\frac{1}{a} - 1 - \frac{1}{a}\log_2\left(\frac{1}{a}\right)\right)}{2(R-1)} + \frac{aR(R-c) - aRc\log_2 c}{2(R-c)^2(R-1)} \\
&= \frac{a}{2}\left(I(x||y) + \frac{2}{(R-c)(2-\frac{2}{R})}\right) + \frac{a}{2}\left(\frac{-I(\frac{1}{a}||1)}{(R-1)} + \frac{R-c-c\log_2 c}{(R-c)^2(1-\frac{1}{R})}\right) \\
&= \frac{a}{2}\left(I(x||y) + \frac{2}{(R-c)(2-\frac{2}{R})}\right) \\
&+ \frac{a}{2}\left(\frac{-I(\frac{1}{a}||1)}{(R-1)} + \frac{R-3c+2+(-c\log_2 c + 2c - 2)}{(R-c)^2(1-\frac{1}{R})}\right)
\end{aligned}$$

$$
\begin{aligned}
&= \frac{a}{2}\left(I(x||y) + \frac{2}{(R-c)(2-\frac{2}{R})}\right) \\
&+ \frac{a}{2}\left(\frac{-I(\frac{1}{a}||1)}{(R-1)} + \frac{(R-2c)+(2-c)+(-c\log_2 c + 2c - 2)}{(R-c)^2(1-\frac{1}{R})}\right) \\
&= \frac{a}{2}\left(I(x||y) + \frac{2}{(R-c)(2-\frac{2}{R})}\right) \\
&- \frac{a}{2}\left(\frac{2RI(\frac{1}{a}||1)}{\left(2-\frac{2}{R}\right)} + 2\frac{(2c-R)+(c-2)+I(c||2)}{(R-c)^2\left(2-\frac{2}{R}\right)}\right),
\end{aligned}
$$

where the last equation was obtained using the following formulas

$$I\left(\frac{1}{a}||1\right) = \frac{1}{a}\log_2\left(\frac{1}{a}\right) - \frac{1}{a} + 1$$

and

$$I(c||2) = c\log_2 c + 2 - 2c.$$

□

4.5.1 Discussion

First, consider the entropy rate as in formula (4.41)

$$\mathscr{H}_\infty(R) = \frac{1 + I\left(1||\frac{2}{R}\right)}{2-\frac{2}{R}} - \frac{I\left(1-\frac{2}{R}||\frac{2}{R}\right)}{2-\frac{2}{R}}, \tag{4.43}$$

which is a sum of two terms. Observe now that at the critical point $R = 4$,

$$I\left(1||\frac{1}{2}\right) = \frac{1}{2}$$

and

$$I\left(\frac{1}{2}||\frac{1}{2}\right) = 0.$$

Thus, the positive term is becomes 1, i.e.,

$$\frac{1 + I\left(1 || \frac{2}{4}\right)}{2 - \frac{2}{4}} = \frac{1 + \frac{1}{2}}{3.5} = 1$$

and the negative term disappears, i.e.,

$$\frac{I\left(1 - \frac{2}{4} || \frac{2}{4}\right)}{2 - \frac{2}{4}} = \frac{I\left(\frac{1}{2} || \frac{1}{2}\right)}{3.5} = 0.$$

Hence, the entropy attains its maximum value.
Similarly, the entropy rate as in formula (4.43)

$$\begin{aligned} \mathscr{H}_\infty(a, c) &= \frac{a}{2}\left(I(x||y) + \frac{2}{(R - c)(2 - \frac{2}{R})}\right) \\ &- \frac{a}{2}\left(\frac{2RI(\frac{1}{a}||1)}{\left(2 - \frac{2}{R}\right)} + 2\frac{(2c - R) + (c - 2) + I(c||2)}{(R - c)^2\left(2 - \frac{2}{R}\right)}\right) \end{aligned}$$

consist of one positive and one negative term. The entropy rate attains its maximum value 1, when $a = 1$ and $c = 2$ (and therefore $R = 4$). As in the case of $\mathscr{H}_\infty(R)$ the negative terms becomes zero at critical points $a = 1$ and $c = 2$.
Interestingly, this representations of both entropy rates $\mathscr{H}_\infty(R)$ and $\mathscr{H}_\infty(a, c)$ allow to eliminate negative terms, when parameters a, c, and R attain critical values 1, 2, and 4, respectively. Note also, that in both formulas the I-divergence $I(x||y)$ clearly plays a role similar to a distance between x and y. In general, this representation can be used for in depth interpretation of the results of Theorems 5, 6, and 7. This suggests a more general information theoretical approach to this type of problems.

Chapter 5: Conclusion

This work was motivated by the growing interest in statistical and complexity characteristics of tree-like structures. Our work was focused on studying entropy rates for dendritic structures that satisfy Horton and Tokunaga self-similarity - the property that can be observed in a variety of hierarchical complex systems. In particular, we considered several subspaces of finite unlabeled rooted planted binary plane trees with no edge length and examined structural complexity of those trees. Specifically, we calculated the number of planted binary trees with particular Horton-Strahler numbers and the number of planted binary trees with given Tokunaga numbers. We extend these results to classes of stemless trees and trees with a ghost edge. Moreover, we defined and evaluated the entropy for a subspace of trees with N vertices. We introduced the entropy rate measure in order to explain the long term behavior of growing tree model and find closed-form formulas for the entropy rate for a subspace of trees with N vertices. Moreover, we found entropy rate for a subspace of trees that satisfy the Horton Law with Horton exponent R and for a subspace of trees that satisfy Tokunaga Law with Tokunaga parameters (a, c). Furthermore, we used I-divergence measure to analyze entropy rates. The author is currently working on other important questions, overlapping theoretical information science and statistical self-similarity.

Chapter 6: Appendix

In this section we present several auxiliary lemmas that are used throughout this work.

6.1 Auxiliary Lemma 1

Lemma 1. *For any $a > 0$ and $i \geq 1$, $j \geq 2$, such that $i < j$ the following holds true*

$$\sum_{i=1}^{j-1} ia^i = \frac{a - ja^j + (j-1)a^{j+1}}{(1-a)^2}.$$

Proof.

$$\begin{aligned}
\sum_{i=1}^{j-1} ia^i &= a\sum_{i=1}^{j-1} ia^{i-1} = a\sum_{i=1}^{j-1} (a^i)' \\
&= a\left(\sum_{i=1}^{j-1} a^i\right)' = a\left(\frac{a-a^j}{1-a}\right)' \\
&= a \times \frac{1 - ja^{j-1} + (j-1)a^j}{(1-a)^2} \\
&= \frac{a - ja^j + (j-1)a^{j+1}}{(1-a)^2}.
\end{aligned}$$

□

6.2 Auxiliary Lemma 2

Lemma 2. *For positive integer k and m the following asymptotic approximation is true*

$$\log_2 \binom{k+m}{k} = (k+m)H\left(\frac{k}{k+m}\right) + \mathcal{O}(\log_2(k+m)), \tag{6.1}$$

as $k \wedge m \to \infty$*, where* $\binom{k+m}{k} = \frac{(k+m)!}{k!m!}$ *and* $H(z) = -z\log_2 z - (1-z)\log_2(1-z)$ *is a binary entropy of* z*.*

Proof. Using the Stirling's approximation $\log_2 n! = n\log_2 n - (\log_2 e)n + \mathcal{O}(\log_2 n)$ we obtain the required approximation as follows

$$\begin{aligned}
\log_2 \binom{k+m}{k} &= \log_2\left(\frac{(k+m)!}{k!m!}\right) \\
&= (k+m)\log_2(k+m) - (\log_2 e)(k+m) - k\log_2 k + (\log_2 e)k \\
&- m\log_2 m + (\log_2 e)m + \mathcal{O}(\log_2(k+m)) - \mathcal{O}(\log_2 k) - \mathcal{O}(\log_2 m) \\
&= (k+m)\log_2(k+m) - k\log_2 k - m\log_2 m \\
&+ k\log_2(k+m) - k\log_2(k+m) + \mathcal{O}(\log_2(k+m)) \\
&= (k+m)\left[-\frac{k}{k+m}\log_2\left(\frac{k}{k+m}\right) - \left[1-\frac{k}{k+m}\right]\log_2\left(1-\frac{k}{k+m}\right)\right] \\
&+ \mathcal{O}(\log_2(k+m)) \\
&= (k+m)H\left(\frac{k}{k+m}\right) + \mathcal{O}(\log_2(k+m)),
\end{aligned}$$

where $k \wedge m \to \infty$. □

6.3 Auxiliary Lemma 3

Lemma 3. *Let* $\mathscr{T}_{N_1,N_2,\cdots,N_K,R} \subset \mathscr{T}_{K,R} \subset \mathcal{T}$ *be a subspace of trees with particular set of Horton-Strahler numbers* $N_1, N_2, \cdots, N_K$*, such that* $N_k = R^{K-k} \pm \alpha^{K-k}$*,* $k = \overline{1,K}$*. Then*

$$\lim_{K\to\infty} \frac{\log_2 |\mathscr{T}_{N_1,N_2,\cdots,N_K,R}|}{2R^{K-1}} = 1 - \frac{1 - H(2/R)}{2 - 2/R},$$

where $|\mathscr{T}_{N_1,N_2,\cdots,N_K,R}|$ *is the number of trees in the subspace* $\mathscr{T}_{N_1,N_2,\cdots,N_K,R}$*. Here* N *grows asymptotically as* $2R^{K-1}$*, i.e.,*

$$\left|1 - \frac{N}{2R^{K-1}}\right| < \left(\frac{\alpha}{R}\right)^{K-1} \to 0$$

as $K \to \infty$*.*

Proof. We begin the proof by using the results of Lemma 1 that gives us the number of trees with a given set of Horton-Strahler numbers $N_1, N_2, \cdots, N_K$. Thus,

$$\begin{aligned}
\lim_{K\to\infty} \frac{\log_2 |\mathscr{T}_{N_1,N_2,\cdots,N_K,R}|}{2R^{K-1}} &= \lim_{K\to\infty} \frac{1}{2R^{K-1}} \log_2 \left(2^{N_1-1-\sum_{i=1}^{K-1} N_{i+1}} \prod_{i=1}^{K-1} \binom{N_i-2}{2N_{i+1}-2} \right) \\
&= \lim_{K\to\infty} \frac{1}{2R^{K-1}} \left[N_1 - 1 - \sum_{i=1}^{K-1} N_{i+1} \right] \\
&+ \lim_{K\to\infty} \frac{1}{2R^{K-1}} \left[\sum_{i=1}^{K-1} \log_2 \binom{N_i-2}{2N_{i+1}-2} \right]. \qquad (6.2)
\end{aligned}$$

Note that the term $\binom{N_i-2}{2N_{i+1}-2}$ in (6.2) can be rewritten in the following way

$$\begin{aligned}
\binom{N_i-2}{2N_{i+1}-2} &= \frac{(N_i-2)!}{(2N_{i+1}-2)!\,(N_i-2N_{i+1})!} \\
&= \frac{(N_i-2)!\,(N_i-2)\,(N_i-1)\,(2N_{i+1}-2)\,(2N_{i+1}-1)}{(2N_{i+1}-2)!\,(N_i-2N_{i+1})!\,(2N_{i+1}-2)\,(2N_{i+1}-1)\,(N_i-2)\,(N_i-1)} \\
&= \frac{(N_i)!}{(2N_{i+1})!\,(N_i-2N_{i+1})!} \frac{(2N_{i+1}-2)\,(2N_{i+1}-1)}{(N_i-2)\,(N_i-1)} \\
&= \binom{N_i}{2N_{i+1}} \frac{(2N_{i+1}-2)\,(2N_{i+1}-1)}{(N_i-2)\,(N_i-1)}.
\end{aligned}$$

Therefore, formula (6.2) can be written as a sum of four terms, as follows

$$\begin{aligned}
\lim_{K\to\infty} \frac{\log_2 |\mathscr{T}_{N_1,N_2,\cdots,N_K,R}|}{2R^{K-1}} &= \lim_{K\to\infty} \frac{1}{2R^{K-1}} \left[N_1 - 1 - \sum_{i=1}^{K-1} N_{i+1} \right] \\
&+ \lim_{K\to\infty} \frac{1}{2R^{K-1}} \sum_{i=1}^{K-1} \log_2 \binom{N_i}{2N_{i+1}} \\
&+ \lim_{K\to\infty} \frac{1}{2R^{K-1}} \sum_{i=1}^{K-1} \log_2 \frac{2N_{i+1}-2}{N_i-2} \\
&+ \lim_{K\to\infty} \frac{1}{2R^{K-1}} \sum_{i=1}^{K-1} \log_2 \frac{2N_{i+1}-1}{N_i-1}. \qquad (6.3)
\end{aligned}$$

We consider each of the four limits in (6.3) separately. Starting with the first limit, we notice that term $N_1 - 1 - \sum_{i=1}^{K-1} N_{i+1}$ can be rewritten in the following way

$$\begin{aligned}
N_1 - 1 - \sum_{i=1}^{K-1} N_{i+1} &= 2N_1 - 1 - \sum_{i=1}^{K} N_i \\
&= N - 1 - \sum_{i=1}^{K} \left(R^{K-i} \pm \alpha^{K-i}\right) \\
&= N - 1 - R^K \sum_{i=1}^{K} R^{-i} - (\pm 1)\alpha^K \sum_{i=1}^{K} \alpha^{-i} \\
&= N - 1 - R^K \frac{\frac{1}{R}\left(1 - \left(\frac{1}{R}\right)^K\right)}{\left(1 - \frac{1}{R}\right)} - (\pm 1)\alpha^K \frac{\frac{1}{\alpha}\left(1 - \left(\frac{1}{\alpha}\right)^K\right)}{\left(1 - \frac{1}{\alpha}\right)} \\
&= N - 1 - \frac{R^K - 1}{R - 1} - (\pm 1)\frac{\alpha^K - 1}{\alpha - 1}. \qquad (6.4)
\end{aligned}$$

Thus, dividing equation (6.4) by $2R^{K-1}$ and taking the limit as $K \to \infty$ we find the value of the first out of four limits in (6.3) as follows

$$\lim_{K\to\infty} \frac{1}{2R^{K-1}} \left[N_1 - 1 - \sum_{i=1}^{K-1} N_{i+1}\right]$$

$$\begin{aligned}
&= \lim_{K\to\infty} \frac{1}{2R^{K-1}} \left[N - 1 - \frac{R^K - 1}{R - 1} - (\pm 1)\frac{\alpha^K - 1}{\alpha - 1}\right] \\
&= 1 - \lim_{K\to\infty} \frac{1}{2R^{K-1}} \frac{R^K - 1}{R - 1} = 1 - \frac{R/2}{R - 1}. \qquad (6.5)
\end{aligned}$$

Consider now the second limit in equation (6.3). Using the result of Lemma 2, provided in Section 6.2, we can rewrite the second limit in (6.3) as follows

$$\lim_{K\to\infty} \frac{1}{2R^{K-1}} \sum_{i=1}^{K-1} \log_2 \binom{N_i}{2N_{i+1}} = \lim_{K\to\infty} \frac{1}{2R^{K-1}} \sum_{i=1}^{K-1} \left[N_i H\left(\frac{2N_{i+1}}{N_i}\right) + \mathcal{O}(\log_2 N_i)\right].$$

To examine the term

$$\sum_{i=1}^{K-1} N_i H\left(\frac{2N_{i+1}}{N_i}\right)$$

we break it into two sums as follows

$$\sum_{i=1}^{K-1} N_i H\left(\frac{2N_{i+1}}{N_i}\right) = \sum_{i=1}^{K'-1} N_i H\left(\frac{2N_{i+1}}{N_i}\right) + \sum_{i=K'}^{K-1} N_i H\left(\frac{2N_{i+1}}{N_i}\right), \tag{6.6}$$

where $K' = \lceil \frac{K}{2} \rceil$.
Consider the first sum in equation (6.6). Using the fact that $1 \leq i \leq K' - 1$, we obtain the following upper bound on term $\frac{2N_{i+1}}{N_i}$

$$\begin{aligned} \frac{2N_{i+1}}{N_i} &\leq 2\frac{R^{K-(i+1)} + \alpha^{K-(i+1)}}{R^{K-i} - \alpha^{K-i}} \\ &= \frac{2}{R} \times \frac{1 + \left(\frac{\alpha}{R}\right)^{K-(i+1)}}{1 - \left(\frac{\alpha}{R}\right)^{K-i}} \\ &\leq \frac{2}{R} \times \frac{1 + \left(\frac{\alpha}{R}\right)^{K-K'}}{1 - \left(\frac{\alpha}{R}\right)^{K-1}}. \end{aligned}$$

Thus,

$$\frac{2N_{i+1}}{N_i} \leq \frac{2}{R}\left(1 + \mathcal{O}\left(\left(\frac{\alpha}{R}\right)^{\frac{K}{2}}\right)\right). \tag{6.7}$$

In a similar fashion we obtain a lower bound on term $\frac{2N_{i+1}}{N_i}$

$$\frac{2}{R}\left(1 - \mathcal{O}\left(\left(\frac{\alpha}{R}\right)^{\frac{K}{2}}\right)\right) \leq \frac{2N_{i+1}}{N_i}. \tag{6.8}$$

Combining two bounds in (6.7) and (6.8) together, we obtain

$$\frac{2}{R}\left(1 - \mathcal{O}\left(\left(\frac{\alpha}{R}\right)^{\frac{K}{2}}\right)\right) \leq \frac{2N_{i+1}}{N_i} \leq \frac{2}{R}\left(1 + \mathcal{O}\left(\left(\frac{\alpha}{R}\right)^{\frac{K}{2}}\right)\right).$$

Since the entropy function $H(\cdot)$ has bounded derivative in any small enough closed neighborhood around $\frac{2}{R}$, we can bound term $H\left(\frac{2N_{i+1}}{N_i}\right)$ as follows

$$H\left(\frac{2}{R}\right)\left(1 - \mathcal{O}\left(\left(\frac{\alpha}{R}\right)^{\frac{K}{2}}\right)\right) \leq H\left(\frac{2N_{i+1}}{N_i}\right) \leq H\left(\frac{2}{R}\right)\left(1 + \mathcal{O}\left(\left(\frac{\alpha}{R}\right)^{\frac{K}{2}}\right)\right). \tag{6.9}$$

Using similar arguments, we obtain the following bounds on the term N_i

$$R^{K-i}\left(1-\left(\frac{\alpha}{R}\right)^{\frac{K}{2}}\right) \leq N_i \leq R^{K-i}\left(1+\left(\frac{\alpha}{R}\right)^{\frac{K}{2}}\right). \tag{6.10}$$

Thus, combining formulas (6.9) and (6.10) we get

$$\begin{aligned} R^{K-i}H\left(\frac{2}{R}\right)\left(1-\mathcal{O}\left(\left(\frac{\alpha}{R}\right)^{\frac{K}{2}}\right)\right) &\leq N_i H\left(\frac{2N_{i+1}}{N_i}\right) \\ &\leq R^{K-i}H\left(\frac{2}{R}\right)\left(1+\mathcal{O}\left(\left(\frac{\alpha}{R}\right)^{\frac{K}{2}}\right)\right). \end{aligned}$$

Therefore,

$$\begin{aligned} \sum_{i=1}^{K'-1} N_i H\left(\frac{2N_{i+1}}{N_i}\right) &\leq \sum_{i=1}^{K'-1} R^{K-i}H\left(\frac{2}{R}\right)\left(1+\mathcal{O}\left(\left(\frac{\alpha}{R}\right)^{\frac{K}{2}}\right)\right) \\ &\leq R^{K}H\left(\frac{2}{R}\right)\left(\sum_{i=1}^{K'-1} R^{-i}\right)\left(1+\mathcal{O}\left(\left(\frac{\alpha}{R}\right)^{\frac{K}{2}}\right)\right) \\ &= R^{K-1}H\left(\frac{2}{R}\right)\frac{1-1/R^{K'-1}}{1-1/R}\left(1+\mathcal{O}\left(\left(\frac{\alpha}{R}\right)^{\frac{K}{2}}\right)\right). \end{aligned} \tag{6.11}$$

Taking the limit as $K \to \infty$ in (6.11), we conclude that

$$\lim_{K\to\infty} \frac{1}{2R^{K-1}} \sum_{i=1}^{K'-1} N_i H\left(\frac{2N_{i+1}}{N_i}\right)$$

$$\begin{aligned} &\leq \lim_{K\to\infty} \frac{R^{K-1}}{2R^{K-1}} H\left(\frac{2}{R}\right)\frac{1-1/R^{K'-1}}{1-1/R}\left(1+\mathcal{O}\left(\left(\frac{\alpha}{R}\right)^{\frac{K}{2}}\right)\right) \\ &= H\left(\frac{2}{R}\right)\frac{R/2}{R-1}. \end{aligned} \tag{6.12}$$

Similarly, we show that

$$\sum_{i=1}^{K^{'}-1} N_i H\left(\frac{2N_{i+1}}{N_i}\right) \;\geq\; R^{K-1} H\left(\frac{2}{R}\right)\frac{1-1/R^{K^{'}-1}}{1-1/R}\left(1-\mathcal{O}\left(\left(\frac{\alpha}{R}\right)^{\frac{K}{2}}\right)\right)$$

and hence

$$\begin{aligned}
&\lim_{K\to\infty}\frac{1}{2R^{K-1}}\sum_{i=1}^{K^{'}-1} N_i H\left(\frac{2N_{i+1}}{N_i}\right) \\
\geq\;& \lim_{K\to\infty}\frac{R^{K-1}}{2R^{K-1}} H\left(\frac{2}{R}\right)\frac{1-1/R^{K^{'}-1}}{1-1/R}\left(1-\mathcal{O}\left(\left(\frac{\alpha}{R}\right)^{\frac{K}{2}}\right)\right) \\
=\;& H\left(\frac{2}{R}\right)\frac{R/2}{R-1}. \qquad (6.13)
\end{aligned}$$

Thus, combining formulas (6.12) and (6.13), we obtain

$$\lim_{K\to\infty}\frac{1}{2R^{K-1}}\sum_{i=1}^{K^{'}-1} N_i H\left(\frac{2N_{i+1}}{N_i}\right) = H\left(\frac{2}{R}\right)\frac{R/2}{R-1}. \qquad (6.14)$$

Consider now the second term in equation (6.6), where $K^{'} \leq i \leq K-1$. Using the fact that the entropy function is always bounded by 1 from above, we obtain

$$\begin{aligned}
\sum_{i=K'}^{K-1} N_i H\left(\frac{2N_{i+1}}{N_i}\right) &\leq \sum_{i=K'}^{K-1} N_i \\
&\leq \sum_{i=K'}^{K-1} \left(R^{K-i}+\alpha^{K-i}\right) \\
&= \sum_{k=1}^{K-K^{'}} \left(R^k+\alpha^k\right) \\
&= R\frac{R^{K-K^{'}}-1}{R-1} + \alpha\frac{\alpha^{K-K^{'}}-1}{\alpha-1} \\
&\leq \frac{R^{K-K^{'}+1}}{R-1} + \frac{\alpha^{K-K^{'}+1}}{\alpha-1}. \qquad (6.15)
\end{aligned}$$

Hence, dividing formula (6.15) by $2R^{K-1}$ and taking a limit as $K \to \infty$, we obtain

$$\begin{aligned}\lim_{K\to\infty} \frac{1}{2R^{K-1}} \sum_{i=K'}^{K-1} N_i H\left(\frac{2N_{i+1}}{N_i}\right) &= \lim_{K\to\infty} \frac{1}{2R^{K-1}} \left(\frac{R^{K-K'+1}}{R-1} + \frac{\alpha^{K-K'+1}}{\alpha - 1}\right) \\ &= \lim_{K\to\infty} \left(\frac{R^{K-1}}{2R^{K-1}}\left(\frac{R^{2-K'}}{R-1}\right) + \frac{\alpha^{K-1}}{2R^{K-1}}\left(\frac{\alpha^{2-K'}}{R-1}\right)\right) \\ &= 0.\end{aligned}$$

Therefore, the second limit in equation (6.2) is

$$\lim_{K\to\infty} \frac{1}{2R^{K-1}} \sum_{i=1}^{K-1} \log_2 \binom{N_i}{2N_{i+1}} = H\left(\frac{2}{R}\right) \frac{R/2}{R-1}. \tag{6.16}$$

Consider now the third term in equation (6.3). Note that since $N_i \geq 2N_{i+1}$

$$\frac{2N_{i+1} - 2}{N_i - 2} \leq 1.$$

Moreover, $\forall i = \overline{1, K-2}$

$$N_{i+1} \geq N_{K-1} \geq 2N_K = 2$$

and

$$N_i - 2 \leq N_i \leq N_1 \leq R^{K-1} + \alpha^{K-1}.$$

Therefore,

$$\begin{aligned}1 \geq \frac{2N_{i+1} - 2}{N_i - 2} &\geq \frac{2 \times 2 - 2}{R^{K-1} + \alpha^{K-1} - 2} \\ &\geq \frac{2}{R^{K-1} + \alpha^{K-1}} \\ &\geq \frac{2}{2R^{K-1}} = \frac{1}{R^{K-1}}.\end{aligned}$$

Thus,

$$0 \geq \log_2\left(\frac{2N_{i+1} - 2}{N_i - 2}\right) \geq (K-1)\log_2\left(\frac{1}{R}\right)$$

and

$$0 \geq \sum_{i=1}^{K-1} \log_2 \left(\frac{2N_{i+1} - 2}{N_i - 2} \right) \geq (K-1)^2 \log_2 \left(\frac{1}{R} \right). \tag{6.17}$$

Now we divide all sides of inequality in (6.17) by $2R^{K-1}$, take a limit as $K \to \infty$, and show that the third term in equation (6.3) is equal to zero

$$0 \geq \lim_{K\to\infty} \frac{1}{2R^{K-1}} \sum_{i=1}^{K-1} \log_2 \left(\frac{2N_{i+1} - 2}{N_i - 2} \right) \geq \lim_{K\to\infty} \frac{(K-1)^2}{2R^{K-1}} \log_2 \left(\frac{1}{R} \right) = 0. \tag{6.18}$$

Similarly, we show that the fourth term in equation (6.3) is equal to zero, i.e.,

$$\lim_{K\to\infty} \frac{1}{2R^{K-1}} \sum_{i=1}^{K-1} \log_2 \left(\frac{2N_{i+1} - 1}{N_i - 1} \right) = 0. \tag{6.19}$$

Thus, combining formulas (6.5), (6.16), (6.18), and (6.19) we find

$$\begin{aligned} \lim_{K\to\infty} \frac{\log_2 |\mathscr{T}_{N_1,N_2,\cdots,N_K,R}|}{2R^{K-1}} &= H\left(\frac{2}{R}\right) \frac{R/2}{R-1} + \left(1 - \frac{R/2}{R-1}\right) \\ &= 1 - \frac{1 - H(2/R)}{2 - 2/R}. \end{aligned}$$

□

Bibliography

[1] David J Aldous et al. Deterministic and stochastic models for coalescence (aggregation and coagulation): a review of the mean-field theory for probabilists. *Bernoulli*, 5(1):3–48, 1999.

[2] Maurice Stevenson Bartlett. *An introduction to stochastic processes: with special reference to methods and applications*. CUP Archive, 1978.

[3] Jean Bertoin. *Random fragmentation and coagulation processes*, volume 102. Cambridge University Press, 2006.

[4] Jeff Bilmes. Submodularity in machine learning applications. *AAAI, Tutorial*, 2015.

[5] Charles G Boncelet. *Probability, Statistics, and Random Signals*. Oxford University Press, 2016.

[6] Leon Brillouin. *Science and information theory*. Courier Corporation, 2013.

[7] Frederick P Brooks Jr. Three great challenges for half-century-old computer science. *Journal of the ACM (JACM)*, 50(1):25–26, 2003.

[8] Gregory A Burd, Edward C Waymire, Ronald D Winn, et al. A self-similar invariance of critical binary Galton-Watson trees. *Bernoulli*, 6(1):1–21, 2000.

[9] Colin Cherry and PJ Davis. On human communication. *Physics Today*, 10:36, 1957.

[10] Evgenia V Chunikhina. Entropy rates for Horton self-similar trees. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 28(8), 2018.

[11] Thomas M Cover and Joy A Thomas. Elements of information theory. 2012.

[12] Imre Csiszár. I-divergence geometry of probability distributions and minimization problems. *The Annals of Probability*, pages 146–158, 1975.

[13] Imre Csiszár, Paul C Shields, et al. Information theory and statistics: A tutorial. *Foundations and Trends® in Communications and Information Theory*, 1(4):417–528, 2004.

[14] M Vauchaussade De Chaumont. *Nombre de Strahler des arbres, langages algébriques et dénombrement de structures secondaires en biologie moléculaire*. Université de Bordeaux 1, 1985.

[15] Luc Devroye and Paul Kruszewski. A note on the Horton-Strahler number for random trees. *Information processing letters*, 52(3):155–159, 1994.

[16] Luc Devroye and Paul Kruszewski. On the Horton-Strahler number for random tries. *RAIRO-Theoretical Informatics and Applications*, 30(5):443–456, 1996.

[17] Peter Sheridan Dodds and Daniel H Rothman. Unified view of scaling laws for river networks. *Physical Review E*, 59(5):4865, 1999.

[18] Peter Sheridan Dodds and Daniel H Rothman. Scaling, universality, and geomorphology. *Annual Review of Earth and Planetary Sciences*, 28(1):571–610, 2000.

[19] Andrei P. Ershov. On programming of arithmetic operations. *Communications of the ACM*, 1(8):3–6, 1958.

[20] Ronald A Fisher. The logic of inductive inference. *Journal of the Royal Statistical Society*, 98(1):39–82, 1935.

[21] Philippe Flajolet, Jean-Claude Raoult, and Jean Vuillemin. The number of registers required for evaluating arithmetic expressions. *Theoretical Computer Science*, 9(1):99–125, 1979.

[22] Jean Françon. Sur le nombre de registres nécessaires à l'évaluation d'une expression arithmétique. *RAIRO. Informatique théorique*, 18(4):355–364, 1984.

[23] A Gabrielov, WI Newman, and DL Turcotte. Exactly soluble hierarchical clustering model: Inverse cascades, self-similarity, and scaling. *Physical Review E*, 60(5):5293, 1999.

[24] IM Gelfand and AM Yaglom. On the general definition of the quantity of information. In *Selected Works of AN Kolmogorov*, pages 2–5. Springer, 1993.

[25] Vijay K Gupta and Edward C Waymire. Some mathematical aspects of rainfall, landforms, and floods. In *Stochastic Methods In Hydrology*, pages 129–171. World Scientific, 1998.

[26] AYA Hinchin. Notion of entropy in the theory of probability. *Math. Sci*, 8(3):2–20, 1953.

[27] AYA Hinchin. On the basic theorems of information theory. *Uspekhi Matematicheckikh Nauk*, 11(1):17–75, 1956.

[28] Robert E Horton. Drainage-basin characteristics. *Eos, transactions american geophysical union*, 13(1):350–361, 1932.

[29] Robert E Horton. Erosional development of streams and their drainage basins; hydrophysical approach to quantitative morphology. *Geological society of America bulletin*, 56(3):275–370, 1945.

[30] Rainer Kemp. The average number of registers needed to evaluate a binary tree optimally. *Acta Informatica*, 11(4):363–372, 1979.

[31] John C Kieffer, E-H Yang, and Wojciech Szpankowski. Structural complexity of random binary trees. pages 635–639, 2009.

[32] James W Kirchner. Statistical inevitability of Horton's laws and the apparent randomness of stream channel networks. *Geology*, 21(7):591–594, 1993.

[33] A Kolmogorov. On the Shannon theory of information transmission in the case of continuous signals. *IRE Transactions on Information Theory*, 2(4):102–108, 1956.

[34] AN Kolmogorov. Theory of information and theory of algorithms. *Science. Moscow*, 1987.

[35] Andrei N Kolmogorov. Three approaches to the quantitative definition ofinformation'. *Problems of information transmission*, 1(1):1–7, 1965.

[36] Yevgeniy Kovchegov and Ilya Zaliapin. Horton law in self-similar trees. *Fractals*, 24(02):1650017, 2016.

[37] Yevgeniy Kovchegov and Ilya Zaliapin. Random self-similar trees and a hierarchical branching process. *arXiv preprint arXiv:1608.05032*, 2016.

[38] Yevgeniy Kovchegov and Ilya Zaliapin. Tokunaga self-similarity arises naturally from time invariance. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 28(4):041102, 2018.

[39] Yevgeniy Kovchegov, Ilya Zaliapin, et al. Horton self-similarity of Kingmans coalescent tree. In *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, volume 53, pages 1069–1107. Institut Henri Poincaré, 2017.

[40] Andreas Krause and Carlos Guestrin. Beyond convexity: Submodularity in machine learning. *ICML Tutorials*, 2008.

[41] Paul Kruszewski. A probabilistic exploration of the Horton-Strahler number for random binary trees. 1995.

[42] Ricardo Mantilla, Brent M Troutman, and Vijay K Gupta. Testing statistical self-similarity in the topology of river networks. *Journal of Geophysical Research: Earth Surface*, 115(F3), 2010.

[43] Bruce T Milne and Vijay K Gupta. Horton Ratios Link Self-Similarity with Maximum Entropy of Eco-Geomorphological Properties in Stream Networks. *Entropy*, 19(6):249, 2017.

[44] Markus E Nebel. On the Horton-Strahler number for combinatorial tries. *RAIRO-Theoretical Informatics and Applications*, 34(4):279–296, 2000.

[45] WI Newman, DL Turcotte, and AM Gabrielov. Fractal trees with side branching. *Fractals*, 5(04):603–614, 1997.

[46] Peter Ossadnik. Branch order and ramification analysis of large diffusion-limited-aggregation clusters. *Physical Review A*, 45(2):1058, 1992.

[47] Scott D Peckham. New results for self-similar trees with applications to river networks. *Water Resources Research*, 31(4):1023–1029, 1995.

[48] Scott D Peckham and Vijay K Gupta. A reformulation of Horton's laws for large river networks in terms of statistical self-similarity. *Water Resources Research*, 35(9):2763–2777, 1999.

[49] Jon D Pelletier and Donald 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):307–311, 2000.

[50] Jim Pitman. *Combinatorial Stochastic Processes: Ecole d'Eté de Probabilités de Saint-Flour XXXII-2002*. Springer, 2006.

[51] I Rodriguez-Iturbe and A Rinaldo. Fractal river networks: chance and self-organization, 1997.

[52] Claude E Shannon, Warren Weaver, and Arthur W Burks. The mathematical theory of communication. 1951.

[53] Claude Elwood Shannon. A mathematical theory of communication. *Bell system technical journal*, 27(3):379–423, 1948.

[54] Ronald L Shreve. Statistical law of stream numbers. *The Journal of Geology*, 74(1):17–37, 1966.

[55] Ronald L Shreve. Infinite topologically random channel networks. *The Journal of Geology*, 75(2):178–186, 1967.

[56] DE Knuth Sorting. Searching, The Art of Computer Programming, vol. 3, 1973.

[57] Arthur N Strahler. Hypsometric (area-altitude) analysis of erosional topography. *Geological Society of America Bulletin*, 63(11):1117–1142, 1952.

[58] Arthur N Strahler. Quantitative analysis of watershed geomorphology. *Eos, Transactions American Geophysical Union*, 38(6):913–920, 1957.

[59] David G Tarboton. Fractal river networks, Horton's laws and Tokunaga cyclicity. *Journal of hydrology*, 187(1-2):105–117, 1996.

[60] David G Tarboton, Rafael L Bras, and Ignacio Rodriguez-Iturbe. The fractal nature of river networks. *Water Resources Research*, 24(8):1317–1322, 1988.

[61] Alejandro Tejedor, Anthony Longjas, Douglas A Edmonds, Ilya Zaliapin, Tryphon T Georgiou, Andrea Rinaldo, and Efi Foufoula-Georgiou. Entropy and optimality in river deltas. *Proceedings of the National Academy of Sciences*, page 201708404, 2017.

[62] Eiji Tokunaga. Consideration on the composition of drainage networks and their evolution. 1978.

[63] DL Turcotte, BD Malamud, G Morein, and WI Newman. An inverse-cascade model for self-organized critical behavior. *Physica A: Statistical Mechanics and its Applications*, 268(3):629–643, 1999.

[64] DL Turcotte, JD Pelletier, and WI Newman. Networks with side branching in biology. *Journal of Theoretical Biology*, 193(4):577–592, 1998.

[65] Seth A Veitzer and Vijay K Gupta. Random self-similar river networks and derivations of generalized Horton laws in terms of statistical simple scaling. *Water resources research*, 36(4):1033–1048, 2000.

[66] G Viennot and M Vauchaussade De Chaumont. Enumeration of RNA secondary structures by complexity. In *Mathematics in Biology and Medicine*, pages 360–365. Springer, 1985.

[67] X Viennot, Georges Eyrolles, Nicolas Janey, and Didier Arques. Combinatorial analysis of ramified patterns and computer imagery of trees. In *ACM SIGGRAPH Computer Graphics*, volume 23, pages 31–40. ACM, 1989.

[68] Xavier Gérard Viennot. Trees everywhere. In *Colloquium on Trees in Algebra and Programming*, pages 18–41. Springer, 1990.

[69] Gleb Yakovlev, William I Newman, Donald L Turcotte, and Andrei Gabrielov. An inverse cascade model for self-organized complexity and natural hazards. *Geophysical Journal International*, 163(2):433–442, 2005.

[70] I Yekutieli and BB Mandelbrot. Horton-Strahler ordering of random binary trees. *Journal of Physics A: Mathematical and General*, 27(2):285, 1994.

[71] Ilia Zaliapin and Yevgeniy Kovchegov. Tokunaga and Horton self-similarity for level set trees of Markov chains. *Chaos, Solitons & Fractals*, 45(3):358–372, 2012.

[72] Ilya Zaliapin, Henry Wong, and Andrei Gabrielov. Hierarchical aggregation in percolation model. *Tectonophysics*, 413(1):93–107, 2006.

[73] Ilya Zaliapin, Henry Hang Lam Wong, and Andrei Gabrielov. Inverse cascade in a percolation model: Hierarchical description of time-dependent scaling. *Physical Review E*, 71(6):066118, 2005.

[74] S Zanardo, I Zaliapin, and E Foufoula-Georgiou. Are American rivers Tokunaga self-similar? New results on fluvial network topology and its climatic dependence. *Journal of Geophysical Research: Earth Surface*, 118(1):166–183, 2013.