TECHNISCHE UNIVERSITÄT WIEN
Vienna University of Technology

DISSERTATION

Asymptotic study of families of unlabelled trees and other unlabelled graph structures

ausgeführt zum Zwecke der Erlangung des akademischen Grades
einer Doktorin der technischen Wissenschaften unter der Leitung von

Ao.Univ.Prof. Dipl.-Ing. Dr.techn. Bernhard Gittenberger

Institut für Diskrete Mathematik und Geometrie
E104

eingereicht an der Technischen Universität Wien
Fakultät für Mathematik und Geoinformation

von

Dipl.-Ing. Veronika Kraus
0125919
Lange Gasse 64/3
1080 Wien

Kurzfassung

Die vorliegende Dissertation beschäftigt sich mit der asymptotischen Analyse von zufälligen Graphenstrukturen, besonders Zufallsbäumen. Wir betrachten dazu die Menge von Objekten von einer festen Größe n (wobei die Größe meist die Anzahl der Knoten eines Graphen bezeichnen wird), und wählen daraus ein bezüglich der Gleichverteilung zufälliges Objekt aus. Es werden Eigenschaften eines solchen zufälligen Vertreters untersucht, wobei die Größe n gegen unendlich strebt.

Sämtliche Ergebnisse werden mit Hilfe von erzeugenden Funktionen und der Analyse ihrer Singularitätenstruktur gewonnen. Die Methodik und eine "Werkzeugbox" aus hilfreichen Theoremen werden im ersten Kapitel vorgestellt. Diese Grundlagen entstammen verschiedener Arbeiten auf dem Gebiet der analytischen Kombinatorik und sind in dem Werk "Analytic Combinatorics" [24] von Flajolet und Sedgewick gesammelt zu finden.

Im zweiten Kapitel beginnen wir unsere Studie mit unmarkierten Wurzelbäumen, auch bekannt als Pólya-Bäume, da sie von George Pólya in seiner Arbeit [60] erstmals eingehend untersucht wurden. Wir untersuchen den Rand von Pólya Bäumen, insbesondere berechnen wir die asymptotische Größe jener Unterbäume eines Pólya-Baumes, deren Wurzel der Vater eines Blattes, also eines Knotens mit keinen Kindern ist.
Weiters wird in diesem Kapitel das sogenannte Gradprofil untersucht. Das Gradprofil eines Baumes ist jene Folge, welche die Anzahl von Knoten eines festen Grades d auf jedem Niveau $k, k \geq 0$, eines Baumes beschreibt, wobei das Niveau k die Menge der Knoten mit Distanz k zur Wurzel ist. Es wird schwache Konvergenz des Gradprofilprozesses nach der lokalen Zeit einer Brown'schen Bewegung gezeigt. Weiters zeigen wir, dass der Korrelationskoeffizient zweier verschiedenen Grade d_1 und d_2 auf einem gemeinsamen Niveau k gegen 1 strebt.

Im dritten Kapitel wird eine Baumklasse vorgestellt, die eine Brücke zum Gebiet der Logik schlägt: Boole'sche Bäume. Boole'sche Bäume repräsentieren einen logischen Ausdruck, bestehend aus Und- und Oder- ($\wedge$ und $\vee$) Verknüpfungen und Literalen (Variablen oder ihren Negationen). Wir untersuchen verschiedene Klassen solcher Boole'schen Bäume, auch die Klasse der assoziativen und kommutativen Boole'schen Bäume, die den Pólya-Bäumen in ihrer Struktur stark ähneln, und die Wahrscheinlichkeitsverteilung auf der Menge der Boole'schen Funktionen, die von ihnen erzeugt wird. Wir zeigen, dass die induzierte Wahrscheinlichkeitsverteilung in allen vorgestellten Modellen dieselbe asymptotische Abhängigkeit von der Komplexität der Funktion zeigt, die exakten Grenzwahrscheinlichkeiten aber dennoch

von Modell zu Modell verschieden sind.

Das vierte und letzte Kapitel schlußendlich beschäftigt sich mit komplexeren unmarkierten Graphen, die zur Familie der subkritischen planaren Graphen zusammengefasst werden können. Wir untersuchen die Gradverteilung in verschiedenen unmarkierten subkritischen Klassen und zeigen einen zentralen Grenzwertsatz, wie er von Drmota, Gimenez und Noy schon für die analogen markierten Klassen gezeigt wurde [18].

Die Ergebnisse dieser Dissertation sind mehreren Forschungsarbeiten entnommen, die bereits publiziert oder zur Veröffentlichung eingereicht wurden. Die Ergebnisse zum Profil von Pólya Bäumen wurde einer gemeinsamen Arbeit mit Bernhard Gittenberger [37] entnommen, Kapitel 3 entstammt einer Zusammenarbeit mit Antoine Genitrini, Bernhard Gittenberger sowie Cécile Mailler [28]. Kapitel 4 schlußendlich entstand teilweise in Zusammenarbeit mit Michael Drmota, Eric Fusy, Miyhun Kang und Juan Jose Rue [16] und wurde durch den Konferenzartikel [49] vervollständigt.
Sämtliche Arbeiten wurden von dem Projekt S9604 der Österreichischen Wissenschaftsfonds FWF finanziell unterstützt.

Abstract

This thesis deals with the asymptotic analysis of diverse random graph structures, especially random trees. For this purpose, we consider the set of objects of fixed size n (where the size is mostly describing the number of vertices of a graph), and choose an object from it uniformly at random. We discuss properties of such a random representative, as the size n tends to infinity.
All results are obtained with the help of generating functions and the analysis of their singular behaviour. The methods and a "tool box" of helpful theorems are presented in the first chapter. These basics originate in different papers on the field of analytic combinatorics and are collected in the book "Analytic combinatorics" [24] by Flajolet and Sedgewick.

In the second chapter we start our study with unlabelled rooted trees, widely known as Pólya trees, as they were for the first time thoroughly studied by George Pólya in his paper [60]. We study the fringe of Pólya trees, in particular we compute the asymptotic size of those subtrees of a Pólya tree whose root is the father of a leaf, that is a vertex with no children.
Later in this chapter we study the degree profile. The degree profile of a tree is the sequence which describes the number of vertices of a fixed degree d on each level k, $k \geq 0$, of a tree, where the level k is the set of all vertices at distance k from the root. Weak convergence of the degree profile process towards the local time of a Brownian motion is proved. Further we show that the correlation of two different degrees d_1 and d_2 on the same level k tends to 1.

In the third chapter we introduce a class of trees which connects our topic to the field of logic, namely Boolean trees. Boolean trees represent a logic expression, consisting of And- and Or- ($\wedge$ and $\vee$) connectors as well as literals (variables or their negations). We study different classes of boolean trees, also the class of associative and commutative Boolean trees, whose structure is much alike the one of Pólya trees, and the probability distribution on the set of Boolean functions which they induce. We show that the induced probability distribution has the same dependence on the complexity of the function in all models introduced, but the exact probabilities are still different.

The fourth and last chapter deals with more complex unlabelled graphs, which can be pooled to the family of subcritical planar graphs. We study the degree distribution in various unlabelled subcritical classes and show a central limit theorem, as it was shown for the corresponding labelled classes by Drmota, Gimenez and Noy [18].

The results of this thesis are taken from several papers which have been published or submitted for publication. The results on the profile of Pólya trees are joint work with Bernhard Gittenberger [37], Chapter 3 origins in a collaboration with Antoine Genitrini, Bernhard Gittenberger as well as Cécile Mailler [28]. Chapter 4, eventually, emerged partially from a joint work with Michael Drmota, Eric Fusy, Miyhun Kang and Juanjo Rue [16] and was completed by the conference paper [49]. The research has been supported by project S9604 of the Austrian Science Foundation FWF.

Personal notes

It goes without saying that, when finishing a Ph.D., there are a lot of people to thank for their support.

When I approached the end of my undergraduate studies three years ago, I had a mind full of ideas what to do after, but I admit applying for a Ph.D. position was not among them. I therefore want to thank my supervisor, Bernhard Gittenberger, first and foremost for offering the position to me, for planting the idea in my head and for awakening the interest and passion for the field of research this thesis is set in. I am grateful for the support and the opportunities to improve and present my work in international conferences I got, I gained a lot from them.

Thanks to Daniele Gardy for agreeing to review this thesis and keeping the tight schedule, for the time she spent on the very careful reading of the manuscript and the helpful comments on improving the presentation.

I thank all the great people I was able to work with for the motivating experiences I could have. I am especially indebted to Michael Drmota, who was always available for questions and discussions and who offered a lot of his time and interest in the progress of my work. The friendly and relaxed working atmosphere at the institute is due to my colleagues who where always open for coffee, discussions, puzzles and more.

I owe a lot to my family, on top of all my parents, for their support and care through many years, my friends for fun and a shoulder whenever needed, my ultimate frisbee team for the uncountably many weekends we spent together on tournaments having a hell lot of fun and exercise. Last but not least I thank a wonderful person for crossing my way somewhere in Slovenija unexpectedly, and for all the great moments we shared since then.

I feel like having a déjà-vu. Again there are a million ideas in my head what I'd like to do in the near future. But who knows what I will end up with?

Contents

CHAPTER 1

Preliminaries

Trees and other graph structures are a basic element in discrete mathematics and theoretical computer science. Trees are used for data storage and search algorithms. Large graphs are used to model real world networks. In this first chapter we will introduce the basic notations and terms and will present a summary of the methods which will be used to prove the results given in the forthcoming chapters.

1.1 Families of graphs

We start with some basic notions on graphs, which should be known to most readers.

- A *graph* G is a set of vertices $V(G)$ together with a set of edges $E(G) \subset V(G) \times V(G)$, where the pairs can be ordered (*directed* graph) or unordered (*undirected* graph).
- The *degree* of a vertex $\nu \in V(G)$ is the number of edges in $E(G)$ containing ν.
- A graph is *connected* if for any pair of nodes there exists a path (a sequence of edges) connecting the two vertices.
- A *family or class of graphs* is a set of graphs sharing certain properties.
- A *tree* is a graph which is connected and does not contain a cycle (that is, if there is a path connecting a pair of vertices, it is unique). We call vertices of degree 1 leaves or external vertices of the tree, while all other vertices are called internal nodes.
- A graph is called *simple*, if it does not contain loops (edges of the form $e = (v, v)$) or double edges.

Remark. Let G be a graph, and $|V| < \infty$. The following 3 statements defining trees are equivalent.

- G is connected and contains exactly $|V| - 1$ edges.
- G does not contain cycles and has exactly $|V| - 1$ edges.

- every pair of nodes $(\nu_1, \nu_2) \in V \times V$ is connected by a unique path (this is exactly the definition of a tree given above).

Let us denote by $\mathcal{A}$ a set of structures sharing some properties, e.g. $\mathcal{T}$ the family of all trees, or $\mathcal{K}$ the set of all quadrangulations, that is graphs where every circle contains exactly 4 vertices.

We further define a function $\omega : \mathcal{A} \to \mathbb{N}_0$, such that the sets $W_n := \{A \in \mathcal{A} | \omega(A) = n\}$ are finite. We call $\omega(A)$ the size of A and denote by A_n the number of objects of size n, $A_n = |W_n|$. In most of the following, the size function will count the number of vertices of an object, but it can also be the number of leaves of a tree, the number of edges or something similar.

1.1.1 Rooting

Consider an object A from a class $\mathcal{A}$. If we mark one vertex ν of A and call it the root of A, that is ν is distinguished and does not belong to the set of vertices of A anymore, we call the new object a derived object. If we point at the vertex ν but still consider it as a vertex of A, we call the new object a rooted object. Every member $A \in \mathcal{A}$ can be derived or rooted at any of its vertices, we call the new class of derived objects of $\mathcal{A}$ the derived class $\mathcal{A}'$ and the class of rooted objects from $\mathcal{A}$ the rooted class $\mathcal{A}^\bullet$. By the same arguments, we can also distinguish an edge and obtain an edge-rooted class $\mathcal{A}^{\circ-\circ}$. Why we use the term *derived* will be clear in the section on generating functions. There is the following correspondence between rooted and unrooted trees:

Theorem 1.1 (dissymmetry theorem on trees)**.** *Let $\mathcal{T}$ be a family of unrooted trees, and let $\mathcal{T}^\bullet$ be the corresponding family of trees rooted at a vertex, $\mathcal{T}^{\circ-\circ}$ the family rooted at an edge and $\mathcal{T}^{\circ\to\circ}$ the family rooted at an oriented edge. Then there exists a size-preserving bijection such that:*

$$\mathcal{T} \cup \mathcal{T}^{\circ\to\circ} = \mathcal{T}^\bullet \cup \mathcal{T}^{\circ-\circ}. \tag{1.1}$$

Proof. For a proof of the theorem, see [2, Chapter 4]. □

1.1.2 The labelled vs. the unlabelled setting

In graph theory, graphs can be either labelled or unlabelled, we denote the according labelled or unlabelled classes by $\mathcal{A}^{(\ell)}$ and $\mathcal{A}^{(u)}$, respectively. In the labelled setting, the vertices of an object $A \in \mathcal{A}^{(\ell)}$ of size n are labelled with distinct numbers $\{1, \dots, n\}$. In the unlabelled setting, vertices of a graph $A \in \mathcal{A}^{(u)}$ are not distinguishable. An unlabelled graph is obviously obtained from a labelled one by removing the labels, but this mapping is not bijective. Two labelled graphs A, B will give the same unlabelled graph $\tilde{A}$ if they are isomorphic, that is, if there exists a permutation σ in the group of permutations $\mathfrak{S}_n$ on the set of labels $\{1, \dots, n\}$, such that applying the permutation σ to the labels of the vertices of A gives B. That is, the unlabelled class $\mathcal{A}^{(u)}$ is the labelled class $\mathcal{A}^{(\ell)}$, considered up to isomorphism.

In the following, we will use the notations $\mathcal{A}^{(\ell)}$ and $\mathcal{A}^{(u)}$ only when required for understanding, while we will stick to the simple notation $\mathcal{A}$ if it is clear from the context which setting we are currently dealing with.

1.1.3 Planarity of graphs

Apart from unlabelling, there is another reason for isomorphisms to appear. An *embedding* into the plane of a graph is, informally speaking, a drawing of the graph where edges only intersect at vertices. An embedding of a graph G is often called a *map* of G. Obviously, this might not exist for all graphs, but it does exist for trees and all other graphs treated in this thesis. Graphs which have an embedding in the plane are called planar graphs. Different maps are isomorphic if they are an embedding of the same graph.

Rooted trees are further called *plane* (sometimes also planar, not to be confounded with the above definition) if we consider every embedding as a different tree, while they are called *non-plane* if we consider the trees before embedding. In Figure 1.1 we see two different plane trees, but the same non-plane tree. The same determination is made for graphs, where we talk of *maps* and *graphs*.

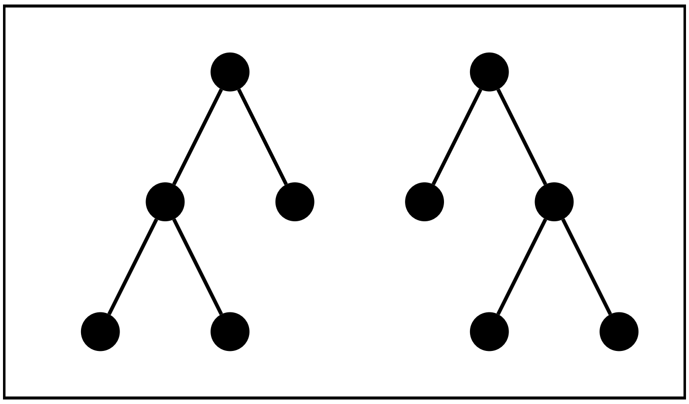

Figure 1.1: 2 different plane trees but the same non-plane tree

1.1.4 Random graphs

The underlying idea of this thesis is the setting of random graphs. Consider a class $\mathcal{A}$ together with a suitable size function ω. Throughout this thesis, we will study a random object of size n instead of a special structure $A \in \mathcal{A}$. That is, on all sets $W_n = \{A \in \mathcal{A} | \omega(A) = n\}$ we define the uniform distribution, i.e. we draw an object from W_n randomly with the constraint that every graph $A \in W_n$ appears with equal probability, which is $\frac{1}{A_n}$. We will study the expected properties of such a randomly drawn graph, as the size n tends to infinity.

1.2 Combinatorial enumeration

One of the basic interests in graph theory is enumeration, that is, obtaining information on the quantities A_n of objects of size n in a family $\mathcal{A}$. From this starting point, graphs can be studied in more detail. Counting techniques to systemize the problem were developed. We will present the symbolic method, which relies on decomposition of graphs into smaller objects and results in recursive relations for graphs of a given size, and generating functions, which are a very powerful tool in explicit as well as asymptotic counting.

1.2.1 The symbolic method

First, we will present the symbolic method, which is described in detail in [24] or in [2]. The symbolic method applies for *decomposable* classes of graphs, the idea is to systematically decompose a family recursively, and to write this decomposition into a general grammar of basic combinatorial structures and operations. This grammar translates to counting series by a given dictionary, from where counting coefficients can be extracted either explicitly or asymptotically.

The grammar includes the following basic classes:

- The neutral class $\mathcal{E}$ contains a single object of size 0.
- The atomic class $\mathcal{X}$ is made of a single object of size 1.

Further it contains the following classes of objects, which are of arbitrary size:

- The sequence class $\mathrm{Seq}(\mathcal{X})$ is an ordered sequence of atoms.
- The set class $\mathrm{Set}(\mathcal{X})$ is an unordered (multi)set of atoms.
- The cyclic class $\mathrm{Cyc}(\mathcal{X})$ is an oriented cycle of atoms.

At last, the grammar consists of the following basic operations

- The sum $\mathcal{A} + \mathcal{B}$ is the disjoint union of the two classes $\mathcal{A}$ and $\mathcal{B}$.
- The product $\mathcal{A} \times \mathcal{B}$ refers to taking a pair $(A, B) \in \mathcal{A} \times \mathcal{B}$ and join their set of atoms, if necessary relabel them with labels from $\{1, \dots, \omega(A) + \omega(B)\}$.
- An object of the composition $\mathcal{B} \circ \mathcal{A}$ is obtained by taking an object $B \in \mathcal{B}$ built of n atomic objects, e.g, vertices. Then pick a n-set of elements from $\mathcal{A}$ and substitute one of them in each atom of B, relabel if necessary.

We will see an example for applying the symbolic method in the following section.

1.2.2 Generating functions

Throughout this work, we will use generating functions to obtain information on the structures we analyze. Generating functions are formal power series, which, interpreted as analytic functions, provide a lot of useful information. Let $\mathcal{A}$ be some set of structures together with a size function ω.

Definition 1.2. *The ordinary generating function of a set $\mathcal{A}$, denoted by $A(z)$ is given by*

$$A(z) = \sum_{A \in \mathcal{A}} z^{\omega(A)} = \sum_{n \geq 0} A_n z^n,$$

while the exponential generating function is given by

$$\tilde{A}(z) = \sum_{n \geq 0} \frac{A_n}{n!} z^n,$$

where we denote by A_n the cardinality of the set $\{A \in \mathcal{A} | \omega(A) = n\}$. The notations $[z^n]A(z)$ and $[z^n]\tilde{A}(z)$ are used to denote the coefficients of z^n, A_n and $\frac{A_n}{n!}$, respectively.

Labelled Counting

Counting structures of a given set of labelled objects, we use exponential generating functions

$$A(z) = \sum_{n \geq 0} \frac{A_n}{n!} z^n,$$

due to the $n!$ possibilities to label the vertices. We look for symbolic decompositions, which then translate into exponential generating functions by the dictionary given in Table 1.1 on page 10. The translations from symbolic language to exponential generating functions are quite obvious, for this reason we do not go into detail here.

Consider a class $\mathcal{A}$ together with its derived class $\mathcal{A}'$ and its rooted class $\mathcal{A}^\bullet$. Note that in a tree of size n, there are n possibilities to root a vertex, hence $A_n^\bullet = nA_n$ and $A'_{n-1} = \frac{nA_n(n-1)!}{n!} = A_n$ by the new labelling on the numbers $\{1, \dots, n-1\}$ in the derived case. Therefore

$$\frac{\partial}{\partial z} A(z) = \sum_{n \geq 0} \frac{A_n}{(n-1)!} z^{n-1} = A'(z)$$
$$zA'(z) = \sum_{n \geq 0} \frac{nA_n}{n!} z^n = A^\bullet(z)$$

In the following we present a counting example for strict labelled plane binary trees, that is trees whose vertices have out-degree either 0 or 2. Those trees are also known as Catalan trees.

Example. Let $\mathcal{T}$ be the family of labelled plane rooted binary trees, with the size function counting the leaves, and $T(z)$ be its exponential generating function. If the tree consists of more than just one leaf, the left and right subtree of the root are also binary trees of smaller size. Hence a binary tree is either just a single leaf or it can be decomposed into smaller binary trees at the root (cf Figure 1.2). This gives a symbolic equation

$$\mathcal{T} = \mathcal{X} + \mathcal{T} \times \mathcal{T} \tag{1.2}$$

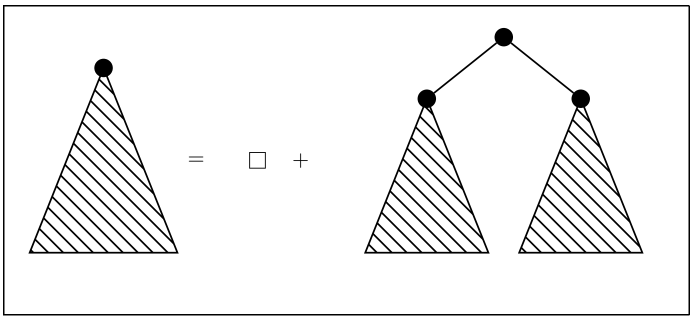

Figure 1.2: The decomposition of a binary tree at its root - $\square$ denotes a leaf while $\bullet$ represents an internal node.

In terms of exponential generating functions $T(z)$, this translates to

$$T(z) = z + T(z)^2.$$

This quadratic equation solves explicitly to

$$T(z) = \frac{1}{2}(1 - \sqrt{1-4z}),$$

where we ignored the second solution as $T(0)$ will give the number of trees with no leaves, which must be 0. From there, we can extract coefficients $[z^n]$. With the help of a binomial series, we get

$$\begin{aligned}
[z^n]T(z) &= [z^n]\frac{1}{2}(1 - \sqrt{1-4z}) \\
&= -\frac{1}{2}[z^n](1-4z)^{\frac{1}{2}} \\
&= -\frac{1}{2}\binom{\frac{1}{2}}{n}(-4)^n = \frac{1}{n}\binom{2(n-1)}{n-1},
\end{aligned}$$

and obtain explicitly the number of labelled binary trees of size n, T_n by $T_n = n![z^n]T(z)$. Note that $[z^n]T(z)$ is the $(n-1)$-st Catalan number, and that a binary tree with n leaves has exactly $n-1$ internal nodes due to the restricted shape.

Unlabelled counting - Pólya theory

To count unlabelled objects, we need to be aware of the symmetries appearing through the unlabelling, that is, we need to observe and quantify isomorphisms. Thus, unlabelled counting is generally more involved than labelled counting. The basic method used in the unlabelled setting was given by George Pólya [60] in 1937. We give here an overview of Pólyas theory of counting, which relies mainly on cycle index sums. Cycle index sums are a refined version of generating functions, defined on an infinite set of variables $\mathbf{s_1} := s_1, s_2, s_3, \ldots$.

Definition 1.3. *Let D be a finite set of size $|D| = n$ and $\mathfrak{S} \subset \mathfrak{S}_D$ a subgroup of the group of all permutations on D, $\mathfrak{S}_D$.*

- *The* cycle type *$ZT(\sigma)$ of a permutation $\sigma \in \mathfrak{S}$ is given by*

$$ZT(\sigma)(s_1, \ldots, s_n) = s_1^{\lambda_1(\sigma)} s_2^{\lambda_2(\sigma)} \cdots s_n^{\lambda_n(\sigma)},$$

  where $\lambda_i(\sigma), i \geq 1$, denotes the number of cycles of length i in σ.

- *The* cycle index *$P(\mathfrak{S})$ of the group $\mathfrak{S} \subset \mathfrak{S}_D$ is given by*

$$P(\sigma)(s_1, \ldots, s_n) = \frac{1}{|\mathfrak{S}|}\sum_{\sigma \in \mathfrak{S}} ZT(\sigma).$$

Note that for every cycle type $ZT(\sigma)$, $\sum_{i=1}^{n} i\lambda_i = n$.

Example. Let $|D| = n$.

- For the trivial group $\mathfrak{S} = \{id\}$,

$$P(\mathfrak{S}) = s_1^n.$$

- For the full symmetric group $\mathfrak{S} = \mathfrak{S}_D$,

$$P(\mathfrak{S}) = \sum_{\sigma\in\mathfrak{S}} ZT(\sigma) = \frac{1}{n!} \sum_{\ell_1+2\ell_2+\ldots n\ell_n=n} \frac{n!}{\ell_1!\ell_2!\cdots\ell_n!1^{\ell_1}2^{\ell_2}\cdots n^{\ell_n}} s_1^{\ell_1}s_2^{\ell_2}\cdots s_n^{\ell_n}.$$

With the help of cycle indices, we count unlabelled objects from a family $\mathcal{A}$. For every structure $A \in \mathcal{A}$ we define the permutation group $\mathfrak{S}_A$ as the subgroup of permutations on the set of atoms of A which do not change the object, that is the set of permutations which represent symmetries. We call $\mathfrak{S}_A$ the set of allowed permutations on A.

Definition 1.4. *Let $\mathcal{A}$ be a class of structures. The cycle index sum $Z_{\mathcal{A}}(\mathbf{s_1})$ of the class $\mathcal{A}$ is defined by*

$$Z_{\mathcal{A}}(\mathbf{s_1}) := \sum_{A\in\mathcal{A}} P(\mathfrak{S}_A)(s_1,\ldots,s_{\omega(A)}),$$

where $\mathfrak{S}_A$ is the set of allowed permutations of A and $\mathbf{s_1}$ denotes the infinite set of variables $\mathbf{s_1} = s_1, s_2, s_3, \ldots$.

Remark. We use the notation $\mathbf{s_1}$ here as we will also define series of variables $\mathbf{s}_\ell$ in the following, given by $\mathbf{s}_\ell = s_\ell, s_{2\ell}, s_{3\ell}, \ldots$.

With this counting series, we keep track of all symmetries the objects of $\mathcal{A}$ have. If we substitute $s_i = z^i, i \geq 0$ in the cycle index sum of a class, we obtain its ordinary generating function. But, when translating equations in the symbolic language to equations on ordinary generating functions we have to deal with cycle index sums, otherwise information will get lost.

First we deal with the substitution $\mathcal{B} \circ \mathcal{A}$. It does not translate to $B(A(z))$, as by replacing every atom of a structure from $\mathcal{B}$ with a new structure from $\mathcal{A}$, we add and destroy symmetries. Let $\nu_1, \ldots, \nu_\ell$ be the elements of a cycle of length ℓ of a permutation of the atoms of an object $B \in \mathcal{B}$. We substitute each of the vertices $\nu_1, \ldots \nu_\ell$ with one element $A_\ell \in \mathcal{A}$. If any two of these substituted structures would be different from each other, the symmetry would be destroyed. Hence we have to substitute ℓ identical copies of an $A \in \mathcal{A}$ into the vertices $\nu_1, \ldots, \nu_\ell$ to maintain the symmetry, then every node of A forms a cycle of length ℓ with all of its identical copies.

The theoretical background therefore is given by Pólya-Redfield theorem, cf also [2]. Let D and R be finite sets and $M = R^D$, further let $\mathfrak{S}$ be a subset of $\mathfrak{S}_D$. A permutation $\sigma \in \mathfrak{S}$ induced a permutation

$$(\tilde{\sigma}(f))(x) := f(\sigma(x)), \qquad f \in M, x \in D,$$

in $\mathfrak{S}_M$, the induced set $\tilde{\mathfrak{S}}$ is a subgroup of $\mathfrak{S}_M$ isomorphic to $\mathfrak{S}$, we call two functions $f, g \in M$ equivalent ($f \sim g$) if there is a permutation $\sigma \in \mathfrak{S}$ with $\tilde{\sigma}(f) = g$. We further consider a weight function $\phi : R \to W$ for some weight set W and define the weight of f by

$$\phi(f) = \prod_{x\in D} \phi(f(x)).$$

Obviously $\phi(f) = \phi(g)$ for $f \sim g$, hence the weight $\phi(\mathbf{c})$ is defined for any equivalence class $\mathbf{c} \in M/_\sim$.

Theorem 1.5 (Pólya-Redfield). *With the above definitions,*

$$\sum_{\mathbf{c}\in M/\sim} \phi(\mathbf{c}) = P_{\mathfrak{G}}\left(\sum_{r\in R}\phi(r), \sum_{r\in R}\phi(r)^2, \ldots, \sum_{r\in R}\phi(r)^{|D|}\right)$$

This theorem can be extended formally to countable sets D, R. Consider a k-tuple $(A_1, \ldots, A_k)$ of objects $A_i \in \mathcal{A}, i = 1, \ldots k$. Its size is given by $\omega((A_1, \ldots, A_k)) = \sum_{i=1}^{k} \omega(A_i)$. Setting $D = \{1, \ldots, k\}$, $R = \mathcal{A}$ and $\phi(A) = z^{\omega(A)}$ we can apply the above theorem to obtain

$$Z_{\mathcal{B}\circ\mathcal{A}}(\mathbf{s_1}) = Z_{\mathcal{B}}(Z_{\mathcal{A}}(\mathbf{s_1}), Z_{\mathcal{A}}(\mathbf{s_2}), \ldots),$$

for $Z_{\mathcal{B}}(\mathbf{s_1})$ the cycle index sum of the class $\mathcal{B}$, $Z_{\mathcal{A}}$ the cycle index sum of the class $\mathcal{A}$ and where $Z_{\mathcal{A}}(\mathbf{s}_\ell)$ denotes the cycle index sum $Z_{\mathcal{A}}(s_\ell, s_{2\ell}, s_{3\ell}, \ldots)$ for all $\ell \geq 1$.

Thus the ordinary generating function $C(z)$ of the class $\mathcal{B} \circ \mathcal{A}$ is given by

$$C(z) = Z_{\mathcal{B}}(A(z), A(z^2), \ldots).$$

Throughout the whole thesis, we will use the following notation. Let $A(z)$ be some generating function, $P(\mathfrak{S}_n)(s_1, \ldots, s_n)$ the cycle index of a group of permutations of n elements, and $Z_{\mathcal{B}}(\mathbf{s_1})$ the cycle index sum of a class of structures. We denote by $P(\mathfrak{S}_n)(\mathbf{A}(\mathbf{z}))$ the substitution $s_1 \leftarrow A(z), s_2 \leftarrow A(z^2), \ldots, s_n \leftarrow A(z^n)$ and by $Z_{\mathcal{B}}(\mathbf{A}(\mathbf{z}))$ the according substitution on infinitely many variables.

With the following lemma, we can handle the Set-operator of the symbolic language. The Set-operator chooses a set of elements of arbitrary size, the group of allowed permutations on a set is the whole permutation group.

Lemma 1.6. *We denote by Z_n the cycle index of symmetric group on n elements. Then,*

$$\sum_{n\geq 0} Z_n(\mathbf{s_1}) = \exp(\sum_{n\geq 1}\frac{s_n}{n}).$$

.

Proof. Denote by $(\ell_1, \ell_2, \ldots)$ a set of arbitrary size with $\ell_i \in \mathbb{N}_0$. We have that

$$\begin{aligned}
\sum_{n\geq 0} Z_n(\mathbf{s_1}) &= \sum_{n\geq 0}\frac{1}{n!}\sum_{\ell_1+2\ell_2+\ldots n\ell_n=n}\frac{n!}{\ell_1!\ell_2!\cdots\ell_n!1^{\ell_1}2^{\ell_2}\cdots n^{\ell_n}}s_1^{\ell_1}s_2^{\ell_2}\cdots s_n^{\ell_n}\\
&= \sum_{(\ell_1,\ell_2,\ldots)}\frac{1}{\ell_1!\ell_2!\cdots\ell_n!}\frac{s_1^{\ell_1}}{1^{\ell_1}}\frac{s_2^{\ell_2}}{2^{\ell_2}}\cdots\frac{s_n^{\ell_n}}{n^{\ell_n}}\\
&= \sum_{\ell_1\geq 0}\frac{1}{\ell_1!}\left(\frac{s_1}{1}\right)^{\ell_1}\sum_{\ell_2\geq 0}\frac{1}{\ell_2!}\left(\frac{s_2}{2}\right)^{\ell_2}\cdots\\
&= \exp\left(\frac{s_1}{1}\right)\exp\left(\frac{s_2}{2}\right)\cdots = \exp\left(\sum_{n\geq 1}\frac{s_n}{n}\right)
\end{aligned}$$

□

The cycle index sum $Z_{\mathcal{B}}(\mathbf{s_1})$ of a set of objects $\mathcal{B} = \mathrm{Set}(\mathcal{A})$ is now given by

$$Z_{\mathcal{B}}(\mathbf{s_1}) = \sum_{n\geq 0}(Z_n \circ Z_{\mathcal{A}})(\mathbf{s_1}) = \exp\left(\sum_{\ell\geq 1}\frac{1}{\ell}Z_{\mathcal{A}}(\mathbf{s}_\ell)\right),$$

and the ordinary generating function $B(z)$ is given by

$$B(z) = \exp\left(\sum_{\ell\geq 1}\frac{1}{\ell}A(z^\ell)\right).$$

Let us consider a class $\mathcal{A}$ together with its derived class $\mathcal{A}'$ and its rooted class $\mathcal{A}^\bullet$. By pointing a vertex ν, we eliminate all permutations not fixing the pointed vertex, as the vertex is distinguished and thus not part of any symmetry. Hence, the cycle index sum of the derived class is given by

$$Z_{\mathcal{A}'}(\mathbf{s_1}) = \frac{\partial}{\partial s_1}Z_{\mathcal{A}}(\mathbf{s_1}).$$

The cycle index sum of the rooted class $\mathcal{A}^\bullet$ is then given by

$$Z_{\mathcal{A}^\bullet}(\mathbf{s_1}) = s_1 Z_{\mathcal{A}'}(\mathbf{s_1}).$$

Remark. Let $\mathcal{A}$ be an unlabelled class and $\tilde{\mathcal{A}}$ be the according labelled class. Further let $Z_{\mathcal{A}}((s_1))$ be the cycle index sum of $\mathcal{A}$ and $\tilde{A}(z)$ be the exponential generating function of $\tilde{\mathcal{A}}$. Note that

$$\tilde{A}(z) = Z_{\mathcal{A}}(z, 0, 0, \ldots).$$

Example. Let us consider the class of rooted unlabelled non-plane binary trees $\mathcal{C}$. Obviously they are the family obtained from the family of labelled binary trees by unlabelling. As described in the example on page 5 and depicted in Figure 1.2, they can be decomposed at the root, which leads to

$$\mathcal{C} = \mathcal{X} + \mathcal{C}^2,$$

when counting external vertices, i.e. leaves. Let us denote by K the binary tree consisting of one root and two leaves. It has cycle index $\frac{1}{2}(s_1^2 + s_2)$, as the permutation group consists of the identity and the reflection exchanging the two leaves. All trees with at least 2 leaves are obtained by substituting the leaves of K for binary trees $C_1, C_2 \in \mathcal{C}$, which is given by the symbolic operation $\mathcal{K} \circ \mathcal{C}$. Hence, the cycle index sum of binary trees is given by

$$Z_{\mathcal{C}}(\mathbf{s_1}) = s_1 + \frac{1}{2}\left(Z_{\mathcal{C}}(\mathbf{s_1})^2 + Z_{\mathcal{C}}(\mathbf{s_2})\right),$$

and the ordinary generating function $C(z)$ by

$$C(z) = z + \frac{1}{2}(A(z)^2 + A(z^2)).$$

Note that this equation is not explicitly solvable as the one in the labelled case is, hence we cannot extract coefficients from this equation directly. To obtain information on the number of unlabelled trees of size n, we will need analytic tools described in the next section.

Construction	Class	Labelled setting	Unlabelled setting
Neutral Class	$\mathcal{C} = 1$	$C(z) = 1$	$Z_{\mathcal{C}}(\mathbf{s_1}) = 1$
Atomic class	$\mathcal{C} = \mathcal{X}$	$C(z) = z$	$Z_{\mathcal{C}}(\mathbf{s_1}) = s_1$
Sequence	$\mathcal{C} = \mathrm{Seq}(\mathcal{X})$	$C(z) = \frac{1}{1-z}$	$Z_{\mathcal{C}}(z) = \frac{1}{1-s_1}$
Set	$\mathcal{C} = \mathrm{Set}(\mathcal{X})$	$C(z) = \exp(z)$	$Z_{\mathcal{C}}(\mathbf{s_1}) = \exp\left(\sum_{i\geq 1} \frac{s_i}{i}\right)$
Cycle	$\mathcal{C} = \mathrm{Cyc}(\mathcal{X})$	$C(z) = \log(\frac{1}{1-z})$	$Z_{\mathcal{C}}(\mathbf{s_1}) = \sum_{r\geq 1} \frac{\varphi(r)}{r} \log(\frac{1}{1-s_1})$
Sum	$\mathcal{C} = \mathcal{A} + \mathcal{B}$	$C(z) = A(z) + B(z)$	$Z_{\mathcal{C}}(\mathbf{s_1}) = Z_{\mathcal{A}}(\mathbf{s_1}) + Z_{\mathcal{B}}(\mathbf{s_1})$
Product	$\mathcal{C} = \mathcal{A} \times \mathcal{B}$	$C(z) = A(z) \cdot B(z)$	$Z_{\mathcal{C}}(\mathbf{s_1}) = Z_{\mathcal{A}}(\mathbf{s_1}) \cdot Z_{\mathcal{B}}(\mathbf{s_1})$
Substitution	$\mathcal{C} = \mathcal{A} \circ \mathcal{B}$	$C(z) = A(B(z))$	$Z_{\mathcal{C}}(\mathbf{s_1}) = Z_{\mathcal{A}}(Z_{\mathcal{B}}(\mathbf{s_1}), Z_{\mathcal{B}}(\mathbf{s_2}), \ldots)$

Table 1.1: The dictionary that translates combinatorial structures and constructions into (operations on) counting series. $\varphi(r)$ denotes the Eulerian totient function.

Walsh series In some counting problems considered in this thesis, we need to be even more precise on symmetries and consider a more refined version of cycle index sums, namely Walsh series. Walsh series track permutations on the set of vertices, but at the same time observe what impact those symmetries have on the edges. It is hence a refinement of the bivariate exponential generating function $A(z, y)$, where z counts vertices and y counts edges, as defined in detail in the next part. Walsh series contain 3 infinite series of variables $\mathbf{s_1} = s_1, s_2, \ldots$, $\mathbf{b_1} = b_1, b_2, \ldots$ and $\mathbf{c_1} = c_1, c_2, \ldots$, where $\mathbf{s_1}$ counts cycles on vertices as before, $\mathbf{b_1}$ indicates cycles on edges and $\mathbf{c_1}$ counts cycles on edges which additionaly change their orientation under the permutation.

Example. Consider a simple edge, that is 2 vertices connected by one edge. The symmetry group contains the identity as well as a reflection. Hence,

$$W_e = s_1^2 b_1 + s_2 c_1,$$

because under the reflection the edge remains fixed but changes its orientation. Now consider a cycle R of n vertices and n edges. The permutation group $\mathfrak{S}_R$ consists of cyclic permutations as well as reflections. The cycle index of a cycle on n elements is given by

$$P(\mathfrak{S}_R) = \begin{cases} \frac{1}{n}\left(\sum_{d|n} \varphi(d) s_d^{n/d} + \frac{1}{2}(s_1^2 s_2^{m-1} + s_2^m)\right) & \text{if } n = 2m \\ \frac{1}{n}\left(\sum_{d|n} \varphi(d) s_d^{n/d} + s_1 s_2^m\right) & \text{if } n = 2m+1 \end{cases}$$

Hence, the contribution to a Walsh series of R is

$$P(\mathfrak{S}_R) = \begin{cases} \frac{1}{n}\left(\sum_{d|n} \varphi(d) s_d^{n/d} b_d^{n/d} + \frac{1}{2}(b_1^m s_1^2 s_2^{m-1} + c_1^2 b_2^{m-2} s_2^m)\right) & \text{if } n = 2m \\ \frac{1}{n}\left(\sum_{d|n} \varphi(d) s_d^{n/d} b_d^{n/d} + s_1 c_1 s_2^m b_2^m\right) & \text{if } n = 2m+1 \end{cases} \tag{1.3}$$

We will need Walsh series only in the very last part of this thesis, where we compute the degree distribution of 2-connected series-parallel graphs.

Counting with additional parameters

In advanced enumeration problems, we want to go beyond counting graphs of size n in a family $\mathcal{A}$. We want to count graphs which have a certain property among those of size n, for example, graphs with n nodes and m edges or with n nodes of which m have a given degree d. Therefore, we introduce multivariate ordinary or exponential generating functions

$$A(z,v) = \sum_{n,m\geq 0} A_{n,m} z^n v^m \qquad \tilde{A}(z,v) = \sum_{n,m\geq 0} A_{n,m} \frac{z^n}{n!} v^m,$$

where the coefficents $A_{n,m}$ denote the number of objects in $\mathcal{A}$ with n vertices and m edges (or m vertices of degree d, respectively), where we suppose in the second case that all vertices are labelled and edges are unlabelled.

Analogously, we define cycle index sums for unlabelled objects $Z_{\mathcal{A}}(\mathbf{s_1}; \mathbf{u_1})$ on the "bivariate" infinite set of variables $(\mathbf{s_1}; \mathbf{u_1}) = (s_1, u_1; s_2, u_2; s_3, u_3; \ldots)$, where the s_i count cycles of vertices and the u_i track cycles of the second parameter. Note that the cycles of the second parameter are strongly related to the cycles of the vertices, in the case of v counting edges we obtain Walsh series mentioned above, in the case of counting degrees a symmetry only appears when all vertices of a cycle have the same degree, thus every circle on vertices is also a circle on vertices of degree d.

Of course, we can not only add one additional parameter, but a tupel of parameters of arbitrary but finite size, $\mathbf{v} = (v_1, \ldots, v_k)$, which in case of cycle index sums gives a multivariate set of variables $(\mathbf{s_1}; \bar{\mathbf{u}}_\mathbf{1}) = (\mathbf{s_1}; \mathbf{u_1}, \ldots, \mathbf{u_k})$.

The grammar and translations given in Table 1.1 can be passed on to multivariate cycle index sums and generating functions under specific conditions, namely that the parameter considered is a so-called *inherited* parameter, as described in [24]. That means, basically, that the value of the parameter remains unchanged under a sum operator, that it is given additively as the sum of the values of the involved objects in case of a product, and that it is given by the values of the substituted structures in case of a substitution. We also adapt the notation introduced earlier in a similar manner, that is let $A(z,v)$ be some generating function, $P(\mathfrak{S}_n)(s_1, \ldots, s_n)$ the cycle index of a group of permutations of n elements, and $Z_{\mathcal{B}}(\mathbf{s_1})$ the cycle index sum of a class of structures. We denote by $P(\mathfrak{S}_n)(\mathbf{A}(\mathbf{z}, \mathbf{v}))$ the substitution $s_1 \leftarrow A(z,v), s_2 \leftarrow A(z^2, v^2), \ldots, s_n \leftarrow A(z^n, v^n)$ and by $Z_{\mathcal{B}}(\mathbf{A}(\mathbf{z}, \mathbf{v}))$ the according substitution on infinitely many variables.

1.3 Analytic combinatorics

Generating functions are formal power series, but they can be considered as power series in a complex variable z. The following part is a presentation of a very powerful toolbox provided by this point of view, which leads us to the results presented in this thesis. More details and proofs of all theorems can be found in [15] or [24], as well as [14].

1.3.1 Singularity analysis

Given a generating function $A(z) = \sum_{n\geq 0} a_n z^n$ with positive radius of convergence $\rho > 0$, we can apply Cauchy's formula (cf e.g. [41])

$$a_n = \frac{1}{2\pi i} \int_\gamma A(z) \frac{\mathrm{d}z}{z^{n+1}}, \tag{1.4}$$

where γ is a closed contour encircling $z = 0$ once and contained in the region of analyticity of $A(z)$. Thus, the analytic behaviour of $A(z)$ provides information on its coefficients a_n. The method of *singularity analysis* shows that certain kinds of singularities give corresponding asymptotics for the coefficients a_n. The method was introduced by Flajolet and Odlyzko [23].

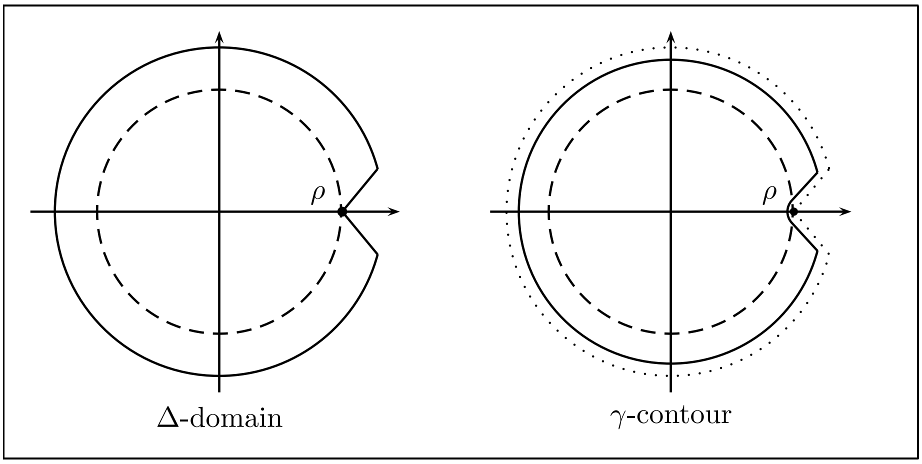

Figure 1.3: The Δ-domain and the integration contour γ

The basic preliminary for the following theorems is that a generating function $A(z)$ has an analytic continuation to a so-called Δ-domain (colloquially often referred to as a "Pac-man region" due to its shape, cf Figure 1.3), that is, there is a unique dominant positive singularity on the circle of convergence:

$$\Delta(\rho, \eta, \delta) = \{z \big| |z| < \rho + \eta, |\arg(\frac{z}{\rho} - 1)| > \delta\} \tag{1.5}$$

Further, for proving the results, we use a truncated contour $\gamma = \gamma_1 \cup \gamma_2 \cup \gamma_3 \cup \gamma_4$, as depicted in Figure 1.3, with

$$\begin{aligned}
\gamma_1 &= \left\{ z = \rho \left(1 + \frac{-i + (\eta n - t)}{n} e^{-i\delta} \right) \bigg| 0 \leq t \leq \eta n \right\}, \\
\gamma_2 &= \left\{ z = \rho \left(1 - \frac{e^{i\varphi}}{n} \right) \bigg| -\frac{\pi}{2} + \delta \leq \varphi \leq \frac{\pi}{2} - \delta \right\}, \\
\gamma_3 &= \left\{ z = \rho \left(1 + \frac{i + t}{n} e^{i\delta} \right) \bigg| 0 \leq t \leq \eta n \right\},
\end{aligned} \tag{1.6}$$

and γ_4 be a circular arc centered at the origin and closing the contour. The integral over γ_4 is exponentially smaller than the contributions of the parts near the singularity, which give the results. We state here the most important result for our purpose, for a complete introduction to singularity analysis and proofs see [24] or [15].

Lemma 1.7 (transfer lemma)**.** *Suppose that a generating function $A(z)$ is analytic in a Δ-domain such that*

$$A(z) \sim C\left(1-\frac{z}{\rho}\right)^{\alpha} \tag{1.7}$$

for $z \to \rho$ with $z \in \Delta$, where α is a complex number not in $\mathbb{N}_0$. Then, as n tends to infinity,

$$[z^n]A(z) \sim C\frac{n^{-\alpha-1}}{\Gamma(-\alpha)}\rho^{-n}$$

Definition 1.8. *Let $A(z)$ be a generating function with a unique dominant positive singularity $\rho > 0$. We call a local singular expansion of the form*

$$A(z) = g(z) - h(z)\left(1-\frac{z}{\rho}\right)^{\alpha}, \tag{1.8}$$

with $\alpha \notin \mathbb{N}$ and $g(z), h(z)$ functions which are analytic around ρ, a local singular expansion of order α *of $A(z)$.*

Note that due to Lemma 1.7, we get asymptotic estimates for the coefficients of a generating function with a singular expansion of order α. Most of the graph classes we are dealing with will have squareroot singular expansions, i.e. singular expansions of order $\alpha = \frac{1}{2}$, as we will see in the next section. The following lemma will be very helpful, as it gives a correspondance between the singular expansion of the generating function $A(z)$ of a family $\mathcal{A}$ and the one of the generating function $A'(z)$ of the derived family $\mathcal{A}'$. The proof can be found in [17].

Lemma 1.9. *Let $A(z)$ be a generating function with a square root singular expansion around a dominant positive singularity ρ. Then the derivative and the integral have local singular expansions of the form*

$$A'(z) = \frac{g_2(z)}{\sqrt{1-\frac{z}{\rho}}} + h_2(z) \tag{1.9}$$

and

$$\int_0^z A(t)\mathrm{d}t = g_3(z) + h_3(z)\left(1-\frac{z}{\rho}\right)^{\frac{3}{2}}, \tag{1.10}$$

where $g_2(z), h_2(z), g_3(z)$ and $h_3(z)$ are analytic at ρ.

Let us denote by $X = \sqrt{1-\frac{z}{\rho}}$. Note that a singular expansion of order $\frac{1}{2}$ rewrites to

$$A(z) = g(z) - h(z)X = \sum_{j\geq 0} a_n X^j = a - bX + \mathcal{O}(X^2),$$

because $z = \rho(X^2-1)$ and $g(z)$ and $h(z)$ are analytic at ρ and hence have a representation as a power series in z.

1.3.2 Functional equations

Throughout this thesis, generating functions will be given by functional equations which are often not explicitly solvable, as we saw in the example of binary unlabelled trees. There is a powerful set of theorems to deal with such implicit equations, showing that solutions of such equations usually have a square root singularity. The following theorem establishes the basis for all of the subsequent theory. The proof can be found in [15, Theorem 2.19], it is actually an application of the theory of functions defined by implicit equations in Analysis, see for example [41].

Theorem 1.10. *Suppose that $F(z, y)$ is an analytic function in z, y around $z = y = 0$ such that $F(0, y) = 0$ and that all Taylor coefficients of F around 0 are real and non-negative. Then there exists a unique analytic solution $y = y(z)$ of the functional equation*

$$y = F(z, y)$$

with $y(0) = 0$ that has non-negative Taylor coeffients around 0.
If the region of convergence of $F(z, y)$ is large enough such that there exist positive solutions $z = z_0$ and $y = y_0$ of the system of equations

$$\begin{aligned} y &= F(z, y) \\ 1 &= F_y(z, y) \end{aligned}$$

with $F_z(z_0, y_0) \neq 0$ and $F_{yy}(z_0, y_0) \neq 0$, then $y(z)$ is analytic for $|z| < z_0$ and there exist functions $g(z)$ and $h(z)$ that are analytic around $z = z_0$ such that $y(z)$ has a square root singular expansion (1.8) *locally around $z = z_0$. We have $g(z_0) = y(z_0)$ and*

$$h(z_0) = \sqrt{\frac{2z_0 F_z(z_0, y_0)}{F_{yy}(z_0, y_0)}}.$$

Moreover, (1.8) *provides a local analytic continuation of $y(z)$ (for $\arg(z - z_0) \neq 0$).*
If we assume that $[z^n]y(z) > 0$ for $n \geq n_0$, then $z = z_0$ is the only singularity of $y(z)$ on the circle $|z| = z_0$ and we obtain an asymptotic expansion for $[z^n]y(z)$ of the form

$$[z^n]y(z) = \sqrt{\frac{z_0 F_z(z_0, y_0)}{2\pi F_{yy}(z_0, y_0)}} z_o^{-n} n^{-\frac{3}{2}} \left(1 + \mathcal{O}(\frac{1}{n})\right). \tag{1.11}$$

1.3.3 Systems of functional equations

Consider a system of equations $\mathbf{y} = \mathbf{F}(z, \mathbf{y})$:

$$\begin{aligned} y_1 &= F_1(z, y_1, \dots, y_N) \\ \vdots &\quad \vdots \\ y_N &= F_N(z, y_1, \dots, y_N) \end{aligned} \tag{1.12}$$

We define a graph $G_{\mathbf{F}}$ on the set of vertices $\{1, \dots, N\}$ and the set of directed edges E, which contains an edge $(i \to j)$ if and only if the function F_i really depends on y_j, that is $\frac{\partial}{\partial y_j} F_i \neq 0$. We call this graph the dependency graph of the system (E).

Definition 1.11. *Consider a system of equations* (1.12). *It is called* strongly connected *if its dependency graph is strongly connected.*

The definition above is equivalent to the fact that no subsystem of (1.12) can be solved before the whole system can be solved. We denote by $\mathbf{F_y}$ the Jacobian matrix of the system (1.12), that is, it is the $N \times N$-matrix whose (i, j)-entry is $\frac{\partial}{\partial y_i} F_j$, and by $\mathbf{I}$ the $N \times N$-identity matrix. The following theorem is a refinement of Theorem 1.10 to systems of equations and is well known as the Drmota-Lalley-Woods theorem, named after the three authors stating the theorem at approximately the same time in different contexts, cf [24, 14, 50, 71].

Theorem 1.12 (Drmota-Lalley-Woods Theorem). *Let $\mathbf{y} = \mathbf{F}(z, \mathbf{y})$ be a nonlinear system of functional equations which is strongly connected, has only nonnegative Taylor coefficients and which is analytic around $z = 0$ and $\mathbf{y} = \mathbf{0}$. Further assume that $\mathbf{F}(0, \mathbf{y}) = \mathbf{0}$, $\mathbf{F}(z, \mathbf{0}) \neq 0$ and $\mathbf{F}_z(z, \mathbf{y}) \neq 0$ and that the region of convergence of $\mathbf{F}$ is large enough such that the system*

$$\begin{aligned}\mathbf{y} &= \mathbf{F}(z, \mathbf{y})\\ 0 &= \det(\mathbf{I} - \mathbf{F}_{\mathbf{y}}(z, \mathbf{y}))\end{aligned}$$

has solutions $z = \rho$ and $\mathbf{y} = \mathbf{y}_0$ that are real, positive and minimal. Let $\mathbf{y} = \mathbf{y}(z)$ denote the analytic solutions of the system $\mathbf{y} = \mathbf{F}(z, \mathbf{y})$ with $\mathbf{y}(0) = \mathbf{0}$. Then all solutions $y_j(z, v), j = 1 \dots, N$ have a square root singular expansion locally around ρ (for $\arg(z - \rho) \neq 0$):

$$y_j(z) = g_j(z) - h_j(z)\sqrt{1 - \frac{z}{\rho}}, \qquad \text{for } i = 1, \dots N,$$

with analytic functions $g_j(z)$ and $h_j(z)$ and with $g_j(\rho) = y_j(\rho) = (\mathbf{y}_0)_j$. Furthermore, if $[z^n]y_j(z) > 0$ for $1 \leq j \leq N$ for sufficiently large $n \geq n_0$ then $\mathbf{y}(z)$ has a unique smallest positive singularity at ρ within $|z| = \rho$.

Note that the above theorem holds for algebraic systems of equations. In counting unlabelled structures, we will often be faced with non-algebraic systems, including terms $\mathbf{y}(z^2), \mathbf{y}(z^3), \cdots$. Drmota-Lalley Woods Theorem can be extended to our needs in such cases by the following lemma:

Lemma 1.13. *Suppose that the system of equations*

$$\begin{aligned}y_1 &= F_1(z, y_1, \dots, y_N, y_1(z^2), \dots, y_N(z^2), y_1(z^3), \dots)\\ &\vdots\\ y_N &= F_N(z, y_1, \dots, y_N, y_1(z^2), \dots, y_N(z^2), y_1(z^3), \dots)\end{aligned}$$

has a solution $\mathbf{y}(z) = (y_1(z), \dots, y_N(z))$ which is analytic and all functions $y(z)$ have non-negative taylor coefficients. Further assume that the system $\boldsymbol{f} = \mathbf{G}(z, \boldsymbol{f})$ given by

$$\begin{aligned}f_1 &= F_1(z, f_1, \dots, f_N, y_1(z^2), \dots, y_N(z^2), y_1(z^3), \dots) =: G_1(z, f_1, \dots, f_N)\\ &\vdots\\ f_N &= F_N(z, f_1, \dots, f_N, y_1(z^2), \dots, y_N(z^2), y_1(z^3), \dots) =: G_N(z, f_1, \dots, f_N)\end{aligned}$$

fulfills the preliminaries of Theorem 1.12 and that $\sum_{i=1}^{N} y_i(z) = y(z)$ has radius of convergence $0 < \rho < 1$. Then, the result of Theorem 1.12 follows for the solution $\boldsymbol{f} = \mathbf{y}(z)$.

Proof. We have $y(z) = \sum_n y_n z^n$. For every $i = 1, \dots, N$, $y_i(z) = \sum_n y_{n,i} z^i$, with $y_{n,i} \leq y_n$, as $y_i(z)$ is a counting series counting a substructure of the structure counted by $y(z)$. Hence the radius of convergence ρ_i of $y_i(z)$ is greater or equal to ρ, $\rho_i \geq \rho$, thus $y_i(z^2)$ is analytic at $|z| \leq \rho + \epsilon$. Hence we can consider the system $\mathbf{G}(z, \boldsymbol{f})$ and apply Drmota-Lalley-Woods theorem to it. □

Having the solution $\mathbf{y}(z)$ of a system of equations (E), we further get the following Lemma.

Lemma 1.14. *Let $\mathbf{y}(z) = (y_1(z), \ldots, y_N(z))$ be the solution of the system of equations (E) and assume that all assumptions of Drmota-Lalley-Woods Theorem are satisfied. Suppose that $G(z, \mathbf{y})$ is a power series such that the point $(z_0, \mathbf{y}_0(z_0))$ is contained in the interior of the region of convergence of $G(z, \mathbf{y})$ and that $G_{\mathbf{y}}(z_0, \mathbf{y}_0(z_0)) \neq 0$.*
Then $y_G(z) := G(z, \mathbf{y}(z))$ has a representation of the form

$$y_G(z) = g(z) - h(z)\sqrt{1 - \frac{z}{z_0}}$$

for $|z - z_0| < \epsilon$, where $g(z)$ and $h(z)$ are analytic functions, and $y_G(z)$ is analytic in a Δ-domain with $z_0 = \rho$.

1.3.4 Additional parameters

In this section we add an additional k-dimensional parameter $\mathbf{v} = v_1, \ldots, v_k$ to the considerations above. For suitable parameters (we will come back to this later), the theory of analytic combinatorics introduced so far can be extended to this concept, which will be helpful in the study of graph parameters such as numbers of edges or degree distribution. Additionally, singularity analysis and some type of quasi power theorem [43] will provide central limit theorems for the distributions of such parameters.

First of all, we extend the definition of singular expansion to additional parameters. Let $A(z, \mathbf{v})$ be a multivariate generating function.

Definition 1.15.

- *A valuation $\mathbf{v}_0$ of $\mathbf{v}$ is called* admissible *if all components of $\mathbf{v}_0$ are positive and if $A(z, \mathbf{v}_0)$ is a valid power series in z, i.e. $[z^n]A(z, \mathbf{v}_0) < \infty$ for every $n \geq 0$.*
- *Consider a fixed point valuation $(z_0, \mathbf{v}_0)$ of $(z, \mathbf{v})$. Then $A(z, \mathbf{v})$ is said to have a singular expansion of order α around $(z_0, \mathbf{v}_0)$ if $\mathbf{v}_0$ is an admissible valuation, z_0 is the radius of convergence of $A(z, \mathbf{v}_0)$, and the expansion*

$$A(z, \mathbf{v}) = g(z, \mathbf{v}) - h(z, \mathbf{v})\left(1 - \frac{z}{\rho(\mathbf{v})}\right)^{\alpha}$$

holds in a neighbourhood of $(z_0, \mathbf{v}_0)$ (except in the part where $1 - \frac{z}{\rho(\mathbf{v})} \in \mathbb{R}^-$), where $g(z, \mathbf{v})$ and $h(z, \mathbf{v})$ are analytic functions at $(z_0, \mathbf{v}_0)$, $h(z_0, \mathbf{v}_0) > 0$, and $\rho(\mathbf{v})$ is an analytic function with $\rho(\mathbf{v}_0) = z_0$. The function $\rho(\mathbf{v})$ is called the singularity function of $A(z, \mathbf{v})$ relative to z. It is the dominant singularity of the mapping $z \mapsto A(z, \mathbf{v})$.

In our applications, we will usually introduce additional parameters by taking a univariate generating function $A(z)$ and refining it to $A(z, \mathbf{v})$ by counting some additional parameters. Therefore, the valuation $\mathbf{v}_0$ we will use is $\mathbf{v}_0 = \mathbf{1} = (1, 1, \ldots, 1)$, which is an admissible valuation if $A(z)$ is the counting series of a class. In this case $A(z, \mathbf{v}_0) = A(z)$. As mentioned above, the theory of singularity analysis can be applied also with multivariate parameters, as we see in the following refinement of Theorem 1.10, cf [15][Theorem 2.21].

Theorem 1.16. *Suppose that $F(z, y, \mathbf{v}) = \sum_{n,m} F_{n,m}(\mathbf{v}) z^n y^m$ is an analytic function in z, y around $z = y = 0$ and $\mathbf{v} = \mathbf{0}$ such that $F(0, y, \mathbf{v}) = 0$, that $F(z, 0, \mathbf{v}) \neq 0$, and that all coefficients $F_{n,m}(\mathbf{1})$ of $F(z, y, \mathbf{1})$ are real and non-negative. Then the unique analytic solution $y = y(z, \mathbf{v}) = \sum_{n \geq 0} y_n(\mathbf{v}) z^n$ of the functional equation*

$$y = F(z, y, \mathbf{v})$$

with $y(0,\mathbf{v}) = 0$ is analytic around $\mathbf{0}$ and $y(z,\mathbf{1}) = \sum_{n\geq 0} y_n(\mathbf{1})z^n$ has non-negative coefficients $y_n(\mathbf{1})$.
If the region of convergence of $F(z,y,\mathbf{v})$ is large enough such that there exist non-negative solutions $z = z_0$ and $y = y_0$ of the system of equations

$$\begin{aligned} y &= F(z,y,\mathbf{1}) \\ 1 &= F_y(z,y,\mathbf{1}) \end{aligned}$$

with $F_z(z_0,y_0,\mathbf{1}) \neq 0$ and $F_{yy}(z_0,y_0,\mathbf{1}) \neq 0$, then there exist functions $f(\mathbf{v}), g(z,\mathbf{v})$ and $h(z,\mathbf{v})$ which are analytic around $z = z_0$, $\mathbf{v} = 1$ such that $y(z,\mathbf{v})$ is analytic for $|z| < z_0$ and $|u_j - 1| \leq \epsilon$ (for some $\epsilon > 0$ and $1 \leq j \leq k$) and has a representation of the form

$$y(z,\mathbf{v}) = g(z,\mathbf{v}) - h(z,\mathbf{v})\sqrt{1 - \frac{z}{f(\mathbf{v})}} \tag{1.13}$$

locally around $z = z_0, \mathbf{v} = \mathbf{1}$.

Moreover, if $y_n(\mathbf{1}) > 0$ for $n \geq n_0$, we also get

$$y_n(\mathbf{v}) = \sqrt{\frac{f(\mathbf{v})F_z(f(\mathbf{v}),y(f(\mathbf{v}),\mathbf{v}),\mathbf{v})}{2\pi F_{yy}(f(\mathbf{v}),y(f(\mathbf{v}),\mathbf{v}),\mathbf{v})}} f(\mathbf{v})^{-n} n^{-\frac{3}{2}} \left(1 + \mathcal{O}(\frac{1}{n})\right). \tag{1.14}$$

uniformly for $|u_j - 1| < \epsilon, 1 \leq j \leq k$.

Stable parameters

Definition 1.17. *Let $\mathbf{v} = (v_1, \ldots, v_k)$ be a tupel of parameters whose values behave linearly in the number of vertices, that is, there exists $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_k)$ with $\alpha_i > 0$ for all i such that the coefficient $a_{n,\mathbf{m}}$ in $A(z,\mathbf{v})$ is 0 if $m_i > \alpha_i n$ for at least one $i \in \{1, \ldots, k\}$. We call such parameters stable parameters.*

Note that every parameter counting a characteristic of nodes is a stable parameter with $\alpha = 1$, as at most every vertex can have the property. Another example for a stable parameter is the number of edges in labelled planar graphs, which is asymptotically bounded by αn with $\alpha = 2.56$ (cf [29, 9, 56]).
From the following lemma we can conclude that the radius of convergence of a mapping $z \mapsto A(z,\mathbf{v})$ is continuous at $\mathbf{v} = \mathbf{1}$ if $\mathbf{v}$ counts a stable parameter.

Lemma 1.18. *Let $A(z,\mathbf{v}) = a_{n,\mathbf{m}}z^n\mathbf{v}^{\mathbf{m}}$ be a power series with non-negative coefficients with the parameters counted by $\mathbf{v}$ behaving at most linear in n with constants $\boldsymbol{\alpha}, \alpha_i > 0\ \forall i$. Let $\rho(\mathbf{v})$ denote the radius of convergence of the mapping $z \mapsto A(z,\mathbf{v})$, Then, for real $v_i > 0$ and for all i*

$$\rho(\mathbf{1})\prod_{i=1}^{k}\min\{1, v_i^{-\alpha_i}\} \leq \rho(\mathbf{v}) \leq \rho(\mathbf{1})\prod_{i=1}^{k}\max\{1, v_i^{-\alpha_i}\}.$$

Proof. If for all $i = 1, \ldots k$ we have $v_i \geq 1$ then

$$\sum_{\mathbf{m}\geq 0} a_{n,\mathbf{m}} \leq \sum_{\mathbf{m}\geq 0} a_{n,\mathbf{m}}\mathbf{v}^m \leq \sum_{\mathbf{m}\geq 0} a_{n,\mathbf{m}}\mathbf{v}^{\alpha n},$$

where $\sum_{\mathbf{m}\geq 0}$ denotes the sum over all tupels $\mathbf{m}$ and $\mathbf{v}^{\alpha n}$ denotes $v_1^{\alpha_1 n} \cdots v_k^{\alpha_k n}$. Hence

$$A(|z|,\mathbf{1}) \leq A(|z|,\mathbf{v}) \leq A(|z\mathbf{v}^{\alpha}|, 1),$$

which implies that $\rho(\mathbf{v}) \geq \rho(\mathbf{1})\mathbf{v}^{-\alpha}$. Similarly, we argue if $0 < v_i < 1$ for all i. If some $v_i \geq 1$ and others $0 < v_j < 1$ we have

$$\sum_{\mathbf{m}\geq 0} a_{n,\mathbf{m}}\mathbf{v}^m \leq \sum_{\mathbf{m}\geq 0} a_{n,\mathbf{m}} \prod_{i:v_i\geq 1} v_i^{\alpha_i} \leq \sum_{\mathbf{m}\geq 0} a_{n,\mathbf{m}} \prod_{i:v_i\geq 1} v^{\alpha_i n},$$

and vice versa. Hence the result follows. □

Given stable parameters, Lemmas 1.9 and 1.14 and Theorems 1.10 and 1.12 generalize to theorems on multivariate generating functions like Theorem 1.16. Those specialisations are stated in Appendix A.

Central limit laws

The concept of multivariate generating functions turns out to be useful to study the distribution of parameters like the number of vertices of a given degree. There is a strong relation between multivariate generating functions with a square root singular expansion and random variables which are asymptotically Gaussian distributed.

Recall that a random variable X is called Gaussian or normally distributed $\mathcal{N}(\mu, \sigma^2)$ if its distribution function is of the form

$$\mathbb{P}(X \leq z) = f\left(\frac{z-\mu}{\sigma}\right),$$

where f is given by

$$f(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{1}{2}t^2} \mathrm{d}t,$$

and μ and σ are real and σ is positive. We have that $\mathbb{E}(X) = \mu$ and $\mathbb{V}\mathrm{ar}(X) = \sigma^2$.

We say that a sequence of random variables $(X_n)_{n\geq 0}$ satisfies a central limit law with mean μ_n and variance σ_n^2, if

$$\mathbb{P}(X_n \leq \mu_n + z\sigma_n) = f(z) + o(1),$$

as n tends to infinity. This is equivalent to

$$\frac{X_n - \mu_n}{\sigma_n} \xrightarrow{d} \mathcal{N}(0,1),$$

where $\xrightarrow{d}$ denotes weak convergence of random variables, cf [4].

The following Quasi-power theorem by H.-K. Hwang [43], cf also [15, 24], is very helpful to prove a central limit theorem for stable parameters on graphs.

Theorem 1.19 (Quasi power theorem)**.** *Let X_n be a sequence of random variables with the property that*

$$\mathbb{E}(v^{X_n}) = e^{\lambda_n \cdot A(v) + B(v)} \left(1 + \mathcal{O}\left(\frac{1}{\phi_n}\right)\right)$$

holds uniformly in a complex neighbourhood of $v = 1$, where λ_n and ϕ_n are sequences of positive real numbers with $\lambda_n, \phi_n \to \infty$ and $A(v)$ and $B(v)$ are analytic functions in a *neighbourhood of $v = 1$ with $A(1) = B(1) = 0$. Then X_n satisfies a central limit theorem of the form*

$$\frac{1}{\sqrt{\lambda_n}}\left(X_n - \mathbb{E}(X_n)\right) \xrightarrow{d} \mathcal{N}(0, \sigma^2)$$

with

$$\mathbb{E}(X_n) = \lambda_n \mu + \mathcal{O}\left(1 + \frac{\lambda_n}{\phi_n}\right), \qquad \mu = A'(1),$$
$$\mathbb{V}ar(X_n) = \lambda_n \sigma^2 + \mathcal{O}\left((1 + \frac{\lambda_n}{\phi_n})^2\right), \qquad \sigma^2 = A''(1) + A'(1).$$

By combining Theorems 1.16 and 1.19 we obtain the following central limit theorem for bivariate generating functions:

Theorem 1.20. *Suppose that X_n is a sequence of random variables such that*

$$\mathbb{E}v^{X_n} = \frac{[z^n]y(z,v)}{[z^n]y(z,1)}, \tag{1.15}$$

where $y(z,v)$ is a power series which is the (analytic) solution of the functional equation $y = F(z,y,v)$, where $F(z,y,v)$ satisfies the assumption of Theorem 1.16. In particular, let $z_0 > 0$ and $y_0 > 0$ be the (minimal) solution of the system of equations

$$y = F(z,y,v)$$
$$1 = F_y(z,y,v)$$

and set

$$\mu = \frac{F_v}{z_0 F_z},$$
$$\sigma^2 = \mu + \mu^2 + \frac{1}{z_0 F_z^3 F_{yy}}\left(F_z^2\left(F_{yy}F_{vv} - F_{yu}^2\right) - 2F_zF_v\left(F_{yy}F_{zv} - F_{yz}F_{yv}\right) + F_v^2\left(F_{yy}F_{zz} - F_{yz}^2\right)\right),$$

where all partial derivatives are evaluated at the point $(z_0, y_0, 1)$. Then we have

$$\mathbb{E}(X_n) = \mu n + \mathcal{O}(1) \quad \text{and} \quad \mathbb{V}ar(X_n) = \sigma^2 n + \mathcal{O}(1),$$

and if $\sigma^2 > 0$ then

$$\frac{X_n - \mathbb{E}(X_n)}{\sqrt{\mathbb{V}ar(X_n)}} \xrightarrow{d} \mathcal{N}(0,1).$$

Note that for a random variable X_n counting the value of an additional parameter in a counting problem, assumption (1.15) is fulfilled: Let $A(z,v)$ be the generating function where z counts the size and v the additional parameter. Then the probability that the value of the parameter is m in an object of size n is the number of objects of size n with parameter value m divided by the number of all objects of size n:

$$\mathbb{P}(X_n = m) = \frac{A_{n,m}}{A_n},$$

where $A_n = [z^n]A(z,1)$. Thus the expectation of v^{X_n} is given by

$$\mathbb{E}(v^{X_n}) = \sum_{m\geq 0} \mathbb{P}(X_n = m)v^m = \sum_{m \geq 0} \frac{A_{n,m}v^m}{A_n} = \frac{[z^n]A(z,v)}{[z^n]A(z,1)}.$$

Remark. Both theorems, Theorems 1.19 and 1.20, can also be stated for multidimensional parameters $\mathbf{v} = (v_1, \ldots, v_k)$ and sequences of vectors of random variables $\mathbf{X}_n = (X_1, \ldots, X_k)_n$. Furthermore, there is an extension to parameters in systems of generating functions. Both extensions can be found in Appendix A.

CHAPTER 2

Random Pólya trees

This chapter is devoted to the study of unlabelled rooted trees, widely known as Pólya trees. Pólya trees have been thorougly studied during many decades, beginning with George Pólya's work [60] in the 1930s. They have been a rich topic not only for their own sake, but also due to their close relation to simply generated trees, which can be interpreted as trees arising from a critical Galton-Watson branching process.

It has been proven that simply generated and Pólya trees behave similarly in many parameters such as asymptotic number [60], height[19, 21], profile process [19, 21] (cf Section 2.1.2) and many more. Marckert and Miermont [52] have shown that binary unlabelled trees converge in some sense to the continuum random tree, which is the same limit as that for simply generated trees. Still, it is well known and proven by simple means in [21] that Pólya trees cannot be generated by a branching process and thus do not belong to the class of simply generated trees, but the question for a structural difference is not answered.

In an attempt to find structural differences between Pólya trees and simply generated trees we study a parameter located at the fringe of Pólya trees in Section 2.2, where we informally speak of the fringe as the structure of the tree close to its leaves. To be more precice, we study the size of Ward-trees of Pólya trees, those are the subtrees of a tree T which are rooted at parent nodes of leaves. It is conjectured that the structural difference between simply generated and Pólya trees is to be found in the fringe. Unfortunately, we could not find a significant difference in the properties of the fringe we studied, but there are parameters left to be studied in forthcoming work.

In Section 2.3 of this chapter, we prove that the degree profile of Pólya trees joins all other parameters and shows a very similar behaviour to the one found in simply generated trees ([20]), namely convergence to a suitably normalized stochastic process known as the local time of a Brownian excursion.

2.1 Preliminaries

We will give a brief introduction to the main results on Pólya trees which will be the base of the upcoming results. The following results have been obtained by different researchers, e.g. Pólya[60], Otter[57], Robinson and Schwenk [62] and many more.

Pólya trees

As for other families of graphs, we decompose a tree $T \in \mathcal{T}$ at its root in order to obtain results on the generating function. Therefore, consider the following. A tree is either just one node, the root, or it is the root together with a set of smaller rooted (sub-)trees attached to the root, cf Figure 2.1. Note that in Pólya trees the subtrees of the root are not ordered,

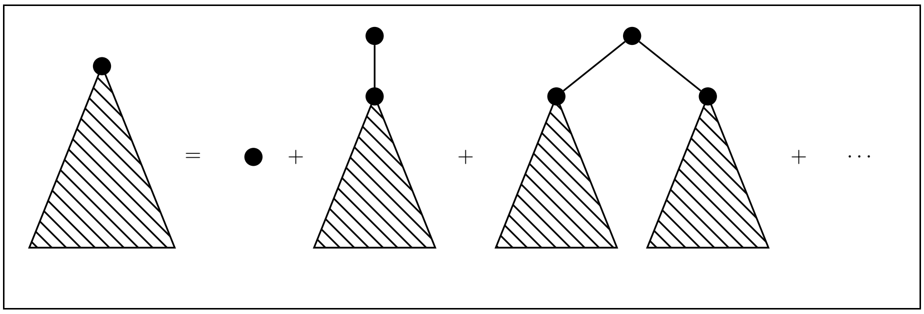

Figure 2.1: The decomposition of a tree at its root.

as they are non-plane. Therefore the recursive decomposition above translates to

$$\mathcal{T} = \mathcal{X} + \mathcal{X} \times \sum_{k\geq 1} \mathrm{Set}_k(\mathcal{T}), \tag{2.1}$$

where $\mathrm{Set}_k(\mathcal{T})$ denotes an unordered set of k elements of $\mathcal{T}$. Translating this to ordinary generating functions via cycle index sums, as described in Chapter 1 (cf Table 1.1), we obtain

$$y(z) = z + z \sum_{k\geq 1} Z_k(y(z), y(z^2), \dots, y(z^k)).$$

Using Lemma 1.6 this leads to

$$y(z) = z \exp\left(\sum_{k\geq 1} \frac{y(z^k)}{k}\right). \tag{2.2}$$

It has been shown that $y(z)$ has a dominant positive singularity at $\rho \approx 0.3383219$, with $y(\rho) = 1$, and around it it has a local expansion of order $\frac{1}{2}$ of the form

$$\begin{aligned} y(z) &= 1 - b\sqrt{\rho - z} + c(\rho - z) - + \cdots \qquad (2.3)\\ &= 1 - b\sqrt{\rho}\sqrt{1 - \frac{z}{\rho}} + c\rho(1 - \frac{z}{\rho}) - + \cdots, \end{aligned}$$

with $b \approx 2.6811266$. From there asymptotic coefficients can be deduced via a transfer lemma (cf Lemma 1.7). We obtain

$$y_n \sim \frac{b\sqrt{\rho}}{2\sqrt{\pi}} n^{-\frac{3}{2}} \rho^{-n}. \tag{2.4}$$

Let Δ be a Δ-domain as given in (1.5). We can prove the following Lemma for the generating function $y(z)$, which will be helpful in the forthcoming parts.

Lemma 2.1. *Provided that η in* (1.5) *is sufficiently small, the generating function of Pólya trees $y(z)$ has the following properties:*

(a) For $z \in \Delta$ we have that $|y(z)| \leq 1$. Equality holds only for $z = \rho$.

(b) Let $z = \rho\left(1 - \frac{1+it}{n}\right)$ and $|t| \leq C \log^2 n$ for some fixed $C > 0$. Then there is a $c > 0$ such that

$$|y(z)| \leq 1 - c\sqrt{\frac{\max(1, |t|)}{n}}.$$

(c) For $|z| \leq \rho$ we have $|y(z)| \leq y(|z|) \leq 1$. Moreover, near $z = 0$ the asymptotic relation $y(z) \sim z$ holds.

(d) There exists an $\epsilon > 0$ such that

$$|y(z)| \geq \min\left(\frac{\epsilon}{2}, \frac{|z|}{2}\right)$$

for all $z \in \Theta$.

For the proof of the Lemma see [21][Lemma 1].

Simply generated trees

Simply generated trees have been introduced by Meir and Moon [54], and are weighted rooted trees, where the weights are given according to the degree distribution.

Definition 2.2. *Let $\mathcal{T}$ denote a family of rooted trees and $T(z) = \sum_{n\geq 0} T_n z^n$ be its generating function. $\mathcal{T}$ is called a simply generated family of trees, if $T(z)$ fulfills*

$$T(z) = z\Phi(T(z)), \qquad \Phi(t) = \sum_{i\geq 0} \phi_i t^i, \quad \phi_i \geq 0, \phi_0 > 0$$

The above definition can be interpreted as assigning the weight ϕ_i to a node of outdegree i.

Example. The following families are simply generated families:

- *labelled trees*
$$T(z) = ze^{T(z)}, \quad \Phi(t) = e^t = \sum_{i\geq 0} \frac{t^i}{i!}.$$

- *binary trees*
$$T(z) = z(1 + T(z))^2, \quad \Phi(t) = 1 + 2t + t^2.$$

- *strict binary trees*
$$T(z) = z(1 + T(z)^2), \quad \Phi(t) = 1 + t^2.$$

It is easily shown that $T(z)$ has a dominant positive singularity at $z = \frac{1}{\Phi'(\tau)}$, where τ fulfills $\tau\Phi'(\tau) = \Phi(\tau)$, and that it has a local singular expansion

$$T(z) \sim \tau - \tilde{b}\sqrt{1 - z\Phi'(\tau)} + \mathcal{O}((1 - z\Phi'(\tau)),$$

where $\tilde{b} = \sqrt{\frac{2\Phi(\tau)}{\Phi''(\tau)}}$. Furthermore

$$T_n \sim \frac{\tilde{b}}{2\sqrt{\pi}} n^{-\frac{3}{2}} \Phi'(\tau)^n = \sqrt{\frac{\Phi(\tau)}{2\pi\Phi''(\tau)}} \frac{\Phi'(\tau)^n}{n^{\frac{3}{2}}} \tag{2.5}$$

2.1.1 The degree distribution

The distribution of degrees in a large random Pólya tree has already been studied in 1975 by Robinson and Schwenk [62]. The results of their work have later been extended by Drmota and Gittenberger in [20], where the degree distribution is studied for several tree classes, including simply generated trees and Pólya trees. The following result holds for Pólya trees, an analogue theorem with different constants holds for simply generated trees.

Theorem 2.3. *Let $X_n^{(d)}$ be the random variable that counts the number of vertices of degree d in a random Pólya tree of size n. Then the expected value of $X_n^{(d)}$ is asymptotically given by*

$$\mathbb{E}X_n^{(d)} = \mu_d n + \mathcal{O}(1) = \frac{2C_d}{b^2\rho}\rho^d n + \mathcal{O}(1),$$

where $C_d = C + \mathcal{O}\left(d\rho^d\right)$ with $C \approx 7.7581604\ldots$ is the constant

$$C = \exp\left(\sum_{\ell\geq 1} \frac{1}{\ell}\left(\frac{y(\rho^\ell)}{\rho^\ell} - 1\right)\right).$$

Furthermore X_n follows a central limit law of the form

$$\frac{X_n^{(d)} - \mathbb{E}(X_n^{(d)})}{\sqrt{\mathbb{V}\mathrm{ar}(X_n^{(d)})}} \xrightarrow{d} \mathcal{N}(0,1),$$

with the above mean $\mathbb{E}(X_n^{(d)}) = \mu_d n$ and the same variance $\mathbb{V}\mathrm{ar}(X_n^{(d)}) = \mu_d n$.

Note that the above theorem holds for fixed degree d. If the given degree grows with the size, a phase transition occurs (cf [55] for simply generated trees and [36] for Pólya trees).

Studying the degree distribution is a first step towards the study of patterns in trees. The question considered here is the occurence of certain trees as a substructure of a large tree as well as the number of such occurences. Note therefore that a vertex of degree d is a star graph and hence an easy pattern, cf Figure 2.2.

A study of general patterns instead of star graphs was carried out by Chyzak et al [12].

2.1.2 Height and profile

Consider the size of level k in a tree, that is the number of certices at distance k from the root, as well as the height of a tree, that is the maximum distance between the root and a leaf. In this section we summarize two results given by Drmota and Gittenberger [21], which provide the base for the results in the Section 2.3. Therefore, we first have to introduce some stochastic processes.

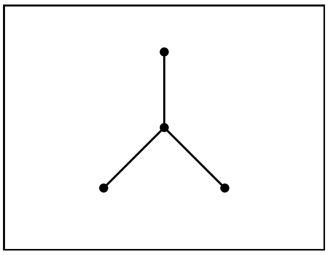

Figure 2.2: A vertex of degree 3 is a certain pattern in a tree.

The Brownian motion

We will state the definition of a Brownian local time, which will appear in the subsequent. A thorough overview of Brownian local times and related processes can be found in [61]. Explicit representations for the moments and the density of the one-dimensional projections of the local time of a Brownian excursion and related processes have been derived by Takács [65, 66]. Multi-dimensional analogues can be found in [38, 39]. For results on density representations for related processes such as occupation times we refer to [19, 44, 42].

A *Brownian Motion* (or Wiener process) is a stochastic process $W(t), t \geq 0$ with the following properties:

- The process starts at 0: $\mathbb{P}(W(0) = 0) = 1$
- The increments are independent.
- For $0 \leq s < t$ the increment $W(t) - W(s)$ is normally distributed with mean 0 and variance $t - s$.

It is proven that such a process exists (see e.g. [4]).

Further, let $W(t), t \geq 0$ be a Brownian motion and let $t_0 > 0, t_1 > 0$ be two consecutive zeros of $W(t)$, that is $|W(t)| > 0$ for $t_0 < t < t_1$.
We define the *Brownian excursion* $B(t), t \in [0, 1]$, associated to $W(t)$ by

- (i) $W(t_0) = B(0) = B(1) = W(t_1) = 0$,
- (ii) $B(t) = |W(t_0 + (t_1 - t_0)t)|$.

That is, $B(t)$ is the part of a Brownian motion $W(t)$ between two positive zeros t_0, t_1, rescaled on the interval $[0, 1]$.

The *local time* $l(s)$ characterises the amount of time the process spends at a given level s. It is defined by

$$T(s, s+\epsilon) := \int_0^1 \chi_{[s,s+\epsilon]}(B(a))\mathrm{d}a$$
$$l(s) := \lim_{\epsilon \to 0} \frac{1}{\epsilon} T(s, s+\epsilon)$$

A possible characterisation of $l(s)$ is via its characteristic function $\phi_\kappa(t) = \mathbb{E}(e^{itl(\kappa)})$.

Height and Profile

We define the following parameters in trees:

Definition 2.4. *The height H_T of a tree T is the length of the longest path, starting at the root, in T, i.e. the maximum number of edges on a path from the root to another vertex.*

Definition 2.5. *The profile $(L_T(k))_{k\geq 0}$ of a tree T is the number of nodes in T at distance k from the root. By linear interpolation we obtain a continuous function $L_T(t)$.*

$$L_T(t) = (\lfloor t \rfloor + 1 - t) L_T(\lfloor t \rfloor) + (t - \lfloor t \rfloor) L_T(\lfloor t \rfloor + 1), \quad t \geq 0$$

For an example, see Figure 2.3.

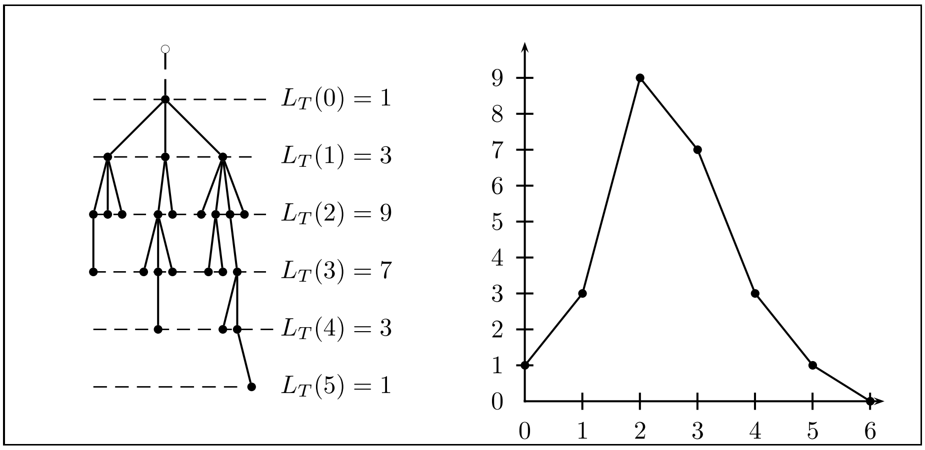

Figure 2.3: The (interpolated) profile of a Pólya tree.

Drawing a tree $T \in \mathcal{T}_n$ uniformly at random as in our usual random model, the parameters above get a random property as well. That is, the height becomes a random variable H_n, the profile a sequence of random variables $(L_n(k))_{k\geq 0}$, and the interpolated profile a stochastic process $L_n(t), t \geq 0$.

It has been proven in [21] that the following holds for the moments of H_n:

Theorem 2.6 (Drmota, Gittenberger, 2010)**.** *Let H_n denote the height of a random Pólya tree with n vertices. Then*

$$\mathbb{E}(H_n) \sim \frac{2\sqrt{\pi}}{b\sqrt{\rho}} \sqrt{n},$$

and

$$\mathbb{E}(H_n^r) \sim \left(\frac{2}{b\sqrt{\rho}} \right)^r r(r-1)\Gamma(\frac{r}{2})\zeta(r) n^{\frac{r}{2}}$$

for every integer $r \geq 2$, as n tends to infinity.

The proof of this theorem is in close relation to the proof of the following statement on the profile process (cf [21]):

Theorem 2.7 (Drmota, Gittenberger, 2010). *Let $l_n(t) = \frac{1}{\sqrt{n}} L_n(t\sqrt{n})$, and $l(t)$ denote the local time of a standard Brownian excursion. Then $l_n(t)$ converges weakly to the local time of a Brownian excursion, i.e., we have*

$$(l_n(t))_{t\geq 0} \xrightarrow{w} \frac{b\sqrt{\rho}}{2\sqrt{2}} \cdot l\left(\frac{b\sqrt{\rho}}{2\sqrt{2}}t\right)_{t\geq 0},$$

where b and ρ are the constants in (2.3).

Similar statements have been proven in [19] for simply generated trees, showing that the height is of asymptotic order $\sqrt{n}$ and the profile converges weakly to a suitably normalized Brownian local time. Note that the scaling factor for the profile process in the above theorem is $\sqrt{n}$, which is due to the fact that the expected height is of order $\sqrt{n}$, and thus no vertices are to be expected at deeper levels.

The above theorem on the profile holds only for levels on height $\kappa\sqrt{n}$. The bahaviour of the profile close to the root has been studied for trees and forests in [33, 35].

2.2 The fringe of trees

In this section we study properties of the fringe of Pólya trees and compare them to results on simply generated trees presented by Drmota *et. al.* in [22]. Let ν be a leaf in a random tree and ζ be its parent node. We call the tree rooted at ζ the Ward tree of the leaf ν, and the set of all Ward trees of a tree T the fringe of T. The name of this parameter origins in the first study of this parameter in tries and suffix tries by Ward in his Ph.D. Thesis [69]. In [22] the size of a Ward tree is studied for several graph classes, including simply generated trees, but not for Pólya trees. We complement this survey with this result. Note that the definition of the fringe we use here differs from the one used in papers on the fringe analysis of search trees, cf e.g. [72, 1].

2.2.1 The Ward-parameter of Pólya trees

Let $\mathcal{T}$ be a family of trees, and $T \in \mathcal{T}$. Further, let ν be a leaf of T. The Ward-parameter $w(\nu)$ counts the number of internal nodes of the Ward tree of ν (alternatively we could count the number of leaves in the Ward tree, $\bar{w}(\nu)$, or the total size of the Ward tree, $\tilde{w}(\nu) = w(\nu) + \bar{w}(\nu)$). We define $W(T) = 0$ for a tree being a single root, $T = \bullet$. As Pólya trees are unlabelled, we cannot distinguish between leaves. As the method of generating functions does not allow us to consider a random leaf, we consider the cumulative Ward parameter $W(T)$, summing up the Ward parameter of all leaves $W(T) = \sum_{\nu \text{leafof} T} w(\nu)$. We denote by W_n the random variable counting the cumulative Ward parameter of a random tree of size n.
Therefore, we define the generating fuction

$$G(z,u,v) = \sum_{n\geq 1}\sum_{\ell\geq 1}\sum_{r\geq 1} y_{n\ell r} z^n u^\ell v^r,$$

where $y_{n\ell r}$ counts the number of trees T with n nodes and ℓ internal nodes with cumulative Ward-parameter equal to $W(T) = r$.

Lemma 2.8. *The generating function $G(z, u, v)$ fulfils the equation*

$$G(z,u,v) = z\left(\sum_{m\geq 0} uz^m v^m \exp\left(\sum_{k\geq 1}\frac{G(z^k,(uv^m)^k,v^k)-z^k}{k}\right)+(1-u)\right),$$

where $Z_{k-m}(\mathbf{G(z,uv^m,v)-z})$ denotes the substitution described on page 8 in Chapter 1.

Proof. Remember that the generating function of Pólya trees is given by a node together with a set of subtrees rooted at that node:

$$y(z) = z(\sum_{k\geq 0} Z_k(y(z), y(z^2), \ldots, y(z^k)) = z\exp\left(\sum_{k\geq 1}\frac{y(z^k)}{k}\right)$$

If a tree T consists of one node only, this node is a leaf and the tree has no internal nodes, just as the Ward-tree of the leaf. Otherwise, k rooted subtrees are attached to the root of the tree. Of those k subtrees, m consist only of their roots, which are therefore leaves, while the remaining $k-m$ subtrees are "real" trees, that are trees which are more than a single vertex z, where $0 \leq m \leq k$. For each of the m leaves, the root and every internal node of the tree T contributes to the Ward parameter of this node, and thus all internal vertices contribute m times to the cumulative Ward parameter $W(T)$. Thus

$$G(z,u,v) = z\left(1+\sum_{k\geq 1}\sum_{m=0}^{k} z^m v^m u Z_{k-m}(\mathbf{G(z,uv^m,v)-z})\right), \tag{2.6}$$

Note that

$$\begin{aligned}\sum_{k\geq 0}\sum_{m=0}^{k} f(m)g(k-m,m) &= \sum_{k\geq 0}\sum_{m\geq 0}\mathbf{1}_{[0,k]}(m)f(m)g(k-m,m)\\ &= \sum_{m\geq 0} f(m)\sum_{k\geq 0}\mathbf{1}_{[0,k]}(m)g(k-m,m)\\ &= \sum_{m\geq 0} f(m)\sum_{k\geq m} g(k-m,m) = \sum_{m\geq 0} f(m)\sum_{\ell\geq 0} g(\ell,m)\end{aligned}$$

and hence (2.6) rewrites to the desired equation with the help of Lemma 1.6. ☐

Remark. Of course, setting the variables u and v to 1 should give the ordinary generating function of Pólya trees, $G(z,1,1) = y(z)$. This is easily proven:

$$\begin{aligned} G(z,1,1) &= z\left(\sum_{m\geq 0} z^m \exp\left(\sum_{k\geq 1}\frac{G(z^k,1,1)-z^k}{k}\right)\right) \\ &= z\left(\sum_{m\geq 0} z^m\right)\exp\left(\sum_{k\geq 1}\frac{G(z^k,1,1)}{k}\right)\exp\left(-\sum_{k\geq 1}\frac{z^k}{k}\right) \\ &= z\frac{1}{1-z}\exp\left(\sum_{k\geq 1}\frac{G(z^k,1,1)}{k}\right)\exp(\log(1-z)) \\ &= z\exp\left(\sum_{k\geq 1}\frac{G(z^k,1,1)}{k}\right), \end{aligned}$$

which is the equation for the ordinary generating function $y(z)$ of Polya trees, given in Equation (2.2).

To obtain results on the expected value $\mathbb{E}(W_n)$ we need to determine

$$\mathbb{E}(X_n) = [z^n]G_v(z,1,1)\frac{1}{y_n},$$

where $G_v(z,u,v)$ denotes the derivative with respect to v of $G(z,u,v)$ and y_n is the number of trees of size n, because

$$\begin{aligned} [z^n]G_v(z,1,1)\frac{1}{y_n} &= [z^n](\sum_{n,m,\ell}\ell y_{nm\ell}z^n)\frac{1}{y_n} = [z^n]\sum_{n,m,\ell}\ell\frac{y_{nm\ell}}{y_n}z^n \\ &= [z^n]\sum_{n,\ell}\ell\mathbb{P}(W_n=\ell)z^n = \sum_{\ell}\ell\mathbb{P}(W_n=\ell) = \mathbb{E}(W_n). \end{aligned}$$

We first set $u = 1$,

$$G(z,1,v) = z\left(\sum_{m\geq 0} z^m v^m \exp\left(\sum_{k\geq 1}\frac{G(z^k,v^{mk},v^k)-z^k}{k}\right)\right),$$

and then derivate with respect to v:

$$\begin{aligned} \frac{\partial}{\partial v}G(z,1,v) = z\Bigg[&\sum_{m\geq 0} m z^m v^{m-1}\exp\left(\sum_{k\geq 1}\frac{G(z^k,v^{mk},v^k)-z^k}{k}\right) \\ &+\sum_{m\geq 0} z^m v^m\exp\left(\sum_{k\geq 1}\frac{G(z^k,v^{mk},v^k)-z^k}{k}\right) \\ &\times\sum_{k\geq 1}\Big(mv^{mk-1}G_u(z^k,v^{mk},v^k)+v^{k-1}G_v(z^k,v^{mk},v^k)\Big)\Bigg]. \end{aligned}$$

Setting $v = 1$ we get

$$\begin{aligned}
G_v(z,1,1) &= z \exp\left(\sum_{k\geq 1} \frac{y(z^k) - z^k}{k}\right) \\
&\quad \times \left[\left(\sum_{m\geq 0} m z^m\right) + \sum_{m\geq 0} z^m \sum_{k\geq 1} \left(m G_u(z^k,1,1) + G_v(z^k,1,1)\right)\right] \\
&= (1-z)y(z)\left[z\left(\frac{1}{1-z}\right)'\left(1 + \sum_{k\geq 1} G_u(z^k,1,1)\right) + \frac{1}{1-z}\sum_{k\geq 1} G_v(z^k,1,1)\right] \\
&= y(z)\left[\frac{z}{1-z}\left(1 + \sum_{k\geq 1} G_u(z^k,1,1)\right) + \sum_{k\geq 1} G_v(z^k,1,1)\right].
\end{aligned}$$

Rewriting this equation gives

$$\begin{aligned}
G_v(z,1,1)(1-y(z)) &= y(z)\left[\frac{z}{1-z}\left(1 + \sum_{k\geq 1} G_u(z^k,1,1)\right) + \sum_{k\geq 2} G_v(z^k,1,1)\right] \\
G_v(z,1,1) &= \frac{y(z)}{1-y(z)}\left[\frac{z}{1-z}\left(1 + G_u(z,1,1)\right) + \underbrace{\sum_{k\geq 2}\left(\frac{z}{1-z} G_u(z^k,1,1) + G_v(z^k,1,1)\right)}_{=:A(z)}\right],
\end{aligned}$$

where $A(z)$ is an analytic function near ρ, which follows immediately from the following Lemma.

Lemma 2.9. *$G_u(z,1,1)$ and $G_v(z,1,1)$ have a dominant singularity at ρ, where ρ is the singularity of $y(z)$.*

Proof. We know that $G(z,1,1) = y(z)$ has a singularity at ρ. As $G(z,u,v) = \sum y_{nm\ell} z^n u^m v^\ell$ has only positive coefficients, it is analytic at $(\rho - \varepsilon_1, 1-\varepsilon_2, 1-\varepsilon_3)$. Now $G_u(z,1,1) = \sum m y_{nm\ell} z^n$ and $G_v(z,1,1) = \sum \ell y_{nm\ell} z^n$ can only have radius of convergence smaller or equal to ρ, but as $G(z,u,v)$ is differentiable infinitely often at $(\rho-\varepsilon_1, 1-\varepsilon_2, 1-\varepsilon_3)$, $G_u(z,u,v)$ and $G_v(z,u,v)$ are as well. Thus ρ has to be the dominant singularity. ☐

We need to determine $G_u(z,1,1)$. Note therefore that

$$\begin{aligned}
G_u(z,1,1) &= \frac{\partial}{\partial u} G(z,u,1)|_{u=1} = \frac{\partial}{\partial u} P^{(1)}(zu, \frac{1}{u})|_{u=1} \\
&= z P_z^{(1)}(z,1) - P_u^{(1)}(z,1)
\end{aligned}$$

where $P^{(1)}(z,u)$ is the generating function of planted[1] Polya trees, where u counts vertices of degree 1, i.e. leaves, cf e.g. [15]. It is given by

$$P^{(1)}(z,u) = z \exp\left(\sum_{k\geq 1} \frac{P^{(1)}(z^k,u^k)}{k}\right) + z(u-1).$$

[1]A planted tree is a rooted tree where we assume that the root node is adjacent to an additional node which is not counted. This assumption does not alter the tree structure, but allows us to treat the root vertex like a normal vertex, that is, a root of degree d has in-degree 1 and out-degree $d-1$.

Derivation gives

$$
\begin{aligned}
P_z^{(1)}(z,1) = y'(z) &= \exp\left(\sum_{k\geq 1}\frac{y(z^k)}{k}\right) + z\exp\left(\sum_{k\geq 1}\frac{y(z^k)}{k}\right)\sum_{k\geq 1} y'(z^k)z^{k-1}\\
y'(z)(1-y(z)) &= \frac{y(z)}{z} + y(z)\sum_{k\geq 2} y'(z^k)z^{k-1}\\
P_u^{(1)}(z,1) &= z\exp\left(\sum_{k\geq 1}\frac{y(z^k)}{k}\right)\sum_{k\geq 1} P_u^{(1)}(z^k,1) + z\\
P_u^{(1)}(z,1)(1-y(z)) &= y(z)\sum_{k\geq 2} P_u^{(1)}(z^k,1) + z,
\end{aligned}
$$

and thus

$$
G_u(z,1,1) = \frac{1}{1-y(z)}\left(y(z)\left(1 + z\sum_{k\geq 2} y'(z^k)z^{k-1} - \sum_{k\geq 2} P_u^{(1)}(z^k,1)\right) - z\right)
$$

We obtain

$$
\begin{aligned}
&G_v(z,1,1)\\
&= \frac{y(z)}{1-y(z)}\left[\frac{z}{1-z}\left(1+\frac{y(z)\left(1+z\sum_{k\geq 2} y'(z^k)z^{k-1} - \sum_{k\geq 2} P_u^{(1)}(z^k,1)\right) - z}{1-y(z)}\right) + A(z)\right]
\end{aligned}
$$

Proposition 2.10. *Let W_n be the random variable that counts the size of the cumulative Ward-parameter in random Pólya trees of size n. Then the expected value of W_n is asymptotically, as n tends to infinity, given by*

$$
\mathbb{E}(X_n) \sim \frac{\sqrt{\pi}}{b^3\sqrt{\rho}(1-\rho)}\left(\rho b^2 - 2(\sum_{k\geq 2} P_u^{(1)}(\rho^k,1) + \rho)\right) n^{\frac{3}{2}} =: Bn^{\frac{3}{2}}, \tag{2.7}
$$

with $B \approx 0.31838978$ being numerically computable.

Proof. To calculate $[z^n]G_v(z,1,1)$ we will use the singular expansion (2.3) of $y(z)$ near its singularity, which by Lemma 2.9 is the same singularity as for $G_v(z,1,1)$, and apply the transfer lemma (Lemma 1.7). Near $z=\rho$ we have

$$
\begin{aligned}
z &\sim \rho\\
y(z) &\sim 1\\
1-y(z) &\sim b\sqrt{\rho}\sqrt{1-\frac{z}{\rho}}
\end{aligned}
$$

Derivating equation (2.2) gives

$$
y'(z) = \exp\left(\sum_{k\geq 1}\frac{y(z^k)}{k}\right) + z\exp\left(\sum_{k\geq 1}\frac{y(z^k)}{k}\right)\sum_{k\geq 1} y'(z^k)z^{k-1}.
$$

Using the singular expansion (2.3) of $y(z)$ near ρ we obtain

$$\lim_{z\to\rho_-} y'(z)(1-y(z)) = \lim_{z\to\rho_-} \frac{y(z)}{z} + y(z)\sum_{k\geq 2} y'(z^k)z^{k-1} = \frac{b^2}{2},$$

which implies that

$$\lim_{z\to\rho_-} y(z)(1+z\sum_{k\geq 2} y'(z^k)z^{k-1}) = \frac{\rho b^2}{2}$$

Expanding near the singularity, we obtain

$$G_v(z,1,1) \sim \frac{1}{b\sqrt{\rho}\sqrt{1-\frac{z}{\rho}}}\left[\frac{\rho}{1-\rho}\left(1+\frac{\rho b^2 - 2(\sum_{k\geq 2} P_u^{(1)}(\rho^k,1)+\rho)}{2b\sqrt{\rho}\sqrt{1-\frac{z}{\rho}}}\right) + A(\rho)\right]$$

And hence, extracting the coefficient $[z^n]$ gives

$$[z^n]G_v(z,1,1) \sim \frac{\rho}{1-\rho}\frac{\rho b^2 - 2(\sum_{k\geq 2} P_u^{(1)}(\rho^k,1)+\rho)}{2b^2\rho}\rho^{-n}\left(1+\mathcal{O}(\frac{1}{n})\right).$$

Dividing by the well known estimate $y_n \sim \frac{b\sqrt{\rho}}{2\sqrt{\pi}}n^{-\frac{3}{2}}\rho^{-n}$, the result follows. □

Note that obviously $\rho b^2 > 2(\sum_{k\geq 2} P_u^{(1)}(\rho^k,1)+\rho)$, as the expected number of leaves in a tree, $\mu_1 n$, is given by

$$\mu_1 n \sim \frac{2(\sum_{k\geq 2} P_u^{(1)}(\rho^k,1)+\rho)}{\rho b^2} n \leq n,$$

which immediately implies that the expected size of the Ward tree of a random leaf is asymptotically of order $\sqrt{n}$.

2.2.2 Comparison with simply generated trees

We want to compare the value we just obtained with the one for simply generated trees. In [22] the Ward parameter has been studied for specific leaves with number j in their left-to-right order. We have to modify this result to be able to compare it with our cumulative result.

Let $F(z,u,v) = \sum f_{nmj} z^n u^j v^m$ be the generating function where the coefficient f_{nmj} counts simply generated trees of size n where the j-th leaf has Ward-parameter m. In [22] it is shown that this generating function fulfils

$$F(z,u,v) = \phi_o zu + \frac{\phi_0 z^2 uv \frac{\phi(L(zv,\frac{1}{v}))-\phi(L(zv,\frac{u}{v}))}{L(zv,\frac{1}{v})-L(zv,\frac{u}{v})}}{1 - z\frac{\phi(T(z))-\phi(L(z,u))}{T(z)-L(z,u)}},$$

where $L(z,u)$ is the generating function counting leaves and $\phi(x)$ is the generating function with non-negative coefficients from the relation

$$T(z) = z\phi(T(z)). \tag{2.8}$$

To sum up over all leaves, we set $u = 1$. In the above equation, two differential quotients appear when u tends to 1.

$$F(z,1,v) = \phi_0 + \frac{\phi_0 z^2 v \phi'(L(zv, \frac{1}{v}))}{1 - z\phi'(T(z))}$$

From equation (2.8) we get

$$\begin{aligned} T'(z) &= \phi(T(z)) + z\phi'(T(z))T'(z) \\ T'(z) &= \frac{\phi(T(z))}{1 - z\phi'(T(z))} \\ \frac{\phi(T(z))}{T'(z)} &= 1 - z\phi'(T(z)) \\ \phi'(T(z)) &= \frac{1}{z} - \frac{\phi(T(z))}{zT'(z)} = \frac{1}{z} - \frac{T(z)}{z^2 T'(z)} \end{aligned}$$

We derive $F(z, 1, v)$ with respect to v and obtain

$$\frac{\partial}{\partial v} F(z,1,v) = \frac{\phi_0 z^3 T'(z)}{T(z)} \left(\phi'(L(zv, \frac{1}{v})) + v\phi''(L(zv, \frac{1}{v})) \left(zL_z(zv, \frac{1}{v}) - \frac{1}{v}^2 L_u(zv, \frac{1}{v}) \right) \right)$$

Hence, setting $v = 1$, gives

$$F_v(z,1,1) = \frac{\phi_0 z^3 T'(z)}{T(z)} \left(\phi'(T(z)) + \phi''(T(z))((zL_z(z,1) - L_u(z,1)) \right).$$

The leaf counting function is given by

$$L(z,u) = \phi_0 z(u-1) + z\phi(L(z,u)),$$

hence we obtain

$$\begin{aligned} \frac{\partial}{\partial z} L(z,u) &= \phi_0(u-1) + \phi(L(z,u)) + z\phi'(L(z,u)) \frac{\partial}{\partial z} L(z,u) \\ \frac{\partial}{\partial u} L(z,u) &= \phi_0 z + z\phi'(L(z,u)) \frac{\partial}{\partial u} L(z,u) \\ zL(z,1) - L_u(z,1) &= \frac{T(z) - \phi_0 z}{1 - z\phi'(T(z))} = \frac{zT'(z)(T(z) - \phi_0 z)}{T(z)}. \end{aligned}$$

This finally gives

$$F_v(z,1,1) = \frac{\phi_0 z^3 T'(z)}{T(z)} \left(\phi'(T(z)) + \phi''(T(z)) \frac{zT'(z)(T(z) - \phi_0 z)}{T(z)} \right)$$

We know that near the singularity $z_0 = \frac{1}{\phi'(\tau)}$, the generating function has a singular expansion $T(z) \sim \tau - b\sqrt{1 - z\phi'(\tau)}$. This implies

$$T(z) \sim \tau - \sqrt{\frac{2\phi(\tau)}{\phi''(\tau)}} \left(1 - z\phi'(\tau)\right)^{\frac{1}{2}} \tag{2.9}$$

$$T'(z) \sim \frac{\phi'(\tau)}{2} \sqrt{\frac{2\phi(\tau)}{\phi''(\tau)}} \left(1 - z\phi'(\tau)\right)^{-\frac{1}{2}} \tag{2.10}$$

$$T''(z) \sim \frac{\phi'(\tau)^2}{4} \sqrt{\frac{2\phi(\tau)}{\phi''(\tau)}} \left(1 - z\phi'(\tau)\right)^{-\frac{3}{2}} \tag{2.11}$$

Further, we need an expression for $\phi''(T(z))$ from (2.8).

$$\phi''(T(z)) = \frac{T(z)T''(z)}{z^2T'(z)^3} - \frac{2}{zT'(z)}\phi'(T(z))$$

Putting everything together, we obtain a singular expansion

$$F_v(z,1,1) \sim \frac{\phi_0}{2\tau\phi'(\tau)}\left(\tau - \frac{\phi_0}{\phi'(\tau)}\right)(1-z\phi'(\tau))^{-1}\left(1+\mathcal{O}\left(1-z\phi'(\tau)\right)^{\frac{1}{2}}\right),$$

which gives an asymptotic coefficient

$$[z^n]F_v(z,1,1) \sim \frac{\phi_0}{2\tau\phi'(\tau)}\left(\tau - \frac{\phi_0}{\phi'(\tau)}\right)\phi'(\tau)^n\left(1+\mathcal{O}\left(\frac{1}{\sqrt{n}}\right)\right).$$

Hence, using (2.5), we obtain

$$\mathbb{E}(W_n) \sim \frac{\phi_0}{2\sqrt{2\pi\phi(\tau)\phi''(\tau)}}\left(\tau - \frac{\phi_0}{\phi'(\tau)}\right)n^{\frac{3}{2}}$$

for the expected value of the cumulative Ward parameter of a random simply generated tree of size n.

We now choose a simply generated family which is very similar to the family of Pólya trees, i.e. labelled nonplane trees, with $\Phi(t) = e^t$. We have the system

$$y = ze^y \qquad 1 = ze^y,$$

hence, for this family, $\tau = 1, \Phi(\tau) = \Phi'(\tau) = \Phi''(\tau) = e$ and $\phi_0 = 1$. We obtain

$$\mathbb{E}(W_n) \sim 0.04638584832n^{\frac{3}{2}}.$$

Recall that in the Pólya case, we obtained $\mathbb{E}(W_n) \sim 0.31838978n^{\frac{3}{2}}$ for the expected value of the cumulative Ward parameter of a Pólya tree of size n (Proposition 2.10). Hence the asymptotic behaviour of the parameters is equivalent, although the constant in the example of a simply generated tree family is significantly smaller.

2.3 The degree profile

As mentioned before (cf Theorem 2.7), Drmota and Gittenberger [21] showed that the profile of random Pólya trees converges weakly to a normalized local time of a standard brownian excursion. In this chapter, we will refine this result and examine the degree profile of random Pólya trees, that is, the number of vertices of given degree d on a level k. The degree profile of simply generated trees has been studied in [14].
We will terminate this section by a study of the correlation of two different degrees on a given level of a Pólya tree. This problem has been adressed by Hofstad et al in [68] and by Gittenberger and Louchard in [38] for simply generated trees.

Definition 2.11. *We define by $L_n^{(d)}(k)$ the number of nodes of degree d at distance k from the root in a randomly chosen unlabelled rooted tree of size n. Obviously, $(L_n^{(d)}(k))_{k\geq 0}$ is another sequence of random variables. By linear interpolation we create a continuous stochastic process, which we call the degree profile process $L_n^{(d)}(t)$ of Pólya trees of size n with respect to degree d.*

$$L_n^{(d)}(t) = (\lfloor t\rfloor + 1 - t)L_n^{(d)}(\lfloor t\rfloor) + (t - \lfloor t\rfloor)L_n^{(d)}(\lfloor t\rfloor + 1), \quad t \geq 0$$

Example. In Figure 2.4 we show an example for the 3-profile of a tree.

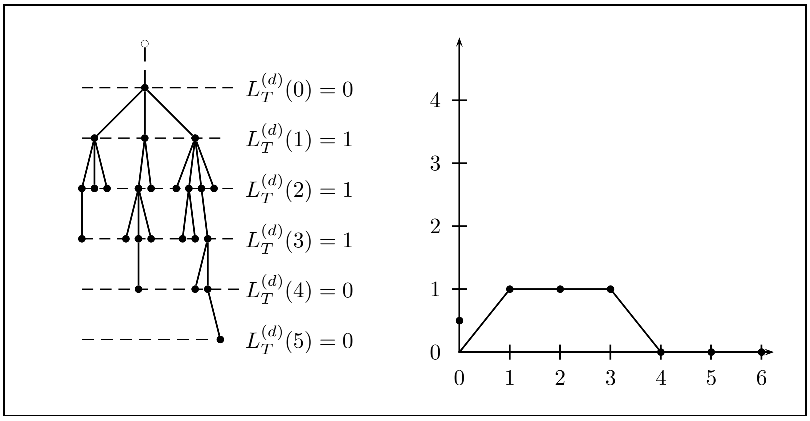

Figure 2.4: The 3-profile of a tree.

As mentioned before, a vertex of given degree corresponds to a star graph, hence defining the degree profile is a first step on the way to examining the profile of arbitrary patterns, that is the number and location of patterns in trees, see Figure 2.5.

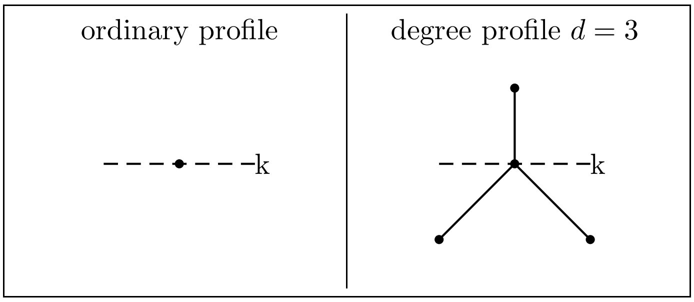

Figure 2.5: The degree profile represents star graphs located at level k of a tree.

In close relation to Theorem 2.7, we prove the following main result:

Theorem 2.12. *Let*

$$l_n^{(d)}(t) = \frac{1}{\sqrt{n}} L_n^{(d)}(t\sqrt{n})$$

and $l(t)$ denote the local time of a standard Brownian excursion. Then $l_n^{(d)}(t)$ converges weakly to the local time of a Brownian excursion, i.e., we have

$$(l_n^{(d)}(t))_{t\geq 0} \xrightarrow{w} \frac{C_d \rho^d}{\sqrt{2\rho b}} \cdot l\left(\frac{b\sqrt{\rho}}{2\sqrt{2}} t\right)_{t\geq 0}, \tag{2.12}$$

where $C_d = C + \mathcal{O}\left(d\rho^d\right)$ with $C = \exp\left(\sum_{i\geq 1} \frac{1}{i}\left(\frac{y(\rho^i)}{\rho^i} - 1\right)\right) \approx 7.7581604\cdots$.

Remark. Recall Theorem 2.7: It has been shown that the general profile of an unlabelled rooted random tree converges to Brownian excursion local time with

$$(l_n(t))_{t\geq 0} \xrightarrow{w} \left(\frac{b\sqrt{\rho}}{2\sqrt{2}} l(\frac{b\sqrt{\rho}}{2\sqrt{2}}t)\right)_{t\geq 0}.$$

The normalising constant in Theorem 2.12 equals $\mu_d \frac{b\sqrt{\rho}}{2\sqrt{2}}$, where $\mu_d n$ is asymptotically equal to the expected value of nodes of degree d in trees of size n, with $\mu_d = \frac{2C_d}{b^2\rho}\rho^d$, as given in Theorem 2.3.

To proof the above statement, weak convergence of the finite dimensional distributions and tightness have to be shown:

Theorem 2.13. *For any choice of fixed numbers $t_1, \ldots, t_m$ and for large d*

$$(l_n^{(d)}(t_1), \ldots, l_n^{(d)}(t_m)) \xrightarrow{w} \frac{C_d\rho^d}{\sqrt{2}\rho b} l\left(\frac{b\sqrt{\rho}}{2\sqrt{2}}t_1, \ldots, \frac{b\sqrt{\rho}}{2\sqrt{2}}t_m\right)$$

as $n \to \infty$.

Remark. We will show this theorem by proving the convergence of the corresponding characteristic functions. It is well known (cf. [13]) that the characteristic function of $\frac{C_d\rho^d}{\sqrt{2}\rho b} l\left(\frac{b\sqrt{\rho}}{2\sqrt{2}}t\right)$ is

$$\psi(t) = 1 + \frac{C_d\rho^d}{ib\sqrt{\rho\pi}} \int_\gamma \frac{t\sqrt{-x}\exp\left(-\frac{\kappa b}{2\sqrt{-\rho x}} - x\right)}{\sqrt{-x}\exp\left(\frac{\kappa b}{2\sqrt{-\rho x}}\right) - \frac{C_d\rho^d t}{b\sqrt{\rho}}\sinh\left(\frac{\kappa b}{2\sqrt{-\rho x}}\right)}\, dx \tag{2.13}$$

where γ is a contour going from $+\infty$ back to $+\infty$ while encircling the origin clockwise.

A sequence of stochastic processes might not converge even if the sequence of their images with respect to every finite-dimensional projection does. Roughly speaking, in order to guarantee convergence in the sense of stochastic processes (i.e., when constructing a sequence by applying an arbitrary continuous bounded functional to the corresponding probability measures, this sequence must converge) the sample paths of the processes must not fluctuate too wildly. Tightness is a technical property of stochastic processes which guarantees this. The next theorem states a technical condition for the profile process which implies tightness (cf. [4] and [45] for the general theory.)

Theorem 2.14. *There exists a constant $C > 0$ such that for all integers r, h, n the inequality*

$$\mathbb{E}\left(L_n(r) - L_n(r+h)\right)^4 \leq C\, h^2 n \tag{2.14}$$

holds.

Remark. According to [4, Theorem 12.3] the inequality

$$\mathbf{E}\left|L_n(r) - L_n(r+h)\right|^\alpha = \mathcal{O}\left(h^\beta(\sqrt{n})^{\alpha-\beta}\right)$$

implies tightness of the process $l_n(t)$ if $\alpha > 0$ and $\beta > 1$. In the theorem above we have $\alpha = 4$ and $\beta = 2$ and thus $l_n(t)$ is tight.

Proving Theorem 2.13, we will start with the one-dimensional case and then extend results to multiple dimensions. Therefore, we introduce generating functions $y_k^{(d)}(z,v)$, which represent trees where all nodes of degree d on level k are marked and counted by v. Note that we consider planted trees instead of 'ordinary' rooted trees, which allows us to treat the root vertex like a normal vertex, that is, a root of degree d has in-degree 1 and out-degree $d-1$.

Refining the decomposition of trees along their root, the $y_k^{(d)}(z,v)$ can be defined recursively:

$$\begin{aligned} y_0^{(d)}(z,v) &= y(z) + (v-1)zZ_{d-1}(\mathbf{y}(\mathbf{z})) \\ y_{k+1}^{(d)}(z,v) &= z\exp\left(\sum_{i\geq 1} \frac{y_k^{(d)}(z^i,v^i)}{i}\right), \end{aligned} \tag{2.15}$$

where $Z_d(s_1,s_2,\ldots,s_d)$ is the cycle index of the permutation group $\mathfrak{S}_d$ on d elements and $Z_{d-1}(\mathbf{y_h^{(d)}}(\mathbf{z},\mathbf{v}))$ denotes the substitution described on page 8 in Chapter 1.

Examining two levels $k, k+h$ at once, we use the generating function $y_{k,h}^{(d)}(z,v_1,v_2)$ where all nodes of degree d on level k are marked by v_1 and nodes of degree d on level $k+h$ are marked by v_2. We get the recursive relation

$$\begin{aligned} y_{0,h}^{(d)}(z,v_1,v_2) &= y_h^{(d)}(z,v_2) + (v_1-1)zZ_{d-1}(\mathbf{y_h^{(d)}}(\mathbf{z},\mathbf{v_2})) \\ y_{k+1,h}^{(d)}(z,v_1,v_2) &= z\exp\left(\sum_{i\geq 1} \frac{y_{k,h}^{(d)}(z^i,v_1^i,v_2^i)}{i}\right). \end{aligned} \tag{2.16}$$

In general, observing levels k_1,$k_2 = k_1+h_1$,..., $k_m = k_{m-1}+h_{m-1}$, we get:

$$\begin{aligned} &y_{0,h_1,\ldots,h_{m-1}}^{(d)}(z,v_1,\ldots,v_m) = \\ &\qquad y_{h_1,\ldots h_{m-1}}^{(d)}(z,v_2,\ldots,v_m) + (v_1-1)zZ_{d-1}(\mathbf{y^{(d)}_{h_1,\ldots h_{m-1}}}(\mathbf{z},\mathbf{v_2},\ldots,\mathbf{v_m})) \\ &y_{k+1,h_1,\ldots,h_{m-1}}^{(d)}(z,v_1,\ldots,v_m) = z\exp\left(\sum_{i\geq 1} \frac{y_{k,h_1,\ldots,h_{m-1}}^{(d)}(z^i,v_1^i,\ldots,v_m^i)}{i}\right) \end{aligned}$$

These functions are related to the process $L_n^{(d)}(t)$ by

$$\begin{aligned} &\mathbb{P}(L_n^{(d)}(k)=\ell_1, L_n^{(d)}(k+h_1)=\ell_2,\ldots,L_n^{(d)}(k+\sum h_i)=\ell_m) \\ &\qquad = \frac{[z^n v_1^{\ell_1} v_2^{\ell_2}\cdots v_m^{\ell_m}]y_{k,h_1,\ldots,h_{m-1}}^{(d)}(z,v_1,\ldots,v_m)}{[z^n]y(z)} \end{aligned}$$

We will give a detailed analysis of the functions $y_k(z,v)$ in Section 2.3.1, providing us with bounds and asymptotic expressions of them near their singularity. The lemmas given in this section will allow us to find a closed expression for the characteristic function of the one-dimensional profile process. In Section 2.3.2 we will extend the bounds computed in Section 2.3.1 to the multidimensional case. The proof of Theorem 2.14 is given in Section 2.3.3.

For this study, we introduce a Δ-domain, cf (1.5), and some other domains

$$\Delta = \Delta(\eta,\theta) = \{z \in \mathbb{C} \big| |z| < \rho + \eta, |\arg(z-\rho)| > \theta\}, \tag{2.17}$$
$$\Delta_\epsilon = \Delta_\epsilon(\theta) = \{z \in \mathbb{C} \big| |z-\rho| < \epsilon, |\arg(z-\rho)| > \theta\}, \tag{2.18}$$
$$\Theta = \Theta(\eta) = \{z \in \mathbb{C} \big| |z| < \rho + \eta, |\arg(z-\rho)| \neq 0\}, \tag{2.19}$$
$$\Xi_k = \Xi_k(\tilde{\eta}) = \{z \in \mathbb{C} \big| |v| \leq 1, k|v-1| \leq \tilde{\eta}\}, \tag{2.20}$$

with $\epsilon, \eta, \tilde{\eta} > 0$ and $0 < \theta < \frac{\pi}{2}$.
In all the proofs in the subsequent sections we will assume (even without explicitely mentioning) that η, θ, ϵ are sufficiently small for all arguments to be valid.

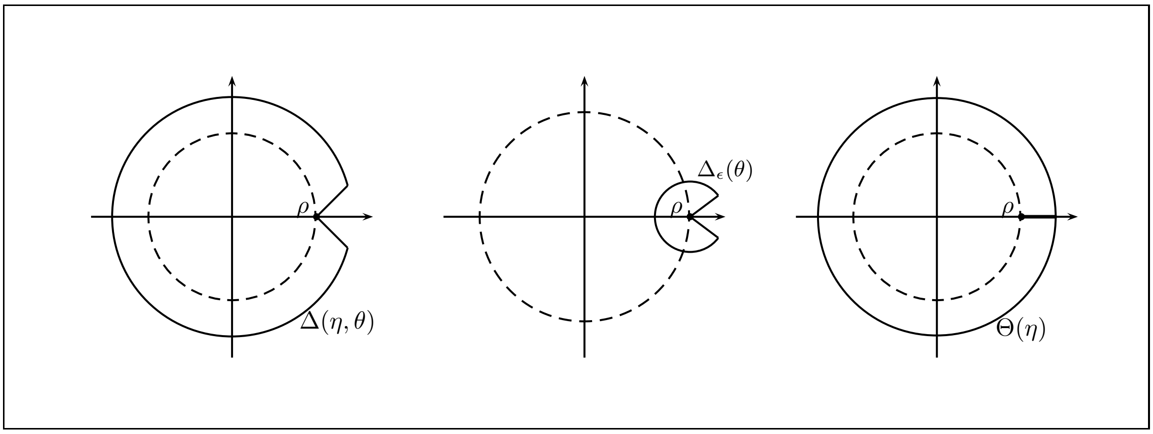

Figure 2.6: The regions used for the proofs

2.3.1 The one dimensional case

In the following, we will use the notations

$$w_k^{(d)}(z,v) = y_k^{(d)}(z,v) - y(z)$$
$$\Sigma_k^{(d)}(z,v) = \sum_{i\geq 2} \frac{w_k^{(d)}(z^i, v^i)}{i}$$
$$\gamma_k^{(d)}(z,v) = \frac{\partial}{\partial v} y_k^{(d)}(z,v)$$
$$\gamma_k^{(d)[2]}(z,v) = \frac{\partial^2}{\partial v^2} y_k^{(d)}(z,v)$$

Our main goal is to prove the following theorem, from where the main result follows by integration.

Theorem 2.15. *Let $z = \rho(1+\frac{s}{n})$, $v = e^{\frac{it}{\sqrt{n}}}$, $k = \lfloor \kappa\sqrt{n} \rfloor$ and d be a fixed integer. Moreover, assume that $|\arg s| \geq \vartheta > 0$ and, as $n \to \infty$, we have $s = \mathcal{O}(\log^2 n)$, whereas κ and t are fixed. Then, $w_k^{(d)}(z,v)$ admits the local representation*

$$w_k^{(d)}(z,v) \sim \frac{C_d \rho^d}{\sqrt{n}} \cdot \frac{it\sqrt{-s}e^{-\frac{1}{2}\kappa b\sqrt{-\rho s}}}{\sqrt{-s}e^{\frac{1}{2}\kappa b\sqrt{-\rho s}} - \frac{itC\rho^d}{b\sqrt{\rho}} \sinh(\frac{1}{2}\kappa b\sqrt{-\rho s})} \tag{2.21}$$

The one-dimensional limiting distribution

Let us first assume that Theorem 2.15 holds. Then, to prove Theorem 2.13 in one dimension, we need to determine the characteristic function

$$\begin{aligned}\phi_{k,n}^{(d)}(t) &= \frac{1}{y_n}[z^n]y_k^{(d)}(z, e^{\frac{it}{\sqrt{n}}}) \\ &= \frac{1}{2\pi i y_n}\int_\Gamma y_k^{(d)}(z, e^{\frac{it}{\sqrt{n}}})\frac{dz}{z^{n+1}} \end{aligned} \tag{2.22}$$

where the contour $\Gamma = \gamma \cup \Gamma'$ consists of the line

$$\gamma = \{z = \rho(1 - \frac{1 + i\tau}{n})| - D\log^2 n \le \tau \le D\log^2 n\}$$

with an arbitrarily chosen constant $D > 0$ and Γ' is a circular arc centered at the origin and closing the curve, see Figure 2.7. The contribution of Γ' is exponentially small since for $z \in \Gamma'$, $\frac{1}{y_n}|z^{-(n+1)}| = \mathcal{O}(n^{\frac{3}{2}}e^{-\log^2 n})$ whereas $|y_k^{(d)}(z, e^{\frac{it}{\sqrt{n}}})|$ is bounded.

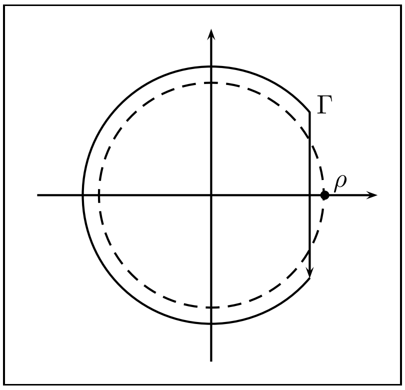

Figure 2.7: The contour Γ

If $z \in \gamma$ the local expansion (2.21) is valid and thus, inserting into (2.22) leads to:

$$\begin{aligned}\lim_{n\to\infty}\phi_{k,n}^{(d)}(t) &= \lim_{n\to\infty}\frac{1}{2\pi i y_n}\Bigg[\int_{\Gamma'} w_k^{(d)}(z,v)\frac{dz}{z^{n+1}} + \underbrace{\int_{\Gamma'} y(z)\frac{dz}{z^{n+1}}}_{=2\pi i y_n}\Bigg] \\ &= 1 + \lim_{n\to\infty}\frac{C_d\rho^d n\rho^n\sqrt{2}}{b\sqrt{2\rho\pi}}\int\limits_{1-i\log^2 n}^{1+i\log^2 n}\frac{t\sqrt{-s}e^{(-\frac{\kappa b\sqrt{-\rho s}}{2})}}{e^{(\frac{\kappa b\sqrt{-\rho s}}{2})} - \frac{itC\rho^d}{\sqrt{\rho b}}\sinh(\frac{\kappa b\sqrt{-\rho s}}{2})}\frac{1}{\rho^n n}e^{-s}ds \\ &= \psi(t),\end{aligned}$$

where $\psi(t)$ is the charateristic function of $\frac{C\rho^d}{\sqrt{2\rho b}}\cdot l\left(\frac{b\sqrt{\rho}}{2\sqrt{2}}\kappa\right)$ given in (2.13).

Now let us turn back to the proof of Theorem 2.15.

The local behaviour of $y_k^{(d)}(z,v)$

For the proof of Theorem 2.15, we study the local behaviour of the functions $y_k^{(d)}(z,v)$ near the singularity. As v is a stable parameter (cf Definition 1.17) and $y_k^{(d)}(z,1) = y(z)$, there is a unique dominant singularity $\rho(v)$ on the circle of convergence of $y_k^{(d)}(z,v)$ which fulfils $\rho(1) = \rho$ due to Lemma 1.18.

Lemma 2.16. *Let $|z| \le \rho^2 + \varepsilon$ for sufficiently small ε and $|v| \le 1$. Then there exists a constant L with $0 < L < 1$ and a positive constant D such that*

$$|w_k^{(d)}(z,v)| \le D|v-1| \cdot |z|^d \cdot L^k$$

Proof. We will only provide a short sketch, since the proof is similar to that of [21, Lemma 2] For $k = 0$ we have

$$|w_0^{(d)}|(z,v)| = |v-1| \cdot |z| \cdot \underbrace{|Z_{d-1}(\mathbf{y(z)})|}_{\mathcal{O}(|y(z)|^{d-1}) = \mathcal{O}(|z|^{d-1})} \le |v-1| \cdot D \cdot |z|^d$$

The result for general $w_k^{(d)}(z,v)$ follows by induction. Starting with the recurrence relation

$$w_{k+1}^{(d)}(z,v) = y(z)\left(\exp\left(w_k^{(d)}(z,v) + \sum_{i\ge 2} \frac{w_k^{(d)}(z^i,v^i)}{i}\right) - 1\right)$$

we use the trivial estimate $|w_k(z,v)| \le 2y(|z|)$ which is valid for $|z| \le \rho$ and $|v| \le 1$, the convexity of $y(z)/z$ on the positive reals, and some elementary estimates for e^z. For the precise details see [21]. ☐

Corollary 2.17. *For $|v| \le 1$ and $|z| \le \rho + \varepsilon$ ($\varepsilon > 0$ small enough) there is a positive constant $\tilde{C}$ such that (for all $k \ge 0, d \ge 1$)*

$$|\Sigma_k^{(d)}(z,v)| \le \tilde{C}|v-1|L^k.$$

Proof. By Lemma 2.16 and with $|v^i - 1| = |1 + v + \cdots v^{i-1}||v-1| \le i|v-1|$ as $|v| \le 1$ we have

$$\begin{aligned} |\Sigma_k^{(d)}(z,v)| &\le \sum_{i\ge 2} \frac{1}{i}|w_k^{(d)}(z^i,v^i)| \le D\sum_{i\ge 2}\frac{1}{i}|v^i - 1||z|^{i\cdot d}L^k \\ &\le D|v-1|L^k \frac{|z|^{2d}}{1-|z|^d} \le D|v-1|L^k \frac{1}{1-(\rho+\epsilon)^d} = \tilde{D}|v-1|L^k \end{aligned}$$

☐

Corollary 2.18. *Let $v \in \Xi_k(\tilde{\eta})$ and $z \in \Theta(\eta)$. Then there exists $L \in (0,1)$ such that*

$$\sum_{i\ge 2} \gamma_k^{(d)}(z^i,v^i) = \mathcal{O}(L^k).$$

Proof. As $i \geq 2$ the functions $\gamma_k^{(d)}(z^i, v^i)$ are analytic in the whole region and $\Gamma_k^{(d)}(z,v) := \sum_{i\geq 2}\gamma_k^{(d)}(z^i,v^i) = \sum_{n,m} y_{nmk}^{(d)} z^n y^m$ with positive coefficients $y_{knm}^{(d)}$, we have $|\Gamma_k^{(d)}(z,v)| \leq \Gamma_k^{(d)}(|z|,|v|)$ where the right-hand side is monotone in $|z|$ and $|v|$.

Now let $z \geq 0$ and $0 < v < 1$. Using Taylor's theorem we get

$$|\Sigma_k^{(d)}(z,v)| = |(v-1)\Gamma_k^{(d)}(z, 1+\vartheta(v-1))| \geq |v-1|\Gamma_k^{(d)}(z,v).$$

In view of Corollary 2.17 this implies for all $z \in \Theta(\eta)$ and $v \in \Xi_k$ the estimate $|\Gamma_k^{(d)}(z,v)| \leq \Gamma_k^{(d)}(|z|,|v|) \leq CL^k$ for some positive constant $L < 1$. □

Now we will refine these a priori bounds. First we show that the first derivate $\gamma_k^{(d)}(z,1)$ is almost a power of $y(z)$. Afterwars we will derive estimates for the second derivative and then obtain a power-like representation for $w_k^{(k)}(z,v)$. Finally, utilizing the recurrence relation for $w_k^{(k)}(z,v)$ we will arrive at the desired result (2.21).

Lemma 2.19. *For $z \in \Theta(\eta)$ (where $\eta > 0$ is sufficiently small) the functions $\gamma_k^{(d)}(z)$ can be represented as*

$$\gamma_k^{(d)}(z) := \gamma_k^{(d)}(z,1) = C_k^{(d)}(z) y(z)^{k+d},$$

where the functions $C_k^{(d)}(z)$ are analytic and converge uniformly to an analytic limit function $C^{(d)}(z)$ (for $z \in \Theta$) with convergence rate

$$C_k^{(d)}(z) = C^{(d)}(z) + \mathcal{O}(L^k)$$

for some $0 < L < 1$, and further $C^{(d)}(\rho) = C_d \rho^d$, where C_d is the constant given in (2.12).

Proof. We define the functions $C_k^{(d)}(z) := \frac{\gamma_k^{(d)}(z)}{y(z)^{k+d}}$.

We prove the analyticity of the functions $\gamma_k^{(d)}(z)$ by induction:

$$\gamma_0^{(d)}(z) = zZ_{d-1}(\mathbf{y}(\mathbf{z})) = z\mathcal{O}(y(z)^{d-1}) = \mathcal{O}(x^d)$$

is analytic in $\Theta(\eta)$ as $y(z)$ is analytic in $\Theta(\eta)$, and so is $C_0^{(d)}(z)$, since it is a quotient of two power series starting with z^d, namely

$$C_0^{(d)}(z) = \frac{z\mathcal{O}(y(z)^{d-1})}{y(z)^d} = \mathcal{O}(1) \tag{2.23}$$

The step of induction works like in [21], as the $\gamma_k^{(d)}$ fulfill the same recursion as the γ_k:

$$\begin{aligned}
\gamma_{k+1}^{(d)}(z,v) &= \frac{\partial}{\partial v} z \exp\left(\sum_{i\geq 1} y_k^{(d)}(z^i,v^i)\right) \\
&= z\exp\left(\sum_{i\geq 1}\frac{y_k^{(d)}(z^i,v^i)}{i}\right)\sum_{i\geq 1}\frac{\partial}{\partial v} y_k^{(d)}(z^i,v^i)v^{i-1} \\
&= y_{k+1}^{(d)}(z,v)\sum_{i\geq 1}\gamma_k^{(d)}(z^i,v^i)v^{i-1},
\end{aligned} \tag{2.24}$$

and for $v = 1$

$$\gamma_{k+1}^{(d)}(z) = y(z)\gamma_k^{(d)}(z) + y(z)\Gamma_k^{(d)}(z),$$

with $\Gamma_k^{(d)}(z) = \sum_{i\geq 2} \gamma_k^{(d)}(z^i)$, which is analytic for $|z| \leq \sqrt{\rho}$ and hence in Θ. Applying the induction hypothesis, this proves the analyticity of $\gamma_k^{(d)}(z)$. Solving the recurrence, we obtain

$$\gamma_k^{(d)}(z) = y(z)^k \gamma_0^{(d)}(z) + \sum_{\ell=0}^{k-1} y(z)^{k-\ell}\Gamma_\ell^{(d)}(z)$$

and hence the analyticity of $\gamma_k^{(d)}$ implies the analyticity of the functions $C_k^{(d)}(z)$ in Θ.
We now want to show that the functions $(C_k^{(d)}(z))_{k\geq 0}$ have a uniform limit $C^{(d)}(z)$, which works analogously as in [21]. Setting $v = 1$, (2.24) translates to

$$C_{k+1}^{(d)}(z)y(z)^{k+d+1} = y(z)\left(C_k^{(d)}(z)y(z)^{k+d} + C_k^{(d)}(z^2)y(z^2)^{k+d} + C_k^{(d)}(z^3)y(z^3)^{k+d}\cdots\right).$$

Hence

$$C_{k+1}^{(d)} = \sum_{\ell\geq 1} C_k^{(d)}(z^\ell)\frac{y(z^\ell)^{k+d}}{y(z)^{k+d}}. \tag{2.25}$$

We set

$$L_k^{(d)} := \sup_{z\in\Theta}\sum_{\ell\geq 2}\frac{|y(z^\ell)|^{k+d}}{|y(z)|^{k+d}}.$$

If η is sufficiently small, we know from Lemma 2.1 that

$$\sup_{z\in\Theta}\frac{|y(z^\ell)|}{|y(z)|} < 1 \quad \text{for all } \ell\geq 2 \quad \text{and} \quad \sup_{z\in\Theta}\frac{|y(z^\ell)|}{|y(z)|} = \mathcal{O}(\bar{L}^\ell)$$

for some $\bar{L}$ with $0 < \bar{L} < 1$. Consequently, we also get $L_k^{(d)} = \mathcal{O}(L^k)$ for some L with $0 < L < 1$ (actually we can choose $L = \bar{L}^2$). We use the notation $\|f\| := \sup_{z\in\Theta}|f(z)|$ and obtain from (2.25)

$$\|C_{k+1}^{(d)}\| \leq \|C_k^{(d)}\|(1 + L_k^{(d)})$$

and also

$$\|C_{k+1}^{(d)} - C_k^{(d)}\| \leq \|C_k^{(d)}\|L_k^{(d)}.$$

The first inequality implies that the functions $C_k^{(d)}(z)$ are uniformly bounded in the given domain by

$$\|C_k^{(d)}\|c_0 := \|C_k^{(d)}\|\prod_{\ell\geq 1}(1 + L_\ell^{(d)}),$$

while the second equation guarantees the existence of a limit $\lim_{k\to\infty} C_k^{(d)}(z) = C^{(d)}(z)$ which is analytic in Θ, and we have a uniform exponential convergence rate

$$\|C_k^{(d)} - C^{(d)}\| \leq c_0\sum_{\ell<k} L_\ell = \mathcal{O}(L^k),$$

hence it follows that the uniform limit exists.

Finally, note that

$$\sum_{k\geq 0}\gamma_k^{(d)}(z,1)=\sum_{k\geq 0}d_n^{(d)}z^n=D^{(d)}(z),$$

where $d_n^{(d)}$ is the total number of vertices of degree d in all trees of size n, and $D^{(d)}(z)$ is the according generating function, introduced in e.g. [62]. On the other hand,

$$\sum_{k\geq 0}\gamma_k^{(d)}(z,1)=\sum_{k\geq 0}(C^{(d)}(z)+\mathcal{O}(L^k))y(z)^k=\frac{C^{(d)}(z)y(z)^d}{1-y(z)}+\mathcal{O}(1),$$

and therefore

$$C^{(d)}(\rho)=\lim_{z\to\rho}\frac{(1-y(z))D^{(d)}(z)}{y(z)^d}.$$

We know that

$$D^{(d)}(z)=\frac{y(z)\sum_{i\geq 2}D^{(d)}(z^i)+zZ_{d-1}(\mathbf{y}(\mathbf{z}))}{1-y(z)}$$

(cf. [62, Eq. (36)] or [48]). Schwenk [63, Lemma 4.1] computed the limit of the cycle index in the numerator. In his proof he provides the speed of convergence as well. In fact, [63, Eq. (32)] says that

$$\left|Z_d\left(\frac{\mathbf{y}(\mathbf{z})}{\mathbf{z}}\right)-\exp\left(\sum_{i=1}^{d}\frac{1}{i}\left(\frac{y(z)}{z}-1\right)\right)\right|\leq z^{d+1}\exp\left(\lambda\sum_{i=1}^{d}\frac{1}{i}\right)$$

with $\lambda=\sup_{0\leq z\leq\rho}\frac{1}{z}\left(\frac{y(z)}{z}-1\right)=\frac{1-\rho}{\rho^2}$. Thus $zZ_{d-1}(\mathbf{y}(\mathbf{z}))=z^dF(z)+\mathcal{O}\left(dz^{2d+1}\right)$. Note further that $D^{(d)}(z)=\mathcal{O}\left(z^{d+1}\right)$ since there are no nodes of degree d in trees of size less than $d+1$. This implies $C^{(d)}(\rho)=C_d\rho^d$ with $C_d=C+\mathcal{O}\left(d\rho^d\right)$ and C as in Theorem 2.12. Hence Lemma 2.19 is proven. ☐

Lemma 2.20. *There exist constants $\epsilon,\theta,\tilde{\eta}>0$ and $\theta<\frac{\pi}{2}$ such that*

$$|\gamma_k^{(d)}(z,v)|=\mathcal{O}(|y(z)|^{k+d})$$

uniformly for $z\in\Delta_\epsilon(\theta)$ and $v\in\Xi_k(\tilde{\eta})$.

Proof. For $\ell\leq k$ we set

$$\bar{C}_\ell^{(d)}=\sup_{\substack{z\in\Delta_\epsilon(\theta)\\ v\in\Xi_k(\tilde{\eta})}}\left|\frac{\gamma_\ell^{(d)}(z,v)}{y(z)^{\ell+d}}\right|.$$

First we derive the following inequality, using the recurrence (2.15) for $y_k^{(d)}(z,v)$:

$$
\begin{aligned}
|y_{\ell+1}^{(d)}(z,v)| &= \left| z \exp\left(\sum_{i\geq 1} \frac{1}{i} \left(y_\ell^{(d)}(z^i,v^i) - y(z^i) + y(z^i) \right) \right) \right| \\
&= \left| y(z) \exp\left(\sum_{i\geq 1} \frac{1}{i} w_\ell^{(d)}(z^i,v^i) \right) \right| \\
&\leq |y(z)| \exp\left(|w_\ell^{(d)}(z,v)| + \sum_{i\geq 2} \frac{1}{i} |w_\ell^{(d)}(z^i,v^i)| \right) \\
&\leq |y(z)| \exp\left(\underbrace{|\gamma_\ell^{(d)}(z,1+\vartheta(v-1))||v-1|}_{\leq \bar{C}_\ell^{(d)}|v-1|} + \sum_{i\geq 2} \frac{|v^i-1|}{i} |\gamma_\ell^{(d)}(z^i,1+\vartheta(v^i-1))| \right)
\end{aligned}
$$

where we did a Taylor expansion in the last step, with some $0<\vartheta<1$ and thus $1+\vartheta(v^i-1) \in \Xi_k$. To get an estimate for the second term, we use that $|v^i-1| = |1+v+\cdots+v^{i-1}||v-1| \leq i|v-1|$ as $|v|\leq 1$ and hence $|\frac{v^i-1}{i}| \leq |v-1| \leq 2$. Further we use $|\gamma_\ell^{(d)}(z^i,1+\vartheta(v^i-1))| \leq |\gamma_\ell^{(d)}(z^i,1)|$, $|y(z)|\leq 1$ and Corollary 2.18 to obtain

$$
|y_{\ell+1}^{(d)}(z,v)| \leq |y(z)| \exp\left(\bar{C}_\ell^{(d)}|v-1| + \mathcal{O}(L^\ell) \right).
$$

Using recurrence (2.24) leads to

$$
\begin{aligned}
\bar{C}_{\ell+1}^{(d)} &= \sup_{\substack{z\in\Delta_\varepsilon(\theta)\\ v\in\Xi_k(\tilde{\eta})}} \left| \frac{y_{\ell+1}^{(d)}(z,v)}{y(z)} \right| \left| \frac{\gamma_\ell^{(d)}(z,v) + \sum_{i\geq 2} \gamma_\ell^{(d)}(z^i,v^i)v^{i-1}}{y(z)^{\ell+d}} \right| \\
&\leq e^{\bar{C}_\ell^{(d)}\frac{\eta}{k} + \mathcal{O}(L^\ell)} (\bar{C}_\ell^{(d)} + \mathcal{O}(L^\ell)) \\
&= \bar{C}_\ell^{(d)} e^{\bar{C}_\ell^{(d)}\frac{\eta}{k}} (1+\mathcal{O}(L^\ell)),
\end{aligned}
\tag{2.26}
$$

where we used Lemma 2.19 to get $|\sum_{i\geq 2} \gamma_\ell^{(d)}(z^i,v^i)v^{i-1}| = \mathcal{O}(\sum_{i\geq 2}|y(z^i)|^{\ell+d})$ and hence

$$
\sup_{\substack{z\in\Delta_\epsilon(\theta)\\ v\in\Xi_k(\tilde{\eta})}} \left| \frac{\sum_{i\geq 2} \gamma_\ell^{(d)}(z^i,v^i)v^{i-1}}{y(z)^{\ell+d}} \right| = \sup\left(\mathcal{O}\left(\sum_{i\geq 2} \left(\frac{y(z^i)}{y(z)} \right)^{\ell+d} \right) \right) = \mathcal{O}(L^{\ell+d}) = \mathcal{O}(L^\ell).
$$

We now set

$$
c_0 = \prod_{j\geq 0} (1+\mathcal{O}(L^j)).
$$

Recall that, by Equation (2.23), $|\frac{\gamma_0^{(d)}(z,v)}{y(z)^d}| = \mathcal{O}(1)$, hence $\bar{C}_0^{(d)} = \sup|\frac{\gamma_0^{(d)}(z,v)}{y(z)^d}| = \mathcal{O}(1)$, too. Thus we can choose $\eta>0$ such that $e^{2\bar{C}_0^{(d)}c_0\eta} \leq 2$. For fixed k we get

$$
\bar{C}_\ell^{(d)} \leq \bar{C}_0^{(d)} \prod_{j<\ell} (1+\mathcal{O}(L^j)) e^{2\bar{C}_0^{(d)}c_0 c\frac{\ell}{k}} \leq 2\bar{C}_0 c_0 = \mathcal{O}(1).
$$

The second estimate is clear by the choice of η and by $\ell \leq k$. The first inequality can be obtained from (2.26) by induction:

$$\begin{aligned}
\bar{C}_1^{(d)} &\leq \bar{C}_0^{(d)}(1+\mathcal{O}(L^0))e^{\bar{C}_0^{(d)}\frac{\eta}{k}} \leq \bar{C}_0^{(d)}\prod_{j<1}(1+\mathcal{O}(L^j))e^{2\bar{C}_0^{(d)}c_0\eta\frac{1}{k}},\\
\bar{C}_{\ell+1}^{(d)} &\leq \bar{C}_\ell^{(d)}e^{\frac{\eta}{k}C_\ell^{(d)}}(1+\mathcal{O}(L^l))\\
&= \prod_{j<\ell}(1+\mathcal{O}(L^j))(1+\mathcal{O}(L^\ell))\bar{C}_0^{(d)}e^{2\bar{C}_0^{(d)}c_0\eta\frac{\ell}{k}}\exp\left(\frac{\eta}{k}\bar{C}_0^{(d)}\Pi_{j<\ell}\underbrace{e^{2\bar{C_0}^{(d)}c_0\eta\frac{\ell}{k}}}_{\leq 2^{\frac{\ell}{k}}\leq 2}\right)\\
&\leq \bar{C}_0\prod_{j<\ell+1}(1+\mathcal{O}(L^j))e^{2\bar{C}_0c_0\eta\frac{\ell+1}{k}})
\end{aligned}$$

And hence, knitting together everything, we obtain the formula of Lemma 2.20. □

For the second derivatives with respect to v of $y_k^{(d)}(z,v)$, $\gamma_k^{(d)[2]}(z,v)$, we find

Lemma 2.21. *Suppose that $|z| \leq \rho - \eta$ for some $\eta > 0$ and $|v| \leq 1$. Then*

$$\gamma_k^{(d)[2]}(z,v) = \mathcal{O}(y(|z|)^{k+d}) \tag{2.27}$$

uniformly in z and v. There also exist constants $\epsilon, \theta, \tilde{\eta}$ such that uniformly for $v \in \Xi_k(\tilde{\eta})$ and $z \in \Delta_\epsilon(\theta)$

$$\gamma_k^{(d)[2]}(z,v) = \mathcal{O}(ky(|z|)^{k+d}). \tag{2.28}$$

Proof. Derivation of (2.24) leads to the recurrence

$$\begin{aligned}
\gamma_{k+1}^{(d)[2]}(z,v) &= y_{k+1}^{(d)}(z,v)\left(\sum_{i\geq 1}\gamma_k^{(d)}(z^i,v^i)v^{i-1}\right)^2\\
&\quad + y_{k+1}^{(d)}(z,v)\sum_{i\geq 1}i\gamma_k^{(d)[2]}(z^i,v^i)v^{2(i-1)}\\
&\quad + y_{k+1}^{(d)}(z,v)\sum_{i\geq 2}(i-1)\gamma_k^{(d)}(z^i,v^i)v^{i-2}
\end{aligned}$$

with initial condition $\gamma_0^{(d)[2]}(z,v) = 0$.

For $|z| < \rho - \eta$, for some $\eta > 0$ and for $|v| \leq 1$ we have $|\gamma_k^{(d)[2]}(z,v)| \leq \gamma_k^{(d)[2]}(|z|,1)$. Thus, in this case we can restrict ourselves to non-negative real $z \leq \rho - \eta$.

By using the bounds $\gamma_k^{(d)}(z,1) \leq C_k^{(d)}y(z)^{k+d}$ from Lemma 2.20, $\sum_{i\geq 2}\gamma_k^{(d)}(z^i,v^i) = \mathcal{O}(L^k)$ from Corollary 2.18 and the induction hypothesis $\gamma_k^{(d)[2]}(z,1) \leq D_k^{(d)}y(z)^{k+d}$, we can derive the following upper bound from the above:

$$\begin{aligned}
\gamma_{k+1}^{(d)[2]}(z) &= y(z)\left(\sum_{i\geq 1}\gamma_k^{(d)}(z^i)\right)^2 + y(z)\sum_{i\geq 1} i\gamma_k^{(d)[2]}(z^i) + y(z)\sum_{i\geq 2}(i-1)\gamma_k^{(d)}(z^i)\\
&\leq y(z)\left[\left(C_k^{(d)}y(z)^{k+d}+\mathcal{O}(L^k)\right)^2 + D_k^{(d)}(\sum_{i\geq 1} iy(z^i)^{k+d}) + \sum_{i\geq 2} C_k^{(d)}y(z^i)^{k+d})\right]\\
&\leq y(z)^{k+d+1}\left[(C_k^{(d)})^2 y(z)^{k+d} + C_k^{(d)}\mathcal{O}(L^k)\right.\\
&\qquad\left. +D_k\left(1+\sum_{i\geq 2} i\frac{y(z^i)^{k+d}}{y(z)^{k+d}}\right) + C\sum_{i\geq 2}(i-1)\frac{y(z^i)^{k+d}}{y(z)^{k+d}}\right]\\
&\leq y(z)^{k+d+1}((C^{(d)})_k^2 y(\rho-\eta)^{k+d} + D_k^{(d)}(1+\mathcal{O}(L^k)) + \mathcal{O}(L^k)).
\end{aligned}$$

Consequently

$$D_{k+1}^{(d)} = ((C_k^{(d)})^2 y(\rho-\eta)^{k+d} + D_k^{(d)}(1+\mathcal{O}(L^k)) + \mathcal{O}(L^k)),$$

which leads to $D_k^{(d)} = \mathcal{O}(1)$ as $k\to\infty$.

To prove the second property we use the same constants $\epsilon, \theta, \tilde{\eta}$ as in Lemma 2.20 and set

$$\bar{D}_\ell^{(d)} = \sup_{\substack{z\in\Delta_\epsilon(\theta)\\ v\in\Xi_k(\tilde{\eta})}} \left|\frac{\gamma_\ell^{(d)[2]}(z,v)}{y(z)^{\ell+d}}\right|$$

for $\ell \leq k$. We use the already known bound $|\gamma_\ell^{(d)}(z,v)| \leq \bar{C}^{(d)}|y(z)^{k+d}|$ and by similar considerations as in the proof of this lemma, we get

$$\begin{aligned}
\bar{D}_{\ell+1}^{(d)} &= \sup_{\substack{z\in\Delta_\epsilon(\theta)\\ v\in x}} \left|\frac{y_{\ell+1}^{(d)}(z,v)}{y(z)}\right|\\
&\quad\times\left|\frac{\left(\sum_{i\geq 1}\gamma_\ell^{(d)}(z^i,v^i)v^{i-1}\right)^2 + \sum_{i\geq 1} i\gamma_\ell^{(d)[2]}(z^i,v^i)v^{2(i-1)} + \sum_{i\geq 2}(i-1)\gamma_\ell^{(d)}(z^i,v^i)v^{i-2}}{y(z)^{\ell+d}}\right|\\
&\leq \bar{D}_\ell^{(d)} e^{\bar{C}_\ell^{(d)}\frac{\eta}{k}}(1+\mathcal{O}(L^\ell)) + (C_\ell^{(d)})^2 e^{\bar{C}_\ell^{(d)}\frac{\eta}{k}} + \mathcal{O}(L^\ell))\\
&\leq \alpha_\ell^{(d)}\bar{D}_\ell^{(d)} + \beta_\ell^{(d)}
\end{aligned}$$

with $\alpha_\ell^{(d)} = e^{\bar{C}_\ell^{(d)}\frac{\eta}{k}}(1+\mathcal{O}(L^\ell))$ and $\beta_\ell^{(d)} = C^{(d)2}e^{\bar{C}_\ell^{(d)}\frac{\eta}{k}} + \mathcal{O}(L^\ell))$. Thus

$$\begin{aligned}\bar{D}_k^{(d)} &\leq \alpha_{k-1}^{(d)}(\alpha_{k-2}^{(d)}(\dots(\alpha_0^{(d)} D_0^{(d)} + \beta_0^{(d)})\dots)\beta_{k-2}^{(d)}) + \beta_{k-1}^{(d)} \\ &= \sum_{j=0}^{k-1} \beta_j^{(d)} \prod_{i=j+1}^{k-1} \alpha_i^{(d)} + \alpha_0^{(d)} \bar{D}_0^{(d)} \\ &\leq k \max_j \beta_j^{(d)} e^{\bar{C}_\ell^{(d)} c} \prod_{i\geq 0}(1+\mathcal{O}(L^i)) \\ &= \mathcal{O}(k),\end{aligned}$$

which completes the proof of the Lemma. □

Lemma 2.22. *Let $\epsilon, \theta, \tilde{\eta}$ and $C_k^{(d)}(z)$ be as in Lemma 2.19 and Lemma 2.20. Then*

$$w_k^{(d)}(z,v) = C_k^{(d)}(z)(v-1)y(z)^{k+d}(1+\mathcal{O}(k|v-1|)) \tag{2.29}$$

uniformly for $z \in \Delta_\epsilon(\theta)$ and $v \in \Xi_k(\tilde{\eta})$. Furthermore we have for $|z| \leq \rho + \eta$ and $|v| \leq 1$

$$\Sigma_k^{(d)}(z,v) = \tilde{C}_k^{(d)}(z)(v-1)y(z^2)^{k+d} + \mathcal{O}(|v-1|^2 y(|z|^2)^{k+d}), \tag{2.30}$$

where the analytic functions $\tilde{C}_k^{(d)}(z)$ are given by

$$\tilde{C}_k^{(d)}(z) = \sum_{i\geq 2} C_k^{(d)}(z^i)\left(\frac{y(z^i)}{y(z^2)}\right)^{k+d}$$

and have a uniform limit $\tilde{C}^{(d)}(z)$ with convergence rate

$$\tilde{C}_k^{(d)}(z) = \tilde{C}^{(d)}(z) + \mathcal{O}(L^k)$$

for some constant L with $0 < L < 1$.

Proof. To prove the first statement, we expand $w_k^{(d)}(z,v)$ into a Taylor polynomial of degree 2 around $v = 1$ and apply Lemmas 2.19 and 2.21.

To prove the second statement, we again use Taylor series. Note that for $i \geq 2$ we have $|z^i| < \rho - \eta$ if $|z| < \rho + \eta$ and η is sufficiently small. We get

$$w_k^{(d)}(z^i, v^i) = C_k^{(d)}(z^i)(v^i-1)y(z^i)^{k+d} + \mathcal{O}(|v^i-1|^2 y(|z^i|)^{k+d})$$

and consequently

$$\begin{aligned}\Sigma_k^{(d)}(z,v) &= \sum_{i\geq 2}\frac{1}{i}C_k^{(d)}(z^i)(v^i-1)y(z^i)^{k+d} + \mathcal{O}(|v-1|^2 y(|z^2|)^{k+d}) \\ &= (v-1)\tilde{C}_k^{(d)}(z)y(z^2)^{k+d} + \mathcal{O}(|v^i-1|^2 y(|z^i|)^{k+d}),\end{aligned}$$

where we used the property that

$$\begin{aligned}\sum_{i\geq 2} C_k^{(d)}(z^i)\frac{v^i-1}{i(v-1)}\frac{y(z^i)^{k+d}}{y(z^2)^{k+d}} &= \sum_{i\geq 2} C_k^{(d)}(z^i)\frac{(1+v+\cdots+v^{i-1})}{i}\frac{y(z^i)^{k+d}}{y(z^2)^{k+d}} \\ &= \tilde{C}_k^{(d)}(z) + \mathcal{O}(\tilde{C}_k^{(d)}(z)(v-1))\end{aligned}$$

represents an analytic function in z and v, and thus its leading term, as $v \to 1$, is our function $\tilde{C}_k^{(d)}(z)$. Finally, since $C_k^{(d)}(z) = C^{(d)}(z) + \mathcal{O}(L^k)$ it follows that $\tilde{C}_k^{(d)}(z)$ has a limit $\tilde{C}^{(d)}(z)$ with the same order of convergence. □

Lemma 2.23. *For $z \in \Delta_\epsilon(\theta)$ and $v \in \Xi_k$ (with the constants $\epsilon, \theta, \tilde{\eta}$ as in Lemma 2.20) we have*

$$w_k^{(d)}(z,v) = \frac{(v-1)y(z)^{k+d}C_k^{(d)}(z)}{1 - \frac{y(z)^d C_k^{(d)}(z)(v-1)}{2}\frac{1-y(z)^k}{1-y(z)} + \mathcal{O}(|v-1|)}.$$

Proof. $w_k^{(d)}(z,v)$ satisfy the recursive relation

$$\begin{aligned} w_{k+1}^{(d)}(z,v) &= z\exp\left(\sum_{i\ge 1}\frac{1}{i}y_k^{(d)}(z^i,v^i)\right) - y(z) \\ &= z\exp\left(\sum_{i\ge 1}\frac{1}{i}\left(w_k^{(d)}(z^i,v^i)+y(z^i)\right)\right) - y(z) \\ &= y(z)\left(\exp\left(w_k^{(d)}(z,v)+\Sigma_k^{(d)}(z,v)\right)-1\right), \end{aligned}$$

and further, since by Lemma 2.22 it follows that $\Sigma_k^{(d)}(z,v) = \mathcal{O}(w_k^{(d)}(z,v)L^k) = \mathcal{O}(w_k^{(d)}(z,v))$ (for brevity, we omit the arguments now),

$$\begin{aligned} w_{k+1}^{(d)} &= y\left[(w_k^{(d)}+\Sigma_k^{(d)}) + \frac{(w_k^{(d)}+\Sigma_k^{(d)})^2}{2} + \mathcal{O}\left((w_k^{(d)}+\Sigma_k^{(d)})^3\right)\right] \\ &= y(w_k^{(d)}+\Sigma_k^{(d)})\left(1+\frac{(w_k^{(d)}+\Sigma_k^{(d)})}{2} + \mathcal{O}\left((w_k^{(d)}+\Sigma_k^{(d)})^2\right)\right) \\ &= yw_k^{(d)}\left(1+\frac{\Sigma_k^{(d)}}{w_k^{(d)}}\right)\left(1+\frac{w_k^{(d)}}{2}+\mathcal{O}(\Sigma_k^{(d)})+\mathcal{O}\left((w_k^{(d)})^2\right)\right). \end{aligned}$$

From there, we obtain

$$\begin{aligned} \frac{y}{w_{k+1}^{(d)}}\cdot\left(1+\frac{\Sigma_k^{(d)}}{w_k^{(d)}}\right) &= \frac{1}{w_k^{(d)}}\frac{1}{\left(1+\frac{w_k^{(d)}}{2}+\mathcal{O}(\Sigma_k^{(d)})+\mathcal{O}\left((w_k^{(d)})^2\right)\right)} \\ &= \frac{1}{w_k^{(d)}}\left(1-\frac{w_k^{(d)}}{2}+\mathcal{O}(\Sigma_k^{(d)})+\mathcal{O}(w_k^{(d)2})\right) \\ &= \frac{1}{w_k^{(d)}} - \frac{1}{2} + \mathcal{O}\left(\frac{\Sigma_k^{(d)}}{w_k^{(d)}}\right) + \mathcal{O}(w_k^{(d)}). \end{aligned}$$

This leads to a recursion

$$\frac{y^{k+1}}{w_{k+1}^{(d)}} = \frac{y^k}{w_k^{(d)}} - \frac{\Sigma_k^{(d)}\cdot y(z)^{k+1}}{w_k^{(d)}w_{k+1}^{(d)}} - \frac{1}{2}y(z)^k + \mathcal{O}\left(\frac{\Sigma_k^{(d)}\cdot y^k}{w_k^{(d)}}\right) + \mathcal{O}(w_k^{(d)}y^k),$$

which we can solve to

$$\begin{aligned}
\frac{y^k}{w_k^{(d)}} &= \frac{1}{w_0^{(d)}} - \sum_{\ell=0}^{k-1} \frac{\Sigma_\ell^{(d)} \cdot y(z)^{\ell+1}}{w_\ell^{(d)} w_{\ell+1}^{(d)}} - \frac{1}{2}\sum_{\ell=0}^{k-1} y^\ell + \mathcal{O}\left(\sum_{\ell=0}^{k-1} \frac{\Sigma_\ell^{(d)} \cdot y^\ell}{w_\ell^{(d)}}\right) + \mathcal{O}(\sum_{\ell=0}^{k-1} w_\ell^{(d)} y^\ell) \\
&= \frac{1}{w_0^{(d)}} \left(1 - w_0^{(d)} \sum_{\ell=0}^{k-1} \frac{\Sigma_\ell^{(d)} \cdot y^{\ell+1}}{w_\ell^{(d)} w_{\ell+1}^{(d)}} - w_0^{(d)} \frac{1}{2}\frac{1-y^k}{1-y} + \mathcal{O}(w_0^{(d)} \frac{1-L^k}{1-L}) + \mathcal{O}((w_0^{(d)})^2 \frac{1-y^{2k}}{1-y^2})\right),
\end{aligned} \tag{2.31}$$

where we used that $\frac{\Sigma_\ell^{(d)} y^\ell}{w_\ell^{(d)}} = \mathcal{O}(L^\ell)$ and that by Lemma 2.22 $w_k^{(d)} = \mathcal{O}(y w_{k-1}^{(d)}) = \mathcal{O}(y^k w_0^{(d)})$. We now analyze the terms of (2.31). Again we apply Lemma 2.22 and (2.15) to obtain

$$\begin{aligned}
w_0^{(d)} \sum_{\ell=0}^{k-1} \frac{\Sigma_\ell^{(d)} \cdot y(z)^{\ell+1}}{w_\ell^{(d)} w_{\ell+1}^{(d)}} &= (v-1) z Z_{d-1} \sum_{\ell=0}^{k-1} \frac{\tilde{C}_\ell^{(d)} (v-1) y(z^2)^{\ell+d} + \mathcal{O}(|v-1|^2 y(|z|^2)^{\ell+d})}{C_\ell^{(d)} C_{\ell+1}^{(d)} y^{2(\ell+d)+1} (v-1)^2 (1 + \mathcal{O}(\ell|v-1|))} y^{\ell+1} \\
&= \frac{z Z_{d-1}}{y(z)^d} \sum_{\ell=0}^{k-1} \frac{\tilde{C}_\ell^{(d)} y(z^2)^{\ell+d} + \mathcal{O}(|v-1|^2 y(|z|^2)^{\ell+d})}{C_\ell^{(d)} C_{\ell+1}^{(d)} y(z)^{\ell+d} (1 + \mathcal{O}(\ell|v-1|))} \\
&= \frac{z Z_{d-1}}{y(z)^d} \left[\sum_{\ell=0}^{k-1} \frac{\tilde{C}_\ell^{(d)}}{C_\ell^{(d)} C_{\ell+1}^{(d)}} \frac{y(z^2)^{\ell+d}}{y(z)^{\ell+d}} + \sum_{\ell=0}^{k-1} \frac{\mathcal{O}(|v-1|^2 y(|z|^2)^{\ell+d})}{C_\ell^{(d)} C_{\ell+1}^{(d)} y(z)^{\ell+d}}\right] \left(\frac{1}{1 + \mathcal{O}(\ell|v-1|)}\right) \\
&= \frac{z Z_{d-1}}{y(z)^d} \left[\sum_{\ell=0}^{k-1} \frac{\tilde{C}_\ell^{(d)}}{C_\ell^{(d)} C_{\ell+1}^{(d)}} \frac{y(z^2)^{\ell+d}}{y(z)^{\ell+d}} + \underbrace{\sum_{\ell=0}^{k-1} \frac{\mathcal{O}(|v-1|^2 y(|z|^2)^{\ell+d})}{C_\ell^{(d)} C_{\ell+1}^{(d)} y(z)^{\ell+d}}}_{=\mathcal{O}(|v-1|^2 L^\ell)}\right] (1 + \mathcal{O}(\ell|v-1|)) \\
&\frac{z Z_{d-1}}{y(z)^d} \left[\sum_{\ell=0}^{k-1} \frac{\tilde{C}_\ell^{(d)}}{C_\ell^{(d)} C_{\ell+1}^{(d)}} \frac{y(z^2)^{\ell+d}}{y(z)^{\ell+d}} + \sum_{\ell=0}^{k-1} \frac{\tilde{C}_\ell^{(d)}}{C_\ell^{(d)} C_{\ell+1}^{(d)}} \underbrace{\frac{y(z^2)^{\ell+d}}{y(z)^{\ell+d}}}_{=\mathcal{O}(L^\ell)} \mathcal{O}(\ell|v-1|) + \mathcal{O}(|v-1|^2)\right] \\
&= c_k^{(d)} + \mathcal{O}(|v-1|),
\end{aligned}$$

where $c_k^{(d)}$ denotes the first sum. Note that by (2.23), $\frac{z Z_{d-1}}{y(z)^d} = \mathcal{O}(1)$.

Now turn to the error terms of (2.31) and observe that $w_0^{(d)} \frac{1-y^{2k}}{1-y^2} = \mathcal{O}(k|v-1| y(z)^d) = \mathcal{O}(y(z)^d) = \mathcal{O}(1)$ if $k|v-1| \le \tilde{\eta}$. Thus, we obtain the following representation for $w_k^{(d)}(z,v)$:

$$w_k^{(d)} = \frac{w_0^{(d)} y^k}{1 - c_k^{(d)}(z) - \frac{w_0^{(d)}}{2} \frac{1-y^k}{1-y} + \mathcal{O}(|v-1|)}.$$

We use the expressions

$$C_{k+1}^{(d)} = \sum_{i\ge1} C_k^{(d)} \frac{y(z^i)^{k+d}}{y(z)^{k+d}} \text{ and } \tilde{C}_k^{(d)} = \sum_{i\ge2} C_k^{(d)} \frac{y(z^i)^{k+d}}{y(z^2)^{k+d}},$$

which are consequences of Lemmas 2.19, Equation (2.24) and 2.22, to obtain

$$\tilde{C}_k^{(d)}(z) = (C_{k+1}^{(d)}(z) - C_k^{(d)}(z)) \left(\frac{y(z)}{y(z^2)}\right)^{k+d}. \tag{2.32}$$

This provides the telescope sum:

$$c_k^{(d)} = \frac{zZ_{d-1}}{y(z)^d} \sum_{\ell=0}^{k-1} \frac{C_{\ell+1}^{(d)} - C_\ell^{(d)}}{C_\ell^{(d)} C_{\ell+1}^{(d)}} \tag{2.33}$$

$$= \frac{zZ_{d-1}}{y(z)^d} \left(\frac{1}{C_0^{(d)}} - \frac{1}{C_k^{(d)}} \right) \tag{2.34}$$

$$\tag{2.35}$$

and hence, since $C_0^{(d)} = \frac{\gamma_0^{(d)}}{y(z)^d} = \frac{zZ_{d-1}}{y(z)^d}$, we get

$$1 - c_k^{(d)} = \frac{zZ(S_{d-1})}{y(z)^d C_k^{(d)}},$$

which provides the result. □

It is now easy to prove Theorem 2.15. With $z = \rho(1 + \frac{s}{n})$, $v = e^{\frac{it}{\sqrt{n}}}$, d and $t \neq 0$ fixed, $k = \kappa\sqrt{n}$ and representation (2.3) of $y(z)$ we obtain the expansions:

$$v - 1 \sim \frac{it}{\sqrt{n}}$$

$$1 - y(z) \sim b\sqrt{\frac{-\rho s}{n}}$$

$$y(z)^k \sim 1 - kb\sqrt{\frac{-\rho s}{n}} + \cdots \sim e^{-\kappa b\sqrt{-\rho s}}$$

$$y(z)^d \sim 1 - db\sqrt{\frac{-\rho s}{n}} + \cdots \sim 1.$$

Since the functions $C_k^{(d)}(z)$ are continuous and uniformly convergent to $C^{(d)}(z)$, they are also uniformly continuous and thus $C_k^{(d)}(z) \sim C^{(d)}(\rho) = C_d\rho^d$. This leads to

$$w_k^{(d)}(z, v) \sim \frac{\frac{it}{\sqrt{n}} C_d\rho^d e^{-\kappa b\sqrt{-\rho s}}}{1 - \frac{it}{\sqrt{n}} C_d\rho^d \left(\frac{1}{2} \frac{1-e^{-\kappa b\sqrt{-\rho s}}}{b\sqrt{\frac{-\rho s}{n}}} \right)}$$

$$= \frac{1}{\sqrt{n}} \cdot \frac{\sqrt{-s}\; it\; C_d\rho^d e^{-\kappa b\sqrt{-\rho s}}}{\sqrt{-s} - \frac{itC_d\rho^d}{2b\sqrt{\rho}} \left(1 - e^{-\kappa b\sqrt{-\rho s}}\right)}$$

$$= \frac{C_d\rho^d}{\sqrt{n}} \cdot \frac{it\sqrt{-s}e^{\frac{1}{2}-\kappa b\sqrt{-\rho s}}}{\sqrt{-s}e^{\frac{1}{2}\kappa b\sqrt{-\rho s}} - \frac{itC_d\rho^{d.}}{b\sqrt{\rho}} \sinh(\frac{1}{2}\kappa b\sqrt{-\rho s})} \tag{2.36}$$

2.3.2 Finite dimensional limiting distributions

First we consider the case $m = 2$. The computation of the 2-dimensional limiting distribution shows the general method of the proof. Iterative applications of the arguments will eventually prove Theorem 2.13.

Theorem 2.24. *Let* $z = \rho(1+\frac{s}{n})$, $v_1 = e^{\frac{it_1}{\sqrt{n}}}, v_2 = e^{\frac{it_2}{\sqrt{n}}}$, $k = \kappa\sqrt{n}$ *and* $h = \eta\sqrt{n}$. *Moreover, assume that* $|\arg s| \geq \Theta > 0$ *and, as* $n \to \infty$, *we have* $s = \mathcal{O}(\log^2 n)$, *whereas* κ, t_1 *and* t_2 *are fixed. Then, for large* d, $w_{k,h}^{(d)}(z,u)$ *admits the local representation*

$$w_{k,h}^{(d)}(z,v_1,v_2) \sim \frac{C\rho^d}{\sqrt{n}} \times \frac{\left(it_2 + \frac{it_1\sqrt{-s}e^{(-\frac{1}{2}\kappa b\sqrt{-\rho s})}}{\sqrt{-s}e^{(\frac{1}{2}\kappa b\sqrt{-\rho s})} - \frac{it_1 C\rho^d}{\sqrt{\rho}b}\sinh\left(\frac{1}{2}\kappa b\sqrt{-\rho s}\right))}\right)\sqrt{-s}e^{(-\frac{1}{2}\eta b\sqrt{-\rho s})}}{\sqrt{-s}e^{(\frac{1}{2}\eta b\sqrt{-\rho s})} - \frac{C\rho^d}{b\sqrt{\rho}}\left(it_2 + \frac{it_1\sqrt{-s}e^{(-\frac{1}{2}\kappa b\sqrt{-\rho s})}}{\sqrt{-s}e^{(\frac{1}{2}\kappa b\sqrt{-\rho s})} - \frac{it_1 C\rho^d}{\sqrt{\rho}b}\sinh\left(\frac{1}{2}\kappa b\sqrt{-\rho s}\right))}\right)\sinh\left(\frac{1}{2}\eta b\sqrt{-\rho s}\right))}. \tag{2.37}$$

Note that $y_{k,h}^{(d)}(z,v_1,1) = y_k^{(d)}(z,v_1)$ and $y_{k,h}^{(d)}(z,1,v_2) = y_{k+h}(z,v_2)$. Considering the first derivative, we denote by

$$\frac{\partial}{\partial v_1} y_{k,h}^{(d)}(z,v_1,v_2) =: \gamma_{k,h}^{(d)[v_1]}(z,v_1,v_2)$$
$$\frac{\partial}{\partial v_2} y_{k,h}^{(d)}(z,v_1,v_2) =: \gamma_{k,h}^{(d)[v_2]}(z,v_1,v_2),$$

and by simple induction, we observe that:

$$\begin{aligned}\gamma_{k,h}^{(d)[v_1]}(z,1,1) &= \gamma_k^{(d)}(z,1) = \gamma_k^{(d)}(z) = C_k^{(d)}(z)y(z)^{k+d}\\ \gamma_{k,h}^{(d)[v_2]}(z,1,1) &= \gamma_{k+h}^{(d)}(z,1) = \gamma_{k+h}^{(d)}(z) = C_{k+h}^{(d)}(z)y(z)^{k+h+d}.\end{aligned} \tag{2.38}$$

As $|\gamma_{k,h}^{(d)[u_i]}(z,v_1,v_2)| \leq \gamma_{k,h}^{(d)[u_i]}(z,1,1)$ for $i = 1,2, v_1 \in \Xi_k, v_2 \in \Xi_{k+h}$, and $|z| \leq \rho$ it follows that $|\gamma_{k,h}^{(d)[v_1]}(z,v_1,v_2)| = \mathcal{O}(y(z)^{k+d})$ and $\gamma_{k,h}^{(d)[v_2]}(z,v_1,v_2) \leq \mathcal{O}(y(z)^{k+h+d})$ in the same regions. To be more precise, we can prove the following analogue to Lemma 2.20.

Lemma 2.25. *There exist constants* $\epsilon, \theta, \tilde{\eta}_1, \tilde{\eta}_2$, *such that for* $z \in \Delta_\epsilon(\theta), v_1 \in \Xi_k(\tilde{\eta}_1)$ *and* $v_2 \in \Xi_{k+h}(\tilde{\eta}_2)$

$$\gamma_{k,h}^{(d)[v_1]}(z,v_1,v_2) + \gamma_{k,h}^{(d)[v_2]}(z,v_1,v_2) = \mathcal{O}(|y(z)|^{k+d})$$

Proof. Set

$$C_{\ell,h}^{(d)[v_1]} = \sup_{\substack{z\in\Delta_\epsilon\\ v_1\in\Xi_k, v_2\in\Xi_{k+h}}} \left|\frac{\gamma_{\ell,h}^{(d)[v_1]}(z,v_1,v_2)}{y(z)^{k+d}}\right|$$
$$C_{\ell,h}^{(d)[v_2]} = \sup_{\substack{z\in\Delta_\epsilon\\ v_1\in\Xi_k, v_2\in\Xi_{k+h}}} \left|\frac{\gamma_{\ell,h}^{(d)[v_2]}(z,v_1,v_2)}{y(z)^{k+h+d}}\right|$$

As in the proof of Lemma 2.21, we apply Taylor's theorem (in two variables) to get

$$\begin{aligned}|y^{(d)}_{\ell+1,h}(z,v_1,v_2)| &= |y(z)|\exp\left(|w^{(d)}_\ell(z,v_1,v_2)| + \sum_{i\ge 2}\frac{|w^{(d)}_\ell(z^i,v_1^i,v_2^i)|}{i}\right)\\
&\le |y(z)|\exp\Big(\gamma^{(d)[v_1]}_{\ell,h}(z,1+\vartheta_1(v_1-1),1+\vartheta_2(v_2-1))(v_1-1)\\
&\qquad + \gamma^{(d)[v_2]}_{\ell,h}(z,1+\vartheta_1(v_1-1),1+\vartheta_2(v_2-1))(v_2-1)\\
&\qquad + \sum_{i\ge 2}\gamma^{(d)[v_1]}_{\ell,h}(z^i,1+\vartheta_1(v_1^i-1),1+\vartheta_2(v_2^i-1))\frac{(v_1^i-1)}{i}\\
&\qquad + \gamma^{(d)[v_2]}_{\ell,h}(z^i,1+\vartheta_1(v_1^i-1),1+\vartheta_2(v_2^i-1))\frac{v_2^i-1}{i}\Big),\end{aligned}$$

$$\left|y^{(d)}_{\ell+1,h}(z,v_1,v_2)\right| \le |y(z)|\exp\left(C^{(d)[v_1]}_{\ell,h}|v_1-1||y(z)|^{\ell+d} + C^{(d)[v_2]}_{\ell,h}|v_2-1||y(z)|^{\ell+d} + \mathcal{O}(L^\ell)\right),$$

where we use that, for $i\ge 2$,

$$\begin{aligned}&|\gamma^{(d)[v_1]}_{\ell,h}(z^i,1+\vartheta_1(v_1^i-1),1+\vartheta_2(v_2^i-1))| \le |\gamma^{(d)[v_1]}_{\ell,h}(z^i,1,1)| \text{ and}\\
&|\gamma^{(d)[v_2]}_{\ell,h}(x^i,1+\vartheta_1(v_1^i-1),1+\vartheta_2(v_2^i-1))| \le |\gamma^{(d)[v_2]}_{\ell,h}(z^i,1,1)|.\end{aligned}$$

By using recursion (2.16) and Lemma 2.20, we obtain

$$\begin{aligned}C^{(d)[v_1]}_{\ell+1,h} &= \sup_{\substack{z\in\Delta_\epsilon\\ v_1\in\Xi_k, v_2\in\Xi_{k+h}}}\left|\frac{y^{(d)}_{\ell+1,h}(z,v_1,v_2)}{y(z)}\right|\left|\frac{\gamma^{(d)[v_1]}_{\ell,h}(z,v_1,v_2)+\sum_{i\ge2}\gamma^{(d)[v_1]}_{\ell,h}(z^i,v_1^i,v_2^i)}{y(z)^{k+d}}\right|\\
&\le \exp\left(C^{(d)[v_1]}_{\ell,h}\frac{\eta_1}{k}+C^{(d)[v_2]}_{\ell,h}\frac{\eta_2}{k}+\mathcal{O}(L^\ell)\right)\left(C^{(d)[v_1]}_{\ell,h}+\mathcal{O}(L^\ell)\right)\\
&= C^{(d)[v_1]}_{\ell,h}\exp\left(C^{(d)[v_1]}_{\ell,h}\frac{\eta_1}{k}+C^{(d)[v_2]}_{\ell,h}\frac{\eta_2}{k}\right)\left(1+\mathcal{O}(L^\ell)\right),\end{aligned}$$

and analogously

$$C^{(d)[v_2]}_{\ell+1,h} \le C^{(d)[v_2]}_{\ell,h}\exp\left(C^{(d)[v_1]}_{\ell,h}\frac{\eta_1}{k}+C^{(d)[v_2]}_{\ell,h}\frac{\eta_2}{k}\right)\left(1+\mathcal{O}(L^\ell)\right).$$

We choose η_1 and η_2 such that $e^{2c_o(C^{(d)[v_1]}_{0,h}\eta_1+C^{(d)[v_2]}_{0,h}\eta_2} \le 2$. Then, by induction we get

$$\begin{aligned}C^{(d)[v_1]}_{\ell,h} &\le C^{(d)[v_1]}_{0,h}\prod_{j<\ell}(1+\mathcal{O}(L^j))e^{2c_o(C^{(d)[v_1]}_{0,h}\eta_1+C^{(d)[v_2]}_{0,h}\eta_2)\frac{\ell}{k}} \le 2C^{(d)[v_1]}_{0,h}c_0 = \mathcal{O}(1),\\
C^{(d)[v_2]}_{\ell,h} &\le C^{(d)[v_2]}_{0,h}\prod_{j<\ell}(1+\mathcal{O}(L^j))e^{2c_o(C^{(d)[v_1]}_{0,h}\eta_1+C^{(d)[v_2]}_{0,h}\eta_2)\frac{\ell}{k}} \le 2C^{(d)[v_2]}_{0,h}c_0 = \mathcal{O}(1).\end{aligned}$$

Note therefore that

$$C_{0,h}^{(d)[v_1]} = \sup_{\substack{z\in\Delta_\epsilon \\ v_1\in\Xi_k, v_2\in\Xi_{k+h}}} \left| \frac{zZ_{d-1}(\mathbf{y_h^{(d)}(z,v_2)})}{y(z)^d} \right| = \mathcal{O}(1)$$

$$C_{0,h}^{(d)[v_1]} = \sup_{\substack{z\in\Delta_\epsilon \\ v_1\in\Xi_k, v_2\in\Xi_{k+h}}} \left| \frac{\gamma_{0,h}^{(d)[v_2]}(z,v_1,v_2) + (v_1-1)z\frac{\partial}{\partial v_2}Z_{d-1}(\mathbf{y_h(z,v_2)})}{y(z)^{h+d}} \right|$$

$$= \mathcal{O}\left(\sup_{\substack{z\in\Delta_\epsilon \\ v_1\in\Xi_k, v_2\in\Xi_{k+h}}} \left| \frac{\gamma_{0,h}^{(d)[v_2]}(z,v_1,v_2)}{y(z)^{h+d}} \right| \right) = \mathcal{O}(1).$$

□

Let

$$\gamma_{k,h}^{(d)[2v_1]}(z,v_1,v_2) := \frac{\partial^2}{\partial v_1^2} y_{k,h}^{(d)}(z,v_1,v_2),$$
$$\gamma_{k,h}^{(d)[2v_2]}(z,v_1,v_2) := \frac{\partial^2}{\partial v_2^2} y_{k,h}^{(d)}(z,v_1,v_2),$$
$$\gamma_{k,h}^{(d)[v_1v_2]}(z,v_1,v_2) := \frac{\partial^2}{\partial v_1 \partial v_2} y_{k,h}^{(d)}(z,v_1,v_2),$$
$$\gamma_{k,h}^{(d)[2]}(z,v_1,v_2) := \gamma_{k,h}^{(d)[2v_1]}(z,v_1,v_2) + \gamma_{k,h}^{(d)[2v_2]}(z,v_1,v_2) + 2\gamma_{k,h}^{(d)[v_1v_2]}(z,v_1v_2).$$

Lemma 2.26. *With the same constants $\epsilon, \theta, \tilde{\eta}_1, \tilde{\eta}_2$ as in Lemma 2.25 and for $|v_1| \le 1, |v_2| \le 1$ and $|z| \le \rho - \eta$ for some $\eta > 0$*

$$\gamma_{k,h}^{(d)[2v_1]}(z,v_1,v_2) = \mathcal{O}(y(|z|)^{k+d})$$
$$\gamma_{k,h}^{(d)[v_1v_2]}(z,v_1,v_2) = \mathcal{O}(y(|z|)^{k+h+2d-1})$$
$$\gamma_{k,h}^{(d)[2v_2]}(z,v_1,v_2) = \mathcal{O}(y(|z|)^{k+h+d})$$

uniformly. Furthermore, for $z \in \Delta_\epsilon, v_1 \in \Xi_k$ and $v_2 \in \Xi_{k+h}$

$$\gamma_{k,h}^{(d)[2]}(z,v_1,v_2) = \mathcal{O}((k+h)y(z)^{k+d})$$

,

Proof. The proof of the statements on the partial derivatives is identical to the one of Lemma 2.21, as we can derive identical recursive relations for $\gamma_{k,h}^{(d)[2v_1]}(z,v_1,v_2)$ and $\gamma_{k,h}^{(d)[2v_2]}(z,v_1,v_2)$ and a similar one for $\gamma_{k,h}^{(d)[v_1v_2]}(z,v_1,v_2)$:

$$\gamma_{k+1,h}^{(d)[v_1v_2]}(z,v_1,v_2) = y_{k+1,h}^{(d)}(z,v_1,v_2)\left(\sum_{i\ge1}\frac{\partial}{\partial v_1}y_{k,h}^{(d)}(z^i,v_1^i,v_2^i)v_1^{i-1}\right)\left(\sum_{i\ge1}\frac{\partial}{\partial v_2}y_{k,h}^{(d)}(z^i,v_1^i,v_2^i)v_2^{i-1}\right)$$
$$+ y_{k+1,h}^{(d)}(z,v_1,v_2)\sum_{i\ge1} i\gamma_{k,h}^{(d)[v_1v_2]}(z^i,v_1^i,v_2^i)v_1^{(i-1)}v_2^{(i-1)}.$$

Then we prove the three statements inductively on k with the following initial conditions

$$\begin{aligned}
&\gamma_{0,h}^{(d)[2v_1]}(z,v_1,v_2) = 0,\\
&\gamma_{0,h}^{(d)[2v_2]}(z,v_1,v_2) \le \gamma_h^{(d)[2]}(z,v_2) = \mathcal{O}(y(z)^{h+d}),
\end{aligned}$$

and with $\frac{\partial}{\partial s_i} Z_n(s_1,\dots,s_n) = \frac{1}{i} Z_{n-i}(s_1,\dots,s_{n-i})$ (cf [48, Chapter 2, page 25])

$$\begin{aligned}
\gamma_{0,h}^{(d)[v_1v_2]}(z,v_1,v_2) &= z\frac{\partial}{\partial v_2} Z_{d-1}(\mathbf{y_h^{(d)}}(\mathbf{z},\mathbf{v_2}))\\
&= \sum_{r=1}^{d-1} \frac{\partial}{\partial s_r} Z_{d-1}(s_1,\dots,s_{d-1})\Bigg|_{s_i=y_h(z^i,v_2^i)} \gamma_h^{(d)}(z^r,v_2^r) r v_2^{r-1}\\
&= \sum_{r=1}^{d-1} \frac{1}{r} Z_{d-r-1}(s_1,\dots,s_{d-1-r})\Bigg|_{s_i=y_h(z^i,v_2^i)} \gamma_h^{(d)}(z^r,v_2^r) r v_2^{r-1}\\
&= \mathcal{O}\left(Z_{d-2}(\mathbf{y_h^{(d)}}(\mathbf{z},\mathbf{v_2}))\gamma_h^{(d)}(z,v_2)\right)\\
&= \mathcal{O}(y(z)^{h+2d-2}).
\end{aligned}$$

For the proof of the last statement we define for $\ell \le k$

$$D_{\ell,h}^{(d)} = \sup_{\substack{z\in\Delta_\epsilon\\ v_1\in\Xi_k, v_2\in\Xi_{k+h}}} \left| \frac{\gamma_{\ell,h}^{(d)[2]}(z,u,v)}{y(z)^{\ell+d}} \right|,$$

as in the proof of the second part of Lemma 2.21. We use the estimate

$$\left|y_{\ell+1,h}^{(d)}(z,v_1,v_2)\right| \le |y(z)| \exp\left(C_{\ell,h}^{(d)[v_1]}\frac{\eta_1}{k} + C_{\ell,h}^{(d)[v_2]}\frac{\eta_2}{k} + \mathcal{O}(L^\ell)\right), \tag{2.39}$$

which we obtained in the proof of Lemma 2.25. From the recursive description, we obtain

$$\begin{aligned}
D_{\ell+1,h}^{(d)} &= \sup_{\substack{z\in\Delta_\epsilon\\ v_1\in\Xi_k, v_2\in\Xi_{k+h}}} \left| \frac{\gamma_{\ell+1,h}^{(d)[2v_1]}(z,v_1,v_2) + \gamma_{\ell+1,h}^{(d)[2v_2]}(z,v_1,v_2) + \gamma_{\ell+1,h}^{(d)[v_1v_2]}(z,v_1v_2)}{y(z)^{k+d+1}} \right|\\
&= \sup_{\substack{z\in\Delta_\epsilon\\ v_1\in\Xi_k, v_2\in\Xi_{k+h}}} \left| \frac{y_{\ell+1,h}^{(d)}(z,v_1,v_2)}{y(z)} \right|\\
&\times \left| \frac{\sum_{r=1}^2 (\sum_{i\ge1} \gamma_{\ell,h}^{(d)[u_r]}(z^i,v_1^i,v_2^i) u_r^{i-1})^2 + \prod_{r=1}^2 (\sum_{i\ge1} \gamma_{\ell,h}^{(d)[u_r]}(z^i,v_1^i,v_2^i) u_r^{i-1})}{y(z)^{\ell+d}} \right.\\
&+ \left. \frac{\sum_{i\ge1} \gamma_{\ell,h}^{(d)[2]}(z,v_1,v_2) + \sum_{r=1}^2 \sum_{i\ge2} (i-1)\gamma_{\ell,h}^{(d)[u_r]}(z^i,v_1^i,v_2^i) u_r^{i-2}}{y(z)^{\ell+d}} \right|
\end{aligned}$$

By applying known bounds from Lemma 2.25 and from the previous statement, and from

(2.38) and (2.39), similar to the proof of Lemma 2.21, we can derive

$$\begin{aligned}
D^{(d)}_{\ell+1,h} &\le \exp\left(C^{(d)[v_1]}_{\ell,h}\frac{\eta_1}{k} + C^{(d)[v_2]}_{\ell,h}\frac{\eta_2}{k} + \mathcal{O}(L^\ell)\right)\\
&\quad\times\left((C^{(d)[v_1]}_{\ell,h})^2|y(z)|^{k+d} + C^{(d)[v_1]}_{\ell,h}C^{(d)[v_2]}_{\ell,h}|y(z)|^{k+h+d} + (C^{(d)[v_2]}_{\ell,h})^2|y(z)|^{k+2h+d} + D^{(d)}_{\ell,h} + \mathcal{O}(L^\ell)\right)\\
&\le D^{(d)}_{\ell,h}\exp\left(C^{(d)[v_1]}_{\ell,h}\frac{\eta_1}{k} + C^{(d)[v_2]}_{\ell,h}\frac{\eta_2}{k}\right)(1+\mathcal{O}(L^\ell))\\
&\quad+\exp\left(C^{(d)[v_1]}_{\ell,h}\frac{\eta_1}{k} + C^{(d)[v_2]}_{\ell,h}\frac{\eta_2}{k}\right)\left((C^{(d)[v_1]}_{\ell,h})^2 + C^{(d)[v_1]}_{\ell,h}C^{(d)[v_2]}_{\ell,h} + (C^{(d)[v_2]}_{\ell,h})^2 + \mathcal{O}(L^\ell))\right)\\
&= D^{(d)}_{\ell,h}\alpha^{(d)}_{\ell,h} + \beta^{(d)}_{\ell,h}.
\end{aligned}$$

As in the proof of Lemma 2.21 we get

$$D^{(d)}_{k,h} \le \alpha_{0,h}D_{0,h} + \sum_{j=0}^{k-1}\beta^{(d)}_{j,h}\prod_{i=j+1}^{k-1}\alpha^{(d)}_{j,h} \tag{2.40}$$

$$= \mathcal{O}(k) + \mathcal{O}(D^{(d)}_{0,h}). \tag{2.41}$$

It remains to prove that $D^{(d)}_{0,h} = \mathcal{O}(h)$:

$$\begin{aligned}
\gamma^{(d)[2v_1]}_{0,h}(z,v_1,v_2) &= 0,\\
\gamma^{(d)[2v_2]}_{0,h}(z,v_1,v_2) &= \gamma^{(d)[2]}_{h}(z,v_2) + \frac{\partial^2}{\partial v_2^2}Z_{d-1}(\mathbf{y^{(d)}_h}(\mathbf{z},\mathbf{v_2}))\\
&= \gamma^{(d)[2]}_{h}(z,v_2) + \sum_{l=1}^{d-1}\sum_{j=1}^{d-l-1}Z_{d-j-l-1}(\mathbf{y^{(d)}_h}(\mathbf{z},\mathbf{v_2}))\gamma^{(d)}_h(z^l,v^l)v^{l-1}\\
&\qquad+\sum_{l=1}^{d-1}Z_{d-l-1}(\mathbf{y^{(d)}_h}(\mathbf{z},\mathbf{v_2}))\gamma^{(d)[2]}_h(z^l,v^l)u^{l-1}\\
&\qquad+\sum_{l=1}^{d-1}Z_{d-l-1}(\mathbf{y^{(d)}_h}(\mathbf{z},\mathbf{v_2}))(l-1)\gamma^{(d)}_h(z^l,v^l)u^{l-2}\\
&= \mathcal{O}(hy(z)^{h+d}) + \mathcal{O}(y(z)^{h+2d-3} + \mathcal{O}(hy(z)^{h+2d-2} + \mathcal{O}(y(z)^{h+2d-2})\\
&= \mathcal{O}(hy(z)^{h+d})
\end{aligned}$$

and

$$\begin{aligned}
\gamma^{(d)[v_1v_2]}_{0,h}(z,v_1,v_2) &= \frac{\partial}{\partial v_2}Z_{d-1}(\mathbf{y^{(d)}_h}(\mathbf{z},\mathbf{v_2})) = \mathcal{O}(y(z)^{h+2d-2})\\
D^{(d)}_{0,h} &= \sup_{\substack{z\in\Delta_\epsilon\\ v_1\in\Xi_k, v_2\in\Xi_{k+h}}}\left|\frac{\gamma^{(d)[2v_1]}_{0,h}(z,v_1,v_2) + \gamma^{(d)[2v_2]}_{0,h}(z,v_1,v_2) + \gamma^{(d)[v_1v_2]}_{0,h}(z,v_1,v_2)}{y(z)^d}\right|\\
&\le \sup_{\substack{z\in\Delta_\epsilon\\ v_1\in\Xi_k, v_2\in\Xi_{k+h}}}\mathcal{O}(hy(z)^h + y(z)^{h+2d-2}) = \mathcal{O}(h).
\end{aligned}$$

□

Remark. Obviously, for $z \in \Delta_\epsilon$ and $v_1 \in \Xi_k, v_2 \in \Xi_{k+h}$ the latter statement also holds for the partial derivatives:

$$\begin{aligned}\gamma_{k,h}^{(d)[2v_1]}(z,v_1,v_2) &= \mathcal{O}(ky(z)^{k+d}),\\ \gamma_{k,h}^{(d)[2v_2]}(z,v_1,v_2) &= \mathcal{O}((k+h)y(z)^{k+h+d}),\\ \gamma_{k,h}^{(d)[v_1v_2]}(z,v_1,v_2) &= \mathcal{O}((k+h)y(z)^{k+d}).\end{aligned}$$

Lemma 2.27. *For $z \in \Delta_\epsilon, v_1 \in \Xi_k$ and $v_2 \in \Xi_{k+h}$, with the same constants as in the previous lemmas, we can approximate*

$$\begin{aligned}w_{k,h}^{(d)}(z,v_1,v_2) =&C_k^{(d)}(z)(v_1-1)y(z)^{k+d} + C_{k+h}^{(d)}(z)(v_2-1)y(z)^{k+h+d}\\ &+ \mathcal{O}((k+h)y(z)^{k+d}(|v_1-1|^2+|v_2-1|^2)).\end{aligned}$$

Furthermore

$$\begin{aligned}\Sigma_{k,h}^{(d)}(z,v_1,v_2) =&\tilde{C}_k^{(d)}(z^2)(v_1-1)y(z^2)^{k+d} + \tilde{C}_{k+h}^{(d)}(z)(v_2-1)y(z^2)^{k+h+d}\\ &+ \mathcal{O}(y(|z|^2)^k|v_1-1|^2 + y(|z|^2)^{k+h}|v_2-1|^2).\end{aligned}$$

Proof. For the first statement, we expand $w_{k,h}^{(d)}(z,v_1,v_2)$ into a Taylor polynomial of degree 2 around $v_1 = v_2 = 1$ and obtain

$$\begin{aligned}w_{k,h}^{(d)}(z,v_1,v_2) &= \gamma_k^{(d)}(z)(v_1-1) + \gamma_{k+h}^{(d)}(z)(v_2-1) + R\\ \text{with}\quad |R| &\le \frac{1}{2}\Big(\gamma_{k,h}^{(d)[2v_1]}(z,1+\vartheta_1(v_1-1),1+\vartheta_2(v_2-1))(v_1-1)^2\\ &\quad + 2\gamma_{k,h}^{(d)[v_1v_2]}(z,1+\vartheta_1(v_1-1),1+\vartheta_2(v_2-1))(v_1-1)(v_2-1)\\ &\quad +\gamma_{k,h}^{(d)[2v_2]}(z,1+\vartheta_1(v_1-1),1+\vartheta_2(v_2-1))(v_2-1)^2\Big).\end{aligned}$$

Hence,

$$\begin{aligned}w_{k,h}^{(d)}(z,v_1,v_2) = &C_k^{(d)}(z)(v_1-1)y(z)^{k+d} + C_{k+h}^{(d)}(z)(v_2-1)y(z)^{k+h+d}\\ &+ \mathcal{O}((k+h)y(z)^{k+d}(|v_1-1|^2+|v_2-1|^2)),\end{aligned}$$

where we can neglect the mixed derivatives as either $(v_1-1)^2$ or $(v_2-1)^2$ will determine the dominant part of the error term. For the second part we again use a Taylor polynomial, using the fact that $|z^i| < \rho < 1$ and $|u_r^i - 1| \le i|u_r - 1|$ for $i > 2, r = 1,2$, hence the result follows immediately. □

Note that the terms $v_1 - 1$ and $v_2 - 1$ are asymptotically proportional: $\frac{v_2}{v_1} = \frac{e^{\frac{it_1}{\sqrt{n}}}-1}{e^{\frac{it_2}{\sqrt{n}}}-1} \sim \frac{t_2}{t_1}$, and that $y(z^2)^{k+h+d}$ is exponentially smaller than $y(z^2)^{k+d}$ as $h = \xi\sqrt{n}$.

Lemma 2.28. *There exist constants $\epsilon, \theta, \tilde{\eta}_1, \tilde{\eta}_2$ such that $w_{k,h}^{(d)} = w_{k,h}^{(d)}(z,v_1,v_2)$ is given by*

$$w_{k,h}^{(d)} = \frac{w_{0,h}^{(d)}y(z)^k}{1 - f_k^{(d)} - \frac{w_{0,h}^{(d)}}{2}\frac{1-y(z)^k}{1-y(z)} + \mathcal{O}(|v_1-1| + |v_2-1|}$$

for $v_1 \in \Xi_k(\tilde\eta_1), v_2 \in \Xi_{k+h}(\tilde\eta_2)$ *and* $z \in \Delta_\epsilon(\theta)$*, where* $f_k^{(d)}$ *is given by*

$$f_k^{(d)}(z,v_1,v_2) = w_{0,h}^{(d)}(z,v_1,v_2)\sum_{l=0}^{k-1}\frac{\Sigma_{l,h}(z,v_1,v_2)y(z)^{l+1}}{w_{l,h}^{(d)}(z,v_1,v_2)w_{l+1,h}(z,v_1,v_2)}. \tag{2.42}$$

Proof. We can argue similarly as in the proof of Lemma 2.23 and derive the recursive description

$$w_{k+1,h}^{(d)} = yw_{k,h}^{(d)}\left(1+\frac{\Sigma_{k,h}^{(d)}}{w_{k,h}^{(d)}}\right)\left(1+\frac{w_{k,h}^{(d)}}{2}+\mathcal{O}(w_{k,h}^{2(d)})+\mathcal{O}(\Sigma_{k,h}^{(d)})\right),$$

and equivalently

$$\frac{y}{w_{k+1,h}^{(d)}}\cdot\left(1+\frac{\Sigma_{k,h}^{(d)}}{w_{k,h}^{(d)}}\right) = \frac{1}{w_{k,h}^{(d)}}-\frac{1}{2}+\mathcal{O}(w_{k,h}^{(d)})+\mathcal{O}\left(\frac{\Sigma_{k,h}^{(d)}}{w_{k,h}^{(d)}}\right).$$

Further we get

$$\frac{y^{k+1}}{w_{k+1,h}^{(d)}} = \frac{y^k}{w_{k,h}^{(d)}}-\frac{\Sigma_{k,h}^{(d)}\cdot y(z)^{k+1}}{w_{k,h}^{(d)}w_{k+1,h}^{(d)}}-\frac{1}{2}y(z)^k+\mathcal{O}(w_{k,h}^{(d)}y^k)+\mathcal{O}\left(\frac{\Sigma_{k,h}^{(d)}\cdot y^k}{w_{k,h}^{(d)}}\right).$$

Solving the recurrence leads to

$$\frac{y^k}{w_{k,h}^{(d)}} = \frac{1}{w_{0,h}^{(d)}}-\sum_{l=0}^{k-1}\frac{\Sigma_{l,h}^{(d)}\cdot y(z)^{l+1}}{w_{l,h}^{(d)}w_{l+1,h}^{(d)}}-\frac{1}{2}\frac{1-y^k}{1-y}+\mathcal{O}\Big(\sum_{l=0}^{k-1}\underbrace{w_{\ell,h}^{(d)}y^\ell}_{=\mathcal{O}(w_{0,h}y^{2\ell})}\Big)+\mathcal{O}\Big(\sum_{l=0}^{k-1}\underbrace{\frac{\Sigma_{l,h}^{(d)}\cdot y^l}{w_{l,h}^{(d)}}}_{=\mathcal{O}(L^l)}\Big)$$

$$= \frac{1}{w_{0,h}^{(d)}}\left(1-\underbrace{w_{0,h}^{(d)}\sum_{l=0}^{k-1}\frac{\Sigma_{l,h}^{(d)}y(z)^{l+1}}{w_{l,h}^{(d)}w_{l+1,h}^{(d)}}}_{=:f_k^{(d)}(z,v_1,v_2)}-\frac{w_{0,h}^{(d)}}{2}\frac{1-y^k}{1-y}+\mathcal{O}(w_{0,h}^{2(d)}\frac{1-y^{2k}}{1-y^2})+\mathcal{O}(w_{0,h}^{(d)}\frac{1-L^k}{1-L})\right).$$

Finally, observe that

$$\begin{aligned} w_{0,h}^{(d)} &= y_h^{(d)}(z,v)+(u-1)zZ_{d-1}(\mathbf{y_h^{(d)}}(\mathbf{z},\mathbf{v}))-y(z)\\ &= w_h^{(d)}+(u-1)zZ_{d-1}(\mathbf{y_h^{(d)}}(\mathbf{z},\mathbf{v}))\\ &= C_h^{(d)}(z)(v_2-1)y(z)^{h+d}+(v_1-1)y(z)^d = \mathcal{O}(|v_1-1|+|v_2-1|). \end{aligned}$$

□

Proof of Theorem 2.24. In the following, we denote by $U := (v_1-1)y(z)^d$ and $W := w_h^{(d)}(z,v_2)$. Note that $w_{0,h} \sim U+W$. By Lemma 2.27 we obtain for $w_{\ell,h}^{(d)}$, $0 \geq \ell \geq k-1$

(note that $C^{(d)}_{\ell+h}(z) = C^{(d)}_h(z)(1+L^\ell)$)

$$\begin{aligned}
w^{(d)}_{\ell,h}(z,v_1,v_2) &= C^{(d)}_\ell(z)(v_1-1)y(z)^{\ell+d} + C^{(d)}_{\ell+h}(z)(v_2-1)y(z)^{\ell+h+d} \\
&\quad + \mathcal{O}\left((\ell+h)y(z)^\ell(|v_1-1|^2+|v_2-1|^2)\right) \\
&= y(z)^\ell\left(C^{(d)}_\ell U + C^{(d)}_h(z)(v_2-1)y(z)^{h+d}\right. \\
&\quad \left.+\mathcal{O}\left((\ell+h)(|v_1-1|^2+|v_2-1|^2)\right)\right) \\
&= y(z)^\ell(C^{(d)}_\ell(z)U+W)(1+\mathcal{O}(h(|v_1-1|+|v_2-1|))
\end{aligned}$$

We use representation (2.32) for $\tilde{C}_\ell(z)$, which we already used in the proof of Lemma 2.23, and omit all error terms, to obtain by telescoping

$$\begin{aligned}
f^{(d)}_k(z,v_1,v_2) &= w^{(d)}_{0,h}(z,v_1,v_2)\sum_{\ell=0}^{k-1}\frac{y(z)^{\ell+1}\left(\tilde{C}^{(d)}_\ell(z)(v_1-1)y(z^2)^{k+d}\right)}{y(z)^\ell(C^{(d)}_\ell(z)U+W)y(z)^{\ell+1}(C^{(d)}_{\ell+1}(z)U+W)} \\
&= Uw^{(d)}_{0,h}(z,v_1,v_2)\sum_{l=0}^{k-1}\frac{\tilde{C}^{(d)}_l(z)\left(\frac{y(z^2)}{y(z)}\right)^{l+d}}{(C^{(d)}_l(z)U+W)(C^{(d)}_{l+1}(z)U+W)} \\
&= w^{(d)}_{0,h}(z,v_1,v_2)\sum_{l=0}^{k-1}\frac{(C^{(d)}_{l+1}(z)U+W)-(C^{(d)}_l(z)U+W)}{(C^{(d)}_l(z)U+W)(C^{(d)}_{l+1}(z)U+W)} \\
&= w^{(d)}_{0,h}(z,v_1,v_2)\left(\frac{1}{C^{(d)}_0(z)U+W}-\frac{1}{C^{(d)}_k(z)U+W}\right).
\end{aligned}$$

As we know from (2.23), $C^{(d)}_0 = \frac{zZ(S_{d-1})}{y(z)^d} = \mathcal{O}(1)$ near $u=1$ (analytic), hence

$$f^{(d)}_k(z,u,v) \sim \left(1-\frac{(U+W)}{C^{(d)}_k(z)U+W}\right).$$

Using

$$\begin{aligned}
C^{(d)}_k(z) &\sim C_d\rho^d \\
(v_1-1) &\sim \frac{it_1}{\sqrt{n}} \\
y(z)^k &\sim e^{-\kappa b\sqrt{-\rho s}} \\
1-y(z) &\sim b\sqrt{\frac{\rho s}{n}}
\end{aligned}$$

and $w^{(d)}_{0,h}(z,v_1,v_2) \sim W+U$, we can derive

$$
\begin{aligned}
w_{k,h}^{(d)} &= \frac{w_{0,h}^{(d)} y(z)^k}{\frac{((U+W)}{C_d\rho^d U+W} - \frac{w_{0,h}^{(d)}}{2}\frac{1-y(z)^k}{1-y(z)}} \\
&= \frac{(C_d\rho^d \frac{it_1}{\sqrt{n}} + w_h^{(d)}(z,v))\sqrt{-s}e^{-\kappa b\sqrt{-\rho s}}}{\sqrt{-s} - ((C_d\rho^d \frac{it_1}{\sqrt{n}} + w_h^{(d)}(z,v))\frac{1}{2b\sqrt{\frac{\rho}{n}}}(1-e^{-\kappa b\sqrt{-\rho s}})} \\
&= \frac{C_d\rho^d}{\sqrt{n}} \frac{(it_1 + w_h^{(d)}(z,v))\sqrt{-s}e^{-\frac{\kappa}{2}b\sqrt{-\rho s}}}{\sqrt{-s}e^{\frac{\kappa}{2}b\sqrt{-\rho s}} - ((C_d\rho^d \frac{it_1}{\sqrt{n}} + w_h^{(d)}(z,v))\frac{1}{b\sqrt{\frac{\rho}{n}}}(\sinh(\frac{\kappa}{2}b\sqrt{-\rho s}))},
\end{aligned}
$$

and with the expansion (2.21) of $w_h^{(d)}(z, v_2)$ with $v_2 = e^{\frac{it_2}{\sqrt{n}}}$ and $h = \eta\sqrt{n}$, given by Theorem 2.15, we can derive the expansion given in Theorem 2.24. □

Proof of Theorem 2.13. The characteristic function of the two dimensional distribution is given by

$$
\begin{aligned}
\phi_{k,k+h,n}^{(d)}(t_1,t_2) &= \frac{1}{y_n}[z^n]y_{k,h}^{(d)}(z, e^{\frac{it_1}{\sqrt{n}}}, e^{\frac{it_2}{\sqrt{n}}}) \\
&= \frac{1}{2\pi i y_n}\int_\Gamma y_{k,h}^{(d)}(z, e^{\frac{it_1}{\sqrt{n}}}, e^{\frac{it_2}{\sqrt{n}}})\frac{dz}{z^{n+1}} \\
&= 1 + \frac{1}{2\pi i y_n}\int_\Gamma w_{k,h}^{(d)}(z, e^{\frac{it_1}{\sqrt{n}}}, e^{\frac{it_2}{\sqrt{n}}})\frac{dz}{z^{n+1}}. \qquad (2.43)
\end{aligned}
$$

We use the same contour as in the one dimensional case. With the same arguments, only integration over γ contributes to the result, hence the representation (2.37) of $w_{k,h}^{(d)}$ leads to:

$$
\begin{aligned}
&\phi_{k,h,n}^{(d)}(t_1,t_2) = 1 + \frac{\sqrt{2}}{\sqrt{\pi}i} \\
&\times \int_{1-i\log^2 n}^{1+i\log^2 n} \frac{\frac{C_d\rho^d}{b\sqrt{2\rho}}i\left(t_1 + \frac{t_2\sqrt{-s}e^{(-\frac{1}{2}\eta b\sqrt{-\rho s})}}{\sqrt{-s}e^{(\frac{1}{2}\eta b\sqrt{-\rho s})} - \frac{it_2 C_d\rho^d}{\sqrt{\rho}b}\sinh\left(\frac{1}{2}\eta b\sqrt{-\rho s}\right))}\right)\sqrt{-s}e^{(-\frac{1}{2}\kappa b\sqrt{-\rho s})}}{\sqrt{-s}e^{(-\frac{\kappa}{2}b\sqrt{-\rho s})} - i\frac{C_d\rho^d}{2b\sqrt{\rho}}\left(t_1 + \frac{t_2\sqrt{-s}e^{(-\frac{1}{2}\eta b\sqrt{-\rho s})}}{\sqrt{-s}e^{(\frac{1}{2}\eta b\sqrt{-\rho s})} - \frac{it_2 C_d\rho^d}{\sqrt{\rho}b}\sinh\left(\frac{1}{2}\eta b\sqrt{-\rho s}\right))}\right)(\sinh\left(\frac{\kappa}{2}b\sqrt{-\rho s}\right))} \\
&\xrightarrow{n\to\infty} \psi_{\kappa,\xi}(t_1,t_2)
\end{aligned}
$$

where $\psi_{\kappa,\xi}(t_1,t_2)$ is the characteristic function of the random variable $\frac{C_d\rho^d}{\sqrt{2\rho}b}\left(l\left(\frac{b\sqrt{\rho}}{2\sqrt{2}}\kappa, \frac{b\sqrt{\rho}}{2\sqrt{2}}\xi\right)\right)$. □

2.3.3 Tightness

We must show the estimate (2.14) in Theorem 2.14. The fourth moment in (2.14) can be obtained by by applying the operator $\left(v\frac{\partial}{\partial v}\right)^4$ and setting $v = 1$ afterwards. Hence, using the transfer lemma of Flajolet and Odlyzko [23] it turns out that it suffices to show that

$$\left[\left(\frac{\partial}{\partial v}+7\frac{\partial^2}{\partial v^2}+6\frac{\partial^3}{\partial v^3}+\frac{\partial^4}{\partial v^4}\right)\tilde{y}_{r,h}\left(z,v,v^{-1}\right)\right]_{v=1}=\mathcal{O}\left(\frac{h^2}{1-|y(z)|}\right) \tag{2.44}$$

uniformly for $z\in\Delta$ and $h\geq 1$ (see [21, pp.2046] for the detailed argument).

Set

$$\gamma_k^{(d)[j]}(z)=\left[\frac{\partial^j y_k(z,v)}{\partial v^j}\right]_{v=1}\qquad\text{and}\qquad \gamma_{k,h}^{(d)[j]}(z)=\left[\frac{\partial^j \tilde{y}_{r,h}\left(z,v,\frac{1}{v}\right)}{\partial v^j}\right]_{v=1}.$$

The left-hand side of (2.44) is a linear combination of $\gamma_{k,h}^{(d)[j]}(z)$ for $j=1,2,3,4$. Therefore we need bound for those quantities. We will derive upper bounds for all j since this more general result is easier to achieve. We start with an auxiliary result.

Lemma 2.29. *Let j be a positive integer. Under the assumption that for all $i\leq j$ the bound $\gamma_k^{(d)[i]}(z)=\mathcal{O}\left(|z/\rho|^k\right)$ holds uniformly for $|z|\leq\rho$, we have $\left[\left(\frac{\partial}{\partial v}\right)^j\Sigma_k^{(d)}\right]_{v=1}=\mathcal{O}\left(L^k\right)$ for some positive constant $L<1$.*

Proof. By Faà di Bruno's formula we have

$$\begin{aligned}\left[\left(\frac{\partial}{\partial v}\right)^j\Sigma_k^{(d)}\right]_{v=1}&=\sum_{i\geq 2}\frac{1}{i}\left[\left(\frac{\partial}{\partial v}\right)^j w_k^{(d)}(z^i,v^i)\right]_{v=1}\\&=\sum_{i\geq 2}\frac{1}{i}\sum_{\sum_{m=1}^j m\nu_m=j}\frac{j!}{\nu_1!\cdots\nu_j!}\gamma_k^{(d)[\nu_1+\cdots+\nu_m]}(z^i,1)\prod_{\lambda=1}^j\left(\frac{1}{\lambda!}\left[\left(\frac{\partial}{\partial v}\right)^\lambda v^i\right]_{v=1}\right)^{\nu_\lambda}.\end{aligned}$$

By our assumption we have $\gamma_k^{(d)[\nu_1+\cdots+\nu_m]}(z^i,1)=\mathcal{O}\left(|z^i/\rho|^k\right)$. The product is essentially a derivative of order $j=\sum\lambda\nu_\lambda$ of v^i and can therefore be estimate by $\mathcal{O}\left(i^j\right)$. So the whole expression is bounded by a constant times $\sum_{i\geq 2}i^{j-1}z^{ik}/\rho^k=\mathcal{O}\left((|z^2|/\rho)^k i\right)=\mathcal{O}\left((\rho+\varepsilon)^k\right)$. Hence we can choose $L=\rho+\varepsilon$ to get the desired bound. ☐

Exactly the same line of arguments yield the analogous result for two levels.

Lemma 2.30. *Let j be a positive integer and set*

$$\tilde{\Sigma}_{k,h}^{(d)}=\sum_{i\geq 2}\frac{1}{i}w_{k,h}^{(d)}(z^i,v^i,v^{-i}). \tag{2.45}$$

Under the assumption that for all $i\leq j$ the bound $\gamma_{k,h}^{(d)[i]}(z)=\mathcal{O}\left(|z/\rho|^k\right)$ holds uniformly for $|z|\leq\rho$ we have $\left[\left(\frac{\partial}{\partial v}\right)^j\tilde{\Sigma}_{k,h}^{(d)}\right]_{v=1}=\mathcal{O}\left(L^k\right)$ for some positive constant $L<1$.

With the auxiliary lemmas we can easily get bounds for $\gamma_k^{(d)[j]}(z)$ and $\gamma_{k,h}^{(d)[j]}(z)$.

Lemma 2.31. *We have*

$$\gamma_k^{(d)[1]}(z)=\begin{cases}\mathcal{O}(1) & \textit{uniformly for } z\in\Delta,\\ \mathcal{O}\left(|z/\rho|^k\right) & \textit{uniformly for } |z|\leq\rho\end{cases} \tag{2.46}$$

and for $\ell>1$

$$\gamma_k^{(d)[\ell]}(z)=\begin{cases}\mathcal{O}\left(\min\left(k^{\ell-1},\frac{k^{\ell-2}}{1-|y(z)|}\right)\right) & \textit{uniformly for } z\in\Delta,\\ \mathcal{O}\left(|z/\rho|^k\right) & \textit{uniformly for } |z|\leq\rho.\end{cases} \tag{2.47}$$

Proof. The estimate (2.46) essentially follows from Lemma 2.19: We know

$$\gamma_k^{(d)[1]}(z) = C^{(d)}(z)y(z)^{k+d}\left(1+\mathcal{O}\left(L^k\right)\right) = \mathcal{O}\left(1\right)$$

with some $0 < L < 1$ and $|y(z)| \leq 1$ and this is sufficient to show the first part of (2.46).

If $|z| \leq \rho$ we can exploit the convexity of $y(z)$ on the positive real line to get $|y(z)| \leq |z/\rho|$. This implies $\gamma_k^{(d)[1]}(z) = \mathcal{O}\left(|z/\rho|^{k+d}\right)$, an even better bound than stated in the assertion.

Now we are left with the induction step. Again we use Faà di Bruno's formula and the fact that $w_k^{(d)}(z,1) = \Sigma_k^{(d)}(z,1) = 0$ and obtain

$$\begin{aligned}
\gamma_k^{(d)[\ell]}(z) &= \left[\frac{\partial}{\partial v} w_k^{(d)}(z,v)\right]_{v=1} = y(z)\left[\frac{\partial}{\partial v}\exp\left(w_{k-1}^{(d)}(z,v) + \Sigma_{k-1}^{(d)}(z,v)\right)\right]_{v=1} \\
&= \sum_{\sum_{i=1}^{\ell} i\lambda_i = \ell} \frac{\ell!}{\lambda_1!\cdots\lambda_\ell!}\prod_{j=1}^{\ell-1}\left(\frac{1}{j!}\left[\left(\frac{\partial}{\partial v}\right)^j\left(w_{k-1}^{(d)}(z,v) + \Sigma_{k-1}^{(d)}(z,v)\right)\right]_{v=1}\right)^{\lambda_j} \\
&\quad + y(z)\left[\left(\frac{\partial}{\partial v}\right)^\ell\left(w_{k-1}^{(d)}(z,v) + \Sigma_{k-1}^{(d)}(z,v)\right)\right]_{v=1} \\
&= \sum_{\sum_{i=1}^{\ell} i\lambda_i = \ell} \frac{\ell!}{\lambda_1!\cdots\lambda_\ell!}\prod_{j=1}^{\ell-1}\left(\frac{\gamma_{k-1}^{(d)[j]}(z) + \Gamma_{k-1}^{(d)[j]}(z)}{j!}\right)^{\lambda_j} + y(z)(\gamma_{k-1}^{(d)[\ell]}(z) + \Gamma_{k-1}^{(d)[\ell]},
\end{aligned} \tag{2.48}$$

where $\Gamma_{k-1}^{(d)[\ell]} = \left[\left(\frac{\partial}{\partial v}\right)^\ell \Sigma_{k-1}^{(d)}(z,v)\right]_{v=1}$.

Consider the case $|z| \leq \rho$. The product comprises only terms which essentially have the form $\gamma_{k-1}^{(d)[j]}(z) + \Gamma_{k-1}^{(d)[j]}(z)$ with $j < \ell$. Thus by the induction hypothesis, $\gamma_{k-1}^{(d)[j]}(z) = \mathcal{O}\left(|z/\rho|^j\right)$. Therefore the assumption of Lemma 2.29 is satisfied and the terms as a whole are bounded by $C \cdot |z/\rho|^j$. Since $\sum_{j=1}^{\ell-1} j\lambda_j = \ell$ we get

$$\gamma_k^{(d)[\ell]}(z) = y(z)(\gamma_{k-1}^{(d)[\ell]}(z) + \Gamma_{k-1}^{(d)[\ell]} + \mathcal{O}\left(|z/\rho|^\ell\right).$$

So we finally get the desired estimate by induction on k and Lemma 2.29, starting with

$$\gamma_0^{(d)[\ell]} = \begin{cases} zZ_{d-1}(\mathbf{y}(\mathbf{z})) & \text{if } \ell = 1, \\ 0 & \text{else.} \end{cases} \tag{2.49}$$

Now let us turn to general $z \in \Delta$. Like before we focus first on the terms of the product of (2.48). Again the induction hypothesis guarantees that the assumption of Lemma 2.29 is satisfied and so $\Gamma_{k-1}^{(d)[j]}(z)$ is exponentially small. Furthermore, the induction hypothesis implies $\gamma_{k-1}^{(d)[j]}(z) = \mathcal{O}\left(\min\left(k^{j-1}, \frac{k^{j-2}}{1-|y(z)|}\right)\right)$. Since $\gamma_{k-1}^{(d)[1]}(z) = \mathcal{O}\left(1\right)$ this implies

$$\prod_{j=1}^{\ell-1}\left(\frac{\gamma_{k-1}^{(d)[j]}(z) + \Gamma_{k-1}^{(d)[j]}(z)}{j!}\right)^{\lambda_j} = \mathcal{O}\left(\min\left(k^{\sum_{j=1}^{\ell-1}(j-1)\lambda_j}, \frac{k^{\sum_{j=2}^{\ell-1}(j-2)\lambda_j}}{(1-|y(z)|)^{\sum_{j=2}^{\ell-1}\lambda_j}}\right)\right). \tag{2.50}$$

Set

$$A = k^{\sum_{j=1}^{\ell-1}(j-1)\lambda_j} \text{ and } B = \frac{k^{\sum_{j=2}^{\ell-1}(j-2)\lambda_j}}{(1-|y(z)|)^{\sum_{j=2}^{\ell-1}\lambda_j}}.$$

Note that $\sum_{j=1}^{\ell-1}(j-1)\lambda_j = \ell - \sum_{j=1}^{\ell-1}\lambda_j$. Since the term corresponding to $\lambda_\ell = 0$ in Faà di Bruno formula is the very last term in (2.48), we must have $\sum_{j=1}^{\ell-1}\lambda_j \geq 2$ and thus $A \leq k^{\ell-2}$. Moreover, we have

$$\sum_{j=2}^{\ell-1}(j-2)\lambda_j = \ell - k_1 - 2\sum_{j=2}^{\ell-1}k_j \leq \ell - 3$$

since $\sum_{j=2}^{\ell-1}\lambda_j < 2$ implies $k_1 > 0$, and in particular if $\sum_{j=2}^{\ell-1}\lambda_j = 0$ then $k_1 = \ell$. Therefore

$$B \leq \frac{k^{\ell-3}}{(1-|y(z)|)^{\sum_{j=2}^{\ell-1}\lambda_j}}.$$

We want to show that

$$B \leq \frac{k^{\ell-3}}{1-|y(z)|}. \tag{2.51}$$

Set $A_j = k^{j-1}$ and $B_j = k^{j-2}/(1-|y(z)|)$. Note that $B_j < A_j$ is equivalent to ${}^1/(1-|y(z)|) < k$. Therefore the term B appears in our upper bound (2.50) if and only if z is such that $1/(1-|y(z)|) < k$. But this implies that $B \leq \frac{k^{\ell-3}}{1-|y(z)|}$ as desired, because the desired bound is equivalent to

$$\frac{1}{1-|y(z)) < k^{\left(-2+2\sum_{j=2}^{\ell-1}\lambda_j\right)/\left(-1+\sum_{j=2}^{\ell-1}\lambda_j\right)}} = k^{1+\alpha}$$

where $\alpha = \sum_{j=1}^{\ell-1}\lambda_j - 1 > 0$ and hence the desired bound (2.51) is weaker than $1/(1-|y(z)|) < k$.

Now let $a_k := \gamma_k^{(d)[\ell]}(z)$. We have shown so far that

$$a_k = y(z)a_{k-1} + y(z)A_k \text{ with } A_k = \mathcal{O}\left(\min\left(k^{\ell-2}, \frac{k^{\ell-3}}{1-|y(z)|}\right)\right)$$

and we know that a_0 is given by (2.49). Solving this recurrence relation gives

$$a_k = y(z)^k a_0 + \mathcal{O}\left(\left|y(z)\frac{1-y(z)^k}{1-y(z)}\right| \cdot \min\left(k^{\ell-2}, \frac{k^{\ell-3}}{1-|y(z)|}\right)\right).$$

Since $\left|y(z)\frac{1-y(z)^k}{1-y(z)}\right| \leq k$ and $a_0 y(z)^k = \mathcal{O}\left(y(z)^{k+d}\right) = \mathcal{O}(1)$ we get the desired bound for a_k and the proof is complete. □

Lemma 2.32. *We have*

$$\gamma_{k,h}^{(d)[1]}(z) = \begin{cases} \mathcal{O}(1) & \textit{uniformly for } z \in \Delta, \\ \mathcal{O}\left(|z/\rho|^k\right) & \textit{uniformly for } |z| \leq \rho, \end{cases} \tag{2.52}$$

and for $\ell > 1$

$$\gamma_{k,h}^{(d)[\ell]}(z) = \begin{cases} \mathcal{O}\left(\min\left(k^{\ell-1}, \frac{k^{\ell-2}}{1-|y(z)|}\right)\right) & \textit{uniformly for } z \in \Delta, \\ \mathcal{O}\left(|z/\rho|^k\right) & \textit{uniformly for } |z| \leq \rho. \end{cases}$$

Proof. As the bounds are precisely like in the previous lemma, the induction step works in an analogous way, using Lemma 2.30 instead of Lemma 2.29. Thus we only have to show the initial step of the induction, Eq. (2.52).

We can use a similar reasoning as in the proof of [21, Lemma 7]. Indeed, by applying the operator $\left[\frac{\partial}{\partial v}\cdot\right]_{v=1}$ to (2.16) we obtain the recurrence relation

$$\gamma_{k+1,h}^{(d)[1]}(z) = y(z)\sum_{i\geq 1}\gamma_{k,h}^{(d)[1]}(z^i)$$

with initial value $\gamma_{0,h}^{(d)[1]}(z) = zZ_{d-1}(\mathbf{y(z)}) - \gamma_h(x)$. Induction on k gives the representation $\gamma_{k,h}^{(d)[1]}(z) = \gamma_k^{(d)[1]}(z) - \gamma_{k+h}^{(d)[1]}(z)$ and using $\gamma_k^{(d)[1]}(z) = C^{(d)}(z)y(z)^{k+d}(1+\mathcal{O}\left(L^k\right))$ from Lemma 2.19 we obtain

$$\gamma_k^{(d)[1]}(z) = \mathcal{O}\left(\sup_{z\in\Delta}|y(z)^{k+d}(1-y(z)^h)| + L^k\right) = \mathcal{O}\left(\frac{h}{k+d+h}\right).$$

Since the last term is bounded the proof is complete. □

Now, applying Lemma 2.32 to (2.44) proves tightness and Theorem 2.14 after all.

2.3.4 The joint distribution of two degrees

We want to gain knowledge on the correlation between two different degrees d_1, d_2 in a certain level $k = \kappa\sqrt{n}$.

The covariance $\mathbb{C}\mathrm{ov}(X_n^{(d_1)}(k), X_n^{(d_2)}(k))$

The covariance of two random variables X and Y is given by

$$\mathbb{C}\mathrm{ov}(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y).$$

In this section, we will prove the following result on the covariance function of the two random variables $X_n^{(d_1)}(k)$ and $X_n^{(d_2)}(k)$, counting the vertices of degree d_1 and d_2, respectively, on level k.

Proposition 2.33. *The covariance $\mathbb{C}\mathrm{ov}(X_n^{(d_1)}(k), X_n^{(d_2)}(k))$ of random variables $X_n^{(d_1)}(k)$ and $X_n^{(d_2)}(k)$ counting vertices of degrees d_1 and d_2, with $d_1 \neq d_2$ fixed, at level $k = \kappa n$ in a random Pólya tree of size n is asymptotically given by*

$$\mathbb{C}\mathrm{ov}(X_n^{(d_1)}(k), X_n^{(d_2)}(k)) = C_{d_1}C_{d_2}\rho^{d_1+d_2}n\left(\frac{2}{b^2\rho}\left(e^{-\frac{\kappa^2b^2\rho}{4}} + e^{-\kappa^2b^2\rho}\right) - \kappa^2 e^{-\frac{\kappa^2b^2\rho}{2}}\right)\left(1+\mathcal{O}\left(\frac{1}{\sqrt{n}}\right)\right), \tag{2.53}$$

as n tends to infinity.

Let $y_k^{(d_1d_2)}(z,v_1,v_2)$ be the generating function of Pólya trees where all vertices of degrees d_1 and d_2 on level k are marked and counted by v_1 and v_2, respectively. The $y_k^{(d_1d_2)}(z,v_1,v_2)$ are given by a similar recursion as (2.15), namely

$$\begin{aligned} y_0^{(d_1d_2)}(z,v_1,v_2) &= y(z) + (v_1-1)zZ_{d_1-1}(y(z),y(z^2),\dots,y(z^{d-1})) \\ &\quad + (v_2-1)zZ_{d_2-1}(y(z),y(z^2),\dots,y(z^{d-1})) \\ y_{k+1}^{(d_1d_2)}(z,v_1,v_2) &= z\exp\left(\sum_{i\geq 1}\frac{y_k^{(d_1d_2)}(z^i,v_1^i,v_2^i)}{i}\right). \end{aligned} \tag{2.54}$$

To compute $\mathbb{E}\left(X_n^{(d_1)}(k) \cdot X_n^{(d_2)}(k))\right)$ we need to determine

$$\frac{1}{y_n}[z^n]\left[\frac{\partial^2}{\partial v_1 v_2} y_k^{(d_1 d_2)}(z, v_1, v_2)\right]_{v_1=v_2=1},$$

while $\mathbb{E}(X_n^{(d_1)}(k))$ and $\mathbb{E}(X_n^{(d_2)}(k))$ are given by

$$\frac{1}{y_n}[z^n]\gamma_k^{(d_1)}(z) \text{ and } \frac{1}{y_n}[z^n]\gamma_k^{(d_2)}(z), \text{respectively.}$$

We use the notations

$$\gamma_k^{(d_1)}(z, v_1, v_2) = \frac{\partial}{\partial v_1} y_k^{(d_1 d_2)}(z, v_1, v_2), \quad \gamma_k^{(d_2)}(z, v_1, v_2) = \frac{\partial}{\partial v_2} y_k^{(d_1 d_2)}(z, v_1, v_2),$$

as well as

$$\tilde{\gamma}_k^{(d_1 d_2)}(z, v_1, v_2) = \frac{\partial^2}{\partial v_1 v_2} y_k^{(d_1 d_2)}(z, v_1, v_2) \text{ and } \tilde{\gamma}_k^{(d_1 d_2)}(z) = \tilde{\gamma}_k^{(d_1 d_2)}(z, 1, 1).$$

We further define $w_k^{(d_1 d_2)}(z, v_1, v_2)$ and $\Sigma_k^{(d_1 d_2)}(z, v_1, v_2)$ by

$$w_k^{(d_1 d_2)}(z, v_1, v_2) = y_k^{(d_1 d_2)}(z, v_1, v_2) - y(z)$$
$$\Sigma_k^{(d_1 d_2)}(z, v_1, v_2) = \sum_{i \geq 2} \frac{w_k^{(d_1 d_2)}(z^i, v_1^i, v_2^i)}{i}.$$

In analogy to Lemma 2.16 and Corollary 2.17 can prove the following results.

Lemma 2.34. *Let $|z| \leq \rho^2 + \varepsilon$ for sufficiently small ε and $|v_1| \leq 1$ and $|v_2| \leq 1$. Then there exists a constant L with $0 < L < 1$ and a positive constant D such that*

$$|w_k^{(d)}(z, v_1, v_2)| \leq D(|v_1 - 1| + |v_2 - 1|) \cdot |z|^d \cdot L^k$$

Corollary 2.35. *For $|v_1| \leq 1$, $|v_2| \leq 1$ and $|z| \leq \rho + \varepsilon$ ($\varepsilon > 0$ small enough) there is a positive constant $\tilde{C}$ such that (for all $k \geq 0, d_1 \geq 1, d_2 \geq 1$)*

$$|\Sigma_k^{(d_1 d_2)}(z, v_1, v_2)| \leq \tilde{C}(|v_1 - 1| + |v_2 - 1|)L^k.$$

Hence it follows, that

$$\sum_{i \geq 2} i\tilde{\gamma}_k^{(d_1 d_2)}(z^i) = \mathcal{O}(L^k) \tag{2.55}$$

Lemma 2.36. *There exist constants ϵ and θ such that for $z \in \Delta(\eta, \theta)$*

$$\tilde{\gamma}_k^{(d_1 d_2)}(z) = C^{(d_1)}(z) \cdot C^{(d_2)}(z) y(z)^{k+d_1+d_2} \sum_{\ell=0}^{k-1} (y(z)^\ell + \mathcal{O}(L^\ell)),$$

where $C^{(d_1)}(z)$ and $C^{(d_2)}(z)$ are given in Lemma 2.19.

Proof. We use the recursive representation (2.24) for $\gamma^{(d_1)}(z, v_1, v_2)$ with the additional variable v_2. This gives

$$\gamma_{k+1}^{(d_1)}(z, v_1, v_2) = y_{k+1}^{(d)}(z, v_1, v_2) \sum_{i \geq 1} \gamma_k^{(d_1)}(z^i, v_1^i, v_2^i) v_1^{i-1}.$$

Derivating with respect to v_2 gives

$$\begin{aligned}
\tilde{\gamma}_{k+1}^{(d_1 d_2)}(z, v_1, v_2) &= \gamma_{k+1}^{(d_2)}(z, v_1, v_2) \sum_{i \geq 1} \gamma_k^{(d_1)}(z^i, v_1^i, v_2^i) v_1^{i-1} \\
&\quad + y_{k+1}(z, v_1, v_2) \sum_{i \geq 1} i \tilde{\gamma}_k^{(d_1 d_2)}(z^i, v_1^i, v_2^i) v_1^{i-1} v_2^{i-1} \\
&= y_{k+1}(z, v_1, v_2) \left(\sum_{i \geq 1} \gamma_k^{(d_1)}(z^i, v_1^i, v_2^i) v_1^{i-1} \right) \left(\sum_{i \geq 1} \gamma_k^{(d_2)}(z^i, v_1^i, v_2^i) v_2^{i-1} \right) \\
&\quad + y_{k+1}(z, v_1, v_2) \sum_{i \geq 1} i \tilde{\gamma}_k^{(d_1 d_2)}(z^i, v_1^i, v_2^i) v_1^{i-1} v_2^{i-1},
\end{aligned}$$

with $\tilde{\gamma}_0^{(d_1 d_2)}(z) = 0$. Setting $v_1 = v_2 = 1$ we obtain

$$\begin{aligned}
\tilde{\gamma}_{k+1}^{(d_1 d_2)}(z) &= y(z) \left[\left(\sum_{i \geq 1} \gamma_k^{(d_1)}(z^i) \right) \left(\sum_{i \geq 1} \gamma_k^{(d_2)}(z^i) \right) + \sum_{i \geq 1} i \tilde{\gamma}_k^{(d_1 d_2)}(z^i) \right] \\
&= y(z) \left((\gamma_k^{(d_1)}(z) + \Gamma_k^{(d_1)}(z))(\gamma_k^{(d_2)}(z) + \Gamma_k^{(d_2)}(z)) + \tilde{\gamma}_k^{(d_1 d_2)}(z) + \tilde{\Gamma}_k^{(d_1 d_2)}(z) \right),
\end{aligned}$$

where we use the notations $\Gamma_k^{(d_1)}(z) = \sum_{i \geq 2} \gamma_k^{(d_1)}(z^i)$ and $\Gamma_k^{(d_2)}(z) = \sum_{i \geq 2} \gamma_k^{(d_2)}(z^i)$ as in the proof of Lemma 2.19, and $\tilde{\Gamma}_k^{(d_1 d_2)}(z) = \sum_{i \geq 2} i \tilde{\gamma}_k^{(d_1 d_2)}(z^i)$. Solving the recurrence, we get

$$\tilde{\gamma}_k^{(d_1 d_2)}(z) = \sum_{\ell=1}^{k-1} y(z)^{k-\ell} \left((\gamma_\ell^{(d_1)}(z) + \Gamma_\ell^{(d_1)}(z))(\gamma_\ell^{(d_2)}(z) + \Gamma_\ell^{(d_2)}(z)) + \tilde{\Gamma}_\ell^{(d_1 d_2)}(z) \right) \tag{2.56}$$

From Corollary 2.18 we know that $\Gamma_\ell^{(d_1)}(z) = \mathcal{O}(L^\ell)$ and $\Gamma_\ell^{(d_2)}(z) = \mathcal{O}(L^\ell)$ in $\Theta(\eta)$ for $v_1 = v_2 = 1$. Together with Equation (2.55) we have

$$\tilde{\gamma}_k^{(d_1 d_2)}(z) = \sum_{\ell=1}^{k-1} y(z)^{k-\ell} \left((C^{(d_1)}(z) y(z)^{\ell+d_1} + \mathcal{O}(L^\ell))(C^{(d_2)} y(z)^{\ell+d_2} + \mathcal{O}(L^\ell)) + \mathcal{O}(L^\ell) \right),$$

and the result follows. □

To extract coefficients we will use Cauchy's formula.

$$[z^n] \tilde{\gamma}^{(d_1 d_2)}(z) = \frac{1}{2\pi i} \int_\delta \tilde{\gamma}^{(d_1 d_2)}(z) \frac{1}{z^{n+1}} \mathrm{d}z,$$

where δ is the truncated contour $\delta = \delta_1 \cup \delta_2 \cup \delta_3 \cup \delta_4$ given by

$$\begin{aligned} \delta_1 &= \left\{ z = a + \frac{\rho i}{n} \middle| \rho \le a \le \rho + \frac{\eta \log^2 n}{n} \right\}, \\ \delta_2 &= \left\{ z = \rho\left(1 - \frac{e^{i\varphi}}{n}\right) \middle| -\frac{\pi}{2} \le \varphi \le \frac{\pi}{2} \right\}, \\ \delta_3 &= \left\{ z = a - \frac{\rho i}{n} \middle| \rho \le a \le \rho + \frac{\eta \log^2 n}{n} \right\} \end{aligned} \tag{2.57}$$

and δ_4 being a circular arc closing the contour, cf Figure 2.8.

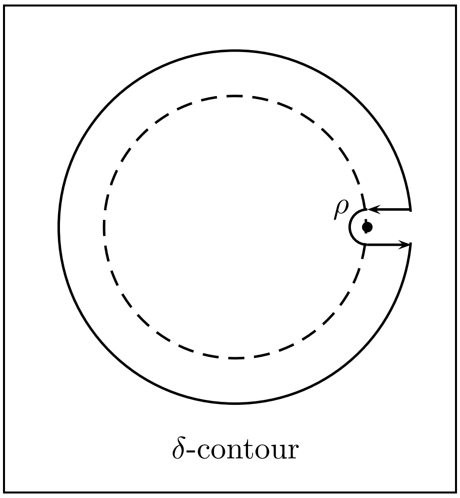

Figure 2.8: The integration contour δ

It can be shown that the circular arc δ_4 gives only negligible contribution to the integral, we omit the details here as the proof is very similar to the one in Section 2.3.1. Near ρ, more precicely for $z = \rho(1 + \frac{s}{n})$ with $z \in \delta_1 \cup \delta_2 \cup \delta_3$ and for $k = \kappa\sqrt{n}$, we have

$$\begin{aligned} y(z)^{d_1+d_2} &= 1 + \mathcal{O}\left(\sqrt{\left|\frac{s}{n}\right|}\right) \\ 1 - y(z) &\sim b\sqrt{\rho}\sqrt{-\frac{s}{n}}\left(1 + \mathcal{O}\left(\sqrt{\left|\frac{s}{n}\right|}\right)\right) \\ y(z)^k &\sim \exp(-\kappa b\sqrt{-\rho s})\left(1 + \mathcal{O}\left(\left|\frac{s}{\sqrt{n}}\right|\right)\right) \\ C^{(d_1)}(z) &\sim C_{d_1}\rho^{d_1} + \mathcal{O}\left(\left|\frac{s}{n}\right|\right), \qquad C^{(d_2)}(z) \sim C_{d_2}\rho^{d_2} + \mathcal{O}\left(\left|\frac{s}{n}\right|\right) \\ \mathrm{d}z &= \frac{\rho}{n}\mathrm{d}s \end{aligned}$$

Hence, the expected value $[z^n]\tilde{\gamma}^{(d_1 d_2)}(z)$ is given by

$$\begin{aligned} &[z^n]\tilde{\gamma}^{(d_1 d_2)}(z) \sim \\ &\sim C_{d_1}C_{d_2}\rho^{d_1+d_2}\frac{1}{2\pi i}\int_{\tilde{\delta}} \frac{\sqrt{n}}{b\sqrt{\rho}\sqrt{-s}} e^{-\kappa b\sqrt{-\rho s}}(1 - e^{-\kappa b\sqrt{-\rho s}})e^{-s}\left(1 + \mathcal{O}\left(\left|\frac{s}{\sqrt{n}}\right|\right)\right)\frac{1}{n}\rho^{-n}\mathrm{d}s, \end{aligned}$$

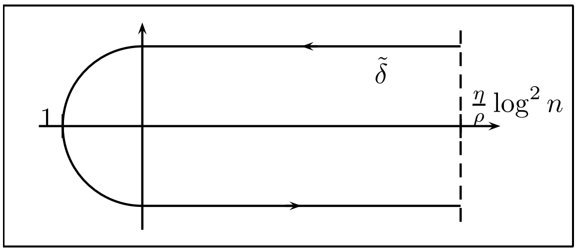

Figure 2.9: The contour $\tilde{\delta}$ obtained by substitution

as $\sum_{\ell=0}^{k-1} y(z)^\ell = \frac{1-y(z)^k}{1-y(z)}$, where $\tilde{\delta}$ is the contour displayed in Figure 2.9 obtained by the substitution $z \mapsto s$.

For an integral of the shape $\int_{\tilde{\delta}} \frac{1}{\sqrt{-s}} e^{-\alpha\sqrt{-s}-s} \mathrm{d}s$ we use the following auxiliary result.

Lemma 2.37. *Let $\hat{\delta}$ be a Hankel contour (cf [24]), that is, informally speaking, a contour like $\tilde{\delta}$, but beginning and ending in infinity. We have for $\beta > 0$*

$$\frac{1}{2\pi i}\int_{\hat{\delta}} \frac{1}{\sqrt{-s}} e^{-\alpha\sqrt{-s}-s} \mathrm{d}s \sim \frac{1}{\sqrt{\pi}} e^{-\frac{\alpha^2}{4}},$$

as n tends to infinity.

Proof. We first substitute $s = u^2$. Then we have

$$-\frac{1}{2\pi i}\int_{-\infty}^{\infty} \frac{1}{iu} e^{-i\alpha u - u^2} 2u\mathrm{d}u,$$

where we get a negative sign due to taking the root $\sqrt{-1} = -i$ to preserve the correct orientation of the contour. We further complete the exponent to a full square and therefore subsitute $v = u + \frac{i\alpha}{2}$ to obtain

$$\frac{1}{\pi} e^{-\frac{\alpha^2}{4}} \int_{-\infty+\frac{\alpha^2}{4}}^{\infty+\frac{\alpha^2}{4}} e^{-v^2}\mathrm{d}v \sim \frac{1}{\sqrt{\pi}} e^{-\frac{\alpha^2}{4}}$$

□

Further, for an integral of the shape $\mathcal{O}\left(\left|\frac{s}{\sqrt{n}}\right|\right)\int_{\hat{\delta}} \frac{1}{\sqrt{-s}} e^{-\alpha\sqrt{-s}-s}\mathrm{d}s$, we obtain

$$\begin{aligned}
\frac{1}{2\pi i}\mathcal{O}\left(\left|\frac{s}{\sqrt{n}}\right|\right)\int_{\hat{\delta}} \frac{1}{\sqrt{-s}} e^{-\alpha\sqrt{-s}-s}\mathrm{d}s &= \frac{1}{2\pi i}\mathcal{O}\left(\frac{1}{\sqrt{n}}\right)\int_{\hat{\delta}} \sqrt{-s} e^{-\alpha\sqrt{-s}-s}\mathrm{d}s \\
&= \frac{1}{2\pi i}\mathcal{O}\left(\frac{1}{\sqrt{n}}\right)\frac{\partial^2}{\partial\alpha^2}\int_{\hat{\delta}} \frac{1}{\sqrt{-s}} e^{-\alpha\sqrt{-s}-s}\mathrm{d}s \\
&= \mathcal{O}\left(\frac{1}{\sqrt{n}}\right)\frac{1}{\sqrt{\pi}} e^{-\frac{\alpha^2}{4}}\frac{1}{2}(1+\alpha) = \mathcal{O}\left(\frac{1}{\sqrt{n}}\right).
\end{aligned}$$

Recall that the coefficients of $y(z)$ are asymptotically given by $y_n \sim \frac{b\sqrt{\rho}}{2\sqrt{\pi}}\rho^{-n}n^{-3/2}$ (cf (2.4)). Hence with $\alpha = \kappa b\sqrt{\rho}$ and $\alpha = 2\kappa b\sqrt{\rho}$, respectively, we obtain for $\mathbb{E}\left(X_n^{(d_1)}(k)\cdot X_n^{(d_2)}(k)\right)$

$$\mathbb{E}\left(X_n^{(d_1)}(k)\cdot X_n^{(d_2)}(k)\right) = C_{d_1}C_{d_2}\rho^{d_1+d_2}\frac{2}{b^2\rho}n\left(e^{-\frac{\kappa^2 b^2\rho}{4}} + e^{-\kappa^2 b^2\rho}\right)\left(1+\mathcal{O}\left(\frac{1}{\sqrt{n}}\right)\right).$$

With the same substitutions as in the proof of Lemma 2.37, we obtain (this has already been proven in [34, Lemma 3.4])

$$\frac{1}{2\pi i}\int_{\hat{\delta}} e^{-\alpha\sqrt{-s}-s}\mathrm{d}s \sim \frac{\alpha}{2\sqrt{\pi}}e^{-\frac{\alpha^2}{2}}.$$

With this auxiliary result, the representation of the covariance given in Proposition 2.33 follows immediately as n tends to infinity. Note that the covariance function is strictly positive, as depicted in Figure 2.10.

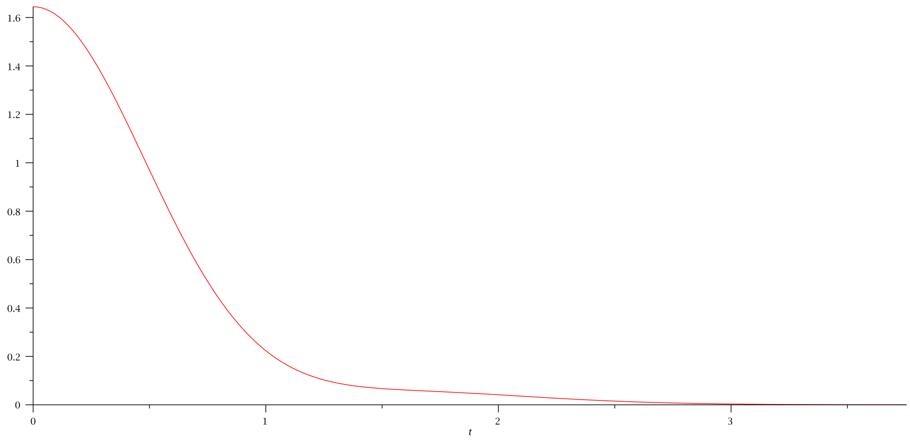

Figure 2.10: The covariance for $\kappa \in [0, \mathbb{E}(H_n) + 3\sqrt{\mathbb{V}\mathrm{ar}(H_n)}]$

The correlation coefficient

To obtain more information on the correlation of two degrees d_1 and d_2 on the same level $k = \kappa\sqrt{n}$, we compute the correlation coefficient, given by

$$\mathrm{Cor}\left(X_n^{(d_1)}(k), X_n^{(d_2)}(k)\right) = \frac{\mathbb{C}\mathrm{ov}(X_n^{(d_1)}(k), X_n^{(d_2)}(k))}{\sqrt{\mathbb{V}\mathrm{ar}(X_n^{(d_1)}(k))}\sqrt{\mathbb{V}\mathrm{ar}(X_n^{(d_2)}(k))}}.$$

Theorem 2.38. *Let $X_n^{(d_1)}(k)$ and $X_n^{(d_2)}(k)$ be the random variables counting the number of vertices of degree d_1 and d_2, respectively, on a level $k = \kappa\sqrt{n}$ in a Pólya tree of size n. Then the correlation coefficient is asymptotically equal to*

$$\mathrm{Cor}\left(X_n^{(d_1)}(k), X_n^{(d_2)}(k)\right) = 1 + \mathcal{O}\left(\frac{1}{\sqrt{n}}\right),$$

as n tends to infinity.

To compute the correlation coefficient, it remains to compute the variance $\mathbb{V}\mathrm{ar}(X_n^{(d_1)}(k))$, given by

$$\mathbb{V}\mathrm{ar}(X_n^{(d_1)}(k)) = \mathbb{E}\left((X_n^{(d_1)}(k))^2\right) - \left(\mathbb{E}(X_n^{(d_1)}(k))\right)^2.$$

We need to determine $\mathbb{E}\left((X_n^{(d_1)}(k))^2\right)$, which can be done very similarly to the previous part.

$$\mathbb{E}\left((X_n^{(d_1)}(k))^2\right) = \frac{1}{y_n}[z^n]\left[\frac{\partial}{\partial v_1}\left(v_1\frac{\partial}{\partial v_1}y_k(z, v_1, 1)\right)\right]_{v_1=1},$$

Proposition 2.39. *The Variance* $\mathbb{V}\mathrm{ar}(X_n^{(d_1)}(k))$ *of the random variable* $X_n^{(d_1)}(k)$ *counting vertices of degree* d_1*, with* d_1 *fixed, at level* $k = \kappa n$ *in a random Pólya tree of size* n *is asymtotically given by*

$$\mathbb{V}\mathrm{ar}(X_n^{(d_1)}(k)) = C_{d_1}^2 \rho^{2d_1} n \left(\frac{2}{b^2\rho} \left(e^{-\frac{\kappa^2 b^2 \rho}{4}} + e^{-\kappa^2 b^2 \rho} \right) - \kappa^2 e^{-\frac{\kappa^2 b^2 \rho}{2}} \right) \left(1 + \mathcal{O}\left(\frac{1}{\sqrt{n}} \right) \right), \tag{2.58}$$

as n *tends to infinity.*

We proceed analogously to the computation of the variance, and obtain the following auxiliary result.

Lemma 2.40. *There exist constants* ϵ *and* θ *such that for* $z \in \Delta(\eta, \theta)$

$$\tilde{\gamma}_k^{(d_1[2])}(z) = (C^{(d_1)}(z))^2 y(z)^{k+2d_1} \frac{1 - y(z)^k}{1 - y(z)} + C^{(d_1)}(z) y(z)^{k+d_1},$$

where $C^{(d_1)}(z)$ *and* $C^{(d_1)}(z)$ *are given in Lemma 2.19.*

Proof. The proof of this lemma is analogous to the proof of Lemma 2.36, derivating recurrence (2.24) a second time. The additional summand $C^{(d_1)}(z)y(z)^{k+d_1}$ origins in derivating twice with respect to the same variable v_1. □

Note that the additional summand $C^{(d_1)}(z)y(z)^{k+d_1}$ in Lemma 2.40, where $\tilde{\gamma}_k^{(d_1[2])}(z)$ and $\tilde{\gamma}_k^{(d_1 d_2)}(z)$ differ from each other, is equal to the expected value $\mathbb{E}\left(X_n^{(d_1)}(k)\right)$ when extracting coefficients $\frac{1}{y_n}[z^n]C^{(d_1)}(z)y(z)^{k+d_1}$. As this is of order $\sqrt{n}$, while the coeffient of the other terms will be of order n, this term is part of the error term, and we obtain

$$\mathbb{E}\left((X_n^{(d_1)}(k))^2 \right) = C_{d_1}^2 \rho^{2d_1} \frac{2}{b^2\rho} n \left(e^{-\frac{\kappa^2 b^2 \rho}{4}} + e^{-\kappa^2 b^2 \rho} \right) \left(1 + \mathcal{O}\left(\frac{1}{\sqrt{n}} \right) \right) \tag{2.59}$$

by using Cauchy's formula and the integration contour δ given in (2.57). Applying the known estimate for $\mathbb{E}(X_n^{(d_1)}(k))$ we obtain the representation given in Proposition 2.39, and with Proposition 2.33 the result given in Theorem 2.38 follows immediately.

Remark. From the result of Theorem 2.38 it follows that asymptotically, $X_n^{(d_1)}(k)$ and $X_n^{(d_2)}(k)$ are linearly related, i.e.

$$X_n^{(d_1)}(k) \sim A \cdot X_n^{(d_2)}(k) + B,$$

with some constants $A(d_1, d_2) > 0$ and $B(d_1, d_2)$. We cannot provide any further informations on these constants at the moment. Their study would propose some forthcoming research.

CHAPTER 3

Boolean trees

Consider the set $\mathcal{F}_n$ of Boolean functions on a set of n variables. There are 2^{2^n} such Boolean functions, as a value from $\{True, False\}$ can be assigned to every variable, which gives 2^n different assignments, for every assignment the function has output $True$ or $False$. In this chapter we consider several models of And/Or trees, i.e. trees where internal nodes carry labels from the set $\{\wedge, \vee\}$ and external nodes (leaves) have labels from the set $\{x_1, \bar{x}_1, \ldots, x_n, \bar{x}_n\}$. Obviously, every such tree represents a function f from $\mathcal{F}_n$. We consider the uniform distribution on the set of And/Or trees of size m (denoting by size the number of leaves) and are interested in the limiting probability $\mathbb{P}_n(f)$ of a given function $f \in \mathcal{F}_n$ being computed by a random tree of size m, as m tends to infinity, if it exists. In all tree models we will study, such a limiting distribution exists.

A lot of work has been going on in this field. Lefmann and Savický [51] were first to prove the existence of the limiting probability of f. The bounds given in their paper were improved by Chauvin *et al.* [11]. A comparison to the probability distribution induced by a critical Galton Watson process as well as various numerical results are given in Gardy [26]. A similar study on implication trees, i.e. Boolean trees where internal nodes carry implication labels ($\Rightarrow$) has been done by Fournier *et al.* [25].
The content of this chapter origins in a collaboration with Antoine Genitrini, Bernhard Gittenberger and Cécile Mailler. It will be subject of a forthcoming paper. The methods and results are based on a recent work by Kozik [47] on pattern languages. In his paper, Kozik proves a strong relation between the limiting probability of a given function f and its complexity $L(f)$ (that is the minimal size of a tree computing the function f) in binary planar And/Or trees, asymptotically as the number of variables tends to infinity. We want to study the impact of removing step by step the restrictions on these tree, that is considering first planar, but non-binary or non-planar but binary And/Or trees, and later non-planar non-binary trees, which relate strongly to Pólya trees. Considering such tree structures seems quite natural, as the new characteristics correspond to adding the properties of associativity (non-binary) and commutativity (non-planar), which are given for the $\wedge$ and $\vee$ operator on the level of Boolean logic.

Kozik has shown that the asymptotic order of $\mathbb{P}_n(f)$ depends on $L(f)$ for binary planar trees (c.f. Kozik's paper [47]). First, we compare the limiting probabilities of the constant function $True$[1]. Supported by numerical results for n equals 1 and 2, we conjectured that commutativity does not matter. Surprisingly to us, we find that both characteristics have impact on the limiting probability $\mathbb{P}_n(True)$. To be more precise, the asymptotic leading coefficient differs from model to model as n tends to infinity. To get more insight, we further compare probabilities of functions of complexity 1, those are the literals $x, \bar{x}$, in a next step. In the last section of this chapter, we prove that for all tree models compared, the asymptotically relevant fraction of trees computing a given function f is given by the set of minimal trees of f expanded once in a given way, and give bounds for the arising probability distribution. A similar result is proved in [47] for planar binary And/Or trees and in [25] for implication trees.

3.1 Associative and commutative trees: definitions, generating functions.

Kozik [47] has shown that in binary planar trees the order of magnitude of the limiting probability of a given Boolean function is related to its complexity. We generalise this result and therefore define And/Or trees and the complexity of a function.

Definition 3.1. *We define an* And/Or *tree as a labelled tree. Each internal node is labelled with one of the connectors $\{\wedge, \vee\}$ and each leaf with one of the literals $\{x_1, \bar{x}_1 \ldots, x_n, \bar{x}_n\}$. We define the* size *of an And/Or tree to be the number of its leaves.*

Definition 3.2. *The* complexity $L(f)$ *of a non-constant function f, i.e. $f \notin \{True, False\}$, is given by the size of a smallest And/Or tree computing f (in the rest of the paper such trees will be called* minimal *for f). We define the complexity of $True$ and $False$ to be $L(True) = L(False) = 0$.*

As it will be clear later, the complexity of a function does not depend on the chosen model.

Definition 3.3. *We are considering sets $\mathcal{T}_{m,n}$ of And/Or trees of size $m = 1, 2, 3 \ldots$. Let $\mathbb{U}_{m,n}$ be the uniform distribution on $\mathcal{T}_{m,n}$, $\mathbb{P}_{m,n}$ its image on the set of Boolean functions. We call*

$$\mathbb{P}_n = \lim_{m \to \infty} \mathbb{P}_{m,n}$$

the limiting distribution, assuming that this limit exists.

At first, we will present the result proven by Kozik. This result will be generalised in the forthcoming parts of this chapter.

3.1.1 The classical model.

Let us consider the set $\mathcal{T}$ of binary planar trees, whose internal nodes are labelled with $\wedge$ or $\vee$, and whose external nodes are labelled with literals chosen in $\{x_1, \bar{x_1}, \ldots, x_n, \bar{x_n}\}$: every such tree computes a Boolean function on n variables. We denote by $T(z) = \sum_{m \geq 0} T_m z^m$

[1]Note that by negation, the probability distribution behaves symmetrically, i.e. the probability of $False$ will be the same as that of $True$, and just as well for all other functions and their negations.

the generating function of this set of trees, and by $T_f(z)$ the generating function of such trees computing the Boolean function f. Let us recall some well known results about this generating function:

Proposition 3.4. *Binary And/Or trees fulfil the symbolic equation*

$$\mathcal{T} = \mathcal{X} \ + \mathcal{T} \wedge \mathcal{T} \ + \mathcal{T} \vee \mathcal{T} \tag{3.1}$$

and thus the generating function $T(z)$ verifies $T(z) = 2nz + 2T(z)^2$. Therefore, we have:

$$T(z) = \frac{1 - \sqrt{1 - 16nz}}{4}$$

and the singularity ρ_n of $T(z)$ is $\frac{1}{16n}$.

Let us consider the uniform distribution over the set of trees of size m and then the probability distribution $\mathbb{P}_{m,n}$ it induces in the set $\mathcal{F}_n$ of Boolean functions on n variables. The limit of this distribution when m tends to infinity, denoted by $\mathbb{P}_n$, has already been studied, in particular by Lefmann and Savický [51], Chauvin *et al.* [11] and Kozik [47] who has shown:

Theorem 3.5. *[47] For all Boolean functions f,*

$$\mathbb{P}_n(f) = \frac{C_f}{n^{L(f)+1}} \text{ as } n \to \infty$$

where $L(f)$ is the complexity of f, i.e. the size of a minimal tree computing f, and C_f is some positive constant, which we will specify later in this chapter.

Remark. Be careful that in this theorem, f (and thus $L(f)$) is fixed, and n tends to infinity. The considered function depends on a finite number of variables.

First of all, let us define associative trees, commutative trees and finally associative and commutative trees, and the induced laws over the set of Boolean functions over $\mathcal{F}_n$. The final aim of this chapter is to generalise Theorem 3.5.

3.1.2 The associative planar model.

Definition 3.6. *An* associative tree *is a planar tree where each node has outdegree chosen in $\mathbb{N} \setminus \{1\}$. A* labelled associative tree *is an associative tree in which each external node has a label in $\{x_1, \bar{x}_1, \ldots, x_n, \bar{x}_n\}$ and each internal node has a $\wedge$-label or a $\vee$-label but cannot have the same label as its father. We denote by $\mathcal{A}$ the family of associative trees and by $\mathcal{A}_m$ the set of such trees of size m.*

Note that the trees are *stratified.* i.e. the root can be labelled either by $\wedge$ or $\vee$ and it determines the labels of all others internal nodes.

We denote by $\mathbb{P}^a_n = \lim_{m\to\infty} \mathbb{P}^a_{m,n}$ the limiting distribution of Boolean functions induced by associative And/Or trees. Our aim is to compare the limiting distributions $\mathbb{P}^a_n$ and $\mathbb{P}_n$.

The generating function of associative trees is given by $A(z) = \hat{A}(z) + \check{A}(z) - 2nz$, where $\hat{A}$ (resp. $\check{A}$) is the generating function of associative trees being a leaf or being rooted at a $\wedge$ (resp. $\vee$). We have to note that $\hat{A} = \check{A}$ and,

$$\hat{A}(z) = 2nz + \sum_{k\geq 2} \check{A}^k = 2nz + \frac{\hat{A}^2(z)}{1 - \hat{A}(z)}.$$

Therefore,

$$A(z) = \frac{1}{2}\left(1 - 2nz - \sqrt{1 - 12nz + 4n^2z^2}\right) \tag{3.2}$$

and its dominant singularity is

$$\alpha_n = \frac{3 - 2\sqrt{2}}{2n}.$$

Moreover, $A(\alpha_n) = \sqrt{2} - 1$.

Remark. Thanks to the Drmota-Lalley-Woods theorem (Theorem 1.12), we can show that $P^a_{m,n}$ has indeed a limit when m tends to infinity. If we denote by $\hat{A}_f(z)$ (resp. $\check{A}_f(z)$) the generating function of the number of associative trees which are a leaf or whose root is labelled by $\wedge$ (resp. $\vee$) computing f, these generating functions satisfy the system:

$$\hat{A}_f(z) = z\mathbb{1}_{\{f \text{ lit}\}} + \sum_{i=2}^{\infty} \sum_{\substack{g_1,\dots,g_i \\ g_1\wedge\dots\wedge g_i = f}} \check{A}_{g_1}(z)\cdots\check{A}_{g_i}(z)$$

$$\check{A}_f(z) = z\mathbb{1}_{\{f \text{ lit}\}} + \sum_{i=2}^{\infty} \sum_{\substack{g_1,\dots,g_l \\ g_1\vee\dots\vee g_i = f}} \hat{A}_{g_1}(z)\cdots\hat{A}_{g_i}(z).$$

This system fulfils all preliminaries of the Drmota-Lalley-Woods theorem, hence $\hat{A}_f(z)$ and $\check{A}_f(z)$ and also $A_f(z) = \hat{A}_f(z) + \check{A}_f(z) - z\mathbb{1}_{\{f \text{ lit}\}}$ have the same singularity α_n as $A(z)$, and therefore the limit

$$\lim_{m\to\infty} \mathbb{P}_{m,n}(f) = \frac{[z^m]A_f(z)}{[z^m]A(z)}$$

exists.

3.1.3 The non-planar binary model.

Definition 3.7. *A* labelled commutative *tree on n variables is a non-plane binary tree where every internal node is labelled with one of the labels $\{\wedge, \vee\}$ and every leaf is labelled by a literal $\{x_i, \bar{x}_i, i = 1, \dots, n\}$. We denote by $\mathcal{C}$ this family of trees.*

We consider the distribution $\mathbb{P}^c_{m,n}$ induced over the set of Boolean functions of n variables by the uniform distribution over such trees of size m.

Binary commutative trees fulfil the same symbolic equation as in the planar case (cf. (3.1)) but because of commutativity, the generating function of all commmutative trees on n variables, counting leaves, is given implicitely by

$$C(z) = 2nz + C(z)^2 + C(z^2), \tag{3.3}$$

where the term $\frac{1}{2}(C^2(z) + C(z^2))$ tracks a possible symmetry if both subtrees of the root are identical. The system of equations for the generating functions $C_f(z)$ computing a given Boolean function f is given by

$$C_f(z) = z\mathbb{1}_{\{f \text{ lit}\}} + \frac{1}{2}\sum_{\substack{g,h\neq f \\ g\wedge h = f}} C_g(z)C_h(z) + \frac{1}{2}\sum_{\substack{g,h\neq f \\ g\vee h = f}} C_g(z)C_h(z) + C_f(z)^2 + C_f(z^2). \tag{3.4}$$

This system is not algebraic due to the term $C_f(z^2)$, hence we cannot apply Drmota-Lalley-Woods theorem in the first place. Still, knowing that for the singularity of $C(z)$ we have $0 < \gamma_n < 1$ we can use Lemma 1.13 to apply the Drmota-Lalley-Woods theorem. Thus, we conclude that all the $C_f(z)$ and $C(z)$ have the same singularity γ_n, and moreover that $\mathbb{P}^c_{m,n}$ converges to a limiting probability distribution P^c_n when m tends to infinity.

3.1.4 The non-planar associative model.

Definition 3.8. *Finally we define* general trees *as non-planar and non-binary trees, with internal nodes labelled by* $\wedge$ *or* $\vee$ *(with the condition that father and sons cannot have the same label), and external nodes labelled by literals chosen in* $\{x_1, \bar{x}_1, \ldots, x_n, \bar{x}_n\}$. *We denote by* $\mathcal{P}$ *this family of trees.*

Remark. Note that general trees are closely related to Pólya trees treated in Chapter 2. In fact, they are Pólya trees with the only restriction of not having nodes with outdegree one, and with the label of the root being chosen from $\{\vee, \wedge\}$ (and thus determining labels for all other internal nodes) and the leaves labelled with literals (if the tree consists only of a single leaf then this is labelled with a literal).

As in the other models, we consider the distribution $\mathbb{P}^{a,c}_{m,n}$ induced over the set of Boolean functions by the uniform distribution over such trees of size m.

Let $P(z)$ be the generating function of general And/Or trees, and $\hat{P}(z)$ (resp. $\check{P}(z)$) the generating function of general trees being a leaf or being rooted by $\wedge$ (or by $\vee$, resp.). Then

$$P(z) = \hat{P}(z) + \check{P}(z) - 2nz, \tag{3.5}$$

with

$$\begin{aligned}
\hat{P}(z) &= \exp\left(\sum_{i\geq 1} \frac{\check{P}(z^i)}{i}\right) - 1 - \hat{P}(z) + 2nz \\
\check{P}(z) &= \exp\left(\sum_{i\geq 1} \frac{\hat{P}(z^i)}{i}\right) - 1 - \check{P}(z) + 2nz.
\end{aligned} \tag{3.6}$$

Moreover, the generating function $\hat{P}_f(z)$ of general trees computing f, and $\check{P}_f(z)$, satisfy the following system:

$$\begin{aligned}
\hat{P}_f(z) &= z\mathbb{1}_{\{\text{f lit}\}} + \sum_{l=2}^{\infty} \sum_{\substack{g_1,\ldots,g_l \\ g_1\wedge\ldots\wedge g_l = f}} \prod_{j=1}^{l} \left(\exp\left(\sum_{i\geq 1} \frac{\check{P}_{g_j}(z^i)}{i}\right) - 1\right) \\
\check{P}_f(z) &= z\mathbb{1}_{\{\text{f lit}\}} + \sum_{l=2}^{\infty} \sum_{\substack{g_1,\ldots,g_l \\ g_1\wedge\ldots\wedge g_l = f}} \prod_{j=1}^{l} \left(\exp\left(\sum_{i\geq 1} \frac{\hat{P}_{g_j}(z^i)}{i}\right) - 1\right).
\end{aligned}$$

Thus, we can check the hypothesis of the Drmota-Lalley-Woods theorem and conclude that the limiting distribution $\mathbb{P}^{a,c}_n$ of $\mathbb{P}^{a,c}_{m,n}$ when m tends to infinity exists, and moreover, that all the $\hat{P}_f, \check{P}_f$ and $\hat{P}, \check{P}$ have the same singularity, denoted by δ_n.

In the forthcoming parts of this chapter, we will prove that Theorem 3.5 still holds in the associative or commutative cases. We start by showing in Section 3.2 that the limiting ratio of tautologies is of order $\frac{1}{n}$. We compute the limit of $\mathbb{P}_n(True)$ when n tends to infinity for the different models. If these limits were the same, we could not conclude anything, but in fact they are all different, which permits us to conclude that asymptotically, when n tends to infinity, the probability distributions induced by the various models are all different. In Part 3.3, we proceed in a similar manner to prove that in all models, the asymptotic ratio of literals, i.e. functions of complexity 1, is of order $\frac{1}{n^2}$ when n tends to infinity, but the limiting ratios are different from model to model. Finally, we generalise Theorem 3.5 in Section 3.4.

3.2 Limiting ratio of tautologies

In this part we compute the limiting probability of the constant function $True$. As suggested by Kozik's results, the limiting probability of tautologies reduces to the limiting probability of so-called *simple tautologies*, defined by the following:

Definition 3.9. *A simple tautology realized by x_i, $i = 1 \dots n$, is a Boolean expression which has the shape $x_i \vee \bar{x}_i \vee f$ for some Boolean function f, i.e. there exists a leaf labelled by x_i and a leaf labelled by $\bar{x}_i$, both connected to the root by an "$\vee$-only-path" (c.f. Figure 3.1). A simple tautology is a simple tautology realized by any variable $x \in \{x_1, \dots, x_n\}$.*

We denote by ST_m the number of simple tautologies of size m (on n variables, n is omitted for simplicity), and let $G_x(z)$ the generating function of simple tautologies realized by x.

Definition 3.10. *Let $\mathcal{V}$ be a set of variables and $ST_m(\mathcal{V})$ the set of simple tautologies realized by all $x \in \mathcal{V}$, but not by any other variable $y \notin \mathcal{V}$, and $\biguplus$ denote the disjoint union.*

- *$K_{1,m}$ is the set of simple tautologies that are realized by exactly one variable: i.e. $K_{1,m} = \biguplus_{i=1}^{n} ST_m(\{x_i\})$,*
- *$K_{2,m}$ is the set of simple tautologies that are realized by exactly two different variables: i.e. $K_{2,m} = \biguplus_{\substack{i,j=1 \\ i \neq j}}^{n} (ST_m(\{x_i, x_j\}))$,*

$$\vdots$$

- *$K_{n,m}$ is the set of simple tautologies that are realized by exactly n different variables : $K_{n,m} = ST(\{x_1, \dots, x_n\})$.*

Let $G(z) = nG_x(z) = \sum_{m \geq 0} G_m z^m$. Obviously, $G_m = K_{1,m} + 2K_{2,m} + \dots + nK_{n,m}$. Note that this function does not count simple tautologies, but the number of simple tautologies of size m is smaller than G_m, and hence we have

$$K_{1,m} \leq ST_m \leq G_m.$$

To calculate limiting probabilities, we use the singular expansions of the considered generating functions around their dominant singularities. Consider the generating function $T(z)$ of a given family of And/Or trees together with the generating function $S(z)$ of a subset $\mathcal{S}$ of such trees.

Lemma 3.11. *We assume that $T(z)$ and $S(z)$ have the same dominant singularity ρ and a squareroot singular expansion*

$$T(z) = a_T - b_T\sqrt{1 - \frac{z}{\rho}} + \mathcal{O}\left(1 - \frac{z}{\rho}\right) \quad S(z) = a_S - b_S\sqrt{1 - \frac{z}{\rho}} + \mathcal{O}\left(1 - \frac{z}{\rho}\right)$$

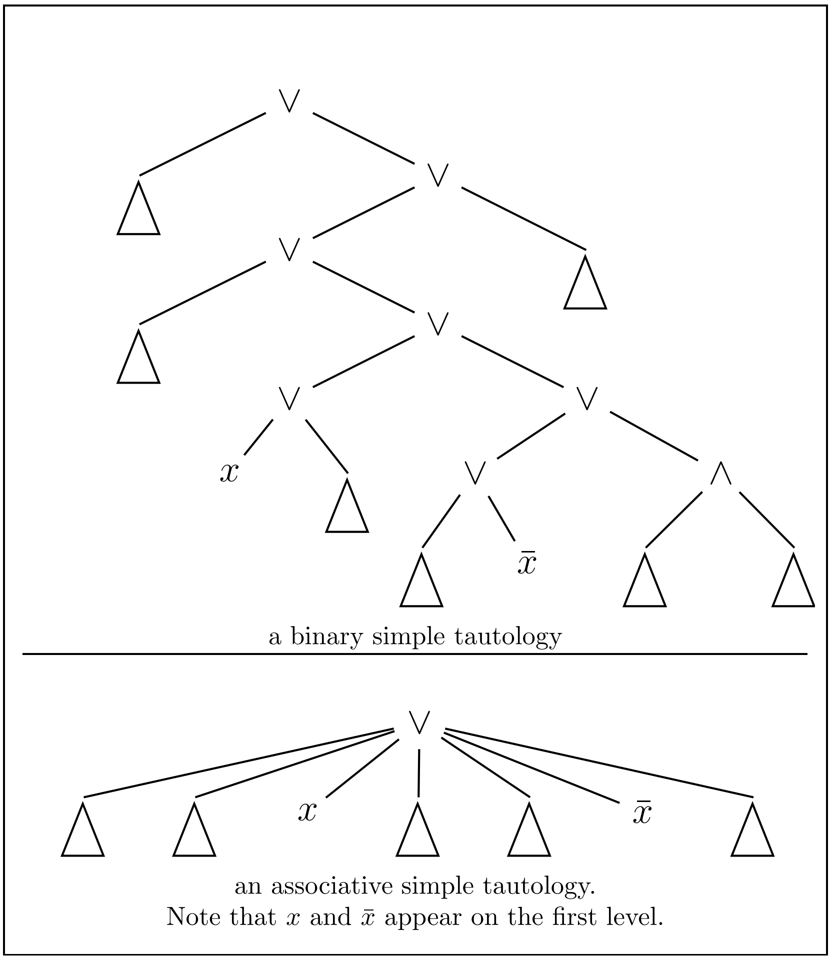

Figure 3.1: Simple tautologies - $\triangle$ stands for arbitrary trees

around ρ. Then

$$\lim_{m\to\infty}\frac{S_m}{T_m} = \lim_{z\to\rho}\frac{S'(z)}{T'(z)}.$$

We call this number the limiting ratio of the set $\mathcal{S}$.

Proof. If m tends to infinity, transfer lemma (cf Lemma 1.7) give

$$\left.\begin{array}{l} T_m \sim \dfrac{b_T}{\Gamma(\frac{1}{2})} n^{-\frac{3}{2}}\rho^{-m} \\ S_m \sim \dfrac{b_S}{\Gamma(\frac{1}{2})} n^{-\frac{3}{2}}\rho^{-m} \end{array}\right\} \Rightarrow \frac{S_m}{T_m} \sim \frac{b_S}{b_T}.$$

Derivation of the singular expansions gives

$$\begin{array}{l} T'(z) \sim \frac{b_T}{2}\left(1-\frac{z}{\rho}\right)^{-\frac{1}{2}}, \\ S'(z) \sim \frac{b_S}{2}\left(1-\frac{z}{\rho}\right)^{-\frac{1}{2}}. \end{array}$$

Hence the result follows. $\square$

Remark. If $\mathcal{S}$ is the set of tree computing a given function f, then, the limiting probability of f is equal to the limiting ratio of $\mathcal{S}$ because

$$\mathbb{P}_{m,n}(f) = \frac{\#\text{ trees of size } m \text{ computing } f}{\#\text{ all trees of size } m} = \frac{S_m}{T_m}.$$

3.2.1 Binary planar trees

In the binary planar model, it has been shown by Woods [70] and again by Kozik [47] that asymptotically, when n tends to infinity, all tautologies are *simple tautologies*. Therefore, to estimate the probability that a binary planar tree computes the function $True$, it suffices to count simple tautologies, and furthermore, thanks to the following proposition, simple tautologies that are realized by only one variable (i.e. the set $K_{1,m}$).

Proposition 3.12. *If n tends to infinity, then*

$$\lim_{m\to\infty} \frac{\sum_{k=1}^{n} k\#K_{k,m}}{T_m} = \lim_{m\to\infty} \frac{\#K_{1,m}}{T_m} + O\left(\frac{1}{n^2}\right).$$

The proof of the proposition is deferred to the end of this section since further technical concepts are required.

Theorem 3.13. *The limit ratio of simple tautologies in the binary planar model, and thus the limit ratio of tautologies in the binary planar model is*

$$\lim_{m\to\infty} \frac{ST_m}{T_m} = \lim_{m\to\infty} \mathbb{P}_{m,n}(True) = \frac{3}{4n} + O\left(\frac{1}{n^2}\right),$$

where T_m is the total number of planar binary trees and ST_m is the number of simple tautologies of size m labelled with n variables.

Proof. Let us compute the generating function of simple tautologies. First, let g_x be the generating function of trees containing a leaf labelled by x which is connected to the root by an $\vee$-only-path (c.f. Figure 3.2) and $\bar{g}_x(z)$ the generating function of trees which are not of such shape. Hence $\bar{g}_x = T - g_x$.

The function $\bar{g}_x$ is given by

$$\bar{g}_x(z) = T(z)^2 + \bar{g}_x^2(z) + (2n-1)z.$$

We obtain this equation by decomposing the tree at its root: if the root is labelled by an $\wedge$, the tree is not of the shape depicted in Figure 3.2 and both subtrees are arbitrary random trees. If the root is labelled by a $\vee$, neither of the two subtrees may have the shape of Figure 3.2. If the root is a single leaf, it must not not be labelled by x. By a symbolic argumentation, the three cases translate to the three terms in the equation. Solving this equation, thanks to the explicit value of $T(z)$ given in Proposition 3.4, we get:

$$\bar{g}_x(z) = \frac{1}{2} - \frac{\sqrt{2 + 2\sqrt{1-16nz} - 16nz + 16z}}{4},$$

and thus

$$g_x(z) = \frac{\sqrt{2 + 2\sqrt{1-16n} - 16nz + 16z} - \sqrt{1-16nz} - 1}{4}. \tag{3.7}$$

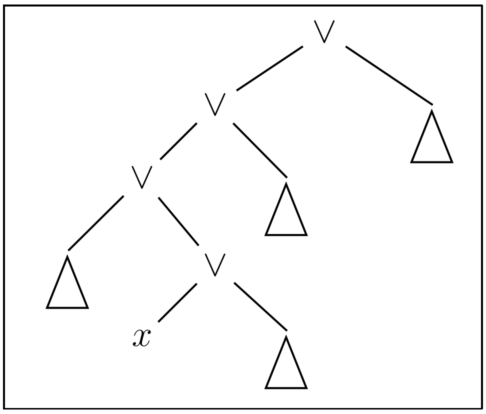

Figure 3.2: A tree counted by the generating function $g_x(z)$, where $\triangle$ denotes an arbitrary tree

Let h_x be the generating function of trees given by $t_1 \vee t_2$ (or $t_2 \vee t_1$) where t_1 is a tree counted by g_x and t_2 is a tree counted by $g_{\bar{x}}$, i.e. simple tautologies realized by x, where x and $\bar{x}$ lie in different subtrees of the root (c.f. Figure 3.3).

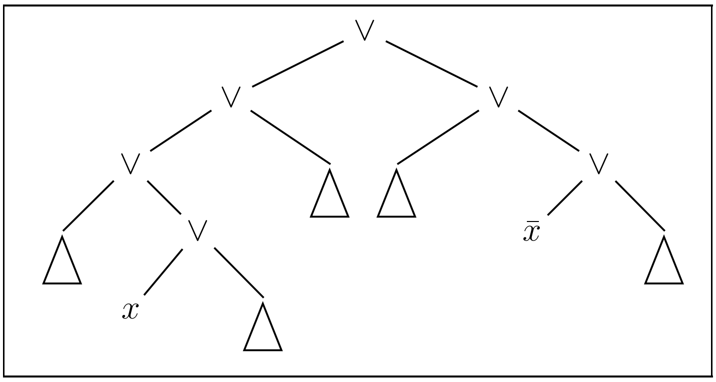

Figure 3.3: A tree counted by the generating function $h_x(z)$, where $\triangle$ denotes an arbitrary tree

Obviously, $h_x(z) = 2g_x^2(z)$. Now, let $G_x(z)$ be the generating function of simple tautologies realized by the variable x, and $\bar{G}_x(z)$ be the generating function of trees that are not simple tautologies realized by x. Again by decomposing and analyzing the label of the root, we get:

$$\bar{G}_x = T(z)^2 + \bar{G}_x(z)^2 - h_x(z) + 2nz.$$

In particular, if the root is labelled by an $\vee$, neither of the two subtrees can be a simple tautology and additionally the whole tree cannot be of the shape depicted in Figure 3.3. Solving this equation, we obtain an explicit expression of $\bar{G}_x(z)$, and $G_x(z) = T(z) - \bar{G}_x(z)$ yields an expression for $G_x(z)$, where Z denotes $Z := \sqrt{1 - 16nz}$:

$$G_x(z) = \frac{1}{4}\Big(-1 - Z + \sqrt{6 + 6Z - 2\sqrt{2 + 2Z - 16nz + 16z} - 2Z\sqrt{2 + 2Z - 16nz + 16z} - 48nz + 16z}\Big) \tag{3.8}$$

By Proposition 3.12, $\lim_{m\to\infty} \frac{ST_m}{T_m} = \lim_{m\to\infty} \frac{G_m}{T_m} + \mathcal{O}\left(\frac{1}{n^2}\right)$ when n tends to infinity. Due to Lemma 3.11 we can compute the ratio

$$\lim_{m\to\infty} \frac{G_m}{T_m} = \lim_{z\to\frac{1}{16n}} \frac{G'(z)}{T'(z)} = \frac{3}{4n} + O\left(\frac{1}{n^2}\right)$$

where $G(z) = nG_x(z)$ as defined earlier. Thus,

$$\lim_{m\to\infty} \frac{ST_m}{T_m} = \frac{3}{4n} + O\left(\frac{1}{n^2}\right)$$

when n tends to infinity. Since, when n tends to infinity, asymptotically almost every tautology is a simple tautology, this implies

$$\lim_{m\to\infty} \mathbb{P}_{m,n}(True) = \lim_{m\to\infty} \frac{ST_m}{T_m} + O\left(\frac{1}{n^2}\right).$$

□

To prove Proposition 3.12, we need to state some definitions that Kozik used for his proof [47] in a binary version.

Definition 3.14. *A pattern language $\tilde{L}$ is a set of planar trees with internal nodes labelled by $\wedge$ or $\vee$, and external nodes labelled by $\bullet$ or $\square$. The leaves labelled by $\square$ are called* placeholders *and the $\bullet$ are called* pattern leaves. *We define $s(x, y)$ as the generating function of $\tilde{L}$, with x marking the pattern leaves and y marking the placeholders.*

Given a pattern language $\tilde{L}$, we will denote by L the set of planar labelled trees with internal nodes labelled by $\wedge$ or $\vee$, and external nodes labelled by literals or placeholders, such that if we replace every literal by a $\bullet$, we obtain a tree of $\tilde{L}$. Therefore, $s(2nx, y)$ is the generating function of L.

Given a set of trees $\mathcal{T}$, we define $\tilde{L}[\mathcal{T}]$ (resp. $L[\mathcal{T}]$) as the set of trees obtained by taking an element of $\tilde{L}$ (resp. L) and plugging an element of $\mathcal{T}$ in every placeholder.

Given two pattern languages L and M, we define the composition $L[M]$ of L and M to be the pattern language obtained by plugging M-patterns into the placeholders of the elements of L.

Definition 3.15. *A pattern language L is* unambiguous *if for every family $\mathcal{T}$ every element of $L[\mathcal{T}]$ can be constructed in only one way.*
A pattern language L is subcritical *for $\mathcal{T}$ if the generating function $t(z)$ of $\mathcal{T}$ has a square root singularity ρ and if $s(x, y)$ is analytic in some set $\{(x, y) : |x| \leq \rho + \epsilon,\ |y| \leq t(\rho) + \epsilon\}$.*

Definition 3.16. *A variable x is essential for a function if putting x to $False$ or $True$ does change the restricted function.*

Remark. An essential variable appears in every tree representation of the function.

Definition 3.17. *If t is an element of $L[\mathcal{T}]$, we say that t has q L-repetitions if q equals the difference between the number of its L-pattern leaves and the number of distinct variables (and not literals) that appear in its L-pattern leaves.*
We say that t has q L-restrictions if q equals the number of its L-repetitions plus the number of essential variables of t that appear at least once in its L-pattern leaves.

For an example of repetitions and restrictions, see Figure 3.4.

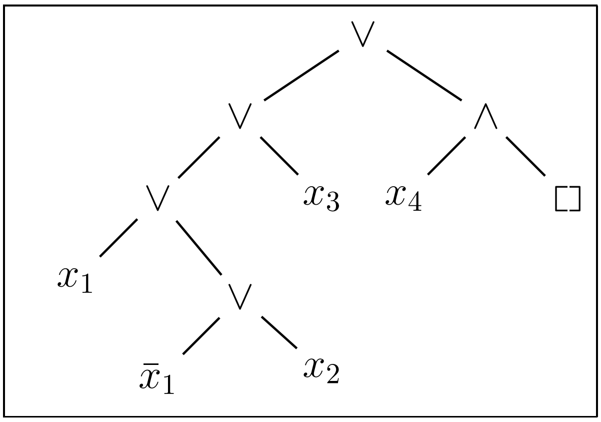

Figure 3.4: A binary tree with one repetition and one restriction - note that none of the variables x_1, x_2, x_3, x_4 are essential as the tree computes a tautology.

Kozik proved the following theorem:

Theorem 3.18 ([47])**.** *Let L be a binary unambiguous pattern language which is subcritical for $\mathcal{T}$. We denote by $L[\mathcal{T}]^{[k]}_{m,n}$ (resp by $L[\mathcal{T}]^{[\geq k]}_{m,n}$) the number of elements of $L[\mathcal{T}]$ of size m which have k (resp. at least k) L-restrictions. Then,*

$$\lim_{m\to\infty} \frac{L[\mathcal{T}]^{[\geq k]}_{m,n}}{T_m} \sim \lim_{m\to\infty} \frac{L[\mathcal{T}]^{[k]}_{m,n}}{T_m} \sim \frac{d}{n^k}$$

when n tends to infinity, and d is a constant.

Due to this theorem, we can now prove Proposition 3.12:

Proof of Proposition 3.12. Let us consider the pattern language $S = \bullet | S \vee S | \square \wedge \square$ (c.f. [47]). The set of all terms computing $True$ with exactly i S-restrictions is exactly K_i. Therefore, thanks to Theorem 3.18, we get that, when n tends to infinity:

$$\lim_{m\to\infty} \frac{\#K_i^m}{T_m} = O\left(\frac{1}{n^i}\right).$$

Therefore,

$$\lim_{m\to\infty} \frac{\#K_2^m + \ldots + \#K_n^m}{T_m} = O\left(\frac{1}{n^2}\right) + (n-2)\, O\left(\frac{1}{n^3}\right) = O\left(\frac{1}{n^2}\right).$$

□

3.2.2 Associative planar trees.

To compute the limit of $\mathbb{P}^a_n(True)$ when n tends to infinity, we again use simple tautologies, and prove that asymptotically every tautology is a simple tautology. Therefore, we will generalize Theorem 3.18 to associative trees. We will prove

Theorem 3.19. *The limiting probability of True in the associative case is:*

$$\lim_{m\to\infty} \mathbb{P}^a_{m,n}(True) = \frac{51-36\sqrt{2}}{n} + O\left(\frac{1}{n^2}\right).$$

Let us first show that Theorem 3.18 can be generalized to the associative case, and then use it to show Theorem 3.19.

Generalization of Kozik's theorem to associative trees

Theorem 3.20. *Let L be an unambiguous pattern language where all nodes have out-degree different from 1, which is subcritical for $\mathcal{A}$. We denote by $L[\mathcal{A}]^{[k]}_{m,n}$ (resp by $L[\mathcal{A}]^{[\geq k]}_{m,n}$) the number of elements of $L[\mathcal{A}]$ of size m which have k (resp. at least k) L-restrictions. Then,*

$$\lim_{m\to\infty} \frac{L[\mathcal{A}]^{[\geq k]}_{m,n}}{A_m} \sim \lim_{m\to\infty} \frac{L[\mathcal{A}]^{[k]}_{m,n}}{A_m} \sim \frac{d}{n^k}$$

when n tends to infinity, and d is a constant.

The proof of the generalization works analogously to the one of Theorem 3.18 in [47], still we will state the main ideas as they will be useful in the following.

Let $\tilde{\mathcal{A}}$ be the family of associative trees with leaves unlabelled, and let $t \in \tilde{L}[\tilde{\mathcal{A}}]_m$ with l L-pattern leaves. For any $r \leq k$, the number of different leaf-labellings of t which give r L-repetitions and k L-restrictions is:

$$\left\{ {l \atop l-r} \right\} \binom{v}{k-r} (l-r)^{\underline{k-r}} (n-v)^{\underline{l-r-(k-r)}} n^{m-l} 2^m.$$

where $x^{\underline{y}} = x(x-1)\dots(x-y+1)$, $\left\{ {y \atop x} \right\}$ are the Stirling numbers of second kind[2], and v stands for the number of essential variables. In this formula, the different terms of the product represent, from left to right:

- the number of partitions of the L-pattern leaves into $l-r$ classes (leaves in the same class will be labelled by the same variable),
- the number of different choices for the $k-r$ essential variables that appear in the L-pattern leaves,
- the number of different assignments of these $k-r$ essential variables to the $l-r$ classes of the first term,
- the number of assignments of non-essential variables to the remaining classes of the L-pattern leaves,

[2]The Stirling number of second kind $\left\{ {n \atop k} \right\}$ counts the number of ways to partition a set of n elements into k non empty sets.

- the number of assignment of variables to the leaves that are not L-pattern leaves,
- and the number of ways to distribute the negations.

Then, the following Proposition is immediate:

Proposition 3.21. *Given an associative tree $t \in \tilde{L}[\tilde{\mathcal{A}}]_m$ with leaves unlabelled, the number of leaf-labellings of t which make it have k L-restrictions is:*

$$(n-v)^{\underline{l-k}}\, n^{m-l}\, 2^m w_{v,k}(l)$$

where $w_{v,k}(l) = \sum_{r=0}^{k} \left\{ {l \atop l-r} \right\} \binom{v}{k-r} (l-r)^{\underline{k-r}}$ *is a polynomial in l.*

As one can check reading the proof of [47, Lemma 2.7], the following proposition holds when $\mathcal{T}$ is a set of (leaf-unlabelled) associative trees:

Proposition 3.22. *Let $\tilde{\mathcal{T}}$ be a set of trees whose generating function $t(z) = \sum t_m z^m$ has a unique dominating singularity ρ in $\mathcal{R}^+$ of the squareroot type. Let $\tilde{L}$ be an unambiguous pattern language, subcritical for $\tilde{\mathcal{T}}$. Let $\tilde{L}[\tilde{\mathcal{T}}](m,l)$ denote the number of trees from $\tilde{L}[\tilde{\mathcal{T}}]$ of size m with exactly l pattern leaves. Finally, let $w(l)$ be a non zero polynolmial. Then,*

$$\lim_{n\to\infty} \frac{\sum_{l\geq 0} \tilde{L}[\tilde{\mathcal{T}}](m,l)w(l)}{t_m} = c_w$$

for some nonnegative real c_w.

Moreover, if $w(l)$ has nonnegative values and is positive at some point l_0, and if $\tilde{L}$ contains a pattern with l_0 non pattern leaves and at least one placeholder, then $c_w \neq 0$.

Thanks to those propositions, we can now prove the extension of Theorem 3.18 to associative trees:

Proof of Theorem 3.20. Let L be an associative pattern and $\tilde{A}$ the family of trees from $\mathcal{A}$ with leaves unlabelled. We have, thanks to Proposition 3.21:

$$\frac{L[\mathcal{T}]^{[k]}_{m,n}}{A_m} = \frac{2^m \sum_{l\geq 0} \tilde{L}[\tilde{\mathcal{T}}](m,l) w_{k,v}(l)(n-v)^{\underline{l-k}} n^{m-l}}{A_m}; \tag{3.9}$$

and this implies:

$$\frac{L[\mathcal{T}]^{[k]}_{m,n}}{A_m} \leq \frac{2^m \sum_{l\geq 0} \tilde{L}[\tilde{\mathcal{T}}](m,l) w_{k,v}(l) n^{l-k} n^{m-l}}{(2n)^m \tilde{A}_m}. \tag{3.10}$$

Thanks to Proposition 3.22, we get:

$$\lim_{m\to\infty} \frac{L[\mathcal{T}]^{[k]}_{m,n}}{A_m} \leq \lim_{m\to\infty} \frac{2^m \sum_{l\geq 0} \tilde{L}[\tilde{\mathcal{T}}](m,l) w_{k,v}(l) n^{m-k}}{(2n)^m \tilde{A}_m} \sim \frac{c_{k,v}}{n^k} \tag{3.11}$$

when n tends to infinity. Moreover, we can check that $c_{k,v}$ is positive. The lower bound can be handled by the same method, but it won't be useful in the commutative cases. Therefore, we do not write it down, but it can be found in [47] for the binary case. The proof is exactly the same in the associative case.

It follows that

$$\lim_{m\to\infty} \frac{L[\mathcal{T}]^{[k]}_{m,n}}{A_m} \sim \frac{d}{n^k}$$

when n tends to infinity. Moreover, we can see that:

$$\frac{L[\mathcal{T}]^{[\geq k]}_{m,n}}{A_m} \leq \frac{2^m \sum_{l\geq 0} \tilde{L}[\tilde{\mathcal{T}}](m,l) w_{k,v}(l) n^{m-k}}{A_m},$$

and since

$$\lim_{m\to\infty} \frac{L[\mathcal{T}]^{[k]}_{m,n}}{A_m} \leq \lim_{m\to\infty} \frac{L[\mathcal{T}]^{[\geq k]}_{m,n}}{A_m},$$

the theorem is proved. □

Associative tautologies

Proposition 3.23. *In the associative model, asymptotically almost every tautology is a simple tautology when n tends to infinity.*

The proof is very similar to the proof of the binary case which can be found in Kozik's paper [47]. First we need to introduce a pattern:

$$\begin{aligned}\hat{N} &= \bullet | \check{N} \wedge \square | \check{N} \wedge \square \wedge \square | \dots \\ \check{N} &= \bullet | \hat{N} \vee \hat{N} | \hat{N} \vee \hat{N} \vee \hat{N} | \dots \\ R &= \{\hat{N}, \check{N}\},\end{aligned} \tag{3.12}$$

where $R = \{\hat{N}, \check{N}\}$ means we start with either an $\vee$- or an $\wedge$-node, and use the according pattern $\check{N}$ or $\hat{N}$, and then use both partial patterns alternatingly until the process finishes. Then R is an unambigous pattern language.

Lemma 3.24. *The pattern R is subcritical for associative trees.*

Proof. The generating function $p(x, y)$ of the unlabelled pattern $\tilde{R}$ is given by

$$p(x, y) = \hat{p}(x, y) + \check{p}(x, y) - x,$$

where $\hat{p}(x, y)$ (resp. $\check{p}(x, y)$) is the generating function of the partial patterns $\hat{N}$ (resp. $\check{N}$). These two generating functions satisfy the following system:

$$\begin{aligned}\check{p} &= x + \frac{\hat{p}^2}{1-\hat{p}} \\ \hat{p} &= x + \frac{y}{1-y}\check{p}\end{aligned} \tag{3.13}$$

Solving this system, we get

$$\hat{p}(x, y) = \frac{1}{2}\left(x - y - 1 - \sqrt{(x-y-1)^2 - 4x}\right). \tag{3.14}$$

and hence

$$\hat{p}(2nx, y) = \frac{1}{2}\left(2nx - y - 1 - \sqrt{(2nx-y-1)^2 - 8nx}\right). \tag{3.15}$$

Recall that (cf. (3.2))

$$A(z) = \frac{1}{2}\left(1 - 2nz - \sqrt{1 - 12nz + 4n^2z^2}\right),$$

$$A(\alpha_n) = \sqrt{2} - 1 \text{ and } \alpha_n = \frac{3 - 2\sqrt{2}}{2n}.$$

To prove that α_n is the dominant singularity of $p(2nz, A(z))$, it is enough to prove that it is the dominant singularity of $\hat{p}(2nz, A(z))$ and $\check{p}(2nz, A(z))$. Actually, $\hat{p}(2nx, y)$ and $\check{p}(2nx, y)$ are analytic in $\mathbb{C}^2 \setminus \{(x, y) \mid (2nx - 1 - y)^2 = 8nx\}$. For all n,

$$(2n\alpha_n - \sqrt{2})^2 = 9(3 - 2\sqrt{2}) > 8n\alpha_n = 4(3 - 2\sqrt{2}),$$

and due to nonnegative coefficients the inequality holds for all $z \in \mathbb{R}, 0 \leq z \leq \alpha_n$. Thus the dominant singularity of $\hat{p}(2nz, A(z))$ is α_n and R is subcritical for associative trees. □

Remark. The R pattern has an interesting property: if one valuates all the R-pattern leaves of a tree to $False$, then the whole tree itself computes $False$. This can be checked by induction. If the pattern is only a leaf, it returns $False$. If the root is an $\vee$-node, then all subtrees of the root are patterns returning $False$ by induction hypothesis. If the root is an $\wedge$-node, the leftmost subtree is a pattern returning $False$ by the induction hypothesis. Thus the whole tree computes $False$ in all cases.

This property is the key point of the following proof.

Remark. The pattern R is a generalization of the pattern $N = \bullet|N \vee N|N \wedge \Box$ defined in Kozik's paper [47] to handle the proof in the binary planar case.

Proof of Proposition 3.23. Let us consider a tree t with exactly one $R[R]$-restriction, which computes $True$. This restriction has to be a repetition.

If the repetition is of the kind x/x, then we can assign all the R-pattern leaves to $False$, and with this assignment the whole tree computes $False$, which is impossible.

Thus the repetition has to be an $x/\bar{x}$ repetition. Let us first assume that the repetition does not appear among the R-pattern leaves. Thus we can assign all these leaves to $False$, and then the whole tree computes $False$. This is impossible. The repetition must occur in the R-pattern leaves. Let us assume that there is a node ν labelled by $\wedge$ between the leaf labelled by x (resp. $\bar{x}$) and the root. Then, the subtree rooted at ν has shape $t_1 \wedge t_2 \wedge \ldots \wedge t_s$. Let us assume that x appears in t_j. Then, we can assign all the $R[R]$-pattern leaves of the other subtrees $(t_i)_{i \neq j}$, and all the $R[R]$-pattern leaves of the whole tree except those in the subtree rooted at ν to $False$. This makes the whole tree compute $False$, which is impossible.

Thus, x and $\bar{x}$ are linked to the root by an $\vee$-only path. As the trees are stratified, the only possibility for t is to be a simple tautology. Thus every term with exactly one $R[R]$-restriction computing $True$ is a simple tautology.

Moreover, there are no terms computing $True$ with zero $R[R]$-restrictions, and the number of trees computing $True$ with more than two $R[R]$-restrictions is negligible in comparison to the number of simple tautologies, by Theorem 3.18 which can be applied thanks to Lemma 3.24. □

We are now able to prove Theorem 3.19 by counting associative simple tautologies.

Proof of Theorem 3.19. Let G_x be the generating function counting the number of simple tautologies realized by x and such that x and $\bar{x}$ appear only once on the first level. Then, $G_x(z) = \sum_{l \geq 2} l(l-1)(A(z)-2)^{l-2}$. If x or $\bar{x}$ appear at least twice in the first generation, the tree has at least two $P[P]$-restrictions, and the set of such trees is negligibly small compared to the set counted by G_x. Thus, asymptotically speaking (when n tends to infinity), G_x counts the set of simple tautologies realized by x.

Finally, note that in view of Theorem 3.20, the assertion of Proposition 3.12 extends to the associative case. So a Maple computation giving

$$\lim_{z\to\alpha_n} \frac{G'(z)}{A'(z)} \sim \frac{51-36\sqrt{2}}{n}$$

completes the proof of Theorem 3.19. □

3.2.3 Binary commutative trees

The generating function of binary commutative And/Or trees, $C(z) = \sum_m C_m z^m$, is given in (3.3), and γ_n is the dominant positive singularity of $C(z)$. Solving the singularity system of equations

$$y = 2nz + y^2 + F(z) \tag{3.16}$$

$$1 = 2y \tag{3.17}$$

we obtain $C(\gamma_n) = \frac{1}{2}$ and $\gamma_n = \frac{1}{8n} - \frac{C(z^2)}{2n}$. As $C(z) = 2nz + \mathcal{O}(z^2)$ by inserting into the equation we can further derive $\gamma = \frac{1}{8n}\left(1 - \frac{1}{8n}\right) + \mathcal{O}(\frac{1}{n^3})$. As we need more terms in some of our calculations, we do a more refined analysis with Maple and further obtain

$$\gamma_n = \frac{1}{8n}\left(1 - \frac{1}{8n} + \frac{7}{256n^3}\right) + \mathcal{O}(\frac{1}{n^4}). \tag{3.18}$$

Theorem 3.25. *The limiting probability of the function True in the binary commutative case $\mathbb{P}^c_{m,n}(True)$ is given by*

$$\lim_{m\to\infty} \mathbb{P}^c_{m,n}(True) = \frac{641}{1024}\frac{1}{n} + \mathcal{O}(\frac{1}{n^2}).$$

To prove the theorem we will extend the method of pattern languages of Kozik to the non-planar case. We consider binary commutative trees, together with a *half-embedding*, that is certain branches of the tree will be planar and some will stay non-planar. We use the planar pattern language known from Section 3.2.1, given by

$$N = \bullet | N \vee N | N \wedge \Box.$$

As N is planar, it is unambigous for any tree family. A tree of $N[\mathcal{C}]$ is a "mobile", that is, the pattern-trees consisting of internal nodes and $\bullet$ and $\Box$-leaves are planar, while the terms substituted into the $\Box$-nodes are still nonplanar trees.

Remark. While in the planar cases the considered pattern was subcritical for the non-leaf labelled family of trees, this is not the case for commutative trees. Therefore, the strategy will be different as before.

Generalization of Kozik's theorem to commutative trees

Theorem 3.26. *Let L be a labelled planar binary unambiguous pattern language with $\ell(x,y)$ its generating function. Further assume that the coefficients $A_l(y)$, given by*

$$\ell(x,y) = \sum_{l\geq 0}\sum_{i\geq 0} s_{i,l} y^i x^l = \sum_{l\geq 0} A_l(y) x^l$$

are subcritical for $C(z)$.
We denote by $L[\mathcal{C}]^{[k]}_{m,n}$ (resp by $L[\mathcal{C}]^{[\geq k]}_{m,n}$) the number of elements of $L[\mathcal{C}]$ of size m which have k (resp. at least k) L-restrictions, and by $L[\mathcal{C}]_m$ the number of elements of $L[\mathcal{C}]$ of size m.
Then,

$$\lim_{m\to\infty} \frac{L[\mathcal{C}]^{[\geq k]}_m}{C_m} = \mathcal{O}\left(\frac{1}{n^k}\right) \quad \text{and} \quad \lim_{m\to\infty} \frac{L[\mathcal{C}]^{[k]}_m}{C_m} = \mathcal{O}\left(\frac{1}{n^k}\right)$$

when n tends to infinity.

Remark. In the planar cases, the patterns N and R we considered fulfilled $N[\mathcal{T}] = \mathcal{T}$ and $R[\mathcal{A}] = \mathcal{A}$, respectively. For commutative trees, this is not the case. The proofs of Theorems 3.18 and 3.20 rely completely on planar structures and subcriticality, which is not given anymore. Hence the above generalization of Theorem 3.18 to mobile structures is indeed different and we will need additional arguments to show that asymptotically almost every tautology is a simple tautology.
Definitions of restrictions are valid in the mobile case, as pattern-leaves appear in planar parts of a mobile. We will adapt the proof, relying on the sketch in Section 3.2.2.

Let $\tilde{L}$ be a planar pattern and $\mathcal{C}$ be a family of commutative trees. Let Λ be an element of $\tilde{L}[\mathcal{C}]$ of size m with l pattern leaves. Note that the leaves of the non-planar parts are labelled while the pattern leaves are unlabelled. For any $r \leq k$, the number of different labellings of the pattern leaves of Λ which give r L-repetitions and k L-restrictions is given by

$$\left\{ {l \atop l-r} \right\} \binom{v}{k-r} (l-r)^{\underline{k-r}} (n-v)^{\underline{l-r-(k-r)}} 2^l,$$

where, as in the planar case, $x^{\underline{y}} = x(x-1)\dots(x-y+1)$, $\left\{ {y \atop x} \right\}$ are the Stirling numbers of second kind, and v stands for the number of essential variables. The different terms of the product again represent, from left to right:

- the number of partitions of the L-pattern leaves into $l-r$ classes (leaves in the same class will be labelled by the same variable),
- the number of different choices for the $k-r$ essential variables that appear in the L-pattern leaves,
- the number of different assignments of these essential variables to $k-r$ of the $l-r$ classes of the first term,
- the number of assignments of non-essential variables to the remaining classes of the L-pattern leaves,

- and the number of distribution of the negations.

With this, the following version of Proposition 3.21 is obvious:

Proposition 3.27. *Given a binary mobile $\Lambda \in \tilde{L}[\mathcal{C}]$ with pattern leaves unlabelled, the number of leaf-labellings of Λ which make it have k L-restrictions satifies*

$$\sharp(\mathit{labellings})_k = (n-v)^{\underline{l-k}}\, n^{m-l}\, 2^l w_{v,k}(l)$$

where $w_{v,k}(l) = \sum_{r=0}^{k} \left\{ {l \atop l-r} \right\} \binom{v}{k-r} (l-r)^{\underline{k-r}}$ *is a polynomial in l.*

We adapt Proposition 3.22 to our needs.

Proposition 3.28. *Let L be an unambiguous labelled pattern language, with $\ell(x,y)$ its generating function, and let $\mathcal{T}$ be a family of leaf-labelled trees with generating function $T(z)$. Further assume that the coefficients $A_l(y)$, given by $\ell(x,y) = \sum_{l\geq 0} A_l(y)x^l$ are subcritical for $T(z)$.*
Let $L[\mathcal{T}](m,l)$ be the number of trees of $L[\mathcal{T}]$ of size m with exactly l pattern leaves and $w(l)$ be a non-zero polynomial of degree λ. Then,

$$\lim_{m\to\infty} \frac{\sum_{l=0}^{N} L[\mathcal{T}](m,l)w(l)}{T_m} = c_w$$

for some nonnegative real c_w, where N is some fixed integer.

Proof. The generating function of the numerator $\sum_{l=0}^{N} L[\mathcal{T}](m,l)w(l)$ is denoted by $\ell_w(x,y)$. Moreover, $w(l) = \sum_{j=0}^{\lambda} w_j l^{\underline{j}}$ is a representation of the polynomial w, and $\ell_N(x,y) = \sum_{l=0}^{N} A_l(y)x^l$ is the truncation of $\ell(x,y) = \sum_{l\geq 0} A_l(y)x^l$. Note that,

$$x^j \frac{\partial^j \ell_N(x,y)}{\partial x^j} = \sum_{l=0}^{N} l^{\underline{j}} A_l(y)x^l.$$

Thus

$$\sum_{j=0}^{\lambda} w_j x^j \frac{\partial^j \ell_N(x,y)}{\partial x^j} = \sum_{l=0}^{N} w(l)A_l(y)x^l.$$

Therefore, the generating function $\ell_w(x,y)$, is a linear combination of the derivatives of $\ell_N(x,y)$, which are finite sums of terms which are subcritical for $T(z)$. Hence, $\ell_w(z,C(z))$ and $T(z)$ have the same radius of convergence. By [47, Observation 3.3] every subcritical summand has a square root expansion around the singularity, if $T(z)$ has a square root singularity, hence the type of singularity of $\ell_w(z,C(z))$ is also of order $1/2$ or of higher order, if there is a cancellation.

Thanks to a transfer lemma (Lemma 1.7), we easily get

$$\frac{[z^m]\ell_w(z,C(z))}{[z^m]T(z)} \sim \mathit{const},$$

when m tends to infinity. Therefore,

$$\lim_{m\to\infty} \frac{\sum_{l\geq 0} L[\mathcal{T}](m,l)w(l)}{T_m} = c_w$$

for some nonnegative constant c_w. Further c_w is positive if there is no cancellation and zero otherwise. ☐

Proof of Theorem 3.26. We have, thanks to Proposition 3.27:

$$\frac{L[\mathcal{C}]_m^{[k]}}{C_m} = \frac{\sum_{l=0}^{N} \tilde{L}[\mathcal{C}](m,l) w_{k,v}(l)(n-v)^{\underline{l-k}}\, 2^l}{C_m},$$

where $N = n - v + k$, because for larger l the factor $(n-v)^{\underline{l-k}}$ gives 0. This implies

$$\frac{L[\mathcal{C}]_m^{[k]}}{C_m} \leq \frac{\sum_{l=0}^{N} \tilde{L}[\mathcal{C}](m,l) w_{k,v}(l) n^{l-k}\, 2^l}{C_m} = \frac{\sum_{l=0}^{N} L[\mathcal{C}](m,l) w_{k,v}(l)}{C_m} \cdot \frac{1}{n^k} \tag{3.19}$$

because $L[\mathcal{C}](m,l) = (2n)^l \tilde{L}[\mathcal{C}](m,l)$. And therefore, by applying Proposition 3.28, we get that

$$\lim_{m\to\infty} \frac{L[\mathcal{C}]_m^{[k]}}{C_m} \leq \frac{c_w}{n^k}.$$

□

Commutative tautologies

Proposition 3.29. *Almost every binary commutative And/Or tree computing the function True is a simple tautology.*

Before proving Proposition 3.29, we introduce some half-embedding of a tree t into the plane: Start at the root and choose a left-right order of the children of the root. If the root is an $\wedge$-node, proceed recursively with the root of the left subtree, the right subtree remains non-planar. If the root is an $\vee$-node, proceed recursively with both subtrees. If doing so we meet a leaf, it is a pattern leaf. Doing this for the whole tree t, we obtain an element of $N[\mathcal{C}]$, where the non-planar subtrees are the structures substituted into the placeholders. Now do the same half-embedding starting at every root of a non-planar subtree. Thus we obtain an element of $N[N][\mathcal{C}]$. Note that different trees $t_1 \neq t_2 \in \mathcal{C}$ will create different patterns $N[t_1]$ and $N[t_2]$, thus the function $\mathcal{C} \to N[\mathcal{C}]$ described above is an injection. Of course, there are several ways to embed a tree t with the above method.

Definition 3.30. *Let t be a non-plane tree and choose a half-embedding of t as described above, such that the resulting $N[N]$-pattern has a minimal number of $N[N]$-restrictions. We call such an embedding a minimal $N[N]$-embedding of t .*

Note that there could be various minimal embeddings for one tree.

Lemma 3.31. *Let t be a tree computing the function True. Then a minimal $[N]$-embedding has at least one restriction.*

Proof. Suppose $N[t]$ has no restriction and set all pattern leaves to *False*. We proceed inductively. If $N[t]$ is just a leaf, it returns *False*. If the root of $N[t]$ is an $\wedge$-node, the left subtree is a pattern and will, by the induction hypothesis, return *False*, thus the whole tree returns *False*. If the root of $N[t]$ is a $\vee$ node, both subtrees are patterns returning *False* by the induction hypothesis. Thus the whole tree returns *False*. □

Lemma 3.32. *Let t be a tree whose minimal $N[N]$-embedding has exactly one $N[N]$-restriction. Then t is a simple tautology.*

Proof. There are two cases to distinguish.

- The restriction is of type x/x. Set all N-pattern leaves to $False$. The same arguments as in the proof of Lemma 3.31 show that t returns $False$.

- The restriction is of type $x/\bar{x}$. Then the restriction appears on the first level, that is, in $N[t]$, as otherwise setting all N-pattern leaves to $False$ would lead to a tree computing $False$ by the same arguments as before. If t is not a simple tautology, then there exists at least one node labelled with $\wedge$ on the path from the root to either x or its negation. Let t_1 be the non-planar subtree rooted at such a node. After the second N-embedding, the $N[t_1]$ pattern contains no repetition as the whole tree $N[N][t]$ had only one $N[N]$-repetition, thus it is easy to have t_1 contribute $False$ by setting all $N[t_1]$-pattern leaves to $False$. Then the $\wedge$ at v gives $False$, thus t does not compute the function $True$. Thus, every tautology t which has a minimal $N[N]$-embedding with a single repetition is a simple tautology.

□

In order to take advantage of the above arguments on $N[N]$-embeddings and hence to prove Proposition 3.29, we need to check if the assumptions of Proposition 3.28 are fulfilled. We prove the following result in full generality, as it will be useful later in this chapter.

Lemma 3.33. *Let L be a pattern language with generating function $\ell(x,y) = \sum_{l\geq 0} A_l(y)x^l$ and with $A_0(y) = 0$, and let L^r be its r-th power for any $r \in \mathbb{N}$, with*

$$\ell^*(x,y) = \underbrace{\ell(x,(\ell(x,\cdots \ell(x}_{r-\text{times}},y)\cdots))) = \sum_{l\geq 0} A^*_l(y)x^l$$

its generating function. Further let $\mathcal{T}$ be a family of trees with generating function $T(z)$. Assume that, for all $l \geq 0$, $A_l(y)$ is subcritical for $T(z)$. Then $A^*_l(y)$ is subcritical for $T(z)$.

Proof. First note that $A_0(y) = 0$ means that every pattern in L has at least one pattern leaf. Obviously, this property still holds for $A^*_0(y)$.

We prove the statement by induction: the case $r = 1$ is true by assumption. Let us assume that the result holds for r, and let $\bar{s}(x,y) = \sum_{l\geq 0} \bar{A}_l(y)x^l$ be the generating function of L^r with $\bar{A}_l(y)$ being subcritical for $T(z)$. We want to show that $[x^l]s(x,\bar{s}(x,y))$ is subcritical for $T(z)$. It is sufficient to show that $[x^\lambda]A_l(\bar{s}(x,y))$ is subcritical for $T(z)$ for all λ, because $s(x,\bar{s}(x,y)) = \sum_{l\geq 0} A_l(\bar{s}(x,y))x^l$ and $A_l(\bar{s}(x,y))$ is a power series in x. Then $[x^l]s(x,\bar{s}(x,y)) = \sum_{j=0}^{l}[x^{l-j}]A_j(\bar{s}(x,y))$, which is a finite sum of such coefficients.

$$\begin{aligned}
[x^\lambda]A_l(\bar{s}(x,y)) &= [x^\lambda]\sum_{j/geq0} s_{l,j}\bar{s}(x,y)^j \\
&= [x^\lambda]\sum_{j\geq 0} s_{l,j}\left(\sum_{\mu} x^\mu \bar{A}_\mu(y)\right)^j \\
&= [x^\lambda]\sum_{j\geq 0} s_{l,j}\sum_{\mu_1,\ldots,\mu_j} x^{\sum \mu_i}\bar{A}_{\mu_1}\cdots\bar{A}_{\mu_j} \\
&= \sum_{j\geq 0} s_{l,j}\sum_{\mu_1+\ldots+\mu_j=\lambda} \bar{A}_{\mu_1}\cdots\bar{A}_{\mu_j}.
\end{aligned}$$

As $\bar{A}_0(y) = 0$, $\mu_i > 0$ for $i = 1, \dots j$, and hence we have a maximum of λ factors in every summand, that is,

$$[x^\lambda]A_l(\bar{s}(x,y)) = \sum_{j=0}^{\lambda} s_{l,j} \sum_{\mu_1+\ldots+\mu_j=\lambda} \bar{A}_{\mu_1} \cdots \bar{A}_{\mu_j}.$$

This is a finite sum of finite products of subcritical factors and hence it is subcritical for $T(z)$. ☐

Additionally, we prove subcriticality of the functions $A_l(z)$ for $C(z)$ in order to apply Theorem 3.26.

Lemma 3.34. *Let $s(x, y) = \sum_{l\geq 0} A_l(y)x^l$ be the generating function of the pattern N. Then the functions $A_l(y)$ are subcritical for $C(z)$.*

Proof. Thanks to symbolic arguments and to the recursive definition of $\tilde{N} = \bullet|\tilde{N}\vee\tilde{N}|\tilde{N}\wedge\Box$, we get that:

$$s(x, y) = 2nx + s(x, y)^2 + ys(x, y).$$

Solving this equation gives $s(x, y) = \frac{1}{2}\left(1 - y - \sqrt{(y-1)^2 - 8nx}\right)$. We want to deduce an explicit formula for the $A_l(y)$ from this expression. With $s(0, 0) = 0$, we obtain the following power series in x.

$$\begin{aligned} s(x,y) &= \frac{1-y}{2} - \frac{1}{2}\sqrt{(y-1)^2}\sqrt{1 - \frac{8nx}{(y-1)^2}} \\ &= \frac{1-y}{2} - \frac{1}{2}(1-y)\sum_{l\geq 0}\binom{1/2}{l}(-8n)^l(y-1)^{-2l}x^l. \end{aligned}$$

Therefore, $A_l(y) = -\frac{1}{2}(1-y)\binom{1/2}{l}(-8n)^l(y-1)^{-2l}$ is a rational function of y and its radius of convergence is 1. Therefore, these functions are subcritical for $C(z)$. ☐

Proof of Proposition 3.29. Let t be a tree in $\mathcal{C}$ which computes $True$. Then there is at least one variable x appearing twice in the leaves of t, because otherwise the tree computes a read-once function where every variable is essential. We half-embed t into the plane as described before.
As this N-embedding represents an injection it follows that $C_m^{(k)} \leq (N[\mathcal{C}])_m^{(k)}$. Hence, by Theorem 3.26, which can be applied thanks to Lemmas 3.33 and 3.34:

$$\frac{C_m^{(k)}}{C_m} \leq \frac{N[\mathcal{C}]_m^{(k)}}{C_m} = \mathcal{O}\left(\frac{1}{n^k}\right),$$

and thus asymptotically almost all tautologies in a binary non-planar And/Or tree are simple (and have a minimal $N[N]$-embedding with one restriction). Proposition 3.29 is thus proved. ☐

Proof of Theorem 3.25. Let $g_x(z)$ be the generating function counting the trees in $\mathcal{C}$ with a path from the root to a leaf labelled with x containing only internal nodes with label $\vee$. It is given by $g_x(z) = C(z) - \bar{g}_x(z)$ with

$$\bar{g}_x(z) = (2n-1)z + \frac{1}{2}\left(C^2(z) + C(z^2)\right) + \frac{1}{2}\left(\bar{g}_x^2(z) + \bar{g}_x(z^2)\right), \tag{3.20}$$

because a tree rooted at an $\wedge$-node cannot contain an $\vee$-only path from the root, while if the root is labelled with $\vee$ both subtrees of the root must not contain an $\vee$-only path to an x-leaf.

The generating function $G_x(z)$ which counts trees which are a simple tautology realized by x is given by $G_x(z) = C(z) - \bar{G}_x(z)$, where $\bar{G}_x(z)$ counts trees which are not simple tautologies realized by x. Similarly to $\bar{g}_x(z)$, such a tree is either rooted at an $\wedge$-node, or it is rooted at an $\vee$-node, and both subtrees of the root are not simple tautologies. Still, it could return $True$ if one of the subtrees contains an $\vee$-only path to x and the other subtree contains an $\vee$-only path to $\bar{x}$. This gives the following implicit equation for $\bar{G}_x(z)$.

$$\bar{G}_x(z) = 2nz + \frac{1}{2}\left(C^2(z) + C(z^2)\right) + \frac{1}{2}\left(\bar{G}_x^2(z) + \bar{G}_x(z^2)\right) - g_x(z)g_{\bar{x}}(z). \tag{3.21}$$

To calculate the limiting ratio of simple tautologies we need to compute $n(1-\lim_{z\to\gamma_n}\frac{\bar{G}_x'(z)}{C'(z)})$, where the factor n is the choice of x in the set of variables, and we use an analogue of Lemma 3.12. We denote by $u_n := \bar{g}_x(\gamma_n)$, $v_n := \bar{g}_x(\gamma_n^2), U_n := \bar{G}_x(\gamma_n)$ and $V_n := \bar{G}_x(\gamma_n^2)$, and compute U_n up to terms of order $\frac{1}{n^2}$. From (3.20) we get

$$u_n = (2n-1)\frac{1}{8n}\left(1-\frac{1}{8n}\right) + \frac{1}{2}\left(\frac{1}{4} + C(\gamma_n^2)\right) + \frac{1}{2}(u_n^2 + v_n) \tag{3.22}$$

$$v_n = (2n-1)\frac{1}{64n^2}\left(1-\frac{1}{8n}\right)^2 + \frac{1}{2}(C^2(\gamma_n^2) + C(\gamma_n^4)) + \frac{1}{2}(v_n^2 + \bar{g}_x(\gamma_n^4)) \tag{3.23}$$

We know that $C(z^2) = 2nz^2 + \mathcal{O}(z^4)$, hence $C(\gamma_n^2) = \frac{1}{32n}(1+\mathcal{O}(\frac{1}{8n}))^2 + \mathcal{O}\left(\frac{1}{n^3}\right)$. Inserting this into (3.23) we can compute $v_n = \frac{1}{32n} + \mathcal{O}\left(\frac{1}{n^2}\right)$, and with (3.22), $u_n = \frac{1}{2} - \frac{1}{4}\frac{1}{n} + \mathcal{O}\left(\frac{1}{n^2}\right)$. Solving the equations for U_n and V_n, up to terms of order $\frac{1}{n^2}$, we get

$$V_n = \frac{1}{32n} - \frac{7}{1024n^2} \quad \text{and} \quad U_n = \frac{1}{2} - \frac{129}{1024n^2} + \mathcal{O}\left(\frac{1}{n^4}\right)$$

Derivating $\bar{G}_x(z)$ and $\bar{g}_x(z)$, we obtain

$$\begin{aligned}
\bar{g}_x'(z) &= 2n - 1 + C(z)C'(z) + zC'(z^2) + \bar{g}_x(z)\bar{g}_x'(z) + z\bar{g}_x'(z^2) \text{ and} \\
\bar{G}_x'(z) &= 2n + C(z)C'(z) + zC'(z^2) + \bar{G}_x(z)\bar{G}_x'(z) + z\bar{G}_x'(z^2) - 2g(z)g'(z).
\end{aligned}$$

Hence

$$\begin{aligned}
\lim_{z\to\gamma_n}\frac{\bar{g}_x'(z)}{C'(z)} &= \frac{1}{1-\bar{g}_x(z)}\left(\frac{2n-1}{C'(z)} + C(z) + \frac{zC'(z^2)}{C'(z)} + \frac{z\bar{g}_x'(z^2)}{C'(z)}\right) \\
&\sim \frac{1}{1-u_n}\frac{1}{2} = 1 - \frac{1}{2n} + \frac{1}{4n^2} + \mathcal{O}(\frac{1}{n^3}) \\
\lim_{z\to\gamma_n}\frac{\bar{G}_x'(z)}{C'(z)} &= \frac{1}{1-\bar{G}_x(z)}\left(\frac{2n}{C'(z)} + C(z) + \frac{zC'(z^2)}{C'(z)} + \frac{z\bar{G}_x'(z^2)}{C'(z)} - \frac{2g(z)g'(z)}{C'(z)}\right) \\
&\sim \frac{1}{1-U_n}\left(\frac{1}{2} - 2u_n\lim_{z\to\gamma_n}\frac{\bar{g}_x'(z)}{C'(z)}\right) \sim \left(2 - \frac{129}{512n^2}\right)\left(\frac{1}{2} - \frac{1}{4n^2}\right) \\
&= 1 - \frac{641}{1024n^2} + \mathcal{O}\left(\frac{1}{n^4}\right)
\end{aligned}$$

The result of Theorem 3.25 follows immediately. □

3.2.4 Associative and commutative trees

Recall that the generating function of general And/Or-trees is given by (cf. Equation (3.5))

$$P(z) = \hat{P}(z) + \check{P}(z) - 2nz,$$

and $\hat{P}(z) = \check{P}(z)$ fulfils (cf Equation (3.6))

$$\hat{P}(z) = \exp\left(\sum_{i\geq 1} \frac{\hat{P}(z^i)}{i}\right) - 1 - \hat{P}(z) + 2nz$$

Let δ_n be the dominant positive singularity of $\hat{P}(z)$, and hence also of $P(z)$. To get $\delta_n, \hat{P}(\delta_n)$ and $P(\delta_n)$ we need to solve the singular system

$$y = e^y \cdot F(z) - 1 - y - 2nz \tag{3.24}$$
$$1 = e^y \cdot F(z) - 1 \tag{3.25}$$

with $F(z) = \exp(\sum_{i\geq 2} \hat{P}(z^i)/i) = 1 + 2n^2z^2 + \mathcal{O}(z^3)$. Therefore $F(z) \sim 1$ for $z = \mathcal{O}\left(\frac{1}{n}\right)$ and n tending to infinity, hence (3.25) gives $e^{y(z)} \sim 2$ or $y(z) \sim \ln(2)$. Inserting (3.25) into (3.24) gives $y = \frac{1+2nz}{2}$and thus the first order asymptotic of δ_n is $\delta_n \sim \frac{2\ln 2 - 1}{2n}$.

Theorem 3.35. *The limiting ratio of tautologies in the binary commutative case, $\mathbb{P}_n^{a,c}(True)$, is given by*

$$\lim_{m\to\infty} \mathbb{P}_{n,m}^{a,c}(True) \sim \frac{(2\ln 2 - 1)^2}{4}\frac{1}{n} + \mathcal{O}\left(\frac{1}{n^2}\right)$$
$$\approx \frac{0.03730583332}{n}.$$

To prove the theorem, we will again use mobiles, relying on the unambiguous pattern $R = \{\hat{N}, \check{N}\}$ from Section 3.2.2, given in (3.12), and will prove an analogue to Theorem 3.26 in the associative case.

Generalization of Kozik's theorem to associative and commutative trees

Theorem 3.36. *Let L be a labelled unambiguous pattern language where all nodes have out-degree different from 1, and $\ell(x, y)$ its generating function. Further assume that the coefficients $A_l(y)$, given by $\ell(x, y) = \sum_{l\geq 0} A_l(y)x^l$, are subcritical for $P(z)$.*
We denote by $L[\mathcal{P}]_{m,n}^{[k]}$ (resp by $L[\mathcal{P}]_{m,n}^{[\geq k]}$) the number of elements of $L[\mathcal{P}]$ of size m which have k (resp. at least k) L-restrictions. Then,

$$\lim_{m\to\infty} \frac{L[\mathcal{P}]_{m,n}^{[\geq k]}}{P_m} = \mathcal{O}\left(\frac{1}{n^k}\right) \quad and \quad \lim_{m\to\infty} \frac{L[\mathcal{P}]_{m,n}^{[k]}}{P_m} = \mathcal{O}\left(\frac{1}{n^k}\right)$$

when n tends to infinity.

The proof of Theorem 3.36 now is an easy generalization of Sections 3.2.2 and 3.2.3. We use mobiles on a non-binary planar pattern L, that is pattern leaves are on planar paths from the root, while non-planar trees have been substituted in the $\square$-nodes of the planar pattern. We can easily prove Proposition 3.27 in the associative case.

Proof of Theorem 3.36. As in previous parts, Proposition 3.27 gives:

$$\frac{L[\mathcal{P}]^{[k]}_{m,n}}{P_m} = \frac{\sum_{l=0}^{N} \tilde{L}[\mathcal{P}](m,l) w_{k,v}(l)(n-v)^{\underline{l-k}} 2^l}{P_m};$$

with $N = n - v + k$, which implies:

$$\frac{L[\mathcal{P}]^{[k]}_{m,n}}{P_m} \leq \frac{\sum_{l=0}^{N} \tilde{L}[\mathcal{P}](m,l) w_{k,v}(l) n^{l-k} 2^l}{P_m} = \frac{\sum_{l=0}^{N} L[\mathcal{P}](m,l) w_{k,v}(l)}{n^k P_m}. \tag{3.26}$$

Hence the result follows from Proposition 3.28, which was proven for all patterns. □

Non-planar associative tautologies

Proposition 3.37. *Almost every non-planar associative And/Or tree computing the function True is a simple tautology.*

Again we introduce a half-embedding of a tree t into the plane: Start at the root and choose a left to right order of the children of the root. If the root is a $\wedge$-node, proceed with the leftmost child of the root. If the root was a $\vee$-node, then do the same for every child of the root. If we end up at a leaf, this is a pattern leaf. By this procedure we obtain an element of $R[\mathcal{P}]$. Appyling the same procedure to every root of a non-planar subtree, we obtain an element of $R[R][\mathcal{P}]$, we call it an $R[R]$-embedding of t. There are several ways to embed t, choose one embedding with a minimal number of $R[R]$-restrictions. Again, the function $t \mapsto R[R]_{min}(t)$ represents an injection.

Now looking at all trees with a minimal $R[R]$-embedding having exactly one restriction, we can proceed in the same way as in the proof of Theorem 3.19 to prove that they are simple tautologies.

We again prove that the functions $A_l(y)$ of the pattern R are subcritical for $P(z)$.

Lemma 3.38. *Let $p(x,y) = \sum_{l\geq 0} A_l(y)x^l$ be the generating function of the pattern language R. The functions $A_l(y)$ are subcritical for $P(z)$.*

Proof. The generating function of the R pattern is $p(x,y) = \hat{p}(x,y) + \check{p}(x,y) - 2nx$ where

$$\hat{p}(x,y) = \frac{1}{2}\left(1 - 2nx - y - 1 - \sqrt{(2nx - y - 1)^2 - 8nx}\right).$$

Therefore,

$$\begin{aligned}
p(x,y) &= -(y+1) - \sqrt{(2nx-y-1)^2 - 8nx} \\
&= -(y+1) - \sqrt{(y+1)^2}\sqrt{1 - \frac{4nx(n-1+y)}{(y+1)^2}} \\
&= -(y+1) + (y+1)\sum_{l\geq 0} \binom{1/2}{l} (y+1)^{-2l}(-4nx)^l(n-1+y)^l
\end{aligned}$$

since $p(0,0) = 0$. Therefore, the $A_l(y)$ are rational functions with radius of convergence 1 which is bigger than δ_n. Thus $A_l(y)$ is subcritical for $P(z)$. □

Proof of Proposition 3.37. Let $t \in \mathcal{P}$ be a tree that computes True. We half-embed t and argue as in the proof of Proposition 3.29 in Section 3.2.3 to prove Proposition 3.37 with the help of Lemma 3.33, the above result and Theorem 3.36. ☐

Proof of Theorem 3.35. We define $G_x(z)$ as previously and obtain

$$G_x(z) = z^2 \sum_{\ell \geq 0} Z_\ell((\hat{\mathbf{P}}(\mathbf{z}) - \mathbf{2z}) = z^2 \exp\left(\sum_{\ell \geq 1} \frac{\hat{P}(z^\ell) - 2z^\ell}{\ell}\right), \tag{3.27}$$

Hence

$$\begin{aligned} G_x'(z) &= 2z(\exp\left(\sum_{\ell\geq 1} \frac{\hat{P}(z^\ell) - 2z^\ell}{\ell}\right) + z^2 \exp\left(\sum_{\ell\geq 1} \frac{\hat{P}(z^\ell) - 2z^\ell}{\ell}\right)\left(\sum_{\ell \geq 1} z^{\ell-1}(\hat{P}'(z^i) - 2)\right) \\ &= \frac{2}{z} G_x(z) + G_x(z)\left(\hat{P}'(z) - 2 + \sum_{\ell\geq 2} z^{\ell-1}(\hat{P}'(z^i) - 2)\right). \end{aligned}$$

At $z = \delta_n$, by (3.25) $G_x(z)$ equals

$$G_x(\delta_n) = \delta_n^2 \underbrace{\exp\left(\sum_{i\geq 1} \frac{\hat{P}(\delta_n^i)}{i}\right)}_{=2} \underbrace{\exp\left(\sum_{i\geq 1} \frac{-2\delta_n^i}{i}\right)}_{=(1-\delta_n)^2 \sim 1} \sim 2\delta_n^2.$$

Hence, due to $P(z) = 2\hat{P}(z) - 2nz$,

$$\begin{aligned} \lim_{z\to\delta_n} \frac{G_x'(z)}{P'(z)} &= \lim_{z\to\delta_n} \frac{G_x(z)\hat{P}'(z)}{P'(z)} \\ &= \lim_{z\to\delta_n} \frac{G_x(z)\hat{P}'(z)}{2\hat{P}'(z) - 2n} \sim \frac{2\delta_n^2}{2} = \frac{(2\ln 2 - 1)^2}{4n^2} \end{aligned}$$

and

$$\frac{G'(z)}{P'(z)} = n\frac{G_x'(z)}{P'(z)} = \frac{(2\ln 2 - 1)^2}{4n}.$$

☐

3.3 Limiting ratio of literals

In this section, we will compute the exact limiting ratios of functions of complexity $L(f) = 1$, that is literals x or $\bar{x}$. Therefore, in analogy to Section 3.2, we will define so called simple x-trees.

Definition 3.39. *A simple x is a tree of the shape $x \wedge ST$, $x \vee SC$, $x \wedge (x \vee \cdots)$ or $x \vee (x \wedge \cdots)$, where ST denotes a simple tautology and SC a simple contradiction[3]. The shape of such trees is depicted in Figures 3.5 and 3.6.*

[3] A simple contradiction is a negated simple tautology, i.e. a Boolean expression of the form $x \wedge \bar{x} \wedge f$ for any variable x and any function f.

For all cases, we will prove the following proposition:

Proposition 3.40. *Asymptotically, almost all trees computing the function x are simple x.*

We state the proposition without a complete proof. The proof is easily done by similar arguments as in the previous section, using the patterns $N[N]$ or $R[R]$, respectively. We can prove that every tree $t \in \mathcal{T}$ or $t \in \mathcal{C}$ with exactly 2 $N[N]$-restrictions, and every tree $t \in \mathcal{A}$ or $t \in \mathcal{P}$ with exactly 2 $R[R]$-restrictions, respectively, computing x is a simple x tree. Theorem 3.18 and its counterparts for the different models implies that those trees give asymptotically almost all trees computing x, as it is an easy task to prove that a large tree computing x will have at least 2 restrictions. Still we suggest a much simpler argument which proves the proposition in Section 3.5.

3.3.1 Binary planar trees

Theorem 3.41. *The limit ratio of functions of complexity* 1 *in the binary planar model is*

$$\lim_{m\to\infty} \mathbb{P}_{m,n}(x) = \frac{5}{16n^2} + O\left(\frac{1}{n^3}\right).$$

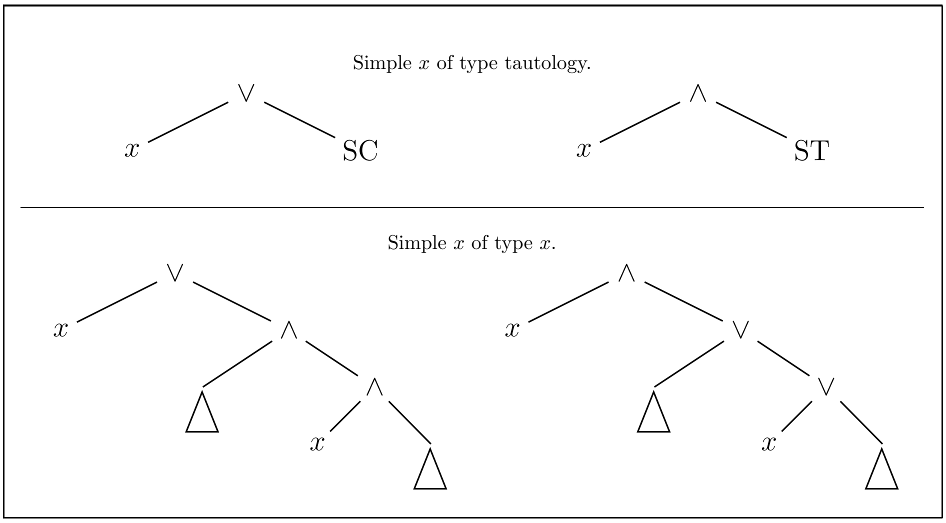

Figure 3.5: The different kinds of simple x - ST denotes a simple tautology and SC a simple contradiction.

Proof. We distinguish between simple x of type tautology, which we denote by x_T, and simple x of type x, denoted by x_X (c.f. Figure 3.5). By Proposition 3.40, we have $\mathbb{P}_n(x) = \mathbb{P}_n(x_T) + \mathbb{P}_n(x_X)$.

First we compute $\mathbb{P}_n(x_T) = \lim_{m\to\infty} \mathbb{P}_{m,n}(x_T)$. Let $G(z)$ be the generating function used for computing simple tautologies, given in Section 3.2.1. Of course, $G(z)$ also counts contradictions. The generating function $\tilde{G}(z)$ of simple x of the first kind is given by $4z \cdot G(z)$, where the factor z counts the leaf labelled with x, and the factor 4 is explained by the constant function being a tautology or a contradiction, the label of the internal node then being fixed, and the constant being positioned left or right. Hence

$$\frac{[z^m]\tilde{G}(z)}{[z^m]T(z)} \sim 4\rho_n \frac{[z^m]G(z)}{[z^m]T(z)} = 4\rho_n \mathbb{P}_n(True) \sim \frac{3}{16n^2}$$

For the computation of $\mathbb{P}_n(x_X)$ we use the function $g_x(z)$ given in (3.7). Let $\tilde{g}_x(z)$ be the function counting simple x of type x. Then $\tilde{g}_x(z) = 4zg_x(z)$ by the same arguments as above, hence

$$\frac{[z^m]\tilde{g}_x(z)}{[z^m]T(z)} \sim 4\rho_n \frac{[z^m]g_x(z)}{[z^m]T(z)} \sim \frac{4}{16n} \lim_{z\to\frac{1}{16n}} \frac{g_x'(z)}{T'(z)}$$

Using Maple, we get $\lim_{z\to\frac{1}{16n}} \frac{g'(z)}{T'(z)} = \frac{1}{2n} + \mathcal{O}\left(\frac{1}{n^2}\right)$, hence

$$\mathbb{P}_n(x) = \mathbb{P}_n(x_T) + \mathbb{P}_n(x_X) \sim \frac{3}{16n^2} + \frac{1}{8n^2} + \mathcal{O}\left(\frac{1}{n^3}\right) = \frac{5}{16n^2} + \mathcal{O}\left(\frac{1}{n^3}\right)$$

□

3.3.2 Associative planar trees

Theorem 3.42. *The limit ratio of functions of complexity* 1 *in the associative planar case is*

$$\lim_{m\to\infty} \mathbb{P}^a_{m,n}(x) = \frac{546 - 386\sqrt{2}}{n^2} + O\left(\frac{1}{n^3}\right) \approx \frac{0.1135651}{n^2} + \mathcal{O}\left(\frac{1}{n^3}\right).$$

Proof. Again we distinguish between simple x of type tautology (x_T), and simple x of type x (x_X), cf. Figure 3.6. Note that a simple x in the associative case is represented by a tree with a binary root.

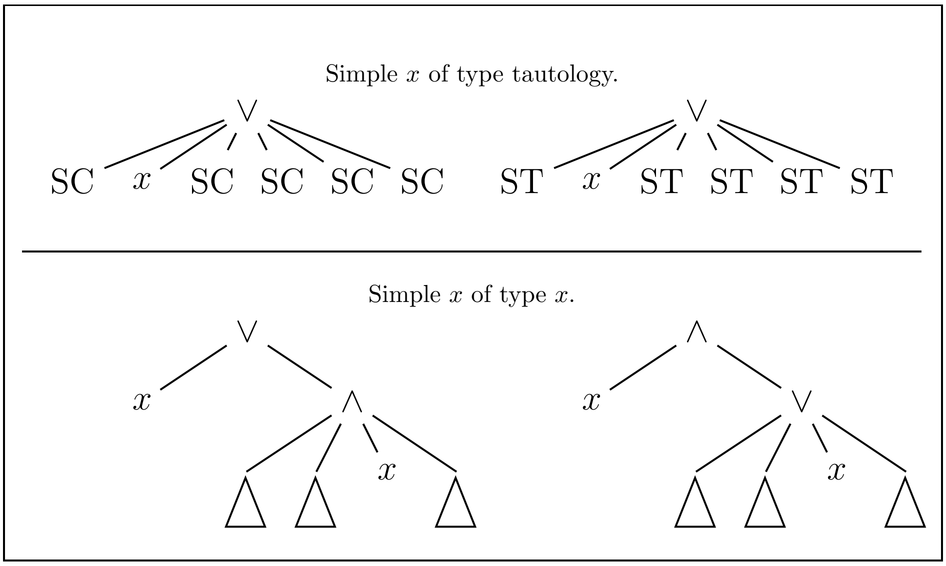

Figure 3.6: The different kinds of simple x in the associative case - ST denotes a simple tautology and SC a simple contradiction.

Calculating $\mathbb{P}_n(x_T) = \lim_{m\to\infty} \mathbb{P}_{m,n}(x_T)$, we obtain $\tilde{G}(z) = 4z \cdot G(z)$ by the same arguments as above and

$$\frac{[z^m]\tilde{G}(z)}{[z^m]A(z)} \sim 4\alpha_n \frac{[z^m]G(z)}{[z^m]A(z)} = 4\alpha_n \mathbb{P}^a_n(True) \sim 4\,\frac{3-2\sqrt{2}}{2n}\,\frac{51-36\sqrt{2}}{n} = \frac{594-420\sqrt{2}}{n^2}.$$

The contribution of x_X is counted by $\tilde{g}_x(z) = 4zg_x(z)$, where $g_x(z)$ counts trees with an $\vee$-root and exactly one leaf labelled by x. Note that the other leaves may not be labelled with x neither with $\bar{x}$, because this would give a simple tautology. Hence, $g_x(z)$ is given by

$$g_x(z) = z \sum_{\ell \geq 2} \ell (A(z) - 2z)^{\ell-1}.$$

Maple computations give $\lim_{z \to \frac{3-2\sqrt{2}}{2n}} \frac{g_x'(z)}{A'(z)} \sim \frac{3\sqrt{2}-4}{n} + \mathcal{O}\left(\frac{1}{n^2}\right)$, and thus

$$\frac{[z^m]\tilde{g}_x(z)}{[z^m]A(z)} \sim 4\alpha_n \frac{[z^m]g_x(z)}{[z^m]A(z)} \sim 4\,\frac{3-2\sqrt{2}}{2n}\,\frac{3\sqrt{2}-2}{n} = \frac{34\sqrt{2}-48}{n^2} + \mathcal{O}\left(\frac{1}{n^3}\right).$$

Adding the two limiting ratios gives the constant in Theorem 3.42. □

3.3.3 Binary non-planar trees

Theorem 3.43. *The limit ratio of functions of complexity* 1 *in the binary non-planar case is*

$$\lim_{m\to\infty} \mathbb{P}^c_{m,n}(x) = \frac{1153}{4096n^2} + O\left(\frac{1}{n^3}\right) \approx \frac{0.2814941406}{n^2} + O\left(\frac{1}{n^3}\right).$$

Proof. Simple x-trees are the same as in the planar binary case, but there is no left-to-right order anymore. Hence, $\tilde{G}(z) = 2\gamma_n G(z)$, and $G(z) = C(z) - \bar{G}(z)$ with $\bar{G}(z)$ given in (3.21). Hence

$$\frac{[z^m]\tilde{G}(z)}{[z^m]C(z)} \sim 2\gamma_n \frac{[z^m]G(z)}{[z^m]C(z)} = 2\gamma_n \mathbb{P}^c_n(True) \sim 2\,\frac{1}{8n}\left(1+\frac{1}{8n}\right)\frac{641}{1024n} = \frac{641}{4096n^2} + \mathcal{O}\left(\frac{1}{n^3}\right)$$

$\tilde{g}_x(z) = 2\gamma_n g_x(z)$, and $g(z) = C(z) - \bar{g}_x(z)$ with $\bar{g}(z)$ given in (3.21) and $\lim_{z\to\gamma_n} \frac{\bar{g}_x'(z)}{C'(z)}$ computed in the proof of Theorem 3.25. Hence

$$\frac{[z^m]\tilde{g}_x(z)}{[z^m]C(z)} \sim 2\gamma_n \frac{[z^m]g_x(z)}{[z^m]C(z)} \sim 2\,\frac{1}{8n}\left(1+\frac{1}{8n}\right)\frac{1}{2n} = \frac{1}{8n^2} + \mathcal{O}\left(\frac{1}{n^3}\right) = \frac{512}{4096n^2} + \mathcal{O}\left(\frac{1}{n^3}\right).$$

□

3.3.4 Associative non-planar trees

Theorem 3.44. *The limit ratio of functions of complexity* 1 *in the associative non-planar case is*

$$\lim_{m\to\infty} \mathbb{P}^{a,c}_{m,n}(x) = \frac{(2\ln 2-1)^2(2\ln 2+1)}{4n^2} + O\left(\frac{1}{n^3}\right) \approx \frac{0.08902269970}{n^2} + \mathcal{O}\left(\frac{1}{n^3}\right).$$

Proof. Again, $\tilde{G}(z) = 2\delta_n G(z)$, with $G(z)$ given in (3.27), and

$$\frac{[z^m]\tilde{G}(z)}{[z^m]P(z)} \sim 2\delta_n \frac{[z^m]G(z)}{[z^m]P(z)} = 2\delta_n \mathbb{P}^{a,c}_m(True)$$
$$\sim 2\frac{(2\ln 2-1)}{2n}\,\frac{(2\ln 2-1)^2}{8n} = \frac{(2\ln 2-1)^3}{8n^2} + \mathcal{O}\left(\frac{1}{n^3}\right).$$

Moreover $g_x(z)$ is given by

$$g_x(z) = z + z\left(\exp\left(\frac{\sum_{\ell\geq 1}\hat{P}(z^\ell) - 2z^\ell}{\ell}\right) - 1\right),$$

and

$$g_x'(z) = 1 + \frac{1}{z}(g_x(z) - z) + g_x(z)\left(\sum_{\ell\geq 1} z^{\ell-1}(\hat{P}(z^\ell) - 2)\right).$$

Since $g_x(z) \sim 2\delta_n$ as $z \to \delta_n$, we get:

$$\lim_{z\to\delta_n}\frac{g_x'(z)}{P'(z)} \sim \lim_{z\to\delta_n}\frac{g_x(z)\hat{P}'(z)}{2\hat{P}'(z) - 2n} \sim \frac{2\delta_n}{2} = \frac{2\ln 2 - 1}{2n},$$

and finally, with $\tilde{g}_x(z) = 2\delta_n g_x(z)$,

$$\frac{[z^m]\tilde{g}_x(z)}{[z^m]P(z)} \sim 2\delta_n\frac{[z^m]g_x(z)}{[z^m]P(z)} \sim 2\,\frac{(2\ln 2 - 1)}{2n}\,\frac{(2\ln 2 - 1)}{4n} = \frac{(2\ln 2 - 1)^2}{4n^2} + \mathcal{O}\left(\frac{1}{n^3}\right).$$

□

3.4 Limiting probability of a general function

In the previous parts, we have studied functions of complexity zero and one. In this part we are interested in the limiting probability of functions of higher complexity. To prove Theorem 3.5 Kozik showed that asymptotically almost all trees computing a function f have a "simple f" shape. To be more precise, they are obtained from a minimal tree by a single well-defined expansion. In this part, we generalize this result to all models and give bounds for the number of such expansions.

3.4.1 The binary planar case

Theorem 3.45 ([47]). *For all Boolean functions f,*

$$\mathbb{P}_n(f) \sim \frac{\lambda_f}{n^{L(f)+1}}$$

when n tends to infinity, where λ_f is the number of possible expansions of a minimal tree computing f.

The idea of this part is to bound λ_f. We show the following result:

Proposition 3.46. *For all Boolean function f,*

$$\frac{10r - 3}{2\cdot 16^r}M_f \leq \lambda_f \leq \frac{8r^2 + 2r - 3}{2\cdot 16^r}M_f$$

where M_f is the number of minimal trees representing f and $r = L(f)$.

The proof of this proposition is based on a result by Kozik [47]. The set of non negligible trees computing f, i.e. the trees determining the asymptotic leading term, is exactly the set of trees obtained by *expanding* a minimal tree of f once.

Definition 3.47. *Let t be an And/Or tree computing f, ν one of its nodes and t_ν the subtree rooted at ν. An* expansion *of t in ν is a tree obtained by replacing the subtree t_ν rooted at ν by a tree $t_\nu \diamond t_e$ where $\diamond \in \{\wedge, \vee\}$ and where t_e is an And/Or tree. Moreover, the expanded tree still has to compute f.*

Kozik has shown that the only non negligible expansions that are to be considered are:

- The T-expansions: an expansion is a T-expansion if the inserted subtree t_e is a simple tautology (resp. a simple contradiction) and if the new label of ν is $\wedge$ (resp. $\vee$).
- The X-expansions: an expansion is an X-expansion if the inserted subtree t_e is (up to commutativity and associativity) of the shape $x \vee ...$ (resp. $x \wedge ...$) where x is an essential variable of f, and if the new label of ν is $\wedge$ (resp. $\vee$).

In the following, we will call a T-expansion an $\wedge$-T-expansion (resp. an $\vee$-T-expansion) if the new label of ν is $\wedge$ (resp. $\vee$), and analogously for X-expansions.

Proof of Proposition 3.46. In a Catalan And/Or tree, a T-expansion is possible in every node without changing the computed function. At each node, either we can expand by an $\vee$-T-expansion on the right side and on the left side, or by an $\wedge$-T-expansion on the right side and on the left side. As a minimal tree of f is of size $L(f)$, it has $2L(f) - 1$ nodes and there are $\lambda_T(f) = 2(2L(f) - 1)M_f$ different kinds of T-expansions that can be done from all minimal trees computing f.

We can now consider $\lambda_X(f)$, the number of different kinds of X-expansions which do not change the computed function f. This number depends heavily on the shape of the minimal trees of f, therefore, we only find bounds for this number. An $\wedge$-X-expansion (resp. $\vee$-X-expansion) according to x_i is allowed at each node linked to a leaf labelled by x_i by an $\vee$-only (resp. $\wedge$-only) path, and at all its sons. Let us note that:

- for each leaf, we can do at least one $\vee$-X-expansion to the right and one to the left and one $\wedge$-X-expansion to the right and to the left at its father, because two different leaves having the same father can be labelled by the same variable; this gives contribution $4L(f)$;
- at each node (internal or external), we can do at most 4 X-expansions (we choose between $\wedge$ and $\vee$ and between right and left side) for every different literal that appears on the leaves. There are at most $L(f)$ different literals appearing on the leaves of a minimal tree and a minimal tree has exactly $2L(f) - 1$ (internal or external) nodes. Therefore, $4L(f)(2L(f) - 1)M_f$ is an upper bound of $\lambda_X(f)$.

Therefore, we have the following bounds:

$$4L(f)M_f \leq \lambda_X(f) \leq 4L(f)(2L(f) - 1)M_f \tag{3.28}$$

To end the proof of Proposition 3.46, we need to note that the expansions of different types are weighted by their limiting ratio, and hence

$$\frac{\lambda_f}{n^{L(f)+1}} = M_f \rho_n^{L(f)}(\lambda_T(f)w_1 + \lambda_X(f)w_2),$$

where w_1 is the limiting ratio of simple tautologies (resp. simple contradictions), and w_2 is the limiting ratio of trees of shape $x \vee ...$ for x a variable. Thanks to the computations made in Section 3.2 (cf. Theorem 3.13), we know that $w_1 = \frac{3}{4n}$. Moreover, the generating function g_x defined in Section 3.2.1 counts exactly the number of trees that can be used for an X-expansion (according to a variable x). Therefore,

$$\lim_{z \to \rho_n} \frac{g'_x(z)}{P(z)} \sim \frac{1}{2n} = w_2,$$

and with (3.28) we prove Proposition 3.46. □

3.4.2 The associative planar case

The associative case appears to be similar to the binary planar case. We show the following theorem.

Theorem 3.48. *In the associative planar case, the probability distribution $\mathbb{P}_n^a(f)$ is asymptotically given by*

$$\mathbb{P}_n^a(f) \sim \frac{\lambda_f^a}{n^{L(f)+1}},$$

as n tends to infinity, and

$$\left(\frac{3-2\sqrt{2}}{2}\right)^r \left[145r + 153 - (102r + 108)\sqrt{2}\right] M_f \leq \lambda_f$$

$$\lambda_f \leq \left(\frac{3-2\sqrt{2}}{2}\right)^r \left[-(12r^2 - 247r + 51) + (9r^2 - 174r + 36)\sqrt{2}\right] M_f$$

where M_f is the number of minimal trees computing f and $r = L(f)$ is the complexity of f.

To show Theorem 3.48, we first have to prove that, as in the binary planar case, the set of non negligible associative trees computing a Boolean function is the set of trees obtained from a minimal tree by expanding once. Moreover, we have to find the non-negligible expansions that have to be considered. Then, we can prove Theorem 3.48 with the same methods as in the binary planar case.

Associative expansions.

Because of the structure of associative trees (they are stratified), we have to be careful with the definition of expansions, which is different to the one in the binary case:

Definition 3.49 (cf. Figure 3.7 on page 109)**.** *Let t be an And/Or associative tree computing f. We define two types of expansions of t.*

- *Let ν be an internal node of t (possibly the root) with subtrees $t_1, \dots, t_j$, $j \geq 2$. An expansion of t in ν of the first kind is a tree obtained by adding a subtree $t_e = t_{j+1}$ to ν.*
- *Let ν be the root or a leaf of the tree. The tree obtained by replacing the subtree t_ν rooted at ν by $t_e \diamond t_\nu$, where $\diamond \in \{\wedge, \vee\}$ is chosen such that the obtained tree is stratified, is an expansion of t in ν of the second kind. In this case, $\diamond$ will be called the* new label *of ν.*

In all cases, the expanded tree still has to compute f.

Remark. We have to keep in mind that both types of expansions are possible at the root of the tree.

Proposition 3.50. *The set of non-negligible trees computing a Boolean function f is the set of trees obtained by expanding a minimal tree of f once. Moreover, the only non-negligible expansions we have to consider are:*

- *The T-expansions: an expansion is a T-expansion if the inserted subtree t_e is a simple tautology (resp. a simple contradiction) and if the new label of ν is $\wedge$ (resp. $\vee$).*
- *The X-expansions: an expansion is an X-expansion if the inserted subtree t_e is (up to commutativity) of the shape $x \vee ...$ (resp. $x \wedge ...$) where x is an essential variable of f and if the new label of ν is $\wedge$ (resp. $\vee$).*

Proof. The proof is inspired by the proof of the corresponding result for binary trees that can be found in Kozik's paper [47]. The idea is to take a well-chosen tree computing f, and to replace every subtree which can be evaluated to $True$ or $False$ independently from the rest of the tree by a $\star$. Then, we state that simplifying the stars gives a minimal tree of the considered Boolean function.

Let f be a Boolean function whose complexity $L(f)$ will be denoted by r. Let us consider the following patterns:

$$\hat{P} = \bullet | \check{P} \wedge \check{P} | \check{P} \wedge \check{P} \wedge \check{P} | \dots$$
$$\check{P} = \bullet | \hat{P} \vee \square | \hat{P} \vee \square \vee \square | \dots$$
$$S = \{\hat{P}, \check{P}\}$$

and R the pattern defined in Section 3.2.2 (see (3.12)).

Remark. The pattern P has the following property: if all the P-pattern leaves of a tree are valuated to $True$, then the whole tree itself computes $True$.

Now consider the patterns $L = R^{(r+1)}[R \oplus S]$ and $\bar{L} = R^{(r+1)}[(R \oplus S)^2]$. We need the following definition:

Definition 3.51. *A leaf which is a $R^{(i)}$-pattern leaf but not a $R^{(i-1)}$-pattern leaf is said to be on level i. A L-pattern leaf which is not an R^{r+1}-pattern leaf is said to be on level $r+2$ and a $\bar{L}$-pattern leaf which is not a L-pattern leaf is said to be on level $r+3$.*

Let t be a tree of size r representing f. If the root of t is labelled with $\vee$ (resp. $\wedge$), then using a simple contradiction (resp. tautology) Φ, the new tree $\Phi \wedge t$ (resp. $\Phi \vee t$) still represents the function f. Since the limiting ratio of simple tautologies or contradictions is equal to $\Theta(1/n)$ and the r nodes of t are counted by z^r, for large enough n we obtain the following lower bound:

$$\mathbb{P}_n^a(f) \geq \frac{\alpha}{n^{r+1}}.$$

Consequently, we can neglect trees with at least $r+2$ $\bar{L}$-restrictions.

By the same arguments as Kozik gives in [47] for binary planar trees, it is easily proven that a tree computing f has to have at least $r+1$ L-restrictions (if it has leaves on level $r+2$, which we can assume), because otherwise it could be simplified to a tree of size smaller than r computing f by the properties of the patterns R and S. Further, we consider trees with exactly $r+1$ L-restrictions and exactly $r+1$ $\bar{L}$-restrictions (those with more $\bar{L}$-restrictions are negligible). Finally, we know that the set of non negligible trees computing f is the set of trees with exactly $r+1$ L-restrictions and $r+1$ $\bar{L}$-restrictions. Thus, we know that every variable appearing in a level $r+3$ pattern leaf is non essential and not repeated among $\bar{L}$ pattern leaves. Therefore, each subtree of t rooted on level $r+3$ with its parent node on level $r+2$ can be replaced by a $\star$, because it can be valuated to $False$ by assigning all the R-pattern leaves to $False$ and to $True$ by assigning all the S-pattern leaves to $True$. Both valuations can be done independently from the rest of the tree because these pattern leaves are non essential and not repeated. After this operation all the remaining leaves are L-pattern leaves.

Moreover, we replace by $\star$ every leaf of the tree which is not an essential variable of f and which appears only once among the leaves of t. We now have obtained a tree $t^\star$.

Then, we can simplify the tree in order to obtain a tree without stars, according to the following rules:

$$\begin{array}{cc} \star \vee \ldots \vee \star \equiv \star & \star \wedge \ldots \wedge \star \equiv \star \\ \star \vee \ldots \star \vee t_1 \vee \ldots \vee t_j \equiv True & \star \wedge \ldots \star \wedge t_1 \wedge \ldots \wedge t_j \equiv False \end{array} \quad (3.29)$$

$$\begin{array}{cc} True \vee t_1 \vee \ldots \vee t_j \equiv True & False \wedge t_1 \wedge \ldots \wedge t_j \equiv False \\ False \vee t_1 \vee \ldots \vee t_j \equiv t_1 \vee \ldots \vee t_j & True \wedge t_1 \wedge \ldots \wedge t_j \equiv t_1 \wedge \ldots \wedge t_j \end{array} \quad (3.30)$$

where $t_1, \ldots, t_j$ are subtrees containing no stars.

The tree $\hat{t}$ obtained after this process still computes f. Let us prove that the obtained tree is a minimal tree of f and that the least common ancestor[4] of the stars in $t^\star$ has been simplified during the process.

The tree $t^\star$ contains at least one $\star$ since t is big enough to have at least one leaf on level $r+3$. Moreover, the tree $t^\star$ has exactly $r+1$ restrictions and no constants. The final tree $\hat{t}$ has no $\star$, and no constant. Therefore, a rule of type (3.29) must have been used at least once during the simplification process, because the rules (3.29) are the only ones simplifying stars. But, using such a rule simplifies at least one subtree with at least one non-star pattern leaf. Therefore, this leaf had to be labelled by a variable which was either essential or repeated. Therefore, the simplifying process has at least simplified one restriction and the obtained tree has at most r leaves. Thus, $\hat{t}$ is a minimal tree of f.

Moreover, let ν be the least common ancestor of the stars in $t^\star$. Let us assume that ν does not disappear during the simplification process. Therefore, two stars have been simplified independently during the process, and thus at least two rules of type (3.29) have been applied, which means that at least two restrictions have disappeared during the simplification process. Therefore, the simplified tree $\hat{t}$ still computes f and contains at most $r-1$ restrictions, i.e. at most $r-1$ leaves, which is impossible since the complexity of f is r. Therefore, the node ν has to disappear during the simplification process. Thus a non-negligible tree computing f is indeed a minimal tree expanded once.

Finally, let us remark that the last simplifying rule applied during the process has to be of the kind (3.30).

The second part of the proof is to understand which are the non-negligible expansions allowed, i.e. which do not change the computed function f.

[4] the least common ancestor of a set of nodes is the node ν farthest from the root such that the tree rooted at ν contains all the nodes of the considered set.

First, let us remark that, thanks to Theorem 3.18, the trees obtained by expanding with a tree t_e with more than two $(R \oplus S)^2$-restrictions are negligible. On the other hand there has to be at least one $(R \oplus S)^2$-restriction in t_e, because if there was none, we could assign this tree to $False$ or $True$ independently from the rest of the tree. Since the expanded tree must still compute the function f, by simplification we would obtain a tree computing f being smaller than the minimal tree, which is impossible.

First case: The tree t_e contains one repetition and no essential variable. Then it has to compute a constant function (i.e. $True$ or $False$). If it does not, the subtree can be valuated to $True$ or $False$ independently from the rest of the tree. Thus, by simplification, we can obtain a tree, smaller than the minimal tree, computing f, which is a contradiction. Therefore, the expanding tree t_e is a simple tautology or a simple contradiction (thanks to Proposition 3.23). Moreover, as the expanded tree still has to compute f, if the father of t_e is a $\wedge$ (resp. $\vee$), t_e is a simple tautology (resp. contradiction), which gives a T-expansion.

Second case: The subtree t_e contains no repetition and one essential variable, let us say x. Then the essential variable has to appear on the first level. If it does not, the Boolean expression has shape $s_1 \wedge (s_2 \vee x)$ or $s_1 \vee (s_2 \wedge x)$ (up to commutativity). Moreover, the terms s_1 and s_2 have no $R \oplus S$-restrictions and therefore we can make them $False$ or $True$ independently from the rest of the tree. Then we can valuate the whole term either to $False$ or $True$ independently from x, which is impossible since x is an essential variable of f.

If a $\wedge$-X-expansion t_e according to the variable x_i is allowed in a node ν, then every $\wedge$-X-expansion t'_e according to this variable x is allowed at ν (and as well for $\vee$-X-expansions). ☐

Proof of Proposition 3.48.

Proof. As in the binary case, we have to compute the limiting ratio of T-expansions and X-expansions, then the number of nodes where each kind of such expansions are allowed. Let us denote by M_f the number of minimal trees representing a given Boolean function f of complexity r.

The limiting ratio of T-expansions is the limit ratio of simple tautologies, which has already been computed in section 3.2.2. We have that $w_1^a = \frac{51-36\sqrt{2}}{n}$.

Let g_x be the generating function of associative trees rooted at $\wedge$ (resp. $\vee$) and containing exactly one x in the first generation. Then,

$$g_x(z) = z \sum_{j \geq 2} j(A(z) - 2z)^{j-1}.$$

Since the set of trees with more than one x in the first generation is negligible in front of the set of trees with exactly one x in the first generation, we can assume that:

$$w_2^a = \lim_{z \to \alpha_n} \frac{g'_x(z)}{A'(z)} = \frac{3\sqrt{2} - 4}{n}$$

is the limiting ratio of $\wedge$-X-expansions (resp. $\vee$-X-expansions).

As in the binary case, the number $\lambda_X(f)$ of different kinds of X-expansions and the number $\lambda_T(f)$ of different T-expansions allowed in a minimal tree depend on the shape of the considered minimal tree. Given a minimal tree t of f, let us number its internal nodes from 1 to N. Let us denote by $s(i)$ the number of sons of the internal node i. Moreover, let us denote by $d(i)$ the number of sons of the node i which are leaves. Then, if $\lambda_T(t)$ is the number of different T-expansions in the minimal tree t of f, we have that:

$$\lambda_T(t) = 2r + \sum_{i=1}^{N}(s(i)+1) + 2$$

where $2r$ is the number of different T-expansions allowed at the leaves of the tree (if the parent node is labelled by $\wedge$ (or $\vee$ respectively), only simple tautology (or contradiction respectively) T-expansions are allowed), $s(i)+1$ is the number of different T-expansions allowed at the node i (the number of different positions at node i is $s(i)+1$); and 2 is the number of expansions allowed at the root by pushing the root to the first generation and adding a new root with two sons. Therefore,

$$\lambda_T(t) = 2r + \sum_{i=1}^{N} s(i) + N + 2 = 2r + (r + N - 1) + N + 2,$$

and since $1 \leq N \leq r-1$, we obtain that:

$$3(r-1)M_f \leq \lambda_T(f) \leq (5r-1)M_f.$$

Further, given a label x_i, an $\wedge$-X-expansion according to x_i is allowed at its father and at all its sisters (brothers that are reduced to a leaf), because two sisters cannot have the same label. Indeed, if two sisters have the same label (or even opposite labels), then the considered tree can be simplified, and since we consider a minimal tree, this is impossible. Therefore,

$$\lambda_X(t) = \sum_{i=1}^{N} d(i)(s(i)+1) + 2d(root) + \sum_{i=1}^{N} d(i)^2.$$

Lemma 3.52. *For all i, $d(i) \leq r - N + 1$.*

Proof. Let us assume that there exist an internal node i_0 such that $d(i_0) > r - N + 1$. It is easy to see that, as each node except the root has a unique father, $\sum_{i=1}^{N} d(i) = r + N - 1$. Moreover,

$$\sum_{i=1}^{N} d(i) > \sum_{i \neq i_0} d(i) + (r - N + 1) > 2(N-1) + (r - N + 1)$$

since every internal node has at least two sons. Therefore, $\sum_{i=1}^{N} d(i) > r + N - 1$, which gives a contradiction. □

Therefore, thanks to the Lemma,

$$\begin{aligned}\lambda_X(t) &\leq \sum_{i=1}^{N} d(i) + (r-N+1)\sum_{i=1}^{N} s(i) + 2(r-N+1) + 2(r-N+1)\sum_{i=1}^{N} d(i)\\ &\leq r + (r-N+1)[(N+r-1) + 2 + 2r]\\ &\leq r(3r+2).\end{aligned}$$

On the other hand,

$$\lambda_X(t) \geq \sum_{i=1}^{N} d(i) + \sum_{i=1}^{N} d(i) = 2r,$$

and

$$\lambda_X(f) \geq 2rM_f.$$

Finally, since

$$\frac{\lambda_f}{n^{L(f)+1}} = M_f \alpha_n^{L(f)} (\lambda_T(f) w_1^a + \lambda_X(f) w_2^a),$$

we get that:

$$\left(\frac{3-2\sqrt{2}}{2}\right)^r \left[145r + 153 - (102r + 108)\sqrt{2}\right] M_f \leq \lambda_f$$

$$\lambda_f \leq \left(\frac{3-2\sqrt{2}}{2}\right)^r \left[-(12r^2 - 247r + 51) + (9r^2 - 174r + 36)\sqrt{2}\right] M_f$$

□

3.4.3 The binary non-planar case

Theorem 3.53. *Let f be a Boolean non-constant function of complexity $L(f) = r$. Then the limiting ratio of f is given by*

$$\lim_{m\to\infty} \mathbb{P}^c_{m,n}(f) \overset{n\to\infty}{\sim} \frac{\lambda^c_f}{2n^{r+1}},$$

where $\lambda^c(f)$ denotes the total number of expansions on all minimal trees of f. Moreover,

$$\frac{2306r - 641}{1024 \cdot 8^r} M_f \leq \lambda^c_f \leq \frac{(2r-1)(1024r + 641)}{1024 \cdot 8^r} M_f$$

where M_f is the number of minimal trees of f.

Proof. The proof relies completely on the binary planar case, doing minimal $[N]$-embeddings and $[N \oplus P]$-embeddings (the planar parts of an $[N \oplus P]$-embeddings are both the planar parts of an $[N]$-embedding or a $[P]$-embedding). It has been proven in Lemmas 3.33 and 3.34 that Proposition 3.28 can be applied to the pattern $[N \oplus P][\mathcal{C}]$, as the generating function of P is the same as the one of N. As in the proof of Theorem 3.25, embedding a tree $t \in \mathcal{C}$ into $N^{(r)}[N \oplus P]$ or $N^{(r)}[N \oplus P]^{(2)}$ represents an injection. Hence asymptotically almost all trees computing a function f are obtained by a single expansion of a minimal tree of f.

The calculation of the bounds can be done in the same way as in the planar binary case. We denote by w_1^c the limiting ratio of simple tautologies and by w_2^c the limiting ratio of X-expansions. Thanks to Section 3.2.3, we know that $w_1^c = \frac{641}{1024n}$. And thanks to Section 3.3, we know that $w_2^c = \frac{1}{2n}$. Moreover, since asymptotically almost all trees computing f are obtained by a single expansion of a minimal tree, we have $\frac{\lambda^c_f}{n^{L(f)+1}} = \gamma_n^{L(f)}(\lambda_T w_1^c + \lambda_X w_2^c)$ and

$$\lambda_T = (2L(f) - 1)M_f$$
$$2L(f)M_f \leq \lambda_X \leq 2L(f)(L(f) - 1)M_f,$$

since $\gamma_n \sim \frac{1}{8n}$ when $n \to \infty$, we obtain the desired result. □

3.4.4 The associative non-planar case

Theorem 3.54. *Let f be a Boolean non-constant function of complexity $L(f) = r$. Then the limiting ratio of f is given by*

$$\lim_{m\to\infty} \mathbb{P}^{a,c}_{m,n}(f) \overset{n\to\infty}{\sim} \frac{\lambda_f^{a,c}}{2n^{r+1}} M_f,$$

where $\lambda_f^{a,c}$ denotes the total number of expansions on all minimal trees of f. Moreover,

$$\left(\frac{2\ln 2 - 1}{2}\right)^r \left(\left(\ln^2 2 - \frac{1}{4}\right) r + \ln^2 2 - 2\ln 2 + \frac{1}{2}\right) M_f \leq \lambda_f^{a,c}$$

$$\lambda_f^{a,c} \leq \left(\frac{2\ln 2 - 1}{2}\right)^r \frac{(2\ln 2 - 1)(r + 1 + 4\ln 2)r}{4},$$

where M_f is the number of minimal trees of f.

Proof. The result is easily proven by using the pattern $R^{(r)}[R \oplus S]$ and applying arguments of Section 3.4.2 and Section 3.4.3. Therefore, we have $\frac{\lambda_f^{a,c}}{n^{L(f)+1}} = \delta_n^{L(f)}(\lambda_T w_1^{a,c} + \lambda_X w_2^{a,c})$. The calculation of the bounds is similar to the calculation done in the planar associative case. Thanks to Section 3.2.4, $w_1^{a,c} = \frac{(2\ln 2-1)^2}{4n}$ and thanks to Section 3.3, $w_2^{a,c} = \frac{2\ln 2-1}{4n}$. Moreover, $\delta_n \sim \frac{2\ln 2-1}{2n}$. We can further show that

$$(r+2)M_f \leq \lambda_T \leq 2rM_f$$
$$2rM_f \leq \lambda_X \leq (r^2 + 3r)M_f$$

and we can conclude the proof. □

3.5 Conclusion

Finally we have understood better the influence of associativity and commutativity on the behaviour of the probability distribution on Boolean functions induced by their tree representations. Indeed, we know that associativity and commutativity do not change the order of $\mathbb{P}_n(f)$ when n tends to infinity. It is still of order $\Theta\left(n^{-(L(f)+1)}\right)$. But, thanks to Sections 3.2 and 3.3, we know that associativity and commutativity change the exact limiting ratio of tautologies and literals. We summarize the different constants computed in Table 3.1. Note that the numerical results obtained for tautologies and literals suggests that

$$\mathbb{P}_n(f) > \mathbb{P}^c_n(f) > \mathbb{P}^a_n(f) > \mathbb{P}^{a,c}_n(f),$$

which is not possible as probabilities have to sum up to 1. Hence for (some) functions of higher complexity the above inequality has to flip.

Section 3.4 gives bounds for the constants for a general f and gives an interesting result about the shape of an average tree computing a function f: it is a minimal tree expanded once. Still it is not obvious from the given bounds that functions of higher complexity will have a bigger limiting distribution in the new models, as further knowledge on the number of minimal trees in the different models would be required.

Finally note that the simple x trees we defined in Section 3.3 are exactly those trees obtained by expanding once a tree consisting of a single leaf x. Hence the proof of Proposition 3.40 is immediate.

	Catalan trees	Commutative binary trees	Associative planar trees	General trees
$True$	$\frac{3}{4} \approx 0.75$	$\frac{641}{1024} \approx 0.626$	$51 - 36\sqrt{2} \approx 0.088$	$\frac{(2\ln 2 - 1)^2}{8} \approx 0.019$
x	$\frac{5}{16} \approx 0.312$	$\frac{1153}{4096} \approx 0.281$	$546 - 396\sqrt{2} \approx 0.114$	$\frac{(2\ln 2 - 1)^2(2\ln 2 + 1)}{4} \approx 0.089$

Table 3.1: The different constants λ such that $\mathbb{P}(True) \sim \frac{\lambda}{n}$ and $\mathbb{P}(x) \sim \frac{\lambda}{n^2}$ when n tends to infinity, depending on the studied model of trees.

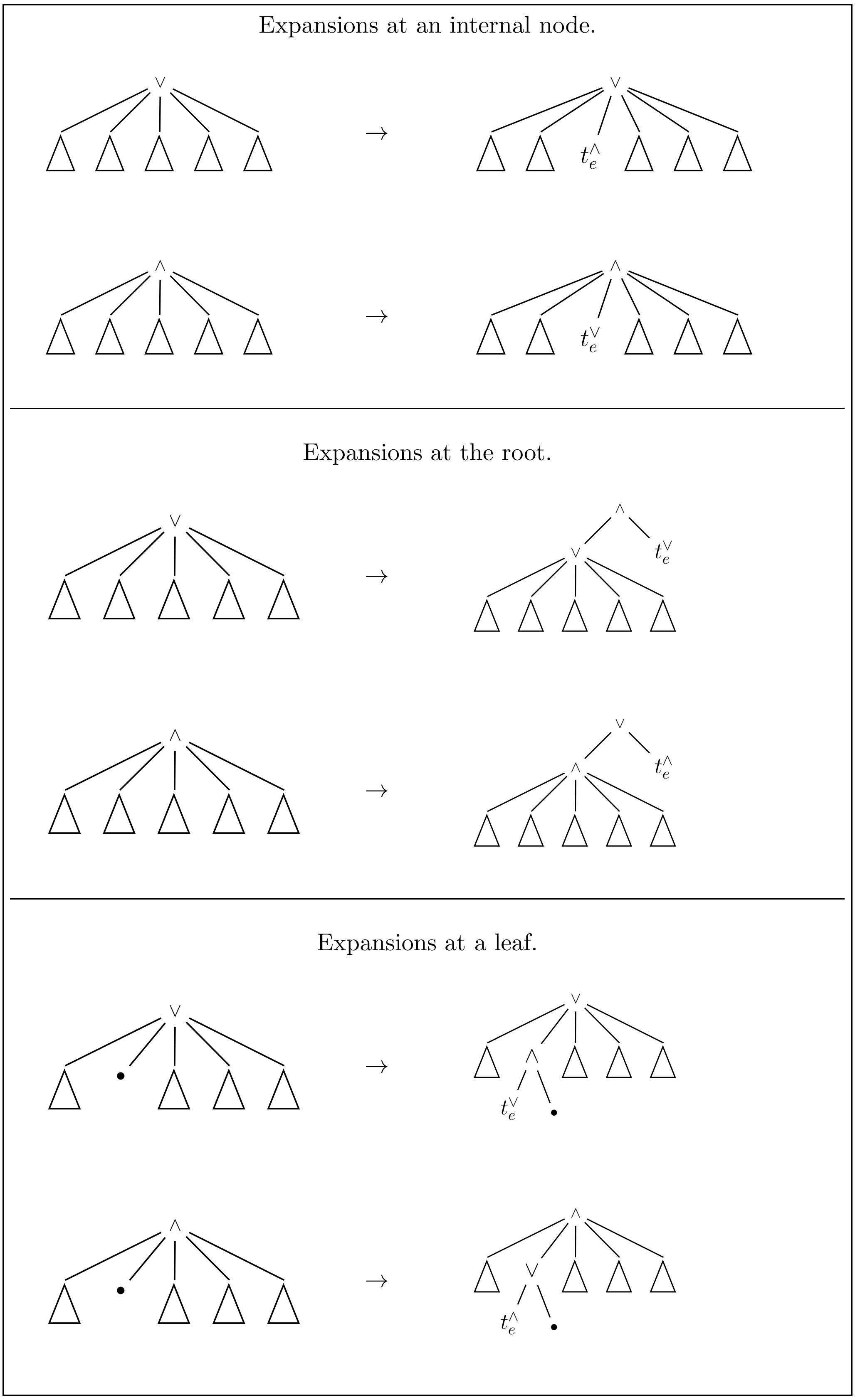

Figure 3.7: possible expansions at a node ν in the associative case. Here, $t_e^{\vee}$ (or $t_e^{\wedge}$, respectively) represents an associative tree rooted by $\vee$ (or $\wedge$, respectively).

CHAPTER 4

Random Graphs

In this chapter we are dealing with certain families of random planar graphs, which contain more symmetries than trees and are thus a bit more difficult to handle. A graph G is called planar if it can be drawn in the plane without crossings edges. Note that we will only study simple graphs, that is, graphs which do not contain multiple edges or loops. An outerplanar graph is a planar graph which can be drawn on the plane such that all nodes lie on the outer face, and a series-parallel graph is a graph obtained from series- and parallel extensions of the edges of a forest.

The study of random planar graphs is a quite active field of research. In the labelled setting, many problems have been solved recently. Bodirsky *et al* [7] asymptotically enumerated series-parallel graphs, an important class of graphs which will be described in detail later in this chapter. Gimenez and Noy solved the problem of enumeration of labelled planar graphs in 2009 [30], to be extended in the same year to count graphs embedded in general surfaces [31]. The degree distributions in labelled outerplanar and series-parallel graphs and later in general planar graphs have been studied in [18] and [17]. In the unlabelled setting, less is known. Outerplanar graphs have been enumerated by Bodirsky *et al* in [5], while the degree distribution of outerplanar graphs as well as enumeration and degree distribution of series-parallel graphs are presented in the following. We will present a general method for enumeration and asymptotic study of families with certain properties. Throughout the presentation, we will also reprove some of the results in the labelled setting as the method applies and understanding of the unlabelled results will be easier. The whole survey is based on a joint paper of Drmota, Fusy, Kang, Rue and myself, [16] and the results of [49].

In the following, we will denote by $\mathcal{G}$ certain families of graphs, and by $G(z)$ its exponential or ordinary generating function. We will not use different notations for the labelled and the unlabelled setting, as it will be clear from the context in which setting we are currently working. Further we will denote by $\mathcal{C}$ the subset of $\mathcal{G}$ containing only connected graphs from the family $\mathcal{G}$ and by $C(z)$ the generating function of this subset.

Let $A(z)$ be some exponential generating function of a labelled class $\mathcal{A}$, $A'(z) = \frac{\partial}{\partial z}A(z)$ be

the generating function of the derived class $\mathcal{A}$ and $A^\bullet(z) = zA'(z)$ the generating function of the rooted class. Recall that the derived class contains all structures from $\mathcal{A}$ where one vertex is distinguished and not labelled, we call this the pointed vertex, while the rooted class $\mathcal{A}^\bullet$ contains those structures from $\mathcal{A}$ which have a pointed vertex which is part of the vertex set.
Analogously, let $A(z)$ be the ordinary generating function and $Z_{\mathcal{A}}(\mathbf{s_1})$ the corresponding cycle index sum of an unlabelled class $\mathcal{A}$, $Z_{\mathcal{A}'}(\mathbf{s_1}) = \frac{\partial}{\partial s_1} Z_{\mathcal{A}}(\mathbf{s_1})$ the cycle index sum of the pointed and $Z_{\mathcal{A}^\bullet}(\mathbf{s_1}) = s_1 Z_{\mathcal{A}'}(\mathbf{s_1})$ the cycle index sum of the rooted class. By substitution we obtain the ordinary generating functions $A'(z)$ and $A^\bullet(z)$.
Recall that in the labelled setting

$$A'(z) = \frac{\mathrm{d}}{\mathrm{d}z} A(z), \qquad A^\bullet(z) = zA'(z),$$

and in the unlabelled setting

$$Z_{\mathcal{A}'}(\mathbf{s_1}) = \frac{\partial}{\partial s_1} Z_{\mathcal{A}}(\mathbf{s_1}), \qquad Z_{\mathcal{A}^\bullet}(\mathbf{s_1}) = s_1 Z_{\mathcal{A}'}(\mathbf{s_1}).$$

4.1 Subcritical graph families

Definition 4.1. *Let $\mathcal{G}$ be a given class of graphs. A graph $G \in \mathcal{G}$ is called k-connected, if we need to delete at least k vertices together with their incident edges to disconnect the graph.*

We can decompose a graph $G \in \mathcal{G}$ into its connected components, such that any graph $G \in \mathcal{G}$ is the set of its connected components. Further we can decompose a connected graph $C \in \mathcal{G}$ into its 2-connected components, which we also call *blocks*. Different blocks of G possibly share common vertices, these are then *cut-vertices* of G.

Definition 4.2. *A vertex $\nu \in V(G)$ from the set of vertices of a graph G is called a cut-vertex, if by removing ν together with its incident edges we disconnect G.*
We say that a vertex ν of a graph $G \in \mathcal{G}$ is incident to a block B of G if ν belongs to B.

Certain families of graphs are determined completely by the set of 2-connected components building them:

Definition 4.3. *We call a family of graphs* block-stable *if a graph belongs to the family if and only if all of its 2-connected components belong to the family.*

The block structure of a connected graph yields a bipartite tree, consisting of two types of nodes, namely blocks and vertices, cf Figure 4.1.

If a family of graphs $\mathcal{G}$ is block-stable, the family is completely determined by the set of 2-connected graphs in $\mathcal{G}$, which we denote by $\mathcal{B}$. A rooted connected graph $C \in \mathcal{G}$ is then a set of blocks from $\mathcal{B}$, where every vertex is replaced by a rooted connected graph from $\mathcal{C}^\bullet$ (which can of course consist of the root only). To get a symbolic equation for this relation, we need to root the graph. We obtain the system

$$\begin{aligned} \mathcal{G} &= \mathrm{Set}(\mathcal{C}) \\ \mathcal{C}^\bullet &= \mathcal{X} \times \mathrm{Set}(\mathcal{B}' \circ \mathcal{C}^\bullet)). \end{aligned} \tag{4.1}$$

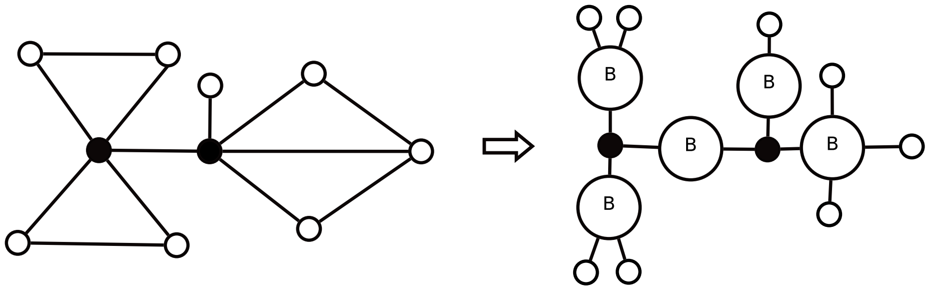

Figure 4.1: An example for the decomposition of an outerplanar graph - the black and white dots denote vertices while the circles labelled with B denote 2-connected components

In block-stable families, there are two phenomena that might appear as the graphs grow big. In some families, which we will later define exactly, in a large random graph of size n almost all blocks will asymptotically be of small, comparable size. In other graph families, on the other hand, asymptotically almost surely there is one giant block, which contains a linear number of nodes, say αn, with $\alpha > 0$, while all other blocks are comparably very small. The families with no giant component phenomena are those families which are subcritical according to the following definitions.

For the results presented in the following section, see also [3]. Let $G(z)$ be the (exponential or ordinary) generating function of a given block-stable graph family, $C(z)$ the (exponential or ordinary) generating function of the family of connected components of $\mathcal{C} \subset \mathcal{G}$, $B(z)$ the GF of blocks $\mathcal{B} \subset \mathcal{C}$, and $C^\bullet(z)$ and $B'(z)$ the generating functions of the according rooted or derived classes.

4.1.1 Subcriticality in the labelled setting

In the labelled setting, (4.1) translates to

$$G(z) = \exp(C(z)) \tag{4.2}$$
$$C^\bullet(z) = z \exp\left(B'(C^\bullet(z))\right). \tag{4.3}$$

Definition 4.4. *Let $\mathcal{G}$ be a block-stable family of graphs and let ρ_B and ρ_C be the radii of convergence of $B'(z)$ and $C^\bullet(z)$. The family $\mathcal{G}$ is called* subcritical*, if $C^\bullet(\rho_C) < \rho_B$.*

Subcritical families have a unified singular behaviour, which gives their asymptotic coefficients.

Lemma 4.5. *Let $\mathcal{G}$ be a labelled subcritical block-stable graph class with $\mathcal{C}$ the connected subclass and $\mathcal{B}$ the 2-connected subclass. Then $C^\bullet(z)$ has a square-root singular expansion around its radius of convergence ρ_C. Furthermore if $[z^n]C^\bullet(z) > 0$ for $n \geq n_0$ then ρ_C is the only singularity on the circle $|z| = \rho_C$ and $C^\bullet(z)$ can be continued analytically to the region $D = \{z \in \mathbb{C} \big| |z| < \rho_C + \varepsilon, 1 - \frac{z}{\rho_C} \notin \mathbb{R}^-\}$ for some $\varepsilon > 0$.*

Proof. The function $y = C^\bullet(z)$ is a solution of

$$y = F(y; z), \text{ with } F(y; z) = z \exp(B'(y)).$$

Let ρ_C and ρ_B be the radii of convergence of $C^\bullet(z)$ and of $B'(y)$, respectively, and let $\tau := C^\bullet(\rho_C)$. Since $\mathcal{G}$ is subcritical, we have $\tau < \rho_B$, hence $F(y;z)$ is analytic at (τ, ρ_C). We conclude from Theorem A.3 that $C^\bullet(z)$ has a square-root expansion at ρ_C and can be continued analytically to D. ☐

Theorem 4.6. *Let $\mathcal{G}$ be a subcritical block-stable graph class with $[z^n]C^\bullet(z) > 0$ for $n \geq n_0$. Then there exist constants $\gamma \geq e \approx 2.71828$ and $c > 0$ such that*

$$[z^n]G(z) = c\, n^{-5/2}\, \gamma^n (1 + o(1)) \quad \text{as } n \to \infty. \tag{4.4}$$

Proof. The function $C(z)$ satisfies

$$C(z) = \int_0^z C^\bullet(t)\frac{\mathrm{d}t}{t},$$

hence $C(z)$ has a singular expansion of order $3/2$ at ρ_C, by Lemma 1.9. Since $G(z) = \exp(C(z))$ and exp is analytic everywhere (in particular at $C(\rho_C)$), we conclude that $G(z)$ also has a singular expansion of order $3/2$ at ρ_C. The transfer lemma (Lemma 1.7) yield an estimate of the form (4.4), where $\gamma = 1/\rho_C$. ☐

Lemma 4.7. *A block-stable family of graphs $\mathcal{G}$ is subcritical if $B'(z)$ has a square-root singular expansion at ρ_B.*

Proof. If $B'(z)$ has a square root singular expansion, then $\lim_{z\to\rho_B-} B''(z) = \infty$. To proof subcriticality, we need to prove that $\tau := C^\bullet(\rho_C) < \rho_B$.
Assume that $\tau > \rho_B$. Then, by continuity of $C^\bullet(z)$, there exists $0 < z_0 < \rho_C$ such that $C^\bullet(z_0) = \rho_B$. Obviously, $C^\bullet(z)$ is regular at z_0 while $B'(C^\bullet(z))$ is singular at z_0. Hence, also $z\exp(B'(C^\bullet(z)))$ is singular at z_0, which is a contradiction, as $C^\bullet(z) = z\exp(B'(C^\bullet(z)))$. Hence $\tau \leq \rho_C$.
Assume now that $\tau = \rho_C$. Derivating Equation (4.3) gives

$$\frac{\mathrm{d}}{\mathrm{d}z}C^\bullet(z) = \frac{C^\bullet(z)}{z} + C^\bullet(z)B''(C^\bullet(z))\frac{\mathrm{d}}{\mathrm{d}z}C^\bullet(z).$$

As $C^\bullet(z)$ starts with a z, $\frac{C^\bullet(z)}{z} > 0$ for $z \in (0, \rho_C)$. Hence,

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}z}C^\bullet(z) &\geq C^\bullet(z)B''(C^\bullet(z))\frac{\mathrm{d}}{\mathrm{d}z}C^\bullet(z),\\ 1 &\geq C^\bullet(z)B''(C^\bullet(z)).\end{aligned}$$

This gives $B''(C^\bullet(z)) \leq \frac{1}{C^\bullet(z)}$ for $C^\bullet(z) \in (0, \rho_B)$, which contradicts the fact that $B''(\rho_B)$ tends to infinity. Hence, $\tau < \rho_B$. ☐

4.1.2 Subcriticality in the unlabelled setting

In the unlabelled setting, (4.1) translates to

$$\begin{aligned} Z_{\mathcal{G}}(\mathbf{s_1}) &= \exp\left(\sum_{i\geq 1}\frac{1}{i}Z_{\mathcal{C}}(\mathbf{s_i})\right)\\ Z_{\mathcal{C}^\bullet}(\mathbf{s_1}) &= s_1\exp\left(\sum_{i\geq 1}\frac{1}{i}Z_{\mathcal{B}'}(Z_{\mathcal{C}^\bullet}(\mathbf{s_i}), Z_{\mathcal{C}^\bullet}(\mathbf{s_{2i}}), \ldots)\right). \end{aligned} \tag{4.5}$$

and thus

$$G(z) = \exp\left(\sum_{i\geq 1}\frac{1}{i}C(z^i)\right), \tag{4.6}$$

$$C^\bullet(z) = z\exp\left(\sum_{i\geq 1}\frac{1}{i}Z_{\mathcal{B}'}(C^\bullet(z^i), C^\bullet(z^{2i}), C^\bullet(z^{3i})\dots)\right). \tag{4.7}$$

Denote by

$$\begin{aligned} g(y,z) &:= Z_{\mathcal{B}'}(y, C^\bullet(z^2), C^\bullet(z^3), \dots) \text{ and} \\ \Sigma(z) &:= \sum_{i\geq 2}\frac{1}{i}Z_{\mathcal{B}'}(C^\bullet(z^i), C^\bullet(z^{2i}), C^\bullet(z^{3i})\dots). \end{aligned}$$

Then equation (4.7) translates to

$$C^\bullet(z) = z\exp\left(g(C^\bullet(z), z) + \Sigma(z)\right)$$

Definition 4.8. *Let $\mathcal{G}$ be a block-stable family of graphs and let ρ_C be the radius of convergence of $C^\bullet(z)$. The family $\mathcal{G}$ is called* subcritical*, if*

(i) ρ_C is non-zero,

(ii) $g(y,z)$ is analytic at $(C^\bullet(\rho_C), \rho_C)$,

(iii) the radius of convergence of $\Sigma(z)$ is larger than ρ_C and

(iv) the radius of convergence of the series $Z_{\mathcal{C}}(0, z^2, z^3, \dots)$ is larger than ρ_C.

Note that for any block stable class $\mathcal{G}$, the class $\mathcal{C}^\bullet$ of rooted connected graphs from $\mathcal{G}$ dominates coefficient-wise the class of Pólya trees by Equation (4.7), thus $\rho_C \leq \rho = 0.33832$, where ρ is the singularity of Pólya trees given in Chapter 2.

Lemma 4.9. *Let $\mathcal{G}$ be an unlabelled subcritical block-stable graph class with $\mathcal{C}$ the connected subclass and $\mathcal{B}$ the 2-connected subclass. Let ρ_C be the radius of convergence of $C^\bullet(z)$. Then $C^\bullet(z)$ has a square-root singular expansion around ρ_C, and $(y,z) = (C^\bullet(\rho_C), \rho_C)$ is a solution of the singular system*

$$\begin{aligned} y &= F(y,z) = z\exp(g(y,z) + \Sigma(z)), \\ 1 &= y g_y(y,z), \end{aligned} \tag{4.8}$$

with $g(y,z)$ and $\Sigma(z)$ as before.

Proof. Note that $h(z) = Z_{\mathcal{B}'}(C^\bullet(z), C^\bullet(z^2), \dots)$ is bounded from above coefficient-wise by $C^\bullet(z)$ as $C^\bullet(z)$ is a solution of equation (4.8), hence the singularity of $h(z)$ is larger than ρ_C. Since $\rho_C \in (0,1)$, as mentioned above, the function $\Sigma(z) = \sum_{i\geq 2} h(z^i)/i$ is analytic at ρ_C. By subcriticality, the function $g(y,z)$ is analytic at $(C^\bullet(\rho_C), \rho_C)$. Hence $F(y,z)$ is analytic at $(C^\bullet(\rho_C), \rho_C)$. Since the system is clearly strongly connected and the function $C^\bullet(z)$ aperiodic, we conclude from Theorem A.3 (see page 144) that $C^\bullet(z)$ has a square-root expansion at ρ_C. ☐

Lemma 4.10. *Define* $R(s,z) := Z_{\mathcal{C}^\bullet}(s, z^2, z^3, \ldots)$. *Then* $R(s,z)$ *has a square-root singular expansion around* (ρ_C, ρ_C), *and the singularity function* $\xi(z)$ *of* $s \mapsto R(s,z)$ *is analytic at* ρ_C *and has a negative derivative at* ρ_C.

Proof. Note that $C^\bullet(z) = R(z,z)$, hence the bivariate series $R(s,z)$ is a refinement of $C^\bullet(z)$. Equation (4.5) implies that $y = F(y,z,s) := s\,\exp(g(y,z) + \Sigma(z))$, with $g(y,z)$ and $\Sigma(z)$ as before. The singular system for $R(s,z)$ is

$$y = s\,\exp(g(y,z) + \Sigma(z)), \quad 1 = y\,g_y(y,z).$$

This is the same as the singular system of $C^\bullet(z)$ (given in Lemma 4.9) except that the variable z on the left-hand side of exp is replaced by the variable s. By Lemma 4.9,$(y,z) = (C^\bullet(\rho_C), \rho_C)$ is a solution of the singular system of $C^\bullet(z)$, hence $(y;z,s) = (C^\bullet(\rho_C); \rho_C, \rho_C)$ is a solution of the singular system of $R(s,z)$, and $F(y;z,s)$ is analytic at $(C^\bullet(\rho_C); \rho_C, \rho_C)$, since $g(y,z)$ is analytic at $(C^\bullet(\rho_C), \rho_C)$. Thus, Theorem A.3 ensures that $R(s,z)$ has a square-root singular expansion at (ρ_C, ρ_C). In addition, the singularity function $\xi(z)$ has a negative derivative, since $\xi(z) = -F_s/F_z$ and $F_s(y;z,s)$ depends only on z. ☐

Theorem 4.11. *Let* $\mathcal{G}$ *be an unlabelled subcritical block-stable graph class. Then there exist constants* $c > 0$ *and* γ *such that*

$$[z^n]G(z) = c\,n^{-5/2}\,\gamma^n(1 + o(1)) \quad \text{as } n \to \infty$$

for $\gamma \geq \gamma^* \approx 2.95576$, *where* $\gamma^* = \frac{1}{\rho}$ *is the exponential growth rate of the number of Pólya trees.*

Proof. First we show that $C(z)$ has a singular expansion of order 3/2 at ρ_C. Define $Q(s,z) := Z_{\mathcal{C}}(s, z^2, z^3, \ldots)$ (note that $C(z) = Q(z,z)$). The general relation $Z_{\mathcal{A}'} = \frac{\partial}{\partial s_1} Z_{\mathcal{A}}$ ensures that $R(s,z) = sQ_s(s,z)$, hence

$$Q(s,z) = Q(0,z) + \int_0^s R(w,z)\frac{\mathrm{d}w}{w}.$$

The term $Q(0,z) = Z_{\mathcal{C}}(0, z^2, z^3, \ldots) = q(z)$ is analytic at ρ_C by subcriticality of $\mathcal{G}$. Since $R(s,z)$ has a square-root expansion at (ρ_C, ρ_C), the integral term has a singular expansion of order 3/2 at (ρ_C, ρ_C) (see Lemma A.2) of the form

$$Q(s,z) = a(s,z) + h(s,z) \cdot (1 - s/\xi(z))^{\frac{3}{2}}.$$

Therefore, $C(z) = Q(z,z)$ has a singular expansion of the form

$$C(z) = a(z,z) + h(z,z) \cdot \left(\frac{\xi(z) - z}{\xi(z)}\right)^{\frac{3}{2}}.$$

Since $\xi(z)$ has a negative derivative at ρ_C and $\xi(\rho_C) = \rho_C$, the function $\frac{\xi(z)-z}{1-\frac{z}{\rho_C}}$ is analytic and nonzero at ρ_C. We conclude that $C(z)$ has a singular expansion of order 3/2, of the form

$$C(z) = \alpha(z) + \beta(z) \cdot \left(1 - \frac{z}{\rho_C}\right)^{3/2},$$

with $\alpha(z) = a(z,z)$ and $\beta(z) = h(z,z) \cdot \left(\frac{\lambda(z)}{\xi(z)}\right)^{\frac{3}{2}}$.

Recall that $G(z)$ and $C(z)$ are related by

$$G(z) = \exp(C(z) + E(z)), \quad \text{with } E(z) := \sum_{i \geq 2} \frac{1}{i} C(z^i).$$

Since $E(z)$ is analytic at ρ_C, the singular expansion of order $\frac{3}{2}$ at ρ_C for $C(z)$ yields also a singular expansion of order $\frac{3}{2}$ at ρ_C for $G(z)$. The transfer lemma then yield the estimate for $[z^n]G(z)$. □

Similarly as in the labelled case, we provide conditions that imply subcriticality, but will be convenient to check on examples:

Lemma 4.12. *For an unlabelled block-stable graph class $\mathcal{G}$, let $\eta(z)$ be the radius of convergence of $y \mapsto g(y, z)$ for $z > 0$. Assume that*

1. *there exist constants c and $\gamma > 0$ such that $[z^n]C^\bullet \leq c\,\gamma^n$,*
2. *the series $g_y(y, z) := \frac{\partial}{\partial y} g(y, z)$ satisfies $\lim_{y \to \eta(\rho_C)^-} g_y(y, \rho_C) = +\infty$,*
3. *the function $\eta(z)$ is continuous at ρ_C, and*
4. *the radius of convergence of $q(z) = Z_{\mathcal{C}}(0, z^2, z^3 \dots)$ is larger than ρ_C.*

Then the unlabelled class $\mathcal{G}$ is subcritical.

Proof. Criterion 4 in the above theorem is identical to (iv) in Definition 4.8, criterion 1 directly implies (i), that is ρ_C is non-zero. We even argued that $\rho_C < 1$ in a remark right after the definition, thus Lemma 4.9 implies that $\Sigma(z)$ is analytic at ρ_C, which proves (ii). It remains to prove (iii) from Definition 4.8, namely that $g(y, z)$ is analytic at $(C^\bullet(\rho_C), \rho_C)$. The proof of this is similar to that of Lemma 4.7, even though more technical. We refer the reader to the proof of [16, Lemma 14].

□

4.1.3 Examples of subcritical graph families

In this section, we present several block-stable families of graphs, which are subcritical both in the labelled and in the unlabelled case (cf [16]).

Definition 4.13. *We define the following families of graphs:*

- *Cacti graphs are graphs where every edge lies in exactly one elementary cycle, that is the class of* 2*-connected components consists of convex polygons together with a simple edge.*
- *Outerplanar graphs are graphs which can be embedded in the plane such that all nodes lie on the outer face, hence the class of* 2*-connected components consists of dissected polygons together with a simple edge.*
- *Series-parallel graphs are graphs which are obtained by series (subdividing) and parallel(doubling) extensions of the edges of a forest.*

There are several other ways to define the families of graphs above, e.g. by sets of avoided minors. Outerplanar graphs are all those graphs which do not contain the complete graph on 4 vertices, K_4, or the complete bipartite graph on 2 sets of 3 vertices each, $K_{2,3}$, as a minor (Note that these graphs are the smallest graphs which do not have an embedding as explained above). Series-parallel graphs are those graphs not containing K_4 as a minor. All these classes do not contain 3-connected components, as K_4 is the smallest 3-connected graph. A result analogous to the following theorem holds also for classes of graphs that are stable under taking connected, 2-connected, *and* 3-connected components, but have only a finite 3-connected subclass (cf. [32]).

Theorem 4.14. *The classes of cacti graphs, outerplanar graphs, and series-parallel graphs are subcritical both in the labelled and unlabelled cases. As a consequence, the counting coefficient g_n of each of these classes – $g_n = |G_n|/n!$ in the labelled case, $g_n = |G_n|$ in the unlabelled case – is asymptotically of the form*

$$g_n = g\, n^{-5/2} \rho_C^{-n} (1 + o(1))$$

for some constants $g > 0$, $\rho_C \in (0, 1)$. The first few digits of $1/\rho_C$ in the labelled case and the approximate values of $1/\rho_C$ in the unlabelled case are resumed in Table 4.1.

As mentioned in the introduction of this chapter, the above asymptotic estimates have been separately obtained by several people (cf [59, 64] for cacti graphs, [7, 5] for outerplanar graphs and [7] for labelled series-parallel graphs. Still, in [16] all results are reproved by a unified method, checking whether the sufficient conditions for subcriticality (Lemma 4.7 in the labelled case, Lemma 4.12 in the unlabelled case) are satisfied, hence we refer the reader to this paper for a proof of Theorem 4.14.

By numerical calculations, it is possible to compute the singularities of the given families of graphs, and hence also the growth rate of their coefficients. These estimates are given in table 4.1.

Family	$1/\rho_C$	
	Labelled	Unlabelled
Acyclic	2.71828	2.95577
Cacti	4.18865	4.50144
Outerplanar	7.32708	7.50360
Series-Parallel	9.07359	9.38527

Table 4.1: Exponential growth for distinct subcritical classes. All constants are referred to connected and general classes.

4.2 The degree distribution in subcritical graph families

In this section we discuss the degree distribution of graphs of a subcritical family $\mathcal{G}$. We denote by X_n^k the random variable which counts the number of vertices of degree k in a randomly chosen member of size n of the family $\mathcal{G}$, either labelled or unlabelled. We further denote by d_k the limiting probability (as n tends to infinity) that the degree of root vertex of a randomly chosen member of $\mathcal{G}'$ is k. The main tools to obtain asymptotic results is the Drmota-Lalley-Woods Theorem on multiple variables (Theorem A.3), giving singular expansions of generating functions, and the refinement of Theorem 1.20 to systems of equations (Theorem A.5), stating sufficient conditions to obtain a central limit law.

4.2.1 The labelled case

Consider the series

$$B'_j(z,\mathbf{v}) = \sum_{n,n_1,\dots,n_k,n_\infty} b'_{j;n;n_1,\dots,n_k,n_\infty} v_1^{n_1}\cdots v_k^{n_k} v_\infty^{n_\infty} \frac{z^n}{n!},$$

where we use the notation $\mathbf{v} = (v_1,\dots,v_k,v_\infty)$ and where $b'_{j;n;n_1,\dots,n_k,n_\infty}$ is the number of derived 2-connected graphs with $1+n = 1+n_1+\cdots+n_k+n_\infty$ vertices such that the distinguished vertex has degree j and the remaining n vertices are labelled by $1,2,\dots,n$ and where n_ℓ vertices have degree ℓ, $1\le \ell \le k$, and n_∞ vertices have degree greater than k.

By definition we have

$$B'(z) = \sum_{2\le j\le\infty} B'_j(z,\mathbf{1}).$$

Hence the radius of convergence of the functions $B'_j(z,\mathbf{1})$ is greater or equal than the radius of convergence of $B'(z)$.

We introduce an analogous series

$$C'_j(z,\mathbf{v}) = \sum_{n,n_1,\dots,n_k,n_\infty} c'_{j;n;n_1,\dots,n_k,n_\infty} v_1^{n_1}\cdots v_k^{n_k} v_\infty^{n_\infty} \frac{z^n}{n!},$$

for derived connected graphs. For convenience, we set

$$C'_0(z,\mathbf{v}) = 1,$$

which corresponds to the case of a graph consisting of a single derived vertex. We further set

$$B'(z,w) = \sum_j B'_j(z,\mathbf{1}) w^j = \sum_{n,j} b'_{n,j} z^n w^j$$

where $b'_{n,j}$ is the number of derived 2-connected graphs with $n+1$ vertices, where one vertex of degree j is marked and the remaining n vertices are labelled by $1,2,\dots,n$. Analogously, we define the function $C'(z,w) = \sum_j C'_j(z,\mathbf{1})w^j$. According to the block decomposition of connected graphs

$$C'(z,w) = \exp(B'(zC'(z),w)). \tag{4.9}$$

In the following, we set $B'_j(z) := B'_j(z,\mathbf{1})$. We will also use the series $B(z,\mathbf{v})$, which is the refined version of the generating function $B(z)$ of unrooted blocks taking into account vertex degrees.

We need the following auxiliary result, proven in [15][Theorem 9.17] with the help of Lemma A.1, to prove the main result of this section.

Lemma 4.15. *The generating function $p(w) = \sum d_k w^k$ satisfies*

$$p(w) = \rho_C e^{B'(z,w)} \frac{\partial}{\partial z} B'(z,w)\bigg|_{z=\rho_C C'(\rho_C)},$$

and $p(1) = 1$, thus the d_k's are indeed a probability distribution.

With the help of this lemma, we can prove the following.

Proposition 4.16. *Let $\mathcal{G}$ be a subcritical family of random labelled graphs and $\mathcal{G}'$ be its derived family. Then for k fixed, the limiting probability d_k that the pointed vertex of a member of $\mathcal{G}'$ has degree k exists and is given by*

$$d_k = \rho_C \left(\sum_{i=1}^{k} \frac{\partial}{\partial z} B_i'(z)\big|_{z=\rho_C C'(\rho_C)} C_{k-i}'(\rho_C) \right).$$

Proof. Using Lemma 4.15, we have

$$\begin{aligned} d_k &= [w^k]p(w) = [w^k]\rho_C \exp(B'(z,w)) \frac{\partial}{\partial z} B'(z,w) \Big|_{z=\rho_C C'(\rho_C)} \\ &= \rho_C \left(\sum_{i=0}^{k} [w^{k-i}] \exp(B'(z,w))[w^i] \frac{\partial}{\partial z} B'(z,w) \right) \Big|_{z=\rho_C C'(\rho_C)} \\ &= \left(\sum_{i=1}^{k} \frac{\partial}{\partial z} B_i'(z)[w^{k-i}] \left(1 + \sum_k B_k'(z) w^k + \frac{(\sum_k B_k'(z) w^k)^2}{2} + \cdots \right) \right) \Big|_{z=\rho_C C'(\rho_C)} \\ &= \left(\sum_{i=1}^{k} \frac{\partial}{\partial z} B_i'(z) \left(\sum_{m=1}^{k-i} \sum_{l_1+\cdots l_m = k-i} B_{l_1}(z) \cdots B_{l_m}(z) \right) \right) \Big|_{z=\rho_C C'(\rho_C)}. \end{aligned}$$

From there the result follows, as the second term is exactly the representation of a connected graph of root degree $k-i$ according to the block decomposition, evaluated at $z = \rho_C$, i.e. $C_{k-i}'(\rho_C)$. ☐

To obtain results on the degree distribution of random vertices, we derive a system of functional equations which is satisfied by $C_j' = C_j'(z, \mathbf{v})$. This is a refined version of the equation

$$C^\bullet(z) = z \exp(B'(C^\bullet(z))).$$

Lemma 4.17. *Consider the series $W_j = W_j(z, \mathbf{v}; C_1', \ldots, C_k', C_\infty')$, $j \in \{1, 2, \ldots, k, \infty\}$, defined by*

$$W_j = \sum_{i=0}^{k-j} v_{i+j} C_i'(z, \mathbf{v}) + v_\infty \left(\sum_{i=k-j+1}^{k} C_i'(z, \mathbf{v}) + C_\infty'(z, \mathbf{v}) \right), \qquad 1 \le j \le k,$$

$$W_\infty = v_\infty \left(\sum_{i=0}^{k} C_i'(z, \mathbf{v}) + C_\infty'(z, \mathbf{v}) \right).$$

Set $\mathbf{W} = (W_1, \ldots, W_k, W_\infty)$. Then the series $C_1', \ldots, C_k', C_\infty'$ satisfy the system of equations

$$C_j'(z, \mathbf{v}) = \sum_{\ell_1 + 2\ell_2 + 3\ell_3 + \cdots j\ell_j = j} \prod_{r=1}^{j} \frac{B_r'(z, \mathbf{W})^{\ell_r}}{\ell_r!}, \qquad 1 \le j \le k,$$

$$\begin{aligned} C_\infty'(z, \mathbf{v}) &= \exp\left(\sum_{j=1}^{k} B_j'(z, \mathbf{W}) + B_\infty'(z, \mathbf{W}) \right) - 1 \\ &\quad - \sum_{1 \le \ell_1 + 2\ell_2 + 3\ell_3 + \cdots k\ell_k \le k} \prod_{r=1}^{k} \frac{B_r'(z, \mathbf{W})^{\ell_r}}{\ell_r!}. \end{aligned}$$

Proof. As already indicated, the proof is a refined version of the functional equation fulfilled by $C^\bullet(z)$, which reflects the decomposition of a rooted connected graph into a finite set of derived 2-connected graphs, where every vertex (different from the root) is substituted by a rooted connected graph. The functions W_j serve the purpose of marking (recursively) the degree of the vertices in the 2-connected blocks which are substituted by other graphs. In the definition of W_j, a connected graph with a distinguished vertex of degree i is plugged into the vertex of degree j, hence this vertex now has degree $i+j$, which is marked accordingly by v_{i+j} in the sum. The analogous holds for W_∞.

□

With the same preliminaries as in Proposition 4.16, we prove the following central limit theorem:

Theorem 4.18. *The random variable X_n^k that counts the number of vertices of degree k in a randomly chosen member of $\mathcal{G}$ satisfies a central limit theorem with mean $\mathbb{E}\, X_n^k = d_k n + \mathcal{O}(1)$ and variance $\mathbb{V}\mathrm{ar}\, X_n^k = \sigma_k^2 n + \mathcal{O}(1)$, with some computable constant $\sigma_k > 0$.*

Proof. To prove Theorem 4.18 we observe

$$C'(z) = \sum_{0\le j\le\infty} C_j'(z,\mathbf{1}).$$

Furthermore, since the above system of equations is strongly connected, all functions $C_j'(z,\mathbf{1})$ have the same radius of convergence as $C'(z)$ by the Drmota-Lalley-Woods Theorem (Theorem 1.12). By subcriticality this radius of convergence is smaller than the radius of convergence of $B'(z)$. Hence, if $\mathbf{v}$ is sufficiently close to $\mathbf{1}$ then the singularities of B_j' and C_j' do not interfere due to Lemma 1.18, in particular we can apply Theorem A.3 and obtain that all functions C_j' have a square-root singularity. Finally, let

$$C^{(k)}(z,v) = \sum_{n,m} c_{k;n,m} v^m \frac{z^n}{n!}$$

be the generating function for the numbers $c_{k;n,m}$ of unrooted connected outerplanar graphs of size n with m nodes of degree k. Then $C^{(k)}(z,v)$ satisfies

$$\frac{\partial C^{(k)}(z,v)}{\partial z} = \sum_{j=1}^{k-1} C_j'(z,1,\ldots,1,v,1) + vC_k'(z,1,\ldots,1,v,1) + C_\infty'(z,1,\ldots,1,v,1),$$

and thus, by integration, $C^{(k)}(z,v)$ has a singular expansion of order $\frac{3}{2}$ around $v=1$.

Furthermore, the central limit theorem for the number of vertices of given degree k with asymptotic mean $\mu_k n$ and variance $\sigma_k^2 n$ follows by the analogue of Theorem A.5 for systems of equations. It immediately follows that $d_k = \mu_k$ as there are exactly n possible ways to root an unrooted object of size n at one of the vertices and thus the probability that a random vertex has degree k is exactly the same as the probability that the root vertex has degree k. Despite that, we can also use formula (A.4) to compute μ_k, we obtain the same result as for d_k here, as we will see in the following:

$$\mu_k = \frac{1}{\rho_C} \frac{\mathbf{b}^T \mathbf{F}_v(\rho_C, \mathbf{C}(\rho_C,1),1)}{\mathbf{b}^T \mathbf{F}_z(\rho_C, \mathbf{C}(\rho_C,1),1)}$$

where $\mathbf{F}(z,\mathbf{y},v)$ is the vector defined by the system of equations $\mathbf{C}(z,v)=\mathbf{F}(z,\mathbf{C}(z,v),v)$, with $\mathbf{C}(z,v)=(C_1(z,v),\ldots,C_k(z,v),C_\infty(z,v))$, and $\mathbf{b}$ is a unique positive left eigenvector of $\mathbf{F_y}$ to the eigenvalue 1, which, in a system like ours, is $\mathbf{b}=(1,\ldots,1)^T$ (cf [15]). For brevity, we denote by $\sum_{j=1}^{k,\infty} B_j=\sum_{j=1}^{k} B_j+B_\infty$ in the following. Note that

$$F_\infty=\underbrace{\exp(\sum_{j=1}^{k,\infty} B_j'(z,\mathbf{W}))}_{=:E}-1-\sum_{j=1}^{k} F_k,$$

as F_∞ counts all graphs with root degrees different from $1,\ldots k$, therefore only the derivative of the exponential term E remains in $\mathbf{b}^T\mathbf{F}_v(\rho_C,\mathbf{C}(\rho_C,1),1)$ and $\mathbf{b}^T\mathbf{F}_z(\rho_C,\mathbf{C}(\rho_C,1),1)$, as all other terms cancel in the vector product. This gives

$$\mathbf{bF}_v(\rho_C,\mathbf{C}(\rho_C,1),1)=\left(E\cdot\sum_{j=1}^{k,\infty}\sum_{i=1}^{k}(B_j')_{v_i}(z,\mathbf{W})y_{k-i}\right)\Bigg|_{\rho_C,\mathbf{C}(\rho_C,1),1)}$$

$$\mathbf{bF}_z(\rho_C,\mathbf{C}(\rho_C,1),1)=\left(E\cdot\sum_{j=1}^{k,\infty}(B_j')_z(z,\mathbf{W})\right)\Bigg|_{\rho_C,\mathbf{C}(\rho_C,1),1)}$$

Note that we can exchange summation in $\mathbf{bF}_v$ and that $B_j'(z,\mathbf{v})=\frac{1}{z}\frac{\partial}{\partial v_j}B(z,v_1,\ldots,v_\infty)$. Further,

$$W_i(\rho_C,1)=C'(\rho_C)\quad\text{for}\quad i=1,\ldots,k,\infty.$$

With that, it is relatively easy to prove

$$\sum_{j=1}^{k,\infty}(B_j)_{v_i}'(\rho_C,C'(\rho_C),\ldots,C'(\rho_C))=\rho_C(B_i)_z(\rho_C C'(\rho_C))$$

and

$$\sum_{j=1}^{k,\infty}(B_j)_z'(\rho_C,C'(\rho_C),\ldots,C'(\rho_C))=C'(\rho_C)B''(\rho_C C'(\rho_C))=\frac{1}{\rho_C},$$

the latter relation following from the implicit representation in a single equation of $C'(z)$. Hence, μ_k is given by the same formula as d_k. □

4.2.2 The unlabelled case

We introduce cycle index sums

$$Z_{\mathcal{B}_j'}(\mathbf{s_1},\bar{\mathbf{u_1}})=Z_{\mathcal{B}_j'}(s_1,s_2,\ldots;u_{1,1},u_{1,2},\ldots;\ldots;u_{k,1},u_{k,2},\ldots;u_{\infty,1},u_{\infty,2},\ldots)$$

for the class of pointed blocks, where the pointed vertex has degree j and is not counted, and where the variables $u_{i,j}$ count the cycles of length j of vertices of degree i, and $u_{\infty,j}$ counts those vertices of degree greater than k. As in Section 4.2.1 let $\mathbf{v}=(v_1,\ldots,v_k,v_\infty)$. Denote the corresponding OGFs by $B_j'(z,\mathbf{v}), j=1,\ldots,k,\infty$ and let

$$B'(z,\mathbf{v}):=\sum_{j=2}^{k} v_j B_j'(z,\mathbf{v})+v_\infty B_\infty'(z,\mathbf{v})$$

Note that

$$Z_{\mathcal{B}'}(\mathbf{s_1}, \mathbf{1}) = Z_{\mathcal{B}'}(\mathbf{s_1}),$$

and thus the singularity $\rho_B(\mathbf{v})$ of $B'(z, \mathbf{v})$ is the same as that of $B'(z)$ at $\mathbf{v} = \mathbf{1}$, and $\rho_B(\mathbf{v})$ is the dominant singularity of the system $\mathbf{B}'(z, \mathbf{v}) = (B'_j(z, \mathbf{v}))_{j=1..k,\infty}$.

We now introduce the multivariate generating functions

$$C'_j(z, \mathbf{v}) = \sum_{n;n_1,\ldots,n_k,n_\infty} c_{i;n;n_1,\ldots,n_k,n_\infty} v_1^{n_1} \cdots v_k^{n_k} v_\infty^{n_\infty} z^n \tag{4.10}$$

where the coefficient $c_{i;n;n_1,\ldots,n_k,n_\infty}$ denotes the number of elements of size n of $\mathcal{C}'$, where the pointed vertex has degree j and with $n_i, i = 1, \ldots, k$, vertices of degree i and n_∞ vertices of degree greater than k. We further set

$$B'(z, w) = \sum_j B'_j(z, \mathbf{1}) w^j = \sum_{n,j} b'_{n,j} z^n w^j$$

and

$$C'(z, w) = \sum_j C'_j(z, \mathbf{1}) w^j = \sum_{n,j} c'_{n,j} z^n w^j.$$

As we need cycle indices for the block decomposition, we set

$$Z_{\mathcal{B}'}(\mathbf{s_1}; w) = \sum_j Z_{\mathcal{B}'_j}(\mathbf{s_1}) w^j.$$

Note that the variable w, which counts the degree of the root, is not involved in any permutation cycle. Then,

$$C'(z, w) = \exp\left(\sum_{\ell \geq 1} \frac{1}{\ell} Z_{\mathcal{B}'}(\mathbf{z}^\ell \mathbf{C}'(\mathbf{z}^\ell); w^\ell) \right).$$

In analogy to the labelled case, we first prove the following auxiliary result.

Lemma 4.19. *The generating function $p(w) = \sum d_k w^k$ satisfies*

$$p(w) = \rho \frac{\partial}{\partial u} \exp\left(Z_{\mathcal{B}'}(u, z^2 C'(z^2), \ldots; w) + \sum_{\ell \geq 2} \frac{1}{\ell} Z_{\mathcal{B}'}(\mathbf{z}^\ell \mathbf{C}'(\mathbf{z}^\ell); w^\ell) \right),$$

and $p(1) = 1$, that is, the d_k are indeed a probability distribution.

Proof. We use Lemma A.1 with

$$f(z, w) = zC'(z)$$

$$H(z, w, u) = \exp\left(Z_{\mathcal{B}'}(u, z^2 C'(z^2), \ldots; w) + \sum_{\ell \geq 2} \frac{1}{\ell} Z_{\mathcal{B}'}(\mathbf{z}^\ell \mathbf{C}'(\mathbf{z}^\ell); w^\ell) \right)$$

and the same line of reasoning as in Lemma 4.15 proves the first part of the lemma. For $p(1) = H_u(\rho, 1, \rho C'(\rho))$ we obtain

$$\rho C'(\rho) \frac{\partial}{\partial u} Z_{\mathcal{B}'}(u, \rho^2 C'(\rho^2), \ldots) \Big|_{u=\rho C'(\rho)} = 1$$

by the implicit function representation (4.7) of $zC'(z)$. ∎

Proposition 4.20. *Let $\mathcal{G}$ be a family of random subcritical graphs and $\mathcal{G}'$ be it's derived family. Further, let d_k be the limiting probability that the root vertex of a member of $\mathcal{G}'$ has degree k and let X_n^k be the random variable that counts the number of vertices of degree k in a randomly chosen member of $\mathcal{G}$. Then*

$$d_k = \rho \left(\sum_{i=1}^{k} \frac{\partial}{\partial z} Z_{\mathcal{B}'_i}(z, \rho^2 C'(\rho^2), \rho^3 C'(\rho^3), \ldots)\big|_{z=\rho C'(\rho)} C'_{k-i}(\rho) \right).$$

Proof. With the help Of Lemma 4.19 we can determine $d_k = [w^k]p(w)$:

$$\begin{aligned} d_k &= \rho\ [w^k] \left[\exp\left(\sum_{\ell \geq 1} \frac{1}{\ell} Z_{\mathcal{B}'}(\rho^\ell \mathbf{C}'(\rho^\ell); w^\ell) \right) \cdot \left(\frac{\partial}{\partial u} Z_{\mathcal{B}'}(u, \rho^2 C'(\rho^2), \ldots; w) \right)_{u=\rho C'(\rho)} \right] \\ &= \rho\ [w^k] \left[\exp\left(\sum_{\ell \geq 1} \frac{1}{\ell} Z_{\mathcal{B}'}(\rho^\ell \mathbf{C}'(\rho^\ell); w^\ell) \right) \cdot \left(\sum_{\ell \geq 1} \frac{\partial}{\partial u} Z_{\mathcal{B}'_\ell}(u, \rho^2 C'(\rho^2), \ldots) w^\ell \right)_{u=\rho C'(\rho)} \right] \\ &= \rho \sum_{i=1}^{k} \frac{\partial}{\partial u} Z_{\mathcal{B}'_i}(u, \rho^2 C'(\rho^2), \ldots)\big|_{u=\rho C'(\rho)} \cdot [w^{k-i}] \exp\left(\sum_{\ell \geq 1} \frac{1}{\ell} Z_{\mathcal{B}'}(\rho^\ell \mathbf{C}'(\rho^\ell); w^\ell) \right) \end{aligned}$$

where the second term translates into $C'_{k-i}(\rho)$ as

$$\exp\left(\sum_{\ell \geq 1} \frac{1}{\ell} Z_{\mathcal{B}'}(\rho^\ell \mathbf{C}'(\rho^\ell); w^\ell) \right) = C'(\rho, w) = \sum_j C'_j(\rho, \mathbf{1}) w^j = \sum_j C'_j(\rho) w^j.$$

□

As in the labelled case, we observe that the functions $C'_j(z, \mathbf{v})$ satisfy a system of equations, using a refinement of the block decomposition. Recall that we denote by $\sum_{i=r}^{k,\infty} F_i = \sum_{i=r}^{k} F_i + F_\infty$.

Lemma 4.21. *For each $j = 1, 2, \ldots, k, \infty$, let W_j be defined by*

$$\begin{aligned} W_j(z, \mathbf{v}) &= \sum_{i=0}^{k-\ell} v_{j+i} C'_j(z, \mathbf{v}) + v_\infty \left(\sum_{i=k-j+1}^{k,\infty} C'_i(z, \mathbf{v}) \right), \\ W_\infty(z, \mathbf{v}) &= v_\infty \left(\sum_{i=0}^{k,\infty} C'_i(z, \mathbf{v}) \right). \end{aligned}$$

Let $W_{j,i} = W_j(z^i, v_1^i, \ldots, v_k^i, v_\infty^i)$ and

$$\mathbf{W}^{(l)} = (W_{1,l}, W_{1,2l}, \ldots; \ldots; W_{k,l}, W_{k,2l}, \ldots; W_{\infty,l}, W_{\infty,2l}, \ldots).$$

$Z_n[Z_\mathcal{B}]$ denotes the substitution $s_l \leftarrow Z_\mathcal{B}(s_l, s_{2l}, \ldots; \mathbf{W}^{(l)}), l \geq 1$ in $S_n(s_1, s_2, \ldots)$.

The series $C'_1, \dots, C'_k, C'_\infty$ satisfy the system of equations

$$C'_j(z,\mathbf{v}) = \sum_{l_1+2l_2+\dots+jl_j=j} \prod_{r=1}^{j} S_{l_r}\left[Z_{\mathcal{B}'_r}\right]_{(s_i=z^i)}, \quad j=1,\dots,k$$

$$C'_\infty(z,\mathbf{v}) = \exp\left(\sum_{l\geq 1}\frac{1}{l}\left(\sum_{r=1}^{k} Z_{\mathcal{B}'_r}\left(z^l, z^{2l}, \dots; \mathbf{W}^{(l)}\right) + Z_{\mathcal{B}'_\infty}\left(z^l, z^{2l}, \dots; \mathbf{W}^{(l)}\right)\right)\right)$$
$$- \sum_{l_1+2l_2+\dots+kl_k\leq k} \prod_{r=1}^{j} S_{l_r}\left[Z_{\mathcal{B}'_r}\right]_{(s_i=z^i)}$$

Proof. As in the labelled case, we refine the recursive decomposition of graphs into their 2-connected components. The functions $W_{j,i}$ play the analogous role as in the labelled case, except that we need a second index for representing the cycles of different length appearing in the cycle indices. This directly leads to the representation of $C'_j(z,\mathbf{v})$ for $j=1,\dots k$. For $C'_\infty(z,\mathbf{v})$ we obtain

$$\sum_{(l_1,l_2,\dots,l_k,l_\infty)} \prod_{r=1}^{k} S_{l_r}\left[Z_{\mathcal{B}'_r}\left(z,z^2,\dots;\mathbf{W}^{(1)}\right)\right] S_{l_\infty}\left[Z_{\mathcal{B}'_\infty}\left(z,z^2,\dots;\mathbf{W}^{(1)}\right)\right]$$
$$- \sum_{l_1+2l_2+\dots l_k\leq k} \prod_{r=1}^{k} S_{l_r}\left[Z_{\mathcal{B}'_r}\left(z,z^2,\dots;\mathbf{W}^{(1)}\right)\right],$$

We can rewrite the first sum to

$$\begin{aligned}
&\sum_{l_1\geq 0} S_{l_1}[Z_{\mathcal{B}'_1}(s_1,s_2,\dots,W_{1,1},W_{1,2},\dots,W_{k,1},W_{k,2},\dots)] \\
&\times \sum_{l_2\geq 0} S_{l_2}[Z_{\mathcal{B}'_2}(s_1,s_2,\dots,W_{1,1},W_{1,2},\dots,W_{k,1},W_{k,2},\dots)] \\
&\times \cdots \\
&= \prod_{r=1}^{k,\infty}\left(\sum_{l_r\geq 0} S_{l_r}[Z_{\mathcal{B}'_r}(s_1,s_2,\dots,W_{1,1},W_{1,2},\dots,W_{k,1},W_{k,2},\dots)]\right) \\
&= \prod_{r=1}^{k,\infty}\exp\left(\sum_{l\geq 1}\frac{s_l}{l}[Z_{\mathcal{B}'_r}(s_1,s_2,\dots,W_{1,1},W_{1,2},\dots,W_{k,1},W_{k,2},\dots)]\right) \\
&= \prod_{r=1}^{k,\infty}\exp\left(\sum_{l\geq 1}\frac{Z_{\mathcal{B}'_r}(s_l,s_{2l},\dots,W_{1,l},W_{1,2l},\dots,W_{k,l},W_{k,2l},\dots)}{l}\right) \\
&= \exp\left(\sum_{r=1}^{k,\infty}\sum_{l\geq 0}\frac{Z_{\mathcal{B}'_r}(s_l,s_{2l},\dots,W_{1,l},W_{1,2l},\dots,W_{k,l},W_{k,2l},\dots)}{l}\right) \\
&= \exp\left(\sum_{l\geq 0}\sum_{r=1}^{k,\infty}\frac{Z_{\mathcal{B}'_r}(s_l,s_{2l},\dots,W_{1,l},W_{1,2l},\dots,W_{k,l},W_{k,2l},\dots)}{l}\right)
\end{aligned}$$

□

Remark. The functions $W_{j,i}$ and thus the whole system can also be considered in terms of cycle index sums, using the cycle index sum $Z_{C'_j}(\mathbf{s_1}, \bar{\mathbf{u}}_\mathbf{1})$ for rooted connected graphs. The root vertices in $W_{j,i}$ are fixed and thus have cycle length 1. We will need this terminology in the proof of Lemma 4.23.

It is easily checked that the system is strongly connected as every $C'_j(z, \mathbf{v})$ depends on $C'_\infty(z, \mathbf{v})$ for $j = 1, \ldots, k$ and $C'_\infty(z, \mathbf{v})$ depends on all $C'_j(z, \mathbf{v}), j = 1, \ldots, k$.

Obviously,

$$\left[\sum_{j=1}^{k} v_j C'_j(z, \mathbf{v}) + v_\infty C'_\infty(z, \mathbf{v}) \right]_{\mathbf{v}=\mathbf{1}} = C'(z). \tag{4.11}$$

Define

$$C'(z, \mathbf{v}) := \sum_{j=1}^{k} v_j C'_j(z, \mathbf{v}) + v_\infty C'_\infty(z, \mathbf{v}).$$

Then the generating function of derived connected graphs of $\mathcal{G}$, where the variable v counts the nodes of degree k, is given by $C'^{(k)}(z, v) = C'(z, \mathbf{v}_k)$, where $\mathbf{v}_k = (1, 1, \ldots, 1, v, 1)$.

Lemma 4.22. *$C'^{(k)}(z, v)$ has a square-root singular expansion around its singularity $\rho_C(v)$ in a neighbourhood of $v = 1$.*

Proof. By Equation (4.11) the singularity of the system $\mathbf{C}'(z, \mathbf{v})$ at $\mathbf{v} = \mathbf{1}$ is ρ_C, the same as that of $C'(z)$. As the system is strongly connected, every $C'_j(z, \mathbf{1})$ has radius of convergence ρ_C. Since $C'^{(k)}(z, 1)$ is a linear combination of these functions, it has the same singularity, which fulfills $\rho_C C'(\rho_C) < \rho_B$ due to the subcriticality assumption. By Lemma 1.18 it follows that $(C'^{(k)}(z, v), B'^{(k)}(z, v))$ is subcritical near 1, and hence we obtain a square-root singular expansion. ☐

Consider the cycles index sums $Z_\mathcal{C}(\mathbf{s_1}, \bar{\mathbf{u}}_\mathbf{1})$ for (unpointed) connected graphs of $\mathcal{G}$. By taking $s_1 = s$, $s_i = z^i$ for $i \geq 2$, $u_{j,i} = 1$ for $1 \leq j < k, i \geq 1$ and $u_{k,i} = v^i$ for $i \geq 1$ we obtain its corresponding OGF $C^{(k)}(z, v) = Z_\mathcal{C}(z, z^2, \ldots; 1, 1, \ldots; \ldots; v, v^2, \ldots; 1, 1, \ldots)$, where the variable v counts the nodes of degree k.

Lemma 4.23. *$C^{(k)}(z, v)$ has a singular expansion of order $\frac{3}{2}$ around its singularity $\rho_C(v)$ in a neighbourhood of $v = 1$.*

Proof. We have to express the system of equations in Lemma 4.21 in terms of cycle index sums and analyze the trivariate generating functions

$$C^{(k)}(s, z, v) = Z_\mathcal{C}(s, z^2, z^3, \ldots; 1, 1, \ldots; \ldots; v, v^2, \ldots; 1, 1, \ldots).$$

Obviously, $C^{(k)}(z, z, v) = C^{(k)}(z, v)$. Analogously we define $C'^{(k)}(s, z, v)$. We obtain

$$C^{(k)}(s, z, v) = C^{(k)}(0, z, v) + \int_0^z C'^{(k)}(s, z, v)\mathrm{d}s,$$

because by rooting, we derivate with respect to the first variable, which we now isolated. Doing so, all those symmetries without fixed points got lost, so we have to take them into account when integrating. Due to the stability condition (Lemma 1.18), we obtain a square-root singular expansion for $C'^{(k)}(s,z,v)$, with a singular term of the form $(1-s/\bar{\rho}(z,v))^{\frac{1}{2}}$. Integration and subcriticality conditions lead to a singular expansion of the form

$$C^{(k)}(s,z,v) = g(s,z,v) + h(s,z,v)\left(1-\frac{s}{\bar{\rho}(z,v)}\right)^{\frac{3}{2}}.$$

At $s=z$ we can represent the singular part as

$$\left(1-\frac{z}{\bar{\rho}(z,v)}\right)^{\frac{3}{2}} = \kappa(z,v)\left(1-\frac{z}{\tilde{\rho}_2(v)}\right)^{\frac{3}{2}},$$

with an analytic factor $\kappa(z,v)$ and some analytic function $\tilde{\rho}(v)$. Since the singular manifold of $C'^{(k)}(s,z,v)$ is the same as that of $C^{(k)}(z,v)$, that is $z=\bar{\rho}(z,v)$ if and only if $z=\rho_2(v)$, it follows that $\tilde{\rho}(v)=\rho_2(v)$ □

Theorem 4.24. *With the same preliminaries as in Proposition 4.20, X_n^k satisfies a central limit theorem with mean $\mathbb{E}X_n^k = \mu_k n + \mathcal{O}(1)$ and variance $\mathbb{V}\mathrm{ar}\ X_n^k = \sigma_k^2 n + \mathcal{O}(1)$.*

Proof. With the singular expansion in $\rho(v)$ given, the proof of the theorem is now immediate by Theorem A.5 for systems of equations. □

Remark. In the unlabelled case, we cannot expect that μ_k equals d_k, as vertex-rooting is only possible at fixed points of permutations, and thus in unlabelled graphs $ng_n \neq g_n'$.

To calculate μ_k we can use (A.4). The derivatives give

$$\mathbf{bF}_v(\rho,\mathbf{C}(\rho,1),1) = \left(E\cdot\sum_{j=1}^{k,\infty}\sum_{i=1}^{k}(B_j')_{v_i}(z,\mathbf{W})y_{k-i}\right)\Bigg|_{(\rho,\mathbf{C}(\rho,1),1)},$$

$$\mathbf{bF}_z(\rho,\mathbf{C}(\rho,1),1) = \left(E\cdot\sum_{j=1}^{k,\infty}(B_j')_{z}(z,\mathbf{W})\right)\Bigg|_{(\rho,\mathbf{C}(\rho,1),1)},$$

where y_{k-i} denotes the $(k-i)$-th coordinate of $\mathbf{y}$ satisfying $\mathbf{y}=\mathbf{F}(\mathbf{y};z,\mathbf{v})$, $(B_j')_{v_i}$ is the derivative of B_j' with respect to v_i, and by E the exponential term appearing in C_∞'. As the formula includes all partial derivatives of (B_j'), we see that the calculation of μ_k will be very involved and it is not equal to d_k.

4.3 The degree distribution in selected families of 2-connected graphs

We have proven in the previous section that a central limit theorem holds for the random variable X_n^k, counting the number of vertices of given degree in a random graph of size n of a subcritical family $\mathcal{G}$. No further knowledge on the family and the shape or degree distribution in the blocks is required to obtain this result. It holds in general for all subcritical families. Due to block-stability and subcriticality, a connected graph $C\in\mathcal{G}$ is built by a large number of (almost independant) blocks of asymptotically the same size and shape. That is, the number of vertices of degree k in C corresponds to a sum of weakly dependant and identically distributed random variables, which suggests a central limit theorem. For a 2-connected graph $B \in \mathcal{G}$ there is no further decomposition into "smaller" components, therefore it is a priori not clear whether a central limit theorem will hold. Therefore, the following section is devoted to the study of the degree distribution in 2-connected outerplanar and series-parallel graphs, and a central limit theorem will be proven for the given classes.

Theorem 4.25. *Let $\mathcal{B}$ be the family of 2-connected outerplanar or series-parallel graphs and let X_n^k be the random variable which counts the number of vertices of degree k in a randomly chosen member of size n in $\mathcal{B}$. X_n^k satisfies a central limit theorem with expected value $\mathbb{E}X_n^k = \mu_k n + \mathcal{O}(1)$ and variance $\mathbb{V}X_n^k = \sigma_k^2 n + \mathcal{O}(1)$, where μ_k and σ_k are real constants.*

4.3.1 Outerplanar graphs

As described before, outerplanar graphs are planar graphs which can be embedded in the plane such that all nodes lie on the outerface. In this section, we will prove Theorem 4.25 for the family of 2-connected unlabelled outerplanar graphs $\mathcal{B}$. As a byproduct, we also obtain a central limit theorem for the derived family $\mathcal{B}'$:

Theorem 4.26. *Let $\mathcal{B}$ be the family of random 2-connected unlabelled outerplanar graphs and $\mathcal{B}'$ be its derived family. Then the random variable that counts the number of vertices of degree k in a randomly chosen member of $\mathcal{B}$, X_n^k, and the corresponding random variable $X_n'^k$ for a member of the derived family, satisfy a central limit theorem with expected value $\mathbb{E}X_n^k \sim \mathbb{E}X_n'^k \sim \mu_k n$ and variance $\mathbb{V}X_n^k \sim \mathbb{V}X_n'^k \sim \sigma_k^2 n$ where $\mu_k = 2(k-1)(\sqrt{2}-1)^k$ and σ_k^2 is a computable constant.*

Proof. The proof is divided into a combinatorial and an analytic part. The goal of the combinatorial part will be to find representations for the generating functions of the desired graph class. In fact, we will find a core system of functional equations with suitable properties for a subfamily of 2-connected outerplanar graphs, which will lead to a local singular expansion for the bivariate generating function $B'^{(k)}(x,v)$ of derived 2-connected graphs in the analytic part, where v counts vertices of degree k. Integration will lead to the result for the unrooted class.

Combinatorial part

Note that a 2-connected outerplanar graph can be interpreted as a dissection of a polygon, cf Figure 4.2.

We will start by counting outer-edge pointed dissections, i.e. dissections where one edge of the outer polygon is pointed and its endpoints not counted, from there it will be easy to count vertex pointed dissections. We first set up a system of functional equations for the cycle index sums of oriented outer-edge rooted dissections, i.e. the pointed edge is given an orientation, because the automorphism group of such dissections consists of the identity only.

An oriented outer-edge rooted dissection A^o can be decomposed in the following way: Consider the (inner) face containing the root edge, which is either a triangle or a k-gon, with $k > 3$. An orientation is implied on the edges of this face by the orientation of the root edge of the dissection. If it is a triangle, the 2 edges other than the root edge can be considered

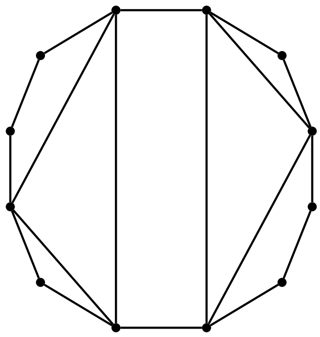

Figure 4.2: A 2-connected outerplanar graph

as root-edges of other oriented outer-edge rooted dissections, if those are not empty. Otherwise they are single edges. If $k > 3$, we insert a virtual edge connecting the endpoint of the oriented root edge with the starting point of the edge connecting to the starting point of the root edge. This imaginary edge is root-edge of another oriented outer-edge rooted dissection, just as the remaining edge is either root-edge or outer-edge (cf. Figure 4.3).

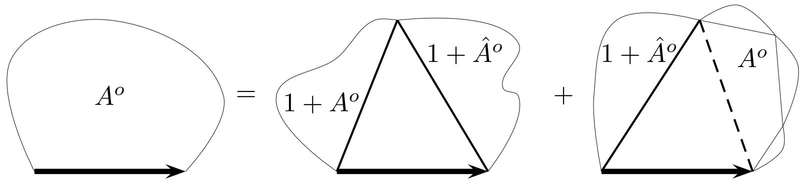

Figure 4.3: Decomposing an oriented outer-edge rooted dissection - the bottom edge is the oriented root edge while the other edges of the triangle are root edges of smaller dissections, denoted by $\hat{A}^o$. In all but one cases those dissections might be empty, denoted by 1.

Translating the above decomposition by the symbolic method, we obtain the following implicit equation for $\mathcal{A}^o$:

$$\mathcal{A}^o = \mathcal{X} \times (1 + \mathcal{A}^o)^2 + \mathcal{X} \times \mathcal{A}^o \times (1 + \mathcal{A}^o) \tag{4.12}$$

We translate this equation to cycle index sums $Z_{\mathcal{A}^o_{i,j}}(\mathbf{s_1}; \bar{\mathbf{u}}_\mathbf{1})$ of oriented outer-edge rooted dissections, where the vertices of the root edge have degrees i and j for $i, j \in \{1, \dots, k\}$ and degree greater than k for $i, j = \infty$ and are not counted.
For brevity, we use the notation. Note that $Z_{\mathcal{A}^o_{1,1}}(\mathbf{s_1}; \bar{\mathbf{u}}_\mathbf{1}) = 1$, $Z_{\mathcal{A}^o_{1,j}}(\mathbf{s_1}; \bar{\mathbf{u}}_\mathbf{1}) = 0$ for $j \neq 1$ and $Z_{\mathcal{A}^o_{i,j}}(\mathbf{s_1}; \bar{\mathbf{u}}_\mathbf{1}) = Z_{\mathcal{A}^o_{j,i}}(\mathbf{s_1}; \bar{\mathbf{u}}_\mathbf{1})$ for $1 \leq i, j \leq k$ or $i, j = \infty$.

Lemma 4.27. *The functions $Z_{\mathcal{A}^o_{i,j}}(\mathbf{s_1}; \bar{\mathbf{u}}_\mathbf{1})$ fulfill the following strongly connected system of* *equations:*

$$
\begin{aligned}
Z_{\mathcal{A}^o_{i,j}}(\mathbf{s_1};\bar{\mathbf{u}}_\mathbf{1}) = &\sum_{l_1+l_2\le k} u_{l_1+l_2,1} Z_{\mathcal{A}^o_{i-1,l_1}}(\mathbf{s_1};\bar{\mathbf{u}}_\mathbf{1}) Z_{\mathcal{A}^o_{j-1,l_2}}(\mathbf{s_1};\bar{\mathbf{u}}_\mathbf{1}) \\
&+ s_1 u_{\infty,1}\left(\sum_{l_1+l_2>k} Z_{\mathcal{A}^o_{i-1,l_1}}(\mathbf{s_1};\bar{\mathbf{u}}_\mathbf{1}) Z_{\mathcal{A}^o_{j-1,l_2}}(\mathbf{s_1};\bar{\mathbf{u}}_\mathbf{1})\right) \\
&+ \sum_{l_1+l_2\le k+1} u_{l_1+l_2-1,1} Z_{\mathcal{A}^o_{i-1,l_1}}(\mathbf{s_1};\bar{\mathbf{u}}_\mathbf{1}) Z_{\mathcal{A}^o_{j-1,l_2}}(\mathbf{s_1};\bar{\mathbf{u}}_\mathbf{1}) \\
&+ s_1 u_{\infty,1}\left(\sum_{l_1+l_2>k+1} Z_{\mathcal{A}^o_{i-1,l_1}}(\mathbf{s_1};\bar{\mathbf{u}}_\mathbf{1}) Z_{\mathcal{A}^o_{j,l_2}}(\mathbf{s_1};\bar{\mathbf{u}}_\mathbf{1})\right), \ \forall\, 2 \le i \le j \le k+1.
\end{aligned}
\tag{4.13}
$$

Proof. A close look at the recursive description depicted in Figure 4.3 together with its symbolic translation (4.12) leads to this system of equations. Strong connectivity is given as every equation depends on $Z_{\mathcal{A}^o_{\infty,\infty}}(\mathbf{s_1};\bar{\mathbf{u}}_\mathbf{1})$ and the equation for $Z_{\mathcal{A}^o_{\infty,\infty}}(\mathbf{s_1};\bar{\mathbf{u}}_\mathbf{1})$ depends on all other variables. □

From the above system, we deduce generating functions $A^o_{i,j}(z,\mathbf{v})$, $\mathbf{v} = (v_1,\ldots,v_k,v_\infty)$, where the variable z counts all vertices while variables $v_i, i \in \{1,\ldots,k\}$, count vertices of degree i and v_∞ counts vertices of degree greater than k, by substituting as previously $s_\ell = z^\ell, u_{i,\ell} = v_i^\ell$. Then, we obtain generating functions $A^o(z,\mathbf{v})$ of oriented outer-edge rooted dissections of arbitrary root degrees by

$$A^o(z,\mathbf{v}) = \sum_{1\le i<j\le\infty} v_i v_j A^o_{i,j}(z,\mathbf{v}) + \sum_{i=1}^{k,\infty} v_i^2 A^o_{i,i}(z,\mathbf{v}).$$

By counting oriented outer-edge pointed dissections, we count every outer-edge pointed dissection twice except those whose automorphism group contain a reflection, because then the root-edge allows only one orientation. So we further have to determine the cycle index sum of symmetric outer-edge rooted dissections, where the automorphism group contains the identity and a reflection which fixes the root edge, to be able to count all outer-edge pointed dissections. Symmetric outer-edge pointed dissections $\mathcal{A}^s$ fulfil the decomposition given in Figure 4.4, which has the symbolic translation

$$\mathcal{A}^s = \mathcal{X} \times (1+\mathcal{A}^o)_{(2)} + \mathcal{X}^2 \times (1+\mathcal{A}^o)_{(2)} \times (1+\mathcal{A}^s) + \mathcal{X}^2 \times (1+\mathcal{A}^o)_{(2)} \times \mathcal{A}^s, \tag{4.14}$$

where $(1+\mathcal{A}^o)_{(2)}$ denotes the choice of two identical copies from $(1+\mathcal{A}^o)$.

Interpreting and refining the decomposition, we obtain the following system of equations for the cycle index sums $Z_{\mathcal{A}^s_{i,i}}(\mathbf{s_1};\bar{\mathbf{u}}_\mathbf{1})$ of reflective dissections, where $Z_{\mathcal{A}^{s+}_{i,i}}(\mathbf{s_1};\bar{\mathbf{u}}_\mathbf{1})$ represents the identity while $Z_{\mathcal{A}^{s-}_{i,i}}(\mathbf{s_1};\bar{\mathbf{u}}_\mathbf{1})$ represents the reflecting part. For shorter notation, we again use $\sum_{\ell=1}^{k,\infty} f_\ell = \sum_{\ell=1}^{k} f_\ell + f_\infty$. In $(\mathbf{s}^\mathbf{r}_\ell, \bar{\mathbf{u}}^\mathbf{r}_\ell)$ every member of the set of variables $(\mathbf{s}_\ell, \bar{\mathbf{u}}_\ell)$ is taken to the power r: $(s^r_\ell, u^r_{1,\ell}, \ldots, u^r_{\infty,\ell}; \ldots)$. Note that indices arising through the constructions below which are greater than k correspond to index ∞ and that for the summand with

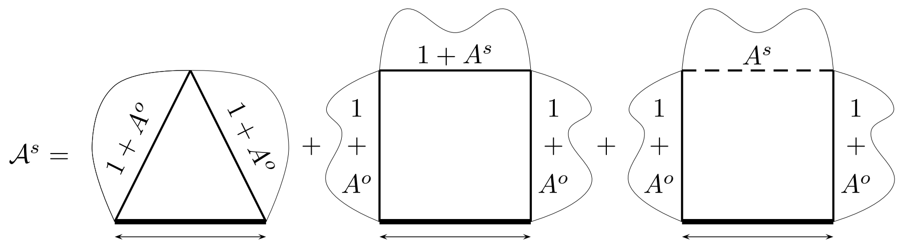

Figure 4.4: The decomposition of a symmetric outer-edge rooted dissections

index $i = \infty$, index $\infty - 1$ means index $\geq k$ (that is k and ∞).

$$\begin{aligned}
Z_{\mathcal{A}^{s+}_{i,i}}(\mathbf{s_1};\mathbf{\bar{u}_1}) = s_1 \sum_{l=1}^{k,\infty} \Big(Z_{\mathcal{A}^o_{i-1,l}}(\mathbf{s_1^2};\mathbf{\bar{u}_1^2}) u_{2l;1} \Big) \\
+ s_1^2 \left(\sum_{l=1}^{k,\infty} Z_{\mathcal{A}^o_{i-1,l}}(\mathbf{s_1^2};\mathbf{\bar{u}_1^2}) \cdot \left(\sum_{l=1}^{k,\infty} u^2_{l+i;1} Z_{\mathcal{A}^{s+}_{i,i}}(\mathbf{s_1};\mathbf{\bar{u}_1}) \right) \right) \\
+ s_1^2 \left(\sum_{l=1}^{k,\infty} Z_{\mathcal{A}^o_{i-1,l}}(\mathbf{s_1^2};\mathbf{\bar{u}_1^2}) \cdot \left(\sum_{i=2}^{k,\infty} z^2_{l+i-1;1} Z_{\mathcal{A}^{s+}_{i,i}}(\mathbf{s_1};\mathbf{\bar{u}_1}) \right) \right),
\end{aligned}$$

$$\begin{aligned}
Z_{\mathcal{A}^{s-}_{i,i}}(\mathbf{s_1};\mathbf{\bar{u}_1}) = s_1 \sum_{l=1}^{k,\infty} \Big(Z_{\mathcal{A}^o_{i-1,l}}(\mathbf{s_2};\mathbf{\bar{u}_2}) u_{2l;1} \Big) \\
+ s_2 \left(\sum_{l=1}^{k,\infty} Z_{\mathcal{A}^o_{i-1,l}}(\mathbf{s_2};\mathbf{\bar{u}_2}) \cdot \left(\sum_{l=1}^{k,\infty} u_{l+i;2} Z^{s-}_{\mathcal{A}_{i,i}}(\mathbf{s_1};\mathbf{\bar{u}_1}) \right) \right) \qquad (4.15) \\
+ s_2 \left(\sum_{l=1}^{k,\infty} Z_{\mathcal{A}^o_{i-1,l}}(\mathbf{s_2};\mathbf{\bar{u}_2}) \cdot \left(\sum_{i=2}^{k,\infty} z_{l+i-1;2} Z_{\mathcal{A}^{s-}_{i,i}}(\mathbf{s_1};\mathbf{\bar{u}_1}) \right) \right).
\end{aligned}$$

Hence, the cycle index sum of symmetric outer-edge pointed dissections is given by

$$Z_{\mathcal{A}^s_{i,i}}(\mathbf{s_1};\mathbf{\bar{u}_1}) = \frac{Z_{\mathcal{A}^{s+}_{i,i}}(\mathbf{s_1};\mathbf{\bar{u}_1}) + Z_{\mathcal{A}^{s-}_{i,i}}(\mathbf{s_1};\mathbf{\bar{u}_1})}{2}.$$

With the appropriate substitution, we obtain generating functions $A^s_i(z, \mathbf{v})$.
The cycle index sum $Z_{\mathcal{A}_{i,j}}$ of all outer edge rooted dissections with root degree (i, j) is be given by

$$\begin{aligned}
Z_{\mathcal{A}_{i,j}}(\mathbf{s_1}, \mathbf{\bar{u}_1}) &= \frac{Z_{\mathcal{A}^o_{i,j}}(\mathbf{s_1}, \mathbf{\bar{u}_1})}{2} \qquad \text{for } i \neq j, \\
Z_{\mathcal{A}_{i,j}}(\mathbf{s_1}, \mathbf{\bar{u}_1}) &= \frac{Z_{\mathcal{A}^o_{i,j}}(\mathbf{s_1}, \mathbf{\bar{u}_1}) + Z_{\mathcal{A}^{s-}_{i,i}}(\mathbf{s_1};\mathbf{\bar{u}_1})}{2} \qquad \text{for } i = j.
\end{aligned}$$

Let $\mathcal{B}'^o_i$ denote oriented vertex pointed dissections and $\mathcal{B}'^s_i$ denote reflective vertex-pointed dissections, where the pointed vertex has degree i in both cases. Note that oriented vertex pointed dissections are in a one-to-one correspondence with oriented outer-edge pointed dissections where one of the endpoints of the pointed edge has arbitrary degree,

$$Z_{\mathcal{B}_i'^o}(\mathbf{s_1};\bar{\mathbf{u}}_\mathbf{1}) = s_1 \sum_{j=1}^{k,\infty} u_{j,1} Z_{\mathcal{A}_{ij}^o}(\mathbf{s_1};\bar{\mathbf{u}}_\mathbf{1}).$$

Again, by orientation every pointed dissection is counted twice except the symmetric ones. Counting vertex-pointed symmetric dissections by outer-edge pointed symmetric dissections is not immediate, but possible with a decomposition similar to the previous ones. We obtain for the reflecting contribution

$$Z_{\mathcal{B}_i'^{s-}}(\mathbf{s_1};\bar{\mathbf{u}}_\mathbf{1}) = \begin{cases} s_1 \sum_{l=1}^{k+1} u_{2l-1;1} Z_{\mathcal{A}_{m,l}^o}(\mathbf{s_2};\bar{\mathbf{u}}_\mathbf{2}) & \text{for } i = 2m-1, \\ s_2 \left(\sum_{l=1}^{k+1} Z_{\mathcal{A}_{m,l}^o}(\mathbf{s_2};\bar{\mathbf{u}}_\mathbf{2}) \cdot \left(\sum_{j=1}^{k+1} Z_{\mathcal{A}_{j,j}^{s-}}(\mathbf{s_1};\bar{\mathbf{u}}_\mathbf{1}) u_{j+l;2} \right.\right. & \\ \left.\left. + \sum_{j=2}^{k+1} Z_{\mathcal{A}_{j,j}^{s-}}(\mathbf{s_1};\bar{\mathbf{u}}_\mathbf{1}) u_{j+l-1;2} \right)\right) & \text{for } i = 2m. \end{cases}$$

Finally, by above arguments and summing all root degrees, we obtain for the cycle index sum of pointed unlabelled 2-connected outerplanar graphs

$$Z_{\mathcal{B}'}(\mathbf{s_1},\bar{\mathbf{u}}_\mathbf{1}) = \sum_{i=1}^{k,\infty} u_{i,1} \frac{Z_{\mathcal{B}_i'^o}(\mathbf{s_1};\bar{\mathbf{u}}_\mathbf{1}) + Z_{\mathcal{B}_i'^{s-}}(\mathbf{s_1};\bar{\mathbf{u}}_\mathbf{1})}{2}, \tag{4.16}$$

and, again by substitution, we obtain a functional equation for the generating function $B'(z,\mathbf{v})$.

Analytic part

In a first step we analyze the core system (4.13) established in the previous part in terms of generating functions $\mathbf{A^o} = \mathbf{F}(z,\mathbf{A^o},\mathbf{v})$, $\mathbf{A^o} = (A_{i,j}^o)_{i,j\in\{1,\ldots,k,\infty\}}$ to obtain information on the singular behaviour. We obtain a distributional result on the derived family $\mathcal{B}'$ from there, which we can extend to the unrooted family $\mathcal{B}$ by analytic integration.
Note that

$$\sum_{i=1}^{k,\infty}\sum_{j=1}^{k,\infty} A_{i,j}^o(z,\mathbf{1}) = A^o(z)$$

is the ordinary generating function of oriented outer-edge rooted dissections with arbitrary root degree. In [5] it has been shown that $A^o(z)$ has singularity $\rho_B = 3-2\sqrt{2}$ and a local singular expansion of order $\frac{1}{2}$ around it: $A^o(x) = g^o(z) - h^o(z)\sqrt{1-\frac{z}{\rho_B}}$. As the system of equations (4.13) is strongly connected and positive, all functions $A_{i,j}^o(z,\mathbf{1})$ have the same singularity. The stability property (Lemma 1.18) leads to a local singular expansion of order $\frac{1}{2}$ for the generating functions $A_{i,j}^{o(k)}(z,v)$ near $(\rho_B(1),1))$, where only vertices of degree k

are counted:

$$\begin{aligned} A^{o(k)}_{i,j}(z,v) &= A^{o}_{i,j}(z,1,\ldots,1,v,1) \\ &= g_{ij}(z,v) - h_{ij}(z,v)\sqrt{1-\frac{z}{\rho(v)}}, \end{aligned}$$

with $\rho_B(1) = 3-2\sqrt{2}$. We further know that

$$\sum_{i=1}^{k,\infty} A^s_{i,i}(z,\mathbf{1}) = A^s(z)$$

has radius of convergence $\sqrt{\rho_B} > \rho_B$. After substitution in system (4.15), the functions $A^o_{i,j}(z,\mathbf{v})$ only appear in variables z^k with exponents $k \geq 2$ and by strong connectivity and positivity the same is true for every function $A^s_{i,i}(z,\mathbf{1})$, and thus, also

$$A^{s(k)}_{i,i}(z,v) = A^s_{i,i}(z,1\ldots,1,v,1)$$

has radius of convergence larger than $\rho_B(v)$ for v sufficiently close to 1. By equation (4.16), the generating function $B'^{(k)}(z,v) = B'(z,1,\ldots,1,v,1)$ is a linear composition of the above functions and hence has the same singular behaviour.

We are interested in the number of nodes of degree k in a member of $\mathcal{B}'$ chosen uniformly at random, X^k_n. The probability, that a random graph of size n has m vertices of degree k is given by $\frac{[z^n v^m]B'^{(k)}(z,v)}{[z^n]B'(z)}$. Thus, X^k_n is of the form

$$\mathbb{E}u^{X^k_n} = \frac{[z^n]B'^{(k)}(z,v)}{[z^n]B'^{(k)}(z,1)}.$$

The local singular expansion of $B'^{(k)}(z,v)$ of order $\frac{1}{2}$ leads to a central limit theorem for the derived family $\mathcal{B}'$ by means of Theorem A.5.

To obtain a result on the unrooted family, we will use integration. Therefore we isolate variable s_1 and count vertices of degree k, obtaining a trivariate generating function:

$$B'^{(k)}_i(s,z,v) = Z_{\mathcal{B}'_i}(s,1,\ldots,1,v,1;z^2,1,\ldots,1,v^2,1;z^3,1,\ldots,1,v^3,1;\ldots).$$

The functions $B'^{(k)}(s,z,v)$ of rooted 2-connected outerplanar graphs with arbitrary root degree are then given by

$$B'^{(k)}(s,z,v) = \sum_{i=1}^{k-1} B'^{(k)}_i(s,z,v) + vB'^{(k)}_k(s,z,v) + B'^{(k)}_\infty(s,z,v).$$

As indicated in Section 1, the generating function of unrooted unlabelled 2-connected outerplanar graphs is then given by $B(z,v) = B(z,z,v)$, where

$$B(s,z,v) = \int_0^s B'(t,z,v)\mathrm{d}t + B(0,z,v). \tag{4.17}$$

It is important to note that $B(0,z,v)$ has bigger radius of convergence than $B'(z,z,v)$ and thus has no influence on the asymptotic behaviour of $B(s,z,v)$, see also the remark

on the next page. The corresponding equations are very lengthy and can be found in the Appendix. With these equations, it can be shown directly that $B(0, z, v)$ is analytic at $\rho(v)$.
Integration of $B'(s, z, v)$ leads to a local singular expansion of order $\frac{3}{2}$

$$B(s, z, v) = g(s, z, v) - h(s, z, v) \left(1 - \frac{s}{\rho(z, v)}\right)^{\frac{3}{2}}.$$

It remains to prove that $B(z, z, v)$ has a local singular expansion of order $\frac{3}{2}$ around its singularity $\rho(v)$. The function $\rho(z, v)$ is analytic and can hence be represented by a power series $\rho(z, v) = \sum_{i \geq 0} a_i(v) z^i$ with $a_0(v) = \rho_B(v)$. Hence, if we set $s = z$, the singular term can be rewritten to

$$\left(1 - \frac{z}{\rho(z, v)}\right)^{\frac{3}{2}} = \kappa(z, v) \left(1 - \frac{z}{\tilde{\rho}_B(v)}\right)^{\frac{3}{2}},$$

where $\kappa(z, v)$ and $\tilde{\rho}_B(v)$ are analytic. Since $z = \rho(z, v)$ if and only if $z = \tilde{\rho}_B(v)$, it follows that $\tilde{\rho}_B(v) = \rho_B(v)$ and $B(z, v)$ has a local singular expansion of order $\frac{3}{2}$. From this expansion, the central limit theorem follows.
The expected value $\mathbb{E}X_n^k \sim \mathbb{E}X_n'^k \sim \mu_k n$ is asymptotically given by $\mu_k = -\frac{\rho_B'(1)}{\rho_B(1)}$. Since the singularity $\rho_B(v)$ is the same as in the labelled case, we can adopt the result from the labelled case [15, Chapter 9], where $\mu_k = 2(k-1)(\sqrt{2}-1)^k$. This matching of the singularities is due to the fact that the core system of equations in the labelled and unlabelled case are exactly the same and thus yield the same singularity. The other terms appearing in the unlabelled equations are analytic at $\rho_B(v)$, as seen before, and have no influence on the asymptotic result.
This result implies that the symmetries arising in 2-connected outerplanar graphs through unlabelling are very few and hence do not influence the asymptotic result. In the connected case, this is no longer true, as an exponential number of symmetries appears which alters the singularity of the generating functions (see the previous section). □

Remark. The step of analytic integration could be omitted by using the dissymmetry theorem on trees and the fact that the dual graph of a dissected polygon is a tree. One then has to set up functional equations for the generating functions of inner-edge and face rooted dissections, which can be decomposed combinatorially into oriented and symmetric outer-edge rooted dissections. In [5] this method is used for counting unlabelled outerplanar graphs. In the case of multivariate cycle index sums like we use to count vertex degrees, systems of equations become very large and it is hardly possible to prove that the singularity is of type $(1 - \frac{z}{\rho_B(v)})^{\frac{3}{2}}$. Still, to prove that the radius of convergence of $B(0, z, v)$ is large enough and for comparison, the corresponding systems of equations can be found in Appendix B.

4.3.2 Series-parallel graphs

Recall that series-parallel graphs are graphs obtained from series-parallel extensions of a forest. 2-connected series parallel graphs can be interpreted as series-parallel networks where the poles are connected by an additional edge. Still we will use a different method to count series-parallel blocks, namely a decompostion presented by Tutte in the 1960's [67]. This method also relies on networks, but provides a unique decomposition, while the mapping described above is not bijective. We proceed in a similar manner as in the previous section and obtain an analogous result for series-parallel graphs, also obtaining a central limit theorem on the derived family as a byproduct.

Theorem 4.28. *Let $\mathcal{B}$ be the family of random 2-connected unlabelled series-parallel graphs and $\mathcal{B}'$ be it's derived family. Then the random variable that counts the number of vertices of degree k in a randomly chosen member of $\mathcal{B}$, X_n^k, and the corresponding random variable $X_n'^k$ for a member of the derived family, satisfy a central limit theorem with expected value $\mathbb{E}X_n^k \sim \mathbb{E}X_n'^k \sim \mu_k n$ and variance $\mathbb{V}X_n^k \sim \mathbb{V}X_n'^k \sim \sigma_k^2 n$ where μ_k and σ_k are computable constants.*

Proof. Again, the proof is divided into a combinatorial part, aiming at setting up suitable systems of equations, and an analytic part where the system and the relating equations are analyzed.

Combinatorial part

For the sake of brevity, equations in this part are given in terms of ordinary generating functions. We denote by $(z^\ell, \mathbf{v}^\ell) = (z^\ell, v_1^\ell, \ldots, v_k^\ell, v_\infty^\ell)$.
The subfamily of 2-connected series-parallel graphs, which provides us with a core system of equations are series-parallel networks. A series-parallel network is obtained via series-parallel extensions of a single edge, thus it is a series-parallel graph with 2 distinguished nodes (the endpoints of the original edge), which we call the poles and which we denote by 0 and ∞, cf Figure 4.5.

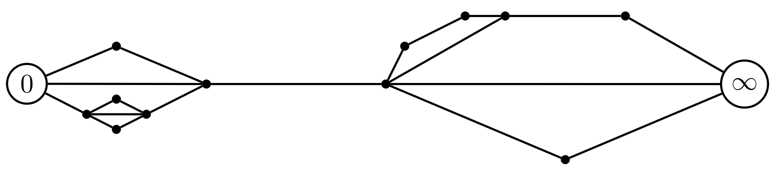

Figure 4.5: A series-parallel network, and more precicely a series network

Let us denote by $\mathcal{D}$ the set of series-parallel networks. We distinguish 2 types: Series networks $\mathcal{S}$, which have a series decomposition, and parallel networks $\mathcal{P}$, which have a parallel decomposition. This is equivalent to the fact that the first extension had been a subdivision or a doubling, respectively. A single edge is also considered a network, which is neither series nor parallel. Performing the decompositions we obtain the following system of equations in the symbolic language:

$$\begin{aligned} \mathcal{D} &= e + \mathcal{P} + \mathcal{S} \\ \mathcal{S} &= \mathcal{S} * \mathcal{X} * \mathcal{D} \\ \mathcal{P} &= e * \mathrm{Set}_{\geq 1}(\mathcal{S}) + \mathrm{Set}_{\geq 2}(\mathcal{S}) \end{aligned} \tag{4.18}$$

For the generating functions $D_{ij}(z, \mathbf{v})$, $S_{ij}(z, \mathbf{v})$ and $P_{ij}(z, \mathbf{v})$ of networks with pole degrees i and j we obtain the following (strongly connected) systems of equations for $1 < i < j$ (note that $D_{ij} = D_{ji}$), by translating the above system (4.18). This systems will be the core system of the proof, as we will see later. Recall that Z_i denotes the cycle index sum of the full symmetric group $\mathfrak{S}_i$ on i elements and the notation $Z_i(\mathbf{G}(\mathbf{z},\mathbf{v})$ denotes the substitution $s_i \leftarrow G(z^i, v_1^i, \ldots, v_k^i, v_\infty^i)$.

$$\begin{aligned}
D_{ij}(z,\mathbf{v}) &= S_{ij}(z,\mathbf{v}) + P_{ij}(z,\mathbf{v}) \\
S_{ij}(z,\mathbf{v}) &= z\left(\sum_{\ell_1=2}^{k,\infty}\sum_{\ell_2=1}^{k,\infty} v_{\ell_1+\ell_2} P_{i\ell_1}(z,\mathbf{v}) D_{\ell_2 j}(z,\mathbf{v})\right) \\
P_{ij}(z,\mathbf{v}) &= \sum_{\ell_1+2\ell_2+\ldots+i\ell_i=i}\left(\sum_{\substack{\sigma\in\mathfrak{S}_i:\\ \sum\sigma(i)\ell_{\sigma(i)}=j}}\prod_{r=1}^{i} Z_{l_r}\left(\mathbf{S}_{\mathbf{r},\sigma(\mathbf{r})}(\mathbf{z},\mathbf{v})\right)\right) \\
&\quad + \sum_{\ell_1+2\ell_2+\ldots+(i-1)\ell_{i-1}=i-1}\left(\sum_{\substack{\sigma\in\mathfrak{S}_i:\\ \sum\sigma(i)\ell_{\sigma(i)}=j}}\prod_{r=1}^{i} Z_{l_r}\left(\mathbf{S}_{\mathbf{r},\sigma(\mathbf{r})}(\mathbf{z},\mathbf{v})\right)\right) \\
&\quad - \sum_{i=1}^{k,\infty}\sum_{j=1}^{k,\infty} S_{ij}(z,\mathbf{v})
\end{aligned} \tag{4.19}$$

For pole-degree $i = 1$ we get:

$$\begin{aligned}
D_{11}(z,\mathbf{v}) &= 1 + S_{11}(z,\mathbf{v}), \\
D_{1j}(z,\mathbf{v}) &= S_{1j}(z,\mathbf{v}) && \text{for } j > 1 \\
S_{11}(z,\mathbf{v}) &= zv_2 + z^2\left(\sum_{\ell_1=1}^{k,\infty}\sum_{\ell_2=1}^{k,\infty} v_{\ell_1+1}v_{\ell_2+1}D_{\ell_1\ell_2}(z,\mathbf{v})\right), \\
S_{1j}(z,\mathbf{v}) &= z\left(\sum_{\ell=1}^{k,\infty} v_{\ell+1}D_{\ell j}(z,\mathbf{v})\right) && \text{for } j > 1.
\end{aligned}$$

Additionally, there are symmetric networks which are invariant under a reflection exchanging the poles 0 and ∞. Let us denote by $\bar{\mathcal{S}}, \bar{\mathcal{P}}$ and $\bar{\mathcal{D}}$ the families of symmetric series, parallel or general networks, respectively. Symbolically, they are given by

$$\begin{aligned}
\bar{\mathcal{D}} &= e + \bar{\mathcal{S}} + \bar{\mathcal{P}} \\
\bar{\mathcal{S}} &= \mathcal{D}_{(2)} * (\mathcal{X} + \mathcal{X}^2 * (e + \bar{\mathcal{P}})) \\
\bar{\mathcal{P}} &= e * \mathrm{Set}_{\geq 1}(\mathcal{S}_{(2)}, \bar{\mathcal{S}}) + \mathrm{Set}_{\geq 2}(\mathcal{S}_{(2)}, \bar{\mathcal{S}}),
\end{aligned} \tag{4.20}$$

where $\mathcal{D}_{(2)}$ and $\mathcal{S}_{(2)}$ denotes the choice of two identical copies from the families $\mathcal{D}$ and $\mathcal{S}$, respectively, and $\mathrm{Set}(\mathcal{S}_2, \bar{\mathcal{S}})$ denotes a set of pairs of arbitrary series networks together with a set of symmetric networks of odd size. This construction appears because a pair of arbitrary but identical series networks forms a symmetric parallel network when we connect one 0-pole with one ∞ pole and vice-versa. Hence, the set of symmetric networks is obtained by a set of pairs of arbitrary networks and symmetric networks. As sets of symmetric networks of even size are already contained in the set of pairs, we only have to count the sets of

uneven size.
For the generating functions $\bar{D}_i(x,\mathbf{v})$, $\bar{S}_i(z,\mathbf{v})$ and $\bar{P}_i(z,\mathbf{v})$, counting symmetric networks with pole-degree i, we obtain the following system:

$$\begin{aligned}
\bar{D}_i(z,\mathbf{v}) &= \bar{S}_i(z,\mathbf{v}) + \bar{P}_i(z,\mathbf{v})\\
\bar{S}_i(z,\mathbf{v}) &= \sum_{\ell=1}^{k,\infty}(D_{i\ell}(z^2,\mathbf{v}^2)\left(zv_{2\ell} + z^2v_{\ell+1}^2 + z\sum_{j=1}^{k,\infty} v_{\ell+j}^2\bar{P}_j(z,\mathbf{v})\right)\\
\bar{P}_i(z,\mathbf{v}) &= \sum_{\ell_1+2\ell_2+\ldots+i\ell_i=i}\left(\prod_{r=1}^{i} Z_{l_r}[s_{2s-1}\leftarrow \bar{S}_r(z^{2s-1},\mathbf{v}^{2s-1}), s_{2s}\leftarrow S_{rr}(z^{2s},\mathbf{v}^{2s}), s\geq 1]\right.\\
&\qquad\left. + \prod_{\substack{\ell_t=\ell_{\tilde{t}}\\ t\neq\tilde{t}}} Z_{\ell_t}\left(\mathbf{S}_{\ell_t\ell_{\tilde{t}}}(\mathbf{z^2},\mathbf{v^2})\right)\right)\\
&\quad + \sum_{\ell_1+2\ell_2+\ldots+(i-1)\ell_{i-1}=i-1}\left(\prod_{r=1}^{i} Z_{l_r}[s_{2s-1}\leftarrow \bar{S}_r(z^{2s-1}\mathbf{v}^{2s-1}), s_{2s}\leftarrow S_{rr}(z^{2s},\mathbf{v}^{2s}), s\geq 1]\right.\\
&\qquad\left. + \prod_{\substack{\ell_t=\ell_{\tilde{t}}\\ t\neq\tilde{t}}} Z_{\ell_t}\left(\mathbf{S}_{\ell_t\ell_{\tilde{t}}}(\mathbf{z^2},\mathbf{v^2})\right)\right) - \sum_{i=1}^{k,\infty}\bar{S}_i(z,\mathbf{v}),
\end{aligned}\tag{4.21}$$

and for pole degree $i=1$ we have

$$\begin{aligned}
\bar{D}_1(z,\mathbf{v}) &= 1 + \bar{S}_1(z,\mathbf{v}),\\
\bar{S}_1(z,\mathbf{v}) &= zv_2 + z^2\left(\sum_{\ell=1}^{k,\infty} v_{\ell+1}^2\bar{D}_i(z,\mathbf{v})\right).
\end{aligned}$$

For 2-connected graphs we use a decomposition into series, parallel and 3-connected components (which do not exist in our case as K_4 is exluded as a minor) which goes back to Tutte [67] and is described in detail in [10]. This decomposition leads to a coloured bipartite tree[1] with nodes colored $\mathcal{R}$ and nodes colored $\mathcal{M}$, where $\mathcal{R}$ denotes cyclic components and $\mathcal{M}$ denotes multi-edge components, both with at least 3 edges. Series-parallel graphs are then obtained by replacing the edges of those cycle- and multi-edge components by parallel and series networks, respectively. To obtain equations for those structures, we use the dissymmetry theorem of trees, described in Chapter 1[Theorem 1.1], which in the case of bipartite trees simplifies to

$$\mathcal{T} = \mathcal{T}^{\bullet} + \mathcal{T}^{\circ} - \mathcal{T}^{\bullet-\circ},$$

where $\mathcal{T}$ denotes unrooted trees, $\mathcal{T}^{\bullet}$ and $\mathcal{T}^{\circ}$ denote trees rooted at the two kinds of vertices, and $\mathcal{T}^{\bullet-\circ}$ denotes trees rooted at an edge. In our case the two types of nodes are the components $\mathcal{R}$ and $\mathcal{M}$, while $\mathcal{R}-\mathcal{M}$ denotes a edge rooted tree of the decomposition.

[1] A bipartite tree is a tree where a colouring of the vertices in two colours, such that adjacent vertices have different colours, is possible.

Furthermore we need to use Walsh series, described in Chapter 1.2, of cycle- and multi-edge structures, to be able to substitute into edges. The Walsh series of cycles is given in the examples of Walsh series on page 10 in Chapter 1, it can be rewritten to

$$\begin{aligned} W_{\mathcal{R}}(\mathbf{a_1},\mathbf{b_1},\mathbf{c_1}) = &\left[\sum_{k\geq 1}\frac{\phi(k)}{2k}\log\left(\frac{1}{1-a_kb_k}\right)-\frac{a_1b_1}{2}-\frac{a_1^2b_1^2}{4}-\frac{a_2b_2}{4}\right] \\ &+\left[\frac{a_2b_2}{1-a_2b_2}\left(\frac{a_2c_1^2}{4}+\frac{a_1^2b_2}{4}+\frac{a_1c_1}{2}\right)\right]. \end{aligned} \tag{4.22}$$

the Walsh series of a multiedge is given by

$$\begin{aligned} W_{\mathcal{M}} = &\frac{s_1^2}{2}\left[\exp\left(\sum_{k\geq 1}\frac{b_k}{k}\right)-1-b_1-\frac{b_1^2}{2}-\frac{b_2}{2}\right] \\ &+\frac{a_2}{2}\left[\exp\left(\sum_{k\geq 1}\frac{b_{2k}}{2k}+\frac{c_{2k-1}}{2k-1}\right)-1-\frac{b_2}{2}-c_1-\frac{c_1^2}{2}\right], \end{aligned} \tag{4.23}$$

where the first part represents the identity while the second term represents reflections.

Tutte's decomposition applies for the rooted class as well as for the unrooted class, with the only difference that cycle and multi-edge components are derived (at vertices) or not. In both cases we obtain systems of equations in terms of the core system on networks.
We proceed as in the outerplanar case and set up systems for the rooted class and use integration for unrooting. Nevertheless, systems of equations for unrooted unlabelled series-parallel graphs can be found in Appendix B, as they will be needed for the analytic part. Applying dissymmetry theorem to the bipartite tree we obtain

$$\mathcal{B}' = 1+\mathcal{B}_{\mathcal{R}'}+\mathcal{B}_{\mathcal{M}'}-\mathcal{B}_{\mathcal{R}\mathcal{M}}, \tag{4.24}$$

where $\mathcal{B}_{\mathcal{R}'} = \mathcal{R}'\circ_e(\mathcal{D}-\mathcal{S})$ denotes pointed rings whose edges are substituted by non-series networks $\mathcal{D}-\mathcal{S}$, $\mathcal{B}_{\mathcal{M}'} = \mathcal{M}'\circ_e(S)$ denotes pointed multiedges whose edges are substituted by series networks $\mathcal{S}$ and at most one edge, and $\mathcal{B}_{\mathcal{R}\mathcal{M}} = \mathcal{X}\times\mathcal{P}\times\mathcal{S}$ denotes an $R-M$ rooted tree, that is an intersection of a ring and multiedge component. The additional 1 counts the single edge which is also considered a 2-connected component.

Due to equation (4.24) the cycle index sum of rooted 2-connected series parallel graphs is given by

$$Z_{\mathcal{B}'}(\mathbf{s_1},\bar{\mathbf{u}}_\mathbf{1}) = s_1+Z_{\mathcal{B}_{\mathcal{R}'}}(\mathbf{s_1},\bar{\mathbf{u}}_\mathbf{1})+Z_{\mathcal{B}_{\mathcal{M}'}}(\mathbf{s_1},\bar{\mathbf{u}}_\mathbf{1})-Z_{\mathcal{B}_{\mathcal{R}\mathcal{M}}}(\mathbf{s_1},\bar{\mathbf{u}}_\mathbf{1}).$$

and for generating functions:

$$B'(z,\mathbf{v}) = z+B_{R'}(z,\mathbf{v})+B_{M'}(z,\mathbf{v})-B_{RM}(z,\mathbf{v}),$$

where $Z_{\mathcal{B}'}(\mathbf{s_1},\bar{\mathbf{u}}_\mathbf{1})$, $Z_{\mathcal{B}_{\mathcal{R}'}}(\mathbf{s_1},\bar{\mathbf{u}}_\mathbf{1})$, $Z_{\mathcal{B}_{\mathcal{M}'}}(\mathbf{s_1},\bar{\mathbf{u}}_\mathbf{1})$, $Z_{\mathcal{B}_{\mathcal{R}\mathcal{M}}}(\mathbf{s_1},\bar{\mathbf{u}}_\mathbf{1})$; $B'(z,\mathbf{v}), B_{R'}(z,\mathbf{v}), B_{M'}(z,\mathbf{v})$ and $B_{RM}(z,\mathbf{v})$ are the cycle index sums and generating functions of the structures appearing in equation (4.24). Again we denote by index i the degree of the root, and obtain a system

$$B_{R'}(z,\mathbf{v}) = z\sum_{i=1}^{k_\infty} v_i B_{R'_i}(z,\mathbf{v})$$

$$B_{R'_i}(z,\mathbf{v}) = \frac{1}{2}\left[\sum_{i_1+j_1=i}\sum_{\ell\geq 3}\left(\sum_{\substack{i_2,i_3,\dots,i_\ell\\ j_2,j_3,\dots,j_\ell}}\prod_{r=1}^{\ell-1}\left(v_{i_r+j_r}(D-S)_{i_r j_{r+1}}(z,\mathbf{v})(D-S)_{i_\ell j_1}(z,\mathbf{v})\right)\right)\right.$$

$$+\sum_{m\geq 1}\left(\sum_{\substack{i_1,i_2,\dots,i_m\\ j_1,j_2,\dots,j_m}}\prod_{r=1}^{m-1} v^2_{j_r+i_r} v^2_{i_m j_m}(D-S)_{\frac{i}{2}j_1}(z^2,\mathbf{v}^2)(D-S)_{i_r j_{r+1}}(z^2,\mathbf{v}^2)(\bar{D}-\bar{S})_{i_m}\right.$$

$$\left.\left. + z\sum_{\substack{i_1,i_2,\dots,i_m\\ j_1,j_2,\dots,j_{m+1}}}\prod_{r=1}^{m} v^2_{j_r+i_r} v^2_{2j_{m+1}}(D-S)_{\frac{i}{2}j_1}(z^2,\mathbf{v}^2)(D-S)_{i_r j_{r+1}}(z^2,\mathbf{v}^2)\right)\right]$$

$$B_{M'}(z,\mathbf{v}) = z\sum_{i=1}^{k_\infty} v_i B_{M'_i}(z,\mathbf{v})$$

$$B_{M'_i}(z,\mathbf{v}) = z\left(\sum_{\ell_1+2\ell_2+\dots+i\ell_i=i}\left(\sum_{\sigma\in\mathfrak{S}_i} v_{\sum_{r=1}^{i}\sigma(i)\ell_{\sigma(i)}}\prod_{r=1}^{i}\mathfrak{S}_{l_r}[S_{r,\sigma(r)}(z,\mathbf{v})]\right)\right.$$

$$+\sum_{\ell_1+2\ell_2+\dots+(i-1)\ell_{i-1}=i-1}\left(\sum_{\sigma\in\mathfrak{S}_i} v_{1+\sum_{r=1}^{i}\sigma(i)\ell_{\sigma(i)}}\prod_{r=1}^{i}\mathfrak{S}_{l_r}[S_{r,\sigma(r)}(z,\mathbf{v})]\right)$$

$$\left.-\sum_{j=1}^{k,\infty} S_{ij}(z,\mathbf{v})\right)$$

$$B_{RM}(z,\mathbf{v}) = \frac{1}{2}z\left(\sum_{\ell_1=1}^{k,\infty}\sum_{\ell_2=1}^{k,\infty}\sum_{\ell_3=1}^{k,\infty}\sum_{\ell_4=1}^{k,\infty} v_{\ell 1+\ell 2}v_{\ell_3+\ell_4}S_{\ell_1\ell_3}(z,\mathbf{v})P_{\ell_2\ell_4}(z,\mathbf{v})\right)$$
$$+\frac{1}{2}z\left(\sum_{\ell_1=1}^{k,\infty}\sum_{\ell_2=1}^{k,\infty} v^2_{\ell_1+\ell_2}\bar{S}_{\ell_1}(z,\mathbf{v})\bar{P}_{\ell_2}(z,\mathbf{v})\right)$$

Analytic part

We can express the core system of equations for networks (4.19) in terms of just one network type $\mathcal{D},\mathcal{S}$ or $\mathcal{P}$. Then we can proceed as in the case of outerplanar graphs and note that $\sum_{i=1}^{k,\infty}\sum_{j=1}^{k,\infty} D_{ij}(z,\mathbf{1}) = D(z)$ is the ordinary generating function of unlabelled series-parallel networks. The same applies for $\sum_i\sum_j S_{ij}(z,\mathbf{1}) = S(z)$ and $\sum_i\sum_j P_{ij}(z,\mathbf{1}) = P(z)$.
We know from [16] that $D(z), S(z)$ and $P(z)$ have a common singularity $0<\rho_B<1$ and a square-root singular expansion around it. Thus, again by the Drmota-Lalley-Woods Theorem 1.12, all functions $D_{ij}(z,v)$, specialized to count only vertices of degree k, have

square root singular expansions

$$D_{ij}(z,v) = D_{ij}(z,1,\dots,1,v,1) = g_{ij}(z,v) - h_{ij}(z,v)\sqrt{1-\frac{z}{\rho(v)}},$$

while $\bar{D}_i$ has bigger radius of convergence. $B'(z,v)$ fulfills functional equations in terms of D_{ij} and $\bar{D}_i$ for $i,j \in \{1,\dots,k,\infty\}$, and thus has a square root expansion around the same singularity. Again, the random variable X_n^k counting the number of vertices of degree k in a random unlabelled rooted 2-connected series-parallel graph of size n is of the form:

$$\mathbb{E}u^{X_n'^k} = \frac{[z^n]B'^{(k)}(z,v)}{[z^n]B'(z,1)},$$

and thus Theorem A.5 applies and leads us to a central limit theorem for the derived family $\mathcal{B}'$ of unlabelled series-parallel graphs.

To obtain the result for the unrooted class, we will again use integration. We have to isolate variable s_1 and obtain trivariate generating functions $B'(s,z,v)$. Then

$$B(s,z,v) = \int_0^s B'(t,z,v)\mathrm{d}t + B(0,z,v)$$

and $B(z,v) = B(z,z,v)$.
We need to check that $B(0,z,v)$ has bigger radius of convergence than $B'(z,z,v)$, which we do directly by using the equations given in Appendix B. Then Integration of $B'(s,z,v)$ leads to a local singular expansion of order $\frac{3}{2}$ around $(B(\rho_B,1),\rho_B(1),1)$. By applying the same arguments as in the case of outerplanar graphs, this expansion translates to a expansion around $(\rho_B(1),1)$ at $s=z$. Thus, we obtain a central limit theorem for the class of unrooted unlabelled 2-connected series-parallel graphs again by Theorem A.5.

□

Remark. In contrary to the outerplanar case, in the series-parallel case, we cannot be sure that the expected value $\mathbb{E}X_n^k$ will be the same as in the labelled case, where it is asymptotically given by $ck^{-\frac{3}{2}}q^k$, with constants c,q. The reason for that is the different core system of equations, where cycle index sums of symmetric groups appear in contrary to the labelled system. Giving an explicit representation of μ_k would require numerical calculation on the singularity $\rho_B(v)$ of the core system (4.19), which could be done with e.g. Maple, but will be very involved.

4.4 Cycle rooting

As indicated in the last remark of the previous section, obtaining numerical values for mean and variance of the degree distribution in the unlabelled case is a difficult question. In the labelled case, values are quite easily obtained by computing the probability d_k that the root of a derived object has degree k. Then d_k immediately gives the probability that a random vertex has degree k in a graph of size n. This is due to the fact that, in the labelled case,

$$A'(z) = \frac{\mathrm{d}}{\mathrm{d}z}A(z),$$

where $A(z)$ is the generating function of a family and $A'(z)$ is the generating function of the derived family. Thus the expected number of nodes of degree k is given by $d_k n$.

In the unlabelled case, this is no longer true, as (cf. Chapter 1)

$$Z_{\mathcal{A}'}(\mathbf{s_1}) = \frac{\partial}{\partial s_1} Z_{\mathcal{A}}(\mathbf{s_1}).$$

We now introduce a new way of rooting unlabelled objects, which was presented by Bodirsky *et. al.* in [6], which is cycle pointing. We denote by $\mathcal{G}^\circ$ the cycle pointed class of the class $\mathcal{G}$. A cycle pointed object $G^\circ \in \mathcal{G}^\circ$ is an representant $G \in \mathcal{G}$ together with a cycle of vertices c, such that there exists at least one automorphism of G which contains the cycle c. The reason to use this technique is the following property.

Lemma 4.29. *Let $G(z) = \sum g_{nm} z^n$ be the generating function of unlabelled graphs of a family $\mathcal{G}$ and $G^\circ(z) = \sum g^\circ_{nm} z^n$ be the generating function of the cycle pointed class. Then $g^\circ_n = n g_n$, that is*

$$G^\circ(z) = z \frac{\partial}{\partial z} G(z)$$

The proof of this Lemma can be found in [6]. As a consequence, to obtain information on the expected degree of a random vertex in a random graph of size n, we can consider a vertex of the root cycle of the cycle pointed family (as there is a symmetry containing this cycle, all vertices in the cycle have the same shape). Then this vertex represents a random node.

Let $Z_{\mathcal{A}}(s_1, s_2, \ldots)$ be the cycle index sum of a family $\mathcal{A}$. Then the cycle index sum of the cycle-rooted family is given by

$$Z_{\mathcal{A}^\circ}(s_1, t_1, s_2, t_2 \ldots) = \sum_{\ell \geq 1} \ell t_\ell \frac{\partial}{\partial s_\ell} Z_{\mathcal{A}}(s_1, s_2, \ldots) \tag{4.25}$$

We can set up equations for the generating function $G^\circ(z, w) = \sum_{n,m} g_{nm} z^n w^m$, where the coefficient g_{nm} counts the number of graphs of size n whose root has degree m, to extract the probability that the root has degree k by

$$d_k = \frac{g_{nm}}{g_n}.$$

In Appendix B.3 the equations for the generating functions $B^\circ(z, w)$ and $C^\circ(z, w)$ of series-parallel graphs are listed. We have not extracted numerical values for the expected values in Theorem 4.24 and Theorem 4.28, but these equations could be used for this purpose due to the above theory.

APPENDIX A

Fundamentals

In this part, we will describe in detail the toolbox of analytic combinatorics and singularity analysis on multivariate generating functions, which we just indicated in the introduction and which is applied to its full extent only in the last chapter on subcritical graph classes. All theorems together with their proofs can be found in [15][Chapter 2].

First of all, we state a result, providing information on the singular behaviour of a function determined by another function with known singularity.

Lemma A.1. *Suppose that $f(z,v)$ has a squareroot singular expansion of the form*

$$f(z,v) = g(z,v) - h(z,v)\sqrt{1 - \frac{z}{\rho(v)}}$$

and that $H(z,v,y)$ is a function that is analytic at $(\rho(1), 1, f(\rho(1),1))$ such that

$$H_y(\rho(1), 1, f(\rho(1),1)) \neq 0.$$

Then

$$f_H(z,v) = H(z,v,f(z,v))$$

has the same kind of singular expansion, that is

$$f_H(z,v) = \tilde{g}(z,v) - \tilde{h}(z,v)\sqrt{1 - \frac{z}{\rho(v}}$$

for certain analytic functions $\tilde{g}(z,v)$ and $\tilde{h}(z,v)$.
If $\rho(v) \neq 0$ and $f(z,v)$ has an analytic continuation to the region $|z| \leq |\rho(v)| + \epsilon, \arg(z/\rho(v) - 1 \neq 0$ for some $\epsilon > 0$, and if $H(z,v,y)$ is also analytic for $|z| < \rho(1) + \epsilon, |v| < 1 + \epsilon$ and $|y| < f(\rho(1),1) + \epsilon$, then $f(z,v)$ and $f_H(z,v)$ have power series expansions

$$f(z,v) = \sum_{n\geq 0} a_n(v) z^n \text{ and } f_H(z,v) = \sum_{n \geq 0} b_n(v) z^n,$$

where $a_n(v)$ and $b_n(v)$ satisfy

$$\lim_{n\to\infty} \frac{b_n(v)}{a_n(v)} = H_y(\rho(v), v, f(\rho(v), v)).$$

Lemma A.2. *Let $A(z, \mathbf{v})$ be a generating function with a square root singular expansion around a dominant positive singularity $\rho(\mathbf{v})$. Then the derivative and the integral have local singular expansions of the form*

$$A_z(z, \mathbf{v}) = \frac{g_2(z, \mathbf{v})}{\sqrt{1 - \frac{z}{\rho(\mathbf{v})}}} + h_2(z, \mathbf{v}) \tag{A.1}$$

and

$$\int_0^z A(t, \mathbf{v})\mathrm{d}t = g_3(z, \mathbf{v}) + h_3(z, \mathbf{v}) \left(1 - \frac{z}{\rho(\mathbf{v})}\right)^{\frac{3}{2}}, \tag{A.2}$$

where $g_2(z, \mathbf{v}), h_2(z, \mathbf{v}), g_3(z, \mathbf{v})$ and $h_3(z, \mathbf{v})$ are analytic near $\rho(\mathbf{v})$ and $\mathbf{v} = \mathbf{1}$.

Systems of equations

We consider a system of equations $\mathbf{y} = \mathbf{F}(z, \mathbf{y}, \mathbf{v})$:

$$\begin{aligned} y_1 &= F_1(z, y_1, \ldots, y_N, \mathbf{v}) \\ \vdots \quad & \quad \vdots \\ y_N &= F_N(z, y_1, \ldots, y_N, \mathbf{v}) \end{aligned} \tag{A.3}$$

Recall that the system is strongly connected if the dependancy graph is strongly connected, cf Definition 1.11, and that we denote by $\mathbf{F_y}$ the Jacobian of the system and by $\mathbf{I}$ the $N \times N$-identity matrix. We have the following refinement of the Drmota-Lalley-Woods theorem (Theorem 1.12).

Theorem A.3. *Let $\mathbf{y} = \mathbf{F}(z, \mathbf{y}, \mathbf{v})$ be a nonlinear system of functional equations which is strongly connected, has only nonnegative Taylor coefficients and which is analytic around $z = 0$, $\mathbf{y} = \mathbf{0}$ and $\mathbf{v} = \mathbf{0}$. Further assume that $\mathbf{F}(0, \mathbf{y}, \mathbf{v}) = \mathbf{0}$, $\mathbf{F}(z, \mathbf{0}, \mathbf{v}) \neq 0$ and $\mathbf{F}_z(z, \mathbf{y}, \mathbf{v}) \neq 0$ and that the region of convergence of $\mathbf{F}$ is large enough such that there exists a complex neighbourhood V of $\mathbf{v} = \mathbf{1}$ where the system*

$$\begin{aligned} \mathbf{y} &= \mathbf{F}(z, \mathbf{y}, \mathbf{v}) \\ 0 &= \det(\mathbf{I} - \mathbf{F_y}(z, \mathbf{y}, \mathbf{v})) \end{aligned}$$

has solutions $z = \rho(\mathbf{v})$ and $\mathbf{y} = \mathbf{y}_0(\mathbf{v})$ which are real, positive and minimal for positive real $\mathbf{v} \in U$. Let $\mathbf{y} = \mathbf{y}(z, \mathbf{v})$ denote the analytic solutions of the system $\mathbf{y} = \mathbf{F}(z, \mathbf{y}, \mathbf{v})$ with $\mathbf{y}(0, \mathbf{v}) = \mathbf{0}$. Then there exists an $\epsilon > 0$ such that all components $y_j(z, v), j = 1 \ldots, N$ have a square root singular expansion

$$y_j(z, \mathbf{v}) = g_j(z, \mathbf{v}) - h_j(z, \mathbf{v})\sqrt{1 - \frac{z}{\rho(\mathbf{v})}}, \quad \text{for } i = 1, \ldots N,$$

for $\mathbf{v} \in V$, $|z - \rho(\mathbf{v})| < \epsilon$ and $|\arg(z - \rho(\mathbf{v}))| \neq 0$, with analytic functions $g_j(z, \mathbf{v}) \neq 0$ and $h_j(z, \mathbf{v}) \neq 0$ with $g_j(\rho(\mathbf{v}), \mathbf{v}) = y_j(\rho(\mathbf{v}), \mathbf{v}) = (\mathbf{y}_0)_j$. Furthermore, if $[z^n]y_j(z, \mathbf{1}) > 0$ for $1 \leq j \leq N$ for sufficiently large $n \geq n_0$ then $z_0 = \rho(\mathbf{1})$ is the only singularity on the radius of convergence $|z| = z_0$ and all components $\mathbf{y}$ can be analytically continued to the region $D = \{z \in \mathbb{C} \big| |z| < |\rho(\mathbf{v})| + \epsilon, 1 - \frac{z}{\rho(\mathbf{v})} \notin \mathbb{R}^-\}$ for $\mathbf{v}$ in some neighbourhood of $\mathbf{1}$.

Lemma 1.14 can be refined to

Lemma A.4. *Let* $\mathbf{y}(z,\mathbf{v}) = (y_1(z.\mathbf{v}),\dots,y_N(z,\mathbf{v}))$ *be the solution of the system of equations* (A.3) *and assume that all assumptions of Theorem A.3 are satisfied. Suppose that* $G(z,\mathbf{y},\mathbf{v})$ *is a power series such that the point* $(z_0,\mathbf{y}_0(z_0,\mathbf{1}),\mathbf{1})$ *is contained in the interior of the region of convergence of* $G(z,\mathbf{y},\mathbf{v})$ *and that* $G_{\mathbf{y}}(z_0,\mathbf{y}_0(z_0,\mathbf{1}),\mathbf{1}) \neq 0$.
Then $y_G(z,\mathbf{v}) := G(z,\mathbf{y}(z,\mathbf{v}),\mathbf{v})$ *has a representation of the form*

$$y_G(z,\mathbf{v}) = g(z,\mathbf{v}) - h(z,\mathbf{v})\sqrt{1-\frac{z}{\rho(\mathbf{v})}}$$

for $\mathbf{v}\in V$ *and* $|z-\rho(\mathbf{v})|<\epsilon$, *where* $g(z,\mathbf{v})\neq 0$ *and* $h(z,\mathbf{v})\neq 0$ *are analytic functions. Moreover,* $G(z,\mathbf{y}(z,\mathbf{v}),\mathbf{v})$ *is analytic in* $(z,\mathbf{v})$ *for* $\mathbf{v}\in V$ *and* $|z-\rho(\mathbf{v})|\geq\epsilon$, *but* $|z|\leq |\rho(\mathbf{v})|+\delta$.

Multivariate central limit laws

We present an extension of Theorem 1.20 to solutions of systems of equations and multiple variables.

Theorem A.5. *Suppose that* $\mathbf{X}_n$ *is a sequence of* k*-dimensional random vectors which are given by*

$$\mathbb{E}(\mathbf{v}^{\mathbf{X}_n}) = \frac{[z^n]y_G(z,\mathbf{v})}{[z^n]y_G(z,\mathbf{1})},$$

where $y_G(z,\mathbf{v}) = G(z,\mathbf{y}(z,\mathbf{v}),\mathbf{v})$ *is a function given as in Lemma A.4, with* $\mathbf{y}(z,\mathbf{v})$ *being the solution of a system of equations of the form* (A.3)*, with* $\mathbf{F}$ *satisfying all preliminaries of Theorem A.3. Then* $\mathbf{X}_n$ *satisfies a central limit theorem of the form*

$$\frac{1}{\sqrt{n}}(\mathbf{X}_n - \mathbb{E}(\mathbf{X}_n)) \xrightarrow{d} \mathcal{N}(\mathbf{0},\mathbf{\Sigma}),$$

with

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

where

$$\boldsymbol{\mu} = \frac{1}{\rho(\mathbf{1})}\frac{\mathbf{b}^T\mathbf{F}_{\mathbf{v}}(z,\mathbf{y}_0,\mathbf{1})}{\mathbf{b}^T\mathbf{F}_z(z,\mathbf{y}_0,\mathbf{1})}, \tag{A.4}$$

where $\mathbf{b}$ *is (up to scaling) the unique positive left eigenvector of the Jacobian* $\mathbf{F}_{\mathbf{y}}$. Σ *is a positive semidefinite matrix computed with the help of second derivatives:*

$$\mathbf{\Sigma} = -\frac{\rho_{\mathbf{vv}}(\mathbf{1})}{\rho(\mathbf{1})} + \boldsymbol{\mu}\boldsymbol{\mu}^T + \mathrm{diag}(\boldsymbol{\mu}), \tag{A.5}$$

where $\rho_{\mathbf{vv}}$ *denotes* $(\rho_{v_iv_j})_{1\leq i,j\leq k}$ *for* $\mathbf{v} = (v_1,\dots,v_k)$.

In many applications $\mathbf{b}$ appears to be $\mathbf{b} = (1,\dots,1)^T$, which is due to the special structure of combinatorial systems of equations. The origin of the above theorems, as we explained in Chapter 1, is given by the singular expansions of the generating functions and mean and variance are determined by their singularities. Therefore, as we state in the following theorem, power series in multiple variables which have singular expansions of order $\alpha\notin\mathbb{N}$ with a singular term $(1-\frac{z}{\rho(\mathbf{v})})$, fulfil a central limit theorem with the same mean and variance independant of α.

Theorem A.6. *Suppose that a sequence of k-dimensional random vectors $\mathbf{X}_n$ satisfies*

$$\mathbb{E}(\mathbf{X}_n) = \frac{[z^n]y(z,\mathbf{v})}{[z^n]y(z,\mathbf{1})},$$

where $y(z,\mathbf{v})$ is a function which has a local singular representation of the form

$$y(z,\mathbf{v}) = g(z,\mathbf{v}) - h(z,\mathbf{v})\left(1 - \frac{z}{\rho(\mathbf{v})}\right)^{\alpha},$$

for some real $\alpha \notin \mathbb{N}$ and functions $g(z,\mathbf{v}) \neq 0, h(z,\mathbf{v}) \neq 0$ and $\rho(\mathbf{v}) \neq 0$ which are analytic around $z = z_0 > 0$ and $\mathbf{v} = \mathbf{1}$. Suppose also that $z = \rho(\mathbf{v})$ is the only singularity of $y(z,\mathbf{v})$ on the disc $|z| \leq |\rho(\mathbf{v})|$, if $\mathbf{v}$ is sufficiently close to $\mathbf{1}$ and that there exists an analytic continuation of $y(z,\mathbf{v})$ to the region $|z| < |\rho(\mathbf{v})| + \delta, |\arg(z - \rho(\mathbf{v})| > \epsilon$ for some $\delta > 0$ and $\epsilon > 0$. Then $\mathbf{X}_n$ satisfies a central limit theorem

$$\frac{1}{\sqrt{n}}(\mathbf{X}_n - \mathbb{E}(\mathbf{X}_n)) \xrightarrow{d} \mathcal{N}(\mathbf{0},\mathbf{\Sigma})$$

with

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

where

$$\boldsymbol{\mu} = -\frac{\rho_{\mathbf{v}}(\mathbf{1})}{\rho(\mathbf{1})}$$

with $\rho_{\mathbf{v}} = (\rho_{v_i})_{1\leq i\leq k}$ for $\mathbf{v} = (v_1, \ldots, v_k)$ and

$$\mathbf{\Sigma} = -\frac{\rho_{\mathbf{vv}}(\mathbf{1})}{\rho(\mathbf{1})} + \boldsymbol{\mu}\boldsymbol{\mu}^T + \mathrm{diag}(\boldsymbol{\mu}).$$

Furthermore there exist positive constants c_1, c_2, c_3 such that

$$\mathbb{P}(||\mathbf{X}_n - \mathbb{E}(\mathbf{X}_n)|| \geq \epsilon\sqrt{n}) \leq c_1 e^{-c_2\epsilon^2}$$

uniformly for $\epsilon \leq c_3\sqrt{n}$.

The above generalizations allow us to deduce that a rooted family and its corresponding unrooted family fulfill central limit theorems with the same asymptotic mean and variance, as we do in Chapter 4.

APPENDIX B

Systems of equations

B.1 Equations to count unrooted 2-connected outerplanar graphs

As mentioned in Remark 1 in Section 4.3.1 the dual graph of a dissection of a polygon is a tree (cf Figure B.1). Vertex degrees are preserved by this duality, as the degree of a vertex in the dissection is equivalent to the distance between the two outer dual vertices neighbouring this vertex.

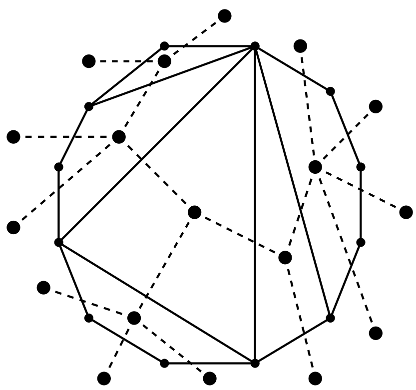

Figure B.1: The dual tree of a dissection

We use the dissymmetry theorem (Theorem 1.1) to set up a system of equations for the multivariate cycle index sums $Z_{\mathcal{D}}(\mathbf{s_1}, \bar{\mathbf{u}}_\mathbf{1})$ of unrooted unlabelled dissections, where the variables $\mathbf{s_1} = (s_1, s_2, \ldots)$ count cycles of vertices and the variables $\bar{\mathbf{u}}_\mathbf{1} = (u_{i,1}, u_{i,2}, \ldots)_{i \in \{1,\ldots,k,\infty\}}$ count cycles of vertices of degree i for $i = 1, \ldots, k$ and vertices of degree greater than k for

$i = \infty$. Translating the dissymetry theorem, we obtain

$$\begin{aligned}\mathcal{T} &= \mathcal{D}\\ \mathcal{T}_{\circ\to\circ} &= 2\mathcal{D}^o + 2\mathcal{D}^i - 2\mathcal{D}^{i(s)} + \mathcal{D}^{f(s)}\\ \mathcal{T}_{\circ} &= \mathcal{D}^o + \mathcal{D}^f\\ \mathcal{T}_{\circ-\circ} &= \mathcal{D}^o + \mathcal{D}^i,\end{aligned}$$

where $\mathcal{D}$ denotes the family of unrooted dissections, $\mathcal{D}^o$ dissections rooted at an outer edge, $\mathcal{D}^i$ dissections rooted at an inner edge, $\mathcal{D}^{i(s)}$ dissections rooted at a symmetry edge, $\mathcal{D}^f$ dissections rooted at a face, $\mathcal{D}^{f(s)}$ rooted at a face containing a symmetry edge. Thus we obtain for $Z_{\mathcal{D}}$:

$$Z_{\mathcal{D}} = Z_{\mathcal{D}^f} - Z_{\mathcal{D}^{f(s)}} - Z_{\mathcal{D}^i} + 2Z_{\mathcal{D}^{i(s)}}, \tag{B.1}$$

In Section 4.3.1 we have set up systems of equations for oriented (4.13) and symmetric (4.15) outer-edge rooted dissections. We can use them to build systems for all other classes needed in the above equation.

We use the same notation as in previous chapters and write $Z_{\mathcal{G}}(\mathbf{s_1}, \bar{\mathbf{u}}_\mathbf{1})$ for the cycle index sum of a structure $\mathcal{G}$ with $(\mathbf{s_1}, \bar{\mathbf{u}}_\mathbf{1})$ being the set of variables $(s_1, u_{1,1}, \ldots, u_{k,1}, u_{\infty,1}; s_2 \ldots)$ and

$$(\mathbf{s}_\ell, \bar{\mathbf{u}}_\ell) = (s_l, u_{1,\ell}, \ldots, u_{k,\ell}, u_{\infty,\ell}; s_{2\ell}, u_{1,2\ell}, \ldots, u_{k,2\ell}, u_{\infty,2\ell}; \ldots)$$

for some $\ell \geq 2$.

Inner edge rooted dissections: Inner edge rooted dissections follow a decomposition into two outeredge rooted dissections, glued together at their root edge. Taking into account all symmetries, we obtain four classes of inner-edge rooted dissections, which we denote by $\mathcal{D}^i_1, \mathcal{D}^i_2, \mathcal{D}^i_3$ and $\mathcal{D}^i_4$(cf. Figure B.2):

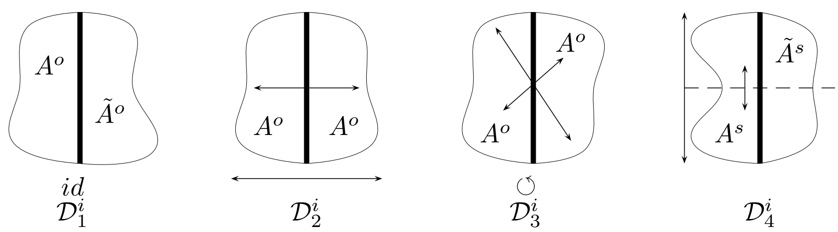

Figure B.2: Decomposition of inner edge rooted dissections

This decomposition translates to the following system in the language of cycle index sums:

$$Z_{\mathcal{D}_1^i}(\mathbf{s_1},\bar{\mathbf{u}}_\mathbf{1}) = s_1^2 \sum_{i_1=2}^{k,\infty}\sum_{j_1=2}^{k,\infty}\sum_{i_2=2}^{k,\infty}\sum_{j_2=2}^{k,\infty} u_{i_1+i_2-1;1}u_{j_1+j_2-1;1}Z_{\mathcal{A}^o_{i_1,j_1}}(\mathbf{s_1},\bar{\mathbf{u}}_\mathbf{1})Z_{\mathcal{A}^o_{i_2,j_2}}(\mathbf{s_1},\bar{\mathbf{u}}_\mathbf{1}),$$

$$Z_{\mathcal{D}_2^i}(\mathbf{s_1},\bar{\mathbf{u}}_\mathbf{1}) = s_1^2 \sum_{i=2}^{k,\infty}\sum_{j=2}^{k,\infty} u_{2i-1;1}u_{2j-1;1}Z_{\mathcal{A}^o_{i,j}}(\mathbf{s_2};\bar{\mathbf{u}}_\mathbf{2}),$$

$$Z_{\mathcal{D}_3^i}(\mathbf{s_1},\bar{\mathbf{u}}_\mathbf{1}) = s_2 \sum_{i=2}^{k,\infty}\sum_{j=2}^{k,\infty} u_{i+j-1;2}Z_{\mathcal{A}^o_{i,j}}(\mathbf{s_2};\bar{\mathbf{u}}_\mathbf{2}),$$

$$Z_{\mathcal{D}_4^i}(\mathbf{s_1},\bar{\mathbf{u}}_\mathbf{1}) = s_2 \sum_{i_1=2}^{k,\infty}\sum_{i_2=2}^{k,\infty} u_{i_1+i_2-1;2}Z_{\mathcal{A}^{s-}_{i_1,i_1}}(\mathbf{s_1};\bar{\mathbf{u}}_\mathbf{1})Z_{\mathcal{A}^{s-}_{i_2,i_2}}(\mathbf{s_1};\bar{\mathbf{u}}_\mathbf{1}),$$

$$Z_{\mathcal{D}^i}(\mathbf{s_1},\bar{\mathbf{u}}_\mathbf{1}) = \frac{Z_{\mathcal{D}_1^i}(\mathbf{s_1},\bar{\mathbf{u}}_\mathbf{1}) + Z_{\mathcal{D}_2^i}(\mathbf{s_1},\bar{\mathbf{u}}_\mathbf{1}) + Z_{\mathcal{D}_3^i}(\mathbf{s_1},\bar{\mathbf{u}}_\mathbf{1}) + Z_{\mathcal{D}_4^i}(\mathbf{s_1},\bar{\mathbf{u}}_\mathbf{1})}{4}.$$

Symmetric inner-edge rooted dissections consist of $\mathcal{D}_2^i, \mathcal{D}_3^i$ and $\mathcal{D}_4^i$ only.

Face rooted dissections: Let $l \geq 3$ be the size of the root face. Then, face rooted dissections follow a decomposition into outer-edge rooted dissections, as outer-edge rooted dissections are attached to the $l \geq 3$ edges of the root-face. Cycle index sums for face rooted dissections fulfill the following system of equations, where $Z_{\mathcal{F}^o}(\mathbf{s_1},\bar{\mathbf{u}}_\mathbf{1})$ denotes oriented face rooted dissections (only cyclic permutations are allowed here), while $Z_{\mathcal{F}_1^m}(\mathbf{s_1},\bar{\mathbf{u}}_\mathbf{1})$,$Z_{\mathcal{F}_2^m}(\mathbf{s_1},\bar{\mathbf{u}}_\mathbf{1})$ and finally $Z_{\mathcal{F}^s}(\mathbf{s_1},\bar{\mathbf{u}}_\mathbf{1})$ denote face rooted dissections where a reflection is applied.

$$Z_{\mathcal{F}^o}(\mathbf{s_1},\bar{\mathbf{u}}_\mathbf{1}) =$$

$$= \sum_{l\geq 3}\frac{1}{l}\left(\sum_{d|l}\varphi(d)\left(\sum_{\substack{i_1,i_2,\dots,i_{\frac{l}{d}}\\ j_1,j_2,\dots,j_{\frac{l}{d}}}}\prod_{m=1}^{\frac{l}{d}-1}\left(u_{i_{m+1}+j_m,d}u_{i_{\frac{l}{d}}+j_1,d}Z_{\mathcal{A}^o_{i_mj_m}}(\mathbf{s_d},\bar{\mathbf{u}}_\mathbf{d})\right)Z_{\mathcal{A}_{i_{\frac{l}{d}}j_{\frac{l}{d}}}}(\mathbf{s_1},\bar{\mathbf{u}}_\mathbf{1})\right)\right).$$

For $l = 2m+2$

$$Z_{\mathcal{F}_2^m}(\mathbf{s_1},\bar{\mathbf{u}}_\mathbf{1}) = s_1^2\left(\sum_{\substack{i_1,i_2,\dots,i_{m+1}\\ j_1,j_2,\dots,j_{m+1}}}\prod_{t=1}^{m} u_{2i_1,1}u_{j_t+i_{t+1},2}u_{2j_{m+1},1}Z_{\mathcal{A}^o_{i_tj_t}}(\mathbf{s_2},\bar{\mathbf{u}}_\mathbf{2})\right)$$

$$Z_{\mathcal{F}_3^m}(\mathbf{s_1},\bar{\mathbf{u}}_\mathbf{1}) = \sum_{i=0}^{k,\infty}Z_{\mathcal{A}^{s-}_{ii}}(\mathbf{s_1},\bar{\mathbf{u}}_\mathbf{1})\sum_{\hat{i}=0}^{k,\infty}Z_{\mathcal{A}^{s-}_{\hat{i}\hat{i}}}(\mathbf{s_1},\bar{\mathbf{u}}_\mathbf{1})$$
$$\times\left(\sum_{\substack{i_1,i_2,\dots,i_m\\ j_1,j_2,\dots,j_m}}\prod_{t=1}^{m-1} u_{i+i_1,2}u_{j_t+i_{t+1},2}u_{j_m+\hat{i},2}Z_{\mathcal{A}^o_{i_tj_t}}(\mathbf{s_2},\bar{\mathbf{u}}_\mathbf{2})\right),$$

for $l = 2m+1$

$$Z_{\mathcal{F}_1^m}(\mathbf{s_1},\bar{\mathbf{u}}_\mathbf{1}) = s_1 \sum_{i=0}^{k,\infty} Z_{\mathcal{A}_{ii}^{s-}}(\mathbf{s_1},\bar{\mathbf{u}}_\mathbf{1}) \left(\sum_{\substack{i_1,i_2,\dots,i_m\\ j_1,j_2,\dots,j_m}} \prod_{t=1}^{m-1} u_{i+i_1,2} u_{j_t+i_{t+1},2} u_{2j_m} Z_{\mathcal{A}^o_{i_t j_t}}(\mathbf{s_2},\bar{\mathbf{u}}_\mathbf{2}) \right),$$

$$Z_{\mathcal{F}^s}(\mathbf{s_1},\bar{\mathbf{u}}_\mathbf{1}) = \sum_{m\geq 1} Z_{\mathcal{F}_1^m}(\mathbf{s_1},\bar{\mathbf{u}}_\mathbf{1}) + Z_{\mathcal{F}_2^m}(\mathbf{s_1},\bar{\mathbf{u}}_\mathbf{1}) + Z_{\mathcal{F}_3^m}(\mathbf{s_1},\bar{\mathbf{u}}_\mathbf{1}).$$

B.2 Equations for unrooted unlabelled 2-connected series-parallel graphs:

The cycle index sum of unrooted 2-connected series parallel graphs is given by a similar equation to the rooted ones. The only difference is that the ring- and multiedge components are no longer rooted:

$$Z_{\mathcal{B}}(\mathbf{s_1},\bar{\mathbf{u}}_\mathbf{1}) = 1 + Z_{\mathcal{B}_\mathcal{R}}(\mathbf{s_1},\bar{\mathbf{u}}_\mathbf{1}) + Z_{\mathcal{B}_\mathcal{M}}(\mathbf{s_1},\bar{\mathbf{u}}_\mathbf{1}) - Z_{\mathcal{B}_{\mathcal{R}\mathcal{M}}}(\mathbf{s_1},\bar{\mathbf{u}}_\mathbf{1}).$$

Again, the equation can be translated into generating functions:

$$B(z,\mathbf{v}) = 1 + B_R(z,\mathbf{v}) + B_M(z,\mathbf{v}) - B_{RM}(z,\mathbf{v}),$$

Also in this part, we will state equations in terms of generating functions. The function $B_{RM}(z,\mathbf{v})$ is nearly identical to the rooted one, with the only difference that we have to count the additional vertex:

$$\begin{aligned} B_{RM}(z,\mathbf{v}) = {} & \frac{1}{2} z^2 \left(\sum_{\ell_1=1}^{k,\infty} \sum_{\ell_2=1}^{k,\infty} \sum_{\ell_3=1}^{k,\infty} \sum_{\ell_4=1}^{k,\infty} v_{\ell 1+\ell 2} v_{\ell_3+\ell_4} S_{\ell_1\ell_3}(z,\mathbf{v}) P_{\ell_2\ell_4}(z,\mathbf{v}) \right) \\ & + \frac{1}{2} z^2 \left(\sum_{\ell_1=1}^{k,\infty} \sum_{\ell_2=1}^{k,\infty} v_{\ell_1+\ell_2}^2 \bar{S}_{\ell_1}(z,\mathbf{v}) \bar{P}_{\ell_2}(z,\mathbf{v}) \right) \end{aligned}$$

The other two generating functions, on the contrary, are more involved. This is due to the cyclic permutations and reflections that can be applied to the rings and multiedges now that they are unrooted:

$$B_R(z,\mathbf{v}) = \sum_{\ell\geq 3}\frac{1}{\ell}z^\ell\sum_{d|\ell}\varphi(d)\left(\sum_{\substack{i_1,i_2,\dots,i_{\frac{\ell}{d}}\\ j_1,j_2,\dots,j_{\frac{\ell}{d}}}}\prod_{r=1}^{\frac{\ell}{d}-1}\Big(v_{j_r+i_{r+1}}v_{j_{\frac{\ell}{d}}+i_1}(D-S)_{i_rj_r}(z^d,\mathbf{v}^d)\Big)\right)$$

$$+\sum_{m\geq 1}\left(z\sum_{i=1}^{k,\infty}\sum_{\substack{i_1,i_2,\dots,i_m\\ j_1,j_2,\dots,j_m}}\prod_{r=1}^{m-1}v_{2i_1}v^2_{j_r+i_{r+1}}u^2_{j_r+1}(D-S)_{i_rj_r}(z^2,\mathbf{v}^2)(\bar{D}-\bar{S})_i(z,\mathbf{v})\right.$$

$$+\sum_{i=1}^{k,\infty}\sum_{j=1}^{k,\infty}\sum_{\substack{i_1,i_2,\dots,i_m\\ j_1,j_2,\dots,j_m}}\prod_{r=1}^{m-1}v^2_{i+i_1}v^2_{j_r+i_{r+1}}v^2_{j+j_m}(D-S)_{i_rj_r}(z^2,\mathbf{v}^2)(\bar{D}-\bar{S})_i(\bar{D}-\bar{S})_j(z,\mathbf{v})$$

$$\left.+z^2\sum_{\substack{i_1,i_2,\dots,i_{m+1}\\ j_1,j_2,\dots,j_{m+1}}}\prod_{r=1}^{m}v^2_{2i_1}v^2_{j_r+i_{r+1}}v^2_{2j_{m+1}}(D-S)_{i_rj_r}(z^2,\mathbf{v}^2)\right)$$

$$B_M(z,\mathbf{v}) = \sum_{i_1,i_2,\dots,i_k,i_\infty}\sum_{\sigma\in\mathfrak{S}_{k+1}}\Big((v_{\sum_{\ell=1}^{k,\infty}\ell i_\ell}v_{\sum_{\ell=1}^{k,\infty}\sigma(\ell)j_{\sigma(\ell)}}+v_{\sum_{\ell=1}^{k,\infty}\ell i_\ell+1}v_{\sum_{\ell=1}^{k,\infty}\sigma(\ell)j_{\sigma(\ell)}+1})$$

$$\times\prod_{r=1}^{k,\infty}\mathfrak{S}_{\ell_r}[S_{r,\sigma(r)}(z,\mathbf{v})]\Big)\qquad-(v_0^2+v_1^2)$$

$$-\sum_{i=1}^{k,\infty}\sum_{j=1}^{k,\infty}(v_iv_j+v_{i+1}v_{j+1})S_{ij}(z,\mathbf{v})-\sum_{\substack{i_1,i_2,j_1,j_2\\ \in\{1,\dots,k,\infty\}}}v_{i_1+i_2}v_{j_1+j_2}\frac{S_{i_1j_1}(z,\mathbf{v})S_{i_2j_2}(z,\mathbf{v})+S_{i_1j_1}(z^2,\mathbf{v}^2)}{2}$$

$$+\sum_{i_1,i_2,\dots,i_k,i_\infty}(v^2_{\sum_{\ell=1}^{k,\infty}\ell i_\ell}+v^2_{\sum_{\ell=1}^{k,\infty}\ell i_\ell+1})$$

$$\times\left(\prod_{r=1}^{i}\mathfrak{S}_{l_r}[s_{2s-1}\leftarrow\bar{S}_r(z^{2s-1},\mathbf{v}^{2s-1}),s_{2s}\leftarrow S_{rr}(z^{2s},\mathbf{v}^{2s}),s\geq 1]\right.$$

$$\left.+\prod_{\substack{\ell_t=\ell_{\tilde{t}}\\ t\neq\tilde{t}}}\mathfrak{S}_{\ell_t}[S_{\ell_t\ell_{\tilde{t}}}(z^2,\mathbf{v}^2)]\right)$$

$$-(v_0+v_1)-\sum_{i=1}^{k,\infty}(v_i^2+v_{i+1}^2)\bar{S}_i(z,\mathbf{v})-\sum_{\substack{i_1,i_2,j_1,j_2\\ \in\{1,\dots,k,\infty\}}}v^2_{i_1+i_2}\frac{\bar{S}_{i_1}(z,\mathbf{v})\bar{S}_{i_2}(z,\mathbf{v})+S_{i_1i_2}(z^2,\mathbf{v}^2)}{2}$$

B.3 Equations for cycle-rooted series parallel graphs

B.3.1 Series-parallel networks

Recall that Series-parallel networks are created by series-parallel extensions of a single edge. Formally, the following decomposition grammar holds for SP-networks:

$$\begin{aligned}\mathcal{D} &= e + \mathcal{P} + \mathcal{S}\\ \mathcal{S} &= \mathcal{S} * \mathcal{X} * \mathcal{D}\\ \mathcal{P} &= e * Set_{\geq 1}(\mathcal{S}) + Set_{\geq 2}(\mathcal{S})\end{aligned} \tag{B.2}$$

We will now translate this equations in terms of cycle index sums, with additional counting variables, tracking vertex degrees.

Notation In the following we count by

- w_1 the degree of the 0-pole,
- w_2 the degree of the ∞-pole,
- w the degree of the root cycle.

As we will need equations for both unrooted and cycle-rooted networks to set up equations for 2-connected series-parallel graphs, we will have various systems stated in the following. First we set up a system for unrooted networks $Z_{\mathcal{D}}(\mathbf{s_1}; w_1, w_2) = Z_{\mathcal{D}}(s_1, s_2, \ldots; w_1, w_2)$. Poles are not counted.

$$\begin{aligned}Z_{\mathcal{D}}(\mathbf{s_1}; w_1, w_2) &= w_1 w_2 + Z_{\mathcal{S}}(\mathbf{s_1}; w_1, w_2) + Z_{\mathcal{P}}(\mathbf{s_1}; w_1, w_2)\\ Z_{\mathcal{S}}(\mathbf{s_1}; w_1, w_2) &= s_1(Z_{\mathcal{P}}(\mathbf{s_1}; w_1, 1) + w_1) Z_{\mathcal{D}}(\mathbf{s_1}; 1, w_2)\\ Z_{\mathcal{P}}(\mathbf{s_1}; w_1, w_2) &= (1 + w_1) \exp\left(\sum_{k\geq 1} \frac{1}{k} Z_{\mathcal{S}}(\mathbf{s_k}; w_1^k, w_2^k)\right)\\ &\quad - (1 + w_1) - Z_{\mathcal{S}}(\mathbf{s_1}; w_1, w_2)\end{aligned}$$

Next, we set up a system for cycle rooted networks (where the root cycle does not contain the poles). We obtain equations by derivation with respect to all $s_i, i \geq 1$ as given in equation (4.25) and inserting the additional parameters $t_i, i \geq 1$ and w, which counts the degree of the root cycle, adequately:

$$\begin{aligned}Z_{\mathcal{D}^\circ}(\mathbf{s_1}, \mathbf{t_1}; w_1, w_2, w) &= Z_{\mathcal{S}^\circ}(\mathbf{s_1}, \mathbf{t_1}; w_1, w_2, w) + Z_{\mathcal{P}^\circ}(\mathbf{s_1}, \mathbf{t_1}; w_1, w_2, w)\\ Z_{\mathcal{S}^\circ}(\mathbf{s_1}, \mathbf{t_1}; w_1, w_2, w) &= t_1(Z_{\mathcal{P}}(\mathbf{s_1}; w_1, w) + w_1 w) Z_{\mathcal{D}}(\mathbf{s_1}; w, w_2)\\ &\quad + s_1 \left(Z_{\mathcal{P}^\circ}(\mathbf{s_1}, \mathbf{t_1}; w_1, w_2, w) Z_{\mathcal{D}}(\mathbf{s_1}; 1, w_2)\right.\\ &\quad + \left.(Z_{\mathcal{P}}(\mathbf{s_1}; w_1, 1) + w_1) Z_{\mathcal{D}^\circ}(\mathbf{s_1}, \mathbf{t_1}; w_1, w_2, w)\right)\\ Z_{\mathcal{P}^\circ}(\mathbf{s_1}, \mathbf{t_1}; w_1, w_2, w) &= (1 + w_1)\left(\sum_{k\geq 1} Z_{\mathcal{S}^\circ}(\mathbf{s_k}, \mathbf{t_k}; w_1^k, w_2^k, w)\right) \exp\left(\sum_{k\geq 1} \frac{1}{k} Z_{\mathcal{S}}(\mathbf{s_k}, w_1^k, w_2^k)\right)\\ &\quad - Z_{\mathcal{S}^\circ}(\mathbf{s_1}, \mathbf{t_1}; w_1, w_2, w)\end{aligned}$$

Besides the above equations for networks, we additionaly need equations for symmetric networks, that is, networks which are invariant under a reflection exchanging the poles 0 and

∞. Recall that $\mathcal{D}_2$ and $\mathcal{S}_2$ denotes two identical copies of a network $\mathcal{D}$ or $\mathcal{S}$, respectively, and $Set(\mathcal{S}_2, \bar{\mathcal{S}})$ denotes a set of pairs of arbitrary series networks together with a set of symmetric networks of odd size. Symmetric networks follow the symbolic description

$$\begin{aligned}
\bar{\mathcal{D}} &= e + \bar{\mathcal{S}} + \bar{\mathcal{P}} \\
\bar{\mathcal{S}} &= \mathcal{D}_2 * (\mathcal{X} + \mathcal{X}^2 * (e + \bar{\mathcal{P}})) \\
\bar{\mathcal{P}} &= e * Set_{\geq 1}(\mathcal{S}_2, \bar{\mathcal{S}}) + Set_{\geq 2}(\mathcal{S}_2, \bar{\mathcal{S}})
\end{aligned} \tag{B.3}$$

Obviously, in symmetric networks both poles are of the same degree, thus we need just one counting variable w_1. We obtain the following system of equations for the cycle index sums $Z_{\bar{\mathcal{D}}}(\mathbf{s_1}; w_1)$ of unrooted symmetric SP-networks:

$$\begin{aligned}
Z_{\bar{\mathcal{D}}}(\mathbf{s_1}; w_1) &= w_1 + Z_{\bar{\mathcal{S}}}(\mathbf{s_1}; w_1) + Z_{\bar{\mathcal{P}}}(\mathbf{s_1}; w_1) \\
Z_{\bar{\mathcal{S}}}(\mathbf{s_1}; w_1) &= Z_{\mathcal{D}}(\mathbf{s_2}; w_1, 1)(s_1 + s_2(1 + Z_{\bar{\mathcal{P}}}(\mathbf{s_1}; 1))) \\
Z_{\bar{\mathcal{P}}}(\mathbf{s_1}; w_1) &= (1 + w_1)\exp\left(\sum_{k\geq 1}\left(\frac{1}{2k}Z_{\mathcal{S}}(\mathbf{s_{2k}}; w_1^k, w_2^k) + \frac{1}{2k-1}Z_{\bar{\mathcal{S}}}(\mathbf{s_{2k+1}}; w_1^{2k+1})\right)\right) \\
&\quad - (1 + w_1) - Z_{\bar{\mathcal{S}}}(\mathbf{s_1}; w_1)
\end{aligned}$$

If we cycle-root a symmetric network (where the root cycle does not contain the poles), that is, we derivate the above system as given in (4.25), we obtain the following system:

$$\begin{aligned}
Z_{\bar{\mathcal{D}}^\circ}(\mathbf{s_1}, \mathbf{t_1}; w_1, w) &= Z_{\bar{\mathcal{S}}^\circ}(\mathbf{s_1}, \mathbf{t_1}; w_1, w) + Z_{\bar{\mathcal{P}}^\circ}(\mathbf{s_1}, \mathbf{t_1}; w_1, w) \\
Z_{\bar{\mathcal{S}}^\circ}(\mathbf{s_1}, \mathbf{t_1}; w_1, w) &= 2Z_{\mathcal{D}^\circ}(\mathbf{s_2}, \mathbf{t_2}; w_1, 1, w)(s_1 + s_2(1 + Z_{\bar{\mathcal{P}}}(\mathbf{s_1}; 1))) \\
&\quad + Z_{\mathcal{D}}(\mathbf{s_2}; w_1, w^2)t_1 \\
&\quad + Z_{\mathcal{D}}(\mathbf{s_2}; w_1, w)2t_2(1 + Z_{\bar{\mathcal{P}}}(\mathbf{s_1}; 1))) \\
&\quad + Z_{\mathcal{D}}(\mathbf{s_2}; w_1, 1)(s_2(Z_{\bar{\mathcal{P}}^\circ}(\mathbf{s_1}, \mathbf{t_1}; 1, w) \\
Z_{\bar{\mathcal{P}}^\circ}(\mathbf{s_1}, \mathbf{t_1}; w_1, w) &= (1 + w_1)\sum_{k\geq 1}\left(Z_{\mathcal{S}^\circ}(\mathbf{s_{2k}}, \mathbf{t_{2k}}; w_1^k, w_2^k, w) + Z_{\bar{\mathcal{S}}^\circ}(\mathbf{s_{2k+1}}, \mathbf{t_{2k+1}}; w_1^{2k+1}, w)\right) \\
&\quad \times \exp\left(\sum_{k\geq 1}\left(\frac{1}{2k}Z_{\mathcal{S}}(\mathbf{s_{2k}}; w_1^k, w_2^k) + \frac{1}{2k-1}Z_{\bar{\mathcal{S}}}(\mathbf{s_{2k+1}}; w_1^{2k+1})\right)\right) \\
&\quad - Z_{\bar{\mathcal{S}}^\circ}(\mathbf{s_1}, \mathbf{t_1}; w_1, w)
\end{aligned}$$

B.3.2 2-connected SP-graphs

Decomposition

To find equations for cycle-rooted 2-connected SP-graphs, we use Tutte's decomposition into cycles, multiedges and substituted networks (cf [67]), introduced in Section 4.3.2. The components are, as described previously, either ring graphs or multiedges, both with at least 3 edges, where the edges are replaced by networks such that the decomposition is unique. That is, edges of ring components are replaced by networks which are not series, while multiedges are replaced by series networks. To translate this substitution procedure to cycle index sums, we first need to describe the underlying cycle and multiedge structures. Therefore we use Walsh series.

The symmetry group of rings $\mathcal{R}$ is determined by the dyhedral group (rotations and reflections), while the symmetry group of multiedges $\mathcal{M}$ is determined by the full symmetric groups on the edges and the remaining symmetries which include exchanging the root vertices. The Walsh series are given in Equations (4.22) and (4.23). Recall that in the decomposition mentioned above, cycle and multiedge components form a bipartide tree, and thus we use dissimetry theorem on trees (Theorem 1.1).

We will denote by $W(\mathbf{a_1}, \mathbf{b_1}, \mathbf{c_1}) \circ_e [\mathbf{s_1}, Z_{\mathcal{A}}(\mathbf{s_1}), Z_{\mathcal{B}}(\mathbf{s_1})]$ the substitution $a_i \leftarrow s_i, b_i \leftarrow Z_{\mathcal{A}}(\mathbf{s_i}), c_i \leftarrow Z_{\mathcal{B}}(\mathbf{s_i})$ for all $i \geq 1$.

To set up equations for cycle-rooted 2-connected SP-graphs, we distinguish whether the root cycle has length 1 or greater than 1, that is, whether the root is a fixed point or a "real cycle". For the cycle index sum of 2-connected cycle-rooted SP-graphs we then obtain

$$Z_{\mathcal{B}^\circ}(\mathbf{s_1}, \mathbf{t_1}; w) = Z_{\hat{\mathcal{B}}}(\mathbf{s_1}, \mathbf{t_1}; w) + Z_{\mathcal{B}^\bullet}(\mathbf{s_1}, \mathbf{t_1}; w),$$

where $\hat{\mathcal{B}}$ denotes the family of vertex rooted 2-connected SP-graphs and $\mathcal{B}^\bullet$ the family of 2-connected SP-graphs rooted at a cycle of length $\ell \geq 2$.

Vertex rooted 2-connected SP-graphs

We know that

$$\hat{\mathcal{B}} = \mathcal{B}_{\hat{\mathcal{R}}} + \mathcal{B}_{\hat{\mathcal{M}}} - \mathcal{B}_{\mathcal{R}\mathcal{M}},$$

where the family $\mathcal{B}_{\hat{\mathcal{R}}}$ consists of vertex rooted ring components $\hat{\mathcal{R}}$, where every edge may be replaced by a parallel network, and the family $\mathcal{B}_{\hat{\mathcal{M}}}$ consists of multiple edges rooted at one of the end vertices $\hat{\mathcal{M}}$, where every but one edge is replaced by a series network and the last edge can be replaced by a series network or not. The family $\mathcal{R}\mathcal{M}$ represents the edges in the dissymetry theorem and is given by the intersection of a ring and a multiedge, that is a series and a parallel network. This relation translates to

$$Z_{\hat{\mathcal{B}}}(\mathbf{s_1}; w) = Z_{\mathcal{B}_{\hat{\mathcal{R}}}}(\mathbf{s_1}; w) + Z_{\mathcal{B}_{\hat{\mathcal{M}}}}(\mathbf{s_1}; w) - Z_{\mathcal{B}_{\mathcal{R}\mathcal{M}}}(\mathbf{s_1}; w),$$

in terms of cycle index sums and

$$\hat{B}(x, w) = B_{\hat{R}}(x, w) + B_{\hat{M}}(x, w) - B_{RM}(x, w),$$

in terms of generating functions. The generating functions of the families appearing in the above equation are given by the substitutions

$$\begin{aligned}
&Z_{\hat{\mathcal{B}}}(\mathbf{s_1}) = W_{\hat{\mathcal{R}}} \circ_e [s_1, Z_{\mathcal{D}-\mathcal{S}}(\mathbf{s_1}), Z_{\bar{\mathcal{D}}-\bar{\mathcal{S}}}(\mathbf{s_1})] && B_{\hat{R}}(x) = W_{\hat{\mathcal{R}}} \circ_e [x, (D-S)(x), (\bar{D}-\bar{S})(x)] \\
&Z_{\mathcal{B}_{\hat{\mathcal{M}}}}(\mathbf{s_1}) = W_{\hat{\mathcal{M}}} \circ_e [s_1, Z_{\mathcal{S}}(\mathbf{s_1}), Z_{\bar{\mathcal{S}}}(\mathbf{s_1})] && B_{\hat{M}}(x) = W_{\hat{\mathcal{M}}} \circ_e [x, S(x), \bar{S}(x)] \\
&Z_{\mathcal{B}_{\mathcal{R}\mathcal{M}}}(\mathbf{s_1}) = s_1 Z_{\mathcal{P}}(\mathbf{s_1}) Z_{\mathcal{S}}(\mathbf{s_1}) && B_{RM}(x) = xP(x)S(x),
\end{aligned}$$

where the substitution on $W_{\hat{\mathcal{M}}}$ is performed on all or one but all edges. To obtain exact equations, containing the new variable w, we first determine $W_{\hat{\mathcal{R}}}(\mathbf{a_1}, \mathbf{b_1}, \mathbf{c_1})$. Rooting a ring graph at a vertex, possible symmetries are either the identity or reflections fixing the root vertex, rotations are eliminated. The Walsh series of rooted ring components can be obtained from the unrooted version by derivating with respect to all fixed points a_1, and is given by

$$\begin{aligned}W_{\hat{\mathcal{R}}}(\mathbf{a_1},\mathbf{b_1},\mathbf{c_1}) &= \frac{1}{2}\left(\frac{a_1^2b_1^3}{1-a_1b_1}+\frac{a_2b_2}{1-a_2b_2}(a_1b_2+c_1)\right)\\ &= \underbrace{\frac{1}{2}a_1^2b_1^2\left(\sum_{k\geq 0}(a_1b_1)^k\right)b_1}_{①}+\underbrace{\frac{1}{2}a_2b_2\left(\sum_{k\geq 0}(a_2b_2)^k\right)(a_1b_2+c_1)}_{②}\end{aligned}$$

where the first term ①) represents the identity and the second term ②) represents reflections fixing the root vertex. In both terms the tupel $a_1^2b_1^2$ and a_2b_2, respectively, represent the edges and according endpoints neighboring the root vertices and thus contributing to the root degree. By substituting, we obtain

$$\begin{aligned}Z_{\mathcal{B}_{\hat{\mathcal{R}}}}(\mathbf{s_1};w) &= \frac{1}{2}s_1^2(Z_{\mathcal{D}-\mathcal{S}}(\mathbf{s_1};w,1))^2\frac{Z_{\mathcal{D}-\mathcal{S}}(\mathbf{s_1};1,1)}{1-s_1(Z_{\mathcal{D}-\mathcal{S}}(\mathbf{s_1};1,1))}\\ &+\frac{1}{2}s_2(Z_{\mathcal{D}-\mathcal{S}}(\mathbf{s_2};w^2,1))\frac{s_1(Z_{\mathcal{D}-\mathcal{S}}(\mathbf{s_2};1,1))+1+Z_{\bar{\mathcal{P}}}(\mathbf{s_1};1)}{1-s_2(Z_{\mathcal{D}-\mathcal{S}}(\mathbf{s_2};1,1))}\end{aligned}\tag{B.4}$$

In terms of generating functions, we obtain

$$\begin{aligned}B_{\hat{R}}(x,w) &= \frac{1}{2}x^2((D-S)(x,w,1))^2\frac{(D-S)(x,1,1)}{1-x((D-S)(x,1,1))}\\ &+\frac{1}{2}x^2((D-S)(x^2,w^2,1))\frac{x((D-S)(x,1,1))+1+\bar{P}(x,1)}{1-x^2((D-S)(x^2,1,1))}\end{aligned}$$

The Walsh series of a rooted multiedge consisting of at least 3 edges is given by

$$W_{\hat{\mathcal{M}}}(\mathbf{a_1},\mathbf{b_1},\mathbf{c_1}) = a_1\left(\exp\left(\sum_{k\geq 1}\frac{b_k}{k}\right)-1-b_1-\frac{1}{2}(b_1^2+b_2)\right)$$

For a multiedge of size ≥ 2 the last term $-\frac{1}{2}(b_1^2+b_2)$ dissappears. In our case, the multiedge components are either of size ≥ 2, where the edges are substituted with series networks, and there is one "real" edge connecting the 2 endpoints, or it is a multiedge of size ≥ 3, where the edges are series networks. Therefore, we obtain

$$\begin{aligned}Z_{\mathcal{B}_{\hat{\mathcal{M}}}}(\mathbf{s_1};w) &= s_1\left((1+w)\left(\exp\left(\sum_{k\geq 1}\frac{Z_{\mathcal{S}}(\mathbf{s_k};w^k,1)}{k}\right)-1-Z_{\mathcal{S}}(\mathbf{s_1};w,1)\right)\right.\\ &\left.-\frac{1}{2}(Z_{\mathcal{S}}(\mathbf{s_1};w,1)^2+Z_{\mathcal{S}}(\mathbf{s_2};w^2,1))\right)\end{aligned}\tag{B.5}$$

$$\begin{aligned}B_{\hat{M}}(x,w) &= x\left((1+w)\left(\exp\left(\sum_{k\geq 1}\frac{S(x^k,w^k,1)}{k}\right)-1-S(x,w,1)\right)\right.\\ &\left.-\frac{1}{2}(S(x,w,1)^2+S(x^2,w^2,1))\right)\end{aligned}$$

Last, for the family $\mathcal{B}_{\mathcal{RM}}$ we obtain equations

$$\begin{aligned}Z_{\mathcal{B}_{\mathcal{RM}}}(\mathbf{s_1};w) &= s_1Z_{\mathcal{P}}(\mathbf{s_1};w,1)Z_{\mathcal{S}}(\mathbf{s_1};w,1)\\ B_{RM}(x,w) &= xP(x,w,1)S(x,w,1)\end{aligned}\tag{B.6}$$

as both networks intersect at the root vertex and thus both degrees contribute.

Symmetric 2-connected SP-graphs

Let us call 2-connected graphs, rooted at a cycle of length $\ell \geq 2$ symmetric. This is a natural term, as containing a cycle of length $\ell \geq 2$ implies that the graph as well as it's underlying $R-M$ tree are symmetric. Counting symmetric bipartide trees we do not need dissymetry theorem anymore, as there is a unique symmetry node, which represents either a $\mathcal{M}$ or a $\mathcal{R}$ component, which we can use as a base for counting. We distinguish 2 cases in counting symmetric 2-connected Sp-graphs. The first case appears if the root cycle of the graph is entirely contained in the central component of the restriced $R-M$ tree, thus it implies an "ordinary" vertex root at the symmetry node of the underlying tree. We call this case the case of being rooted at a "central" cycle, and denote the according family by $\mathcal{B}^c$. The second possible case is that the root cycle of the graph implies a "real" root cycle, that is a cycle of length $\ell \geq 2$, on the vertices of the underlying $R-M$ tree. We call this second case the case of being rooted at a "noncentral" cycle, and denote the family by $\mathcal{B}^n$. Obviously,

$$Z_{\mathcal{B}^\bullet}(\mathbf{s_1},\mathbf{t_1},w) = Z_{\mathcal{B}^c}(\mathbf{s_1},\mathbf{t_1},w) + Z_{\mathcal{B}^n}(\mathbf{s_1},\mathbf{t_1},w) \qquad \text{and} \qquad B^\bullet(x,w) = B^c(x,w) + B^n(x,w)$$

- **rooted at a central cycle** The root cycle coincides with a vertex-cycle in the central component of the $R-M$-tree of the graph, that is, all vertices of the root cycle are contained in the same basic component. Therefore the cycle index sums $Z_{\mathcal{B}^c}(\mathbf{s_1},w)$ is determined by a cycle rooted object of type $\mathcal{R}$ or $\mathcal{M}$, where the edges are again substituted by networks. Of course, the root cycle still is of length $\ell \geq 2$. That is,

$$\begin{aligned} Z_{\mathcal{B}^c}(\mathbf{s_1},w) &= W_{\mathcal{R}^\bullet}(\mathbf{a_1},\mathbf{t_1},\mathbf{b_1},\mathbf{c_1}) \circ_e [\mathbf{s_1},\mathbf{t_1},Z_{\mathcal{D}-\mathcal{S}}(\mathbf{s_1},w),Z_{\bar{\mathcal{D}}-\bar{\mathcal{S}}}(\mathbf{s_1},w)] \\ &\quad + W_{\mathcal{M}^\bullet}(\mathbf{a_1},\mathbf{t_1},\mathbf{b_1},\mathbf{c_1}) \circ_e [\mathbf{s_1},\mathbf{t_1},Z_{\mathcal{S}}(\mathbf{s_1},w),Z_{\bar{\mathcal{S}}}(\mathbf{s_1},w)] \\ &=: Z_{\mathcal{B}^c_{\mathcal{R}}}(\mathbf{s_1},\mathbf{t_1},w) + Z_{\mathcal{B}^c_{\mathcal{M}}}(\mathbf{s_1},\mathbf{t_1},w), \end{aligned}$$

or, in terms of generating functions

$$\begin{aligned} B^c(x,w) &= W_{\mathcal{R}^\bullet}(\mathbf{a_1},\mathbf{t_1},\mathbf{b_1},\mathbf{c_1}) \circ_e [x,x,(D-S)(x,w),(\bar{D}-\bar{S})(x,w)] \\ &\quad + W_{\mathcal{M}^\bullet}(\mathbf{a_1},\mathbf{t_1},\mathbf{b_1},\mathbf{c_1}) \circ_e [x,x,S(x,w),\bar{S}(x,w)] \\ &=: B^c_R(x,w) + B^c_M(x,w). \end{aligned}$$

To obtain the Walsh series of cycle rooted rings, we have to derivate the Walsh series with respect to all $a_i, i \geq 2$ and introduce marking variables t_i. We obtain

$$\begin{aligned} W_{\mathcal{R}^\bullet}(\mathbf{a_1},\mathbf{t_1},\mathbf{b_1},\mathbf{c_1}) &= \left(\sum_{k\geq 2} \frac{\varphi(k)}{2}\frac{t_k b_k}{1-a_k b_k} - \frac{t_2 b_2}{2}\right) \\ &\quad + \frac{2t_2 b_2}{(1-a_2 b_2)^2}\left(\frac{a_2 c_1^2}{4} + \frac{a_1^2 b_2}{4} + \frac{a_1 c_1}{2}\right) + \frac{a_2 b_2}{1-a_2 b_2}\frac{t_2 c_1^2}{2} \\ &= \underbrace{\sum_{k\geq 2}\frac{\varphi(k)}{2} t_k b_k (1 + \sum_{\ell\geq 1}(a_k b_k)^\ell) - \frac{t_2 b_2}{2}}_{①} \\ &\quad \underbrace{+ 2t_2\left(\left(\sum_{\ell_1\geq 0}(a_2 b_2)^{\ell_1}\right) b_2 \left(\sum_{\ell_2\geq 0}(a_2 b_2)^{\ell_2}\right)\left(\frac{a_2 c_1^2}{4} + \frac{a_1^2 b_2}{4} + \frac{a_1 c_1}{2}\right)\right) + \frac{a_2 b_2}{1-a_2 b_2}\frac{t_2 c_1^2}{2}}_{②} \end{aligned}$$

where the first term (①) again represents rotations, i.e., one circle of length $k \geq 2$ is marked. This circle either contains all vertices of the ring, or just some of them. In the first case, the cycle has to be of length at least 3, whis is secured by the term $-\frac{t_2 b_2}{2}$, and the degree of both poles 0 and ∞ contribute to our new parameter w. In the second case all other nodes and edges form cycles of equal length k, and the exponent of w is determined by both networks neighboring the root-vertices, with one pole-type each. The second term (②) represents reflections, where, naturally, the root cycle has length exactly 2. Every reflection fixes either 2 vertices, or 2 edges, which change orientation, or it fixes 1 vertex and 1 edge. Rooting at a cycle of a reflection, there is the possibility to root in the "middle" of the reflection, which is represented by the first part in②, or at the "edge" of the reflection, that is, at the endpoints of an edge which is fixed under the reflection but changes it's orientation. Substitution leads to

$$
\begin{aligned}
Z_{\mathcal{B}^c_{\mathcal{R}}}(\mathbf{s_1}, \mathbf{t_1}, w) = & \sum_{k \geq 2} \frac{\varphi(k)}{2} t_k \left[Z_{\mathcal{D}-\mathcal{S}}(\mathbf{s_k}, w, w) + Z_{\mathcal{D}-\mathcal{S}}(\mathbf{s_k}, w, 1) \cdot \frac{s_k Z_{\mathcal{D}-\mathcal{S}}(\mathbf{s_k}, 1, w)}{1 - s_k Z_{\mathcal{D}-\mathcal{S}}(\mathbf{s_k}, 1, 1)} \right] \\
& - \frac{1}{2} t_2 Z_{\mathcal{D}-\mathcal{S}}(\mathbf{s_2}, w, w) \\
& + 2t_2 \left(Z_{\mathcal{D}-\mathcal{S}}(\mathbf{s_2}, w, 1) \frac{1}{1 - s_2 Z_{\mathcal{D}-\mathcal{S}}(\mathbf{s_2}, 1, 1)} \right. \\
& \times \left[\frac{Z_{\mathcal{D}-\mathcal{S}}(\mathbf{s_2}, 1, w)}{1 - s_2 Z_{\mathcal{D}-\mathcal{S}}(\mathbf{s_2}, 1, 1)} \left(\frac{s_2 Z_{\bar{\mathcal{D}}-\bar{\mathcal{S}}}(\mathbf{s_1}, 1)^2}{4} + \frac{s_2 Z_{\mathcal{D}-\mathcal{S}}(\mathbf{s_2}, 1, 1)}{4} + \frac{s_1 Z_{\bar{\mathcal{D}}-\bar{\mathcal{S}}}(\mathbf{s_1}, 1)}{2} \right) \right. \\
& + \left. \left(\frac{s_2 Z_{\bar{\mathcal{D}}-\bar{\mathcal{S}}}(\mathbf{s_1}, w) Z_{\bar{\mathcal{D}}-\bar{\mathcal{S}}}(\mathbf{s_1}, 1)}{4} + \frac{s_2 Z_{\mathcal{D}-\mathcal{S}}(\mathbf{s_2}, 1, w)}{2} + \frac{s_1 Z_{\bar{\mathcal{D}}-\bar{\mathcal{S}}}(\mathbf{s_1}, w)}{2} \right) \right] \\
& + \left. \frac{s_2 Z_{\mathcal{D}-\mathcal{S}}(\mathbf{s_2}, w, 1)}{1 - s_2 Z_{\mathcal{D}-\mathcal{S}}(\mathbf{s_2}, 1, 1)} \frac{Z_{\bar{\mathcal{D}}-\bar{\mathcal{S}}}(\mathbf{s_1}, w) Z_{\bar{\mathcal{D}}-\bar{\mathcal{S}}}(\mathbf{s_1}, 1)}{4} \right)
\end{aligned}
$$

and further to

$$
\begin{aligned}
B^c_R(x, w) = & \sum_{k \geq 2} \frac{\varphi(k)}{2} x^k \left[(D-S)(x^k, w, w) + (D-S)(x^k, w, 1) \cdot \frac{x^k (D-S)(x^k, 1, w)}{1 - x^k (D-S)(x^k, 1, 1)} \right] \\
& - \frac{1}{2} x^2 (D-S)(x^2, w, w) \\
& + 2x^2 \left((D-S)(x^2, w, 1) \frac{1}{1 - x^2 (D-S)(x^2, 1, 1)} \right. \\
& \times \left[\frac{(D-S)(x^2, 1, w)}{1 - x^2 (D-S)(x^2, 1, 1)} \left(\frac{x^2 (\bar{D}-\bar{S})(x, 1)^2}{4} + \frac{x^2 (D-S)(x^2, 1, 1)}{4} + \frac{x (\bar{D}-\bar{S})(x, 1)}{2} \right) \right. \\
& + \left. \left(\frac{x^2 (\bar{D}-\bar{S})(x, w)(\bar{D}-\bar{S})(x, 1)}{4} + \frac{x^2 (D-S)(x^2, 1, w)}{2} + \frac{x (\bar{D}-\bar{S})(x, w)}{2} \right) \right] \\
& + \left. \frac{x^2 (D-S)(x^2, w, 1)}{1 - x^2 (D-S)(x^2, 1, 1)} \frac{(\bar{D}-\bar{S})(x, w)(\bar{D}-\bar{S})(x, 1)}{4} \right)
\end{aligned}
$$

The Walsh series of multiedges are also obtained by derivation, the only possible root cycle is part of a symmetry exchanging the endpoints of the multiedge.

$$
W_{\mathcal{M}^\bullet}(\mathbf{a_1}, \mathbf{t_1}, \mathbf{b_1}, \mathbf{c_1}) = t_2 \left(\exp \left(\sum_{k \geq 1} \left(\frac{b_{2k}}{2k} + \frac{c_{2k-1}}{2k-1} \right) \right) - 1 - \frac{b_2}{2} - c_1 - \frac{c_1^2}{2} \right)
$$

Substitution, either for all or for all but one edges lead to

$$Z_{\mathcal{B}^c_{\mathcal{M}}}(\mathbf{s_1}, w) = t_2 \Bigg[(1+w)\exp\left(\sum_{k\geq 1}\left(\frac{Z_{\mathcal{S}}(\mathbf{s_{2k}}, w^k, w^k)}{2k} + \frac{Z_{\bar{\mathcal{S}}}(\mathbf{s_{2k-1}}, w^{2k-1})}{2k-1}\right)\right) - (1+w)(1 - Z_{\bar{\mathcal{S}}}(\mathbf{s_1}, w)) - \frac{Z_{\mathcal{S}}(\mathbf{s_2}, w, w)}{2} - \frac{Z_{\bar{\mathcal{S}}}(\mathbf{s_1}, w)^2}{2}\Bigg]$$

$$B^c_M(x, w) = x^2 \Bigg[(1+w)\exp\left(\sum_{k\geq 1}\left(\frac{S(x^{2k}, w^k, w^k)}{2k} + \frac{\bar{S}(x^{2k-1}, w^{2k-1})}{2k-1}\right)\right) - (1+w)(1 - \bar{S}(x, w)) - \frac{S(x^2, w, w)}{2} - \frac{\bar{S}(x, w)^2}{2}\Bigg]$$

- **rooted at a noncentral cycle** The root cycle does not coincide with a cycle in the central component of the $R-M$-tree of the graph. Therefore we count $R-M$ trees, which are rooted at a symmetry node. The vertices of the root cycle are either part of ring or multiedge components. Those are parts of networks which have been substituted to the edges of the symmetry node of the $R-M$ tree, thus the graph is rooted at a cycle of edges of length ≥ 2 of the central component of the tree, which are then substituted with networks rooted at cycles not containing the poles (the root cycle of the network could have length 1). The length of the root cycle is the product of the length of the two root cycles, as described in [5]. The remaining edges of the symmetry component are substituted by non rooted networks. Thus, $B^n(x, w)$ is given by

$$\begin{aligned} Z_{\mathcal{B}^n}(\mathbf{s_1}, \mathbf{t_1}, w) &= W_{\mathcal{R}^{\bullet e}} \odot_e [Z_{\mathcal{D}^\circ - \mathcal{S}^\circ}(\mathbf{s_1}, \mathbf{t_1}, w), Z_{\bar{\mathcal{D}}^\circ - \bar{\mathcal{S}}^\circ}(\mathbf{s_1}, \mathbf{t_1}, w), \mathbf{s_1}, Z_{\mathcal{D}-\mathcal{S}}(\mathbf{s_1}), Z_{\bar{\mathcal{D}}-\bar{\mathcal{S}}}(\mathbf{s_1})] \\ &\quad + W_{\mathcal{M}^{\bullet e}} \odot_e [Z_{\mathcal{D}^\circ - \mathcal{S}^\circ}(\mathbf{s_1}, \mathbf{t_1}, w), Z_{\bar{\mathcal{D}}^\circ - \bar{\mathcal{S}}^\circ}(\mathbf{s_1}, \mathbf{t_1}, w), \mathbf{s_1}, Z_{\mathcal{D}-\mathcal{S}}(\mathbf{s_1}), Z_{\bar{\mathcal{D}}-\bar{\mathcal{S}}}(\mathbf{s_1})] \\ &=: Z_{\mathcal{B}^n_{\mathcal{R}}}(\mathbf{s_1}, w) + Z_{\mathcal{B}^n_{\mathcal{M}}}(\mathbf{s_1}, w) \end{aligned}$$

where $W_{\mathcal{R}^{\bullet e}}(\mathbf{t_1}, \bar{\mathbf{t_1}}, \mathbf{a_1}, \mathbf{b_1}, \mathbf{c_1})$ denotes the Walsh series of ring components rooted at an edge cycle which is counted by t_l for "ordinary" edges and $\bar{t}_l$ for cycles of edges which change their orientation. $\odot_e$ denotes the substitution described above, that is $t_\ell \leftarrow Z_{\mathcal{D}^\circ - \mathcal{S}^\circ}(\mathbf{s}_\ell, \mathbf{t}_\ell, w), \bar{t}_\ell \leftarrow Z_{\bar{\mathcal{D}}^\circ - \bar{\mathcal{S}}^\circ}(\mathbf{s}_\ell, \mathbf{t}_\ell, w), a_\ell \leftarrow s_\ell, b_\ell \leftarrow Z_{\mathcal{D}-\mathcal{S}}(\mathbf{s}_\ell)$ and $c_\ell \leftarrow Z_{\bar{\mathcal{D}}-\bar{\mathcal{S}}}(\mathbf{s}_\ell)$. The Walsh series of edge rooted components is obtained by derivation with respect to all variables b_ℓ and c_ℓ for
$ell \geq 2$, that is

$$W_{\mathcal{R}^{\bullet e}}(\mathbf{t_1}, \bar{\mathbf{t_1}}, \mathbf{a_1}, \mathbf{b_1}, \mathbf{c_1}) = \sum_{\ell \geq 2} \ell t_\ell \frac{\partial}{\partial b_\ell} W_{\mathcal{R}}(\mathbf{a_1}, \mathbf{b_1}, \mathbf{c_1}) + \ell \bar{t}_\ell \frac{\partial}{\partial c_\ell} W_{\mathcal{R}}(\mathbf{a_1}, \mathbf{b_1}, \mathbf{c_1})$$

Doing so, for ring components we obtain

$$W_{\mathcal{R}^{\bullet e}}(\mathbf{t_1}, \bar{\mathbf{t}}_\mathbf{1}, \mathbf{a_1}, \mathbf{b_1}, \mathbf{c_1}) = \underbrace{\sum_{k\geq 2}\left(\frac{\varphi(k)}{2}\frac{a_k t_k}{1 - a_k b_k}\right) - \frac{a_2 t_2}{2}}_{①} + \underbrace{\frac{a_2 t_2}{(1 - a_2 b_2)^2}\left(\frac{a_2 c_1^2}{2} + \frac{a_1^2 b_2}{2} + a_1 c_1\right) + \frac{a_2 b_2}{1 - a_2 b_2}\left(\frac{a_1^2 t_2}{2}\right)}_{②}$$

where again,① represents rotations and② reflections. The series can be interpreted analogously to the vertex-rooted case, but as we substitute cycle rooted networks, the root degree w is now counted directly by the generating functions of the networks. We obtain for $Z_{\mathcal{B}^n_{\mathcal{R}}}(\mathbf{s_1}, w)$

$$\begin{aligned}
Z_{\mathcal{B}^n_{\mathcal{R}}}(\mathbf{s_1}, \mathbf{t_1}, w) = &\sum_{k\geq 2}\left(\frac{\varphi(k)}{2}\frac{s_k(Z_{\mathcal{D}^\circ-\mathcal{S}^\circ})(\mathbf{s_k}, \mathbf{t_k}, 1, 1, w)}{1 - s_k Z_{\mathcal{D}-\mathcal{S}}(\mathbf{s_1}, 1, 1)}\right) - \frac{s_2 Z_{\mathcal{D}^\circ-\mathcal{S}^\circ}(\mathbf{s_2}, \mathbf{t_2}, 1, 1, w)}{2}\\
&+ \frac{s_2 Z_{\mathcal{D}^\circ-\mathcal{S}^\circ}(\mathbf{s_2}, \mathbf{t_2}, 1, 1, w)}{(1 - Z_{\mathcal{D}-\mathcal{S}}(\mathbf{s_1}, 1, 1))^2}\left(\frac{s_2(Z_{\bar{\mathcal{D}}-\bar{\mathcal{S}}}(\mathbf{s_1}, 1)^2}{2} + \frac{s_2 Z_{\mathcal{D}-\mathcal{S}}(\mathbf{s_2}, 1, 1)}{2} + s_1 Z_{\bar{\mathcal{D}}-\bar{\mathcal{S}}}(\mathbf{s_1}, 1)\right)\\
&+ \frac{s_2 Z_{\mathcal{D}-\mathcal{S}}(\mathbf{s_2}, 1, 1)}{1 - s_k Z_{\mathcal{D}-\mathcal{S}}(\mathbf{s_1}, 1, 1)}\left(\frac{s_2 Z_{\mathcal{D}^\circ-\mathcal{S}^\circ}(\mathbf{s_2}, \mathbf{t_2}, 1, 1, w)}{2}\right),
\end{aligned}$$

and for $B^n_R(x, w)$

$$\begin{aligned}
B^n_R(x, w) = &\sum_{k\geq 2}\left(\frac{\varphi(k)}{2}\frac{x^k(D^\circ - S^\circ)(x^k, 1, 1, w)}{1 - x^k(D - S)(x, 1, 1)}\right) - \frac{x^2(D^\circ - S^\circ)(x^2, 1, 1, w)}{2}\\
&+ \frac{x^2(D^\circ - S^\circ)(x^2, 1, 1, w)}{(1 - x^k(D - S)(x, 1, 1))^2}\left(\frac{x^2(\bar{D} - \bar{S})(x, 1)^2}{2} + \frac{x^2(D - S)(x^2, 1, 1)}{2} + x(\bar{D} - \bar{S})(x, 1)\right)\\
&+ \frac{x^2(D - S)(x^2, 1, 1)}{1 - x^k(D - S)(x, 1, 1)}\left(\frac{x^2(D^\circ - S^\circ)(x^2, 1, 1, w)}{2}\right).
\end{aligned}$$

The Walsh series of edge-rooted multiedges is given by

$$\begin{aligned}
W_{\mathcal{M}^{\bullet_e}}(\mathbf{t_1}, \bar{\mathbf{t}}_\mathbf{1}, \mathbf{a_1}, \mathbf{b_1}, \mathbf{c_1}) = &\frac{a_1^2}{2}\left[\sum_{k\geq 2} t_k \exp\left(\sum_{\ell\geq 1}\frac{b_\ell}{\ell}\right) - t_2\right]\\
&+ \frac{a_2}{2}\left[\sum_{k\geq 1}(t_{2k} + \bar{t}_{2k+1})\exp\left(\sum_{\ell\geq 1}\frac{b_{2\ell}}{2\ell} + \frac{c_{2\ell-1}}{2\ell-1}\right) - t_2\right]
\end{aligned}$$

The root degree is given by the root cycle of the substituted network again, a simple edge cannot have a root cycle not containing the poles, thus to count those multiedges where all edges are subsituted plus those where one is still an edge, we can take the above twice.

$$\begin{aligned}
Z_{\mathcal{B}^n_{\mathcal{M}}}(\mathbf{s_1}, \mathbf{t_1}, w) = s_2&\left(\sum_{k\geq 2} Z_{\mathcal{S}^\circ}(\mathbf{s_k}, \mathbf{t_k}, 1, 1, w)\exp\left(\sum_{\ell\geq 1}\frac{Z_{\mathcal{S}}(\mathbf{s_1}, 1, 1)}{\ell}\right)\right)\\
&+ s_2\left(\sum_{k\geq 1}(Z_{\mathcal{S}^\circ}(\mathbf{s_{2k}}, \mathbf{t_{2k}}, 1, 1, w) + Z_{\bar{\mathcal{S}}^\circ}(\mathbf{s_{2k+1}}, \mathbf{t_{2k+1}}, 1, 1, w))\right.\\
&\left.\times \exp\left(\sum_{\ell\geq 1}\frac{Z_{\mathcal{S}}(\mathbf{s_{2\ell}}, 1, 1}{2\ell} + \frac{Z_{\bar{\mathcal{S}}}(\mathbf{s_{2\ell-1}}, 1, 1)}{2\ell - 1}\right)\right) - \frac{s_2}{2} Z_{\mathcal{S}^\circ}(\mathbf{s_2}, \mathbf{t_2}, 1, 1, w)
\end{aligned}$$

$$B_M^n(x,w) = x^2\left(\sum_{k\geq 2} S^\circ(x^k,1,1,w)\exp\left(\sum_{\ell\geq 1}\frac{S(x,1,1)}{\ell}\right)\right)$$
$$+ x^2\left(\sum_{k\geq 1}\Big(S^\circ(x^{2k},1,1,w)+\bar{S}^\circ(x^{2k+1},1,1,w)\Big)\exp\left(\sum_{\ell\geq 1}\frac{S(x^{2\ell},1,1}{2\ell}+\frac{\bar{S}(x^{2\ell-1},1,1)}{2\ell-1}\right)\right)$$
$$-\frac{x^2}{2}S^\circ(x^2,1,1,w)$$

Now, combining all equations obtained in this chapter, that is

$$\begin{aligned}
Z_{\mathcal{B}^\circ}(\mathbf{s_1},\mathbf{t_1},w) &= t_1 Z_{\hat{\mathcal{B}}}(\mathbf{s_1},w) + Z_{\mathcal{B}^\bullet}(\mathbf{s_1},\mathbf{t_1},w)\\
&= t_1\left(Z_{\mathcal{B}_{\hat{\mathcal{R}}}}(\mathbf{s_1};w)+Z_{\mathcal{B}_{\hat{\mathcal{M}}}}(\mathbf{s_1};w)-Z_{\mathcal{B}_{\mathcal{RM}}}(\mathbf{s_1};w)\right)+Z_{\mathcal{B}^c}(\mathbf{s_1},\mathbf{t_1},w)+Z_{\mathcal{B}^n}(\mathbf{s_1},\mathbf{t_1},w)\\
&= t_1\left(Z_{\mathcal{B}_{\hat{\mathcal{R}}}}(\mathbf{s_1};w)+Z_{\mathcal{B}_{\hat{\mathcal{M}}}}(\mathbf{s_1};w)-Z_{\mathcal{B}_{\mathcal{RM}}}(\mathbf{s_1};w)\right)\\
&\quad + Z_{\mathcal{B}^c_{\mathcal{R}}}(\mathbf{s_1},\mathbf{t_1},w)+Z_{\mathcal{B}^c_{\mathcal{M}}}(\mathbf{s_1},\mathbf{t_1},w)+Z_{\mathcal{B}^n_{\mathcal{R}}}(\mathbf{s_1},\mathbf{t_1},w)+Z_{\mathcal{B}^n_{\mathcal{M}}}(\mathbf{s_1},\mathbf{t_1},w),
\end{aligned}$$

we obtain one equation for $Z_{\mathcal{B}^\circ}(\mathbf{s_1},\mathbf{t_1},w)$ in terms of series-parallel networks.

B.3.3 Connected series-parallel graphs

Let $\mathcal{C}^\circ$ denote all cycle rooted series-parallel graphs, $\hat{\mathcal{C}}$ connected series-parallel graphs rooted at a cycle of length 1 (i.e. at a vertex), $\hat{\mathcal{C}}_s$ connected SP-graphs rooted at a non-separating vertex, and $C^\bullet$ those rooted at a cycle of length at least 2. Further let $C^\circ(x,w), \hat{C}(x,w), \hat{C}_s(x,w)$ and $C^\bullet(x,w)$ denote the according generating functions. Of course,

$$C^\circ(x,w) = x\hat{C}(x,w) + C^\bullet(x,w).$$

To set up the system of equations for $C^\circ(x,w)$ we will need the cycle index sum $Z_{\mathcal{B}^\circ}(\mathbf{s_1};w)$, which are given in the previous section.

Vertex-rooted connected SP-graphs

Recall the decomposition of connected graphs into 2-connected ones, given in (4.1). For our new variable w, the root degree of every 2-connected graph attached to the root contributes additively, which leads to

$$\begin{aligned}
\hat{C}(x,w) &= \exp\left(\sum_{\ell\geq 1}\frac{1}{l}Z_{\hat{\mathcal{B}}}(x^\ell\hat{C}(x^\ell,1), x^{2\ell}\hat{C}(x^{2\ell},1),\ldots;w^\ell)\right) \qquad \text{(B.7)}\\
&= \exp\left(\sum_{\ell\geq 1}\frac{1}{l}\hat{C}_s(x^\ell,w^\ell)\right),
\end{aligned}$$

in our notation, as we substituted every s_i by $x^i\hat{C}(x^i,1)$ in

$$\exp\left(\sum_{\ell\geq 1}\frac{1}{l}Z_{\hat{\mathcal{B}}}(\mathbf{s}_\ell;w^\ell)\right).$$

As only the term $x\hat{C}(x,1)$ will contribute asymptotically by methods of singularity analysis, we isolate the variable s_1, that is, we substitute $s_1 \leftarrow s$ and $s_i \leftarrow x^i\hat{C}(x^i,1)$ for $i \geq 2$. Thus we obtain a trivariate generating function $\hat{C}(s,x,w)$ given by

$$\begin{aligned}\hat{C}(s,z,w) &= \\ &= \exp\left(Z_{\hat{\mathcal{B}}}(s, z^2\hat{C}(z^2,1), z^3\hat{C}(z^3,1),\ldots; w) + \sum_{\ell\geq 2}\frac{1}{l}Z_{\hat{\mathcal{B}}}(x^\ell\hat{C}(x^\ell,1), x^{2\ell}\hat{C}(x^{2\ell},1),\ldots; w^\ell)\right) \\ &= \exp\left(\hat{B}(s,z,w) + h(z,w)\right),\end{aligned}$$

where $h(z,w)$ is some function not depending on the isolated variable s.

Symmetric cycle rooted 2-connected SP-graphs

When rooting a connected SP-graph at a cycle, the graph is definitely symmetric, with the center of the symmetry being either a single vertex (marked in the equation by ①) or a symmetric block (labelled by ②). Let $l \geq 2$ be the length of the root cycle.
If the center is a vertex, then the root cycle is part of several members of $\hat{\mathcal{C}}_s$, that is it implies a root cycle of length at least $2 \leq l_s \leq l$ on the set of structures attached to this vertex, where every vertex of the cycle is replaced by a connected graph rooted at a non-separating vertex with an additional cycle of length $l_c = \frac{l}{l_s} \geq 1 \in \mathbb{N}$ marked, $(\hat{\mathcal{C}}_s)^\circ$, and every other vertex of the set is replaced by an "ordinary" member of $\hat{\mathcal{C}}_s$.
If the center of symmetry is a block, then there is a cycle of length $2 \leq l_b \leq l$ marked in this block, where every vertex of the root cycle is replaced by a rooted connected graph with either the root itself marked (if $l = l_b$) or an additional cycle of length $l_c = \frac{l}{l_b} \geq 1 \in \mathbb{N}$ different from the root marked. This procedure corresponds to the symbolic equation

$$(\mathcal{X} * \hat{\mathcal{C}})^\circ = \hat{\mathcal{C}} + \mathcal{X} * \hat{\mathcal{C}}^\circ,$$

as described in more detail in [5]. In the first case the degree of w is given by both the degree of the vertex in the block and the degree as a root of a connected graph, in the second case the degree of the substituted vertex in the block does not matter.
Finally, we obtain for $C^\bullet(x,w)$:

$$\begin{aligned}C^\bullet(x,w) &= \underbrace{xZ_{Set^\bullet}(s_i \leftarrow \hat{C}_s(x^i,1), t_i \leftarrow \hat{C}_s^\circ(x^i,w))}_{①} \\ &\quad + \underbrace{Z_{\mathcal{B}^\bullet}[s_i \leftarrow x^i\hat{C}(x^i,1), t_i \leftarrow x^i\hat{C}(x^i,w); w] + Z_{\mathcal{B}^\bullet}[s_i \leftarrow x^i\hat{C}(x^i,1), t_i \leftarrow x^i\hat{C}^\circ(x^i,w); 1]}_{②} \\ &= \exp\left(\sum_{\ell\geq 1}\frac{\hat{C}_s(x^\ell,1)}{\ell}\right)\left(\sum_{k\geq 2}\hat{C}_s^\circ(x^k,w)\right) \\ &\quad + Z_{\mathcal{B}^\bullet}[s_i \leftarrow x^i\hat{C}(x^i,1), t_i \leftarrow x^i\hat{C}(x^i,w); w] + Z_{\mathcal{B}^\bullet}[s_i \leftarrow x^i\hat{C}(x^i,1), t_i \leftarrow x^i\hat{C}^\circ(x^i,w); 1]\end{aligned}$$

To finish, we need equations for $\hat{C}_s^\circ(x,w)$ and $\hat{C}^\circ(x,w)$, where the same ideas as above are used.

$$\hat{C}^{\circ}(x,w) = \sum_{\ell \geq 1} \hat{C}_s^{\circ}(x^\ell, w) \exp\left(\sum_{\ell \geq 1} \frac{1}{\ell} \hat{C}_s(x^\ell, 1) \right)$$

$$\hat{C}_s^{\circ}(x,w) = Z_{\hat{\mathcal{B}}^{\circ}}(s_i \leftarrow x^i \hat{C}(x^i,1), t_i \leftarrow x^i \hat{C}(x^i,w); w) + Z_{\hat{\mathcal{B}}^{\circ}}(s_i \leftarrow x^i \hat{C}(x^i,1), t_i \leftarrow x^i \hat{C}^{\circ}(x^i,w); 1)$$

Bibliography

[1] R. Beaza-Yates. Fringe analysis revisited. *ACM Computing Surveys*, pages 109–119, 1995.

[2] F. Bergeron, G. Labelle, and P. Leroux. *Combinatorial Species and Tree-like Structures.* Cambridge University Press, 1998.

[3] N. Bernasconi, K. Panagiotou, and A. Steger. The degree sequence of random graphs from subcritical classes. *Combinatorics, Probability and Computing*, 18(5):647–681, 2009.

[4] P. Billingsley. *Convergence of probability measures.* John Wiley & Sons Inc., New York, 1968.

[5] M. Bodirsky, E. Fusy, M. Kang, and S. Vigerske. Enumeration of unlabeled outerplanar graphs. *Electronic Journal of Combinatorics*, 14(1, research paper 66), 2007.

[6] M. Bodirsky, E. Fusy, M. Kang, and S. Vigerske. Maximal biconnected subgraphs of random planar graphs. In *SODA '07 Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 356–365, 2007.

[7] M. Bodirsky, O. Giménez, M. Kang, and M. Noy. Enumeration and limit laws for series-parallel graphs. *European J. Combin.*, 28(8):2091–2105, 2007.

[8] M. Bóna and P. Flajolet. Isomorphism and symmetries in random phylogenetic trees. *Journal of Applied Probability*, 46(4):1005–1019, 2009.

[9] N. Bonichon, C. Gavoille, N. Hanusse, D. Poulalhon, and G. Schaeffer. Planar graphs, via well-orderly maps and trees. *Graphs Combin.*, 22(2):185–202, 2006.

[10] G. Chapuy, E. Fusy, M. Kang, and B. Shoilekova. A complete grammar for decomposing a family of graphs into 3-connected components. *Electron. J. Combin.*, 15(1):R148, 2008.

[11] B. Chauvin, P. Flajolet, D. Gardy, and B. Gittenberger. And/or trees revisited. *Combinatorics, Probability and Computing*, 13(4-5):475–497, 2004.

[12] F. Chyzak, M. Drmota, T. Klausner, and G. Kok. The distribution of patterns in random trees. *Combin. Probab. Comput.*, 17(1):21–59, 2008.

[13] J. W. Cohen and G. Hooghiemstra. Brownian excursion, the $M/M/1$ queue and their occupation times. *Math. Oper. Res.*, 6(4):608–629, 1981.

[14] M. Drmota. Systems of functional equations. *Random Structures and Algorithms*, 10:103–124, 1997.

[15] M. Drmota. *Random Trees - An interplay between combinatorics and probability.* Springer - Wien, NewYork, 2009.

[16] M. Drmota, E. Fusy, M. Kang, V. Kraus, and J. Rue. Asymptotic study of subcritical graph families. to appear in SIAM Journal on Discrete Mathematics. Available at http://arxiv.org/abs/1003.4699, 2010.

[17] M. Drmota, O. Giménez, and M. Noy. Degree distribution in random planar graphs. to appear in Journal of Combinatorial Theory, manuscript available at *http://www.dmg.tuwien.ac.at/drmota/DegreeDist.pdf.*

[18] M. Drmota, O. Giménez, and M. Noy. Vertices of given degree in series-parallel graphs. *Random Structures Algorithms*, 36(3):273–314, 2008.

[19] M. Drmota and B. Gittenberger. On the profile of random trees. *Random Structures and Algorithms*, 10(4):421–451, 1997.

[20] M. Drmota and B. Gittenberger. The distribution of nodes of given degree in random trees. *Journal of Graph theory*, 3(31):227–253, 1999.

[21] M. Drmota and B. Gittenberger. The shape of unlabeled rooted random trees. *European J. Combinat.*, 31:2028–2063, 2010.

[22] M. Drmota, B. Gittenberger, A. Panholzer, H. Prodinger, and M. Ward. On the shape of the fringe of various types of random trees. *Math. Meth. Appl. Sci.*, 32:1207–1245, 2009.

[23] P. Flajolet and A.Odlyzko. Singularity analysis of generating functions. *SIAM Journal on discrete mathematics*, 3(2):216–240, 1990.

[24] P. Flajolet and R. Sedgewick. *Analytic Combinatorics.* Cambridge U.P., Cambridge, 2009.

[25] H. Fournier, D. Gardy, A. Genitrini, and B. Gittenberger. The fraction of large random trees representing a given Boolean function in implicational logic. To appear in Random Structures Algorithms, 2011.

[26] D. Gardy. Random Boolean expressions. In *Colloquium on Computational Logic and Applications*, volume AF, pages 1–36. DMTCS Proceedings, 2006.

[27] A. Genitrini and B. Gittenberger. No Shannon effect on probability distributions on Boolean functions induced by Boolean expressions. In *Analysis of Algorithms*, Wien, Austria, Juillet 2010. DMTCS proceedings.

[28] A. Genitrini, B. Gittenberger, V. Kraus, and C. Mailler. Associative and commutative tree representations for Boolean functions. submitted.

[29] S. Gerke and C. McDiarmid. On the number of edges in random planar graphs. *Combin. Probab. Comput.*, 13(2):165–183, 2004.

[30] O. Giménez and M. Noy. Asymptotic enumeration and limit laws of planar graphs. *J. Amer. Math. Soc.*, 22:309–329, 2009.

[31] O. Giménez and M. Noy. Counting planar graphs and related families of graphs. In *Surveys in combinatorics 2009*, pages 169–210. Cambridge University press, 2009.

[32] O. Giménez, M. Noy, and J. Rué. Graph classes with given 3-connected components: asymptotic counting, limit laws and critical phenomena. In *Sixth Conference on Discrete Mathematics and Computer Science (Spanish)*, pages 369–376. Univ. Lleida, Lleida, 2008.

[33] B. Gittenberger. Convergence of branching processes to the local time of a Bessel process. In *Proceedings of the Eighth International Conference "Random Structures and Algorithms" (Poznan, 1997)*, volume 13, pages 423–438, 1998.

[34] B. Gittenberger. On the contour of random trees. *SIAM Journal Discrete Math.*, 12(4):434–458, 1999.

[35] B. Gittenberger. On the profile of random forests. In *Mathematics and computer science, II (Versailles, 2002)*, Trends Math., pages 279–293. Birkhäuser, Basel, 2002.

[36] B. Gittenberger. Nodes of large degree in random trees and forests. *Random Structures Algorithms*, 28(3):374–385, 2006.

[37] B. Gittenberger and V. Kraus. The degree profile of Pólya trees. submitted.

[38] B. Gittenberger and G. Louchard. The Brownian excursion multi-dimensional local time density. *J. Appl. Probab.*, 36(2):350–373, 1999.

[39] B. Gittenberger and G. Louchard. On the local time density of the reflecting Brownian bridge. *J. Appl. Math. Stochastic Anal.*, 13(2):125–136, 2000.

[40] F. Harary and E. Palmer. *Graphical Enumeration.* Academic Press New York and London, 1973.

[41] H. Heuser. *Lehrbuch der Analysis - Teil 2.* B.G. Teubner Stuttgart-Leipzig-Wiesbaden, 11 edition, 2000.

[42] G. Hooghiemstra. On the occupation time of Brownian excursion. *Electron. Comm. Probab.*, 4:61–64 (electronic), 1999.

[43] H.-K. Hwang. *Théorèmes limites pour les structures combinatoires et les fonctions arithmétiques.* PhD thesis, École Polytechnique, 1994.

[44] K. M. Jansons. The distribution of time spent by a standard excursion above a given level, with applications to ring polymers near a discontinuity in potential. *Electron. Comm. Probab.*, 2:53–58 (electronic), 1997.

[45] I. Karatzas and S. Shreve. *Brownian motion and stochastic calculus*, volume 113 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1988.

[46] V. Kolchin. *Random Mappings*. Springer New York, 1986.

[47] J. Kozik. Subcritical pattern languages for and/or trees. In *DMTCS proceedings from Fifth Colloquium on Mathematics and Computer Science Algorithms, Trees, Combinatorics and Probabilities*, pages 437–448, 2008.

[48] V. Kraus. *Diverse families of rooted random trees - a compilation of characteristics*. Diploma Thesis, 2008. available at *http://media.obvsg.at/AC05037032-2001*.

[49] V. Kraus. The degree distribution in unlabelled 2-connected graph families. In *AofA'10 Proceedings of the 21st International Meeting on Probabilistic, Combinatorial and Asymptotic Methods for the Analysis of Algorithms*, pages 455–474, 2010.

[50] S. P. Lalley. Finite range random walk on free groups and homogeneous trees. *Ann. Probab.*, 21(4):2087–2130, 1993.

[51] H. Lefmann and P. Savický. Some typical properties of large And/Or Boolean formulas. *Random Structures and Algorithms*, 10:337–351, 1997.

[52] J.-F. Marckert and G. Miermont. The CRT is the scaling limit of unordered binary trees. *Random Structures and Algorithms*, 38(3):1–35, 2011.

[53] C. McDiarmid, A. Steger, and D. Welsh. Random planar graphs. *J.Combin.Theory*, 93:187–205, 2005.

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

[55] A. Meir and J. W. Moon. On nodes of large out-degree in random trees. In *Proceedings of the Twenty-second Southeastern Conference on Combinatorics, Graph Theory, and Computing (Baton Rouge, LA, 1991)*, volume 82, pages 3–13, 1991.

[56] M. Noy. Random planar graphs and the number of planar graphs. In *Combinatorics, complexity, and chance*, volume 34 of *Oxford Lecture Ser. Math. Appl.*, pages 213–233. Oxford Univ. Press, Oxford, 2007.

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

[58] K. Panagiotou and A. Steger. Maximal biconnected subgraphs of random planar graphs. In *Proceedings of the 20th Annual ACM-SIAM Symposium on Discrete Algorithms(SODA)*, pages 432–440, 2009.

[59] K. Panagiotou and A. Weißl. Properties of random graphs via Boltzmann samplers. In *2007 Conference on Analysis of Algorithms, AofA 07*, Discrete Math. Theor. Comput. Sci. Proc., AH, pages 159–168. Assoc. Discrete Math. Theor. Comput. Sci., Nancy, 2007.

[60] G. Pólya. Kombinatorische Anzahlbestimmungen für Gruppen, Graphen und chemische Verbindungen. *Acta Mathematica*, 68:145–254, 1937.

[61] D. Revuz and M. Yor. *Continuous martingales and Brownian motion*, volume 293 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1991.

[62] R. Robinson and A. Schwenk. The distribution of degrees in a large random tree. *Discrete Mathematics*, 12:359–372, 1975.

[63] A. J. Schwenk. An asymptotic evaluation of the cycle index of a symmetric group. *Discrete Math.*, 18(1):71–78, 1977.

[64] B. Shoilekova. Unlabelled enumeration of cacti graphs. Manuscript, 2007.

[65] L. Takács. On the local time of the Brownian motion. *Ann. Appl. Probab.*, 5(3):741–756, 1995.

[66] L. Takács. On the local time of the Brownian bridge. In *Applied probability and stochastic processes*, volume 19 of *Internat. Ser. Oper. Res. Management Sci.*, pages 45–62. Kluwer Acad. Publ., Boston, MA, 1999.

[67] W. Tutte. *Connectivity in graphs.* Oxford U.P, 1966.

[68] R. van der Hofstad, G. Hooghiemstra, and P. Van Mieghem. On the covariance of the level sizes in random recursive trees. *Random Structures Algorithms*, 20(4):519–539, 2002.

[69] M. Ward. *Analysis of the multiplicity matching parameter in suffix tries.* Ph.D. Thesis, Purdue University, West Lafayette, IN, USA, 2005.

[70] A. Woods. On the probability of absolute truth for And/Or Boolean formulas. *The Bulletin of Symbolic Logic*, 3(12), 2006.

[71] A. R. Woods. Coloring rules for finite trees, and probabilities of monadic second order sentences. *Random Structures Algorithms*, 10(4):453–485, 1997.

[72] A. C. C. Yao. On random $2-3$ trees. *Acta Informat.*, 9(2):159–170, 1977/78.

Index