Symmetry in Randomness

Additive Functionals and Symmetries
of Random Trees and Tree-Like Graphs

Christoffer Olsson

Dissertation presented at Uppsala University to be publicly examined in Siegbahnsalen, Ångströmslaboratoriet, Lägerhyddsvägen 1, Uppsala, Tuesday, 21 November 2023 at 09:15 for the degree of Doctor of Philosophy. The examination will be conducted in English. Faculty examiner: Professor Ralph Neininger (Goethe-Universität Frankfurt).

Abstract
Olsson, C. 2023. Symmetry in Randomness. Additive Functionals and Symmetries of Random Trees and Tree-Like Graphs. *Uppsala Dissertations in Mathematics* 134. 45 pp. Uppsala: Department of Mathematics. ISBN 978-91-506-3020-6.

Properties of symmetries in random trees and tree-like graphs are explored. The primary structures studied are Galton-Watson trees, unlabeled unordered trees as well as labeled subcritical graphs. The most significant results are exponential decay of the probability that two trees are isomorphic for some types of Galton-Watson trees and a central limit theorem for the logarithm of the size of the automorphism in all of the three models listed above, but a number of related theorems are also given including the limiting distribution of the number of labelings of an unlabeled unordered tree. An important tool is that of additive functionals of rooted trees and we also show how to extend the definition to the case of subcritical graphs together with powerful results that were previously only known for trees. Methods from both probability theory and analytic combinatorics are used.

Keywords: Random trees, Random graphs, Additive functionals, Automorphisms

Christoffer Olsson, Department of Mathematics, Probability Theory and Combinatorics, Box 480, Uppsala University, SE-751 06 Uppsala, Sweden.

List of papers

This thesis is based on the following papers, which are referred to in the text by their Roman numerals.

I   Olsson, C. The Probability of Random Trees Being Isomorphic. *Submitted*. https://arxiv.org/abs/2205.06625

II Olsson, C., Wagner, S. The Distribution of the Number of Automorphisms of Random Trees. *La Matematica* 2, 743–771 (2023). https://doi.org/10.1007/s44007-023-00064-z

III Olsson, C., Stufler, B. Additive functionals of subcritical graphs. *Preprint*.

Reprints were made with permission from the publishers.

Contents

1. Introduction

This is a study of symmetry in random objects making it an investigation into the interplay between chaos and pattern. The objects of interest are primarily random trees. To get a feeling for what a random tree is, consider the case of a family tree where one person gets a (beforehand) random number of children that then each gets a random number of children on their own and so on. We can then think of this family tree growing, randomly, over time. As mathematicians, we would be interested in the overall shape rather than the exact names of each person in the tree and might picture such a tree as in Figure 1.1. Even though we sweep exactly what "random" means under the rug for now, this is perhaps the most classical of all random tree models and this type of structure will occur frequently in the thesis. The example of a family tree might indicate possible connections between a random tree, seen as an abstract mathematical object, and fields such as biology and sociology, but it is also frequently occurring in the field of data science where the tree structure is used for more efficient data storage and retrieval. Thus, interest for random trees comes from different directions, motivating a deeper analysis of their properties. As is often the case in mathematics, we take the study one step back from the specific domain and study their properties from a general perspective where we hope to understand what is actually going on a fundamental level. It is then possible to take what we have learned and apply it to any field that might be interested in random trees when the need occurs.

Perhaps the first questions one would ask about random trees is what the probability is of picking one specific tree or the same tree twice when performing our random experiment, and this is certainly well-studied for all the classical models. A natural variant is to ask what the probability is when picking two trees at random, of getting two trees that look the same in some way that we decide beforehand. Certainly, one way of looking the same is to be the same tree, but this question is more general than the previous one since there are other ways for two trees to look the same. For example, we might consider
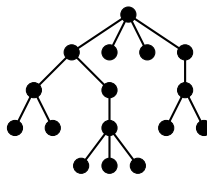

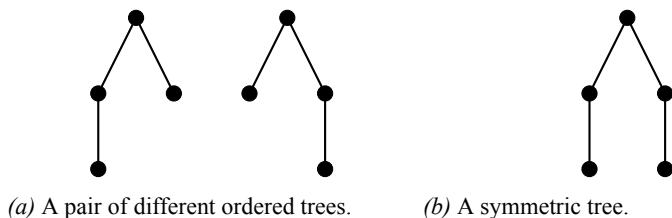*Figure 1.1.* A visual example of a rooted tree.

Figure 1.2.

two trees to be different yet look the same if they are mirror images of each other when we draw them in the plane (see e.g. Figure 1.2a). Studying these probabilities is one of the main topics of this thesis. Another question of interest is how symmetric a random tree is. Intuitively, we might argue that the tree in Figure 1.2b exhibits some sort of symmetry as it would look the same if we switched the order of the two branches. Mathematics has developed strong tools to study questions of symmetry so that it is possible to specify what this means. The (random) behavior of the number of symmetries in large random trees is another of the main topics in this thesis. While these questions have been the motivation for much of the research, we will also see that they lead to many generalizations and related problems that are also answered in the text.

After this, hopefully, more accessible introduction, we go on to define the mathematical objects, and problems, under study in a rigorous way. Recall that a graph is a set of vertices V together with a set of edges E between them, see Figure 1.3. The edges can themselves be thought of as pairs $e = \{v_1, v_2\}$ of vertices since they naturally correspond to such pairs. A tree is a graph without cycles, see Figure 1.1 for an example. Sometimes we mark one vertex as special, and call it the root of the tree.

Both graphs and trees can be unlabeled or labeled. In the latter case, we have some function $l : V \to [n]$ that assigns a number to each vertex. See Figure 1.4 for an example. For trees, especially, we also make a distinction between the unordered and ordered case, i.e., whether we place an ordering on the children of each vertex or not. An ordering of a tree also implies an embedding in the plane since we can decide to place the first child furthest to
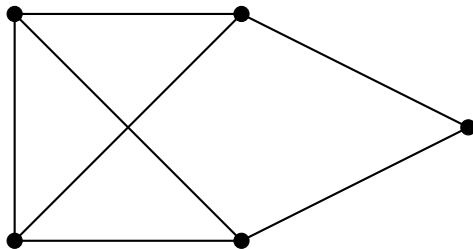

Figure 1.3. A visual example of a graph with five vertices and seven edges.
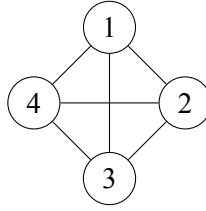
Figure 1.4. A labeled graph.

the left, the next one just to the right of it, and so on. The two trees in Figure 1.2a are different as ordered trees but not as unordered ones.

When we talk about a random tree or graph, we are considering some set of trees (or graphs) as a measure space, together with a probability measure to assign probabilities to different events, such as picking a specific tree. This means that picking a random tree can mean different things depending on which probability measure we are using, but for many models we pick one element uniformly at random among all objects of a given size n, meaning that all choices are equally likely. This corresponds to the uniform probability measure on the space of objects. Some examples of this would be random full rooted binary trees, random ordered rooted trees and random labeled unordered rooted trees, where we pick a tree of size n from the corresponding set uniformly at random. In this setting, studying the probabilistic properties of a random object lies very close to enumerating them, since the probability of picking a particular one is 1 divided by the total number of objects.

Due to the close connection to enumerative combinatorics, we will have reason to use the combinatorial notion of generating functions to study the trees and graphs that we are interested in. If $\{C_n\}_{n=0}^{\infty}$ is a sequence enumerating the number of objects, e.g. a type of tree of size n, we formally define the *generating function $f(x)$* of the object (or the sequence) to be

$$f(x) = \sum_{n=0}^{\infty} C_n x^n.$$

When dealing with labeled objects, we often use an *exponential generating function*,

$$g(x) = \sum_{n=0}^{\infty} \frac{C_n}{n!} x^n,$$

instead, where the factor $1/n!$ compensates for the large number of possible labelings that the object might have, which could otherwise lead to rapidly growing coefficients and series with poor convergence properties. We will use $[x^n]f(x)$ to denote the coefficient in front of x^n in the series $f(x)$. In many practical situations, we have some recursive formula for C_n which we can then translate to a functional equation for $f(x)$. If we can also show that this

function is an analytic function in some region of the complex plane, then we have a wealth of analytic tools available to obtain information about the nature of $\{C_n\}$. As we shall soon see, this will be one of our main tools throughout the thesis.

Fairly often, we are more interested in the behavior of the random objects when $n \to \infty$, rather than what happens for one particular size n. For example, we might study the asymptotic probability of picking one particular tree of large size n. Because of this, we will have use of the standard O notation as well as the following, well known, definitions. For two sequences $\{a_n\}$ and $\{b_n\}$, we will use $a_n = o(b_n)$ to mean that

$$\frac{a_n}{b_n} \to 0 \text{ as } n \to \infty,$$

and let $a_n \sim b_n$ denote the fact that

$$\frac{a_n}{b_n} \to 1 \text{ as } n \to \infty.$$

We will also make frequent use of *asymptotic expansions*. Let $\{c_j\}$ be some sequence of numbers, and let $\{\phi_j\}$ be a sequence of functions that satisfies $\phi_{j+1}(n) = o(\phi_j(n))$ as $n \to \infty$. Then we say that the formal series

$$\sum_{j=0}^{N} c_j \phi_j(n)$$

is an asymptotic expansion of a_n of order J if

$$a_n - \sum_{j=0}^{J} c_j \phi_j(n) = O(\phi_{J+1}(n)), \tag{1.1}$$

as n goes to infinity. Note that this means that $a_n \sim \sum_{j=0}^{J} c_j \phi_j(n)$, except in degenerate cases when the sum is equal to 0. If (1.1) holds for all J, we will write

$$a_n \sim \sum_{j=0}^{\infty} c_j \phi_j(n), \tag{1.2}$$

and call the (formal) series a *full asymptotic expansion*.

1.1 Models of random trees

While many types of random trees exist, we will focus on two of the most classical models, those being Galton–Watson trees (sometimes called Bienaymé–Galton–Watson trees), together with their close relative simply generated trees,

as well as unlabeled unordered trees, also called Pólya trees. We will primarily think of these trees as rooted, since not all types of Galton–Watson trees can be thought of as unrooted in a natural way. The book [7] serves as a good introduction to the families of trees encountered in this thesis.

Galton–Watson trees have a rich history, where early studies include [3] (see also [16] from which the reference is taken) and [37]. They can be thought of as a type of growth process that evolves over time, where we start with a single vertex, the root, and let it have a number of children according to some fixed random variable ξ that takes values in a subset of the non-negative integers that include at least 0 and some number larger than 1 (the restrictions are done to avoid degenerate cases involving infinitely large trees or trees without branch-points). We then let each of the children have a number of children of its own according to the same probability distribution and independently of all other vertices. By repeating this procedure, we obtain a tree that grows generation by generation. We are mainly interested in trees where we condition on the size of the tree being n, so called *conditioned Galton–Watson trees*. Many of our results then concern the behavior of conditioned Galton–Watson trees as $n \to \infty$.

There is a close connection between conditioned Galton–Watson trees and *simply generated trees*, first introduced in [24]. Simply generated trees are defined in terms of generating functions, and can thus be seen as the combinatorial counterpart to the probabilistically defined Galton–Watson trees. The generating function $T(x)$ of the simply generated tree associated with the *weight sequence* $\{w_k\}_{k=0}^{\infty}$ satisfies

$$T(x) = x\Phi(T(x)), \tag{1.3}$$

where

$$\Phi(x) = \sum_{k=0}^{\infty} w_k x^k$$

is called the *weight generating function*. The relation (1.3) captures the notion of recursion and can be read as "any simply generated tree consists of a root with a number of smaller trees attached to it as branches". Then, any tree T has a weight $w(T)$ associated with it, which is the product of the weights associated with each of its vertices. This can be seen by recursively expanding the definitions in (1.3). To get a random tree of size n from a simply generated class we pick one according to its weight

$$\mathbb{P}(\mathcal{T}_n = T) = \frac{w(T)}{\sum_{S \in \mathcal{T}_n} w(S)} = \frac{w(T)}{[x^n]T(x)}.$$

The connection between Galton–Watson trees and simply generated trees becomes apparent if we let $w_k = \mathbb{P}(\xi = k)$. Then, the weight $w(T)$ is the same as the probability of obtaining that particular tree from the growth process
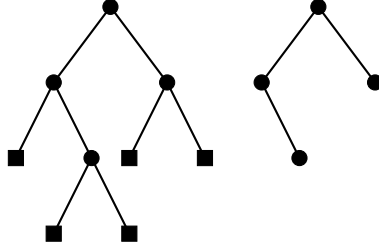
Figure 1.5. Example of a full binary tree and the corresponding pruned binary tree.

described above. We can choose different distributions for ξ (alternatively, different weight sequences $\{w_k\}$) to obtain different types of trees. Some examples of classical types of trees obtained in this fashion are full rooted binary trees, labeled unordered rooted trees and ordered rooted trees with the distribution generated by picking a tree uniformly at random. In the three cases, the distribution and weight sequence can, respectively, be taken to be: $\mathrm{Bin}(2, 1/2)$ and $\{1/2, 0, 1/2, 0, 0, \ldots\}$, $\mathrm{Pois}(1)$ and $\{1/k!\}_k$ (with the generating function being an exponential one), as well as $\mathrm{Ge}(1/2)$ and $\{1\}_k$.

As an aside, we recall that full binary trees (i.e., ordered trees where all vertices have outdegree 0 or 2) of size $2n + 1$ are in bijection with what we call *pruned* binary trees of size n. For pruned binary trees all vertices have outdegree 0, 1 or 2 and for vertices with only one child we distinguish between left children and right children. The reason that they are called pruned trees (and the source of the bijection) is that we can obtain them by removing all the leaves from a full binary tree. See Figure 1.5 for an example.

If we assume that there exists a positive τ within the radius of convergence of $\Phi(z)$ such that

$$\Phi(\tau) = \tau \Phi'(\tau) < \infty, \tag{1.4}$$

we can also find $\rho = \frac{\tau}{\Phi(\tau)}$ such that [7, Theorem 3.6]

$$[x^n]T(x) \sim C n^{-3/2} \rho^{-n}. \tag{1.5}$$

This shows that the weight or number of simply generated trees grows exponentially and in the context of Galton–Watson trees, this is a statement on the probability of obtaining a tree of size n.

If $\mathbb{E}\,\xi \leq 1$, the tree is almost surely finite, while $\mathbb{E}\,\xi > 1$ implies that the tree is infinite with some positive probability. The case $\mathbb{E}\,\xi = 1$ is called *critical* (we also speak of critical Galton–Watson trees when this holds), and is of special interest as then the probability $P(|\mathcal{T}| = n)$ decays polynomially in n (i.e., exponentially with rate 1), as opposed to the other cases where the probability decays exponentially. This is a direct consequence of the fact that in the critical case (1.4) is satisfied with $\tau = \rho = 1$. Assuming that the tree is critical turns out to not be a very strong restriction since, if we can find a τ as above, we can modify the weight sequence without affecting the probability measure on

the set of trees in such a way that $\Phi'(1) = 1$, i.e., such that the corresponding Galton–Watson trees are critical [7, Subsection 1.2.7]. This does not only hold for conditioned Galton–Watson trees, but also for other simply generated trees satisfying the condition, and it is often possible to translate the results for one type of random tree to the other. Because of this, we will often use the terms interchangeably.

One way to study the behavior of conditioned Galton–Watson trees as $n \to \infty$ is to restrict attention to the first M levels (or generations) of the tree. This procedure will give a new tree, of size at most n, and one can try to categorize the probability distribution on trees we get by this procedure for large conditioned Galton–Watson trees. Ideally, we want to do this for all M at the same time. It is possible to define a metric space on rooted trees (related to the so called Ulam-Harris tree) where convergence of trees is equivalent to convergence of every restriction $T_n^{(M)} \to T^{(M)}$. In this space, the conditioned and critical Galton–Watson tree $\mathcal{T}_n$ converges weakly to the *size-biased Galton–Watson tree*, also called Kesten's tree. See [20] and [23] for more details in this direction. This tree can be constructed by a modification of the Galton–Watson growth process where we have two different types of vertices which we call normal and special, respectively. The normal vertices get a number of children in the same way as described for the usual Galton–Watson trees, while special vertices get a number of children according to the size-biased random variable $\hat{\xi}$ with distribution $\mathbb{P}(\hat{\xi} = k) = k\,\mathbb{P}(\xi = k)$ with exactly one of them being designated as special. We define the root to be special and observe that the size-biased tree always has an infinite spine of vertices, starting at the root, of special vertices. The size-biased tree will primarily occur in intermediate steps in some of our arguments and can, in the context of this thesis, be considered a technicality.

Since Galton–Watson trees are defined through a growth process that starts with the root, there is no natural way to consider unrooted Galton–Watson trees in the general case. However, some special cases have an unrooted counterpart through their combinatorial interpretation. For example, we can root any labeled unordered unrooted tree in n different ways by simply picking one of its vertices as the root. This means that picking any unrooted such tree uniformly at random has the same probability as picking any of its rooted versions when picking a labeled unordered rooted tree uniformly at random. In other words, we get a natural connection between the unrooted and rooted case by considering the former to be a sub-σ-algebra of the latter.

Unlabeled unordered rooted trees, or *Pólya trees*, do not belong to the family of Galton–Watson trees, although they are not too far off as they essentially contain a Galton–Watson tree at its core [27] and share many of the same properties. From the combinatorial perspective, we can describe Pólya trees in terms of their generating function $P(x)$ which keeps track of the number of unlabeled unordered rooted trees of size n. This generating function satisfies

the functional equation

$$P(x) = x \exp\left(\sum_{k=1}^{\infty} \frac{P(x^k)}{k}\right). \tag{1.6}$$

One can show that the number of trees of size n is asymptotic to $An^{-3/2}\rho_p^{-n}$, where A is a constant and $\rho_p = 0.33832\ldots$ is the dominant (i.e., closest to the origin) singularity of $P(x)$. For this singularity, we have $P(\rho_p) = 1$.

The relationship between rooted and unrooted unlabeled unordered trees can be described by the functional equation

$$U(x) = P(x) - \frac{1}{2}P(x)^2 + \frac{1}{2}P(x^2) \tag{1.7}$$

that connects the generating function of unrooted Pólya trees $U(x)$ to the one for rooted trees $P(x)$. We will use Pólya tree and unlabeled unordered tree interchangeably, even in the unrooted case. The equation relies on a bijection between Pólya trees and the union of unrooted unlabeled trees together with pairs of distinct Pólya trees, a classical result from [25]. The number of unrooted Pólya trees of size n is asymptotic to $Bn^{-5/2}\rho_p^{-n}$ for a constant B. Thus, the asymptotic behavior of unrooted Pólya trees is similar to that of rooted ones but with a different constant and a factor $n^{-5/2}$ instead of $n^{-3/2}$.

Sometimes we will be interested in trees with out-degrees restricted to lie in some finite set. This is natural in the context of simply generated trees as we can simply set all large weights to 0. For Pólya trees $\mathcal{P}_D$ with degrees restricted to lie in some set D we will instead define the generating function

$$P_D(x) = \sum_{P \in \mathcal{P}_D} x^{|P|},$$

with a corresponding modification to the functional equation (1.6)

$$P_D(x) = x \sum_{k \in D} \sum_{\lambda \vdash k} \prod_j \frac{P_D(x^j)^{\lambda_j}}{j^{\lambda_j}\lambda_j!}.$$

We will use $\mathcal{T}$ to denote a class of Galton–Watson trees or a random such tree, while T denotes a specific realization of it. We will use $\mathcal{P}$ and P in the same way for Pólya trees. We also let $\deg(v)$ denote the outdegree of a vertex v and $\deg(T)$ denote the outdegree of the root. We will use the term *root branches* (or branches) to denote the trees obtained as connected components if we remove the root.

1.2 Models of random graphs

While random trees, seen as an area in random graph theory, has received significant attention, the wider subject of random graph models is a flourishing topic in its own right. Here the most famous structure is undoubtedly the

Erdős-Rényi-Gilbert graph $G(n, p)$ where we start with the complete graph on n vertices and then, for each edge, keep it with probability p, independently of all others ([10],[14]). While much effort has gone into studying this, and other random graph models, many questions are much harder than they are for random trees due to the more complex structure. A way of amending this problem is to study graphs that are sufficiently tree-like. One could then hope to adapt tools that have been successful in the study of trees to this setting as well.

One approach to this is the study of block-stable graph classes and, their specification, subcritical graphs [2], [8], [15]. We will use *block* to denote a 2-connected component of a graph. A graph class $\mathcal{G}$ is *block-stable* if it contains the graph K_2, consisting of a single edge, and has the property that a graph G belongs to $\mathcal{G}$ if and only if all the blocks of G also belong to the class.

Subcritical graphs are a special type of block-stable graphs that we define in terms of analytic properties of their generating functions. It is possible to study both labeled and unlabeled graphs in this framework. The conditions imposed are similar so, to avoid inconvenient notation, we give the definition in the labeled case and refer the interested reader to the original article [8] for a proper definition in the unlabeled case. We therefore let $\mathcal{G}$ be a block-stable class of labeled graphs. Define $\mathcal{C}$ to be the class of connected graphs in $\mathcal{G}$ and $\mathcal{B}$ be the subclass of blocks (together with K_2). We will make use of rooted graphs in the definition and use $\bullet$ to denote rooted objects so that $\mathcal{C}^{\bullet}$ is the class of rooted connected graphs. We use $\mathcal{G}(z)$, $\mathcal{C}^{\bullet}(z)$ and $\mathcal{B}(z)$ to denote the exponential generating functions of the relevant graph classes. It can be shown that the exponential generating functions of any block-stable class satisfies

$$\mathcal{G}(z) = \exp(\mathcal{C}(z)),$$
$$\mathcal{C}^{\bullet}(z) = z \exp(\mathcal{B}'(\mathcal{C}^{\bullet}(z))).$$

This recursive description is very similar to what we have for some types of trees and one way to think of connected block-stable graphs is, in fact, as a tree where the vertices themselves are blocks and the edges between them symbolize vertices lying in two blocks (and thereby connecting them).

A subcritical graph has some additional analytic requirements attached to it. Let ρ_B be the radius of convergence of $\mathcal{B}'(z)$ and ρ_C be the radius of convergence of $\mathcal{C}^{\bullet}(z)$. We now say that the class $\mathcal{G}$ is subcritical if

$$\mathcal{C}^{\bullet}(\rho_C) < \rho_B.$$

Some examples of subcritical graph classes are block graphs, cacti, outerplanar graphs and series-parallel graphs. On the other hand, the class of planar graphs is an example of a block-stable class that is not subcritical. It was shown in [26] that the largest block of a subcritical graph of size n is of size $O(\log n)$. This can be compared to planar graphs where a graph of size n has one giant block that contains a fraction of n of the vertices.

There is actually more than one way to interpret a block-stable graph as a tree structure. An alternate viewpoint can be found in [28] and [35] (see also

[33] for the more complicated case of unlabeled graphs), where it is shown that a subcritical graph (or, more generally, a block-stable graph) can be seen as a random *decorated tree*. A decorated (also called enriched) tree is a rooted tree where we add some object from a specified combinatorial class C to each vertex in the tree. Specifically, for every vertex v, we divide its children into subsets and then add the structure of an object from the class to each such subset together with v. See Figure 1.6 for an example including block graphs where divide the children of a vertex into a number of subsets and decorate each of them with the structure of a complete graph. Based on the results in the papers, we can see a random labeled subcritical graph as a Galton–Watson tree with decorations added randomly (depending only on the outdegree of each vertex). Since the outdegree of any vertex is finite and since there are only a finite number of objects in C, a suitable probability measure is easy to define. This explicit representation of subcritical graphs in terms of trees will let us transfer tools developed in the study of random trees to the case of graphs. Note that the definition of a decorated tree is very general and also covers other structures than block-stable graphs as long as we use the proper decorations.

Unlabeled graphs can instead be seen as decorated sesqui-type trees, first introduced in [22]. This type of tree can be described by a branching process with two types of vertices that we call L and S. The vertices of type L get a number of children of type L according to a random variable ξ and a number of children of type S according to a different random variable ζ. We can also view this as a random vector $\mathbf{X} = [\xi, \zeta]$. The offspring of different vertices are assumed to be independent but ξ and ζ for a given vertex can depend on each other. The vertices of type S are infertile and do not get any offspring. Alternatively, we can think of sesqui-type trees as a Galton–Watson tree consisting of vertices of type L to each of whose vertices we attach a random number of additional children of type S. The connection to unlabeled structures comes from viewing the vertices of type L as being the fixed-points under some automorphism of the object and decorating the vertices of type S to have the structure of a number of isomorphic branches. Again, we can also describe other objects than subcritical graphs in this way. One example (or a special case) is that of Pólya trees.

Figure 1.6. A rooted block graph together with its underlying decorated tree. The colors highlight how the children of the root are divided into subsets.
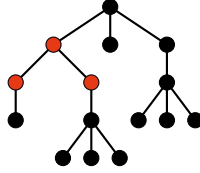
Figure 1.7. A cherry occurring as a pattern in a tree (in red).

1.3 Additive functionals of random trees

Many parameters of trees can naturally be thought of in a recursive way. Consider, for example, the number of leaves in a tree. This number can be found by counting the number of leaves in each of the root branches and adding them up, except if there are no branches. In the latter case we find the number of leaves by adding up the contribution from the branches (which is 0 in this case) and then adding 1 to account for the fact that the root is a leaf. This phenomenon is captured by the concept of an *additive functional* $F(T)$ of rooted trees T. To the additive functional $F(T)$ we associate a so-called *toll function* $f(T)$. Then, the definition of an additive functional is

$$F(T) = f(T) + \sum_B F(B),$$

where the sum is over all root branches. Intuitively, $f(T)$ describes the contribution coming from the root, while the sum captures the contribution from the branches. In the example given above, $f(T) = \mathbf{1}\{|T| = 1\}$, an indicator function for the root being the only vertex in the tree.

If we expand the definition, we find that we can also write the additive functional as a sum over all fringe subtrees T_v:

$$F(T) = \sum_v f(T_v),$$

where we recall that the fringe subtree T_v at the vertex v is a subtree of T that contains v as the root together with all of its descendants.

Other parameters that can be described as additive functionals are the number of vertices of outdegree $m \in \{0, 1, 2, \ldots\}$ ($m = 0$ gives leaves, as above), the number of occurrences of a given tree as fringe subtree, the number of occurrences of more general patterns like cherries (see Figure 1.7), and the logarithm of the total number of subtrees in a tree. In fact, all parameters of rooted trees can be described in this way by the right choice of toll function. However, it is not always the case that the toll function is easy to understand and the framework of additive functionals is therefore not always suitable to use.

Due to their generality, much effort has gone into studying additive functionals in a number of different models of random trees, both when it comes

to moments and central limit theorems. For additive functionals in the context of Galton–Watson trees, see e.g. [11], [12], [21], [31] and [36]. Less work has been done in the context of Pólya trees, but see e.g. [36]. The topic has also been studied for models of trees of lesser interest to this thesis, such as m-ary search trees and recursive trees (see for example [6], [11], [17], [18], [19]) as well as m-ary increasing trees and generalized plane oriented recursive trees (GPORTS) ([30]). In conclusion, we can find the moments and even central limit theorems for various random tree models under fairly general conditions. We will both use and prove these types of results and additive functionals on random trees can be considered one of the main themes of this thesis.

One result that will be of particular interest to us is the central limit theorem for *almost local* additive functionals from [31] which builds on results from [21].

Theorem 1.3.1. *Let $\mathcal{T}_n$ be a conditioned Galton–Watson tree of size n with offspring distribution ξ, with $\mathbb{E}\,\xi = 1$ and $0 < \sigma^2 := \operatorname{Var} \xi < \infty$. Assume further that $\mathbb{E}\,\xi^{2\alpha+1} < \infty$ for some integer $\alpha \geq 0$. Consider a functional F of finite rooted ordered trees with the property that*

$$f(T) = O(\deg(T)^\alpha),$$

where f is the toll function associated with the functional.

Furthermore, assume that there exists a sequence $(p_M)_{M \geq 1}$ of positive numbers with $p_M \to 0$ as $M \to \infty$, such that
  • *for every integer $M \geq 1$,*

$$\mathbb{E}\left| f(\hat{\mathcal{T}}^{(M)}) - \mathbb{E}\left(f(\hat{\mathcal{T}}^{(N)}) | \hat{\mathcal{T}}^{(M)} \right) \right| \leq p_M,$$

  for all $N \geq M$,
  • *there is a sequence of positive integers $(M_n)_{n \geq 1}$ such that for large enough n,*

$$\mathbb{E}\,|f(\mathcal{T}_n) - f(\mathcal{T}_n^{(M_n)})| \leq p_{M_n}.$$

If $a_n = n^{-1/2}(n^{\max\{\alpha,1\}} p_{M_n} + M_n^2)$ satisfies

$$\lim_{n\to\infty} a_n = 0, \text{ and } \sum_{n=1}^{\infty} \frac{a_n}{n} < \infty,$$

then

$$\frac{F(\mathcal{T}_n) - \mu n}{\sqrt{n}} \xrightarrow{d} N(0, \gamma^2),$$

where $\mu = \mathbb{E}\,f(\mathcal{T})$ and $0 \leq \gamma^2 < \infty$.

The proof shows that the result still holds if we replace $(F(\mathcal{T}_n) - \mu n)/\sqrt{n}$ by $(F(\mathcal{T}_n) - \mathbb{E}\,F(\mathcal{T}_n))/\sqrt{n}$. While the formulation of the theorem is fairly

technical, it strives to capture the intuitive notion of functionals that are almost local in the sense that if we look at only the first M levels of the tree we obtain a sufficiently good approximation of the toll function, with the approximation becoming better and better as $M \to \infty$. This can be compared to the notion of *local* additive functionals (see [21], again), where the toll function only depends on a fixed-size neighborhood of the root.

1.4 Symmetries and automorphisms

Counting objects up to symmetry is a classical subject in combinatorics (the Redfield–Pólya theorem from [29] and [32] is an important result in this direction). In many situations, this amounts to studying the *automorphism group* of the combinatorial object, where an object with many symmetries has an automorphism group of large cardinality. An example of the connection between enumeration and automorphisms is the case of graphs, where the number of unique labelings of a graph G equals $\frac{n!}{|\operatorname{Aut} G|}$. Studying symmetry in random combinatorial objects is one of the main themes of this thesis.

To define the automorphism group of a graph or rooted tree, we first need to define the concept of an *isomorphism* between such objects. An isomorphism between two graphs $G_1 = \{V_1, E_1\}$ and $G_2 = \{V_2, E_2\}$ is a bijection h between their vertex sets that preserves edges, i.e., if $e = \{v_1, v_2\}$ is an edge in E_1, then we must have that $\{h(v_1), h(v_2)\}$ is an edge in E_2. For isomorphisms between rooted trees T_1 and T_2, we also require that h sends the root of T_1 to the root of T_2. If there is an isomorphism between two objects we say that they are *isomorphic*. Intuitively, two trees are isomorphic if they look the same, see Figure 1.2a.

An automorphism of a graph (or rooted tree) is then an isomorphism from the object to itself. The set of automorphisms is also a group: the automorphism that sends all vertices to themselves is the identity element, a composition of automorphisms is again an automorphism, and since an automorphism is a bijection between vertex sets any element has an inverse.

The cardinality of the automorphism group of rooted trees satisfies a nice recursive formula. For any rooted tree T, let $T_1, T_2, \ldots, T_k$ be its root branches up to isomorphism, having multiplicities $m_1, m_2, \ldots, m_k$, respectively. Then we have

$$|\operatorname{Aut} T| = \prod_{i=1}^{k} m_i! |\operatorname{Aut} T_i|^{m_i}. \tag{1.8}$$

This follows from the observation that the automorphism group of a rooted tree is obtained from symmetric groups by iterated direct and wreath products (see [1], Proposition 1.15). Formulated differently, the tree is invariant under the automorphisms of each of the root branches as well as under permutation of isomorphic branches. Furthermore, if we take logarithms, we see that it

satisfies

$$\log|\operatorname{Aut} T| = \sum_{i=1}^{k} \log(m_i!) + \sum_{i=1}^{k} m_i \log|\operatorname{Aut} T_i|, \qquad (1.9)$$

so that $\log|\operatorname{Aut} T|$ is, in fact, an additive functional with toll function $f(T) = \sum_{i=1}^{k} \log(m_i!)$.

When it comes to random trees, there are several relevant questions related to iso- and automorphisms. Parameters of interest include the probability of two random trees being isomorphic as well as the distribution for the size of the automorphism group. This has previously been studied in [4] for a certain type of simply generated tree called phylogenetic trees. Phylogenetic trees can be described as unordered rooted full binary trees that are labeled at the leaves (with the size of the tree being the number of leaves). The article shows, among other things, that the probability that two phylogenetic trees are isomorphic decays exponentially and that the size of the automorphism group is asymptotically log-normal which means that the logarithm of the size satisfies a central limit theorem. That the distribution is log-normal is related to the logarithm in (1.9).

It turns out that the results for random phylogenetic trees extend naturally to some other types of rooted binary trees (in this paragraph, all trees are considered to be full binary trees unless otherwise stated). In these cases, all phylogenetic trees correspond to the same number of trees of the different type. For example, this is the case for labeled unordered binary trees of size n. Note that any full binary tree have the same number of leaves and therefore also the same number of internal vertices and also that any phylogenetic tree has a canonical ordering of its leaves based on an ordering of its labels. To find a labeled unordered tree that corresponds to a given phylogenetic tree of size (i.e., number of leaves) n, we first label the internal vertices in one of $(n-1)!$ different ways and then relabel all the vertices in one of $\binom{2n-1}{n-1}$ ways by deciding which of the labels correspond to leaves. This means that all phylogenetic trees correspond to the same number of labeled unordered trees. As a next step, any labeled *unordered* binary tree corresponds to the same number of labeled *ordered* trees since, to go from one to the other, we permute the branches at each of the internal vertices. This gives a total of 2^{n-1} ordered trees for each unordered one. Now, every *unlabeled* ordered binary tree corresponds to $n!$ *labeled* ordered trees, and, finally, we have also seen how there is a bijection between full binary trees and pruned binary trees obtained by removing the leaves. In conclusions, any given tree of the listed types corresponds to the same number of trees of any other type. Neither the size of the automorphism group nor the isomorphism class is affected by the labelings and reorderings described above so the probability of picking a tree with certain properties is unaffected by going between the different models. Thus, the results valid for phylogenetic trees also hold for any of the other types.

When studying the probability of two trees being isomorphic to each other, one naturally encounters *isomorphism classes* of trees. The isomorphism class of a given tree is the set of all trees that are isomorphic to it. Thus, the probability of two trees being isomorphic is highly related to the probability of a random tree belonging to a given isomorphism class. Isomorphism of rooted trees disregards order and labelings, which means that the isomorphism classes of Galton–Watson trees are the unlabeled unordered rooted trees or, in other words, Pólya trees. This means that many of the results in this thesis say something fundamental about how Galton–Watson trees and Pólya trees relate to each other.

1.5 Analytic combinatorics and singularity analysis

Analytic combinatorics is the mathematical field concerned with attaining combinatorial information from generating functions of combinatorial objects using analytic tools. We have seen how the families of trees we aim to study can be described in terms of generating functions, and a large part of this thesis, and the field of probabilistic combinatorics in general, is devoted to estimating the asymptotic number or weight of trees based on properties of these functions. The strength in the generating function approach is that the functions, while symbolically defined, often can be shown to be analytic in some region of the complex plane. This translates the problem of enumerating combinatorial objects to one of estimating the growth of coefficients of an analytic function, which is a classical and well studied subject (see for example the foundational work of Cauchy [5]). For an analytic function, the coefficients will grow at an exponential rate inverse to the radius of convergence (at this point the coefficient and x^n-factor cancels out so that the underlying mechanism is related to the observation that the geometric series $\sum_n x^n$ has radius of convergence 1).

Here, Cauchy's coefficient formula is a fundamental tool, saying that if $f(z)$ is analytic in a region containing 0, and if λ is a simple positively oriented loop around 0 in the region, then

$$[z^n]f(z) = \frac{1}{2\pi i} \oint_\lambda \frac{f(z)}{z^{n+1}} dz.$$

Singularity analysis is a method that lets us extract the asymptotic information about the coefficients in a standardized way, without having to compute the value of the integral directly. It does not work in all situations, but successful applications are numerous (see [13], which also serves as a general reference to the subject, for a wealth of examples). For the method to work, the generating function needs to be analytic in a so-called Δ-*region* (see Figure 1.8). We first define a Δ-region at 1 as an open set

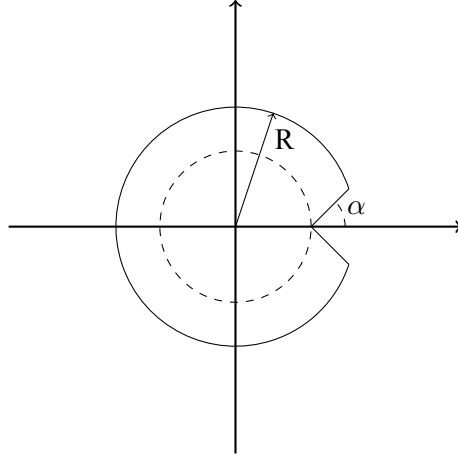$$\{z \, | \, |z| < R, z \neq 1, |\arg(z-1)| > \alpha\},$$

21

Figure 1.8. An example of a Δ-region in the complex plane.

for two given numbers $R > 1$ and $0 < \alpha < \frac{\pi}{2}$. We can define a Δ-region at ζ for any complex $\zeta \neq 0$ as the image of a Δ-region at 1 under the map $z \mapsto \zeta z$.

Since the generating function of a type of combinatorial object only has positive coefficients, Pringsheim's theorem implies that it always has a dominant singularity on the positive real axis (but it is possible that it has more than one dominant singularity), where, again, dominant means that it has smallest the absolute value of all singularities.

The central step in applying singularity analysis to a problem is to find a *singular expansion* of the corresponding generating function. The expansion is taken around the singularity and is valid inside the Δ-region where the generating function is analytic. We will restrict the definition slightly to the cases most relevant to us. For now, we assume that the functions only have one dominant singularity. We will call the set of singular functions

$$\mathcal{S} = \{(1-z)^{-\alpha} | \alpha \in \mathbb{C}\},$$

the *standard scale*. A singular expansion of a function $f(z)$ around its dominant singularity ρ is

$$f(z) = \sigma(z/\rho) + O(\tau(z/\rho)),$$

where τ is a function in $\mathcal{S}$ and σ is a finite linear combination of such functions.

It is then possible to use general theorems (Theorems VI.1 and VI.3 in [13]) to translate the expression term-wise (including the error term) to an asymp-

totic expansion for the coefficients according to

$$f(z) = \left(1 - \frac{z}{\rho}\right)^{-\alpha} \longrightarrow f_n \sim \frac{n^{\alpha-1}}{\Gamma(\alpha)}\left(1 + \sum_{k=1}^{K} \frac{e_k}{n^k}\right)$$

$$f(z) = O\left(\left(1 - \frac{z}{\rho}\right)^{-\alpha}\right) \longrightarrow f_n = O(n^{\alpha-1})$$

for some constants $\{e_k\}_k$ depending on α and some K depending on the linear combination σ. Of course, not all functions have an expansion of the form required (nor are all functions analytic in a Δ-region), but in the context of trees, this approach has turned out to be (perhaps) surprisingly successful.

Example 1.5.1 (See Section VI.7 of [13])**.** If the generating function $T(x)$ of a type of simply generated trees has ρ as its sole dominant singularity, then the singular expansion has the shape

$$T(x) = T(\rho) - c_1\left(1 - \frac{x}{\rho}\right)^{1/2} + \sum_{k=2}^{K}(-1)^k c_k\left(1 - \frac{x}{\rho}\right)^{k/2}$$
$$+ O\left(\left(1 - \frac{x}{\rho}\right)^{(K+1)/2}\right)$$

for some computable constants $c_1, c_2, \ldots, c_K$. Here, we can actually take K to be arbitrarily large and we write

$$T(x) = T(\rho) + \sum_{k=1}^{\infty}(-1)^k c_k\left(1 - \frac{x}{\rho}\right)^{k/2}$$

to symbolize this.

The asymptotic behavior (1.5) of simply generated trees with weight generating function $\Phi(x)$ can then be obtained using the method of singularity analysis

$$[x^n]T(x) \sim \frac{\rho^{-n}}{\sqrt{\pi n^3}}\left(\sqrt{\frac{\Phi(\tau)}{2\Phi''(\tau)}} + \sum_{k=2}^{\infty}\frac{d_k}{n^k}\right)$$

for computable constants $d_2, d_3, \ldots$. Recall also the definition of τ from (1.4). Here, again, we are writing an infinite sum to suggest that any truncation of it is asymptotic to $[x^n]T(x)$. The asymptotic growth rate is ρ^{-1}. Note that the condition (1.4) on τ and ρ is equivalent to

$$\tau = \rho\Phi(\tau)$$
$$1 = \rho\Phi'(\tau),$$

which, in words, means that the implicit function theorem fails at the point (ρ, τ) and indicates that there is, indeed, a singularity at ρ.

We have now defined the concepts necessary to understand our results and explained the most important tools used to obtain them. We go on to describe the contribution of this thesis.

2. Summary of articles

We have already indicated that this thesis has two main themes: symmetry of random objects and additive functionals of random trees. It consists of three articles that study these themes from various angles. A summary of the papers follows.

2.1 Article I: The probability of random trees being isomorphic

This article concerns a number of questions related to isomorphisms and isomorphism classes of Galton–Watson trees, with one of the main points being the asymptotic behavior of the probability that two such trees are isomorphic. As such, this is a natural extension of some of the results in [4], valid for phylogenetic trees (and also some other types of binary trees as shown in the previous chapter). Recall that the isomorphism classes of Galton–Watson trees are Pólya trees, so that they, too, play a central role in the paper. When it comes to the titular probabilities, we have a few different results, valid for different classes of Galton–Watson trees and using different methods.

For labeled unordered rooted trees it is proven that the probability that two trees are isomorphic is, asymptotically, exponentially small, and a full asymptotic expansion is derived.

Theorem 2.1.1. *The probability p_n that two labeled rooted trees are isomorphic has the full asymptotic expansion*

$$p_n \sim A n^{3/2} c_l^n \left(1 + \sum_{k=1}^{\infty} \frac{e_k}{n^k} \right),$$

where $A \approx 2.397678$, $c_l \approx 0.354379$ and the e_k are constants that can be calculated.

In this case, the probabilities can be extracted as coefficients from a certain generating function and this function satisfies a functional equation derived from the recursion (1.8). A deeper analysis of this generating function lets us apply singularity analysis to derive the asymptotics.

At first, one might suspect that this should hold for all Galton–Watson trees since this is fundamentally a question of how simply generated trees relate

to their isomorphism classes, i.e., Pólya trees. We observe that the weight of simply generated trees of size n grows exponentially (possibly with rate 1, as is the case for critical Galton–Watson trees) and that the isomorphism classes that they are divided into are also exponentially many. However, it might be the case that the weight concentrates on only one or a few isomorphism classes in such a way that they decay slower than any exponential function. Somewhat surprisingly, we show that this happens in the case of ordered rooted trees which we can also call plane trees due to their canonical embedding in the plane.

Theorem 2.1.2. *The probability that two plane trees are isomorphic decays subexponentially. Thus, we cannot obtain exponential bounds on the probability that two conditioned Galton–Watson trees are isomorphic in general.*

The core of the proof is finding a sufficiently large set of trees of the same size and from different isomorphism classes. This can be done by using the pigeonhole principle. We can then attach all of them to a common root to obtain a tree belonging to a large isomorphism class since we get a unique embedding in the plane for each choice of embedding for its branches together with a permutation of them.

The theorem shows that we cannot obtain exponential decay for general Galton–Watson trees. However, if we restrict ourselves to Galton–Watson trees with bounded degrees, we can still prove exponential decay of the probabilities. However, as opposed to the labeled case in Theorem 2.1.1, we have not been able to obtain a full asymptotic expansion except in special cases.

Theorem 2.1.3. *The probability g_n that two Galton–Watson trees with degrees in a finite set D are isomorphic satisfies*

$$g_n \leq B c_g^n,$$

for some constants B and $c_g < 1$.

The result relies on the fact that sufficiently large Galton–Watson trees have a giant branch (i.e., one that contains a fraction of the vertices that approaches 1 as $n \to \infty$) which sets the stage for a proof by induction.

A natural follow-up question to Theorem 2.1.1 is whether trees conditioned on being isomorphic exhibit a different structure than regular labeled trees. For example, we can compare the expected number of leaves (or vertices of any other fixed outdegree) in the two models. This is possible since a pair of isomorphic trees will necessarily have the same number of vertices of each degree, so that we can condition on the event that two labeled trees are isomorphic and study the number of vertices with degree d which will be common to both of them. In the general case, we have the following result on the degree distribution of the vertices.

26

Theorem 2.1.4. *Let $\mathbf{X}_n$ be a random vector that counts the number of vertices of (out)degree $\mathbf{d} = (d_1, d_2, \ldots, d_k)$ in either of a pair of isomorphic labeled trees of size n. Then*

$$\mathbb{E}\,\mathbf{X}_n = \boldsymbol{\mu} n + O(1),$$
$$\mathrm{Cov}\,\mathbf{X}_n = \boldsymbol{\Sigma} n + O(1),$$

for a vector $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_k)$ and a matrix $\boldsymbol{\Sigma} = (\sigma_{i,j})_{1 \leq i,j \leq k}$. Furthermore, we have joint convergence to a normal distribution

$$\frac{X_n - \mathbb{E}\,\mathbf{X}_n}{\sqrt{n}} \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma}).$$

The proof, once again, relies on functional equations and singularity analysis. As an application we can see that the expected number of leaves in isomorphic trees is asymptotic to μn with $\mu \approx 0.340252$, which is slightly lower than for the usual model of labeled unordered rooted trees where we instead have $\mu = e^{-1} \approx 0.367879$. This implies that the structure of the trees from the two models do indeed differ.

Recall that Pólya trees (possibly with some degree restrictions) correspond to the isomorphism classes of simply generated trees. This means that we can associate a weight to any Pólya tree in the form of the total weight of the associated isomorphism class. This is particularly interesting when the weights w_i are integers and the coefficients of the generating function correspond to the number of trees in the set as opposed to a weight in the more general sense. Then the weight of a given isomorphism class is equal to the number of representations of the underlying unordered, unlabeled tree as a tree from the simply generated class. The prime examples are the number of orderings and labelings of a random Pólya tree. We can show that the logarithm of the weight of a randomly chosen Pólya tree satisfies a central limit theorem under various conditions. To highlight one of the themes of the thesis, we make a short remark that both the number of vertices of given degrees, which was discussed above, and the logarithm of the weight of isomorphism classes are examples of additive functionals.

Theorem 2.1.5. *Let $\mathcal{P}_n$ be a random tree of size n from the class of unordered, unlabeled trees with degrees in the finite set D and $\mathcal{T}_D$ be a class of Galton–Watson trees with the same degree restrictions. Then the weight $W(\mathcal{P}_n)$ of $\mathcal{P}_n$ seen as an isomorphism class of $\mathcal{T}_D$ has expected value and variance*

$$\mathbb{E}[\log W(\mathcal{P}_n)] = \mu n + O(1),$$
$$\mathrm{Var}[\log W(\mathcal{P}_n)] = \sigma^2 n + O(1),$$

for some constants μ and σ and satisfies

$$\frac{\log W(\mathcal{P}_n) - \mu n}{\sqrt{n}} \xrightarrow{d} N(0, \sigma^2).$$

Just as for several of the results above, the proof of this theorem proceeds by analyzing a functional equation and applying singularity analysis. Note that this theorem is only valid for trees with bounded out-degrees. We can also derive an analogue for the set of labeled trees without any degree restrictions.

Theorem 2.1.6. *Let $\mathcal{P}_n$ be a random Pólya tree of size n, then the number of labelings $\mathrm{L}(\mathcal{P}_n)$ of $\mathcal{P}_n$ has expected value and variance*

$$\mathbb{E}[\log \mathrm{L}(\mathcal{P}_n)] = n \log n - (\mu + 1)n + \frac{\log n}{2} + O(1),$$
$$\mathrm{Var}[\log \mathrm{L}(\mathcal{P}_n)] = \sigma^2 n + O(1),$$

for numerical constants $\mu \approx 0.137342$ and $\sigma^2 \approx 0.196770$ and satisfies

$$\frac{\log \mathrm{L}(\mathcal{P}_n) - \mathbb{E}[\log \mathrm{L}(\mathcal{P}_n)]}{\sqrt{n}} \xrightarrow{d} N(0, \sigma^2).$$

Observe that the order of the mean, $n \log n$, is unusual for combinatorial limit laws where means of order n are abundant. An example of this is the case of labeled trees with bounded degrees, which has mean of order n by Theorem 2.1.5 above. Since the number of labelings of a tree T is $\frac{|T|!}{|\mathrm{Aut}\,T|}$, the proof of this last theorem is a direct consequence of a result on the size of the automorphism group of labeled trees that is discussed in the next article.

2.2 Article II: The distribution of the number of automorphisms of random trees

This paper is concerned with the size of the automorphism group of some classes of random trees, both rooted and unrooted. In fact, we show asymptotic normality of $\log |\mathrm{Aut}\,\mathcal{T}_n|$, for Galton–Watson trees (under fairly general conditions) and labeled unrooted trees, as well as for Pólya trees, both rooted and unrooted. Note that normality of $\log |\mathrm{Aut}\,\mathcal{T}_n|$ means that $|\mathrm{Aut}\,\mathcal{T}_n|$ is asymptotically log-normal.

Recall from (1.9), that the logarithm of the size of the automorphism group is an additive functional. This is central to the proofs in the article, but we use different methods for Galton–Watson trees as compared to Pólya trees. However, fundamentally they both rely on the same idea of approximating the additive functionals by simpler ones. In the first case, the underlying approximation (hidden in the proof of Theorem 1.3.1) is done by ignoring too large fringe subtrees, while for Pólya trees, the approximation is done by ignoring vertices with too many isomorphic branches.

We prove the following theorem on the automorphism group of Galton–Watson trees.

Theorem 2.2.1. *Let $\mathcal{T}_n$ be a conditioned Galton–Watson tree of size n with offspring distribution ξ, where $\mathbb{E}\,\xi = 1$, $0 < \operatorname{Var}\xi < \infty$ and $\mathbb{E}\,\xi^5 < \infty$. Then there exist constants μ and $\sigma^2 \geq 0$, depending on $\mathcal{T}$, such that*

$$\frac{\log|\operatorname{Aut}\mathcal{T}_n| - \mu n}{\sqrt{n}} \xrightarrow{d} \mathrm{N}(0, \sigma^2).$$

The proof relies on an application of Theorem 1.3.1, but verifying the conditions requires some technical work. It is the application of this theorem that is the reason behind the condition on $\mathbb{E}\,\xi^5$, and the exponent is probably not the best possible. However, the assumption is valid for combinatorially significant examples such as labeled unordered trees, ordered trees and d-ary trees. In general, it appears difficult to obtain numerical estimates of the mean and variance constants μ and σ^2, but we show how to do it for Galton–Watson trees with bounded degrees and labeled unordered trees. We also calculate concrete estimates in some cases, see Table 2.1.

Because of the simple relationship between rooted and unrooted *labeled* trees that was mentioned in the introduction, we can extend the results to unrooted trees in this case. The same argument works for labeled trees with degree restrictions.

Theorem 2.2.2. *Let $\mathcal{T}_n$ be a uniformly random unrooted labeled tree of size n. Then, $\mathbb{E}(\log|\operatorname{Aut}\mathcal{T}_n|) = \mu n + O(1)$ and $\operatorname{Var}(\log|\operatorname{Aut}\mathcal{T}_n|) = \sigma^2 n + O(1)$, with $\mu = 0.0522901\ldots$ and $\sigma^2 = 0.0394984\ldots$. Furthermore, we have*

$$\frac{\log|\operatorname{Aut}\mathcal{T}_n| - \mu n}{\sqrt{n}} \xrightarrow{d} \mathrm{N}(0, \sigma^2).$$

We can also prove asymptotic log-normality for the size of the automorphism group of unlabeled unordered trees, both rooted and unrooted.

Theorem 2.2.3. *Let $\mathcal{P}_n$ be a uniformly random Pólya tree of size n, rooted or unrooted. Then, $\mathbb{E}(\log|\operatorname{Aut}\mathcal{P}_n|) = \mu n + O(1)$ and $\operatorname{Var}(\log|\operatorname{Aut}\mathcal{P}_n|) = \sigma^2 n + O(1)$, with $\mu = 0.1373423\ldots$ and $\sigma^2 = 0.1967696\ldots$. Furthermore,*

Class of tree	μ	σ^2
Labeled trees (rooted or unrooted)	0.0522901	0.0394984
Full binary trees	0.0939359	0.0252103
Pruned binary trees	0.0145850	0.0084835
Pólya trees (rooted or unrooted)	0.1373423	0.1967696

Table 2.1. *Numerical estimates of the mean and variance constants for some types of trees.*

we have

$$\frac{\log|\operatorname{Aut}\mathcal{P}_n| - \mu n}{\sqrt{n}} \xrightarrow{d} \mathrm{N}(0, \sigma^2).$$

The tools used to prove the result for Galton–Watson trees do not apply here, and we instead use a more direct approach in terms of generating functions, where we define

$$P(x,t) = \sum_{P \in \mathcal{P}} |\operatorname{Aut} P|^t x^{|T|}.$$

However, notice that even if we, for each size $|T| = n$, only consider the star on n vertices we have

$$\sum_n (n-1)!^t x^n,$$

which is not analytic for any choice of $t > 0$. This is a big obstacle in applying methods from analytic combinatorics. Recall, for example, that singularity analysis requires that the generating function is analytic in a Δ-domain which is defined as a region around 0. We can circumvent this problem by introducing a cut-off that ignores highly symmetric vertices. This makes the analysis more involved, but some technical calculations and the application of a known result on the approximation of random variables gets the work done. Going from the rooted to the unrooted case follows immediately from a very general result in [34] on how unrooted unlabeled unordered trees can be approximated by rooted ones.

2.3 Article III: Additive functionals of subcritical graphs

In this article we study subcritical graphs using the decorated tree approach. In particular, we extend the definition of an additive functional to this setting and prove results on functionals of random graphs using methods coming from probability theory. While we primarily focus on the labeled case, we also prove some results on unlabeled graphs along the way.

The main theorem is a central limit theorem for almost local additive functionals of labeled subcritical graphs. This is an extension of the main theorem from [31], valid for Galton–Watson trees, which was discussed in the introduction.

Theorem 2.3.1. *Let $\mathcal{D}_n$ be a conditioned decorated Galton–Watson tree of size n with offspring distribution ξ, satisfying $\mathbb{E}\,\xi = 1$ and $0 < \sigma^2 := \operatorname{Var}\xi < \infty$. Assume further that $\mathbb{E}\,\xi^{2\alpha+1} < \infty$ for some integer $\alpha \geq 0$. Consider a functional F of finite decorated rooted trees with the property that*

$$f(D) = O(\deg(D)^\alpha),$$

where f is the toll function associated with the functional.

Furthermore, assume that there exists a sequence $(p_M)_{M \geq 1}$ of positive numbers with $p_M \to 0$ as $M \to \infty$, such that

- *for every integer $M \geq 1$,*

$$\mathbb{E}\left| f(\hat{\mathcal{D}}^{(M)}) - \mathbb{E}\left(f(\hat{\mathcal{D}}^{(N)}) | \hat{\mathcal{D}}^{(M)} \right) \right| \leq p_M, \qquad (2.1)$$

  for all $N \geq M$,

- *there is a sequence of positive integers $(M_n)_{n \geq 1}$ such that for large enough n,*

$$\mathbb{E}\left| f(\mathcal{D}_n) - f(\mathcal{D}_n^{(M_n)}) \right| \leq p_{M_n}.$$

If $a_n = n^{-1/2}(n^{\max\{\alpha,1\}} p_{M_n} + M_n^2)$ satisfies

$$\lim_{n \to \infty} a_n = 0, \; and \; \sum_{n=1}^{\infty} \frac{a_n}{n} < \infty,$$

then

$$\frac{F(\mathcal{D}_n) - \mu n}{\sqrt{n}} \xrightarrow{d} N(0, \beta^2),$$

where $\mu = \mathbb{E} f(\mathcal{D})$ and $0 \leq \beta^2 < \infty$.

The proof combines the idea that the underlying tree can be seen as a Galton–Watson tree with the methods of proof from [31] (see also [21] which inspired that paper). It is a theorem with many possible applications, both in the study of subcritical graphs and, since it is formulated for general decorated trees, in other models that can be described as such. We give two applications to subcritical graphs in the paper. The first one concerns the number of occurrences of a fixed 2-connected subgraph. This, for example, shows that the number of triangles in a connected labeled subcritical graph, chosen randomly, is asymptotically normal.

Theorem 2.3.2. *Fix a 2-connected graph H. Let $\mathcal{C}_n$ be a random connected graph of size n from some fixed labeled subcritical graph class and let X_n^H be the number of copies of H in $\mathcal{C}_n$. Then*

$$\mathbb{E} X_n^H = \mu_H n + o(\sqrt{n}),$$
$$\operatorname{Var} X_n^H = \sigma_H^2 n + o(n),$$

for some constants μ_H and σ_H^2. Moreover, we have

$$\frac{X_n^H - \mu_H n}{\sqrt{n}} \xrightarrow{d} N(0, \sigma_H^2).$$

The theorem is a special case of the results in [9], but the method of proof is novel since the arguments in that paper relied on an intricate study of analytic properties of generating functions. Our other example is completely new,

as far as we know, and concerns the number of automorphisms of random labeled subcritical graphs. This means that they extend the results from Article II where we obtained similar results for Galton–Watson trees.

Theorem 2.3.3. *Let $\mathcal{C}_n$ be a random connected graph of size n from some fixed labeled subcritical graph class. Then there exist constants μ and $\sigma^2 \geq 0$ such that*

$$\frac{\log |\operatorname{Aut} \mathcal{C}_n| - \mu n}{\sqrt{n}} \xrightarrow{d} \mathrm{N}(0, \sigma^2).$$

In both examples we show that the results are also valid for unrooted graphs by using structural properties of labeled graphs. In the case of unlabeled objects, we instead study decorated sesqui-type trees and derive the following result on the mean of additive functionals with finite support, meaning that the toll function f only takes non-zero values for a finite set of objects.

Theorem 2.3.4. *Let $\mathcal{U}$ be a set of unlabeled combinatorial objects associated with a decorated sesqui-type tree with offspring distribution $\mathbf{X} = [\xi, \zeta]$ where ξ and ζ have some exponential moment and $\mathbb{E}\,\xi = 1$. Let F be an additive functional of such objects that is finitely supported and assume that $\mathcal{U}_n$ is a randomly chosen object from $\mathcal{U}$ of size n. Then*

$$\mathbb{E}\,F(\mathcal{U}_n) = n\mu + o(n).$$

for some constant μ.

The primary models we have in mind are those of Pólya trees and subcritical graphs and it, for example, applies to give an estimate of the expected number of occurrences of a given tree T as a fringe subtree in a Pólya tree. Just as for labeled structures, our main contribution is a novel method to study this type of question.

3. Sammanfattning på svenska - Summary in Swedish

I denna avhandling studerar vi egenskaper hos vissa typer av slumpträd och slumpgrafer, d.v.s. träd (eller grafer) som vi drar på något slumpartat sätt. I praktiken innebär det att ett slumpträd är ett rum av träd tillsammans med ett sannolikhetsmått som anger sannolikheterna för olika utfall av ett slumpförsök där resultatet är ett träd från mängden. De typer av träd som studeras är främst Galton–Watsonträd och omärkta oordnade träd, ofta kallade Pólyaträd. Boken [7] kan agera som referens till de typer av slumpträd som återfinns i denna avhandling.

Galton–Watsonträd definieras genom en växtprocess där vi börjar med en nod som vi kallar för roten. Roten får ett antal avkommor enligt någon diskret slumpvariabel ξ vars värdemängd är en delmängd av de icke-negativa heltalen och innefattar talet 0 och åtminstone ett tal större än 1. Vi låter sedan trädet växa fram genom att varje avkomma får ett slumpmässigt antal egna avkommor enligt en oberoende kopia av ξ. Tidiga studier av dessa träd återfinns i [3] (se även [16]) och [37]. Av särskilt intresse är de *betingade Galton–Watsonträden*, som är betingade på att trädet har storlek n. De betingade Galton–Watsonträden kan, under ganska generella antagande, ses som ekvivalenta med *simpelt genererade träd* [24]. De senare är en kombinatoriskt specificerad familj av träd som definieras genom att deras genererande funktion $T(x)$ uppfyller en funktionalekvation av typen

$$T(x) = x\Phi(x), \tag{3.1}$$

där

$$\Phi(x) = \sum_{k=0}^{\infty} w_k x^k$$

är en genererande funktion förknippad med en följd av vikter $\{w_k\}_{k=0}^{\infty}$. För Galton–Waltsonträd ska vikten w_k ses som sannolikheten för en nod att få k barn. Sannolikheten för ett dra ett givet simpelt genererat träd T är proportionell mot dess vikt:

$$\mathbb{P}(\mathcal{T}_n = T) = \frac{w(T)}{\sum_{S \in \mathcal{T}_n} w(S)} = \frac{w(T)}{[x^n]T(x)}.$$

Pólyaträd kan inte beskrivas av samma växtprocess som Galton–Watsonträd (men se [27] för en koppling mellan dem) och faller därför inte under definitionen. Däremot har de många liknande egenskaper som Galton–Watsonträd.

Pólyaträd kan beskrivas genom dess genererande funktion som uppfyller funktionalekvationen

$$P(x) = x \exp\left(\sum_{k=1}^{\infty} \frac{P(x^k)}{k}\right). \tag{3.2}$$

Vi använder $\mathcal{T}$ för att ange en klass av Galton–Watsonträd (alternativt simpelt genererade träd) eller ett slumpmässigt träd taget från denna klass. Vi använder T för att ange ett specifikt träd eller en specifik realisation av slumpträdet $\mathcal{T}$. Vi använder $\mathcal{P}$ och P på motsvarande sätt för Pólyaträd.

Många kombinatoriska strukturer uppfyller någon rekursion som kan översättas till en funktionalekvation för dess genererande funktion. I grunden är funktionen symboliskt definierad, men i många sammanhang kan man visa att den är analytisk i någon region av det komplexa talplanet. Då kan vi använda analytiska metoder för att dra slutsatser om de asymptotiska egenskaperna hos funktionens koefficienter C_n som ju, i ett kombinatoriskt sammanhang, numrerar antalet objekt, t.ex. träd, av storlek n. En vanligt förekommande metod för att göra detta kallas för singularitetsanalys, vilket man kan läsa mer om i [13].

Ett av denna avhandlings teman kan anses vara så kallade *additiva funktionaler* av rotade träd. En additiv funktional $F(T)$ av ett träd T är en egenskap av trädet som kan beskrivas rekursivt genom formeln

$$F(T) = f(T) + \sum_B F(B).$$

Här är $f(T)$ en annan funktion av trädet som kallas för *kostnadsfunktionen* och summan tas över alla grenar B fästa vid roten. Observera att alla egenskaper för rotade träd kan beskrivas på denna form genom rätt val av f, men det är inte alltid som f kan beskrivas på något enkelt och användbart sätt. Ett exempel på en additiv funktional (med en enkel sluten formel för f) är antalet löv i trädet. I detta fall är $f(T) = \mathbf{1}\{T$ är ett löv$\}$ och vi kan utläsa rekursionen som: antalet löv i ett träd är summan av löven i varje delträd, såvida inte trädet endast består av roten i vilket fall roten är det enda lövet. Detta kan generaliseras till förekomst av noder av godtycklig grad. Additiva funktionaler för slumpträd har studerats flitigt. Några exempel på artiklar som tagit upp ämnet är [11], [12], [21], [31] och [36] för Galton–Watsonträd samt [36] för Pólyaträd.

Ifall additiva funktionaler är ett av avhandlingens teman så är det andra temat symmetriska egenskaper hos slumpträd och andra kombinatoriska strukturer. Egenskaper hos isomorfier och automorfigrupper för fylogenetiska träd (en viss typ av simpelt genererade träd) har tidigare studerats i [4]. Stora delar av denna avhandling syftar till att generalisera resultat från den artikeln till andra träd och typer av kombinatoriska strukturer.

3.1 Artikel I

I denna artikel studeras ett antal egenskaper förknippade med isomorfi-klasser hos Galton–Watsonträd, med främsta fokus på sannolikheten att två Galton–Watsonträd är isomorfa. En grundläggande observation är att dessa isomorfiklasser motsvarar Pólyaträd. Detta gäller eftersom en isomorfi bortser från egenskaper så som ordning och märkning, och Pólyaträden är, per definition, oordnade och omärkta träd. På samma sätt har Galton–Watsonträd av begränsad grad familjen av Pólyaträd med motsvarande begränsning som isomorfiklasser. Denna observation innebär att artikeln, på ett djupare plan, avslöjar kopplingar mellan Galton–Watson- och Pólyaträd.

Vårt första resultat är en fullständig asymptotisk utveckling för sannolikheten att två slumpmässigt valda märkta träd är isomorfa. Av extra intresse är det faktum att sannolikheten avtar exponentiellt.

Sats 3.1.1. *Sannolikheten p_n att två märkta rotade träd är isomorfa har följande asymptotiska utveckling*

$$p_n \sim A n^{3/2} c_l^n \left(1 + \sum_{k=1}^{\infty} \frac{e_k}{n^k} \right),$$

där $A \approx 2.397678$, $c_l \approx 0.354379$ och e_k är konstanter vars numeriska värde kan uppskattas.

Resultatet utgår från en funktionalekvation och använder sig av singularitetsanalys. Sannolikheten att två simpelt genererade träd är isomorfa är nära förknippad med sannolikheten att ett träd ligger i en given isomorfiklass, d.v.s. den relativa vikten av alla träd som ligger i den klassen. Då isomorfiklasserna motsvarar Pólyaträd, av vilka det finns exponentiellt många, och då vikten för en given klass av simpelt genererade träd växer eller avtar exponentiellt (möjligtvis med konstant 1 i det kritiska fallet), är det knappast förvånande att denna sannolikhet (för märkta träd) också avtar exponentiellt. Av denna anledning får nästa resultat anses vara en överraskning.

Sats 3.1.2. *Sannolikheten att två ordnade rotade träd är isomorfa avtar subexponentiellt.*

Eftersom ordnade träd är en typ av Galton–Watsonträd, agerar detta som motexempel till den annars, till synes, rimliga förmodan att exponentiellt avtagande gäller för alla Galton–Watsonträd. Beviset går ut på att hitta ett stort antal icke-isomorfa träd som alla har ungefär samma antal representationer som ordnade träd. Genom att fästa alla dessa träd vid samma rot får vi ett nytt träd som har väldigt många sådana representationer.

Även om vi inte kan påvisa exponentiellt avtagande i det generella fallet, kan vi åstadkomma det under begränsningen att mängden av tillåtna grader i träden är ändlig.

Sats 3.1.3. *Sannolikheten g_n att två Galton–Watsonträd med grader i en ändlig mängd D är isomorfa uppfyller*

$$g_n \leq B c_g^n,$$

där B och $c_g < 1$ är konstanter.

Beviset använder sig av andra metoder (induktion samt kända strukturella egenskaper hos betingade Galton–Watson träd) än de som vi använde när vi studerade märkta träd och resultatet är inte lika starkt då vi inte kan få en fullständig asymptotisk utveckling. I artikeln visar vi dock hur vi kan erhålla sådana utvecklingar för olika typer av binära träd (se även [4]).

Som en uppföljning till Sats 3.1.1, kan vi fråga ifall två träd som är betingade att vara isomorfa uppvisar en annan struktur än vanliga märkta träd. Notera här att två träd som är isomorfa nödvändigtvis har samma antal noder av varje given grad, t.ex. måste de ha samma antal löv. Som ett led i att besvara denna fråga kan vi bevisa följande sats.

Sats 3.1.4. *Låt $\mathbf{X}_n$ vara en slumpvektor som räknar antalet noder av grad $\mathbf{d} = (d_1, d_2, \ldots, d_k)$ i ett av träden i ett par av isomorfa märkta rotade träd. Då har vi*

$$\mathbb{E} \mathbf{X}_n = \boldsymbol{\mu} n + O(1),$$
$$\mathrm{Cov} \, \mathbf{X}_n = \boldsymbol{\Sigma} n + O(1),$$

för någon vektor $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_k)$ och någon matris $\boldsymbol{\Sigma} = (\sigma_{i,j})_{1 \leq i,j \leq k}$. Vi har också gemensam konvergens till en normalfördelning enligt

$$\frac{X_n - \mathbb{E} \mathbf{X}_n}{\sqrt{n}} \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma}).$$

I kontexten av singularitetsanalys är detta en naturlig utvidgning av Sats 3.1.1 och beviset använder sig av samma metoder. Med hjälp av denna sats kan vi numeriskt uppskatta det asymptotiska antalet löv i ena delen av ett par av isomorfa träd och finna att detta antal uppför sig som $\mu n + O(1)$, där $\mu \approx 0.340252$. Detta kan jämföras med antalet löv i ett märkt träd där vi istället har $\mu = e^{-1} \approx 0.367879$. Med andra ord medför betingningen en annan struktur för de märkta träden.

Vi har redan påpekat att resultaten i denna artikel säger något om hur Galton–Watsonträd och Pólyaträd förhåller sig till varandra. En naturlig fråga om denna koppling är hur många (eller hur stor vikt av) Galton–Watsonträd som

hör till ett givet Pólyaträd. Detta är särskilt intressant när Galton–Watsonträden är antingen märkta eller ordnade eftersom frågan då rör hur många märkningar eller planära inbäddningar ett Pólyaträd har. I en stokastisk kontext blir detta antal en slumpvariabel och vi kan visa att den uppfyller en central gränsvärdessats under olika antaganden.

Sats 3.1.5. *Låt $\mathcal{P}_n$ vara ett slumpmässigt omärkt oordnat rotat träd av storlek n med grader i någon ändlig mängd D och $\mathcal{T}_D$ vara en typ av Galton–Watsonträd med samma gradbegränsningar. Om vi då ser $\mathcal{P}_n$ som isomorfiklasser hos $\mathcal{T}_D$ gäller att vikten $W(\mathcal{P}_n)$ har väntevärde och varians*

$$\mathbb{E}[\log W(\mathcal{P}_n)] = \mu n + O(1),$$
$$\mathrm{Var}[\log W(\mathcal{P}_n)] = \sigma^2 n + O(1),$$

för några konstanter μ och σ. Dessutom uppfyller den

$$\frac{\log W(\mathcal{P}_n) - \mu n}{\sqrt{n}} \xrightarrow{d} N(0, \sigma^2).$$

Återigen är beviset baserat på en funktionalekvation som kan studeras med analytiska metoder. För märkta träd utan några begränsningar av grader kan vi visa följande sats.

Sats 3.1.6. *Låt $\mathcal{P}_n$ vara ett slumpmässigt omärkt oordnat rotat träd av storlek n. Då gäller att trädets antal distinkta märkningar $\mathrm{L}(\mathcal{P}_n)$ har väntevärde och varians*

$$\mathbb{E}[\log \mathrm{L}(\mathcal{P}_n)] = n \log n - (\mu + 1)n + \frac{\log n}{2} + O(1),$$
$$\mathrm{Var}[\log \mathrm{L}(\mathcal{P}_n)] = \sigma^2 n + O(1),$$

med numeriska konstanter $\mu \approx 0.137342$ och $\sigma^2 \approx 0.196770$. Vidare har vi

$$\frac{\log \mathrm{L}(\mathcal{P}_n) - \mathbb{E}[\log \mathrm{L}(\mathcal{P}_n)]}{\sqrt{n}} \xrightarrow{d} N(0, \sigma^2).$$

Notera särskilt att väntevärdet är av ordning $n \log n$, en relativ ovanlighet bland kombinatoriska gränsvärdessatser. Eftersom antalet möjligt märkningar av ett träd T är $\frac{|T|!}{|\mathrm{Aut}\,T|}$ så är resultatet en direkt konsekvens av satsen om storleken på märkta träds automorfigrupper som finns i nästa artikel.

3.2 Artikel II

Ämnet för denna artikel är symmetri hos slumpträd. Mer specifikt studeras fördelningen för storleken hos automorfigruppen $|\mathrm{Aut}\,T|$ hos både Galton–Watsonträd och Pólyaträd. Grunden för studien är följande rekursion som

gäller för alla rotade träd T

$$|\operatorname{Aut} T| = \prod_{i=1}^{k} m_i! |\operatorname{Aut} T_i|^{m_i}, \qquad (3.3)$$

där $T_1, T_2, \ldots, T_k$ är delträden fästa vid roten, upp till isomorfi, med $m_1, m_2, \ldots, m_k$ som antalet förekomster för respektive träd. Ifall vi tar logaritmen av detta uttryck får vi

$$\log|\operatorname{Aut} T| = \sum_{i=1}^{k} \log(m_i!) + \sum_{i=1}^{k} m_i \log|\operatorname{Aut} T_i|. \qquad (3.4)$$

Detta är en additiv funktional med $\sum_{i=1}^{k} \log(m_i!)$ som kostnadsfunktion $f(T)$. I både fallet för Galton–Watson- och för Pólyaträd använder vi i grunden allmänna egenskaper hos additiva funktionaler och approximationsargument för att nå vårt resultat, men i övrigt är bevisen ganska annorlunda. För Galton–Watsonträd kan vi bevisa följande sats.

Sats 3.2.1. *Låt $\mathcal{T}_n$ vara ett betingat Galton–Watsonträd av storlek n med avkomma givet av slumpvariabeln ξ, där $\mathbb{E}\,\xi = 1$, $0 < \operatorname{Var}\xi < \infty$ och $\mathbb{E}\,\xi^5 < \infty$. Då kan vi, beroende på $\mathcal{T}$, hitta konstanter μ och $\sigma^2 \geq 0$ så att*

$$\frac{\log|\operatorname{Aut} \mathcal{T}_n| - \mu n}{\sqrt{n}} \xrightarrow{d} \operatorname{N}(0, \sigma^2).$$

Beviset grundar sig på en tillämpning av en central gränsvärdessats i [31] (som bygger vidare på resultat i [21]), men för att verifiera att alla krav håller krävs en del tekniskt arbete. Satsen vi applicerar bygger på ett approximationsargument där vi uppskattar värdet på kostnadsfunktionen genom att trunkera den och bortse från värdet på delträd som är för stora. Kravet på $\mathbb{E}\,\xi^5$ är en teknisk rest från satsen vi använder oss av och exponenten är troligtvis inte den bästa möjliga. Notera att kravet är uppfyllt för kombinatoriskt intressanta exempel så som märkta och ordnade träd.

Då Galton–Watsonträd är rotade per definition är det inte alltid lämpligt att tala om en orotad version av trädet, men i vissa fall, t.ex. för märkta träd, finns det en naturlig tolkning av vad det betyder. Då varje märkt rotat träd av storlek n motsvarar exakt $n!$ märkta orotade träd är det rättframt att utvidga föregående sats till detta fall eftersom sannolikheterna inte påverkas när vi går mellan de två modellerna.

Sats 3.2.2. *Låt $\mathcal{T}_n$ vara ett märkt orotat träd av storlek n med likformig fördelning. Då gäller att*

$$\mathbb{E}(\log|\operatorname{Aut} \mathcal{T}_n|) = \mu n + O(1), \ och$$
$$\operatorname{Var}(\log|\operatorname{Aut} \mathcal{T}_n|) = \sigma^2 n + O(1)$$

där $\mu = 0.0522901\ldots$ och $\sigma^2 = 0.0394984\ldots$. Dessutom har vi att

$$\frac{\log|\operatorname{Aut}\mathcal{T}_n| - \mu n}{\sqrt{n}} \xrightarrow{d} \mathrm{N}(0, \sigma^2).$$

Vi kan även bevisa en central gränsvärdessats för antalet automorfier hos både rotade och orotade Pólyaträd.

Sats 3.2.3. *Låt $\mathcal{P}_n$ vara ett rotat eller orotat omärkt och oordnat träd av storlek n med likformig fördelning. Då gäller att*

$$\mathbb{E}(\log|\operatorname{Aut}\mathcal{P}_n|) = \mu n + O(1), \; och$$
$$\operatorname{Var}(\log|\operatorname{Aut}\mathcal{P}_n|) = \sigma^2 n + O(1)$$

där $\mu = 0.1373423\ldots$ och $\sigma^2 = 0.1967696\ldots$. Vidare har vi att

$$\frac{\log|\operatorname{Aut}\mathcal{P}_n| - \mu n}{\sqrt{n}} \xrightarrow{d} \mathrm{N}(0, \sigma^2).$$

På grund av skillnaden mellan Pólyaträd och Galton–Watsonträd kan vi inte applicera metoderna från [31] här. Istället använder vi oss, för rotade träd, av singularitetsanalys och ett approximationsargument där vi uppskattar värdet på kostnadsfunktionen genom att bortse från noder med för många isomorfa grenar. Resultatet för orotade träd är sedan en direkt konsekvens av ett resultat i [34] som förknippar egenskaper hos ett orotat Pólyaträd med egenskaper hos ett rotat Pólyaträd som det innehåller.

3.3 Artikel III

Även ifall slumpgrafer, i vid bemärkelse, kan ha ett mycket mer komplicerat beteende än slumpträd så finns det vissa typer av grafer som är tillräckligt lika träd för att vi ska kunna använda metoder som fungerar väl i det senare fallet för att studera grafer också. Ett exempel på detta är så kallade subkritiska grafer ([2], [8], [15]) som definieras genom analytiska egenskaper hos deras genererande funktioner. Dessa egenskaper påminner om de vi kan observera hos träd och det finns till och med flera sätt att beskriva subkritiska grafer i termer av en trädstruktur.

Vi definierar ett *block* som en 2-sammanhängande komponent i en graf. En grafklass $\mathcal{G}$ är *block-stabil* ifall den har egenskapen att en graf G tillhör klassen om och endast om den även innehåller samtliga block i G. Vi kräver också att den innehåller grafen K_2 som endast består av en kant. Det är möjligt att definiera block-stabila grafklasser i både det märkta och det omärkta fallet. Definitionerna är snarlika, så vi väljer att fokusera på den märkta versionen för att det är något enklare. Säg därför att $\mathcal{G}$ är en klass av märkta grafer. Vi låter

$\mathcal{C}^\bullet$ vara mängden av alla sammanhängande och roterade grafer i $\mathcal{G}$ och låt $\mathcal{B}$ vara alla block i grafklassen. Vi låter även $\mathcal{G}(z)$, $\mathcal{C}^\bullet(z)$ och $\mathcal{B}(z)$ vara motsvarande exponentiella genererande funktioner. Alla block-stabila grafklasser uppfyller då följande ekvationssystem:

$$\mathcal{G}(z) = \exp(\mathcal{C}(z)),$$
$$\mathcal{C}^\bullet(z) = z \exp(\mathcal{B}'(\mathcal{C}^\bullet(z))).$$

Vi kan tolka detta som att en block-stabil grafklass har en trädstruktur där noderna består av block och kanterna utgörs av de noder som ligger i två block och på så sätt sammanbinder dem.

En *subkritisk graf* är en speciell typ av block-stabil graf vars genererande funktioner uppfyller vissa egenskaper utöver ovanstående ekvationssystem. Låt ρ_B vara konvergensradien hos $\mathcal{B}'(z)$ och låt ρ_C vara konvergensradien hos $\mathcal{C}^\bullet(z)$. En grafklass är subkritisk ifall

$$\mathcal{C}^\bullet(\rho_C) < \rho_B.$$

Några exempel på subkritiska grafer är blockgrafer, kaktusgrafer och serie-parallella grafer. En grafklass som är block-stabil men inte subkritisk är den för planära grafer.

Ett annat sätt att beskriva subkritiska grafer som en trädstruktur är i termer av *dekorerade träd*. Ett dekorerat träd är ett rotat träd där vi associerar något objekt från en kombinatorisk klass C till varje nod. Mer specifikt så delar vi, för varje nod v, upp dess avkommor i delmängder och ger varje delmängd, tillsammans med v, strukturen av något objekt från C. Vi kan, till exempel, se blockgrafer på detta sätt genom att dela upp varje nods avkomma i block och sedan lägga till alla möjliga kanter mellan noderna i varje block (se Figur 1.6). Tillvägagångssättet för att beskriva subkritiska grafer beskrivs i [28] och [35]. Det omärkta fallet diskuteras i [33] där graferna istället kan beskrivas som dekorerade *sesqui-träd* så som de definieras i [22]. Denna typ av träd beskrivas med hjälp av en förgreningsprocess med två typer av noder som vi kallar för L och S. Den första typen, L, får ett antal avkommor av typ L enligt någon slumpvariabel ξ och ett antal avkommor av typ S enligt någon slumpvariabel ζ. Även om ξ och ζ kan bero på varandra antar vi att avkommorna från olika noder är oberoende. Noder av typ S är infertila och får ingen avkomma. Kopplingen till omärkta strukturer fås genom att se trädet bestående endast av noder av typ L som fix-punkterna för någon automorfi av objektet. I båda fallen kan vi använda trädstrukturen för att utvidga definitionen av additiva funktionaler av roterade träd till block-stabila grafer. Detta låter oss utöka generella resultat från [31] som rör additiva funktionaler av slumpträd till märkta subkritiska slumpgrafer.

Sats 3.3.1. *Låt $\mathcal{D}_n$ vara ett dekorerat Galton–Watsonträd betingat på att ha storlek n med avkommefördelning ξ. Anta att ξ uppfyller $\mathbb{E}\,\xi = 1$ och $0 <$*

$\sigma^2 := \mathrm{Var}\,\xi < \infty$ *samt att* $\mathbb{E}\,\xi^{2\alpha+1} < \infty$ *för något heltal* $\alpha \geq 0$. *Låt F vara en funktional av dekorerade träd som uppfyller att*

$$f(D) = O(\deg(D)^\alpha),$$

där f är funktionalens kostnadsfunktion.

Anta också att vi kan hitta någon sekvens $(p_M)_{M \geq 1}$ *av positiva tal som uppfyller att* $p_M \to 0$ *as* $M \to \infty$ *och att,*
- *vi, för varje heltal $M \geq 1$, har*

$$\mathbb{E}\left| f(\hat{\mathcal{D}}^{(M)}) - \mathbb{E}\left(f(\hat{\mathcal{D}}^{(N)})|\hat{\mathcal{D}}^{(M)}\right) \right| \leq p_M, \qquad (3.5)$$

så länge $N \geq M$,
- *det finns någon sekvens $(M_n)_{n \geq 1}$ av positiva heltal så att vi, för tillräckligt stora n, har*

$$\mathbb{E}\,|f(\mathcal{D}_n) - f(\mathcal{D}_n^{(M_n)})| \leq p_{M_n}.$$

Om $a_n = n^{-1/2}(n^{\max\{\alpha,1\}}p_{M_n} + M_n^2)$ uppfyller

$$\lim_{n\to\infty} a_n = 0, \ and \ \sum_{n=1}^{\infty} \frac{a_n}{n} < \infty,$$

så har vi att

$$\frac{F(\mathcal{D}_n) - \mu n}{\sqrt{n}} \xrightarrow{d} N(0, \beta^2),$$

där $\mu = \mathbb{E}\,f(\mathcal{D})$ och $0 \leq \beta^2 < \infty$.

Vi tillämpar satsen för att ge några exempel. Ett av dem rör antalet förekomster av någon given 2-sammanhängande graf H i en slumpmässigt utvald märkt subkritisk graf.

Sats 3.3.2. *Fixera någon 2-sammanhängande graf H. Låt $\mathcal{C}_n$ vara en slumpmässigt vald sammanhängande graf av storlek n från någon fixerad märkt subkritisk grafklass och låt X_n^H vara antalet förekomster av H som en delgraf av $\mathcal{C}_n$. Då gäller*

$$\mathbb{E}\,X_n^H = \mu_H n + o(\sqrt{n}),$$
$$\mathrm{Var}\,X_n^H = \sigma_H^2 n + o(n),$$

för några konstanter μ_H och σ_H^2. Dessutom har vi att

$$\frac{X_n^H - \mu_H n}{\sqrt{n}} \xrightarrow{d} N(0, \sigma_H^2).$$

Satsen är ett specialfall av resultaten från [9] och vårt bidrag är att använda en ny typ av metoder för att tackla denna typ av problem. I den nämnda artikeln användes istället invecklade resonemang baserade på analytiska egenskaper hos genererande funktioner. Nästa resultat är, oss veterligen, helt nytt och utökar resultaten om automorfier hos Galton–Watsonträd från Artikel II till subkritiska grafer.

Sats 3.3.3. *Låt $\mathcal{C}_n$ vara en slumpmässigt vald sammanhängande graf av storlek n från någon fixerad märkt subkritisk grafklass. Då kan vi hitta konstanter μ och $\sigma^2 \geq 0$ så att*

$$\frac{\log |\operatorname{Aut} \mathcal{C}_n| - \mu n}{\sqrt{n}} \xrightarrow{d} \mathrm{N}(0, \sigma^2).$$

I båda de nämnda exemplen visar vi hur vi kan utvidga resultaten till att täcka fallet med orotade grafer trots att Sats 3.3.1 rör rotade strukturer. I det omärkta har vi inte lika starka resultat men vi bevisar följande sats om väntevärdet av funktionaler som har ändligt stöd, d.v.s. vars kostnadsfunktion f endast antar nollskilda värden för ett ändligt antal träd.

Sats 3.3.4. *Låt $\mathcal{U}$ vara en mängd av omärkta kombinatoriskt objekt förknippad med något dekorerat sesqui-träd som har avkommefördelning $\mathbf{X} = [\xi, \zeta]$ där ξ och ζ har något exponentiellt moment och $\mathbb{E}\,\xi = 1$. Låt F vara en additiv funktional av sådana objekt som har ändligt stöd och låt $\mathcal{U}_n$ vara ett slumpmässigt objekt från $\mathcal{U}$ av storlek n. Då har vi att*

$$\mathbb{E}\, F(\mathcal{T}_n) = \mu n + o(n)$$

för någon konstant μ.

I första hand tänker vi oss att de kombinatoriska objekten är Pólyaträd eller subkritiska grafer och satsen kan tillämpas för att uppskatta det förväntade antalet förekomster av ett träd T på randen i ett Pólyaträd. Precis som för märkta grafer ovan är vårt främsta bidrag att tillämpa probabilistiska metoder till ett problem som tidigare främst har studerats med hjälp av analytiska metoder.

4. Acknowledgments

My time at the math department at Uppsala university has been a fulfilling one and there are many people who deserve thanks for making them as good as they have been. Above all else, this thesis owes a huge deal of credit to my supervisor Stephan Wagner – a spring of mathematical knowledge and always a pleasure to work with. This thesis would not have been nearly as good if it was not for you. I would also like to thank my co-supervisor Svante Janson and the the entire research group in combinatorics and probability for help, encouragement and a fruitful environment to work in. Additionally, I am grateful to Michael Drmota for inviting me to a reaserch visit in Vienna and inspiring the final paper of this thesis.

While not technically related to the thesis, I would like to thank my teaching assistants, Rolf Larsson, the departments great administrative staff as well as (of course) my students for making my lecturing experiences at the department such a pleasure.

When it comes to the social life at the department, my fellow PhD students have been big source of positive energy. Here, Joel Dahne deserves extra credit. The numerous walks and fika breaks that you have invited me to have meant a lot to me. Other PhDs (and friends) that I would like to highlight are Darius, Elin, Fabian, Helena, Martin, Taya and Yu. Along the same lines, I am grateful to everyone in the Food&math group for allowing me to join your lunches every now and then.

There are also people outside of academia that have had a big impact on my life during the past few years. Here, I would like to thank my parents and my brother Jacob. A and T, your friendship has always meant a lot to me, not least your support during my switch in academic careers. I want to thank Wilhelm for being a dear friend from Umeå who has also been my friend in Uppsala from the very start. Jacob, Johan, Marcus, Oskar and Rasmus, our adventures (whether virtual, imaginary, or actual) have always been a source of joy. I also want to thank the DnD group for taking my acquaintances at work to friends outside of it. People that have not already been mentioned are Anna, Aron, Filipe and Gabri.

Last, but certainly not least, I would like to thank Maja – for all of the reasons.

Bibliography

[1] László Babai. Automorphism groups, isomorphism, reconstruction. In *Handbook of combinatorics, Vol. 2*, pages 1447–1540. Elsevier, Amsterdam, 1995.

[2] Nicla Bernasconi, Konstantinos Panagiotou, and Angelika Steger. The degree sequence of random graphs from subcritical classes. *Combin. Probab. Comput.*, 18(5):647–681, 2009.

[3] I. J. Bienaymé. De la loi de multiplication et de la duree des familles. *Soc. Philomath. Paris Extraits*, Ser. 5:37–9, 1845.

[4] Miklós Bóna and Philippe Flajolet. Isomorphism and symmetries in random phylogenetic trees. *J. Appl. Probab.*, 46(4):1005–1019, 2009.

[5] Augustin Louis Cauchy. *Œuvres complètes. Series 2. Volume 2*. Cambridge Library Collection. Cambridge University Press, Cambridge, 2009. Reprint of the 1958 original.

[6] Luc Devroye. Limit laws for sums of functions of subtrees of random binary search trees. *SIAM J. Comput.*, 32(1):152–171, 2002/03.

[7] Michael Drmota. *Random trees*. SpringerWienNewYork, Vienna, 2009.

[8] Michael Drmota, Éric Fusy, Mihyun Kang, Veronika Kraus, and Juanjo Rué. Asymptotic study of subcritical graph classes. *SIAM J. Discrete Math.*, 25(4):1615–1651, 2011.

[9] Michael Drmota, Lander Ramos, and Juanjo Rué. Subgraph statistics in subcritical graph classes. *Random Structures Algorithms*, 51(4):631–673, 2017.

[10] P. Erdős and A. Rényi. On random graphs. I. *Publ. Math. Debrecen*, 6:290–297, 1959.

[11] James Allen Fill, Philippe Flajolet, and Nevin Kapur. Singularity analysis, Hadamard products, and tree recurrences. *J. Comput. Appl. Math.*, 174(2):271–313, 2005.

[12] James Allen Fill and Nevin Kapur. Limiting distributions for additive functionals on Catalan trees. *Theoret. Comput. Sci.*, 326(1-3):69–102, 2004.

[13] Philippe Flajolet and Robert Sedgewick. *Analytic combinatorics*. Cambridge University Press, Cambridge, 2009.

[14] E. N. Gilbert. Random graphs. *Ann. Math. Statist.*, 30:1141–1144, 1959.

[15] Omer Giménez, Marc Noy, and Juanjo Rué. Graph classes with given 3-connected components: asymptotic enumeration and random graphs. *Random Structures Algorithms*, 42(4):438–479, 2013.

[16] C. C. Heyde and E. Seneta. Studies in the history of probability and statistics. xxxi. the simple branching process, a turning point test and a fundamental inequality: A historical note on i. j. bienayme. *Biometrika*, 59(3):680–683, 1972.

[17] Cecilia Holmgren and Svante Janson. Limit laws for functions of fringe trees for binary search trees and random recursive trees. *Electron. J. Probab.*, 20:no. 4, 51 pp., 2015.

[18] Cecilia Holmgren, Svante Janson, and Matas Šileikis. Multivariate normal limit laws for the numbers of fringe subtrees in m-ary search trees and preferential attachment trees. *Electron. J. Combin.*, 24(2):Paper No. 2.51, 49 pp., 2017.

[19] Hsien-Kuei Hwang and Ralph Neininger. Phase change of limit laws in the quicksort recurrence under varying toll functions. *SIAM J. Comput.*, 31(6):1687–1722, 2002.

[20] Svante Janson. Simply generated trees, conditioned Galton-Watson trees, random allocations and condensation. *Probab. Surv.*, 9:103–252, 2012.

[21] Svante Janson. Asymptotic normality of fringe subtrees and additive functionals in conditioned Galton-Watson trees. *Random Struct. Algorithms*, 48(1):57–101, 2016.

[22] Svante Janson, Oliver Riordan, and Lutz Warnke. Sesqui-type branching processes. *Stochastic Process. Appl.*, 128(11):3628–3655, 2018.

[23] Harry Kesten. Subdiffusive behavior of random walk on a random cluster. *Ann. Inst. H. Poincaré Probab. Statist.*, 22(4):425–487, 1986.

[24] A. Meir and J. W. Moon. On the altitude of nodes in random trees. *Canadian J. Math.*, 30(5):997–1015, 1978.

[25] Richard Otter. The number of trees. *Ann. of Math. (2)*, 49:583–599, 1948.

[26] Konstantinos Panagiotou and Angelika Steger. Maximal biconnected subgraphs of random planar graphs. *ACM Trans. Algorithms*, 6(2):Art. 31, 21, 2010.

[27] Konstantinos Panagiotou and Benedikt Stufler. Scaling limits of random Pólya trees. *Probab. Theory Related Fields*, 170(3-4):801–820, 2018.

[28] Konstantinos Panagiotou, Benedikt Stufler, and Kerstin Weller. Scaling limits of random graphs from subcritical classes. *Ann. Probab.*, 44(5):3291–3334, 2016.

[29] G. Pólya. Kombinatorische Anzahlbestimmungen für Gruppen, Graphen und chemische Verbindungen. *Acta Math.*, 68(1):145–254, 1937.

[30] Dimbinaina Ralaivaosaona and Stephan Wagner. A central limit theorem for additive functionals of increasing trees. *Combin. Probab. Comput.*, 28(4):618–637, 2019.

[31] Dimbinaina Ralaivaosaona, Matas Šileikis, and Stephan Wagner. A central limit theorem for almost local additive tree functionals. *Algorithmica*, 82(3):642–679, 2020.

[32] J. Howard Redfield. The Theory of Group-Reduced Distributions. *Amer. J. Math.*, 49(3):433–455, 1927.

[33] Benedikt Stufler. Random enriched trees with applications to random graphs. *Electron. J. Combin.*, 25(3):Paper No. 3.11, 81, 2018.

[34] Benedikt Stufler. The continuum random tree is the scaling limit of unlabeled unrooted trees. *Random Structures Algorithms*, 55(2):496–528, 2019.

[35] Benedikt Stufler. Limits of random tree-like discrete structures. *Probab. Surv.*, 17:318–477, 2020.

[36] Stephan Wagner. Central limit theorems for additive tree parameters with small toll functions. *Combin. Probab. Comput.*, 24(1):329–353, 2015.

[37] H. W. Watson and Francis Galton. On the probability of the extinction of families. *The Journal of the Anthropological Institute of Great Britain and Ireland*, 4:138–144, 1875.