35th International Conference on Probabilistic, Combinatorial and Asymptotic Methods for the Analysis of Algorithms

AofA 2024, June 17–21, 2024, University of Bath, UK

Edited by
Cécile Mailler
Sebastian Wild

LIPICS

Editors

Cécile Mailler (ID)
University of Bath, UK
c.mailler@bath.ac.uk

Sebastian Wild (ID)
University of Liverpool, UK
wild@liverpool.ac.uk

LIPIcs – Leibniz International Proceedings in Informatics

LIPIcs is a series of high-quality conference proceedings across all fields in informatics. LIPIcs volumes are published according to the principle of Open Access, i.e., they are available online and free of charge.

ISSN 1868-8969

https://www.dagstuhl.de/lipics

Dedicated to the memory of Philippe Flajolet.

Contents

Regular Papers

◼ Preface

The 35th International Conference on Probabilistic, Combinatorial and Asymptotic Methods for the Analysis of Algorithms (AofA 2024) was held at the University of Bath, UK, during June 17 – 21, 2024.

Analysis of algorithms is a scientific basis for quantifying the efficiency of computation, providing a link between abstract algorithms and the performance characteristics of their implementations in the real world. The general effort to predict precisely the performance of algorithms and data structures – the amount of time, storage, or other resources needed – has produced mathematical methods of sweeping generality that unify and simplify making such predictions in a rigorous way, as well as software tools supporting their application. In enabling this progress, AofA has come to involve research in analytic combinatorics, the analysis of random discrete structures, asymptotic analysis, exact and limiting distributions, and other fields of inquiry in computer science, probability theory, and enumerative combinatorics. See the AofA community websites for more details: `https://www.math.aau.at/AofA/`.

The Call for Papers invited papers in

- analytic algorithmics and combinatorics,
- probabilistic analysis of algorithms,
- randomized algorithms.

We also welcomed papers addressing problems such as: combinatorial algorithms, string searching and pattern matching, sublinear algorithms on massive data sets, network algorithms, graph algorithms, caching and memory hierarchies, indexing, data mining, data compression, coding and information theory, and computational finance. Papers were also welcomed that address bridges to research in related fields such as statistical physics, computational biology, computational geometry, and simulation.

The conference program featured the 2024 Philippe Flajolet Lecture by Michael Drmota, 30 contributed papers, which are collected in this volume, as well as 8 invited lectures:

- Antoine Genitrini (Sorbonne Université): "Varieties of Trees with Constrained Labelings"
- Leslie Goldberg (University of Oxford): "The Complexity of Approximate Counting"
- Daniel Krenn (Paris Lodron University of Salzburg): "Multi-pivot quicksort and how to compute precise asymptotics"
- László Kozma (Freie Universität Berlin): "Analysis of algorithms via extremal combinatorics"
- Alessandra Caraceni (Scuola Normale Superiore Pisa): "Growing random geometries: making trees blossom and triangulations flip"
- Markus Lohrey (Universität Siegen): "Grammar-based tree compression: combinatorics and algorithms"
- Thomas Sauerwald (University of Cambridge): "Balanced Allocations: The Power of Choice versus Noise"
- Sylvie Corteel (CNRS and Université Paris Cité): "Combinatorics of k-tilings"

As a typical convention in the field is to list authors in alphabetical order, we randomized the order of papers in these proceedings to avoid biases; the same order was used for the conference, so the proceedings volume gives paper in chronological order of presentation.
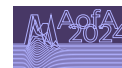
Flajolet Lecture

The *Philippe Flajolet Lecture Prize* for outstanding contributions to analytic combinatorics and analysis of algorithms is awarded every two years by the Analysis of Algorithms (AofA) community.

At the AofA 2024 conference, Michael Drmota presented the sixth Flajolet Lecture, entitled "The Moment Method Revisited". Previous Flajolet Lectures Prize recipients are Donald E. Knuth, Robert Sedgewick, Luc Devroye, Wojciech Szpankowski, and Svante Janson.

The prize is named in honor and recognition of the extraordinary accomplishments of the late Philippe Flajolet and his formative influence on the growth and flourishing of the AofA community. Philippe spent most of his scientific life at INRIA, France. He is best known for fundamental advances in mathematical methods for the analysis of algorithms. His research laid the foundation of a subfield of mathematics now known as analytic combinatorics. Analytic combinatorics is a modern basis for the quantitative study of combinatorial structures (such as words, trees, mappings, and graphs), with applications to probabilistic study of algorithms that are based on these structures. It also strongly influences research in other scientific domains, such as statistical physics, computational biology, and information theory. Flajolet's work takes the field forward by introducing original approaches in combinatorics based on two types of methods: symbolic and analytic. The symbolic side is based on the automation of decision procedures in combinatorial enumeration to derive characterizations of generating functions. The analytic side treats those functions as functions in the complex plane and leads to precise characterization of limit distributions. Beyond these foundational contributions, Philippe's research opened new avenues in various domains of applied computer science, including streaming algorithms, communication protocols, database access methods, data mining, symbolic manipulation, text-processing algorithms, and random generation.

Cécile Mailler and Sebastian Wild,
on behalf of the Program and Steering Committees

AofA 2024 Program Committee

- Cécile Mailler (co-chair), University of Bath, UK
- Sebastian Wild (co-chair), University of Liverpool, UK
- Frédérique Bassino, Université Sorbonne Paris Nord, France
- Gabriel Berzunza, University of Liverpool, UK
- Benjamin Doerr, École Polytechnique, France
- Laura Eslava, UNAM, Mexico
- Hsien-Kuei Hwang, Academia Sinica, Taiwan
- Conrado Martínez, Universitat Politècnica de Catalunya, Spain
- Stephen Melczer, University of Waterloo, Canada
- Ralph Neininger, Goethe-Universität Frankfurt, Germany
- Cyril Nicaud, Université Gustave Eiffel, France
- Pablo Rotondo, Université Gustave Eiffel, France
- Fiona Skerman, Uppsala University, Sweden
- Andrea Sportiello, Université Sorbonne Paris Nord, France
- Benedikt Stufler, TU Wien, Austria
- Michael Wallner, TU Wien, Austria

AofA Steering Committee

- Bruno Salvy (chair), INRIA, Lyon, France
- Frédérique Bassino, Université Sorbonne Paris Nord, France
- James Allen Fill, Johns Hopkins University, USA
- Clemens Heuberger, Alpen-Adria-Universität Klagenfurt, Austria
- Cecilia Holmgren, Uppsala Universitet, Sweden
- Mark Daniel Ward, Purdue University, IN, USA

Fringe Trees for Random Trees with Given Vertex Degrees

Gabriel Berzunza Ojeda ✉ ⌂
Department of Mathematical Sciences, University of Liverpool, UK

Cecilia Holmgren ✉ ⌂
Department of Mathematics, Uppsala University, Sweden

Svante Janson ✉ ⌂
Department of Mathematics, Uppsala University, Sweden

──── **Abstract** ────────────────────────────────

We prove that the number of fringe subtrees, isomorphic to a given tree, in uniformly random trees with given vertex degrees, asymptotically follows a normal distribution. As an application, we establish the same asymptotic normality for random simply generated trees (conditioned Galton-Watson trees). Our approach relies on an extension of Gao and Wormald's (2004) theorem to the multivariate setting.

1 Introduction and main results

In this paper, we consider fringe trees of random plane trees with given vertex statistics, i.e., a given number of vertices of each degree. As an application, we also give corresponding result for random simply generated trees (or conditioned Galton–Watson trees). The main results are laws of large numbers and central limit theorems for the number of fringe trees of a given type.

Let $\mathbb{T}$ be the set of all (finite) plane rooted trees (also called ordered rooted trees); see e.g., [9]. Denote the size, i.e. the number of vertices, of a tree T by $|T|$. The (out)degree of a vertex $v \in T$, denoted $d_T(v)$, is its number of children in T; thus leaves have degree 0 and all other vertices have strictly positive degree. The *degree statistic* of a rooted tree T is the sequence $\mathbf{n}_T = (n_T(i))_{i \geq 0}$, where $n_T(i) := |\{v \in T : d_T(v) = i\}|$ is the number of vertices of T with i children. We have

$$|T| = \sum_{i \geq 0} n_T(i) = 1 + \sum_{i \geq 0} i n_T(i). \tag{1}$$

A sequence $\mathbf{n} = (n(i))_{i \geq 0}$ is the degree statistic of some tree if and only if $\sum_{i \geq 0} n(i) = 1 + \sum_{i \geq 0} i n(i)$. For such sequences, we let $|\mathbf{n}| := \sum_{i \geq 0} n(i)$ be the size of $\mathbf{n}$, and we write $\mathbb{T}_\mathbf{n}$ for the set of plane rooted trees with degree statistic $\mathbf{n}$. We let $\mathcal{T}_\mathbf{n}$ be a uniformly random

element of the set $\mathbb{T}_{\mathbf{n}}$, and we denote this by $\mathcal{T}_{\mathbf{n}} \sim \mathrm{Unif}(\mathbb{T}_{\mathbf{n}})$. It is also well known that the total number of plane rooted trees with degree statistic $\mathbf{n}$ is given by (see [23, Exercise 6.2.1])

$$|\mathbb{T}_{\mathbf{n}}| = \frac{1}{|\mathbf{n}|}\binom{|\mathbf{n}|}{\mathbf{n}} = \frac{1}{|\mathbf{n}|}\frac{|\mathbf{n}|!}{\prod_{i \geq 0} n(i)!}. \tag{2}$$

For $T \in \mathbb{T}$ and a vertex $v \in T$, let T_v be the subtree of T rooted at v consisting of v and all its descendants. We call T_v a fringe (sub)tree of T. We regard T_v as an element of $\mathbb{T}$ and let, for $T, T' \in \mathbb{T}$,

$$N_{T'}(T) := |\{v \in T : T_v = T'\}| = \sum_{v \in T}\mathbf{1}_{\{T_v = T'\}}, \tag{3}$$

i.e., the number of fringe subtrees of T that are equal (i.e., isomorphic to) to T'. A random fringe subtree T^{fr} of $T \in \mathbb{T}$ is the random rooted tree obtained by taking the fringe subtree T_v at a uniform random vertex $v \in T$. Thus, the distribution of T^{fr} is given by

$$\mathbb{P}(T^{\mathrm{fr}} = T') = \frac{N_{T'}(T)}{|T|}, \quad \text{for } T' \in \mathbb{T}. \tag{4}$$

We prove an asymptotic result on the distribution of a random fringe subtree in a random rooted plane tree with a given degree statistic. In order to state the theorem, we need a little more terminology. (See also Section 1.2 for some notation.) For a degree statistic $\mathbf{n}$, denote by $\mathbf{p}(\mathbf{n}) = (p_i(\mathbf{n}))_{i \geq 0}$ its (empirical) degree distribution, i.e.,

$$p_i(\mathbf{n}) := \frac{n(i)}{|\mathbf{n}|}, \quad \text{for } i \geq 0. \tag{5}$$

In this paper, we assume for convenience the following condition.

▶ **Condition 1.** $\mathbf{n}_\kappa = (n_\kappa(i))_{i \geq 0}$, $\kappa \geq 1$, *are degree statistics such that as* $\kappa \to \infty$:
  (i) $|\mathbf{n}_\kappa| \to \infty$,
  (ii) *For every* $i \geq 0$, *we have* $p_i(\mathbf{n}_\kappa) \to p_i$, *where* $\mathbf{p} = (p_i)_{i \geq 0}$ *is a probability distribution on* $\mathbb{N}_0$.

▶ Remark 2. The condition that $\mathbf{p}$ is a probability distribution is no restriction. In fact, the degree distribution $\mathbf{p}(\mathbf{n}_\kappa)$ has mean

$$\sum_{i \geq 0} i p_i(\mathbf{n}_\kappa) = \frac{1}{|\mathbf{n}_\kappa|}\sum_{i \geq 0} i n_\kappa(i) = \frac{|\mathbf{n}_\kappa| - 1}{|\mathbf{n}_\kappa|} < 1, \tag{6}$$

and thus the sequence of distributions $\mathbf{p}(\mathbf{n}_\kappa)$ is always tight. Hence, if $p_i(\mathbf{n}_\kappa) \to p_i$, for every $i \geq 0$, then $\mathbf{p} = (p_i)_{i \geq 0}$ is a probability distribution. Note also that **(ii)** says that $\mathbf{p}(\mathbf{n}_\kappa)$ converges weakly to $\mathbf{p}$, as $\kappa \to \infty$. (As is well known, this is equivalent to convergence in total variation.)

By (6) and Fatou's lemma, if Condition 1 holds, then $\sum_{i \geq 0} i p_i \leq 1$. Conversely, it is easily seen that any such probability distribution $\mathbf{p}$ is the limit of $\mathbf{p}(\mathbf{n}_\kappa)$ for some sequence of degree statistics $\mathbf{n}_\kappa$. In other words, the set of probability distributions $\mathbf{p}$ that can appear as limits in Condition 1 is precisely the set of probability distributions $\mathbf{p}$ on $\mathbb{N}_0$ with mean $\sum_{i \geq 0} i p_i \leq 1$; we denote this set by $\mathcal{P}_1(\mathbb{N}_0)$.

For a probability distribution $\mathbf{p} = (p_i)_{i \geq 0} \in \mathcal{P}_1(\mathbb{N}_0)$, let $\mathcal{T}_{\mathbf{p}}$ be a Galton–Watson tree with offspring distribution $\mathbf{p}$, and define $\pi_{\mathbf{p}}$ as the distribution of $\mathcal{T}_{\mathbf{p}}$, i.e., (with $0^0 := 1$ as usual)

$$\pi_{\mathbf{p}}(T) := \mathbb{P}(\mathcal{T}_{\mathbf{p}} = T) = \prod_{i \geq 0} p_i^{n_T(i)} = \prod_{i \in \mathcal{D}(T)} p_i^{n_T(i)}, \quad \text{for } T \in \mathbb{T}, \tag{7}$$

where

$$\mathcal{D}(T) := \{i : n_T(i) > 0\} = \{d_T(v) : v \in T\}, \tag{8}$$

the set of degrees that appear in T. Note that $\pi_{\mathbf{p}}(T) = 0 \iff p_i = 0$ for some $i \in \mathcal{D}(T)$. In particular, if $\mathbf{n}_\kappa$ and $\mathbf{p}$ are as in Condition 1, then $\pi_{\mathbf{p}}(T) = 0$ if and only if $n_\kappa(i) = o(|\mathbf{n}_\kappa|)$ for some $i \in \mathcal{D}(T)$.

We first give a law of large numbers for the number of fringe trees of a given type in a random rooted plane tree with a given degree statistic. The proofs of this and the following theorem are given in later sections.

▶ **Theorem 3.** *Let* $\mathbf{n}_\kappa$, $\kappa \geq 1$, *be some degree statistics that satisfy Condition 1, and let* $\mathcal{T}_{\mathbf{n}_\kappa} \sim \mathrm{Unif}(\mathbb{T}_{\mathbf{n}_\kappa})$. *For every fixed* $T \in \mathbb{T}$, *as* $\kappa \to \infty$:

  (i) *(Annealed version)* $\quad \mathbb{P}(\mathcal{T}_{\mathbf{n}_\kappa}^{\mathrm{fr}} = T) = \dfrac{\mathbb{E}[N_T(\mathcal{T}_{\mathbf{n}_\kappa})]}{|\mathbf{n}_\kappa|} \to \pi_{\mathbf{p}}(T)$.

  (ii) *(Quenched version)* $\quad \mathbb{P}(\mathcal{T}_{\mathbf{n}_\kappa}^{\mathrm{fr}} = T \mid \mathcal{T}_{\mathbf{n}_\kappa}) = \dfrac{N_T(\mathcal{T}_{\mathbf{n}_\kappa})}{|\mathbf{n}_\kappa|} \to \pi_{\mathbf{p}}(T)$ *in probability.*

In other words, the random fringe tree converges in distribution as $\kappa \to \infty$: **(i)** *says* $\mathcal{T}_{\mathbf{n}_\kappa}^{\mathrm{fr}} \xrightarrow{\mathrm{d}} \mathcal{T}_{\mathbf{p}}$, *or equivalently* $\mathcal{L}(\mathcal{T}_{\mathbf{n}_\kappa}^{\mathrm{fr}}) \to \mathcal{L}(\mathcal{T}_{\mathbf{p}})$, *and* **(ii)** *is the conditional version* $\mathcal{L}(\mathcal{T}_{\mathbf{n}_\kappa}^{\mathrm{fr}} \mid \mathcal{T}_{\mathbf{n}_\kappa}) \xrightarrow{\mathrm{p}} \mathcal{L}(\mathcal{T}_{\mathbf{p}})$.

▶ Remark 4. Similar results are known for several other models of random trees. In particular, a version of Theorem 3 was proved by Aldous [2] for conditioned Galton–Watson trees with finite offspring variance; this was extended to general simply generated trees in [19, Theorem 7.12]. In those cases, the degree statistic is random, but Condition 1 holds in probability, with a non-random limiting probability distribution $\mathbf{p}$. We return to simply generated trees in Section 5. Another standard example is family trees of Crump–Mode–Jagers branching processes (which includes e.g. random recursive trees, binary search trees and preferential attachment trees); see e.g. [2] and [17, Theorem 5.14].

Theorem 3 is thus a law of large numbers for the number of fringe trees of a given type. In this work, we also study the fluctuations and prove a central limit theorem for this number; we furthermore show that this holds jointly for different types of fringe trees.

For a probability distribution $\mathbf{p} = (p_i)_{i \geq 0} \in \mathcal{P}_1(\mathbb{N}_0)$ and $T, T' \in \mathbb{T}$, let

$$\eta_{\mathbf{p}}(T, T') := (|T| - 1)(|T'| - 1) - \sum_{i \geq 0} \frac{n_T(i) n_{T'}(i)}{p_i}, \tag{9}$$

where we interpret $0/0 := 0$, and, for $T \neq T'$,

$$\gamma_{\mathbf{p}}(T, T) := \pi_{\mathbf{p}}(T) + \eta_{\mathbf{p}}(T, T)(\pi_{\mathbf{p}}(T))^2, \tag{10}$$

$$\gamma_{\mathbf{p}}(T, T') := N_{T'}(T)\pi_{\mathbf{p}}(T) + N_T(T')\pi_{\mathbf{p}}(T') + \eta_{\mathbf{p}}(T, T')\pi_{\mathbf{p}}(T)\pi_{\mathbf{p}}(T'). \tag{11}$$

Note that $\eta_{\mathbf{p}}(T, T') = -\infty$ if $p_i = 0$ for some $i \in \mathcal{D}(T) \cap \mathcal{D}(T')$. In this case, $\pi_{\mathbf{p}}(T) = \pi_{\mathbf{p}}(T') = 0$, and we interpret $\infty \cdot 0 := 0$ in (10)–(11); thus $\gamma_{\mathbf{p}}(T, T')$ is always finite.

▶ **Theorem 5.** *Let* $\mathbf{n}_\kappa$, $\kappa \geq 1$, *be some degree statistics that satisfy Condition 1 and let* $\mathcal{T}_{\mathbf{n}_\kappa} \sim \mathrm{Unif}(\mathbb{T}_{\mathbf{n}_\kappa})$. *For a fixed* $m \geq 1$, *let* $T_1, \ldots, T_m \in \mathbb{T}$ *be a fixed sequence of rooted plane trees. Then, as* $\kappa \to \infty$,

$$\mathbb{E}\, N_{T_i}(\mathcal{T}_{\mathbf{n}_\kappa}) = \pi_{\mathbf{p}}(T_i)|\mathbf{n}_\kappa| + o(|\mathbf{n}_\kappa|), \tag{12}$$

$$\mathrm{Var}(N_{T_i}(\mathcal{T}_{\mathbf{n}_\kappa})) = \gamma_{\mathbf{p}}(T_i, T_i)|\mathbf{n}_\kappa| + o(|\mathbf{n}_\kappa|), \tag{13}$$

$$\mathrm{Cov}\big(N_{T_i}(\mathcal{T}_{\mathbf{n}_\kappa}), N_{T_j}(\mathcal{T}_{\mathbf{n}_\kappa})\big) = \gamma_{\mathbf{p}}(T_i, T_j)|\mathbf{n}_\kappa| + o(|\mathbf{n}_\kappa|), \tag{14}$$

for $1 \leq i, j \leq m$, and

$$\left(\frac{N_{T_j}(\mathcal{T}_{\mathbf{n}_\kappa}) - \mathbb{E}[N_{T_j}(\mathcal{T}_{\mathbf{n}_\kappa})]}{\sqrt{|\mathbf{n}_\kappa|}} \right)_{j=1}^m \xrightarrow{\mathrm{d}} \mathrm{N}(0, \Gamma_{\mathbf{p}}), \tag{15}$$

where the covariance matrix is defined by $\Gamma_{\mathbf{p}} := (\gamma_{\mathbf{p}}(T_i, T_j))_{i,j=1}^m$. Furthermore, in (15), we can replace $\mathbb{E}[N_{T_i}(\mathcal{T}_{\mathbf{n}_\kappa})]$ by $|\mathbf{n}_\kappa| \pi_{\mathbf{p}(\mathbf{n}_\kappa)}(T_i)$.

If $T \in \mathbb{T}$ with $\pi_{\mathbf{p}}(T) > 0$ and $|T| > 1$, then $\gamma_{\mathbf{p}}(T, T) > 0$ and thus (13) and (15) (with $m = 1$) show that $N_T(\mathcal{T}_{\mathbf{n}_\kappa})$ is asymptotically normal, with

$$\frac{N_T(\mathcal{T}_{\mathbf{n}_\kappa}) - \mathbb{E}[N_T(\mathcal{T}_{\mathbf{n}_\kappa})]}{\sqrt{\mathrm{Var}(N_T(\mathcal{T}_{\mathbf{n}_\kappa}))}} \xrightarrow{\mathrm{d}} \mathrm{N}(0, 1), \qquad \kappa \to \infty. \tag{16}$$

The case $|T| = 1$ is trivial, with $N_T(\mathcal{T}_{\mathbf{n}_\kappa}) = n_\kappa(0)$ non-random. Ignoring this case, Theorem 5 shows that $N_T(\mathcal{T}_{\mathbf{n}_\kappa})$ is asymptotically normal when $\pi_{\mathbf{p}}(T) > 0$. On the other hand, if $\pi_{\mathbf{p}}(T) = 0$, then also $\gamma_{\mathbf{p}}(T, T) = 0$, and the theorems above do not give precise information on the asymptotic distribution of $N_T(\mathcal{T}_{\mathbf{n}_\kappa})$. In this case, [3, Theorem 1.7] in the full version is more precise.

In the case of critical conditioned Galton–Watson trees with finite offspring variance, (joint) normal convergence of the subtree counts in analogy to (15) was proved in [20, Corollary 1.8] (together with convergence of mean and variance). Indeed, [20, Theorem 1.5] proved, more generally, asymptotic normality of additive functionals that are defined via toll functions (under some conditions); see [3, Section 8] in the full version for further discussion on additive functionals.

▶ Remark 6. Results on asymptotic normality for fringe tree counts have also been proved earlier for several other classes of random trees. For example, for binary search trees see [7], [8], [6], [12], [16]; for random recursive trees see [11], [16]; for increasing trees see [13]; for m-ary search trees and preferential attachment trees see [18]; for random tries see [21].

Our approach relies on a multivariate version of the Gao–Wormald theorem [14, Theorem 1]; see [3, Theorem A.1]. The original Gao–Wormald theorem [14] provides a way to show asymptotic normality by analysing the behaviour of sufficiently high factorial moments. (Typically, factorial moments are more convenient than standard moments in combinatorics.) The multivariate version [3, Theorem A.1] extends this by considering joint factorial moments. In our framework, this is very convenient since we can precisely compute the joint factorial moments of the subtree counts in (3) for random trees with given degree statistics. (Another, closely related, multivariate version of the Gao–Wormald theorem has independently been shown recently by Hitczenko and Wormald [15].)

The (one dimensional) Gao–Wormald theorem has been used before by Cai and Devroye [5] to study large fringe trees in critical conditioned Galton–Watson trees with finite offspring variance. Indeed, they considered fringe subtree counts of a sequence of trees instead of a fixed tree. In particular, they showed that asymptotic normality still holds in some regimes, while in others there is a Poisson limit. In a forthcoming work, we will study the case of not fixed fringe trees in the framework of random trees with given degrees.

1.1   Organization of the paper

In Section 2 we provide exact formulas for factorial moments of $N_T(\mathcal{T}_{\mathbf{n}})$. These formulas are then used in Sections 3–4 to prove our main results. An application to simply generated trees is given in Sections 5.

1.2 Some notation

In addition to the notation introduced above, we use the following standard notation.

We let $\mathbb{Z} := \{\ldots, -1, 0, 1, \ldots\}$, $\mathbb{N} := \{1, 2, \ldots\}$, $\mathbb{N}_0 := \{0, 1, 2, \ldots\}$. We let 0 denote also vectors and matrices with all elements 0 (the dimension will be clear from the context). We use standard o and O notation, for sequences and functions of a real variable.

$\mathbf{1}_{\mathcal{E}}$ is the indicator function of an event $\mathcal{E}$, and $\delta_{ij} := \mathbf{1}_{\{i=j\}}$ is Kronecker's delta.

For $x \in \mathbb{R}$ and $q \in \mathbb{N}_0$, we let $(x)_q := x(x-1)\cdots(x-q+1)$ denote the qth falling factorial of x. (Here $(x)_0 := 1$. Note that $(x)_q = 0$ whenever $x \in \mathbb{N}_0$ and $x - q + 1 \leq 0$.)

We interpret $0/0 = 0$ and $0 \cdot \infty = 0$.

We use $\xrightarrow{\mathrm{d}}$ for convergence in distribution, and $\xrightarrow{\mathrm{p}}$ for convergence in probability, for a sequence of random variables in some metric space. Also, $\mathcal{L}(X)$ denotes the distribution of X, and $\overset{\mathrm{d}}{=}$ means equal in distribution. We write $\mathrm{N}(0, \Gamma)$ for the multivariate normal distribution with mean vector 0 and covariance matrix $\Gamma := (\gamma_{ij})_{i,j=1}^m$, for $m \in \mathbb{N}$. (This includes the case $\Gamma = 0$; in this case $X \sim \mathrm{N}(0, \Gamma)$ means that $X = 0 \in \mathbb{R}^m$ a.s.)

Unspecified limits are as $\kappa \to \infty$.

2 Moment computations

In this section, we compute the joint factorial moments of $N_{T_1}(\mathcal{T}_\mathbf{n}), \ldots, N_{T_m}(\mathcal{T}_\mathbf{n})$, for $m \geq 1$ and a sequence of distinct rooted plane trees $T_1, \ldots, T_m \in \mathbb{T}$, where $\mathcal{T}_\mathbf{n}$ is a uniformly random tree of $\mathbb{T}_\mathbf{n}$, for a degree statistic $\mathbf{n}$. Before that, we need to introduce some notation. For $1 \leq i, j \leq m$, let

$$\tau_{ij} := N_{T_i}(T_j)\mathbf{1}_{\{i \neq j\}} \tag{17}$$

be the number of proper fringe subtrees of T_j that are equal to T_i. (Note that many of these terms are 0. In particular, if we order $T_1, \ldots, T_m$ according to their sizes, the matrix $(\tau_{ij})_{i,j=1}^m$ is strictly triangular.)

For $q_1, \ldots, q_m \in \mathbb{N}_0$, note that the product $(N_{T_1}(\mathcal{T}_\mathbf{n}))_{q_1} \cdots (N_{T_m}(\mathcal{T}_\mathbf{n}))_{q_m}$ is the number of sequences of $q := q_1 + \cdots + q_m$ distinct fringe subtrees of $\mathcal{T}_\mathbf{n}$, where the first q_1 are copies of T_1, the next q_2 are copies of T_2, and so on. Given such a sequence of fringe subtrees, we say that these fringe subtrees are *marked*. Furthermore, for each such sequence of marked fringe subtrees of $\mathcal{T}_\mathbf{n}$, say that a tree in the sequence is *bound* if it is a fringe subtree of another tree in the sequence; otherwise it is *free*. Note that the free trees are disjoint. Furthermore, each bound tree in the sequence is a fringe subtree of exactly one free tree. For a sequence $b = (b_1, \ldots, b_m) \in \mathbb{N}_0^m$, let $S_b(\mathcal{T}_\mathbf{n})$ be the number of such sequences of q fringe trees such that exactly b_i of the fringe trees T_i are bound, for $1 \leq i \leq m$. We thus have

$$\mathbb{E}\big[(N_{T_1}(\mathcal{T}_\mathbf{n}))_{q_1} \cdots (N_{T_m}(\mathcal{T}_\mathbf{n}))_{q_m}\big] = \sum_{b \in \mathbb{N}_0^m} \mathbb{E}[S_b(\mathcal{T}_\mathbf{n})]. \tag{18}$$

The sum is really only over $b = (b_1, \ldots, b_m) \in \mathbb{N}_0^m$ such that $0 \leq b_i \leq q_i$ for $1 \leq i \leq m$, since otherwise $S_b(\mathcal{T}_\mathbf{n}) = 0$. This sum can be computed by the following lemma.

▶ **Lemma 7.** *Let $\mathbf{n}$ be a degree statistic and let $\mathcal{T}_\mathbf{n} \sim \mathrm{Unif}(\mathbb{T}_\mathbf{n})$. For $m \geq 1$ and $q_1, \ldots, q_m \in \mathbb{N}$, let $T_1, \ldots, T_m \in \mathbb{T}$ be a sequence of distinct rooted plane trees such that $|\mathbf{n}| \geq \sum_{j=1}^m (q_j - b_j)(|T_j| - 1) + 1$. Then $\mathbb{E}[S_b(\mathcal{T}_\mathbf{n})]$ is equal to*

$$\frac{|\mathbf{n}|}{(|\mathbf{n}|)_{1+\sum_{j=1}^m (q_j - b_j)(|T_j|-1)}} \prod_{i \geq 0} (n(i))_{\sum_{j=1}^m (q_j - b_j) n_{T_j}(i)} \prod_{j=1}^m \frac{(q_j)_{b_j} \left(\sum_{k=1}^m (q_k - b_k)\tau_{jk}\right)_{b_j}}{b_j!}, \tag{19}$$

for every $b = (b_1, \ldots, b_m) \in \mathbb{N}_0^m$ such that $0 \leq b_i \leq q_i$, for $1 \leq i \leq m$.

Proof. If $\sum_{j=1}^{m}(q_j - b_j)n_{T_j}(i) > n(i)$ for some $i \geq 0$, then both $\mathbb{E}[S_b(\mathcal{T_n})]$ and (19) are 0. We may thus assume that $\sum_{j=1}^{m}(q_j - b_j)n_{T_j}(i) \leq n(i)$ for all $i \geq 0$.

First, let us consider the case when all fringe trees are free, that is, the case $b = 0 = (0, \ldots, 0) \in \mathbb{N}_0^m$. Replace each marked fringe subtree in $\mathcal{T_n}$ by a single leaf; moreover, mark this leaf and order all marked leaves into a sequence, corresponding to the order of the fringe subtrees. This yields another tree $\widetilde{\mathcal{T}}$, which we call a *reduced tree*, with a sequence of q marked leaves. Since $\mathcal{T_n}$ has $n(i)$ vertices of degree i, for $i \geq 0$, and we have replaced q_j copies of T_j by leaves, the degree statistic $\tilde{\mathbf{n}} = (\tilde{n}(i))_{i \geq 0}$ of $\widetilde{\mathcal{T}}$ is given by

$$\tilde{n}(i) := \begin{cases} n(i) - \sum_{j=1}^{m} q_j n_{T_j}(i), & i \geq 1, \\ n(0) - \sum_{j=1}^{m} q_j n_{T_j}(0) + \sum_{j=1}^{m} q_j, & i = 0, \end{cases} \tag{20}$$

and has size

$$|\tilde{\mathbf{n}}| := \sum_{i \geq 0} \tilde{n}(i) = |\mathbf{n}| - \sum_{j=1}^{m} q_j(|T_j| - 1). \tag{21}$$

There is a one-to-one correspondence between trees in $\mathbb{T_n}$ with a sequence of marked fringe subtrees as above, and reduced trees with the degree statistic (20) and a sequence of q marked leaves. If we ignore the marks, the number of possible reduced trees is given by (2) with the degree statistic $\tilde{\mathbf{n}}$ in (20). In each unmarked reduced tree, the number of ways to choose sequences of marked leaves is $(\tilde{n}(0))_{q_1 + \cdots + q_m}$. Thus, the number of trees in $\mathbb{T_n}$ with marked sequences of free fringe subtrees is the product of these numbers, i.e.,

$$\frac{(|\tilde{\mathbf{n}}| - 1)!}{\prod_{i \geq 0} \tilde{n}(i)!}(\tilde{n}(0))_{\sum_{j=1}^{m} q_j} = \frac{(|\tilde{\mathbf{n}}| - 1)!}{\prod_{i \geq 0}(n(i) - \sum_{j=1}^{m} q_j n_{T_j}(i))!}. \tag{22}$$

By dividing with $|\mathbb{T_n}|$, which is given by (2), and using (21), we find

$$\mathbb{E}[S_0(\mathcal{T_n})] = \frac{1}{(|\mathbf{n}| - 1)_{\sum_{j=1}^{m} q_j(|T_j| - 1)}} \prod_{i \geq 0} (n(i))_{\sum_{j=1}^{m} q_j n_{T_j}(i)}. \tag{23}$$

Now consider the general case with a sequence $b = (b_1, \ldots, b_m)$ telling the number of bound fringe subtrees. There are thus $q_j - b_j$ free trees of type T_j. The number of ways to choose the positions of the bound trees in the sequences of fringe trees is $\prod_{j=1}^{m} \binom{q_j}{b_j}$, and for each choice of free trees, there are $\sum_{k=1}^{m}(q_k - b_k)\tau_{jk}$ possible bound trees of type T_j; thus the number of choices of the bound trees is

$$\prod_{j=1}^{m} \frac{(q_j)_{b_j} \left(\sum_{k=1}^{m}(q_k - b_k)\tau_{jk} \right)_{b_j}}{b_j!}. \tag{24}$$

The number of trees in $\mathbb{T_n}$ with sequences of $q_j - b_j$ free trees T_j, for $1 \leq j \leq m$, is given by replacing q_j by $q_j - b_j$ in (20)–(22). Hence, we obtain (19), extending (23).   ◄

We record two important special cases of Lemma 7 (see the proof of [3, Lemma 3.3] in the full version for details).

▶ **Lemma 8.** *Let* $\mathbf{n}$ *be a degree statistic and let* $\mathcal{T_n} \sim \mathrm{Unif}(\mathbb{T_n})$.
  (i) *For* $q \in \mathbb{N}$ *and* $T \in \mathbb{T}$ *such that* $|\mathbf{n}| \geq q|T| - q + 1$,

$$\mathbb{E}[(N_T(\mathcal{T_n}))_q] = \frac{|\mathbf{n}|}{(|\mathbf{n}|)_{q|T| - q + 1}} \prod_{i \geq 0} (n(i))_{qn_T(i)}. \tag{25}$$

(ii) *For distinct $T, T' \in \mathbb{T}$ such that $|\mathbf{n}| \geq |T| + |T'| - 1$,*

$$\mathbb{E}[N_T(\mathcal{T}_\mathbf{n}) N_{T'}(\mathcal{T}_\mathbf{n})] = N_T(T') \, \mathbb{E}[N_{T'}(\mathcal{T}_\mathbf{n})] + N_{T'}(T) \, \mathbb{E}[N_T(\mathcal{T}_\mathbf{n})]$$

$$+ \frac{|\mathbf{n}|}{(|\mathbf{n}|)_{|T|+|T'|-1}} \prod_{i \geq 0} (n(i))_{n_T(i) + n_{T'}(i)}. \tag{26}$$

3 Proof of Theorems 3

In this section we prove Theorem 3. In what follows we will frequently use the following well-known estimate (see for example, [3, Lemma 4.1]).

▶ **Lemma 9.** *If $x \geq 1$ is a real number and $0 \leq k \leq x/2$ is an integer, then*

$$(x)_k = x^k \exp\left(-\frac{k(k-1)}{2x} + O\left(\frac{k^3}{x^2}\right)\right). \tag{27}$$

We start by proving the following theorem.

▶ **Theorem 10.** *Let $T \in \mathbb{T}$ be a fixed tree. Then, uniformly for all degree statistics $\mathbf{n} = (n(i))_{i \geq 0}$,*

$$\mathbb{E} \, N_T(\mathcal{T}_\mathbf{n}) = |\mathbf{n}| \pi_{\mathbf{p}(\mathbf{n})}(T) + O(1), \tag{28}$$

$$\mathrm{Var} \, N_T(\mathcal{T}_\mathbf{n}) = |\mathbf{n}| \gamma_{\mathbf{p}(\mathbf{n})}(T, T) + O(1). \tag{29}$$

More generally, if $T, T' \in \mathbb{T}$, then

$$\mathrm{Cov}\big(N_T(\mathcal{T}_\mathbf{n}), N_{T'}(\mathcal{T}_\mathbf{n})\big) = |\mathbf{n}| \gamma_{\mathbf{p}(\mathbf{n})}(T, T') + O(1). \tag{30}$$

Proof. Note first the trivial bound

$$N_T(\mathcal{T}_\mathbf{n}) \leq \frac{n(i)}{n_T(i)} \leq n(i), \qquad i \in \mathcal{D}(T), \tag{31}$$

since the copies of T in $\mathcal{T}_\mathbf{n}$ are distinct. Furthermore, by (7) and (5),

$$|\mathbf{n}| \pi_{\mathbf{p}(\mathbf{n})}(T) \leq |\mathbf{n}| p_i(\mathbf{n}) = n(i), \qquad i \in \mathcal{D}(T). \tag{32}$$

Hence, (28) is trivial if $n(i) = O(1)$ for some $i \in \mathcal{D}(T)$. In particular, we may in the sequel assume $n(i) \geq 2n_T(i)$ for every $i \geq 0$, and thus $|\mathbf{n}| \geq 2|T|$. Then, by (25) (with $q = 1$) and Lemma 9,

$$\mathbb{E} \, N_T(\mathcal{T}_\mathbf{n}) = |\mathbf{n}|^{1-|T|} \prod_{i \in \mathcal{D}(T)} n(i)^{n_T(i)}$$

$$\times \exp\left(\frac{|T|(|T|-1)}{2|\mathbf{n}|} - \sum_{i \in \mathcal{D}(T)} \frac{n_T(i)(n_T(i)-1)}{2n(i)} + O\Big(\sum_{i \in \mathcal{D}(T)} \frac{1}{n(i)^2}\Big)\right)$$

$$= |\mathbf{n}| \pi_{\mathbf{p}(\mathbf{n})}(T)$$

$$\times \exp\left(\frac{|T|(|T|-1)}{2|\mathbf{n}|} - \sum_{i \in \mathcal{D}(T)} \frac{n_T(i)(n_T(i)-1)}{2n(i)} + O\Big(\sum_{i \in \mathcal{D}(T)} \frac{1}{n(i)^2}\Big)\right), \tag{33}$$

which implies (28) by (32).

Similarly, taking $q = 2$ in (25), and now assuming as we may $n(i) \geq 4n_T(i)$ for every $i \geq 0$,

$$
\begin{aligned}
\mathbb{E}(N_T(\mathcal{T}_{\mathbf{n}}))_2 &= \frac{|\mathbf{n}|}{(|\mathbf{n}|)_{2|T|-1}} \prod_{i \in \mathcal{D}(T)} (n(i))_{2n_T(i)} \\
&= |\mathbf{n}|^{2-2|T|} \prod_{i \in \mathcal{D}(T)} n(i)^{2n_T(i)} \\
&\quad \times \exp\left(\frac{(2|T|-1)(2|T|-2)}{2|\mathbf{n}|} - \sum_{i \in \mathcal{D}(T)} \frac{2n_T(i)(2n_T(i)-1)}{2n(i)} + O\left(\sum_{i \in \mathcal{D}(T)} \frac{1}{n(i)^2} \right) \right) \\
&= \left(|\mathbf{n}| \pi_{\mathbf{p(n)}}(T) \right)^2 \\
&\quad \times \exp\left(\frac{(2|T|-1)(|T|-1)}{|\mathbf{n}|} - \sum_{i \in \mathcal{D}(T)} \frac{n_T(i)(2n_T(i)-1)}{n(i)} + O\left(\sum_{i \in \mathcal{D}(T)} \frac{1}{n(i)^2} \right) \right),
\end{aligned}
\tag{34}
$$

Hence, using also (33),

$$
\begin{aligned}
\mathbb{E}(N_T(\mathcal{T}_{\mathbf{n}}))_2 &= \left(\mathbb{E}\, N_T(\mathcal{T}_{\mathbf{n}}) \right)^2 \\
&\quad \times \exp\left(\frac{(|T|-1)^2}{|\mathbf{n}|} - \sum_{i \in \mathcal{D}(T)} \frac{n_T(i)^2}{n(i)} + O\left(\sum_{i \in \mathcal{D}(T)} \frac{1}{n(i)^2} \right) \right).
\end{aligned}
\tag{35}
$$

Consequently, using (28) and noting that $\mathbb{E}\, N_T(\mathcal{T}_{\mathbf{n}}) = O(n(i))$ for $i \in \mathcal{D}(T)$ by (28) and (32),

$$
\begin{aligned}
\mathrm{Var}[N_T(\mathcal{T}_{\mathbf{n}})] &= \mathbb{E}(N_T(\mathcal{T}_{\mathbf{n}}))_2 + \mathbb{E}\, N_T(\mathcal{T}_{\mathbf{n}}) - \left(\mathbb{E}\, N_T(\mathcal{T}_{\mathbf{n}}) \right)^2 \\
&= \left(\mathbb{E}\, N_T(\mathcal{T}_{\mathbf{n}}) \right)^2 \left(\frac{(|T|-1)^2}{|\mathbf{n}|} - \sum_{i \in \mathcal{D}(T)} \frac{n_T(i)^2}{n(i)} \right) + \mathbb{E}\, N_T(\mathcal{T}_{\mathbf{n}}) + O(1) \\
&= \left(|\mathbf{n}| \pi_{\mathbf{p(n)}}(T) \right)^2 \left(\frac{(|T|-1)^2}{|\mathbf{n}|} - \sum_{i \in \mathcal{D}(T)} \frac{n_T(i)^2}{n(i)} \right) + |\mathbf{n}| \pi_{\mathbf{p(n)}}(T) + O(1),
\end{aligned}
\tag{36}
$$

which yields (29) by the definitions (10), (9) and (5).

For the proof of (30) we use (26). The first two terms are handled by (28), and the final term is treated as in (34)–(36) with mainly notational differences; we omit the details. ◄

Proof of Theorem 3. By Condition 1, we have $p_i(\mathbf{n}_\kappa) \to p_i$ for every $i \geq 0$, and thus $\pi_{\mathbf{p(n_\kappa)}}(T) \to \pi_{\mathbf{p}}(T)$. Hence, **(i)** follows from (28).

Moreover, it follows from (9)–(10) that $\gamma_{\mathbf{p(n_\kappa)}}(T, T) = O(1)$ (for a fixed T), and thus (29) yields $\mathrm{Var}\, N_T(\mathcal{T}_{\mathbf{n}_\kappa}) = O(|\mathbf{n}_\kappa|)$. Therefore, **(ii)** follows from **(i)** and Chebyshev's inequality. ◄

4 Proof of Theorems 5

We have now all the ingredients to prove Theorem 5.

Proof of Theorem 5. First note that Condition 1 implies

$$
\pi_{\mathbf{p(n_\kappa)}}(T_i) \to \pi_{\mathbf{p}}(T_i) \quad \text{and} \quad \gamma_{\mathbf{p(n_\kappa)}}(T_i, T_j) \to \gamma_{\mathbf{p}}(T_i, T_j), \qquad \text{for } 1 \leq i, j \leq m.
\tag{37}
$$

Hence, (12)–(14) follow from (28)–(30) in Theorem 10.

We next prove the asymptotic normality result in (15). Note first that (28) implies that it does not matter whether we use $\mathbb{E}[N_{T_i}(\mathcal{T}_{\mathbf{n}_\kappa})]$ or

$$\mu_{\mathbf{n}_\kappa}(T) := |\mathbf{n}_\kappa| \pi_{\mathbf{p}(\mathbf{n}_\kappa)}(T) = |\mathbf{n}_\kappa| \prod_{i \geq 0} p_i(\mathbf{n}_\kappa)^{n_T(i)} = |\mathbf{n}_\kappa| \prod_{i \in \mathcal{D}(T)} p_i(\mathbf{n}_\kappa)^{n_T(i)}. \tag{38}$$

in (15).

If $\pi_{\mathbf{p}}(T_i) = 0$, for some $1 \leq i \leq m$, then it follows from (10) that $\gamma_{\mathbf{p}}(T_i, T_i) = 0$, and thus (13) yields $\mathrm{Var}[N_{T_i}(\mathcal{T}_{\mathbf{n}_\kappa})] = o(|\mathbf{n}_\kappa|)$; consequently, (28) and Chebyshev's inequality yield, as $\kappa \to \infty$,

$$\frac{N_{T_i}(\mathcal{T}_{\mathbf{n}_\kappa}) - \mathbb{E}[N_{T_i}(\mathcal{T}_{\mathbf{n}_\kappa})]}{\sqrt{|\mathbf{n}_\kappa|}} \xrightarrow{\mathrm{p}} 0. \tag{39}$$

Hence, convergence of the i-th component in (15) is trivial in this case. Furthermore, $\pi_{\mathbf{p}}(T_i) = 0$ also implies $\gamma_{\mathbf{p}}(T_i, T_j) = 0$ for every $1 \leq j \leq m$ by (11), noting that if $N_{T_i}(T_j) > 0$ then also $\pi_{\mathbf{p}}(T_j) = 0$. Thus, we may ignore all i in (15) with $\pi_{\mathbf{p}}(T_i) = 0$ and show (joint) convergence for the remaining ones, because then (15) in general will follow from [4, Theorem 3.9 in Chapter 1]. Consequently, we henceforth assume that $\pi_{\mathbf{p}}(T_i) > 0$ for all $1 \leq i \leq m$. Equivalently, $p_k > 0$ for every $k \in \bigcup_{i=1}^m \mathcal{D}(T_i)$. We may also assume that $T_1, \ldots, T_m$ are distinct.

To see the main idea of the proof, we consider only the univariate case $m = 1$. The general case follows similarly by a multidimensional version of the Gao–Wormald theorem [3, Theorem A.1] in the full version. The main complication in the multivariate case is the possibility that fringe trees of type T_j may contain fringe trees of type T_k for some $1 \leq j, k \leq m$; we thus use the decomposition in (18) and estimate the terms separately; we refer to the proof of [3, Theorem 1.5] in the full version for details.

We then consider $m = 1$ and omit the index 1 and write T instead of T_1. In this case, we can use the Gao–Wormald theorem [14, Theorem 1] and the following estimate. For any $q_\kappa = O(|\mathbf{n}_\kappa|^{1/2})$, (25) and Lemma 9 yield, recalling the definitions (5), (7), (9), (10), and (38) of $p_i(\mathbf{n})$, $\pi_{\mathbf{p}}(T)$, $\eta_{\mathbf{p}}(T, T)$, $\gamma_{\mathbf{p}}(T, T)$, and $\mu_{\mathbf{n}_\kappa}(T)$,

$$\begin{aligned}
\mathbb{E}[(N_T(\mathcal{T}_{\mathbf{n}_\kappa}))_{q_\kappa}] &= \frac{\prod_{i \geq 0} n_\kappa(i)^{q_\kappa n_T(i)}}{|\mathbf{n}_\kappa|^{q_\kappa(|T|-1)}} \exp\left(\frac{\left(q_\kappa(|T|-1)\right)^2}{2|\mathbf{n}_\kappa|} - \sum_{i \geq 0} \frac{\left(q_\kappa n_T(i)\right)^2}{2n_\kappa(i)} + o(1) \right) \\
&= |\mathbf{n}_\kappa|^{q_\kappa} \prod_{i \geq 0} p_i(\mathbf{n}_\kappa)^{q_\kappa n_T(i)} \exp\left(\frac{\left(q_\kappa(|T|-1)\right)^2}{2|\mathbf{n}_\kappa|} - \sum_{i \geq 0} \frac{\left(q_\kappa n_T(i)\right)^2}{2n_\kappa(i)} + o(1) \right) \\
&= \left(|\mathbf{n}_\kappa| \pi_{\mathbf{p}(\mathbf{n}_\kappa)}(T)\right)^{q_\kappa} \exp\left(\frac{q_\kappa^2}{2|\mathbf{n}_\kappa|} \eta_{\mathbf{p}(\mathbf{n}_\kappa)}(T, T) + o(1) \right) \\
&= \mu_{\mathbf{n}_\kappa}(T)^{q_\kappa} \exp\left(\frac{(\gamma_{\mathbf{p}(\mathbf{n}_\kappa)}(T, T) - \pi_{\mathbf{p}(\mathbf{n}_\kappa)}(T))|\mathbf{n}_\kappa|}{2\mu_{\mathbf{n}_\kappa}(T)^2} q_\kappa^2 + o(1) \right) \\
&= \mu_{\mathbf{n}_\kappa}(T)^{q_\kappa} \exp\left(\frac{\gamma_{\mathbf{p}}(T, T)|\mathbf{n}_\kappa| - \mu_{\mathbf{n}_\kappa}(T)}{2\mu_{\mathbf{n}_\kappa}(T)^2} q_\kappa^2 + o(1) \right).
\end{aligned} \tag{40}$$

If $\gamma_{\mathbf{p}}(T, T) > 0$, we may now apply the Gao–Wormald theorem [14, Theorem 1] with $\mu_\kappa := \mu_{\mathbf{n}_\kappa}(T)$ and $\sigma_\kappa^2 := \gamma_{\mathbf{p}}(T, T)|\mathbf{n}_\kappa|$ and conclude (16), which by (13) is equivalent to (15) (with $m = 1$). The case $\gamma_{\mathbf{p}}(T, T) = 0$ is trivial, since then (13) implies (39). Alternatively, for any $\gamma_{\mathbf{p}}(T, T)$, we may take the same μ_κ but $\sigma_\kappa^2 := |\mathbf{n}_\kappa|$ in the case $m = 1$ of our version [3, Theorem A.1] of the Gao–Wormald theorem. ◀

5 Application to simply generated trees

Let $\mathbb{T}_n$ denote the (finite) subset of all plane rooted trees of size $n \in \mathbb{N}$. Let $\mathbf{w} = (w_i)_{i \geq 0}$ be a sequence of non-negative real weights with $w_0 > 0$ and $w_i > 0$ for at least one $i \geq 2$. For a finite rooted plane tree $T \in \mathbb{T}$, we define the weight of T to be

$$w(T) := \prod_{v \in T} w_{d_T(v)} = \prod_{i \geq 0} w_i^{n_T(i)}. \tag{41}$$

For $n \in \mathbb{N}$, let $Z_n(\mathbf{w}) = \sum_{T \in \mathbb{T}_n} w(T)$. If $Z_n(\mathbf{w}) > 0$, then we define the random tree $\mathcal{T}_{\mathbf{w},n}$ by picking an element of $\mathbb{T}_n$ at random with probability proportional to its weight, i.e.,

$$\mathbb{P}(\mathcal{T}_{\mathbf{w},n} = T) = \frac{w(T)}{Z_n(\mathbf{w})}, \quad \text{for } T \in \mathbb{T}_n. \tag{42}$$

The random tree $\mathcal{T}_{\mathbf{w},n}$ is called simply generated tree of size n and weight sequence $\mathbf{w}$; see e.g. [9] and [19]. If $\mathbf{w}$ is a probability distribution (i.e., $\sum_{i \geq 0} w_i = 1$), then $\mathcal{T}_{\mathbf{w},n}$ is a Galton–Watson tree with offspring distribution $\mathbf{w}$ conditioned to have n vertices.

Let $\Phi_{\mathbf{w}}(z) = \sum_{i \geq 0} w_i z^i$ be the generating function of the weight sequence $\mathbf{w}$, and let $\rho_{\mathbf{w}} \in [0, \infty]$ be its radius of convergence. For $0 \leq s < \rho_{\mathbf{w}}$, we let

$$\Psi_{\mathbf{w}}(s) := \frac{s \Phi_{\mathbf{w}}'(s)}{\Phi_{\mathbf{w}}(s)} = \frac{\sum_{i \geq 0} i w_i s^i}{\sum_{i \geq 0} w_i s^i}. \tag{43}$$

Furthermore, if $\Phi_{\mathbf{w}}(\rho_{\mathbf{w}}) < \infty$, we define also $\Psi_{\mathbf{w}}(\rho_{\mathbf{w}})$ by (43); if $\Phi_{\mathbf{w}}(\rho_{\mathbf{w}}) = \infty$ then we define $\Psi_{\mathbf{w}}(\rho_{\mathbf{w}}) := \lim_{s \uparrow \rho_{\mathbf{w}}} \Psi_{\mathbf{w}}(s)$; the limit exists by [19, Lemma 3.1 (i)]. Let $\nu_{\mathbf{w}} := \Psi_{\mathbf{w}}(\rho_{\mathbf{w}}) \in [0, \infty]$, and define

$$\tau_{\mathbf{w}} = \begin{cases} \rho_{\mathbf{w}} & \text{if } \nu_{\mathbf{w}} < 1, \\ \Psi_{\mathbf{w}}^{-1}(1) & \text{if } \nu_{\mathbf{w}} \geq 1. \end{cases} \tag{44}$$

It follows from [19, Lemma 3.1] that

$$\rho_{\mathbf{w}} > 0 \iff \nu_{\mathbf{w}} > 0 \iff \tau_{\mathbf{w}} > 0. \tag{45}$$

The following result from [19] shows that simply generated trees satisfy Condition 1 in probability.

▶ **Theorem 11** ([19, Theorem 7.1 and Theorem 7.11]). *Let $\mathbf{w}$ be a sequence of non-negative real weights with $w_0 > 0$ and $w_i > 0$ for at least one $i \geq 2$. Define*

$$\theta_i(\mathbf{w}) = \frac{w_i \tau_{\mathbf{w}}^i}{\Phi_{\mathbf{w}}(\tau_{\mathbf{w}})}, \quad \text{for } i \geq 0. \tag{46}$$

Then, $\theta(\mathbf{w}) = (\theta_i(\mathbf{w}))_{i \geq 0}$ is a probability distribution with expectation $\mu_{\mathbf{w}} = \min(1, \nu_{\mathbf{w}})$ and variance $\sigma_{\mathbf{w}}^2 = \tau_{\mathbf{w}} \Psi_{\mathbf{w}}'(\tau_{\mathbf{w}}) \in [0, \infty]$. Moreover, for $n \in \mathbb{N}$ with $Z_n(\mathbf{w}) > 0$, let $\mathcal{T}_{\mathbf{w},n}$ be a simply generated tree of size n and weight sequence $\mathbf{w}$. Then, the (empirical) degree distribution $\mathbf{p}(\mathbf{n}_{\mathcal{T}_{\mathbf{w},n}})$ of $\mathcal{T}_{\mathbf{w},n}$ satisfies, for every $i \geq 0$, $p_i(\mathbf{n}_{\mathcal{T}_{\mathbf{w},n}}) \xrightarrow{\mathrm{p}} \theta_i(\mathbf{w})$, as $n \to \infty$ (along integers n such that $Z_n(\mathbf{w}) > 0$).

Note that if $\rho_{\mathbf{w}} = 0$, then $\theta_0(\mathbf{w}) = 1$ and $\theta_i(\mathbf{w}) = 0$ for $i \geq 1$; otherwise, $\tau_{\mathbf{w}} > 0$ and (46) shows that $\theta_i(\mathbf{w}) > 0 \iff w_i > 0$ for $i \geq 0$.

Using Theorem 11, we can show that Theorem 5 implies the following version for conditioned Galton–Watson trees. The asymptotic normality (49) was proved in case **(i)** by different methods in [20, Corollary 1.8]; **(ii)** and **(iii)** are new.

▶ **Theorem 12** (partly [20]). *Let* $\mathbf{w}$ *be a sequence of non-negative real weights with* $w_0 > 0$ *and* $w_i > 0$ *for at least one* $i \geq 2$. *Moreover, for* $n \in \mathbb{N}$ *with* $Z_n(\mathbf{w}) > 0$, *let* $\mathcal{T}_{\mathbf{w},n}$ *be a simply generated tree of size* n *and weight sequence* $\mathbf{w}$. *For fixed* $m \geq 1$, *let* $T_1, \ldots, T_m \in \mathbb{T}$ *be a fixed sequence of rooted plane trees. Then, as* $n \to \infty$ *(along integers* n *such that* $Z_n(\mathbf{w}) > 0$*),*

$$\left(\frac{N_{T_j}(\mathcal{T}_{\mathbf{w},n}) - \mathbb{E}[N_{T_j}(\mathcal{T}_{\mathbf{w},n}) \mid \mathbf{n}_{\mathcal{T}_{\mathbf{w},n}}]}{\sqrt{n}} \right)_{j=1}^m \xrightarrow{\mathrm{d}} \mathrm{N}(0, \Gamma_{\theta(\mathbf{w})}), \tag{47}$$

where the covariance matrix $\Gamma_{\theta(\mathbf{w})}$ *is defined by* (10)–(11), *and for* $1 \leq j \leq m$,

$$\mathbb{E}[N_{T_j}(\mathcal{T}_{\mathbf{w},n}) \mid \mathbf{n}_{\mathcal{T}_{\mathbf{w},n}}] = \frac{n}{(n)_{|T_j|}} \prod_{i \geq 0} (n_{\mathcal{T}_{\mathbf{w},n}}(i))_{n_{T_j}(i)}. \tag{48}$$

Furthermore, suppose that the weight sequence $\mathbf{w}$ *satisfies one of the following conditions:*
  (i) $\nu_{\mathbf{w}} \geq 1$ *and* $\sigma_{\mathbf{w}}^2 \in (0, \infty)$.
 (ii) $\nu_{\mathbf{w}} \geq 1$, $\sigma_{\mathbf{w}}^2 = \infty$ *and* $\theta(\mathbf{w})$ *belongs to the domain of attraction of a stable law of index* $\alpha \in (1, 2]$. *(The last condition is equivalent to that there exists a slowly varying function* $L : \mathbb{R}_+ \to \mathbb{R}_+$ *such that* $\sum_{i=0}^k i^2 \theta_i(\mathbf{w}) = k^{2-\alpha} L(k)$, *as* $k \to \infty$ *[10, Theorem XVII.5.2].)*
(iii) $0 < \nu_{\mathbf{w}} < 1$ *and* $\theta_i(\mathbf{w}) = ci^{-\beta} + o(i^{-\beta})$, *as* $i \to \infty$, *with fixed* $c > 0$ *and* $\beta > 2$.
Then, as $n \to \infty$ *(along integers* n *such that* $Z_n(\mathbf{w}) > 0$*),*

$$\left(\frac{N_{T_j}(\mathcal{T}_{\mathbf{w},n}) - n\pi_{\theta(\mathbf{w})}(T_j)}{\sqrt{n}} \right)_{j=1}^m \xrightarrow{\mathrm{d}} \mathrm{N}(0, \widetilde{\Gamma}_{\theta(\mathbf{w})}), \tag{49}$$

where the covariance matrix $\widetilde{\Gamma}_{\theta(\mathbf{w})} = (\widetilde{\gamma}_{\theta(\mathbf{w})}(T_i, T_j))_{i,j=1}^m$ *is given by, for* $T, T' \in \mathbb{T}$ *such that* $T \neq T'$,

$$\widetilde{\gamma}_{\theta(\mathbf{w})}(T, T) = \pi_{\theta(\mathbf{w})}(T) - \left(2|T| - 1 + \varsigma_{\mathbf{w}}^{-2} \right) (\pi_{\theta(\mathbf{w})}(T))^2, \tag{50}$$

$$\widetilde{\gamma}_{\theta(\mathbf{w})}(T, T') = N_{T'}(T)\pi_{\theta(\mathbf{w})}(T) + N_T(T')\pi_{\theta(\mathbf{w})}(T')$$
$$- \left(|T| + |T'| - 1 + \varsigma_{\mathbf{w}}^{-2} \right) \pi_{\theta(\mathbf{w})}(T)\pi_{\theta(\mathbf{w})}(T'), \tag{51}$$

with $\varsigma_{\mathbf{w}}^2 = \sigma_{\mathbf{w}}^2$ *in case* (i), *and* $\varsigma_{\mathbf{w}}^2 = \infty$ *in cases* (ii) *and* (iii).

▶ **Remark 13.** Recall that for any weight sequence $\mathbf{w}$ and any constants $a, b > 0$, the weight sequence $\widehat{\mathbf{w}} = (\widehat{w}_i)_{i \geq 0}$ with $\widehat{w}_i := ab^i w_i$ is equivalent to $\mathbf{w}$, i.e., it satisfies that $\mathcal{T}_{\mathbf{w},n} \stackrel{\mathrm{d}}{=} \mathcal{T}_{\widehat{\mathbf{w}},n}$, for all n for which either (and thus both) of the random trees are defined; this is a consequence of (42). In the setting of Theorem 11, if $\rho_{\mathbf{w}} > 0$, then the weight sequence $\mathbf{w}$ is equivalent to the weight sequence $\theta(\mathbf{w}) = (\theta_i(\mathbf{w}), i \geq 0)$, which is a probability distribution with mean $\mu_{\mathbf{w}} = \min(1, \nu_{\mathbf{w}})$; see further [19, Section 7]. Thus, if $\rho_{\mathbf{w}} > 0$ we can regard $\mathcal{T}_{\mathbf{w},n}$ as a Galton–Watson tree $\mathcal{T}_{\theta(\mathbf{w}),n}$ with offspring distribution $\theta(\mathbf{w})$ conditioned to have n vertices. This explains the appearance of $\theta(\mathbf{w})$ in Theorem 12, and it shows that there is no real loss of generality to consider (as is often done) only the case $\tau_{\mathbf{w}} = 1$ when $\theta(\mathbf{w}) = \mathbf{w}$. Note that the conditioned Galton–Watson tree $\mathcal{T}_{\theta(\mathbf{w}),n}$ is critical if $\nu_{\mathbf{w}} \geq 1$, and subcritical if $0 < \nu_{\mathbf{w}} < 1$.

The complete proof of Theorem 12 is given in [3, Section 7] of the full version. Here, we only comment on the main ideas. Indeed, for any fixed degree statistic $\mathbf{n}$ with $\mathbb{P}(\mathbf{n}_{\mathcal{T}_{\mathbf{w},n}} = \mathbf{n}) > 0$, (42) implies that conditionally given $\mathbf{n}_{\mathcal{T}_{\mathbf{w},n}} = \mathbf{n}$, $\mathcal{T}_{\mathbf{w},n} \sim \mathrm{Unif}(\mathbb{T}_{\mathbf{n}})$; see e.g., [1, Proposition 8]. By the Skorohod coupling theorem [22, Theorem 4.30], we can assume that the convergence in Theorem 11 holds a.s.; in other words, Condition 1 holds a.s. for the degree statistics $\mathbf{n}_{\mathcal{T}_{\mathbf{w},n}}$, with $\mathbf{p} = \theta(\mathbf{w})$. Moreover, e.g. by resampling $\mathcal{T}_{\mathbf{w},n}$ conditioned on $\mathbf{n}_{\mathcal{T}_{\mathbf{w},n}}$, we may assume

that also conditioned on the entire sequence of degree statistics $(\mathbf{n}_{\mathcal{T}_{\mathbf{w},n}})_{n=1}^\infty$, the random trees $\mathcal{T}_{\mathbf{w},n}$, $n \geq 1$, have the (conditional) distributions $\mathrm{Unif}(\mathbb{T}_{\mathbf{n}_{\mathcal{T}_{\mathbf{w},n}}})$. It follows that we may apply Theorem 5 conditioned on the sequence of degree statistics $(\mathbf{n}_{\mathcal{T}_{\mathbf{w},n}})_{n=1}^\infty$; this shows that (47) holds conditioned on $(\mathbf{n}_{\mathcal{T}_{\mathbf{w},n}})_{n=1}^\infty$. Then, (47) also holds unconditionally by the dominated convergence theorem. Furthermore, (48) follows from Lemma 8 (with $q = 1$). On the other hand, the central idea to obtain the unconditional limit (49) is by combining the conditional limit (47) with a limit result for the conditional expectations in (48). For this, one uses a theorem on asymptotic normality of the degree statistics, which is proved in [20] and [24] (see also [3, Theorem 7.6] for a different approach).

Theorem 12 gives a partial solution to [19, Problem 21.4], but the general case remains open.

▶ **Problem 14.** *Does* (49) *in Theorem 12 hold for any weight sequence* $\mathbf{w}$*, with some covariance matrix* $\widetilde{\Gamma}_{\theta(\mathbf{w})} = (\widetilde{\gamma}_{\theta(\mathbf{w})}(T_i, T_j))_{i,j=1}^m$*?*

────── **References** ──────

1   Louigi Addario-Berry and Serte Donderwinkel. Random trees have height $O(\sqrt{n})$. *arXiv e-prints*, January 2022. `arXiv:2201.11773`.

2   David Aldous. Asymptotic fringe distributions for general families of random trees. *Ann. Appl. Probab.*, 1(2):228–266, 1991.

3   Gabriel Berzunza Ojeda, Cecilia Holmgren, and Svante Janson. Fringe trees for random trees with given vertex degrees. *arXiv e-prints*, December 2023. `arXiv:2312.04243`.

4   Patrick Billingsley. *Convergence of probability measures.* Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons, Inc., New York, second edition, 1999. A Wiley-Interscience Publication. `doi:10.1002/9780470316962`.

5   Xing Shi Cai and Luc Devroye. A study of large fringe and non-fringe subtrees in conditional Galton-Watson trees. *ALEA Lat. Am. J. Probab. Math. Stat.*, 14(1):579–611, 2017. `doi:10.30757/alea.v14-29`.

6   Huilan Chang and Michael Fuchs. Limit theorems for patterns in phylogenetic trees. *J. Math. Biol.*, 60(4):481–512, 2010. `doi:10.1007/s00285-009-0275-6`.

7   Luc Devroye. Limit laws for local counters in random binary search trees. *Random Structures Algorithms*, 2(3):303–315, 1991. `doi:10.1002/rsa.3240020305`.

8   Luc Devroye. Limit laws for sums of functions of subtrees of random binary search trees. *SIAM J. Comput.*, 32(1):152–171, 2002/03. `doi:10.1137/S0097539701383923`.

9   Michael Drmota. *Random trees, An interplay between combinatorics and probability.* Springer-WienNewYork, Vienna, 2009. `doi:10.1007/978-3-211-75357-6`.

10   William Feller. *An introduction to probability theory and its applications. Vol. II.* John Wiley & Sons, Inc., New York-London-Sydney, second edition, 1971.

11   Qunqiang Feng and Hosam M. Mahmoud. On the variety of shapes on the fringe of a random recursive tree. *J. Appl. Probab.*, 47(1):191–200, 2010. `doi:10.1239/jap/1269610825`.

12   Philippe Flajolet, Xavier Gourdon, and Conrado Martínez. Patterns in random binary search trees. *Random Structures Algorithms*, 11(3):223–244, 1997. `doi:10.1002/(SICI) 1098-2418(199710)11:3<223::AID-RSA2>3.0.CO;2-2`.

13   Michael Fuchs. Limit theorems for subtree size profiles of increasing trees. *Combin. Probab. Comput.*, 21(3):412–441, 2012. `doi:10.1017/S096354831100071X`.

14   Zhicheng Gao and Nicholas C. Wormald. Asymptotic normality determined by high moments, and submap counts of random maps. *Probab. Theory Related Fields*, 130(3):368–376, 2004. `doi:10.1007/s00440-004-0356-9`.

15   Pawel HItczenko and Nick Wormald. Multivariate asymptotic normality determined by high moments. *arXiv e-prints*, December 2023. `arXiv:2312.04246`.

16 Cecilia Holmgren and Svante Janson. Limit laws for functions of fringe trees for binary search trees and random recursive trees. *Electron. J. Probab.*, 20:no. 4, 51, 2015. `doi:10.1214/EJP.v20-3627`.

17 Cecilia Holmgren and Svante Janson. Fringe trees, Crump-Mode-Jagers branching processes and m-ary search trees. *Probab. Surv.*, 14:53–154, 2017. `doi:10.1214/16-PS272`.

18 Cecilia Holmgren, Svante Janson, and Matas Šileikis. Multivariate normal limit laws for the numbers of fringe subtrees in m-ary search trees and preferential attachment trees. *Electron. J. Combin.*, 24(2):Paper No. 2.51, 49, 2017. `doi:10.37236/6374`.

19 Svante Janson. Simply generated trees, conditioned Galton-Watson trees, random allocations and condensation. *Probab. Surv.*, 9:103–252, 2012. `doi:10.1214/11-PS188`.

20 Svante Janson. Asymptotic normality of fringe subtrees and additive functionals in conditioned Galton-Watson trees. *Random Structures Algorithms*, 48(1):57–101, 2016. `doi:10.1002/rsa.20568`.

21 Svante Janson. Central limit theorems for additive functionals and fringe trees in tries. *Electron. J. Probab.*, 27:Paper No. 47, 1–63, 2022. `doi:10.1214/22-ejp776`.

22 Olav Kallenberg. *Foundations of modern probability*. Probability and its Applications (New York). Springer-Verlag, New York, second edition, 2002. `doi:10.1007/978-1-4757-4015-8`.

23 J. Pitman. *Combinatorial stochastic processes*, volume 1875 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2006. Lectures from the 32nd Summer School on Probability Theory held in Saint-Flour, July 7–24, 2002.

24 Paul Thévenin. Vertices with fixed outdegrees in large Galton-Watson trees. *Electron. J. Probab.*, 25:Paper No. 64, 25, 2020. `doi:10.1214/20-ejp465`.

Enumeration and Succinct Encoding of AVL Trees

Jeremy Chizewer ✉ 🆔
University of Waterloo, Canada

Stephen Melczer ✉ 🏠 🆔
University of Waterloo, Canada

J. Ian Munro ✉ 🏠 🆔
University of Waterloo, Canada

Ava Pun ✉ 🏠 🆔
University of Waterloo, Canada

─── **Abstract** ───

We use a novel decomposition to create succinct data structures – supporting a wide range of operations on static trees in constant time – for a variety of tree classes, extending results of Munro, Nicholson, Benkner, and Wild. Motivated by the class of AVL trees, we further derive asymptotics for the information-theoretic lower bound on the number of bits needed to store tree classes whose generating functions satisfy certain functional equations. In particular, we prove that AVL trees require approximately 0.938 bits per node to encode.

Introduction

AVL trees [1] (named for their discoverers, G. Adelson-Velsky and E. Landis) are a subclass of binary search trees with logarithmic height, a property they maintain with updates during insertions and deletions in logarithmic time. Indeed, AVL trees are the oldest class of binary search trees maintaining logarithmic height and are characterized by the key property that any pair of sibling subtrees differ in height by at most 1. In this paper, we examine the amount of storage needed to encode AVL trees with n nodes, a property intimately related to the number of AVL trees on n nodes. Odlyzko [13] gave a conjectural form for the number of AVL trees on n nodes in the 1980s, anticipating a forthcoming proof, but this proof did not appear in the literature.

If $\mathcal{C} = \bigsqcup_{n=0}^{\infty} \mathcal{C}_n$ is a family of objects, with $\mathcal{C}_n$ denoting the objects of size n in $\mathcal{C}$ then a representation of $\mathcal{C}$ is called *succinct* if it maps each object of $\mathcal{C}_n$ to a unique string of length $\log_2 |\mathcal{C}_n| + o(\log |\mathcal{C}_n|)$. A succinct representation is thus one whose space complexity asymptotically equals, up to lower-order terms, the information-theoretic lower bound. A *succinct data structure* [11, 12] for $\mathcal{C}$ is a succinct representation of $\mathcal{C}$ that supports a range of operations on $\mathcal{C}$ under reasonable time constraints.
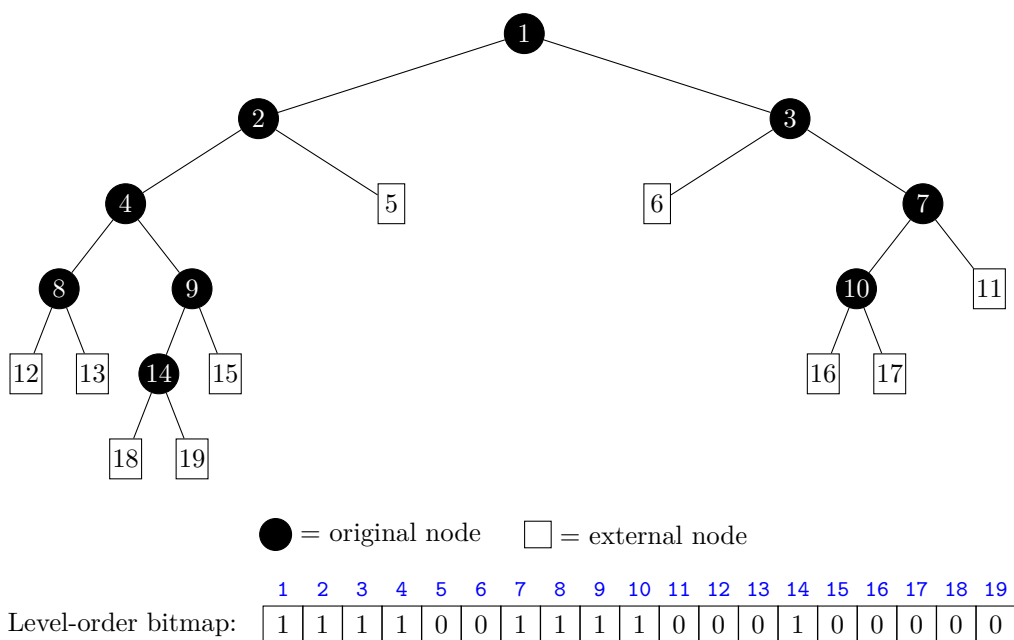
Figure 1 A binary tree and its level-order bitmap representation.

1 Representations of Trees

The theory of succinct data structures has a long history, much of it focused on representations of trees. We first describe some important classes of trees in this context, and then discuss our main results.

Binary Search Trees

Let $\mathcal{B}$ be the class of rooted binary trees, so that the number $|\mathcal{B}_n|$ of objects in $\mathcal{B}$ of size n is the nth Catalan number $b_n = \frac{1}{n+1}\binom{2n}{n}$. The class $\mathcal{B}$ lends itself well to storing ordered data in a structure called a *binary search tree*. The general idea is that for each node in the tree, the data stored in its left subtree will be smaller than the data at that node, and the data stored in the right subtree will be larger. To retrieve elements, one can recursively navigate through the tree by comparing the desired element to the current node, and moving to the left or right subtree if the element is respectively smaller or larger than the current node. As a result, it is desirable to efficiently support the navigation operations of moving to parent or child nodes in whatever representation is used.

A naive representation of $\mathcal{B}$ gives each node a label (using roughly $\log_2 n$ space) and stores the labels of each node's children and parent. The resulting data structure supports operations like finding node siblings in constant time, but is *not* succinct as it uses $\Theta(n \log n)$ bits while the information-theoretic lower bound is only $\log_2(b_n) = 2n + o(n)$. Somewhat conversely, a naive space-optimal representation of $\mathcal{B}$ is obtained by listing the objects of $\mathcal{B}_n$ in any canonical order and referencing a tree by its position $\{1, \ldots, b_n\}$ in the order, but asking for information like the children or parents of a node in a specific tree is then expensive as it requires building parts of the tree.

Practical succinct representations of binary trees supporting efficient navigation date back to Jacobson [6], who encoded a tree by storing the binary string of length $2n + 1$ obtained by adding *external* vertices so that every node has exactly two children, then taking a level-order

traversal of the tree and recording a 1 for each original *internal* node encountered and a 0 for each external node encountered (see Figure 1). If each node is labelled by its position in a level-order traversal then, for instance, the children of the node labelled x in the tree encoded by the string σ have labels $2 \operatorname{rank}_x(\sigma)$ and $2 \operatorname{rank}_x(\sigma) + 1$, where $\operatorname{rank}_x(\sigma)$ is the number of ones in σ up to (and including) the position x. By storing $o(n)$ bits, the rank operation (and similar supporting operations used to retrieve information about the trees) can be implemented in $O(1)$ time. Jacobson's results allow finding a parent or child using $O(\log_2 n)$ bit inspections; Clark [2] and Munro [8] improved this to $O(1)$ inspections of $\log_2 n$ bit words.

AVL Trees

Because the time taken to access elements in a binary search tree typically depends on the height of the tree, many data structures *balance* their trees as new data is added. The balance operation requires rearranging the tree while preserving the underlying property that, for each node, the elements in the left subtree are smaller and the elements in the right subtree are larger. One of the most popular balanced tree structures – for theoretical study and practical application – are *AVL trees* [1]. Roughly speaking, AVL trees have balancing rules that force the subtrees rooted at the children of any node differ in height by at most one. Throughout this paper we let $\mathcal{A}$ denote the class of AVL trees, so that $\mathcal{A}_n$ consists of all binary trees on n vertices such that the subtrees of any vertex differ in height by at most one (including empty subtrees).

Due to the way they are constructed, AVL trees have mainly been enumerated under height restrictions, and enumeration by number of vertices (which is crucial for determining space-efficient representations, but not as important for other applications) is less studied. A 1984 paper [13] of Odlyzko describes the behaviour of a family of trees whose generating functions satisfy certain equations. It ends by stating that the generating function of AVL trees "appears not to satisfy any simple functional equation, but by an intensive study... it can be shown" that $|\mathcal{A}_n| \sim n^{-1}\alpha^{-n}u(\log n)$ where $\alpha = 0.5219\ldots$ is "a certain constant" and u is a periodic function, referencing for details a paper that was planned to be published but was never written.[1]

Efficiently Representing Tree Classes

Let B be a function satisfying $B(n) = \Theta(\log n)$. In [9] the authors give a method to construct a succinct encoding, and corresponding data structure, for any class of binary trees $\mathcal{T}$ satisfying the following four conditions.

1. *Fringe-hereditary*: For any tree $\tau \in \mathcal{T}$ and node $v \in \tau$ the *fringe subtree* $\tau[v]$, which consists of v and all of its descendants in τ, also belongs to $\mathcal{T}$.
2. *Worst-case B-fringe dominated*: Most nodes in members of $\mathcal{T}$ do not generate large fringe subtrees, in the sense that

$$\left|\left\{v \in \tau : \left|\tau[v]\right| \geq B(n)\right\}\right| = o(n/\log B(n))$$

for every binary tree τ in the subset $\mathcal{T}_n \subset \mathcal{T}$ containing the members of $\mathcal{T}$ with n nodes, where $|\tau|$ denotes the number of nodes in τ.
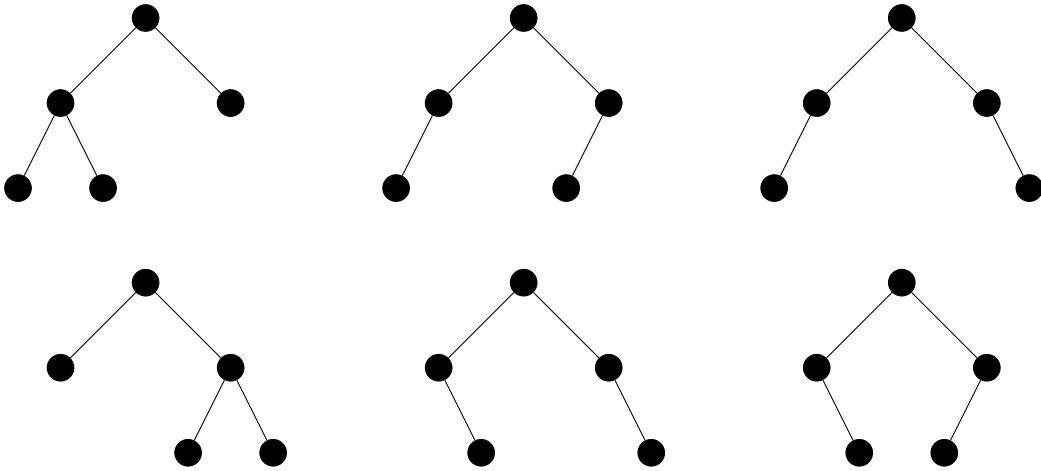
■ **Figure 2** The six types of AVL trees with $n = 5$ nodes.

3. *Log-linear*: There is a constant $c > 0$ and a function $\vartheta(n) = o(n)$ such that

$$\log |\mathcal{T}_n| = cn + \vartheta(n). \tag{1}$$

4. *B-heavy twigged*: If v is a node of any $\tau \in \mathcal{T}$ with $|\tau[v]| \geq B(n)$, and $\tau_\ell[v]$ and $\tau_r[v]$ are the left and right subtrees of v in τ, then $|\tau_\ell[v]|, |\tau_r[v]| = \omega(1)$.

We present a new construction that gives a succinct encoding for all classes of trees satisfying only the first three conditions. By using constant time rank and select operations already supported by a succinct encoding for binary trees, we can also eliminate the use of so-called "portal nodes" and thus relax the second condition to the following.

2′. *Worst-case weakly fringe dominated:* Most nodes in members of $\mathcal{T}$ do not generate large fringe subtrees, in the sense that there is a $B'(n)$ satisfying $B'(n) = d \log n + o(\log n)$ for some $d < 1$ such that

$$\left| \{ v \in \tau : \left| \tau[v] \right| \geq B'(n) \} \right| = o(n) \tag{2}$$

for every binary tree $\tau \in \mathcal{T}_n$.

Adopting terminology similar to that of [9], we call a class of binary trees *weakly tame* if it is fringe-hereditary, worst-case weakly fringe dominated, and log-linear.

▶ **Theorem 1.** *If $\mathcal{T}$ is a weakly tame class of binary trees then there exists a succinct encoding for $\mathcal{T}$ that supports the operations in Table 1 in $O(1)$ time using the $(\log n)$-bit word RAM model.*

▶ Remark 2. We support operations on static trees, leaving extensions to trees with updates (such as in [10]) to future work.

Proof. See Section 2.   ◀

▶ **Corollary 3.** *There exists a succinct encoding for AVL trees that supports the operations in Table 1 in $O(1)$ time using the $(\log n)$-bit word RAM model.*

Proof. AVL trees are weakly tame (see [9, Example F.2]) so the result follows immediately from Theorem 1.   ◀

▶ Remark 4. In [9] the log-linearity of AVL trees is inferred from the stated exponential growth of a_n in Odlyzko [13]. This growth is proven in Theorem 6 below.

A minor modification of the arguments in [9] show that *Left-Leaning AVL (LLAVL) Trees*, which are AVL trees with the added restriction that at every node the height of the left subtree is at least the height of the right subtree, are also weakly tame, giving the following.

▶ **Corollary 5.** *There exists a succinct encoding for LLAVL trees that supports the operations in Table 1 in $O(1)$ time using the $(\log n)$-bit word RAM model.*

■ **Table 1** Operations discussed in [5, 9] which can be done in $O(1)$ time in the $(\log n)$-bit word RAM model in a succinct encoding of a binary tree.

`parent(`v`)`	the parent of v, same as `anc(`$v, 1$`)`
`degree(`v`)`	the number of children of v
`left_child(`v`)`	the left child of node v
`right_child(`v`)`	the right child of node v
`depth(`v`)`	the depth of v, i.e., the number of edges between the root and v
`anc(`v, i`)`	the ancestor of node v at depth `depth(`v`)` $- i$
`nbdesc(`v`)`	the number of descendants of v
`height(`v`)`	the height of the subtree rooted at node v
`LCA(`v, u`)`	the lowest common ancestor of nodes u and v
`leftmost_leaf(`v`)`	the leftmost leaf descendant of v
`rightmost_leaf(`v`)`	the rightmost leaf descendant of v
`level_leftmost(`ℓ`)`	the leftmost node on level ℓ
`level_rightmost(`ℓ`)`	the rightmost node on level ℓ
`level_pred(`v`)`	the node immediately to the left of v on the same level
`level_succ(`v`)`	the node immediately to the right of v on the same level
`node_rank`$_X$`(`v`)`	the position of v in the X-order, $X \in \{$`PRE, POST, IN`$\}$, i.e., in a preorder, postorder, or inorder traversal of the tree
`node_select`$_X$`(`i`)`	the ith node in the X-order, $X \in \{$`PRE, POST, IN`$\}$
`leaf_rank(`v`)`	the number of leaves before and including v in preorder
`leaf_select(`i`)`	the ith leaf in preorder

To characterize how much space is required by a succinct encoding, we derive an asymptotic bound on the number of AVL trees using techniques from *analytic combinatorics* [4, 7]. To this end, let $a_n = |\mathcal{A}_n|$ be the counting sequence of $\mathcal{A}$ and let $A(z) = \sum_{n \geq 0} a_n z^n$ be its associated generating function. The key to enumerating AVL trees is to let $A_h(z)$ be the generating function for the subclass of AVL trees with *height h*. The balance condition on subtrees implies that an AVL tree of height $h + 2$ is a root together with a subtree of height $h + 1$ and a subtree of height either $h + 1$ or h, giving rise to the recursive equation

$$A_{h+2}(z) = A_{h+1}(z)(A_{h+1}(z) + 2A_h(z)) \tag{3}$$

for all $h \geq 0$, where the factor of 2 indicates that the shorter subtree can be on the left or right side. This recursion, along with the initial conditions $A_0(z) = z$ (encoding the only AVL tree with height zero, which is a single vertex) and $A_1(z) = z^2$ (encoding the only AVL tree with height one, which is a root with two children) uniquely determines $A_h(z)$ for all h. Summing over all possible heights gives the generating function

$$A(z) = \sum_{h=0}^{\infty} A_h(z)$$

for AVL trees.
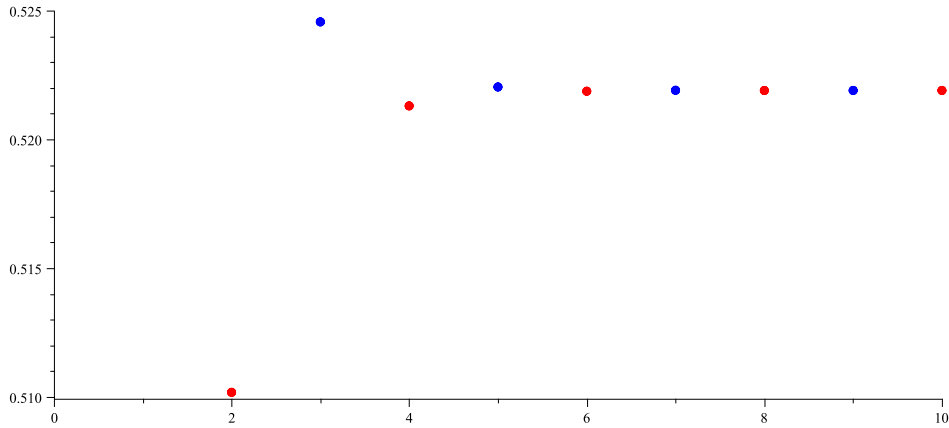
■ **Figure 3** Values α_h converging to $\alpha = 0.5219\ldots$ monotonically from below among even h (red) and monotonically from above among odd h (blue).

Equation (3) implies that $A_h(z)$ is a non-constant polynomial with positive coefficients for all h, so the equation $A_h(z) = 1/3$ has a unique positive solution for all $h \in \mathbb{N}$ (see Figure 3 for values of these solutions). We prove the following.

▶ **Theorem 6.** *If α_h is the unique positive solution to $A_h(z) = 1/3$ then the limit*

$$\alpha = \lim_{h \to \infty} \alpha_h = 0.5219\ldots$$

exists. Furthermore,

$$\log_2(a_n) = \underbrace{n \log_2(\alpha^{-1})}_{n(0.938\ldots)} + \log \theta(n)$$

for a function θ growing at most sub-exponentially (meaning $\theta(n) = o(\kappa^n)$ for all $\kappa > 1$).

Proof. The result follows immediately from applying Theorem 13 below with $f(x_1, x_2) = x_1^2 + 2x_1 x_2$, since the unique positive solution to $f(C, C) = C$ is $C = 1/3$. ◀

▶ Remark 7. A full proof of the claimed asymptotic behaviour $a_n \sim n^{-1} \alpha^{-n} u(\log n)$ in Odlyzko [13], which characterizes *sub-dominant* asymptotic terms for the bitsize, requires a more intense study of the recursion (3) and is outside the scope of this discussion. It is postponed to future work.

Our approach derives asymptotics for a family of generating functions satisfying recursive equations similar to (3). For instance, if $L_h(z)$ is the generating function for LLAVL trees with height h then

$$L_{h+2}(z) = L_{h+1}(z)(L_{h+1}(z) + L_h(z)) \tag{4}$$

for all $h \geq 0$, as an LLAVL tree of height $h + 2$ is a root together with a left subtree of height $h + 1$ and a right subtree of height $h + 1$ or h. Note that the only difference between this recurrence and the recursive equation (3) for AVL trees is the coefficient of $L_h(z)$, since there is now only one way to have an unbalanced pair of subtrees.

▶ **Theorem 8.** *If γ_h is the unique positive solution to $L_h(z) = 1/2$ then the limit*

$$\gamma = \lim_{h \to \infty} \gamma_h = 0.67418\ldots$$

is well-defined. Furthermore, the number ℓ_n of LLAVL trees on n nodes satisfies

$$\log_2(\ell_n) = \underbrace{n \log_2(\gamma^{-1})}_{n(0.568\ldots)} + \log \theta(n)$$

for a function θ growing at most sub-exponentially.

Proof. The result follows by applying Theorem 13 below with $f(x_1, x_2) = x_1^2 + x_1 x_2$, since the unique positive solution to $f(C, C) = C$ is $C = 1/2$. ◀

2 A New Succinct Encoding for Weakly Tame Classes

We now prove Theorem 1, first describing our encoding and then showing it has the stated properties.

2.1 Our encoding

Let $\mathcal{E}$ denote a succinct data structure representing all binary trees that supports the operations in Table 1, and denote the encoding of a binary tree τ in this data structure by $\mathcal{E}(\tau)$. We now fix a weakly tame class of binary trees $\mathcal{T}$ and, given a binary tree $\tau \in \mathcal{T}$ of size n, define the *upper tree*

$$\tau' = \left\{ v \in \tau : \left| \tau[p(v)] \right| \geq d \log n \right\}$$

where $p(v)$ denotes the parent of a vertex v in the tree τ and d is a constant such that $B'(n) = d \log n + o(\log n)$ satisfies (2) in the definition of worst-case weakly fringe dominated.

Our succinct data structure for $\mathcal{T}$ is constructed as follows.

1. We simply copy the encodings $\mathcal{E}(\tau')$ for upper trees.
2. For every $1 \leq j < d \log n$ we write down a lookup table mapping the trees in $\mathcal{T}_j$ (with j nodes) to their corresponding $\mathcal{E}$ encoding. We can do this, for example, by enumerating the $\mathcal{T}_j$ in lexicographic order by the $\mathcal{E}$ encoding using integers of bitsize $\log |\mathcal{T}_j| = cj + o(j)$, where c is the constant in the definition of log-linearity (1).
3. For each leaf node $\ell \in \tau'$ the fringe subtree $\tau[\ell]$ has size $|\tau[\ell]| < d \log n$ by definition of τ'. We call these trees *lower trees*, and write them down using their encoding in a lookup table in `leaf_rank` order of their roots in τ', storing the root locations in an indexable dictionary.
4. Lastly, we store additional information in (fully) indexable dictionaries to support operations like `node_rank/select`, `level_succ/pred`, and `leaf_rank/select`. For instance, for `node_rank/select` we store a fully indexable dictionary that maps the `node_rank` for a node in τ' to the `node_rank` of the node in τ. The techniques to support the other operations are similar, and are analogous to constructions used in [5, 3].

2.2 Proof of Size and Operation Time Bounds

Navigation through the upper tree follows standard navigation using $\mathcal{E}$, which supports the desired operations in constant time. When a leaf node ℓ is reached in the upper tree, the operation $x = \texttt{leaf\_rank}(\ell)$ gives the index of the child tree in the indexable dictionary.

Then the operation $\mathtt{select}(x)$ gives the location of the string encoding the child tree. Finally, using the table mapping our encoding to the $\mathcal{E}$ encoding gives us the ability to perform all the navigation operations on the smaller tree. In order to perform the lookup using the mapping, it is necessary to know the size of the tree. This can be inferred from the space in memory allocated to the naming, which can be calculated by the operation $\mathtt{select}(x+1)$ in the indexable dictionary to find the starting location of the next child tree. To navigate back to the upper tree from a child tree, we use the reverse operations of $y = \mathtt{rank}(x)$ in the indexable dictionary followed by $\mathtt{select\_leaf}(y)$ in the upper tree.

To get the $\mathtt{node\_rank}$ of a node in τ' we use the fully indexable dictionary, and to get the $\mathtt{node\_rank}$ of a node not in τ' we simply get the $\mathtt{node\_rank}$ of the root of the child tree and the $\mathtt{node\_rank}$ of the node within the child tree and perform the appropriate arithmetic depending on the desired rank order ($\mathtt{pre, post, in}$). For $\mathtt{node\_select}$, if the node is in τ' then selecting using the indexable dictionary is sufficient. Otherwise, the node is in a child tree and the initial $\mathtt{node\_select}$ will return the predecessor node in τ' which will be the root of the child tree when using $\mathtt{preorder}$ (the argument is similar for $\mathtt{postorder}$ and $\mathtt{inorder}$). Using the rank of this root and appropriate arithmetic, we can then select the desired node in the child tree. Implementing the other operations is analogous. It is clear that all of these operations are supported in constant time, since they involve a constant number of calls to the constant-time operations in the existing data structures, and lookups using $(\log n)$-bit words.

Space Complexity

The space used by $\mathcal{E}(\tau')$ is $o(n)$ by the weakly tame property. The space used by the lookup tables is $O(n^d \log n) = o(n)$ by definition of τ' and d, and the space used by all of the encodings of the child trees is $cn + o(n)$ by log-linearity. Lastly, the space needed for the indexable dictionaries is $o(n)$ for each [3, Lemmas 1 and 2]. Summing these requirements shows that the total storage required is $cn + o(n)$ many bits, so the encoding is succinct. ◄

3   Asymptotics for a Family of Recursions

We derive the asymptotic behaviour of a family of generating functions which includes Theorem 6 as a special case. Let $\mathcal{F}$ be a combinatorial class decomposed into a disjoint union of finite subclasses $\mathcal{F} = \bigsqcup_{h=0}^{\infty} \mathcal{F}_h$ whose generating functions $F_h(z)$ are non-constant and satisfy a recursion

$$F_h(z) = f(F_{h-1}(z), F_{h-2}(z), \ldots, F_{h-c}(z)) \quad \text{for all} \quad h \geq c, \tag{5}$$

where c is a positive integer and f is a multivariate polynomial with non-negative coefficients.

▶ Remark 9. The elements of $\mathcal{F}_h$ are usually *not* the objects of $\mathcal{F}$ of size h (in our tree applications they contain trees of height h, not trees with h nodes). The fact that each $\mathcal{F}_h$ is finite implies that the $F_h(z)$ are polynomials with non-negative coefficients. The coefficient of z^n in $F_h(z)$ counts the number of objects of size n within the subclass indexed by h.

We assume that there exists a (necessarily unique) positive real solution C to the equation $C = f(C, C, \ldots, C)$, and for each $h \geq 0$ we let α_h be the unique positive real solution to $F_h(z) = C$. In order to rule out degenerate cases and cases where the counting sequence has periodic behaviour, we need another definition.
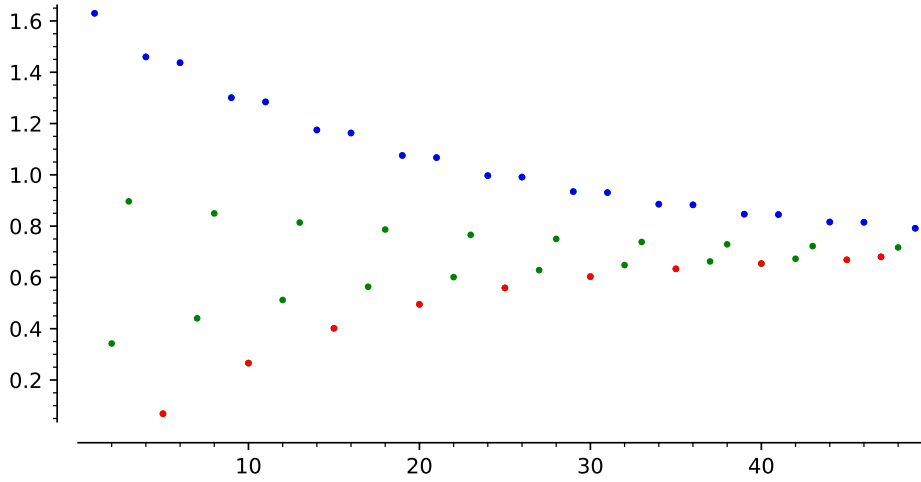
Figure 4 Values α_i converging with u_is shown in blue and ℓ_js shown in red.

▶ **Definition 10** (recursive-dependent). *We call the polynomial f recursive-dependent if there exists a constant k (depending only on f) such that for any indices $i, j \geq c$ with $i \geq j + k$ there exists a sequence of applications of the recurrence* (5) *resulting in a polynomial P with $F_i = P(F_{\ell_1}, \ldots, F_{\ell_m})$ for some $0 \leq \ell_1 < \cdots < \ell_m \leq i$ where $\frac{\partial P}{\partial F_j} \neq 0$.*

▶ **Example 11.** The polynomial $f(x, y) = y$ is not recursive-dependent because it leads to the recursion $F_h(z) = F_{h-2}(z)$, meaning that the values of F_h when h is even can be independent of those where h is odd.

▶ **Lemma 12.** *If f is recursive-dependent with non-negative coefficients and a positive fixed point then the limit $\alpha = \lim_{h \to \infty} \alpha_h$ exists.*

Proof. We start by defining two subsequences of α_h to give upper and lower bounds on its limit, then prove that these are equal. First, we let

  ▬ u_0 be the smallest index $j \in \{0, \ldots, c-1\}$ such that $\alpha_j = \max\{\alpha_0, \ldots, \alpha_{c-1}\}$

and for all $i \geq 0$ let

  ▬ u_{i+1} be the smallest index $j \in \{u_i + 1, \ldots, u_i + c\}$ such that $\alpha_j = \max\{\alpha_{u_i+1}, \ldots, \alpha_{u_i+c}\}$,

so that the u_i denote the indices of the maximum values of the α_h as h ranges over intervals of size at most c. Conversely, we let

  ▬ ℓ_0 be the index $j \in \{0, \ldots, c-1\}$ such that $\alpha_j = \min\{\alpha_0, \ldots, \alpha_{c-1}\}$

and for all $j \geq 0$ let

  ▬ ℓ_{i+1} be the index $j \in \{u_i + 1, \ldots, u_i + c\}$ such that $\alpha_j = \min\{\alpha_{u_i+1}, \ldots, \alpha_{u_i+c}\}$,

so that the ℓ_j denote the indices of the minimum values of the α_h as h ranges over intervals of size at most c.

We claim that the subsequence α_{u_i} is non-increasing. To establish this, we fix $i \geq 1$ and consider α_{u_i}. By definition, $\alpha_{u_i} \geq \alpha_{u_j}$ for all $j \in \{u_{i-1} + 1, \ldots, u_{i-1} + c\}$. Thus, if $u_{i+1} \in \{u_{i-1} + 1, \ldots, u_{i-1} + c\}$ then $\alpha_{u_i} \geq \alpha_{u_{i+1}}$ as claimed. If, on the other hand, $u_{i+1} > u_{i-1} + c$ then repeated application of the recursion (5) implies

$$F_{u_{i+1}}(\alpha_{u_i}) = f\Big(F_{u_{i+1}-1}(\alpha_{u_i}), \ldots, F_{u_{i+1}-c}(\alpha_{u_i})\Big)$$

$$\vdots$$

$$= Q\Big(F_{u_{i-1}+1}(\alpha_{u_i}), \ldots, F_{u_{i-1}+c}(\alpha_{u_i})\Big),$$

where Q is a multivariate polynomial with non-negative coefficients such that $Q(C, \ldots, C) = C$. All the F_h are monotonically increasing as non-constant polynomials with non-negative coefficients, so $F_j(\alpha_{u_i}) \geq F_j(\alpha_{u_j}) = C$ for all $j \in \{u_{i-1} + 1, \ldots, u_{i-1} + c\}$ and

$$F_{u_{i+1}}(\alpha_{u_i}) \geq Q\Big(C, \ldots, C\Big) = C.$$

Since $F_{u_{i+1}}$ is monotonically increasing and $F_{u_{i+1}}(\alpha_{u_{i+1}}) = C$, we once again see that $\alpha_{u_i} \geq \alpha_{u_{i+1}}$. As i was arbitrary, we have proven that α_{u_i} is non-increasing. The same argument, reversing inequalities, proves that the subsequence α_{ℓ_j} is non-decreasing.

As α_{ℓ_j} is non-decreasing and α_{u_i} is non-increasing, either $\alpha_{\ell_j} \leq \alpha_{u_i}$ for all $i, j \geq 0$ or $\alpha_{\ell_j} > \alpha_{u_i}$ for all sufficiently large i and j. The second case implies the existence of indices $a, b > 0$ such that $\alpha_{\ell_b} > \alpha_{u_a}$ but $\ell_b \in \{u_{a-1} + 1, \ldots, u_{a-1} + c\}$ so that u_a is not the maximum index of α_j in this range, giving a contradiction. Thus, $\alpha_{\ell_j} \leq \alpha_{u_i}$ for all $i, j \geq 0$ and the limits

$$\alpha_u = \lim_{i \to \infty} \alpha_{u_i} \quad \text{and} \quad \alpha_\ell = \lim_{j \to \infty} \alpha_{\ell_j}$$

exist. To prove that the limit of α_h exists as $h \to \infty$, it is now sufficient to prove that $\alpha_u = \alpha_\ell$.

Suppose toward contradiction that $\alpha_u \neq \alpha_\ell$, and define $a = \alpha_u - \alpha_\ell > 0$. For any $\epsilon > 0$, we pick i, j, k sufficiently large so that $\ell_j > u_i > \ell_k + c$ and $|\alpha_{u_i} - \alpha_u|, |\alpha_{\ell_j} - \alpha_\ell|, |\alpha_{\ell_k} - \alpha_\ell| < \epsilon$. Then by recursive-dependence we can recursively decompose F_{ℓ_j} in terms of F_{u_i}, and possibly some other terms $F_{h_1}, \ldots, F_{h_r}$ where each $|h_n - u_i| \leq c$, to get

$$C = F_{\ell_j}(\alpha_{\ell_j}) = P(F_{u_i}(\alpha_{\ell_j}), F_{h_1}(\alpha_{\ell_j}), \ldots, F_{h_r}(\alpha_{\ell_j}))$$

where $P(F_{u_i}, F_{h_1}, \ldots, F_{h_r})$ is a polynomial with non-negative coefficients that depends on F_{u_i} and satisfies $P(C, \ldots, C) = C$. Because P is monotonically increasing in each coordinate, and $\alpha_{\ell_k} + \epsilon > \alpha_\ell \geq \alpha_{\ell_j}$, we see that

$$C \leq P(F_{u_i}(\alpha_{\ell_k} + \epsilon), F_{h_1}(\alpha_{\ell_k} + \epsilon), \ldots, F_{h_r}(\alpha_{\ell_k} + \epsilon)).$$

Furthermore, each $\alpha_{h_n} \geq \alpha_{\ell_k}$ so

$$C \leq P(F_{u_i}(\alpha_{\ell_k} + \epsilon), F_{h_1}(\alpha_{h_1} + \epsilon), \ldots, F_{h_r}(\alpha_{h_r} + \epsilon))$$
$$\leq P(F_{u_i}(\alpha_{\ell_k} + \epsilon), C + \text{poly}(\epsilon), \ldots, C + \text{poly}(\epsilon)).$$

Finally, $\alpha_{u_i} - a \geq \alpha_{\ell_k}$ so

$$C \leq P(F_{u_i}(\alpha_{u_i} - a + \epsilon), C + \text{poly}(\epsilon), \ldots, C + \text{poly}(\epsilon)).$$

Because a is fixed, P is monotonically increasing in each variable, and $F_{u_i}(\alpha_{u_i}) = C$, taking $\epsilon \to 0$ shows that the right-hand side of this last inequality is strictly less than $P(C, \ldots, C) = C$, a contradiction. Thus, $a = 0$ and the limit $\alpha = \alpha_u = \alpha_\ell$ exists. ◀

▶ **Theorem 13.** *If f is recursive-dependent with non-negative coefficients and a positive fixed point, then the number a_n of objects in $\mathcal{F}$ of size n satisfies*

$$a_n = \alpha^{-n} \theta(n),$$

where α is the limit described in Lemma 12 and $\theta(n)$ is a function growing at most sub-exponentially.

Proof. We prove that the generating function $F(z)$ is analytic for $|z| < \alpha$ by showing that the series $\sum_{h=0}^{\infty} F_h(z)$ converges for these values of z. Because $|F(z)| \to \infty$ as $z \to \alpha$, the point $z = \alpha$ is then a singularity of $F(z)$ of smallest modulus, and thus (by the root test for series convergence) the reciprocal of the exponential growth of a_n.

First, assume that there exists some $k \geq 0$ and $0 < \lambda < 1$ such that $F_h(z) < \lambda C$ for every $h \in \{k, k+1, \ldots, k+c-1\}$. Let A be the sum of the coefficients of all degree 1 terms of f. Since f has non-negative coefficients and a positive real fixed point, we must have $A < 1$. Let $g(x_1, \ldots, x_c)$ be the function created by removing all degree one terms from f. Observe that $C = AC + g(C, \ldots, C)$, and thus $g(\lambda C, \ldots, \lambda C) \leq \lambda^2 g(C, \ldots, C) = \lambda^2 (1-A)C$, so that

$$f(\lambda C, \ldots, \lambda C) \leq A\lambda C + \lambda^2 (1-A)C.$$

Algebraic manipulation shows that $A\lambda + \lambda^2(1-A) \leq \lambda$, and since f has non-negative coefficients we can conclude that for every $h \in \{k+c, k+1+c, \ldots, k+2c-1\}$ we have $F_h(z) \leq A\lambda C + \lambda^2(1-A)C$. Let $\lambda_0 = \lambda$ and define $\lambda_i = \lambda_{i-1}(A + \lambda_{i-1} - A\lambda_{i-1})$ for all $i \geq 1$. By the above argument we have

$$F_{ch+k}(z) \leq \lambda_h C,$$

so it remains to show that $\sum_{i=0}^{\infty} \lambda_i$ converges. We will show that $\lambda_i \leq \lambda(A + \lambda - A\lambda)^i$ by induction on i. The result holds by definition for $i = 1$. If the result holds for some $j \geq 1$ then

$$\begin{aligned}
\lambda_{j+1} &= \lambda_j(A + \lambda_j - A\lambda_j) \\
&\leq \lambda(A + \lambda - A\lambda)^j(A + \lambda_j - A\lambda_j) \\
&\leq \lambda(A + \lambda - A\lambda)^{j+1},
\end{aligned}$$

where the last inequality follows from the fact that $\lambda_j < \lambda$ since $A + \lambda - A\lambda < 1$. The sum $\sum_{i=0}^{\infty} \lambda(A + \lambda - A\lambda)^i$ converges as a geometric series, and thus $\sum_{h=0}^{\infty} F_h(z)$ converges.

It remains to show that if $|z| < \alpha$ then such a k and λ exist. For any $|z| < \alpha$ there is some N sufficiently large such $|z| < \alpha_n$ for all $n \geq N$. By the definition of α_n, and since the coefficients of F_n are all positive, we must have $F_n(z) < C$. Hence $F_n(z) < \lambda_n C$ for some $0 < \lambda_n < 1$. Taking $k = N$ and letting λ be the largest λ_n for $n \in \{N, N+1, \ldots, N+c-1\}$ proves our final claim. ◄

---- **References** ---

1 G. Adelson-Velsky and E. Landis. An algorithm for the organization of information. In *Proceedings of the USSR Academy of Sciences (in Russian)*, 1962.

2 Clark, David. *Compact PAT trees*. PhD thesis, University of Waterloo, 1997. URL: `http://hdl.handle.net/10012/64`.

3 Arash Farzan and J. Ian Munro. A uniform paradigm to succinctly encode various families of trees. *Algorithmica*, 68(1):16–40, January 2014. `doi:10.1007/s00453-012-9664-0`.

4 Philippe Flajolet and Robert Sedgewick. *Analytic combinatorics*. Cambridge University Press, Cambridge, 2009.

5 Meng He, J. Ian Munro, and S. Srinivasa Rao. Succinct ordinal trees based on tree covering. In *Automata, Languages and Programming*, pages 509–520. 2007.

6 G. Jacobson. Space-efficient static trees and graphs. In *30th Annual Symposium on Foundations of Computer Science*, pages 549–554, 1989.

7 Stephen Melczer. *An Invitation to Analytic Combinatorics: From One to Several Variables*. Texts and Monographs in Symbolic Computation. Springer International Publishing, 2021.

8   J. Ian Munro. Tables. In Vijay Chandru and V. Vinay, editors, *Foundations of Software Technology and Theoretical Computer Science, 16th Conference, Hyderabad, India, December 18-20, 1996, Proceedings*, volume 1180 of *Lecture Notes in Computer Science*, pages 37–42. Springer, 1996. `doi:10.1007/3-540-62034-6_35`.

9   J. Ian Munro, Patrick K. Nicholson, Louisa Seelbach Benkner, and Sebastian Wild. Hypersuccinct trees – new universal tree source codes for optimal compressed tree data structures and range minima. In *29th Annual European Symposium on Algorithm*, pages 70:1–70:18, 2021.

10   J. Ian Munro, Venkatesh Raman, and Adam J. Storm. Representing dynamic binary trees succinctly. In S. Rao Kosaraju, editor, *Proceedings of the Twelfth Annual Symposium on Discrete Algorithms, January 7-9, 2001, Washington, DC, USA*, pages 529–536. ACM/SIAM, 2001. URL: `http://dl.acm.org/citation.cfm?id=365411.365526`.

11   J. Ian Munro and S. Srinivasa Rao. Succinct representation of data structures. In Dinesh P. Mehta and Sartaj Sahni, editors, *Handbook of Data Structures and Applications*. Chapman and Hall/CRC, 2004. `doi:10.1201/9781420035179.ch37`.

12   Gonzalo Navarro. *Compact Data Structures: A Practical Approach*. Cambridge University Press, 2016.

13   A. M. Odlyzko. Some new methods and results in tree enumeration. In *Proceedings of the thirteenth Manitoba conference on numerical mathematics and computing (Winnipeg, Man., 1983)*, volume 42, pages 27–52, 1984.

Maximal Number of Subword Occurrences in a Word

Wenjie Fang ✉ 🏠 ⓘ

Univ Gustave Eiffel, CNRS, LIGM, F-77454 Marne-la-Vallée, France

—————— **Abstract** ——————

We consider the number of occurrences of subwords (non-consecutive sub-sequences) in a given word. We first define the notion of subword entropy of a given word that measures the maximal number of occurrences among all possible subwords. We then give upper and lower bounds of minimal subword entropy for words of fixed length in a fixed alphabet, and also showing that minimal subword entropy per letter has a limit value. A better upper bound of minimal subword entropy for a binary alphabet is then given by looking at certain families of periodic words. We also give some conjectures based on experimental observations.

1 Introduction

Enumeration problems concerning patterns have been rich sources of interesting combinatorics. The most famous examples are classes of permutations avoiding a given pattern. We refer readers to [9, 14] for an exposition of such results. In this article, we will consider enumeration about patterns in a word, which is in general easier than that for permutations.

There are two different widely used notions of patterns for words. The first notion is that of a *factor*. A word v occurs in another word w *as a factor* if there is a consecutive segment of w equal to v. The second notion is that of a *subword*. A word v occurs in another word w *as a subword* if we can obtain v by deleting letters in w. A factor of w is always a subword of w, but not *vice versa*. There are also other notions of patterns, such as the one in [2] that generalizes both factors and subwords, but we will not discuss them here.

Unlike for permutations, the enumeration of classes of words avoiding a (set of) given subwords or factors is already known in the sense that, for a given subword or factor, we can express their avoidance in regular expressions, leading automatically (no pun intended) to the generating function of such classes, which is always rational and can be effectively computed [5, Section V.5]. There is also some work on counting words with a fixed number of occurrences of a given pattern, for example [2, 11]. For the other end of the spectrum, the problem of maximal density of certain patterns in words is considered by Burstein, Hästö and Mansour in [1]. Readers are referred to the survey-book of Kitaev [9] for more of such results. In general, such results are non-trivial, due to the possible overlap of patterns.

We may also consider all patterns that occur in a given word. For the notion of factors, this idea leads to the notion of factor complexity of a word w, first defined by Morse and Hedlund in [13] and also called "subword complexity", which is a function f_w such that $f_w(k)$ is the number of distinct factors in w of length k. In [7], Gheorghiciuc and Ward studied the factor complexity of random words. We may also want to consider the number of occurrences of a given pattern in a word. The work of Flajolet, Szpankowski and Vallée [6] establishes a Gaussian limit law and large deviations for the number of occurrences of a given subword in a long random word, again by analyzing overlap of occurrences of subwords. The number of occurrences of a given pattern is of particular interest in algorithmics with applications in data mining, in which researchers propose algorithms finding patterns with large number of occurrences [8] and study complexity of such problems [15].

In this article, we take a further step on enumeration problems on patterns by considering the number of occurrences of all subwords in a word. More precisely, we can see a given word w that permits frequent occurrences of some subword w' as having some "large space" for such a subword, and we would like to measure the "extend" of such space, or from the opposite direction, the "disorder" generated by the possible different occurrences. To this end, we define a notion of *subword entropy*, which measures the maximal number of times that any subword can occur in a given word. We delay its precise definition to later sections. We then look at the minimal subword entropy of all words of a given length n in an alphabet of k letters, denoted by $\min S_{\mathrm{sw}}^{(k)}(n)$, as it is easy to find the ones with maximal subword entropy. Using the super-additivity of minimal subword entropy, we show that $\min S_{\mathrm{sw}}^{(k)}(n)/n$ has a finite limit L_k. We then concentrate on the binary case, showing some upper bounds of L_2 by looking at certain families of periodic words, inspired by experimental data. As a by-product, we also show that, given two words w and v, the generating function of the number of occurrences of v^r in w^m is rational.

The rest of this article is organized as follows. We first give necessary definitions in Section 2, then some basic results on subword occurrences and minimal subword entropy in Section 3, including the proof of the existence of the limit L_k of $\min S_{\mathrm{sw}}^{(k)}(n)/n$, and bounds of L_k. Then in Section 4, we focus on the case of binary alphabet, and shows a better upper bound of L_2 than the one given in Section 3. We end in Section 5 with a discussion on open problems partially inspired by experimental results obtained for the binary case.

2   Preliminaries

A *word* w of length n is a sequence $w = (w_1, \ldots, w_n)$ of elements in a finite set $\mathcal{A}$ called the *alphabet*. We denote by $|w|$ the length of w, and $|w|_a$ the number of letters a in w. For two words v, w, their concatenation is denoted by $v \cdot w$. We also denote by ϵ the empty word of length 0. A *run* in a word w is a maximal consecutive segment in w formed by only one letter in $\mathcal{A}$. Given a word w, if there is another word $w' = (w'_1, \ldots, w'_k)$ such that there is some set $P = \{p_1 < \cdots < p_k\}$ of integers from 1 to n satisfying $w_{p_j} = w'_j$ for all $1 \leq j \leq k$, then we say that w' is a *subword* of w, and we call the set P an *occurrence* of w' in w. We denote by $\mathrm{occ}(w, w')$ the number of occurrences of w' in w. For instance, for $w = 011001$ and $w' = 01$, there are 5 occurrences of w' in w, which are $\{1, 2\}, \{1, 3\}, \{1, 6\}, \{4, 6\}, \{5, 6\}$. When w' is not a subword of w, we have $\mathrm{occ}(w, w') = 0$, and when $w' = \epsilon$, we have $\mathrm{occ}(w, \epsilon) = 1$.

It is easy to find words who have a subword with a large number of occurrences. For instance, with $w = a^n$ for some letter $a \in \mathcal{A}$, the subword $w' = a^{\lfloor n/2 \rfloor}$ appears $\binom{n}{\lfloor n/2 \rfloor} \sim \left(\frac{2}{\pi n}\right)^{1/2} 2^n$ times. It is more difficult to find words in which no subword occurs frequently. To quantify such intuition, we define the *maximal subword occurrences* $\mathrm{maxocc}(w)$ of a word

w to be the maximal value of $\mathrm{occ}(w, w')$, and subwords w' reaching this value are called *most frequent subwords* of w. We note that a word w may have several most frequent subwords. We then define the *subword entropy* $S_{\mathrm{sw}}(w)$ of w in an alphabet of size k by

$$S_{\mathrm{sw}}(w) := \log_2 \mathrm{maxocc}(w).$$

We note that this definition does not depend on the size of the alphabet, as subword occurrences are fundamentally about subsets of positions, and the size of the alphabet is implicit in the word w. Now, finding words in which no subword occurs frequently is to find words minimizing their subword entropy. We define the minimal subword entropy for words of length n in an alphabet of size k by

$$\min S_{\mathrm{sw}}^{(k)}(n) := \min_{w \in \mathcal{A}^n, |\mathcal{A}|=k} S_{\mathrm{sw}}^{(k)}(w).$$

3   Some basic results

We start with some simple properties of $\mathrm{occ}(w, u)$.

▶ **Lemma 3.1.** *For words w, w', u, u', we have $\mathrm{occ}(w \cdot w', u \cdot u') \geq \mathrm{occ}(w, u) \, \mathrm{occ}(w', u')$.*

Proof. Let P (resp. P') be an occurrence of u (resp. u') in w (resp. w'). The set $Q = P \cup \{p' + |w| \mid p' \in P'\}$ is an occurrence of $u \cdot u'$ in $w \cdot w'$, and the map $(P, P') \mapsto Q$ is clearly injective. ◀

▶ **Lemma 3.2.** *For a word w, it has a most frequent subword u with $w_1 = u_1$ and $w_{|w|} = u_{|w'|}$.*

Proof. Let v be a most frequent subword of w. If $v_1 \neq w_1$, then for all occurrences P of v in w, we have $1 \notin P$. Then, $\{1\} \cup P$ is an occurrence of $v' = w_1 \cdot v$, which is thus also a most frequent subword. Otherwise, we take $v' = v$. We repeat the same reasoning on v' for the last letter of w to obtain u. ◀

We now give simple upper and lower bounds for $\mathrm{maxocc}(w)$ for any word w.

▶ **Proposition 3.3.** *Given an alphabet $\mathcal{A}$ of size k and $n \geq 1$, for any word $w \in \mathcal{A}^n$, we have $\mathrm{maxocc}(w) \leq \binom{n}{\lceil n/2 \rceil}$, and it is realized exactly by $w = a^n$ for any letter $a \in \mathcal{A}$.*

Proof. For w' of length k, as occurrences of w' in w are subsets of $\{1, \ldots, n\}$, we have $\mathrm{occ}(w, w') \leq \binom{n}{k} \leq \binom{n}{\lceil n/2 \rceil}$. It is clear that only words composed by the same letter reach this bound. ◀

▶ **Proposition 3.4.** *Given an alphabet $\mathcal{A}$ of size k and $n \geq 1$, when $n \to \infty$, for any word $w \in \mathcal{A}^n$, we have $\ln \mathrm{maxocc}(w) \geq \max_{0 \leq \ell \leq n} \ln \left(\binom{n}{\ell} k^{-\ell} \right)$, with $\ell = \lfloor \frac{n}{k+1} \rfloor$ giving the asymptotically maximized value $n \ln(1 + k^{-1}) - \frac{1}{2}(\ln n) + O(1)$.*

Proof. Let u be a uniformly chosen word of length ℓ. We have

$$\mathbb{E}[\mathrm{occ}(w, u)] = \sum_{P \subseteq \{1,\ldots,n\}, |P|=\ell} \mathbb{P}[u \text{ occurs in } w \text{ at positions } P] = \binom{n}{\ell} k^{-\ell}.$$

The first equality is from linearity of expectation, and the second from the fact that u is uniformly chosen at random, and the probability does not depend on P. Hence, there is some u^* with $\mathrm{occ}(w, u^*) \geq \mathbb{E}[\mathrm{occ}(w, u)]$, implying the non-asymptotic part of our claim.

For the asymptotic part, take $\alpha = \ell/n$. Using Stirling's approximation, we have

$$\ln\left(\binom{n}{\ell}k^{-\ell}\right) = n\left[-\alpha - \ln\alpha - (1-\alpha)\ln(1-\alpha) - \alpha\ln k\right] - \frac{1}{2}\ln n + O(1).$$

The coefficient of n above is maximized for $\alpha = (k+1)^{-1}$, with value $\ln(1+k^{-1})$. We thus have our claim on the asymptotic growth. ◀

▶ **Corollary 3.5.** *There are constants c_1, c_2 such that, for all $n \in \mathbb{N}$ and $w \in \mathcal{A}^n$ with $|\mathcal{A}| = k \geq 2$, we have*

$$\log_2(1+k^{-1})n - \frac{1}{2}\log_2 n + c_1 \leq \min S_{\mathrm{sw}}^{(k)}(n) \leq S_{\mathrm{sw}}(w) \leq n - \frac{1}{2}\log_2 n + c_2.$$

Proof. The bounds on $S_{\mathrm{sw}}(w)$ result from combining Propositions 3.3 and 3.4 with $\ln\binom{n}{\lceil n/2\rceil} = n\ln 2 - \frac{1}{2}\ln n + O(1)$. The bounds for $\min S_{\mathrm{sw}}^{(k)}(n)$ then follows. ◀

We now show that there is a limit for $\min S_{\mathrm{sw}}^{(k)}(n)/n$. To this end, we need the well-known Fekete's lemma [4] for super-additive sequences.

▶ **Lemma 3.6.** *Suppose that a sequence $(g_n)_{n\geq 1}$ satisfies that, for all $n, m \geq 1$, we have $g_{n+m} \geq g_n + g_m$. Then, for $n \to +\infty$, the value of g_n/n either tends to $+\infty$ or converges to some limit L.*

We first show that the function $\min S_{\mathrm{sw}}^{(k)}(n)$ is super-additive.

▶ **Proposition 3.7.** *Given $k \geq 2$, for any $n, m \geq 1$, we have*

$$\min S_{\mathrm{sw}}^{(k)}(n+m) \geq \min S_{\mathrm{sw}}^{(k)}(n) + \min S_{\mathrm{sw}}^{(k)}(m).$$

Proof. Let w be a word of length $n + m$ achieving minimal subword entropy $\min S_{\mathrm{sw}}^{(k)}(n+m)$. We write $w = w' \cdot w''$, with $|w'| = n$ and $|w''| = m$. Let v' (resp. v'') be a most frequent subword of w' (resp. w''). We have

$$\mathrm{maxocc}(w) \geq \mathrm{occ}(w' \cdot w'', v' \cdot v'') \geq \mathrm{occ}(w', v')\,\mathrm{occ}(w'', v'') = \mathrm{maxocc}(w')\,\mathrm{maxocc}(w'').$$

The first inequality is from the definition of maxocc, the second from Lemma 3.1, and the equality comes from the definition of v' and v''. By the definition of w, we have

$$\min S_{\mathrm{sw}}^{(k)}(n+m) \geq \log_2 \mathrm{maxocc}(w') + \log_2 \mathrm{maxocc}(w'') \geq \min S_{\mathrm{sw}}^{(k)}(n) + \min S_{\mathrm{sw}}^{(k)}(m).$$

The second inequality is from the definition of $\min S_{\mathrm{sw}}^{(k)}$. ◀

▶ **Theorem 3.8.** *For any $k \geq 2$, the sequence $(\min S_{\mathrm{sw}}^{(k)}(n)/n)_{n\geq 1}$ converges to a certain limit $L_k < +\infty$.*

Proof. Proposition 3.7 shows that $\min S_{\mathrm{sw}}^{(k)}(n)$ is super-additive. We then apply Lemma 3.6, and as $\min S_{\mathrm{sw}}^{(k)}(n)/n$ is bounded above by some constant according to Corollary 3.5, we have the existence of the limit L_k which is finite. ◀

With the existence of the limit L_k, we can use known values of $\min S_{\mathrm{sw}}^{(k)}(n)$ to give lower bounds for L_k.

▶ **Proposition 3.9.** *Given $k \geq 2$, we have $L_k \geq \min S_{\mathrm{sw}}^{(k)}(n)/n$ for all n.*

Proof. By iterating Proposition 3.7, we have $\min S_{\text{sw}}^{(k)}(rn) \geq r \min S_{\text{sw}}^{(k)}(n)$ for all $r \geq 1$. Diving both sides by rn, it means that the limit L_k of $\min S_{\text{sw}}^{(k)}(rn)/rn$ is also larger than $\min S_{\text{sw}}^{(k)}(n)$. ◀

From Corollary 3.5, we know that

$$\log_2(1 + k^{-1}) \leq L_k \leq 1.$$

When $k \to \infty$, the lower bound is asymptotically $(\ln 2)^{-1}k^{-1}$, which tends to 0, while the upper bound stays constant. The next natural step is to try to give better bounds for L_k, and eventually compute the precise value of L_k. However, it seems to be a formidable task.

4 Better upper bound for binary alphabet

After the general basic results given in Section 3, we will focus hereinafter on the case of binary alphabet $\mathcal{A} = \{0, 1\}$. In this case, the bounds in Corollary 3.5 become $\log_2(3/2) \leq L_2 \leq 1$ for the limit L_2 in Theorem 3.8. The gap between the two bounds are significant, as $\log_2(3/2) \approx 0.585$. We now give a better upper bound of L_2 by constructing a family of periodic words with a small value of maximal subword occurrences.

▶ **Proposition 4.1.** *For $w = (0011)^m$, there is a most frequent subword w' of the form $(01)^r$.*

Proof. Take a most frequent subword u of w of length ℓ. Suppose that u has the form $u = s \cdot 00 \cdot t$. We take $u^{(1)} = s \cdot 010 \cdot t$ and $u^{(2)} = s \cdot 0 \cdot t$. Let $P = \{p_1, \ldots, p_\ell\}$ be an occurrence of u in w, and we suppose that the 00 occurs at p_i, p_{i+1}. Let $\mathcal{P}$ be the set of occurrences of u in w, which is divided into $\mathcal{P} = \mathcal{P}_1 \cup \mathcal{P}_2$, where $\mathcal{P}_1$ contains those P's with $p_i + 1 \neq p_{i+1}$, while $\mathcal{P}_2$ contains those with $p_i + 1 = p_{i+1}$. For any $P \in \mathcal{P}_1$, the two 0's occur in different runs, meaning that there is at least one run 11 in between. This leads to at least two choices for the extra 1 added in $u^{(1)}$. Therefore, $\text{occ}(w, u^{(1)}) \geq 2|\mathcal{P}_1|$. For any $P \in \mathcal{P}_2$, the two 0's occur in the same run, meaning that replacing them by a single 0 leaves us two choices. We thus have $\text{occ}(w, u^{(2)}) \geq 2|\mathcal{P}_2|$, meaning that $\text{occ}(w, u^{(1)}) + \text{occ}(w, u^{(2)}) \geq 2|\mathcal{P}| = 2\text{maxocc}(w)$. We deduce that at least one of the $u^{(j)}$'s satisfies $\text{occ}(w, u^{(j)}) = \text{maxocc}(w)$. We observe that both $u^{(1)}$ and $u^{(2)}$ have one less pair of identical consecutive letters than u. We may then do the same for consecutive 1's. By iterating such a process, we get a most frequent subword without identical consecutive letters, thus alternating between 0 and 1. Then we conclude by Lemma 3.2. ◀

▶ **Remark 4.2.** We want to highlight the importance of Proposition 4.1 here. The main difficulty in the study of maximal subword occurrences is, in a sense, algorithmic. To the author's knowledge, we don't know whether there is a polynomial time algorithm to compute a most frequent subword of a given word, or to decide whether there is a subword that occurs at least a given number of times. However, in the case of words of the form $w = (0011)^m$, we manage to show some structure of their most frequent subwords, which then allows us to compute $\text{maxocc}(w)$.

Let $a_{m,r} = \text{maxocc}((0011)^m, (01)^r)$, and $f(x, y) = \sum_{m,r \geq 0} a_{m,r} x^m y^r$ be their generating function. We have the following counting result.

▶ **Proposition 4.3.** *We have*

$$f(x, y) = \frac{1 - x}{(1 - x)^2 - 4xy}, \quad a_{m,r} = 4^r \binom{m + r}{m - r}.$$

Proof. For an occurrence $P = \{p_1, \ldots, p_{2r}\}$ of $(01)^r$ in $(0011)^m$, we have two cases.

- $p_{2r} < 4(m-1)$, meaning that P is also an occurrence of $(01)^r$ in $(0011)^{m-1}$;
- $p_{2r} \in \{4m-2, 4m-1\}$, meaning that the last letter 1 of $(01)^r$ occurs at the last segment of 0011. As the $(2r-1)$-st letter of $(01)^r$ is 0, we have $p_{2r-1} \in \{4m'+1, 4m'+2\}$ for some $0 \le m' \le m-1$. By removing both p_{2r-1} and p_{2r}, we obtain P', which is an occurrence of $(01)^{r-1}$ in $(0011)^{m'}$. To go back from P' to P given m', we have two choices for both p_{2r} and p_{2r-1}.

We thus have the recurrence for $m \ge 1$ that

$$a_{m,r} = a_{m-1,r} + \sum_{m'=0}^{m-1} 4 a_{m',r-1}. \tag{1}$$

Subtracting Equation (1) for $a_{m,r}$ with that for $a_{m-1,r}$, we have

$$a_{m,r} - 2a_{m-1,r} + a_{m-2,r} - 4a_{m-1,r-1} = 0.$$

By the standard symbolic method, and with the initial conditions $a_{m,0} = 1$ and $a_{m,r} = 0$ for $r > m$, we obtain the claimed expression of $f(x, y)$. We can then compute $a_{m,r}$ by simply extracting the coefficient of y^r first, then that of x^m. ◄

▶ **Theorem 4.4.** *There is some constant c_3 such that, for all $n \in \mathbb{N}$, we have*

$$\min S_{\mathrm{sw}}^{(2)}(n) \le \frac{1}{2} \log_2(1 + \sqrt{2})n - \frac{1}{2} \log_2 n + c_3.$$

Proof. For the case $n = 4m$, we have

$$\min S_{\mathrm{sw}}^{(2)}(4m) \le S_{\mathrm{sw}}((0011)^m) = \max_{0 \le r \le m} \log_2 \mathrm{occ}((0011)^m, (01)^r)$$

$$= \frac{1}{\ln 2} \max_{0 \le r \le m} \ln\left(4^r \binom{m+r}{m-r}\right).$$

The first equality comes from Proposition 4.1, and the second from Proposition 4.3. We take $r = \alpha m$ for some fixed α with $0 < \alpha < 1$. Using Stirling's approximation, we have

$$\ln\left(4^r \binom{m+r}{m-r}\right) = s(\alpha)m - \frac{1}{2}\ln m + O(1),$$

where

$$s(\alpha) = \alpha \ln 4 + (1 + \alpha)\ln(1 + \alpha) - (1 - \alpha)\ln(1 - \alpha) - 2\alpha \ln(2\alpha).$$

The function $s(\alpha)$ is maximized at $\alpha = 2^{-1/2}$, with value $2\ln(1 + \sqrt{2})$. We thus have, for some constant c_3, and in terms of $n = 4m$,

$$\min S_{\mathrm{sw}}^{(2)}(n) \le \frac{\ln(1 + \sqrt{2})}{2\ln 2}n - \frac{1}{2\ln 2}\ln n + c_3 - \ln 4.$$

For the case $n = 4m + i$ with $1 \le i \le 3$, let u be a most frequent subword of $w = (0011)^m 010$. For an occurrence P of u in w, we take $P' = P \cap \{n-2, n-1, n\}$. Then, $j = |P'|$ can be $0, 1, 2$ or 3. In each case, we define $u^{(j)}$ to be u with the last j letters removed, and there are at most 2 possibilities for P'. We also notice that $P \setminus P'$ is an occurrence of $u^{(j)}$ in $(0011)^m$. We thus have

$$\mathrm{maxocc}(w) = \mathrm{occ}(w, u) = 2\,\mathrm{occ}((0011)^m, u^{(1)}) + \mathrm{occ}((0011)^m, u^{(2)}) + \mathrm{occ}((0011)^m, u^{(3)})$$

$$\le 4\,\mathrm{maxocc}((0011)^m).$$

We conclude by

$$\min S_{\mathrm{sw}}^{(2)}(4m+i) \le S_{\mathrm{sw}}^{(2)}(w) \le \ln 4 + S_{\mathrm{sw}}^{(2)}((0011)^m) = \frac{\ln(1+\sqrt{2})}{2\ln 2}n - \frac{1}{2\ln 2}\ln n + c_3.$$

For the first inequality, we take w' to be the first $(4m+i)$ letters of w, and it is clear that $\mathrm{maxocc}(w') \le \mathrm{maxocc}(w)$, as each occurrence of some subword v' of w' is also one for w. ◄

The asymptotic upper bound of $\min S_{\mathrm{sw}}^{(2)}(n)/n$ given by Theorem 4.4 is $\frac{1}{2}\log_2(1+\sqrt{2}) \approx 0.636\ldots$, which is much better than that in Corollary 3.5. Furthermore, by regarding $(0011)^m$ as a word in a bigger alphabet, we have the following corollary, which also gives a better upper bound than that in Corollary 3.5.

▶ **Corollary 4.5.** *For all $k \ge 2$ and $n \in \mathbb{N}$, with the constant c_3 from Theorem 4.4, we have*

$$\min S_{\mathrm{sw}}^{(k)}(n) \le \frac{1}{2}\log_2(1+\sqrt{2})n - \frac{1}{2}\log_2 n + c_3.$$

There are also other families of words with which we have some knowledge on its most frequent subwords, with results similar to Proposition 4.1. Two of the families we have studied are $(01)^m$ and $(000111)^m$.

▶ **Proposition 4.6.** *For $w = (01)^m$, there is a most frequent subword w' of the form $(01)^r$. Furthermore, $\mathrm{maxocc}((01)^m, (01)^r) = \binom{m+r}{m-r}$, which is maximized asymptotically for $r = \lfloor n/\sqrt{5} \rfloor$, with the asymptotic maximal value $\exp\left(n \ln \frac{3+\sqrt{5}}{2} - \frac{\ln n}{2} + O(1)\right)$.*

▶ **Proposition 4.7.** *For $w = (000111)^m$, there is a most frequent subword w' of the form $(0011)^r$. Furthermore, let $f_{000111}(x,y) = \sum_{m,r\ge 0} \mathrm{occ}((000111)^m, (0011)^r)x^m y^r$, we have*

$$f_{000111}(x,y) = \frac{(1-x)^3}{(1-x)^4 - 9x(1+2x)^2 y}.$$

For $m \to \infty$, the value of $\mathrm{occ}((000111)^m, (0011)^r)$ is asymptotically maximized for $r = \alpha m$, with α an explicit value around $0.6597177\ldots$. The asymptotic maximal value is $\exp\left(\gamma m - \frac{\ln m}{2} + O(1)\right)$, where γ is an explicit value around $2.7182400\ldots$.

While the proof of Proposition 4.6 does not need heavy machinery, the proof of Proposition 4.7 needs the saddle-point estimates of large powers [5, Theorem VIII.8], during which a polynomial equation of degree 5 appears. Fortunately, the needed solution of the said equation has a radical expression, albeit complicated and making the explicit expressions of α and γ in Proposition 4.7 too long to fit here.

While interesting, the upper bounds of L_2 given by Propositions 4.6 and 4.7, which are approximately $0.6942\ldots$ and $0.6536\ldots$ respectively, are worse than the one from Theorem 4.4. It is natural to try to look at other families of periodic words. This is encouraged by the following theorem.

▶ **Theorem 4.8.** *For any words w, v in an alphabet $\mathcal{A}$ of size k, the generating function $f_{w,v}(x,y) = \sum_{m,r\ge 0} \mathrm{occ}(w^m, v^r)x^m y^r$ is rational in x, y.*

Proof. We define $a_{w,v}^{s,t}(m)$ with $1 \le s,t \le m$ to be the number of occurrences $P = \{p_1,\ldots,p_{|v|}\}$ of v in w^m such that $p_1 = s$ and $p_{|v|} = (m-1)|w| + t$. In other words, $a_{w,v}^{s,t}(m)$ counts the occurrences of v in w^m such that the first (resp. last) letter of v occurs in the first (resp. last) copy of w at position s (resp. t). Let $g_{w,v}^{s,t}(x) = \sum_{m\ge 1} a_{w,v}^{s,t}(m)x^{m-1}$. Note the extra -1 in the exponent of x in $g_{w,v}^{s,t}(x)$. We first show that $g_{w,v}^{s,t}(x)$ is rational.

For an occurrence P of v in w^m, some consecutive letters may occur in the same copy of w. We say that such letters form a cluster, and we denote by σ the integer composition of the number of letters in each cluster from left to right. We denote by $\ell(\sigma)$ the length of σ, which is also the number of clusters. We denote the clusters by $v_\sigma^{(1)}, \ldots v_\sigma^{(\ell(\sigma))}$, and it is clear that they are obtained by cutting v into pieces whose lengths are the parts of σ. We then have

$$g_{w,v}^{s,t}(x) = a_{w,v}^{s,t}(1)$$
$$+ \sum_{m \geq 2} \sum_{\sigma \vDash |v|} x^{m-1} \binom{m-2}{\ell(\sigma)-2} \left(\sum_{t'=s}^{|w|} a_{w,v_\sigma^{(1)}}^{s,t'}(1) \right) \left(\sum_{s'=1}^{t} a_{w,v_\sigma^{(\ell(\sigma))}}^{s',t}(1) \right) \prod_{i=2}^{\ell(\sigma)-1} \mathrm{occ}(w, v_\sigma^{(i)}).$$

Here, $\sigma \vDash |v|$ means that we go over all integer compositions of $|v|$. The first term is for $m = 1$. For the second term, we simply count all possibilities of how clusters of v appear in w^m with $m \geq 2$ while fixing the first and the last cluster. We observe that each $a_{w,v_{\sigma(i)}}^{s',t'}(1)$ for any s', t', i is a constant, and the same holds for $\mathrm{occ}(w, v_\sigma^{(i)})$. By exchanging the two summations, and observing that $\sum_{m \geq 2} \binom{m-2}{d-2} x^{m-1} = x^{d-1}(1-x)^{-(d-1)}$, we see that $g_{w,v}^{s,t}(x)$ is rational in x with $(1-x)^{|v|-1}$ as denominator, as $\ell(\sigma) \leq |v|$ for $\sigma \vDash |v|$.

Now, for $1 \leq t \leq |w|$, we define $f_{w,v}^{(t)}(x,y) = \sum_{m \geq 1} \sum_{r \geq 1} b_{w,v}^{(t)}(m,r) x^{m-1} y^r$ with $b_{w,v}^{(t)}(m,r)$ counting the number of occurrences $P = \{p_1, \ldots p_{|v|r}\}$ of v^r in w^m such that $p_{|v|r} = (m-1)|w|+t$. Again, we note the extra -1 in the exponent of x. We see that $b_{w,v}^t(m,r)$ is defined similarly as $a_{w,v}^{s,t}(m)$, except that we consider subwords of the form v^r, and we do not fix the position of the first letter of v^r in w^m. We thus have $b_{w,v}^{(t)}(m,1) = \sum_{s=1}^{|w|} a_{w,v}^{s,t}(m)$. Now, let P be an occurrence of v^r in w^m counted by $b_{w,v}^{(t)}(m,r)$. By considering the copies of w spanned by the last copy of v, we have

$$f_{w,v}^{(t)}(x,y) = y \sum_{s=1}^{|w|} g_{w,v}^{s,t}(x) + \frac{xy}{1-x} \left(\sum_{t'=1}^{|w|} f_{w,v}^{(t')}(x,y) \right) \left(\sum_{s=1}^{|w|} g_{w,v}^{s,t}(x) \right)$$
$$+ y \sum_{t'=1}^{|w|-1} \left(f_{w,v}^{(t')}(x,y) \sum_{s=t'+1}^{|w|} g_{w,v}^{s,t}(x) \right).$$

Here, the first term is for $r = 1$, and the rest is for $r \geq 2$. There are two cases: either letters in the r-th and the $(r-1)$-st copies of v do not occur in the same copy of w in w^m, or they do. The first case is counted by the second term above, with the factor $(1-x)^{-1}$ for copies of w between the occurrences of the two last copies of v in w^m. The second case is accounted by the third term above, where we have the constraint that the last letter of the $(r-1)$-st copy of v occurs before the first letter of the r-th copy in the same copy of w.

Let $\mathbf{f} = {}^t(f_{w,v}^{(1)}, \ldots, f_{w,v}^{(|w|)})$. The equation above can be seen as $\mathbf{Af} = \mathbf{b}$ for some matrix $\mathbf{A} = (A_{i,j})_{1 \leq i,j \leq |w|}$ and some row vector $\mathbf{b}$, both with coefficients that are linear in y and rational in x, and with only powers of $(1-x)$ as denominators. We also observe that $A_{i,i}$ is of the form $1 + R(x)y$ with $R(x)$ rational in x, while $A_{i,j}$ for $i \neq j$ is of the form $R(x)y$. Hence, $\mathbf{A}$ is non-singular, and $f_{w,v}^{(i)}$ is rational in x, y for all $1 \leq i \leq |w|$. We conclude by observing that $f_{w,v}(x,y) = \frac{1}{1-x}(1 + x \sum_{t=1}^{|w|} f_{w,v}^{(t)}(x,y))$, with the 1 taking care of the case $m = 0$, then the factor $(1-x)^{-1}$ for the copies of w after the last cluster of v^r. ◀

Therefore, in principle, for any word w and v, we can first compute $f_{w,v}(x,y)$ effectively as in the proof of Theorem 4.8, then use analytic combinatorics in several variables [10, 12] to compute the asymptotically maximal value of $\mathrm{occ}(w^m, v^r)$ for fixed m. Although the

computation of $f_{w,v}(x,y)$ would be tedious, it is still feasible in principle. The only problem is that, for w in general, we do not have results like Proposition 4.1 for the structure of most frequent subwords of w^m, meaning that $\mathrm{maxocc}(w^m)$ is not necessarily achieved for subwords of the form v^r.

5 Open questions

Generally, the "minimum of maximums" structure in the definition of $\min S_{\mathrm{sw}}^{(k)}(n)$ makes estimates difficult. Hence, not a lot is known about $\min S_{\mathrm{sw}}^{(k)}(n)$. Intuitively, we have the following conjecture that is surprisingly difficult to tackle.

▶ **Conjecture 5.1.** *For fixed $k \geq 2$, there is a value N such that the function $\min S_{\mathrm{sw}}^{(k)}(n)/n$ is increasing for $n \geq N$.*

We already know experimentally that Conjecture 5.1 is not universally true for all n (see Table 1). This point needs to be addressed in possible proofs.

We are also interested by better bounds of L_2. Inspired by Theorem 4.8, we looked at some families experimentally, and there are some with potential to give a slightly better upper bound of L_2 according to numerical evidence. However, we don't know how to prove the observed structure on most frequent subwords of such periodic words in general, which leads us to the following conjecture.

▶ **Conjecture 5.2.** *For a given word w, there is a word v such that, for all m large enough, there is a most frequent subword of w^m that takes the form $u \cdot v^r \cdot u'$, with u and u' of lengths bounded by $|v|$.*

If Conjecture 5.2 holds, then using arguments similar to those in Theorem 4.4, we can reduce the computation of $\mathrm{maxocc}(w^m)$ to that of maximizing $\mathrm{occ}(w^m, v^r)$ while losing only a multiplicative constant. We can then apply Theorem 4.8 and tools in analytic combinatorics in several variables to obtain a better upper bound for L_k. Again, Conjecture 5.2 seems natural, intuitive and supported by experimental evidence, but we don't see how to settle it.

Another intuitive idea on most frequent subwords of a given word w is that their length should be smaller than $|w|/2$. The reasoning is that longer subwords have more letters, thus more possible occurrences, but this effect only works up till length $|w|/2$. However, even such an intuitive idea, supported by Proposition 3.4, seems difficult to prove.

▶ **Conjecture 5.3.** *For a given word w of length at least 2, there is no most frequent subword of w with length at least $|w|/2$.*

We now present some concrete experimental results, based on which we have other conjectures. We denote by $\overline{w}$ the word obtained from w by switching 0 and 1, and $\overleftarrow{w}$ the reverse of w. By symmetry between the two letters, we have the following simple observation.

▶ **Lemma 5.4.** *For any $w \in \{0,1\}^n$ with $n \geq 0$, we have $\mathrm{maxocc}(w) = \mathrm{maxocc}(\overline{w}) = \mathrm{maxocc}(\overleftarrow{w})$.*

We now give the words achieving minimal subword entropy of length up to 35 in Table 1, up to the symmetries in Lemma 5.4. These results are computed using a program written in C on one core on a local computation server, and it took around 11 days for $n = 35$. The source code can be found at `https://github.com/fwjmath/maxocc-subword`. We do not include the most frequent subwords we find, because there may be several of them for a word, taking up too much space in the table. There are several observations we can draw from Table 1, but few is without exception.

■ **Table 1** Binary words achieving minimal subword entropy of length from 1 to 35. In each equivalent class defined by the symmetries in Lemma 5.4, only one representative is given. Numerical values are rounded to three digits after the decimal point when needed.

n	Words w with lowest $S_{\mathrm{sw}}^{(2)}(w)$	$\mathrm{maxocc}(w)$	$S_{\mathrm{sw}}^{(2)}(w)$	$S_{\mathrm{sw}}^{(2)}(w)/n$	#runs
1	0	1	0	0	1
2	01	1	0	0	2
3	001	2	1	0.333	2
4	0110	2	1	0.25	3
5	01110	3	1.585	0.317	3
6	011001	5	2.322	0.387	4
7	0110001	6	2.585	0.369	4
8	01110001	9	3.170	0.396	4
9	011000110	16	4	0.444	5
10	0110001110	22	4.459	0.446	5
11	01110001110	33	5.044	0.459	5
12	011000111001	52	5.700	0.475	6
13	0111001001110	72	6.170	0.475	7
14	01100010111001	108	6.755	0.482	8
15	011000101110001	162	7.340	0.489	8
16	0111000101110001	252	7.977	0.499	8
17	01100011111000110	390	8.607	0.506	7
18	0111001001011110001	588	9.200	0.511	10
19	0110001011101000110 0110001110110001110	900	9.814	0.517	11
20	01110001011011000110	1320	10.366	0.518	11
21	011100011011010001110	2049	11.000	0.524	11
22	0110001110101000111001	2958	11.530	0.524	12
23	01110001011011010001110	4473	12.127	0.527	13
24	011000111010101000111001	6979	12.769	0.532	14
25	0111000101101101000111001	10602	13.372	0.535	14
26	01110001011011001000111001	15962	13.962	0.537	14
27	011100010101110101000111001	24150	14.560	0.539	16
28	0110001111010010010111000110 0111000101110101000101110001	36450	15.154	0.541	15 16
29	01100011101010001010111000110	53671	15.712	0.542	17
30	011000111001100010101111000110	83862	16.356	0.545	15
31	0110001110101000101011110001110	127998	16.966	0.547	17
32	01100011101010001010111010001110	189131	17.529	0.548	19
33	011000111101010001011011010001110	288900	18.140	0.550	19
34	0110001110101000101011101001001110	442386	18.755	0.552	21
35	01110001011011001000110111001001110	681966	19.379	0.554	19

- The words of length n achieving $\min S_{\text{sw}}^{(2)}(n)$ are palindromic, *i.e.*, $w = \overleftarrow{w}$, or anti-palindromic, *i.e.*, $\overline{w} = \overleftarrow{w}$, for many values of n, such as 1, 2, 4, 5, 6, 8, 9, 11, 12, 13, 14, 16, 17, 22, 23, 24, 29. Moreover, for $n = 19$ (resp. $n = 28$), one of the two words is palindromic (resp. anti-palindromic), the other not.
- The value of $\min S_{\text{sw}}^{(2)}(n)/n$ increases with n in general, but with the exceptions of $n = 3, 4$, $n = 6, 7$ and $n = 12, 13$ (although the rounded numbers are the same). We believe that the exceptions are due to the effect of small size, and should not reproduce for larger n.
- The number of runs for words of length n achieving $\min S_{\text{sw}}^{(2)}(n)$ is increasing with n, with the exception of $n = 17, 28, 30, 35$. Moreover, for $n = 28$ only one word among the two that has less runs than the word for $n = 27$.
- The maximal run length for words of length n achieving $\min S_{\text{sw}}^{(2)}(n)$ is at most 3, with the exception of $n = 17, 28, 30, 31, 33$. Moreover, for $n = 28$, one of the two words has maximal run length 3, and the other 4.
- There is only one word of length n up to symmetries in Lemma 5.4 that achieves $\min S_{\text{sw}}^{(2)}(n)$, with the exception of $n = 19, 28$, where there are two such words.

However, we should note that we only have very limited data, as we were only able to perform exhaustive search for small values of n. A naïve method requires looking at $\Theta(4^n)$ word-subword pairs. Although some optimizations are possible, such as using Lemma 5.4 to reduce the number of words to examine, the time taken remains exponential, against which we cannot push too far. An evidence is that, although asymptotically $\min S_{\text{sw}}^{(2)}(n)/n$ should be bounded from below by $\log_2(3/2) \approx 0.585$ by Corollary 3.5, all the values of $\min S_{\text{sw}}^{(2)}(n)/n$ in Table 1 are smaller than this asymptotic bound, meaning that the values of n tested here are not large enough. Nevertheless, we can still formulate reasonable conjectures based on these observations.

▶ **Conjecture 5.5.** *For $k \geq 2$, let w be a word of length $n \geq 1$ achieving the minimal subword entropy $\min S_{\text{sw}}^{(k)}(n)$. Then, except for a finite number of n, the longest run in w has length 3. Furthermore, the average run length converges when $n \to +\infty$.*

Given Proposition 3.9, we may be tempted to use experimental results to give better lower bound for L_2. However, all such bounds are worse than the one in Corollary 3.5, which is around 0.585 for $k = 2$. Judging from their gap, it seems impractical or even impossible to obtain a better bound in this way. In fact, with examples obtained from searches using various heuristics, it seems that $\text{maxocc}(w)$ for w of length n achieving the lowest subword entropy has an exponential growth in n with a growth constant close to but slightly larger than 1.5, which is the value given by the lower bound. We thus have the following conjecture.

▶ **Conjecture 5.6.** *We have $L_2 > \log_2(3/2)$.*

The value in Conjecture 5.6 comes from the lower bound in Corollary 3.5, which is in fact the expectation of the number of occurrences of a random subword of length $n/3$. Hence, Conjecture 5.6 implies that, for all large values of n, there are binary words of length n in which each subword of length $n/3$ occurs much more often than others. The question remains on how to find such subwords, which probably have relatively high self-correlations.

─── **References** ───────────────────────

1   A. Burstein, P. Hästö, and T. Mansour. Packing Patterns into Words. *Eletron. J. Combin.*, 9(2), 2003. `doi:10.37236/1692`.

2   A. Burstein and T. Mansour. Counting occurrences of some subword patterns. *Discrete Mathematics & Theoretical Computer Science*, Vol. 6 no. 1, January 2003. `doi:10.46298/dmtcs.320`.

3   Wenjie   Fang.   fwjmath/maxocc-subword.   Software,   swhId: `swh:1:dir:fef689a6896632f63f67b460e989fc106d5899e0` (visited on 2024-07-05).   URL: `https://github.com/fwjmath/maxocc-subword`.

4   M. Fekete. Über die Verteilung der Wurzeln bei gewissen algebraischen Gleichungen mit ganzzahligen Koeffizienten. *Math. Z.*, 17(1):228–249, 1923. `doi:10.1007/bf01504345`.

5   Ph. Flajolet and R. Sedgewick. *Analytic combinatorics*. Cambridge University Press, Cambridge, 2009. `doi:10.1017/CB09780511801655`.

6   Ph. Flajolet, W. Szpankowski, and B. Vallée. Hidden word statistics. *Journal of the ACM*, 53(1):147–183, 2006. `doi:10.1145/1120582.1120586`.

7   I. Gheorghiciuc and M. D. Ward. On Correlation Polynomials and Subword Complexity. *Discrete Mathematics & Theoretical Computer Science*, DMTCS Proceedings vol. AH, 2007 Conference on Analysis of Algorithms (AofA 2007), 2007. `doi:10.46298/dmtcs.3553`.

8   K. Iwanuma, R. Ishihara, Y. Takano, and H. Nabeshima. Extracting Frequent Subsequences from a Single Long Data Sequence: A Novel Anti-Monotonic Measure and a Simple On-Line Algorithm. In *Fifth IEEE International Conference on Data Mining (ICDM'05)*. IEEE, 2005. `doi:10.1109/icdm.2005.60`.

9   S. Kitaev. *Patterns in Permutations and Words*. Springer Berlin Heidelberg, 2011. `doi:10.1007/978-3-642-17333-2`.

10   S. Melczer. *Algorithmic and Symbolic Combinatorics: An Invitation to Analytic Combinatorics in Several Variables*. Springer International Publishing, 2021. `doi:10.1007/978-3-030-67080-1`.

11   K. Menon and A. Singh. Subsequence frequency in binary words. *Discrete Mathematics*, 347(5):113928, May 2024. `doi:10.1016/j.disc.2024.113928`.

12   M. Mishna. *Analytic combinatorics: a multidimensional approach*. Discrete Mathematics and its Applications (Boca Raton). CRC Press, 2020.

13   M. Morse and G. A. Hedlund. Symbolic dynamics. *Amer. J. Math.*, 60(4):815, October 1938. `doi:10.2307/2371264`.

14   V. Vatter. Permutation classes. In *Handbook of Enumerative Combinatorics*. CRC Press, 2015.

15   G. Yang. The complexity of mining maximal frequent itemsets and maximal frequent patterns. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD04. ACM, August 2004. `doi:10.1145/1014052.1014091`.

Sparsification of Phylogenetic Covariance Matrices of k-Regular Trees

Sean Svihla ✉
Department of Applied Mathematics, University of Colorado, Boulder, CO, USA

Manuel E. Lladser[1] ✉ 🄳
Department of Applied Mathematics, University of Colorado, Boulder, CO, USA

— **Abstract** —————————————————————————————

Consider a tree $T = (V, E)$ with root $\circ$ and an edge length function $\ell : E \to \mathbb{R}_+$. The phylogenetic covariance matrix of T is the matrix C with rows and columns indexed by L, the leaf set of T, with entries $C(i, j) := \sum_{e \in [i \wedge j, o]} \ell(e)$, for each $i, j \in L$. Recent work [Gorman & Lladser 2023] has shown that the phylogenetic covariance matrix of a large but random binary tree T is significantly sparsified, with overwhelmingly high probability, under a change-of-basis to the so-called Haar-like wavelets of T. Notably, this finding enables manipulating the spectrum of covariance matrices of large binary trees without the necessity to store them in computer memory but instead performing two post-order traversals of the tree [Gorman & Lladser 2023]. Building on the methods of the aforesaid paper, this manuscript further advances their sparsification result to encompass the broader class of k-regular trees, for any given $k \geq 2$. This extension is achieved by refining existing asymptotic formulas for the mean and variance of the internal path length of random k-regular trees, utilizing hypergeometric function properties and identities.

2012 ACM Subject Classification Mathematics of computing $\to$ Trees; Mathematics of computing $\to$ Generating functions; Mathematics of computing $\to$ Random graphs

Keywords and phrases cophenetic matrix, Haar-like wavelets, hierarchical data, hypergeometric functions, metagenomics, phylogenetic covariance matrix, sparsification, ultrametric matrix

Digital Object Identifier 10.4230/LIPIcs.AofA.2024.4

1 Introduction

Hierarchical datasets are described, or presumed to be, by a rooted tree that recursively organizes data into clusters so that its leaves are in a one-to-one correspondence with the data points. Such datasets are common in various fields such as microbial ecology [5], where they arise as models of inter-species covariance [10, 21]. In this context, the associated covariance matrices are often large and dense, making their manipulation computationally challenging. Nevertheless, these matrices bear redundancies induced by their hierarchical structure, which may be exploited to sparsify them and make such manipulations tenable, if not trivial.

Ultrametric matrices, which arise in probability theory and statistical physics among other fields, are often also dense. A symmetric matrix $C \in \mathbb{R}_+^{n \times n}$ is called *ultrametric* when $C(i, j) \geq \min\{C(i, k), C(k, j)\}$, for all $i, j, k \in \{1, \ldots, n\}$. If, in addition, $C(i, i) > C(i, j)$ for all $j \neq i$ when $n > 1$, or $C(1, 1) > 0$ when $n = 1$, C is called *strictly ultrametric*. These matrices are fully dense, that is, all their entries are nonzero, but have a myriad of mathematical properties [6, 18].

---

[1] Corresponding author

A matrix C is ultrametric if and only if there is a rooted binary tree $T = (V, E)$ and edge length function $\ell : E \to \mathbb{R}_+$ such that [14, 20]:

$$C = \sum_{e \in E} \ell(e)\, \delta_e\, \delta'_e,$$

where δ_e is the binary column vector, with entries indexed by the leaves of T, indicating the leaves that descend from e, and δ'_e is the transpose of δ_e. If L denotes the leaf set of T, the above identity is equivalent to having

$$C(i, j) = \sum_{e \in [i \wedge j, \circ]} \ell(e), \text{ for all } i, j \in L, \tag{1}$$

where $[i \wedge j, \circ]$ denotes the set of edges that connect $(i \wedge j)$, the least common ancestor of i and j, with the root of the tree, denoted as $\circ$. Ultrametric matrices have, therefore, a recursive structure, and their entries are redundant, suggesting that they may be amenable to some form of compression [14].

The formulation in (1) arises naturally as a model of "phylogenetic" covariance wherein the genetic drift of a particular trait follows a Brownian motion [15]. Under this model, each leaf represents a microbial species (or some notion thereof), and the trait variation among different species is a function of time since they diverged evolutionarily. Intuitively, since species sharing more of their evolutionary history should thrive or struggle in similar environments accordingly, a natural measure of trait covariance between two different species i and j is the length of their shared evolutionary history, namely the quantity in (1).

In the general setting of rooted trees – not necessarily binary – a matrix with entries such as in (1) is called the *phylogenetic covariance matrix* (or cophenetic matrix) of a tree. (The term "phylogenetic" is usually omitted from now on.) Tree covariance matrices arise naturally in the context of hierarchical datasets; in particular, the class of covariance matrices associated with datasets having a binary hierarchy is precisely the class of ultrametric matrices.

Recent work [14] has demonstrated that, in the case of large datasets with a binary hierarchy, or equivalently, weighted and rooted binary trees, the associated covariance matrices become asymptotically diagonal with overwhelmingly high probability after changing basis to the so-called Haar-like wavelets [11] of the tree. (By "asymptotically diagonal with overwhelmingly high probability," we mean that the fraction of non-zero off-diagonal entries of the covariance matrix, with respect to the wavelet basis, converges in probability to 0 as the tree size approaches infinity.) The sparsification of such covariance matrices facilitates manipulations that may be infeasible otherwise. For instance, the spectrum of ultrametric matrices can be derived, whether exactly or approximately depending on the matrix size, from just two post-order traversals of the associated tree without having to store the actual matrix in computer memory [14]. In addition, the subclass of ultrametric matrices diagonalized by Haar-like wavelets has been characterized, and their spectra shown to be in bijection with non-negative decreasing functions on the interior nodes of a binary tree [14].

Nevertheless, many hierarchical datasets, for example, in the context of phylogenomic studies [24], are non-binary and suffer the same downfall of having unmanageably large and dense covariance matrices. It is a natural question, then, whether the same technique may be used to sparsify covariance matrices belonging to a broader class of hierarchy. This manuscript extends some of the ideas in [14] to the broader context of k-regular trees, i.e., rooted trees for which each interior node has exactly k children. Specifically, Theorem 2, Theorem 4, Corollary 5, and Corollary 15 in this manuscript are generalizations of [14, Theorem 2.3,

Theorem 3.4, Corollary 3.5, and Corollary 3.8], respectively, such that they are applicable to any $k \geq 2$. In addition, Theorem 10 supplies more precise asymptotic formulas for the mean and variance of the internal path length of random k-regular trees – beyond what is currently available in the literature.

This manuscript is based on the recent MS thesis [22].

Notation and Terminology

Depending on the context, we regard functions with finite domains as finite-dimensional column vectors, and vice versa. Throughout, $[\![\cdot]\!]$ is used to denote indicator functions.

We use standard terminology for trees unless otherwise stated. In particular, $T = (V, E)$ represents a rooted tree with vertex set V and edge set E. The size of T is the quantity $|T| := |V|$. The sets of leaves and interior nodes of T are denoted as L and I, respectively, and the internal path length of T is the quantity

$$\mathrm{IPL}(T) = \sum_{v \in I} \mathrm{depth}(v).$$

For $u, v \in V$, we denote by $[u, v]$ the set of edges in the shortest path from u to v and by $(u \wedge v)$ the least common ancestor of u and v.

$\mathring{T}$ denotes the *interior* of T, obtained by trimming the leaves of T.

In the subsequent discussion, T denotes a planted, ordered, and unlabelled k-regular tree endowed with an edge length function $\ell : E \to \mathbb{R}_+$. In particular, T is an ordered unlabelled k-regular tree whose root has been appended to a phantom one (denoted as $\circ$), which acts as the parent of the original root. (The term "out-rooted" instead of "planted" was introduced in [14].) See Figure 1(a). By definition, $\circ$ is an interior node of T (i.e., $\circ \in I$) and it has degree one. Furthermore, since each $v \in I \setminus \{\circ\}$ has k children, we denote these generically as $v_1, \ldots, v_k$ from left-to-right.

For a given $v \in V$, $T(v)$ denotes the sub-tree of T rooted at v and containing all of its descendants. We let $L(v)$ denote the sets of leaves of $T(v)$. In addition, we orient edges away from the root, that is, if $e = (u, v) \in E$, then u is understood to be the parent of v (and v is understood to be a child of u). We define $T(e) := T(v)$ and $L(e) := L(v)$.

The *trace length* of T is the function $\ell^* : E \to \mathbb{R}_+$ defined as [14]:

$$\ell^*(e) := |L(e)| \, \ell(e).$$

We also define for $u, v \in V$:

$$\ell(u, v) := \sum_{e \in [u,v]} \ell(e), \text{ and } \ell^*(u, v) := \sum_{e \in [u,v]} \ell^*(e),$$

and, given non-empty $J \subset V$, we denote by $\ell(J, v)$ the column vector with entries $\ell(j, v)$, for all $j \in J$, having dimension $|J|$. We give an analogous definition to $\ell^*(J, v)$.

For the remainder of this manuscript, the term "k-regular tree" encompasses planted, unlabelled, ordered, k-regular trees.

2 Haar-like Wavelets on k-Regular Trees

In this section, we specialize the Haar-like wavelet basis given in [11] to our setting of k-regular trees and present a useful interaction between the basis and the phylogenetic covariance matrix associated with any such tree (Theorem 2).

(a) 3-Regular Tree.

(b) Associated Haar-Like Wavelets.
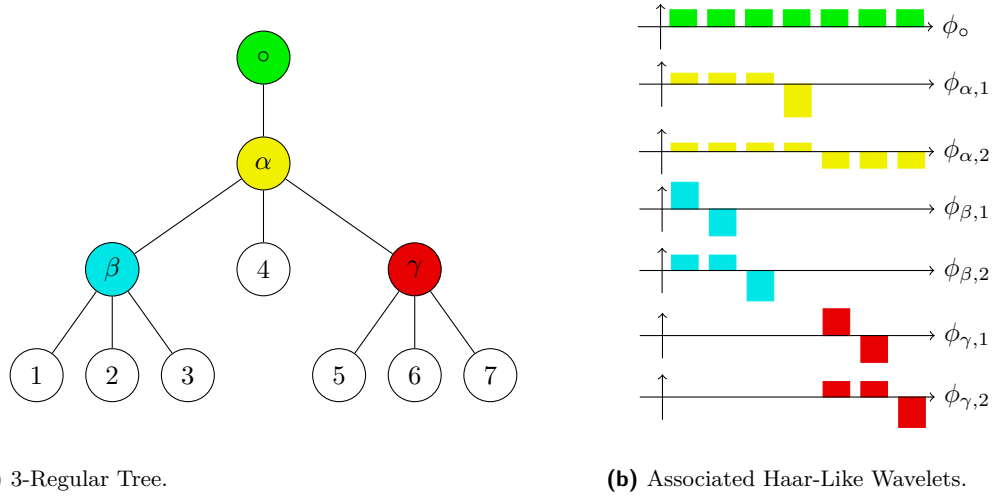
■ **Figure 1** Haar-like wavelets (right panel) associated with a 3-regular tree (left panel). As per our convention, the tree leaves are labeled $1, 2, \ldots$ following a depth-first search (DFS), while the non-root internal nodes are leaf-labeled. Thus $\alpha := \{1, 2, 3, 4, 5, 6, 7\}$, $\beta := \{1, 2, 3\}$, and $\gamma := \{5, 6, 7\}$. Except for the planted root (i.e., green node), each internal node has two wavelets associated with it. By construction, the wavelets are orthonormal piece-wise constant functions over the leave set and, except for $\phi_\circ$, each has mean zero.

While our manuscript focuses on ordered unlabeled trees, systematically naming their leaves is beneficial for constructing the wavelets. Henceforth, **we label the leaves of T as $1, \ldots, |L|$, following the order encountered in a depth-first search (DFS) of T.** Moreover, aside from the planted root, **we label interior nodes by the set of leaves descending from them.**

Wavelets are usually functions defined on a Euclidean space and derive their name from their commonly wave-like shape. They are prevalent in time series and image analysis to localize information across various scales.

In our context, the Haar-like wavelets associated with T are functions from L to $\mathbb{R}$ that are in a one-to-one correspondence with the elements in the set $\{\circ\} \cup (I \setminus \{\circ\} \times \{1, \ldots, k-1\})$; in particular, there are $1 + (k-1)(|I| - 1)$ wavelets associated with a k-regular tree. The precise definition follows.

Given a $v \in I$ and integer $1 \leq n < k$, define

$$L_{v,n} := \bigcup_{j=1}^{n} L(v_j), \text{ and } L_{v,n}^+ := L_{v,n} \cup L(v_{n+1}).$$

In particular, $L_{v,n}^+ = L_{v,n+1}$. Nevertheless, we introduced the notation of $L_{v,n}^+$ because this is precisely the support of the wavelet with index (v, n) in the next definition.

▶ **Definition 1** (Haar-like wavelets of a k-regular tree). *The (mother) wavelet associated with the root of T is the constant function $\phi_\circ : L \to \mathbb{R}$ defined as*

$$\phi_\circ(i) := \frac{1}{\sqrt{|L|}}, \text{ for } i \in L.$$

Instead, for each $v \in I \setminus \{\circ\}$ and integer $1 \leq n < k$, the wavelet with index (v, n) is the function $\phi_{v,n} : L \to \mathbb{R}$ defined as

$$\phi_{v,n}(i) := \sqrt{\frac{|L(v_{n+1})|}{|L_{v,n}| \cdot |L_{v,n+1}|}} \, [\![i \in L_{v,n}]\!] - \sqrt{\frac{|L_{v,n}|}{|L(v_{n+1})| \cdot |L_{v,n+1}|}} \, [\![i \in L(v_{n+1})]\!], \text{ for } i \in L.$$

It turns out that the Haar-like wavelets are mutually orthogonal [11]. Moreover, from well-known facts about graphs and trees,

$$|I| + |L| - 1 = \frac{1}{2} \sum_{v \in V} \deg(v) = \frac{1}{2}\Big[|I| + |L| + k\big(|I| - 1\big)\Big],$$

hence $|L| = 1 + (k-1)(|I| - 1)$. That is, there are as many Haar-like wavelets as leaves on the tree, and the wavelets form therefore an orthonormal basis of the linear space of functions from L to $\mathbb{R}$.

We emphasize that the wavelet with index (v, n) assigns identical positive values to the leaves in $L_{v,n}$, and identical negative values to those in $L(v_{n+1})$. These values are chosen so that $\phi_{v,n}$ has mean zero and a unit ℓ^2-norm; namely,

$$\sum_{i \in L} \phi_{v,n}(i) = 0, \text{ and } \sum_{i \in L} \phi^2_{v,n}(i) = 1.$$

To fix ideas, Figure 1(b) provides an illustration of the Haar-like wavelets of a 3-regular tree.

Notice that $\phi_{v,n}$ is supported on $L^+_{v,n}$. As such, wavelets associated with nodes nearer the root capture coarser information about the leaves of the tree. As noted in [13, Theorem 1.1] for $k = 2$, given a function $\varphi : L \to \mathbb{R}$ and wavelet $\phi_{v,n}$, we have

$$\langle \varphi, \phi_{v,n} \rangle = c_{v,n} \cdot \left\{ \frac{1}{|L_{v,n}|} \sum_{i \in L_{v,n}} \varphi(i) - \frac{1}{|L(v_{n+1})|} \sum_{i \in L(v_{n+1})} \varphi(i) \right\},$$

where

$$c_{v,n} := \sqrt{\frac{|L_{v,n}|\,|L(v_{n+1})|}{|L^+_{v,n}|}}.$$

Consequently, projecting a function onto $\phi_{v,n}$ is the same, up to a constant factor, as computing the difference between the average values of φ over $L_{v,n}$ and $L(v_{n+1})$. Hence, the projections of a real-valued function defined over L onto the Haar-like wavelets can be computed efficiently – which is relevant for applications involving large trees.

The following result highlights a remarkably simple action of the covariance matrix of a k-regular tree over its Haar-like wavelets. This property was first noticed in [14] for the case of 2-regular trees.

▶ **Theorem 2.** *If ψ is a Haar-like wavelet associated with $v \in I$, then $C\,\psi = \mathrm{diag}(\ell^*(L, v))\,\psi$.*

Proof. We first show the result for $\psi = \phi_\circ$. For this, note that for all $j \in L$

$$(C\psi)(j) = \frac{1}{\sqrt{L}} \sum_{i \in L} \ell(i \wedge j, \circ) = \psi(j) \sum_{i \in L} \sum_{e \in [i \wedge j, \circ]} \ell(e),$$

yet we have the logical equivalence:

$$\forall i, j \in L \,\forall v \in I, e \in [i \wedge j, v] \iff i \in L(e) \text{ and } e \in [j, v]. \tag{2}$$

Hence, $\ell(e)$ occurs $|L(e)|$ times in the previous double-sum, and

$$(C\psi)(j) = \psi(j) \sum_{e \in [j,v]} |L(e)|\,\ell(e) = \psi(j)\,\ell^*(j, v),$$

which is precisely the j-th entry of the vector $\mathrm{diag}(\ell^*(L, v))\,\psi$.

Next, we consider $\psi = \phi_{v,n}$, for a $v \in I \setminus \{\circ\}$ and $1 \le n < k$. Then, for each $j \in L$, we have the following cases:

(a) Assume $j \notin L_{v,n}^+$. Then, for all $i \in L_{v,n}^+$, $(i \wedge j) = (v \wedge j)$, so

$$(C\psi)(j) = \sum_{i \in L_{v,n}^+} \ell(i \wedge j, \circ)\, \phi_{v,n}(i) = \ell(v \wedge j, \circ) \sum_{i \in L_{v,n}^+} \phi_{v,n}(i) = 0,$$

because $\phi_{v,n}$ has mean zero. On the other hand, the entry associated with j in the vector $\operatorname{diag}(\ell^*(L, v))\, \psi$ also vanishes because ψ is supported on $L_{v,n}^+$.

(b) Instead, if $j \in L_{v,n}^+$ then

$$
\begin{aligned}
(C\psi)(j) &= \sum_{i \in L_{v,n}^+} \ell(i \wedge j, \circ)\, \phi_{v,n}(i) \\
&= \sum_{i \in L_{v,n}^+} \ell(i \wedge j, v)\, \phi_{v,n}(i) + \ell(v, \circ) \sum_{i \in L_{v,n}^+} \phi_{v,n}(i) \\
&= \sum_{i \in L_{v,n}^+} \phi_{v,n}(i) \sum_{e \in [i \wedge j, v]} \ell(e),
\end{aligned}
$$

where for the middle identity we have reused that $\phi_{v,n}$ has mean zero. But notice that $(i \wedge j) = v$ when $\ell(i \wedge j, v) = 0$. Otherwise, if $(i \wedge j) \neq v$ then $\phi_{v,n}(i) = \phi_{v,n}(j)$. Hence, reusing the logic equivalence in (2), we find that

$$(C\psi)(j) = \phi_{v,n}(j) \sum_{i \in L_{v,n}^+ : (i \wedge j) \neq v} \sum_{e \in [i \wedge j, v]} \ell(e) = \psi(j)\, \ell^*(j, v).$$

which corresponds to the entry associated with j in the vector $\operatorname{diag}(\ell^*(L, v))\, \psi$.
The theorem now follows from parts (a)-(b). ◀

3   Sparsification of k-Regular Covariance Matrices

We now leverage the insights from the previous section to obtain a lower bound on the proportion of entries in the covariance matrix of a k-regular tree that vanish when switching to the Haar-type basis. To establish our main result in this section, we require the following definition, which is analogous to the one used in [14] for 2-regular trees.

▶ **Definition 3** (Haar-like matrix of a k-regular tree). *The Haar-like matrix of T is the square matrix Φ whose columns are the wavelets associated with the tree; that is, its rows are indexed by L and its columns by the wavelets.*

Let ψ_u, ψ_v be wavelets associated with interior nodes $u, v \in I$, respectively. We have as a direct consequence of Theorem 2 that

$$(\Phi' C \Phi)(\psi_u, \psi_v) = \psi_u \operatorname{diag}\big(\ell^*(L, v)\big)\, \psi_v. \tag{3}$$

But note that the support of $\operatorname{diag}\big(\ell^*(L, v)\big)\, \psi_v$ is contained in the support of ψ_v; therefore, the entry associated with row ψ_u and column ψ_v of $\Phi' C \Phi$ vanish when ψ_u and ψ_v have disjoint supports. Since $\Phi' C \Phi$ corresponds to C after changing basis to the Haar-like wavelets, (3) shows that the wavelets may be used to sparsify the covariance matrix of k-regular trees. Of course, there is no reason why such interactions should occur often enough to meaningfully sparsify the matrix. As detailed next, however, a minimum level of sparsification is guaranteed by the tree's size and internal path length.

We emphasize that the following result is a conservative bound on the number of vanished entries under the Haar-like basis. In practice, it is not uncommon to see better sparsification; however, we have found in all cases that the overwhelming majority of sparsification arises from the interactions considered in the following theorem. We discuss this topic further in Section 4.

▶ **Theorem 4.** *Let Φ and C denote the Haar-like matrix and (phylogenetic) covariance matrix of a k-regular tree T, respectively. If ζ is the fraction of vanishing entries of $\Phi' C \Phi$ then*

$$(1 - \zeta) \le (k - 1)^2 \frac{1}{|I|} + 2(k - 1)^2 \frac{\mathrm{IPL}(T)}{|I|^2}.$$

Proof. Note that C has dimensions $|L| \times |L|$ because T has as many Haar-like wavelets as leaves. Observe if $u, v \in I$ and ψ_u, ψ_v are any of their corresponding wavelets, identity (3) implies

$$\big(\Phi' C \Phi\big)(\psi_u, \psi_v) = \sum_{i \in L(u) \cap L(v)} \psi_u(i) \, \ell^*(i, v) \, \psi_v(i).$$

As $\big(\Phi' C \Phi\big)(\psi_u, \psi_v) = 0$ when $L(u) \cap L(v) = \emptyset$, then ζ obeys:

$$(1 - \zeta) \le \frac{(k - 1)^2}{|L|^2} \Big| \Big\{ (u, v) \in I \times I \; ; \; L(u) \cap L(v) \ne \emptyset \Big\} \Big|.$$

Notice, however, that $L(u) \cap L(v) \ne \emptyset$ only if u descends from v (or vice versa) [14]. Hence, accounting for pairs of the form (u, u) and, when $u \ne v$, (u, v) and (v, u), we obtain that

$$(1 - \zeta) \le \frac{(k - 1)^2}{|L|^2} \left(|I| + 2 \sum_{u \in I} \big(|\mathring{T}(u)| - 1 \big) \right) \le \frac{(k - 1)^2}{|L|^2} \left(2 \sum_{u \in I} |\mathring{T}(u)| - |I| \right).$$

Now, notice $\sum_{u \in I} |\mathring{T}(u)| = \sum_{u \in I} \sum_{v \in I} [\![v \in T(u)]\!]$; that is, each node is counted once for each of its ancestors, or

$$\sum_{u \in I} |\mathring{T}(u)| = \sum_{v \in I} \big(1 + \mathrm{depth}(v) \big) = |I| + \mathrm{IPL}(T).$$

Hence, we further obtain that

$$(1 - \zeta) \le \frac{(k - 1)^2}{|L|^2} \Big(|I| + 2\,\mathrm{IPL}(T) \Big).$$

As mentioned in Section 2, $|L| = 1 + (k - 1)(|I| - 1)$. Hence $|L| \ge |I|$; otherwise, $|I| < 1$, which is not possible because $k \ge 2$ and $\circ \in I$. The theorem is now a direct consequence of the above inequality.   ◀

▶ **Corollary 5.** *If $\mathrm{IPL}(T) \ll |I|^2$ as $|T| \to \infty$, then $\zeta = 1 - o(1)$.*

▶ **Example 6** (Perfect k-Regular Trees.)**.** These are k-regular trees in which every leaf has the same depth. Let T be one such tree with height $(h + 1)$. At each depth $j \ge 1$, there are k^{j-1} nodes. So

$$|I| = 1 + \sum_{j=1}^h k^{j-1} = \frac{k + k^h - 2}{k - 1}, \text{ and } \mathrm{IPL}(T) = \sum_{j=1}^h j k^{j-1} = \frac{h k^{h+1} - (h + 1) k^h + 1}{(k - 1)^2},$$

and we find that

$$\frac{\mathrm{IPL}(T)}{|I|^2} = \frac{h k^{h+1} - (h + 1) k^h + 1}{(k + k^h - 2)^2} \sim \frac{h}{k^{h-1}}.$$

Hence $\mathrm{IPL}(T) = o\big(|I|^2\big)$, as $h \to \infty$, and Corollary 5 implies that the Haar-like basis asymptotically diagonalizes the covariance matrix of perfect k-regular trees as their height tends to infinity.

▶ **Example 7** (k-Regular Caterpillar Trees). These are k-regular trees in which the parent of every leaf is a node in a central path graph. Let T be a k-regular caterpillar tree of height $(h+1)$ so that its central path has length h. Since there is only one node at depth 1, and at each depth $2 \le j \le h$, there are $(k-1)$ leaves and one interior node, we have

$$|I| = 2 + h, \text{ and } \text{IPL}(T) = \sum_{j=1}^{h} j = \frac{h(h+1)}{2}.$$

So

$$\frac{\text{IPL}(T)}{|I|^2} = \frac{h(h+1)}{2(h+2)^2} \sim \frac{1}{2},$$

as $h \to \infty$, and the lower-bound for ζ in Theorem 4 is trivial (in fact, strictly negative), and we cannot guarantee that the covariance matrix associated with T is sparsified to a significant degree as $|T| \to \infty$.

The last example shows that covariance matrices of some k-regular trees do not meet the criterion for significant sparsification given by Theorem 4; moreover, the tree in Example 6 is exceptionally constrained. So, the question remains whether trees meeting the sparsification criterion are at all common. To this end, we note that the guarantee provided by Theorem 4 is solely based on the tree's topology; i.e., the edge lengths are irrelevant. Therefore, we may investigate our remaining question by considering random k-regular trees irrespective of their edge lengths.

3.1 Interlude on Hypergeometric Functions

For a concise introduction to hypergeometric functions, see [1, 23].

A hypergeometric function is one whose power series is hypergeometric; that is, its ratio of consecutive coefficients indexed by n is a rational function of n. In particular, a power series $\sum_{n=0}^{\infty} f_n z^n$ is hypergeometric when there are constants $a_1, \ldots, a_p, b_1, \ldots, b_q$, for some integers $(q+1) \ge p \ge 1$, such that

$$\frac{f_{n+1}}{f_n} = \frac{1}{n+1} \cdot \frac{(n+a_1) \cdots (n+a_p)}{(n+b_1) \cdots (n+b_q)}.$$

(The constraints on p and q assume that the series coefficients are not eventually zero nor undefined due to division by zero.) Accordingly, the coefficients f_n may be written in terms of the Pochhammer symbol (defined such that $(c)_0 := 1$ and $(c)_n := \prod_{i=0}^{n-1}(c+i)$ for each integer $n \ge 1$) as

$$f_n = \frac{1}{n!} \cdot \frac{(a_1)_n \cdots (a_p)_n}{(b_1)_n \cdots (b_q)_n}, \text{ for all } n \ge 0.$$

The above hypergeometric function is denoted $_pF_q(a_1, a_2, \ldots, a_p; b_1, b_2, \ldots, b_q; z)$. A hypergeometric function of this form is said to be *s-balanced* if

$$\sum_{j=1}^{q} b_j - \sum_{j=1}^{p} a_j = s,$$

and if $s > 0$, the power series converges at $z = 1$ [8].

The following result will be crucial to identifying dominant singularities of the generating function enumerating internal path length of k-regular trees, as well as addressing their uniqueness in the closure of the disk of convergence. In stating this result, we use the following notation for the remainder of Section 3:

$$p(t) := t(1-t)^{k-1}, \text{ for all } t \in \mathbb{C}.$$

and

$$z_k := p\left(\frac{1}{k}\right) = \frac{(k-1)^{k-1}}{k^k}.$$

▶ **Proposition 8** (Reformulation of [23, Identity (25)]). *If $k \geq 2$ then the hypergeometric function $_{k-1}F_{k-2}(\frac{1}{k}, \ldots, \frac{k-1}{k}; \frac{k-2}{k-1}, \frac{k}{k-1}; \frac{z}{z_k})$ is analytic in the disc $|z| < z_k$ and continuous in $|z| \leq z_k$. Furthermore,*

$$_{k-1}F_{k-2}\left[\begin{array}{cccc} \frac{1}{k} & \cdots & \frac{k-1}{k} & \\ \frac{2}{k-1} & \cdots & \frac{k-2}{k-1} & \frac{k}{k-1} \end{array}; \frac{p(t)}{p\left(\frac{1}{k}\right)}\right] = \frac{1}{1-t}, \text{ for all } 0 \leq t \leq \frac{1}{k}.$$

Proof. We attribute some of the ideas in this proof to I. Gessel [12].

Consider the functional equation

$$F(z) = 1 + z\{F(z)\}^k, \tag{4}$$

with F analytic in an open neighborhood of $z = 0$. The Lagrange Inversion Theorem implies that (4) has a unique analytic solution in some open neighborhood of $z = 0$, with coefficients given by:

$$[z^n] F(z) = \frac{1}{(k-1)n+1}\binom{kn}{n}. \tag{5}$$

Therefore

$$\lim_{n\to\infty} \frac{[z^{n+1}] F(z)}{[z^n] F(z)} = \frac{k^k|z|}{(k-1)^{k-1}},$$

implying that $F(z)$ has radius of convergence z_k and, due to the Vivanti-Pringsheim Theorem [16, Theorem 5.7.1], z_k is a singular point of F.

On the other hand, the ratio of consecutive terms in $F(z)$ is

$$\frac{z^{n+1}[z^{n+1}] F(z)}{z^n [z^n] F(z)} = \frac{\left(n+\frac{1}{k}\right)\cdots\left(n+\frac{k-1}{k}\right)}{\left(n+\frac{2}{k-1}\right)\cdots\left(n+\frac{k-2}{k-1}\right)\cdot\left(n+\frac{k}{k-1}\right)} \cdot \frac{z/z_k}{n+1},$$

which reveals

$$F(z) = {_{k-1}F_{k-2}}\left[\begin{array}{cccc} \frac{1}{k} & \cdots & \frac{k-1}{k} & \\ \frac{2}{k-1} & \cdots & \frac{k-2}{k-1} & \frac{k}{k-1} \end{array}; \frac{z}{z_k}\right], \tag{6}$$

is the only analytic solution of (4) in the disk $|z| < z_k$. Moreover, since the balance of this hypergeometric function is

$$s = \left(\sum_{j=2}^{k-2} \frac{j}{k-1}\right) + \frac{k}{k-1} - \sum_{j=1}^{k-1} \frac{j}{k} = \frac{1}{2} > 0,$$

it follows from [8] that the series of $F(z)$ converges at $z = z_k$. Since F has non-negative coefficients, the series is absolutely convergent for all $|z| \leq z_k$, implying that F is analytic for $|z| < z_k$ and continuous for $|z| \leq z_k$.

To complete the proof, notice that $F(z) > 0$ for all $0 \leq z \leq z_k$. From the functional equation (4) we have that

$$z = \left(1 - \frac{1}{F(z)}\right)\left(\frac{1}{F(z)}\right)^{k-1} = p\left(\frac{F(z)-1}{F(z)}\right), \text{ for all } 0 \leq z \leq z_k,$$

but $p'(t) \geq 0$ for all $0 \leq t \leq k^{-1}$, with equality only at $t = k^{-1}$, and $p(k^{-1}) = z_k$. Therefore,

$$p^{-1}(t) = \frac{F(t)-1}{F(t)}, \text{ i.e., } F(t) = \frac{1}{1-p^{-1}(t)}, \text{ for all } t \in \left[0, \frac{1}{k}\right],$$

or, equivalently,

$$F\big(p(t)\big) = \frac{1}{1-t}, \text{ for all } t \in \left[0, \frac{1}{k}\right],$$

which finalizes the proof. ◄

▶ **Remark 9.** The sequence in (5) belongs to the class of Fuss-Catalan numbers. They are cataloged in the OEIS as A000108 ($k = 2$), A001764 ($k = 3$), A002293 ($k = 4$), and so forth.

3.2 Expectation and Variance of Internal Path Length

Corollary 5 establishes that a sufficient condition for the Haar-like basis of a k-regular tree to diagonalize its (phylogenetic) covariance matrix asymptotically, as the tree grows, is that its internal path length becomes negligible compared to the square of the number of its interior points. To assess the prevalence and extent of trees meeting this criterion, we need asymptotic estimates for the mean and variance of large, uniformly at random, k-regular trees, which is precisely what our next result addresses. We emphasize that related results exist in the literature – see [9, Proposition VII.3], [19, Theorem 3.1], and [7, Theorem 2.19]. Our new contribution is the derivation of leading asymptotic estimates – with explicit multiplicative constants – in terms of k.

▶ **Theorem 10.** *For a uniformly at random k-regular tree $\mathbb{T}$ with $|I|$ internal nodes, the expectation and variance of its internal path length satisfy the asymptotic estimates*

$$\mathbb{E}\big[\mathrm{IPL}(\mathbb{T})\big] = \sqrt{\frac{\pi k |I|^3}{2(k-1)}}\left(1 + O(|I|^{-1/2})\right), \tag{7}$$

and

$$\mathbb{V}\big[\mathrm{IPL}(\mathbb{T})\big] = \frac{k}{2(k-1)}\left(\frac{10}{3} - \pi\right)|I|^3\left(1 + O(|I|^{-1/2})\right). \tag{8}$$

Proof (Sketch). Given that the asymptotic formulas in the Theorem agree with those in [14] for $k = 2$, we assume henceforth that $k \geq 3$.

Let $Q(z, u)$ be the generating function for the class of k-regular trees, where the variable z marks the number of internal nodes and u the internal path length of a tree. We find using the methods of [9, Section III] that $Q(z, u)$ obeys the functional equation

$$Q(z, u) = G\big(z, Q(zu, u)\big), \text{ where } G(z, w) := 1 + z\,w^k. \tag{9}$$

For brevity, define $Q(z) := Q(z,1)$, $Q_u(z) := \frac{\partial Q}{\partial u}(z,1)$, and $Q_{uu}(z) := \frac{\partial^2 Q}{\partial u^2}(z,1)$. In particular:

$$\mathbb{E}\big[\text{IPL}(\mathbb{T})\big] = \frac{[z^n]Q_u(z)}{[z^n]Q(z)}; \tag{10}$$

$$\mathbb{V}\big[\text{IPL}(\mathbb{T})\big] = \frac{[z^n]Q_{uu}(z)}{[z^n]Q(z)} + \frac{[z^n]Q_u(z)}{[z^n]Q(z)} - \left(\frac{[z^n]Q_u(z)}{[z^n]Q(z)}\right)^2. \tag{11}$$

The asymptotic formulas in (7)-(8) follow from a detailed asymptotic analysis of the numerators and denominators above. To this effect, we first note from (9) that

$$Q(z) = 1 + z\{Q(z)\}^k. \tag{12}$$

In particular, the proof of Proposition 8 implies that $Q(z)$ is hypergeometric. Namely:

$$Q(z) = {}_{k-1}F_{k-2}\left[\begin{array}{ccc} \frac{1}{k} & \cdots & \frac{k-1}{k} \\ \frac{2}{k-1} & \cdots & \frac{k-2}{k-1} \quad \frac{k}{k-1} \end{array}; \frac{z}{z_k}\right], \text{ for all } |z| \le z_k.$$

Moreover, since $z_k = p(k^{-1})$, the following result is immediate from Proposition 8.

▶ **Corollary 11.** *If $k \ge 3$ then $Q(z_k) = \frac{k}{k-1}$; in particular, $k\, z_k\, \{Q(z_k)\}^{k-1} = 1$.*

We next show that $Q(z)$ fits the "smooth implicit function schema" [9], and determine the asymptotic order of the denominator in (10).

▶ **Lemma 12.** *z_k is the only singularity of $Q(z)$ on the disk $|z| \le z_k$ and, locally around z_k, $Q(z)$ admits the representation*

$$Q(z) = 1 + g(z) - h(z) \cdot \sqrt{1 - \frac{z}{z_k}},$$

for functions $g(z)$ and $h(z)$ analytic near z_k. Furthermore

$$[z^n]\, Q(z) = \sqrt{\frac{k}{2\pi n^3(k-1)^3}} \cdot z_k^{-n}\big(1 + O(n^{-1})\big). \tag{13}$$

As seen in Figure 2, the asymptotic formula in (13) is highly accurate even for small values of n when $k = 3, 4$.

Proof. Define $P(z) := Q(z) - 1$. $P(0) = 0$ and $P(z) = F\big(z, P(z)\big)$, where $F(z,p) := z\{1+p\}^k$. Note that $F(0,p) = 0$ and F has only non-negative Taylor coefficients around $(0,0)$. Moreover, due to Corollary 11, $P(z_k) = \frac{1}{k-1}$, hence

$$\begin{array}{llll} F\big(z_k, P(z_k)\big) &= \frac{1}{k-1}; & F_z\big(z_k, P(z_k)\big) &= \frac{k^k}{(k-1)^k} \neq 0; \\ F_p\big(z_k, P(z_k)\big) &= 1; & F_{pp}\big(z_k, P(z_k)\big) &= \frac{(k-1)^2}{k} \neq 0. \end{array}$$

In particular, $F\big(z_k, P(z_k)\big) = P(z_k)$ and, since $[z^n]\, P(z) > 0$ for all $n \ge 1$, [7, Theorem 2.19] implies that that z_k is the only singularity of $P(z)$ and hence of $Q(z)$ in the disk $|z| \le z_k$. Furthermore, it admits the singular expansion

$$P(z) = g(z) - h(z) \cdot \sqrt{1 - \frac{z}{z_k}}$$

(a) $k = 3$.   **(b)** $k = 8$.
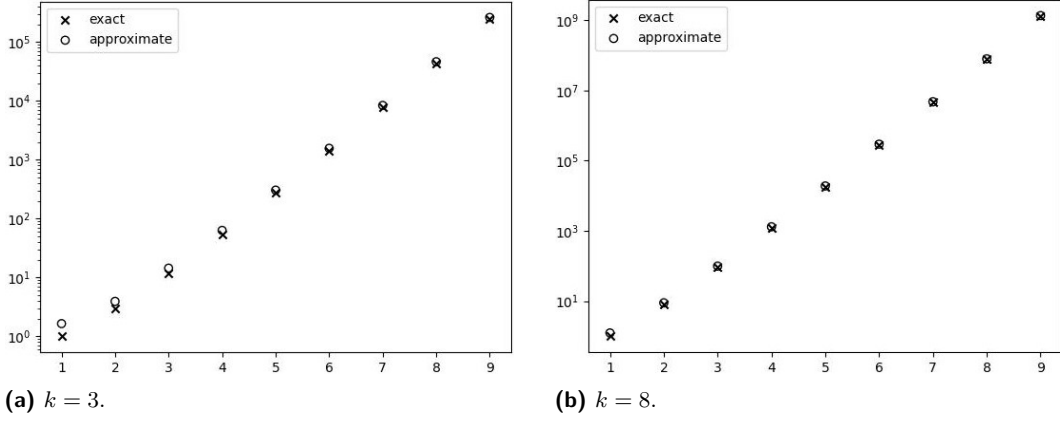
■ **Figure 2** Plots of the first ten coefficients of the counting sequence associated with $Q(z)$, for $k = 3$ (left) and $k = 8$ (right), computed both by the exact formula and according to the asymptotic estimate in (13).

where $g(z)$ and $h(z)$ are analytic in an open neighborhood of z_k, and

$$[z^n] \, P(z) = \sqrt{\frac{z_k \, F_z\big(z_k, P(z_k)\big)}{2\pi \, F_{pp}\big(z_k, P(z_k)\big)}} \, z_k^{-n} \, n^{-3/2} \, \big(1 + \mathrm{O}(n^{-1})\big),$$

from which the lemma follows.   ◀

We now address the asymptotic behavior of the numerators in (10)-(11). For this, note that by implicit differentiation in (12), we find that

$$Q_u(z) = \frac{k \, z^2 \, \{Q(z)\}^{2k-1}}{\big(1 - kz \, \{Q(z)\}^{k-1}\big)^2}; \tag{14}$$

$$Q_{uu}(z) = \frac{k(7k-1)z^3\{Q(z)\}^{3k-2}}{(1 - kz\{Q(z)\}^{k-1})^3}$$
$$+ \frac{k^2(7k-1)z^4\{Q(z)\}^{4k-3}}{(1 - kz\{Q(z)\}^{k-1})^4} + \frac{5k^3(k-1)z^5\{Q(z)\}^{5k-4}}{(1 - kz\{Q(z)\}^{k-1})^5}. \tag{15}$$

(For details about these derivations refer to Appendix A of [22].) Our next result implies that z_k is the only singularity of the above partial derivatives in the disk $|z| \leq z_k$.

▶ **Lemma 13.** *The equation $kz \, \{Q(z)\}^{k-1} = 1$, with $|z| \leq z_k$, has only z_k as a solution.*

Proof. Per Corollary 11, we know that z_k solves the equation $kz \, \{Q(z)\}^{k-1} = 1$. On the other hand, because $Q(z)$ has non-negative coefficients, $|kz \, \{Q(z)\}^{k-1}| < k \, z_k \, \{Q(z_k)\}^{k-1} = 1$ for $|z| < z_k$; in particular, any solution to the equation must lie on the circle $|z| = z_k$. But, if $|z| = z_k$ and $z \neq z_k$ then $|Q(z)| < Q(|z|) = Q(z_k)$ because Q has all powers of z with strictly positive coefficients. Hence, $|kz\{Q(z)\}^{k-1}| = kz_k|Q(z)|^{k-1} < kz_k\{Q(z_k)\}^{k-1} = 1$, which completes the proof of the lemma.   ◀

Notice that the generating functions in (14)-(15) share a common form, which as we see next may be exploited to yield a general asymptotic expansion of their coefficients.

▶ **Lemma 14.** *If $f : \mathbb{C} \to \mathbb{C}$ is an entire analytic function such that $f(z_k) \neq 0$, and $a \geq 0$ and $b \geq 1$ are integers, then the function*

$$[z^n] \, \frac{f(z)\{Q(z)\}^a}{(1 - kz\{Q(z)\}^{k-1})^b} = \frac{f(z_k)}{2^{b/2} \, \Gamma(b/2)} \left(\frac{k}{k-1} \right)^{a+b/2} n^{(b-2)/2} \, z_k^{-n} \left(1 + \mathrm{O}\left(n^{-1/2} \right) \right).$$

Proof (Sketch). For further details, refer to the proof of [22, Lemma 3.2.5].

Define

$$P(z) := \frac{A(z)}{\left(1 - B(z) \right)^b},$$

where $A(z) := f(z)\{Q(z)\}^a$ and $B(z) := kz\{Q(z)\}^{k-1}$. By Lemma 12, we can write

$$Q(z) = g(z) - h(z)\sqrt{1 - \frac{z}{z_k}},$$

where $g(z)$ and $h(z)$ are analytic in an open neighborhood of z_k. In particular, locally around z_k, we have that

$$f(z) = f_0 + \mathrm{O}\left(1 - \frac{z}{z_k} \right); \quad g(z) = g_0 + \mathrm{O}\left(1 - \frac{z}{z_k} \right); \text{ and } h(z) = h_0 + \mathrm{O}\left(1 - \frac{z}{z_k} \right);$$

where f_0, g_0, and h_0 are $f(z_k)$, $g(z_k)$, and $h(z_k)$, respectively.

To determine an asymptotic formula for the coefficients of $P(z)$, we first seek singular expansions for its numerator and denominator. For this, observe that locally around z_k:

$$A(z) = \alpha_0 + \mathrm{O}\left(\sqrt{1 - \frac{z}{z_k}} \right);$$

$$B(z) = \beta_0 - \beta_1 \sqrt{1 - \frac{z}{z_k}} + \mathrm{O}\left(1 - \frac{z}{z_k} \right);$$

where $\alpha_0 := f_0 g_0^a$, $\beta_0 := kz_k g_0^{k-1}$, and $\beta_1 := k(k-1)z_k g_0^{k-2} h_0$. But we know from Lemma 13 that $B(z_k) = 1$; in particular, $\beta_0 = 1$, and as a result

$$P(z) = \frac{f_0 g_0^a}{\{k(k-1)z_k g_0^{k-2} h_0\}^b} \left(1 - \frac{z}{z_k} \right)^{-b/2} \left(1 + \mathrm{O}\left(\sqrt{1 - \frac{z}{z_k}} \right) \right).$$

Finally, note that $g_0 = Q(z_k)$, hence $g_0 = \frac{k}{k-1}$ due to Corollary 11. On the other hand, due to Lemma 12, $h_0 = \frac{1}{k-1}\sqrt{\frac{2k}{k-1}}$. Therefore

$$P(z) = \frac{f(z_k)}{2^{b/2}} \cdot \left(\frac{k}{k-1} \right)^{a+b/2} \left(1 - \frac{z}{z_k} \right)^{-b/2} \left(1 + \mathrm{O}\left(\sqrt{1 - \frac{z}{z_k}} \right) \right),$$

and the lemma follows from [9, Theorem VI.4]. ◀

Finally, by applying Lemma 13 to each of the terms in (14)-(15), we obtain that

$$[z^n] \, Q_u(z, 1) = \frac{k}{2(k-1)^2} \cdot z_k^{-n} \left(1 + \mathrm{O}(n^{-1/2}) \right);$$

$$[z^n] Q_{uu}(z) = \frac{5k}{6(k-1)^2} \sqrt{\frac{2kn^3}{\pi(k-1)}} \cdot z_k^{-n} \left(1 + \mathrm{O}(n^{-1/2}) \right);$$

from which the asymptotic formulas for the expected value and variance of the internal path length of a uniformly at random k-regular tree in Theorem 10 follow. ◀

(a) Dense covariance matrix.

(b) Sparsified covariance matrix.
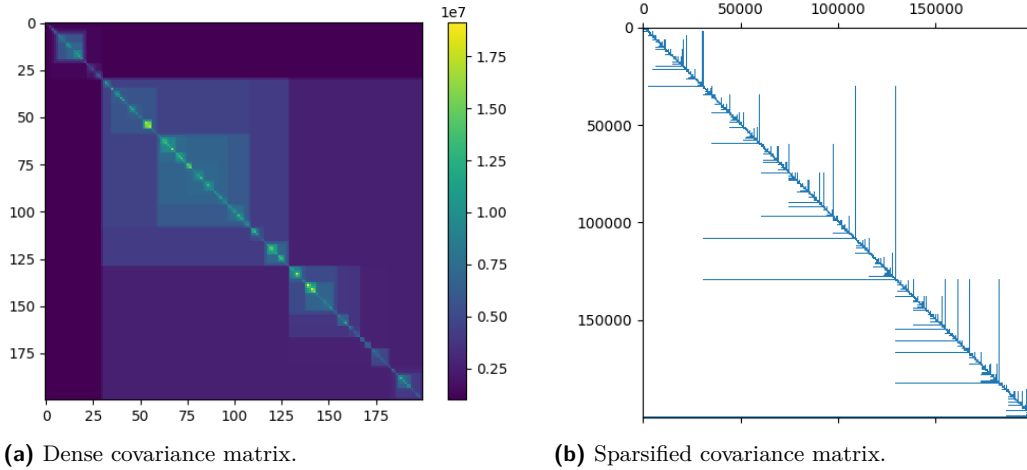
Figure 3 (a) Heat-map visualization of a dense covariance matrix C, and (b) sparsity pattern of $\Phi'C\Phi$ for a 3-regular tree with 200,001 leaves. The dense matrix has over 40 billion non-zero entries. The sparse matrix, on the other hand, has only about 0.03% as many non-zero entries. The heatmap of the dense matrix was produced downsampling the sparse representation by a factor of 1000, prior to undoing the change-of-basis.

3.3   Sparsification of Large, Random, k-Regular Covariance Matrices

So far, we have seen that the Haar-like wavelets can partially sparsify the (phylogenetic) covariance matrix of a k-regular tree by changing the basis, obtaining a lower bound on the proportion of vanishing entries under the basis change. A remaining challenge is determining whether trees meeting the criteria for a high degree of sparsification are common. In this section, we find that such trees are, in fact, abundant and that large random k-regular trees are highly sparsified by the Haar-like wavelets with overwhelmingly high probability.

▶ **Corollary 15.** *Let $k \geq 2$ and $\mathbb{T}$ be a uniformly at random k-regular tree with $|I|$ internal nodes, and C its covariance matrix. If Φ is the Haar-like matrix associate with $\mathbb{T}$, and ζ the fraction of vanishing entries in $\Phi'C\Phi$, then $\zeta \to 1$ in probability, as $|I| \to \infty$.*

Proof. The argument mirrors the one for [14, Corollary 3.8].

Let μ and σ^2 denote the expectation and variance of the internal path length of $\mathbb{T}$, respectively. Cantelli's inequality states that $\mathbb{P}\big(\mathrm{IPL}(\mathbb{T}) > \mu + t\sigma\big) \leq (1 + t^2)^{-1}$, for all $t > 0$. But, from Theorem 10, we know that $\mu + t\sigma = \Omega(t|I|^{3/2})$; that is, there is a constant $c > 0$ such that $\mathbb{P}\big(\mathrm{IPL}(T) > c\,t|I|^{3/2}\big) \leq (1 + t^2)^{-1}$, or equivalently

$$\mathbb{P}\left(\frac{\mathrm{IPL}(\mathbb{T})}{|I|^2} \leq \frac{c\,t}{\sqrt{|I|}} \right) \geq \frac{t^2}{1 + t^2}, \text{ for all } t > 0.$$

The result follows by choosing $t \to \infty$ so that $t = o(\sqrt{|I|})$.   ◀

4   Discussion

Phylogenetic covariance matrices are often large and dense to the point of being computationally unmanageable; however, we have demonstrated that expressing them in the Haar-like basis can significantly sparsify them. In particular, we have shown that the covariance matrix of a random k-regular tree will be highly sparsified with overwhelmingly high probability as
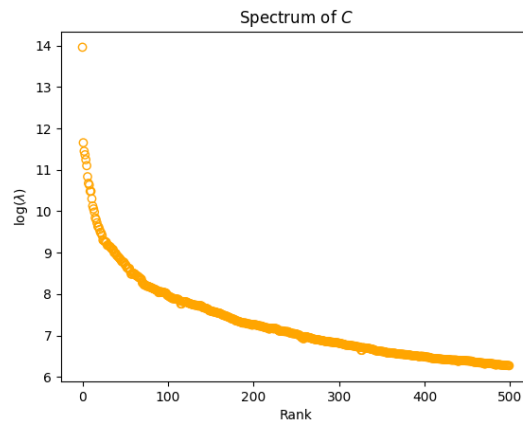
Figure 4 The largest 500 eigenvalues of a covariance matrix C associated with a 3-regular tree with 200,001 leaves. The spectrum was computed from its sparse representation, obtained after changing to the Haar-like basis.

the size of the tree tends to infinity. In this section, we illustrate the process of sparsifying a large phylogenetic covariance matrix and discuss some of the practical considerations of the method.

Consider Figure 3, which illustrates the application of the Haar-like wavelet transform on a random 3-regular tree with 200,001 leaves. The tree was generated by simulating a Galton-Watson process until a desired population size (i.e., number of leaves) was reached. We find that, although the original, dense matrix contains over 40 billion non-zero entries, the resulting sparse matrix contains a comparatively meager 11.3 million; that is, over 99.97% of the entries in the original matrix were zeroed by the transform. Importantly, because $\Phi'C\Phi$ and C are similar matrices, we can quickly calculate the spectrum of C from its sparse representation, illustrated in Figure 4. While this random tree model may not precisely induce actual phylogenetic trees [2–4], we emphasize that the sparsification observed results in an overwhelming part from the tree's topology, irrespective of its edge lengths (which effectively store covariances). Hence, the simulated data gives reasonable insight into the performance of this technique in general.

As already mentioned, the degree of sparsification guaranteed by Theorem 4 is conservative. For one, it does not account for the fact that the Haar-like wavelets are orthonormal, which would suggest that at least some entries associated with two wavelets ψ_u and ψ_v, where u is an ancestor of v, vanish. In fact, if a k-regular tree is *trace balanced* [14] (that is, for all $v \in I$ and $i, j \in L(v)$, $\ell^*(i, v) = \ell^*(j, v)$), Theorem 2 implies that the Haar-like wavelets fully diagonalize its covariance matrix. As a result, we often see a better degree of sparsification than what is guaranteed by Theorem 4.

Our analysis shows that the technique of sparsifying dense phylogenetic covariance matrices by a change-of-basis with the Haar-like wavelets extends to the broader class of k-regular trees. While further work is required to verify the exact performance of this method on random k-ary trees (i.e., ones for which each interior node contains *at most k* children), well-known properties of generating functions enumerating simple varieties of trees suggest that such a generalization is possible by the methods employed here. Comparison with k-regular trees indicates sparsification for what one may call "almost" k-regular trees (i.e., k-ary trees which are k-regular except at a relatively small number of internal nodes); however, initial investigation suggests that the worst-case k-ary tree might result in poor sparsification.

To conclude, it is worth noting that the k-regular and k-ary trees belong to the class of "simply generated trees" [17], a generalization which may prove helpful for further expanding our methodology. Exploring this avenue of research could grant access to new datasets characterized by more intricate yet richer hierarchical structures.

References

1   K. M. Abadir. An introduction to hypergeometric functions for economists. *Econometric Reviews*, 18(3):287–330, 1999.

2   D. Aldous. Probability distributions on cladograms. In D. Aldous and R. Pemantle, editors, *Random Discrete Structures*, pages 1–18, New York, NY, 1996. Springer New York.

3   D. Aldous and B. Pittel. The critical beta-splitting random tree: Heights and related results, 2023. `arXiv:2302.05066`.

4   D. J. Aldous. The critical beta-splitting random tree II: Overview and open problems, 2023. `arXiv:2303.02529`.

5   L. L. Cavalli-Sforza and A. W. Edwards. Phylogenetic analysis: models and estimation procedures. *Evolution*, 21(3):550, 1967. `doi:10.2307/2406616`.

6   C. Dellacherie, S. Martinez, and J. San Martín. *Inverse M-Matrices and Ultrametric Matrices*, volume 2118 of *Lecture Notes in Mathematics*. Springer, 2014.

7   M. Drmota. *Random Trees: An Interplay between Combinatorics and Probability*. Springer-Verlag/Wein, 2009.

8   R. J. Evans and D. Stanton. Asymptotic formulas for zero-balanced hypergeometric series. *SIAM J. Math. Anal.*, 1984. `doi:10.1137/0515078`.

9   P. Flajolet and R. Sedegwick. *Analytic Combinatorics*. Cambridge University Press, 2009. URL: `http://www.cambridge.org/uk/catalogue/catalogue.asp?isbn=9780521898065`.

10   J. Fukuyama, P. J. McMurdie, L. Dethlefsen, D. A. Relman, and S. Holmes. Comparisons of distance methods for combining covariates and abundances in microbiome studies. *Biocomputing*, pages 213–224, 2012. URL: `http://psb.stanford.edu/psb-online/proceedings/psb12/fukuyama.pdf`.

11   Matan Gavish, Boaz Nadler, and Ronald R. Coifman. Multiscale wavelets on trees, graphs and high dimensional data: theory and applications to semi supervised learning. In *Proceedings of the 27th International Conference on International Conference on Machine Learning*, ICML'10, pages 367–374, Madison, WI, USA, 2010. Omnipress.

12   I. Gessel. How can I verify this family of values for hypergeometric functions? MathOverflow, september 12 2023. URL: `https://mathoverflow.net/q/454420`.

13   E. Gorman and M. E. Lladser. Interpretable metric learning in comparative metagenomics: The adaptive Haar-like distance. PLoS Comput Biol, 2024 (to appear). URL: `https://www.biorxiv.org/content/10.1101/2023.09.27.559681v1`.

14   E. Gorman and M. E. Lladser. Sparsification of large ultrametric matrices: insights into the microbial Tree of Life. *Proc. R. Soc. A*, 479:20220847, 2023. `doi:10.1098/rspa.2022.0847`.

15   L. J. Harmon. *Phylogenetic Comparative Methods*. CreateSpace Independent Publishing Platform, 2019.

16   E. Hille. *Analytic function theory. Vol. 1*. Introduction to Higher Mathematics. Ginn and Company, 1959.

17   Svante Janson. Simply generated trees, conditioned Galton–Watson trees, random allocations and condensation. *Probability Surveys*, 9:103–252, 2012.

18   S. Martinez, G. Michon, and J. San Martín. Inverse of strictly ultrametric matrices are of Stieltjes type. *SIAM J. Matrix Anal. Appl.*, 15(1):98–106, 1994. `doi:10.1137/s0895479891217011`.

19   A. Meir and J. W. Moon. On the altitude of nodes in random trees. *Canadian Journal of Mathematics*, 30, 1978.

20   R. Nabben and R. S. Varga. A linear algebra proof that the inverse of a strictly ultrametric matrix is a strictly diagonally dominant stieltjes matrix. *SIAM J. Matrix Anal. Appl.*, 15(1):107–113, 1994. `doi:10.1137/s0895479892228237`.

21   S. Pavoine, A.-B. Dufour, and D. Chessel. From dissimilarities among species to dissimilarities among communities: a double principal coordinate analysis. *Journal of Theoretical Biology*, 228(4):523–537, 2004. `doi:10.1016/j.jtbi.2004.02.014`.

22   Sean Svihla. Sparsification of covariance matrices of k-regular trees. Master's thesis, The University of Colorado, 2024.

23   E. W. Weisstein. Hypergeometric function. `https://mathworld.wolfram.com/HypergeometricFunction.html`. Accessed: September 2023.

24   Q. Zhu, U. Mai, and W. Pfeiffer et al. Phylogenomics of 10,575 genomes reveals evolutionary proximity between domains bacteria and archaea. *Nature Communications*, 10:5477, 2019.

Bit-Array-Based Alternatives to HyperLogLog

Svante Janson ✉ 🏠 🆔
Department of Mathematics, Uppsala University, Sweden

Jérémie Lumbroso ✉ 🏠 🆔
Department of Computer Science, University of Pennsylvania, Philadelphia, PA, USA

Robert Sedgewick[1] ✉ 🏠 🆔
Department of Computer Science, Princeton University, NJ, USA

───── **Abstract** ─────

We present a family of algorithms for the problem of estimating the number of distinct items in an input stream that are simple to implement and are appropriate for practical applications. Our algorithms are a logical extension of the series of algorithms developed by Flajolet and his coauthors starting in 1983 that culminated in the widely used `HyperLogLog` algorithm. These algorithms divide the input stream into M substreams and lead to a time-accuracy tradeoff where a constant number of bits per substream are saved to achieve a relative accuracy proportional to $1/\sqrt{M}$. Our algorithms use just one or two bits per substream. Their effectiveness is demonstrated by a proof of approximate normality, with explicit expressions for standard errors that inform parameter settings and allow proper quantitative comparisons with other methods. Hypotheses about performance are validated through experiments using a realistic input stream, with the conclusion that our algorithms are more accurate than `HyperLogLog` when using the same amount of memory, and they use two-thirds as much memory as `HyperLogLog` to achieve a given accuracy.

1 Introduction

Counting the number of distinct items in a data stream is a classic computational challenge with many applications. As an example, consider the stream of strings taken from a web log shown in the left column of Table 1 (we will use 1 million strings from this log of which $N = 368,217$ are distinct values as a running example in this paper). There is no bound on the length of the stream, but maintaining an estimate of the number of different strings is useful for many purposes.

---

[1] Corresponding author

One classic application is in computer networks. The ability to estimate the number of different visitors of a website is certainly of interest, and can be critical in maintaining the integrity of the site. For example, a significant drop in the percentage of distinct visitors in a given time period might be an indication that the site is under a denial-of-service attack.

Another classic application is found in database systems, where estimating the number of different strings having each attribute is a critical piece of knowledge in implementing certain common data base operations. In this case, the length of the streams is available, but may be very large, and a rough estimate suffices, so using a streaming algorithm is appropriate.

Elementary algorithms for solving the problem are standard in introductory computer science classes. Perhaps the simplest is to use a *hash table*, but that requires saving all the items in memory, which is far too high a cost to be useful in typical applications. In fact, *any* method for computing an exact count must save all the items in memory (trivial proof: any item not saved might or might not be distinct from all the others, and that fact cannot be known until the last item is seen).

Accordingly, we focus on *estimating* the count. In typical applications, exact counts are actually not needed – the estimates are being used to make relative decisions that do not require full accuracy.

Since the seminal research by Flajolet and Martin in the 1980s [5][6] it has been known that we actually can get by with a surprisingly small amount of memory. The *practical cardinality estimation problem* is to estimate the number of distinct items in a data stream under the following constraints:

- Each item is examined only once.
- The time to process each item is a very small constant multiple of the size of the item.
- The amount of memory used is very small, no matter how large the stream.
- The estimate is expected to be within a small percentage of the real count.

A solution to this problem typically is defined by an implementation that makes clear its time and space requirements and an analysis that provides a precise characterization of how the estimate compares to the actual value.

For many years, the state of the art in solving the practical cardinality estimation problem has been `HyperLogLog`, the last in the series of algorithms developed by Flajolet and colleagues from the 1980s through the 2000s [4] [7] [9] [14]. `HyperLogLog` is based on four main ideas: *Hashing* is used to convert each item in the stream into a fixed-length binary number; the position of the *rightmost zero* is computed, taking the maximum value found as an estimate of the binary logarithm of the count; a technique known as *stochastic averaging* splits the stream into M independent substreams so that an average of experimental results can be computed; and the *harmonic mean* is used to properly handle outlying values. One reason `HyperLogLog` is so widely used is that *precise analysis* of the bias in the estimate provides the basis for formulating hypotheses about how the algorithm will perform in practical situations, and the results of experiments that validate the hypotheses are presented. The analysis exposes a *space-accuracy tradeoff*, allowing practitioners to choose with confidence the amount of memory needed to achieve a given accuracy or the accuracy achieved for a given amount of memory use: For a stream with N distinct values and using M substreams, `HyperLogLog` uses $M \lg \lg N$ bits and typically produces an estimate with a relative standard error of $c/\sqrt{M}$ where $c \doteq 1.04$.

A series of theory papers have proven that $O(M)$ bits are necessary and sufficient to achieve estimates with asymptotic accuracy on the same order as `HyperLogLog`, an important and significant accomplishment [1][10][11]. However, these papers lack implementations, likely because the implied constants in the proofs are much too large for the methods to be

■ **Table 1** Computing a sketch for `HyperBitT` (with $M = 8$ and $T = 1$).

s	x	k	r(x)	sketch[]
81.95.186.98.freenet.com.ua	111...1111101111011110011101011	7	2	00000001
lsanca.dsl-w.verizon.net	011...1010100010001111110100000	3	0	00000001
117.222.48.163	110...0111001100000111011101101	6	1	00000001
1.23.193.58	100...0100101001000011101100011	4	2	00001001
188.134.45.71	101...0101111000101101000111001	5	1	00001001
gsearch.CS.Princeton.EDU	010...1010011011000011010000100	2	0	00001001
81.95.186.98.freenet.com.ua	011...1011110001110000111010000	3	0	00001001
81.95.186.98.freenet.com.ua	111...1111101111011110011101011	7	2	00001001
1.23.193.58	000...0100111001111101011100111	0	3	10001001
lnse3.cht.bigpond.net.au	111...0110011001011011101001110	7	0	10001001
117.211.88.36	000...1001100010100010010111010	0	0	10001001
1.23.193.58	000...0100111001111101011100111	0	3	10001001
lsanca.dsl-w.verizon.net	011...1010100010001111110100000	3	0	10001001
81.95.186.98.freenet.com.ua	111...1111101111011110011101011	7	2	10001001
gsearch.seas.upenn.edu	000...1000100011011010100001000	0	0	10001001
109.108.229.102	010...1010111010101101100111111	2	5	10101001
msnbot.search.msn.com	001...1001110110111001001101100	1	0	10101001

viable in practice . The theory literature also makes the implicit assumption that strong assumptions on the hash functions are necessary (even to the point of dismissing algorithms like `HyperLogLog` as illegitimate [1]). Strong hash functions add to the expense of processing each item, and the idea that using one makes any difference at all in practice is tenuous at best (see, for example, [3] for a discussion of this issue). In this paper, we focus on algorithms with the potential to be useful in practice – we use hash functions that are widely used in practice and hypothesize that any differences from the ideal are relatively insignificant. *Any* practical application of hashing, however perfect in theory, must assume, at least, that random bits exist, and therefore requires such a hypothesis.

`HyperLogLog` uses $5M$ bits for $N < 2^{32}$, but much higher values are typical in modern applications. Since it is safe to assume that $N < 2^{64}$, `HyperLogLog` demonstrates that $6M$ bits suffice for the practical cardinality estimation problem. Some improvements to `HyperLogLog` and some interesting new approaches to the problem have been studied in recent years [16] [19] [15] [12] [17] but we are still left with the following question: can we find a practical algorithm as simple as `HyperLogLog` with comparable accuracy that uses cM bits for some constant c that is significantly less than 6?

In this paper, we provide answers to this question. The algorithms we present have the same structure as `HyperLogLog` but use much less memory – instead of recording the maximum number of trailing ones, we focus on *one* bit per sub-stream indicating whether a threshold has been hit. In Section 2, we use a rough estimate of the cardinality as an input parameter in order to set the threshold to be the logarithm of the extimated number of distinct items per substream. As such, the resulting algorithm is *not* a streaming algorithm, but it serves as a basis for the streaming algorithms in Section 3 and Section 4 that do solve the practical cardinality estimation problem, using just *two* bits per substream. In Section 5 we conclude by discussing how these algorithms match up against those in the literature.

2   HyperBitT

Our first algorithm uses the standard technique of starting with a rough estimate of the cardinality and is therefore not properly a streaming algorithm, as no fixed estimate can remain accurate as the cardinality grows without bound. We consider this algorithm because, as we will see, it is sometimes useful in its own right, and it admits a precise analysis that we can use to develop the streaming algorithms in Section 3 and Section 4.

We start with hashing and stochastic averaging with M substreams precisely as does `HyperLogLog`, but use just *one* bit per substream, as follows. Of course, we expect each substream to have about N/M distinct values, and it has been known since the original work of Flajolet and Martin [5] that the maximum number of trailing 1s found among the items in a stream is a good estimator of the logarithm of the number of distinct items in the stream. (Indeed, this is the same as the length of the rightmost path in a random trie, a quantity that was studied in the 1970s.) In this spirit, we use a parameter T as an estimate of $\lg(N/M)$. That is, 2^T is an estimate of N/M, and $2^T M$ is an estimate of the cardinality N. Now, we maintain a *sketch* comprising an array of M bits, one per substream, and set the bit corresponding to a substream to 1 when an item from that substream has more than `T` trailing 1s. When we want to estimate the number of distinct values in the stream, it turns out that we can use a simple function of the number of 0 bits in the sketch to improve our estimate. The algorithm may produce an inaccurate result or fail completely if the rough estimate `T` is poorly chosen, but, as we will see, it is remarkably forgiving.

Implementation

We start with a bit array `sketch[]` with one bit per substream, initialized to all 0s. For clarity, we use a `bit[]` type to describe our algorithms – although few programming languages support an explicit `bit[]` type, the abstraction is easily implemented. For small M, we can use integer values; for large M, we can use shifting and masking on arrays of integers (see Appendix B). We typically use a power of two for convenience.

For each new item `s` in the stream, we compute a hash value `x` to represent it and a second hash value `k` to identify its substream (typically, one might compute a 64-bit hash and use the leading $\lg M$ bits for `k` and the rest for `x`). Then we compute `r(x)`, the number of trailing 1s in `x`. As described in Appendix B, this operation can be implemented with only a few machine-language instructions. If `r(x)` is larger than `T`, we set `sketch[k]` to 1. Table 1 is a trace of the process for a small sequence of hash values with $M = 8$ and $T = 1$.

When the stream is exhausted, we compute a correction to the rough estimate of $N = 2^T M$ that takes into account some bias, as a function of the bit values in the sketch. Specifically, we are interested in the parameter β, the proportion of 0s in the sketch. As indicated by the analysis below, the appropriate correction factor is $\ln(1/\beta)$. If the sketch is small enough to fit in a computer word, computing the number of 1s in the sketch is a classic machine-language programming exercise and is actually a single instruction in many modern machine architectures. For clarity, we use the function `p(sketch)`; for large M it is preferable to just increment a counter each time a sketch bit is changed from 0 to 1, as described in Appendix B. The implementation in Algorithm 1 follows immediately and is easily translated to any programming language.

If T is too small or too large, the algorithm fails because the estimate cannot be reasonably corrected (when β is close to 0 or 1, the correction factor is too large or too small to be useful). But, as we shall see, the algorithm does produce accurate results for a remarkably large range of cardinality values, and we can precisely characterize that range and the accuracy.

🟨 **Algorithm 1** `HyperBitT`.

```
public static int estimateHBT (Iterable<String> stream, int M, int T)
{
   bit[] sketch[M];
   for (String s : stream)
   {
      long x = hash1(s);                 // 64-bit hash
      int  k = hash2(s, M);              // (lg M)-bit hash
      if (r(x) > T) sketch[k] = 1;       // more than T trailing 1s?
   }
   double beta = 1.0 - 1.0*p(sketch)/M;  // fraction of 0s in sketch
   return (int) (Math.pow(2, T)*M*Math.log(1.0/beta));
}
```

Analysis

As a basis for developing an intuition about the problem, we start with an approximate analysis for the mean value of the number of distinct values in the stream. After N distinct values have been processed from the input stream, we have seen an average of N/M distinct values in each substream. As an approximation, assume that *exactly N/M* values go to each substream. The probability that a given value has at least T trailing `1`s is $1/2^T$ so the probability that a given bit in `sketch[]` remains `0` after N/M values are processed in its corresponding substream is given by a Poisson approximation

$$\left(1 - \frac{1}{2^T}\right)^{N/M} \sim e^{-N/(M2^T)}$$

(see for example, [18]). The number of `0`s in `sketch[]` is a binomially distributed random variable, so this value is also (approximately) β, the expected proportion of `0`s in `sketch[]` after N values have been processed. Thus, $N/M \sim 2^T \ln(1/\beta)$ and the expected number of values processed is $N \sim M2^T \ln(1/\beta)$. In other words, we need to correct our rough estimate of the number of values per stream by the factor $\ln(1/\beta)$.

A full detailed analysis provides much more information, which is critical for studying the performance of the algorithm. Specifically, we are able to approximate the *distribution* of the reported cardinality, which gives us the information needed to estimate how accurate it will be for given values of M.

The proof is based on the idea of *Poissonization* – instead of assuming that we have a fixed given number N of distinct items, we assume that the number is random with a Poisson distribution. It uses two technical lemmas from probability theory:

▶ **Lemma 1.** *Suppose that $X_n \geq 0$ are random variables and a_n, b_n, and σ^2 numbers such that, as $n \to \infty$, we have $a_n \to a > 0$, $b_n \to 0$, and $(X_n - a_n)/b_n \xrightarrow{d} \mathbb{N}(0, \sigma^2)$. If f is a continuously differentiable function on $(0, \infty)$ with $f'(a) \neq 0$, then $(f(X_n) - f(a_n))/b_n \xrightarrow{d} \mathbb{N}(0, f'(a)^2\sigma^2)$.*

Proof. See Appendix A. ◀

▶ **Lemma 2.** *Let $X \sim \text{Binomial}(n, p)$ and let $Y \in \text{Poisson}(np)$ where $n > 0$ and $p \in [0, 1]$. Then the total variation distance between them $d_{TV}(X, Y)$ is no greater than p; in other words there exists a coupling of X and Y such that $\mathbb{P}(X \neq Y) \leq p$.*

Proof. See Theorem 2.M and pages 1–8 in [2]. ◀

▶ **Theorem 3.** *Suppose that a stream S has N distinct items and that* `HyperBitT` *processes S using M substreams with parameter T and terminates with βM 0s left in the sketch. Then the statistic $M2^T \ln(1/\beta)$ is approximately Gaussian with mean N and relative standard error $c_\beta/\sqrt{M}$ where $c_\beta = \sqrt{1/\beta - 1}/\ln(1/\beta)$. Formally,*

$$\frac{\sqrt{M}}{c_\beta}\left(\frac{M2^T \ln(1/\beta)}{N} - 1\right) \xrightarrow{d} \mathbb{N}(0,1) \tag{1}$$

as $N, M, T \to \infty$ with $N = \Theta(M2^T)$.

Proof. Assume first that $N \sim aM2^T$ for some $a \in (0, \infty)$. Pretend that the distinct items in the stream arrive according to a Poisson process with rate 1. We then may consider the process at a given time $\tilde{N}$. If we keep $\tilde{N}$ fixed, then the number of distinct items seen so far is a random variable obeying a Poisson distribution $\text{Poisson}(\tilde{N})$. We let $\tilde{N} \sim N \sim aM2^T$. For reference, we summarize here the notations used in this proof:

- $N \sim aM2^T$, the cardinality of the stream seen so far when Algorithm 1 terminates
- a, a positive number
- $\hat{N} = M2^T \ln(1/\beta)$, the reported estimate of N
- $\tilde{N} \sim aM2^T$, the Poisson parameter

Our goal is to approximate the distribution of $\hat{N}$.

We begin by finding, in the Poisson model, the distribution of βM, the number of 0s in the sketch. Since a randomly thinned Poisson process is a new Poisson process, it follows that each of the M substreams is a Poisson process with rate $1/M$, and thus the number of distinct items in each of them is $\text{Poisson}(\tilde{N}/M)$. These random numbers are independent, and each item in the kth substream has probability 2^{-T} to set `sketch[k]` to `1`. It follows that if the number of such items is Y_k, then Y_k is also Poisson, with $Y_k \in \text{Poisson}(2^{-T}\tilde{N}/M) = \text{Poisson}(\tilde{N}/(M2^T))$. Now, let q be the probability that `sketch[k]=0` (which is the same for all k). Then

$$q = \mathbb{P}(Y_k = 0) = \exp\left(-\frac{\tilde{N}}{M2^T}\right) \to e^{-a}. \tag{2}$$

Since the numbers Y_k are independent, the total number of 0s in the sketch is

$$\beta M \in \text{Binomial}(M, q). \tag{3}$$

with mean Mq and variance $Mq(1-q)$.

As $M \to \infty$, we have the normal approximation to the binomial:

$$\sqrt{M}(\beta - q) = \frac{M\beta - Mq}{\sqrt{M}} \xrightarrow{d} \mathbb{N}\left(0, e^{-a}(1 - e^{-a})\right). \tag{4}$$

Now, applying Lemma 1 with the function $f(x) = \ln(1/x)$ gives

$$\sqrt{M}\left(\ln(1/\beta) - \ln(1/q)\right) \xrightarrow{d} \mathbb{N}(0, e^a - 1). \tag{5}$$

Consequently, since $\hat{N} = M2^T \ln(1/\beta)$, $M2^T/\tilde{N} \to 1/a$, and $\ln(1/q) = \tilde{N}/M2^T$, we have

$$\sqrt{M}\left(\frac{\hat{N}}{\tilde{N}} - 1\right) = \sqrt{M}\frac{M2^T}{\tilde{N}}\left(\ln\frac{1}{\beta} - \ln\frac{1}{q}\right) \xrightarrow{d} \mathbb{N}\left(0, a^{-2}(e^a - 1)\right). \tag{6}$$

Furthermore, (5) implies $\ln(1/\beta) - \ln(1/q) \xrightarrow{p} 0$, and thus, using (2), $\ln(1/\beta) \xrightarrow{p} a$; hence (6) implies (1) (with $\tilde{N}$ instead of N).
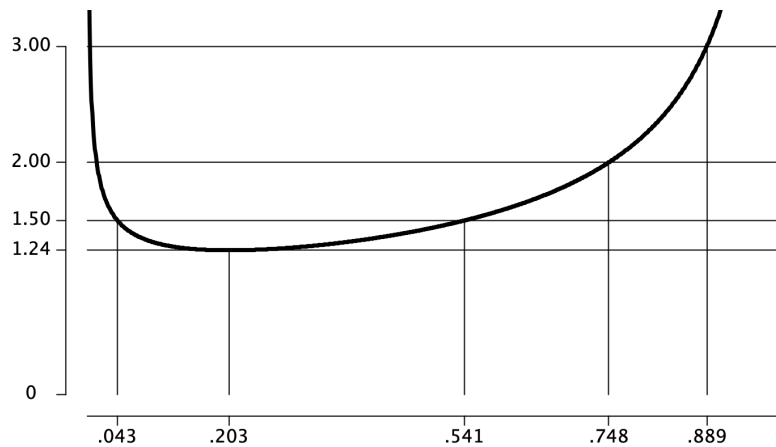
Figure 1 This plot shows the coefficient of $1/\sqrt{M}$ in the relative standard error $c_\beta = \sqrt{1/\beta - 1}\big/\ln(1/\beta)$ (y-coordinate) for β (fraction of 0s in the sketch) between 0 and 1 (x-coordinate). The value of c_β goes to infinity as β approaches 0 or 1, but it is relatively small when β is not close to these extremes. For example, $c_\beta < 1.5$ when $.043 < \beta < .541$, $c_\beta < 2$ when $.014 < \beta < .748$, and $c_\beta < 3$ when $.0035 < \beta < .888$.

This is the desired result for the Poisson model. To prove the result for a given number N of items, we use Lemma 2. We may assume that we start by selecting all items with at least $\mathtt{T}$ trailing 1s. Since each item is selected with probability 2^{-T}, the number of selected items is $\mathrm{Binomial}(N, 2^{-T})$. Similarly, if we consider the Poisson model with $\mathrm{Poisson}(N)$ items (thus choosing $N = \tilde{N}$ above) then the number of selected items is $\mathrm{Poisson}(N2^{-T})$. By Lemma 2. we may couple the two versions such that the number of selected items agree with probability no less than $1 - 2^{-T} \to 1$. Hence, (1) for a fixed N follows from the Poisson version.

We have proved that (1) holds when $N/(M2^T)$ converges to a limit in $(0, \infty)$. The more general assumption $N = \Theta(M2^T)$ implies that every subsequence has a subsubsequence such that $N/(M2^T)$ converges, and thus (1) holds for the subsubsequence. As is well known, this implies that the full sequence converges (see Section 5.7 in [8]). ◀

To summarize, the goal of `HyperBitT` is to compute an estimate of N, the cardinality of the input stream. To do so, it takes two parameters

- M, the number of substreams (and the number of bits used)
- T, a rough estimate of $\lg(N/M)$

and, using an M-bit sketch, computes a value

- β, the fraction of 0s in the sketch.

Theorem 3 provides formulas for two important pieces of information, as functions of β:

- the correction factor $\ln(1/\beta)$, leading to the estimate $2^T M \ln(1/\beta)$ for N
- the coefficient of $1/\sqrt{M}$ in the relative standard error $c_\beta = \sqrt{1/\beta - 1}\big/\ln(1/\beta)$

This is the information that we need to properly choose the value of T. Of most interest is the fact that c_β is relatively small and is large only when β is close to 0 or 1 (see Figure 1). If T is too small, then the sketch will be predominately 1s, and β will be close to 0; if T is too large, the sketch will be predominantly 0s and β will be close to 1.

As an example, suppose that we take $M = 1024$ and aim to keep $c_\beta < 1.5$, which is the case when $.043 < \beta < .541$ (see Figure 1). As indicated in this table, each value of T leads to an accurate answer for a rather large range of values of N.

■ **Table 2** Since it is based on hash values, `HyperBitT` produces a different result every time it is run. The following table shows the result of five consecutive runs of `HyperBitT` for our sample web log with these parameter values. The last line compares the estimated cardinality with the actual value 368,217. Since our estimate of the standard error is conservative (c_β is usually smaller than 1.5), four of the five runs produced estimates well within the desired 5%. Since the distribution is Gaussian, the outlier in the first experiment is not unexpected.

# of 0s in `sketch[]`	$M\beta$	228	253	257	261	265
estimated cardinality	$2^T M \ln(1/\beta)$	393,773	366,498	362,386	358,338	354,351
estimated relative accuracy	$c_\beta/\sqrt{M}$	3.9%	3.9%	3.9%	3.9%	3.9%
actual relative accuracy		6.9%	0.5%	1.6%	2.7%	3.8%

T	6	7	8	9	10	11
$M2^T \ln(1/\beta)$ for $\beta = .541$	40,261	80,522	161,044	322,089	644,177	1,288,356
$M2^T \ln(1/\beta)$ for $\beta = .043$	206,212	412,425	824,850	1,649,701	3,299,402	6,598,804

Validation

The purpose of our analysis is to enable us to hypothesize that the cardinality returned by `HyperBitT` behaves as described by Theorem 3 and to set parameter values that keep the error low. As with any scientific study, our confidence in the result grows with the number of experiments that validate it, so we can only give an initial indication. (For example, practitioners have confidence in a similar hypothesis for `HyperLogLog` because it has been used in a wide variety of practical situations for years.)

The hypothesis rests on three main assumptions. First, we assume that the data we have and that the hash functions we use have the idealized properties stipulated in the analysis, or that deviations from this ideal are relatively insignificant. Second, we assume that the second hash function splits the stream into each substream with equal probability, or that deviations from this ideal are relatively insignificant. Third, we assume that deviations from approximations in the analysis are relatively insignificant.

For example, suppose that we wish to use `HyperBitT` to estimate the number of distinct strings in the web log described in Section 1. To do so, we need to specify the values of the two parameters: M (the number of bits of memory we need to use to achieve the accuracy that we want) and T (where $2^T M$ is our rough guess of the cardinality).

First, we choose the value of M. As an example, suppose that we are looking for an accurate answer, say with 5% relative error. Referring to Figure 1, if β is in the range $(.043, .541)$, then $c_\beta < 1.5$ and $M = 1024$ will do the trick, because $1.5/\sqrt{1024} \doteq .0469$. This is a conservative choice because c_β is usually much smaller than 1.5 in that range.

Next, we choose the value of T. Suppose we decide that it is a reasonable guess that the unique values comprise somewhere between 20% and 80% of the stream (a rather wide range). This leads to the choice $T = 8$ because $M2^T \ln(1/\beta)$ is between 161,044 and 824,851 (and $c_\beta < 1.5$) when β is between .541 and .043.

Table 2 shows the experimental results that constitute a quick validation check. Figure 2 describes two experiments that each run it *10 thousand* times, which both are strong evidence of the validity of our analysis and our hypotheses about the performance of `HyperBitT`.

(a) 100 trials every 10,000 inputs up to 1 million.

(b) 10,000 trials with 1 million inputs.
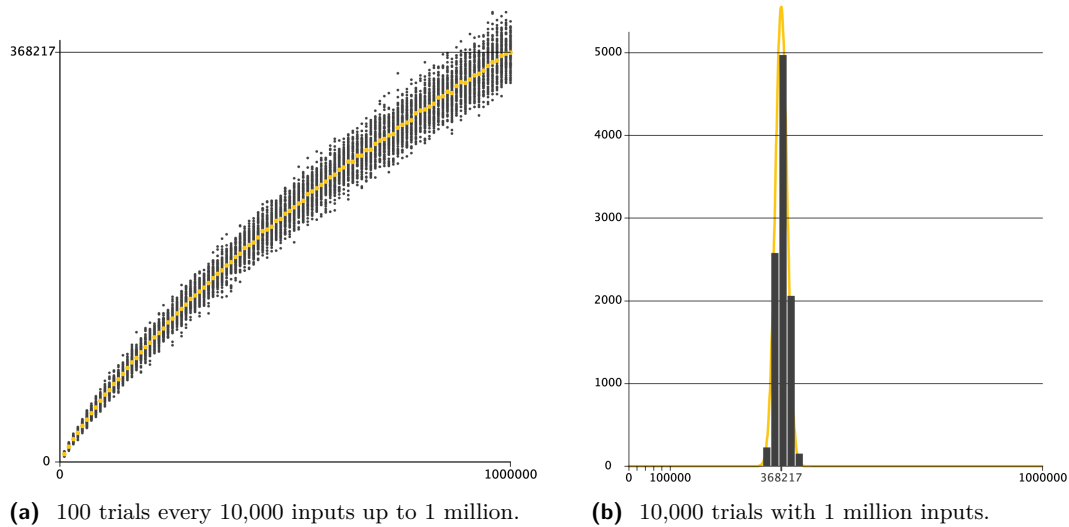
Figure 2 Results of estimating cardinalities in a web log, each with 10,000 trials. In Figure 2(a) `HyperBitT` was run 100 times for the first 10,000, 20,000, 30,000, ... items in the log, up to 1 million. Each grey dot shows the result of one experiment and the colored dots are the average of the values for each set of 100 experiments. A black line that shows the actual number of distinct items in the stream is completely hidden by the colored dots. The histogram in Figure 2(b) plots the estimates returned by `HyperBitT` for 10,000 runs on the first 1 million strings in the web log. The distribution matches a Gaussian, centered on the true number of distinct values, with relative standard deviation about $1.25/\sqrt{M} \doteq 0.039$ (plotted in color), thus validating Theorem 3 and our hypothesis that the estimated cardinality is likely to be within within 5% of the true value.

It is important to reiterate that `HyperBitT` is *not* a streaming algorithm. For example, it could not be used without some periodic adjustments for our web log example, where the log may be monitored for weeks, months, or even years, and therefore could consists of billions or trillions of strings or more. But there are many situations where `HyperBitT` may be useful because the estimate need not be very accurate and there are reasonable approaches to coming up with one. In a database or similar application, one might take a random sample. In a web log or similar application, one might take a small sample from initial values, or run multiple offsetting streams, using the estimate from one as the rough guess for another. For example, in protecting against a denial-of-service attack, the whole point might be to just set off an alarm when the cardinality deviates significantly from an expected range.

3 HyperBitBit and HyperBitBitBit

In this section, we describe variants of the algorithm that can *adapt* as the number of unique values grows, by making `T` a *variable* and then increasing it as needed.

Obviously, `T` needs to increase when the sketch becomes nearly full of `1`s. The first approach that comes to mind is to plan to increase `T` by one when the sketch becomes nearly full and to maintain a second sketch with `1` bits corresponding to whether or not an item with at least `T+1` trailing `1`s has been seen. Then, when the sketch is nearly full, we can increment `T` and replace the first sketch with the second one. But then we need to replace the second sketch. We could use a third sketch (and we will, when M is not small), but then do we need a fourth sketch? Moreover, when the sketch for `T` is nearly full of `1`s, so is the sketch for `T+1`, so incrementing `T` by 1 does not help much.

■ **Table 3** Fraction of zeros in sketches for `T+i` when the sketch for `T` is 97% full. The sketch for `T` is 3% 0s, the sketch for `T+4` is 80% 0s and the sketch for `T+8` is 99% 0s.

i	0	1	2	3	4	5	6	7	8
$\beta_i = \exp(-\ln(1/\beta)/2^i)$	.03	.17	.42	.64	**.80**	.90	.95	.97	**.99**

So we want to increment `T` by *more* than one. But by how much? Recall that our analysis indicates that the accuracy degrades as the number of 0s in the sketch grows, and incrementing `T` corresponds to increasing the number of 0s. Eventually we can stop when we encounter sketches that are all 0s, but we are faced with a delicate balance between the amount by which we increment `T` and the number of sketches we might need. Theorem 3 gives us precisely the information we need to make an intelligent choice.

To fix ideas, take $M = 64$ and suppose that we consider the sketch to be "nearly full" when 62 of its bits are 1 (and therefore $\beta = 2/64 \doteq 0.032$). Now, we want to choose an increment i for `T` – we will maintain a second sketch for `T+i` and increment `T` by i when the sketch for `T` is 97% full of `1`s. Our goal is to choose i such that we do not need to maintain a third sketch.

Let β_i be the fraction of 0s in the sketch for `T+i`. Because the estimated value of N does not change, we must have $\ln(1/\beta) = \ln(1/\beta_i)/2^i$. Solving for β_i gives $\beta_i = \exp(-\ln(1/\beta)/2^i)$. Table 3 shows these values for possible increments up to 8 (after that point, the sketches are increasingly likely to be all 0s).

Specifically, Table 3 tells us something very important: for increments 4 or greater, there is no need to maintain a third sketch, *because it would be nearly all zeros*. With our choice to increment `T` by 4 when the sketch is 97% 0s, we know that at that time the sketch for `T+4` is about 80% 0s and the sketch for `T+8` would be about 99% 0s, so we can increment `T`, update our sketch for `T` using the sketch for `T+4`, and set the sketch for `T+4` to all 0s. We may be ignoring a few `1`s that would be in the sketch for `T+8` had we maintained it, but the likelihood that ignoring them would noticeably affect the final estimate is very small. If we want to be very conservative, we could maintain the indices of these `1`s, at a very small (if not negligible) extra cost, but few practitioners would bother.

This discussion brings us to `HyperBitBit64` (Algorithm 2). It uses $M = 64$, maintains two sketches, increments `T` by 4, and updates the sketches when the first sketch becomes 97% full of `1`s. The implementation also illustrates how to use 64-bit words for the sketches, which eliminates the overhead of maintaining bit arrays and leads to very simple and efficient code in typical programming environments, even machine language. For clarity, Algorithm 2 uses the call `p(sketch)` to count the number of 1s in the sketch. If this is not available as an atomic operation, one might choose the alternative of counting as the bits are set, as described in Appendix B and illustrated in the code at `https://github.com/robert-sedgewick/hyperbitbit`.

From the above discussion, it is reasonable to hypothesize that when Algorithm 2 terminates, `sketch0` *is the same as the sketch when Algorithm 1 is used with the current value of* `T`. In other words, Theorem 3 applies throughout. As we saw in Table 3, just before incrementing `T`, `sketch0` has about 97% `1`s and `sketch1` has about 20% `1`s. Thus, the fraction of 0s in the sketches stays in the range $.03 < \beta < .80$, so the value of c_β is in the flat part of its curve (see Figure 1) – it is always less than 2.25 with average value about $\frac{1}{.77} \int_{.03}^{.80} c_\beta \mathrm{d}\beta \doteq 1.48$. This is conservative – the number of 0s quickly increases when it is small, so c_β is more often than not less than this average.

Algorithm 2 `HyperBitBit64`.

```
public static int estimateHBB64(Iterable<String> stream)
{
   int T = 1;
   int M = 64;
   long sketch0;
   long sketch1;
   for (String s : stream)
   {
      long x = hash1(s);                          // 64-bit hash
      int  k = hash2(s, M);                       // 6-bit hash
      if (r(x) > T)   sketch0 = sketch0 | 1L << k; // >T trailing 1s?
      if (r(x) > T+4) sketch1 = sketch1 | 1L << k;
      if (p(sketch0) > .97*M)                     // >62 1s?
      { sketch0 = sketch1; sketch1 = 0; T += 4; }
   }
   double beta = 1.0 - 1.0*p(sketch0)/M;          // fraction of 0s
   return (int) (Math.pow(2, T)*M*Math.log(1.0/beta));
}
```

The end result is that `HyperBitBit64` is a true streaming algorithm that uses just 128 bits (plus six bits for `T`) to achieve an expected standard error which is usually lower than $1.48/\sqrt{64} \doteq 18.5\%$ even for streams having billions or trillions or more distinct items. As we will see in Section 5, this accuracy is substantially better than that achieved by `HyperLogLog` for the same number of bits. The cost of processing each element is the cost of hashing plus a few machine-language instructions. In applications where 18.5% accuracy suffices (and developing a rough guess that would enable use of `HyperBitT` is infeasible), `HyperBitBit64` is likely to be the method of choice because of these low costs. For example, it would be quite useful in an application where maintaining large number of different cardinality counters are needed, each responding to some different filter of the input stream.

For larger values of M (say 128 or 256) we can implement `HyperBitBit` with a bit array (perhaps implemented with an array of 64-bit integers as described in Appendix B) and do even better. Specifically, it makes sense to *set the cutoff to increment `T` when the relative standard error for the new value is equal to the current relative standard error*. That is, with $a = \ln(1/\beta)$ and $c(a) = \sqrt{e^a - 1}/a$, we increment `T` by 4 when $c(a) = c(a/16)$. The solution to this equation is $a = \ln(1/\beta) \doteq 4.41$ so $\beta = e^{-a} \doteq .012$. That is, we should increment `T` by 4 and update the sketches when `sketch0` has $.988M$ 1 bits. At that point, the proportion of 0s in the sketch for `T+4` will be about $e^{-a/2^4} \doteq .759$. The proportion of 0s in the sketch for `T+8` would be about $e^{-a/2^8} \doteq .983$, so we are ignoring (2, 4, 9) 1 bits for (128, 256, 512) respectively, which is likely tolerable. The fraction of 0s in the sketches stays in the range $.012 < \beta < .759$, so the value of c_β is always less than 2.05 with average value about $\frac{1}{.747} \int_{.012}^{.759} c_\beta \mathrm{d}\beta \doteq 1.46$.

HyperBitBitBit

For even larger values of M, we can go to a third sketch, marking the subarrays with at least `T`, `T+4`, and `T+8` trailing 1s and define `HyperBitBitBit` in a straightforward manner. The implementation is omitted because we present a significant improvement in Section 4. The proportion of 0s in the sketch for `T+12` would be about $e^{-a/2^{12}} \doteq .996$, so we are ignoring (1, 2, 4) 1 bits for (1024, 2048, and 4096) respectively, again likely tolerable.

Algorithm 3 HyperTwoBits.

```
public static int estimateHTB(Iterable<String> stream, int M)
{  // for M = 1024, 2048, or 4096
   int T = 1;
   twobit[] sketch = new twobit[M];
   for (String s : stream)
   {
      long x = hash1(s);       // 64-bit hash
      int  k = hash2(s, M);    // (lg M)-bit hash
      if (r(x) >= T  ) if (sketch[k] < 1) sketch[k] = 1;
      if (r(x) >= T+4) if (sketch[k] < 2) sketch[k] = 2;
      if (r(x) >= T+8) if (sketch[k] < 3) sketch[k] = 3;
      if (pnz(sketch) > .988*M)
      {
         T = T+4;
         for (int i = 0; i < M; i++)
            if (sketch[i] > 0) sketch[i]--;
      }
   }
   double beta = 1.0 - 1.0*pnz(sketch)/M;
   return (long) (Math.pow(2, T)*M*Math.log(1/beta));
}
```

As just noted for `HyperBitBit`, the fraction of 0s in the sketches stays in the range $.012 < \beta < .759$, so the value c_β is always less than 2.05 with average value about $\frac{1}{.747}\int_{.012}^{.759} c_\beta \mathrm{d}\beta \doteq 1.46$. In summary, `HyperBitBitBit` is a true streaming algorithm, effective for M up to at least 4096, that uses $3M$ bits and achieves relative standard error of about $1.46/\sqrt{M}$.

4   HyperTwoBits

Remarkably, we can produce the same result as `HyperBitBitBit` but using just $2M$ bits. The trick is to note that if a bit is set in the sketch for `T+4`, the bit in the corresponding position in the sketch for `T` must be set, and if a bit is set in the sketch for `T+8`, the bits in the corresponding positions in the sketches for both `T+4` and `T` must be set. This observation means that we can represent the three sketches with an array of two-bit values that encode in binary the number of 1s in each position in the three sketches in `HyperBitBitBit`, as shown in this example:

	sketch for T	111111111110111011111111111111111111110111011111100111111111
before	*sketch for* T+4	000100111010000000000001000011001011000001110000100000000
	sketch for T+8	0000000100000000000000000000000000001000000001100001000000000
	two-bit values	111211232120111011111112111122112123101110223311003111111111
	sketch for T	000100111010000000000001000011001011000001110000100000000
after	*sketch for* T+4	0000000100000000000000000000000000001000000001100001000000000
T+=4	*sketch for* T+8	0000000000000000000000000000000000000000000000000000000000000
	two-bit values	000100121010000000000001010011001012000000112200002000000000

Maintaining this array while streaming is simple: for each data item, we identify its stream and set its value as appropriate. Then when the number of nonzero values reaches the threshold, we increment `T` by 4 and simply *decrement the nonzero values in the array.*

From this description, the implementation in Algorithm 3 is immediate. For clarity, we use a `twobit[]` type to describe the algorithm – although no programming languages support an explicit `twobit[]` type, the abstraction can be implemented with shifting and masking on arrays of integers, an amusing exercise in bit logic (see Appendix B). For clarity, we use a method `pnz()` to count the nonzero entries in the array – its implementation is omitted because it is better to maintain the count dynamically (also see Appendix B).

In summary, `HyperTwoBits` is a true streaming algorithm, effective for M up to at least 4096, that uses $2M$ bits and achieves relative standard error of about $1.46/\sqrt{M}$. As described in Appendix B, it can be implemented such that processing each item in a stream requires only a few machine-language operations.

Figure 3 presents the results of two experiments for Algorithm 3 corresponding to those presented for Algorithm 1 in Figure 2, which validate our hypothesis that the relative accuracies of the algorithms are comparable and are strong evidence of the utility of the algorithm in practice.
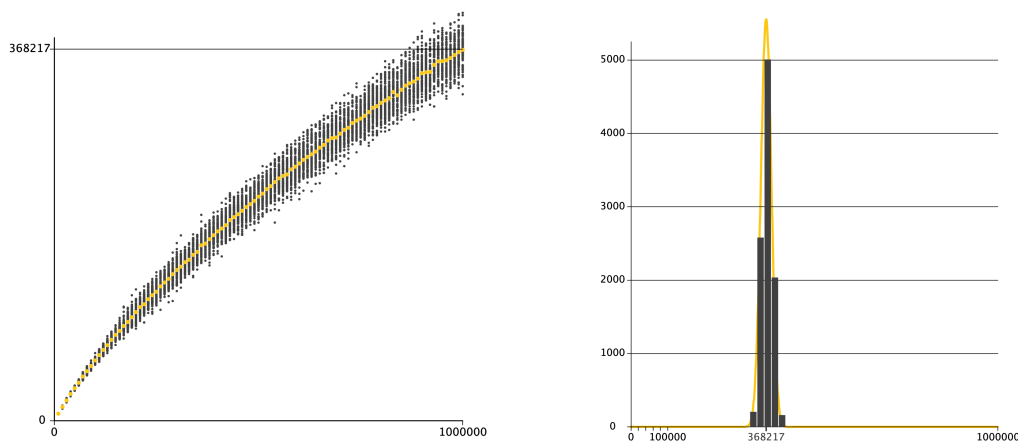
(a) 100 trials every 10,000 inputs up to 1 million.

(b) 10,000 trials with 1 million inputs.

Figure 3 Results of estimating cardinalities in a web log using Algorithm 3 with $M = 1000$, for comparison with Figure 2 (where the details of the experiments are described). Given the same inputs (and the same random numbers), the figures for `HyperBitBitBit` would be identical.

5 Performance comparisons

Comparing the performance of our algorithms with each other and with cardinality estimation algorithms in the literature needs to be done carefully for several reasons.

First, many papers from the theoretical computer science literature study algorithms implemented in pseudocode (or just described in English). While these papers often introduce interesting ideas, they cannot be evaluated as solutions to the practical cardinality estimation problem for two reasons. First, the methods described have never been implemented (and are sufficiently complicated that implementing them is not likely to be worthwhile) so the time required to process each item while streaming cannot be determined. Second, the analyses generally define complexity results that use O-notation and are not sufficiently precise to compare the relative accuracy with other methods.

Second, even among methods that have been implemented and tested, practitioners might prefer algorithms that are much simpler to implement and maintain over more complicated methods that perform slightly better. Some methods are sufficiently complicated to implement that practitioners might shy away from (or may not be able to afford) actually doing so. For example, `HyperLogLog` is easy to implement with 8-bit bytes, but 6-bit bytes are sufficient. Implementing a 6-bit byte array with arrays of 64-bit words is not difficult, but may be too cumbersome from the point of view of some practitioners.

Third, many papers use the parameter M to count the number of bytes or words (of varying length) of memory used. Proper comparisons necessitate counting *total number of bits* of memory in all cases. As an extreme example, suppose that two algorithms achieve standard error $2/\sqrt{M}$ but one uses M bits and the other uses M 64-bit words. The first is *eight times* more accurate for a given number of bits of memory. In general, if we know that the accuracy of an algorithm is $c/\sqrt{M}$ and that it stores Mb bits, we express the accuracy in terms of $M^\star$, the total number of bits used, or $c\sqrt{b}/\sqrt{M^\star}$. Inverting this equation gives the number of bits needed to achieve a given accuracy $x : M^\star = b(c/x)^2$. We ignore relatively inconsequential small fixed costs such as the six bits required to store the value of `T` in our adaptive algorithms.

Fourth, few papers actually *prove* anything about the distribution of the reported values, with the notable exception of [13]. Typically, normality is instead presented as a reasonable hypothesis, which may often be the case, but our proof of asymptotic normality of the reported cardinalities is significant.

Fifth, the accuracy of our algorithms depend on the coefficient c_β of $1/\sqrt{M}$ in the relative standard error, which varies. We use the average value of c_β over the interval of values β might take on during the execution of the algorithm. For `HyperBitT` we (somewhat arbitrarily) use the interval where $c_\beta < 1.5$; our other algorithms calculate an appropriate interval. As we have noted, the curve in Figure 1 is quite flat, so it is likely that the value encountered in practice is smaller than the value cited.

Sixth, it is important to remember that we are dealing with random fluctuations and approximate analyses. It may be tempting to use more precision, but any differences indicated would not be noticed in practice. For example, one might conclude that `HyperLogLog` with 6-bit bytes should be very slightly better than `LogLog` with 6-bit bytes because its standard error of $1.02/\sqrt{M}$ is very slightly better than $1.05/\sqrt{M}$, but it would be extremely challenging to develop experimental validation of that hypothesis.

With all these caveats, Table 4 presents a comparison of the algorithms we have discussed. Our simplest and perhaps most useful implementation is `HyperBitBit64`, which achieves 18.5% accuracy on a stream on any length with just two 64-bit words and can be implemented with a few dozen machine instructions. `HyperBitT` is the best by far when starting with a rough estimate is feasible. More generally, if a straightforward and easy to maintain implementation is desired, `HyperBitBit` and `HyperBitBitBit` are arguably simpler than the 8-bit version of `HyperLogLog` and substantially more efficient. If a careful implementation with improved efficiency is desired, `HyperTwoBits` is substantially more efficient than the 6-bit version of `HyperLogLog`. In both cases our algorithms provide much better accuracy for the same number of bits and use two-thirds as many bits to achieve the same accuracy.

6 Further Improvements

We conclude by briefly mentioning some opportunities that may lead to variants of our algorithms that may be worthy of study in various particular situations.

Table 4 Performance of cardinality estimation algorithms.

algorithm	range for M	b	c	$c\sqrt{b}$	$M^\star = b(c/x)^2$ bits needed for 2%	20%	$c\sqrt{b}/M^\star$ accuracy with 128 bits	8K bits
Adaptive sampling[5]		64	1.20	9.60	230400	2304	85%	10.6%
Prob. counting[6]		64	0.78	6.24	97344	973	55%	6.9%
LogLog[4]		6	1.05	2.57	16538	165	23%	3.5%
HyperLogLog8[7]		8	1.04	2.94	21632	216	26%	3.3%
HyperLogLog[7]		6	1.02	2.55	16224	162	23%	2.8%
ExtHyperLogLog[16]		7	0.88	2.33	13552	136	21%	2.6%
HyperBitT		1	1.32	1.32	4356	44	12%	1.5%
HyperBitBit64	64	2	1.48	2.09	—	128	19%	—
HyperBitBit	64–512	2	1.46	2.06	—	128	18%	—
HyperBitBitBit	128–4096	3	1.46	2.53	15987	128	22%	2.8%
HyperTwoBits	128–4096	2	1.46	2.06	10658	128	18%	2.3%

- *Sparse arrays.* Precise characterization of the "transition cost" just after incrementing `T` (when the sketches are mostly `0s`) may lead to slight performance improvements.
- *Use two sketches.* The second sketch contains information that may lead to a more accurate estimate. Analyzing this effect is tractable, but not likely to improve the accuracy by more than a percentage point or two.
- `HyperThreeBits`. Using 3-bit counters instead of the 2-bit counters in `HyperTwoBits` allows implementation of seven layers of bit arrays and may be useful for specialized applications needing very high accuracy (requiring huge values of M) for the kinds of truly huge streams seen in modern computing.
- `HyperBit`. We have studied many approaches to modifying `HyperBitT` to just increment `T`, reset the sketch to `0s`, and then characterizing the error due to the "transition cost". Despite some promising empirical results, the problem of developing a mathematical model admitting proper comparison of such an algorithm with the ones described here remains open.
- *Mergeability.* Many applications can benefit from being able to merge sketches built from two different streams. Our sketches are not difficult to merge, as indicated by the following argument for `HyperBitBit`. A sketch is a triple (`T, sketch0, sketch1`). To merge (`T_A, sketch0_A, sketch1_A`) with (`T_B, sketch0_B, sketch1_B`) consider the following three cases:
  - If $T_A = T_B = T$ use (`T, sketch0_A|sketch0_B, sketch1_A|sketch1_B`).
  - If the values of `T` differ by 8 or more, use the larger value and its sketches.
  - Otherwise, suppose wlog that $T_A = T_B + 4$. Use (`T_A, sketch0_A|sketch1_B, sketch1_A`).
  In the first and third cases, check whether the first sketch is nearly full. If so, increment `T` (by 4) and update the sketches as usual. This result is not precisely the same as if the two streams had actually been merged, but the difference is likely acceptably small in many practical situations. The argument for `HyperBitT` is similar, but simpler; the argument for `HyperBitBitBit` is similar, but more complicated.

──── **References** ────

1 Noga Alon, Yossi Matias, and Mario Szegedy. The space complexity of approximating the frequency moments. In Gary L. Miller, editor, *Proceedings of the Twenty-Eighth Annual ACM Symposium on the Theory of Computing, Philadelphia, Pennsylvania, USA, May 22-24, 1996*, pages 20–29. ACM, 1996. `doi:10.1145/237814.237823`.

2   A. D. Barbour, Lars Holst, and Svante Janson. *Poisson Approximation.* Oxford University Press, 1992.

3   Kai-Min Chung, Michael Mitzenmacher, and Salil P. Vadhan. Why simple hash functions work: Exploiting the entropy in a data stream. *Theory Comput.*, 9:897–945, 2013. `doi:10.4086/TOC.2013.V009A030`.

4   Marianne Durand and Philippe Flajolet. Loglog counting of large cardinalities (extended abstract). In Giuseppe Di Battista and Uri Zwick, editors, *Algorithms - ESA 2003, 11th Annual European Symposium, Budapest, Hungary, September 16-19, 2003, Proceedings*, volume 2832 of *Lecture Notes in Computer Science*, pages 605–617. Springer, 2003. `doi:10.1007/978-3-540-39658-1_55`.

5   Philippe Flajolet and G. Nigel Martin. Probabilistic counting. In *24th Annual Symposium on Foundations of Computer Science, Tucson, Arizona, USA, 7-9 November 1983*, pages 76–82. IEEE Computer Society, 1983. `doi:10.1109/SFCS.1983.46`.

6   Philippe Flajolet and G. Nigel Martin. Probabilistic counting algorithms for data base applications. *J. Comput. Syst. Sci.*, 31(2):182–209, 1985. `doi:10.1016/0022-0000(85)90041-8`.

7   Philippe Flajolet, Éric Fusy, Olivier Gandouet, and Frédéric Meunier. Hyperloglog: the analysis of a near-optimal cardinality estimation algorithm. In Philippe Jacquet, editor, *AofA 07— Proceedings of the 2007 Conference on Analysis of Algorithms, Juan-les-pins, France, June 18-22, 2007, DMTCS Proceedings volume AH*, pages 127–146. DMTCS, 2007. URL: `https://algo.inria.fr/flajolet/Publications/FlFuGaMe07.pdf`.

8   Allan Gut. *Probability: A Graduate Course (2nd edition).* Springer Texts in Statistics, 75, 2013.

9   Stefan Heule, Marc Nunkesser, and Alexander Hall. Hyperloglog in practice: algorithmic engineering of a state of the art cardinality estimation algorithm. In Giovanna Guerrini and Norman W. Paton, editors, *Joint 2013 EDBT/ICDT Conferences, EDBT '13 Proceedings, Genoa, Italy, March 18-22, 2013*, pages 683–692. ACM, 2013. `doi:10.1145/2452376.2452456`.

10   Piotr Indyk and David P. Woodruff. Tight lower bounds for the distinct elements problem. In *44th Symposium on Foundations of Computer Science (FOCS 2003), 11-14 October 2003, Cambridge, MA, USA, Proceedings*, pages 283–288. IEEE Computer Society, 2003. `doi:10.1109/SFCS.2003.1238202`.

11   Daniel M. Kane, Jelani Nelson, and David P. Woodruff. An optimal algorithm for the distinct elements problem. In Jan Paredaens and Dirk Van Gucht, editors, *Proceedings of the Twenty-Ninth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems, PODS 2010, June 6-11, 2010, Indianapolis, Indiana, USA*, pages 41–52. ACM, 2010. `doi:10.1145/1807085.1807094`.

12   Matti Karppa and Rasmus Pagh. Hyperlogloglog: cardinality estimation with one log more. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 753–761, 2022.

13   Jérémie Lumbroso. An optimal cardinality estimation algorithm based on order statistics and its full analysis. *Discrete Mathematics & Theoretical Computer Science*, AM, 2010. `doi:10.46298/dmtcs.2780`.

14   Jérémie Lumbroso. How Flajolet processed streams with coin flips. *CoRR*, abs/1805.00612, 2018. `arXiv:1805.00612`.

15   Jérémie Lumbroso and Conrado Martínez. Affirmative Sampling: Theory and Applications. In Mark Daniel Ward, editor, *33rd International Conference on Probabilistic, Combinatorial and Asymptotic Methods for the Analysis of Algorithms (AofA 2022)*, volume 225 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 12:1–12:17, Dagstuhl, Germany, 2022. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. `doi:10.4230/LIPIcs.AofA.2022.12`.

16   Tal Ohayon. Extendedhyperloglog: Analysis of a new cardinality estimator. *CoRR*, abs/2106.06525, 2021. `arXiv:2106.06525`.

17 Seth Pettie and Dingyu Wang. Information theoretic limits of cardinality estimation: Fisher meets shannon. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, pages 556–569, 2021.

18 Robert Sedgewick and Philippe Flajolet. *An Introduction to the Analysis of Algorithms, Second Edition.* Addison-Wesley-Longman, 2013.

19 Dingyu Wang and Seth Pettie. Better cardinality estimators for hyperloglog, pcsa, and beyond. In Floris Geerts, Hung Q. Ngo, and Stavros Sintos, editors, *Proceedings of the 42nd ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems, PODS 2023, Seattle, WA, USA, June 18-23, 2023*, pages 317–327. ACM, 2023. `doi:10.1145/3584372.3588680`.

A   Proof of Lemma 1

Suppose that $X_n \geq 0$ are random variables and a_n, b_n, and σ^2 numbers such that, as $n \to \infty$, we have $a_n \to a > 0$, $b_n \to 0$, and $(X_n - a_n)/b_n \xrightarrow{d} \mathbb{N}(0, \sigma^2)$. If f is a continuously differentiable function on $(0, \infty)$ with $f'(a) \neq 0$, then $(f(X_n) - f(a_n))/b_n \xrightarrow{d} \mathbb{N}(0, f'(a)^2 \sigma^2)$.

Proof. This is well known, but we include this proof for completeness.

By the mean value theorem,

$$\frac{f(X_n) - f(a_n)}{b_n} = f'(X_n^*)\frac{X_n - a_n}{b_n} \tag{7}$$

for some X_n^* with $X_n \leq X_n^* \leq a_n$ or $a_n \leq X_n^* \leq X_n$. Since $(X_n - a_n)/b_n \xrightarrow{d} \mathbb{N}(0, \sigma^2)$ and $b_n \to 0$, we have $X_n - a_n \xrightarrow{p} 0$. Furthermore, $a_n \to a$, and hence $X_n \xrightarrow{p} a$. Consequently, also $X_n^* \xrightarrow{p} a$. Thus, since f' is continuous, $f'(X_n^*) \xrightarrow{p} f'(a)$. The result follows from (7) and the assumption.   ◄

B   Implementation details

The abstract operations we have used in expressing our algorithms can be implemented efficiently on most computers, as described in the following paragraphs. Our code makes liberal use of Java's left and right shift operators `<<` and `>>` and bitwise logical operations (`&`, `|`, and `~`) for bitwise (*AND*, *OR*, and *NOT*) respectively. Algorithm 4 is a full low-level implementation of `HyperBitBit64` that solves the practical cardinality estimation problem.

Sketches

As we have noted, few programming languages support an efficient `bit[]` type (even Java does not guarantee that boolean arrays use one bit per entry). As we saw in `HyperBitBit64` (Algorithm 2), shifting and masking on 64-bit long values is an easy way to implement the abstraction. For larger values of `M`, we use arrays of 64-bit values. In Java, for example, we maintain the sketch as an array of `long` values:

```
long[] sketch = new long [M/64];
```

Then the Java code

```
if ((sketch[k/64] & (1L << (k % 64))) != 0)
```

tests whether the `k`th bit in the sketch is `1` and the Java code

```
sketch[k/64] = sketch[k/64] | (1L << (k % 64));
```

sets the `k`th bit in the sketch to `1`.

Trailing 1s

The key abstract operation in our implementations involves computing the function `r(x)`, so that we can test whether a 64-bit value `x` has at least `T` trailing 1s. Rather than maintaining the parameter T, we maintain $U = 2^T$. The reason for doing so is that the value `U-1` has `T` trailing 1s, which enables us to test whether a value `x` has at least `T` trailing ones with the bitwise logical operation `(x & (U-1)) == (U-1)`, which is easy to implement with a few machine-language instructions.

Population count

The second abstract operation in our implementations is the function `p(x)`, the so-called "population count" – the number of `1` bits in a binary value. This function has a long and interesting history, but, for our purposes, it is easy to avoid, by maintaining a count of the number of `1` bits in the sketches, incrementing when each bit is set.

Two-bit counters

Again, we use shifting and masking on arrays of 64-bit `long` values. We keep one `long` array `s1` for the more significant bit and a second `long` array `s0` for the less significant bit. To make the code more readable, we define the following methods to test and set the bit corresponding to bit `k`:

```
public static long val(long[] s1, long[] s0, int k)
{ return 2*((s1[k/64] >> (k % 64)) & 1L)+((s0[k/64] >> (k % 64)) & 1L); }
public static void setval(long[] s1, long[] s0, int k, long v)
{
    s1[k/64] = (s1[k/64] & ~(1L << (k % 64))) | ((v/2) & 1L) << (k % 64);
    s0[k/64] = (s0[k/64] & ~(1L << (k % 64))) | (v & 1L) << (k % 64);
}
```

In a tightly efficient or machine-code version, this code would be used inline.

The final abstract operation to consider is to decrement all the non-zero counters. Consider the following table, which gives all possibilities for a given bit position, where s_1s_0 is the value before incrementing and t_1t_0 is the value after decrementing.

	before			after		
value	s1	s0	*value*	t1	t0
0	0	0	0	0	0
1	0	1	0	0	0
2	1	0	1	0	1
3	1	1	2	1	0

Considering these as truth tables on boolean values, it is easy to check that `t1 = s1` *AND* `s0` and `t0 = s1` *AND NOT* `s0`. Furthermore, we can eliminate the temporary variables by doing the operations in the order `s0 = s1` *AND NOT* `s0` and then `s1 = s1` *AND NOT* `s0`. Implementing these operations with bitwise operations on our arrays of `long` values is straightforward.

■ **Algorithm 4** `HyperTwoBits` (full low-level implementation).

```
public static int estimateHTB (String [] stream , int N, int M)
{
   int U = 2;
   double alpha = .988;
   long[] s0 = new long [M/64];
   long[] s1 = new long [M/64];
   int count = 0;
   for (int i = 0; i < N; i++)
   {
      long x = hash1(s);        // 64-bit hash
      int  k = hash2(s, M);     // (lg M)-bit hash
      if ((x & (U-1)) == (U-1)) count ++;
      if ((x & (U-1)) == (U-1))
         if (val(s1, s0, k) < 1) setval(s1, s0, k, 1);
      if ((x & (16*U-1)) == (16*U-1))
         if (val(s1, s0, k) < 2) setval(s1, s0, k, 2);
      if ((x & (256*U-1)) == (256*U-1))
         if (val(s1, s0, k) < 3) setval(s1, s0, k, 3);
      if (count >= alpha*M)
      {
         for (int j = 0; j < M/64; j++)
         { s0[j] = s1[j] & ~s0[j]; s1[j] = s1[j] & ~s0[j]; }
         count = 0;
         for ( int j = 0; j < M; j++)
           if (val(s1, s0, j) > 0) count ++;
         U = 16*U;
      }
   }
   double beta = 1.0 - 1.0*count/M;
   double bias = Math.log(1.0/beta);
   return (int) (U*M*bias);
}
```

Phase Transition for Tree-Rooted Maps

Marie Albenque ✉
IRIF, Université Paris Cité, France

Éric Fusy ✉
Univ Gustave Eiffel, CNRS, LIGM, F-77454 Marne-la-Vallée, France

Zéphyr Salvy ✉
Univ Gustave Eiffel, CNRS, LIGM, F-77454 Marne-la-Vallée, France

───── **Abstract** ─────

We introduce a model of tree-rooted planar maps weighted by their number of 2-connected blocks. We study its enumerative properties and prove that it undergoes a phase transition. We give the distribution of the size of the largest 2-connected blocks in the three regimes (subcritical, critical and supercritical) and further establish that the scaling limit is the Brownian Continuum Random Tree in the critical and supercritical regimes, with respective rescalings $\sqrt{n/\log(n)}$ and $\sqrt{n}$.

1 Introduction

A *planar map* $\mathfrak{m}$ is the proper embedding into the two-dimensional sphere of a connected planar finite multigraph, considered up to homeomorphisms. In recent years, models of random planar maps with weighted 2-connected blocks [5, 23] have been introduced. In particular, the model with a Boltzmann weight u per block exhibits a phase transition at $u_C = 9/5$, with a "tree phase" for $u > u_C$ with only small blocks and having the Brownian Continuum Random Tree (CRT) as scaling limit [23], a "map-phase" for $u < u_C$ characterized by the existence of a giant block and having the Brownian sphere as scaling limit, and with mesoscopic blocks and the stable tree of parameter $3/2$ as scaling limit at the critical point u_C [14].

Here, we study such a model in the context of decorated maps and consider the emblematic case of *tree-rooted maps*, *i.e.*, maps endowed with a spanning tree. In theoretical physics, decorated maps are instrumental to provide models of two-dimensional quantum gravity coupled with matter. They lead to new asymptotic behaviours, and the study of scaling limits in that context is currently a very challenging topic in random maps [16]. Among decorated maps, tree-rooted maps have very rich combinatorial properties and their enumeration goes back to Mullin [22], who obtained the formula

$$m_n = \sum_{k=0}^{n} \binom{2n}{2k} \mathrm{Cat}_k \mathrm{Cat}_{n-k} = \mathrm{Cat}_n \mathrm{Cat}_{n+1} \tag{1}$$

for the number of tree-rooted maps with n edges, by observing that a tree-rooted map is a shuffle of two plane trees (the spanning tree and its dual). A direct bijective proof that $m_n = \text{Cat}_n\text{Cat}_{n+1}$ was later obtained by Bernardi [2], who subsequently extended his bijection to maps endowed with a Potts model [3].

Our contributions here are both enumerative and probabilistic. First, we show an asymptotic estimate for 2-connected tree-rooted maps, and that the enumeration of tree-rooted maps with a weight u per 2-connected block undergoes a phase transition at an explicit (transcendental) value u_C. On the probabilistic side, we obtain limit laws for the sizes of the largest blocks, with the existence of a giant block if and only if $u < u_C$. Furthermore, we show that the scaling limit is the CRT for all $u \geq u_C$, with a discontinuity at u_C for the order of magnitude of the rescaling, which is $\sqrt{n/\log n}$ at u_C whereas it is $\sqrt{n}$ for $u > u_C$ (the scaling limit result for $u > u_C$ also follows from [23, Th 6.63], as further commented in Section 4.3). Finally, we discuss possible extensions in the concluding section.

2 Block-decomposition of tree-rooted maps

Let $\mathfrak{m}$ be a planar map. We denote, respectively, by $E(\mathfrak{m})$, $V(\mathfrak{m})$ and $F(\mathfrak{m})$ its sets of edges, vertices and faces. Any edge is made of two half-edges (which meet at the middle of the edge). All the maps considered in this paper will be *rooted*, meaning that one of their half-edges is distinguished (and is represented by an oriented edge on figures), and a rooted planar map will be simply called a *map* from now on. The *size* of a map $\mathfrak{m}$ – denoted by $|\mathfrak{m}|$ – is defined as its number of edges.

A map $\mathfrak{m}$ is said to be *separable* if $E(\mathfrak{m})$ can be partitioned into two non-empty subsets E_1 and E_2 such that there exists exactly one vertex – called *cut vertex* – incident to both an element of E_1 and an element of E_2. It is is said to be 2-*connected* otherwise. By convention, the vertex map (*i.e.*, the map reduced to a single vertex) is considered to be 2-connected. A *block* of $\mathfrak{m}$ is a maximal 2-connected submap *of positive size*. The number of blocks of $\mathfrak{m}$ is denoted by $b(\mathfrak{m})$, so that if $|\mathfrak{m}| > 0$, then $b(\mathfrak{m}) = 1$ if and only if $\mathfrak{m}$ is 2-connected.

Fix $\mathfrak{m}$ a map, and let τ be one of its spanning trees, then one calls $(\mathfrak{m}, \tau)$ a tree-rooted map. We denote, respectively, by $\mathcal{M}$ and $\mathcal{B}$ the set of tree-rooted maps and of 2-connected tree-rooted maps, by $\mathcal{M}_n$ and $\mathcal{B}_n$ the subset of $\mathcal{M}$ and $\mathcal{B}$ restricted to elements of size n, and by $M(z)$ and $B(y)$ the associated generating series. In the following, a tree-rooted map will be denoted by $\mathfrak{m}$ instead of being explicitly written as a pair, and, for $\mathfrak{m} \in \mathcal{M}$, we write $\tau(\mathfrak{m})$ for its distinguished spanning tree.

To enumerate 2-connected maps, Tutte [25] formulated a decomposition to relate the generating series of maps and of 2-connected maps, as follows. Fix $\mathfrak{m}$ a map and let $\mathfrak{b}$ be the block containing its root. For each half-edge e of $\mathfrak{b}$ incident to a vertex u, let c be the corner of $\mathfrak{b}$ incident to u and following e in counterclockwise order around u. The *pendant submap* $\mathfrak{m}_e$ of e is defined as the maximal submap of $\mathfrak{m}$ disjoint from $\mathfrak{b}$ except at u, and located in the area of c. Unless $\mathfrak{m}_e$ is reduced to the vertex map, its root is defined at the half-edge following e in counterclockwise order around u in $\mathfrak{m}$. From $\mathfrak{b}$ and the collection of pendant submaps $\{\mathfrak{m}_e\}$, we can bijectively reconstruct $\mathfrak{m}$.

This decomposition extends readily to tree-rooted maps as follows. Fix $\mathfrak{m} \in \mathcal{M}$. Consider $(\mathfrak{b}, \tau_\mathfrak{b})$, where $\mathfrak{b}$ is, as before, the block of $\mathfrak{m}$ containing its root and $\tau_\mathfrak{b} = \tau(\mathfrak{m}) \cap \mathfrak{b}$. We claim that $\tau_\mathfrak{b}$ is a spanning tree of $\mathfrak{b}$: clearly, $\tau_\mathfrak{b}$ is acyclic since $\tau(\mathfrak{m})$ is. Then, for any $u, v \in \mathfrak{b}$, since $\mathfrak{b}$ is 2-connected, any path between u and v that is not included in $\mathfrak{b}$ has to visit the same cut vertex at least twice, and in particular is not simple. Any simple path between u and v in $\mathfrak{m}$ is then included in $\mathfrak{b}$, and so is the unique simple path between u and v in $\tau(\mathfrak{m})$. This proves
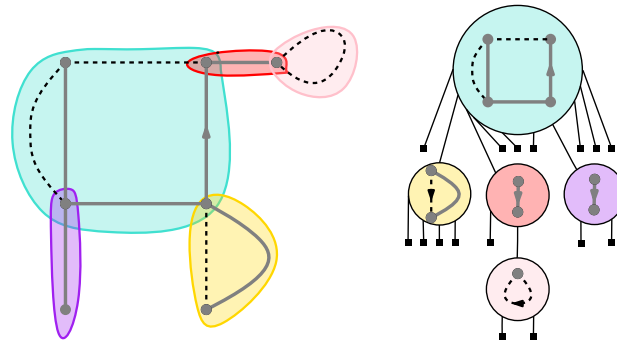
Figure 1 Block tree corresponding to a tree-rooted planar map. Full grey (resp. dashed black) edges represent edges that are part (resp. not part) of the decorating spanning tree.

that $\tau_{\mathfrak{b}}$ is connected. The same reasoning can be applied to all the pendant submaps $\mathfrak{m}_e$, to get a similar decomposition in the tree-rooted case, which induces the following identity of generating series.

▶ **Proposition 1.** *The generating series satisfy the following equality:*

$$M(z) = B\left(zM(z)^2\right). \tag{2}$$

Moreover, this equation can be refined to account for the number of blocks in a tree-rooted map. Writing $M(z,u) = \sum_{\mathfrak{m}\in\mathcal{M}} z^{|\mathfrak{m}|}u^{b(\mathfrak{m})}$, one has:

$$M(z,u) = uB(zM(z,u)^2) + 1 - u. \tag{3}$$

Note that these relations are exactly the same as the ones obtained in the non–tree-rooted case [25, 14].

Tutte's decomposition can also be applied recursively, by considering first the root block and then applying the block decomposition to each of the pendant submaps. This can be encoded by a decomposition tree $T_{\mathfrak{m}}$, which was first explicitly described by Addario-Berry in the non–tree-rooted case [1, §2], but which can also be extended to the tree-rooted case; see Figure 1.

▶ **Proposition 2.** *The block tree $T_{\mathfrak{m}}$ of a tree-rooted map $\mathfrak{m}$ satisfies the following properties:*
- *Edges of $T_{\mathfrak{m}}$ correspond to half-edges of $\mathfrak{m}$;*
- *Internal nodes of $T_{\mathfrak{m}}$ correspond to blocks of $\mathfrak{m}$: if an internal node v of $T_{\mathfrak{m}}$ has r children, then the corresponding block $\mathfrak{b}_v$ of $\mathfrak{m}$ has size $r/2$;*
- *The map $\mathfrak{m}$ is entirely determined by $\left(T_{\mathfrak{m}}, (\mathfrak{b}_v, v \in T_{\mathfrak{m}})\right)$ where $\mathfrak{b}_v$ is the block of $\mathfrak{m}$ represented by v in $T_{\mathfrak{m}}$ if v is an internal node and is the vertex map otherwise.*

3 Asymptotic enumeration

3.1 Asymptotic enumeration of 2-connected tree-rooted maps

We obtain here an asymptotic estimate for the number $b_n := [y^n]B(y)$ of 2-connected tree-rooted maps of size n. The steps are as follows: we first lift (Lemma 3) the asymptotic estimate $m_n \sim \frac{4}{\pi n^3}16^n$ for tree-rooted maps to a singular expansion for the generating function $M(z)$. Then via (2), we get in Proposition 4 the radius of convergence ρ_B and the singular expansion of $B(y)$ around ρ_B. In order to transfer the singular expansion to an asymptotic estimate for b_n, we also show that ρ_B is the unique dominant singularity of $B(y)$, using a combinatorial argument.

▶ **Lemma 3.** *When $z \to \rho_M = \frac{1}{16}$ in $\mathbb{C} \setminus \{z \geq \rho_M\}$, one has, with $Z = 1 - 16z$,*

$$M(z) = 8 - \frac{64}{3\pi} - 8\left(\frac{10}{3\pi} - 1\right) Z - \frac{2}{\pi} \ln(Z) Z^2 + O\left(Z^2\right). \tag{4}$$

Proof. The explicit expression of m_n translates to a D-finite equation satisfied by $M(z)$:

$$M'''(z) + \frac{6}{z} M''(z) + \frac{6(18z - 1)}{z^2(16z - 1)} M'(z) + \frac{12}{z^2(16z - 1)} M(z) = 0. \tag{5}$$

D-finite equation theory [13, Sec. VII.9.1, p. 518] gives that the finite singularities of a solution $f(z)$ of (5) are among the zeroes of the denominators of the coefficients: $\mathcal{S} = \{0, 1/16\}$; and any solution of (5) is analytically continuable along any path avoiding $\mathcal{S}$. In particular, the solution $M(z) = \sum_{n \geq 0} \text{Cat}_n \text{Cat}_{n+1} z^n$, which is clearly analytic at 0, is continuable to the whole complex plane slit by the half-line $z \geq 1/16$.

Moreover, $1/16$ is a so-called *regular* singularity; and, using the `DEtools` package of the `Maple` computer algebra software, one can compute singular expansions for a basis of solutions of (5). The singular expansion of $M(z)$ is then a linear combination of the basis' singular expansions, which gives:

$$M(z) = \sum_{k=0}^{\infty} a_k Z^k - \ln(Z) \sum_{k=2}^{\infty} b_k Z^k, \quad \text{with} \quad Z = 1 - 16z,$$

holding in a slit neighborhood of $1/16$. From the explicit expression $[z^n]M(z) = \text{Cat}_n \text{Cat}_{n+1}$ it follows that $a_0 = M(1/16) = 8 - \frac{64}{3\pi}$ and $a_1 = M'(1/16)/(-16) = -8\left(\frac{10}{3\pi} - 1\right)$. By Pringsheim's theorem, $M(z)$ is singular at its radius of convergence $1/16$ so there exists a smallest integer $k \geq 2$ such that $b_k \neq 0$. By applying transfer theorems [13, Chap. VI], one has

$$[z^n]M(z) \sim (-1)^k \frac{b_k k!}{n^{k+1}} 16^n.$$

Since $\text{Cat}_n \text{Cat}_{n+1} \sim \frac{4}{\pi n^3} 16^n$, one must have $k = 2$ and $b_2 = \frac{2}{\pi}$, which concludes the proof.   ◀

▶ **Proposition 4.** *The radius of convergence of $B(y)$ is*

$$\rho_B := \rho_M M^2(\rho_M) = \frac{4(3\pi - 8)^2}{9\pi^2} \approx 0.091, \tag{6}$$

and, when $y \to \rho_B$ in a Δ-neighbourhood of ρ_B, one has, with $Y = 1 - y/\rho_B$,

$$B(y) = 8 - \frac{64}{3\pi} - \frac{2\pi(10 - 3\pi)}{(3\pi - 8)(4 - \pi)} Y - \frac{2(3\pi - 8)^3}{27\pi(4 - \pi)^3} \ln(Y) Y^2 + O(Y^2). \tag{7}$$

Moreover, ρ_B is the unique dominant singularity of $B(y)$.

▶ **Remark 5.** The generating series $B(y)$ is not D-finite (having a transcendental radius of convergence), but from Equations (2) and (5) it is D-algebraic.

Proof of Proposition 4. Let $H(z) = zM(z)^2$, so that one has $M(z) = B(y)$, where $y = H(z)$. Note that $H(z)$ has radius of convergence $1/16$, and it inherits from $M(z)$ a singular expansion of the form (with $Z = 1 - 16z$):

$$H(z) = \tau - \kappa Z + \xi \ln(Z) Z^2 + O(Z^2)$$

for τ, κ, ξ explicit (*e.g.* $\tau = (8 - \frac{64}{3\pi})^2/16$).

The function H is analytic on $D(0, 1/16)$ and since $H'(z) > 0$ for any $z \in [0, 1/16)$, one can apply the analytic local inversion theorem at any such value of z. Moreover, H maps the interval $[0, 1/16]$ to the interval $[0, \tau]$, so one can define a functional inverse g of H on a neighborhood of $[0, \tau)$, which is analytic on this domain. Furthermore, H' is continuous in a Δ-neighbourhood of $1/16$, with positive value at $1/16$, hence $H(z)$ is injective on a Δ-neighbourhood U of $1/16$, and maps U to an open region containing a Δ-neighbourhood V of τ.

Using bootstrapping, the singular expansion of $g(y)$ at τ (valid in V) is easily obtained from the singular expansion of $H(z)$. With $Y = 1 - y/\tau$, one gets

$$z = g(y) = \frac{1}{16} - \frac{\tau}{16\kappa} Y - \frac{\tau^2}{16\kappa^3} \ln(Y) Y^2 + O(Y^2).$$

With $Z = 1 - 16z$, this gives

$$Z = \frac{\tau}{\kappa} Y + \frac{\tau^2}{\kappa^3} \ln(Y) Y^2 + O(Y^2).$$

Then, $B(y) = M(g(y))$ is analytic at every point in $[0, \tau)$, and the claimed singular expansion of $B(y)$ at $\rho_B := \tau$ is obtained by composing the singular expansion of $M(z)$ with the singular expansion of $g(y)$, *i.e.*, injecting the above expansion of Z into the expansion in Lemma 3. By Pringsheim's theorem, ρ_B is the radius of convergence of $B(y)$.

It remains to prove that ρ_B is the unique dominant singularity of $B(y)$. To do so, we use the trick of writing $B(y)$ as a supercritical composition scheme (in the sense of [15]), which we achieve thanks to a decomposition into series-parallel components. Doing so, we prove in Lemma 6 that $B(y)$ can be written as

$$B(y) = 1 + 2y + 2yA(y) + yA'(y)Q(A(y)), \tag{8}$$

for some generating functions $A(y)$ and $Q(w)$ with nonnegative coefficients, such that $A(y)$ is non-periodic and has radius of convergence larger than ρ_B. This implies that the radius of convergence of $Q(w)$ is $A(\rho_B)$. Moreover, by the Daffodil Lemma, see [13, Lem. IV.1, p. 266], for any $y \neq \rho_B$ such that $|y| = \rho_B$, we have that $|A(y)| < A(\rho_B)$. Hence, $A(y)$ belongs to the disk of convergence of Q and y cannot be a singularity, which concludes the proof. ◀

Let $Q(w)$ be the generating function of 2-connected tree-rooted maps with no face of degree 2 nor vertex of degree 2, with w counting the number of non-root edges. A 2-connected map with at least 2 edges is called *series-parallel* if it has no K_4 minor. Let $A(y)$ (resp. $\overline{A}(y)$) be the generating function of 2-connected tree-rooted series-parallel maps such that the root-edge is not (resp. is) in the spanning tree, the variable y counting the number of non-root edges.

▶ **Lemma 6.** *The generating series $B(y)$, $A(y)$ and $Q(w)$ satisfy the identity* (8). *Moreover, the radius ρ_A of convergence of A satisfies $\rho_A = 2 - 3 \cdot 2^{-2/3} \approx 0.11 > \rho_B$.*

Proof. A *series-parallel network* N is obtained by deleting the root-edge e of a series-parallel map, the two extremities of e being called the *poles* of N (which are distinguished as the *source*, the origin of e, and the *sink*, the end of e). Note that $A(y)$ is also the generating function of series-parallel networks endowed with a spanning tree, while $\overline{A}(y)$ is the generating function of series-parallel networks endowed with a spanning forest made of two trees containing each of the two poles. These two cases are respectively called *crossing* and *non-crossing*.

The *core* $\mathfrak{c}$ of a 2-connected tree-rooted map $\mathfrak{b}$ of size $|\mathfrak{b}| \geq 2$ is obtained by repeatedly collapsing faces of degree 2 and erasing vertices of degree 2 (turning the two incident edges into a single edge). This process is actually well-behaved only if $\mathfrak{b}$ is not series-parallel (otherwise it ends at a loop-edge with no vertex). Conversely, a 2-connected map $\mathfrak{b}$ is obtained from its core $\mathfrak{c}$ where every edge is replaced by a series-parallel network. By convention, the root-edge of $\mathfrak{c}$ is chosen as the one bearing the series-parallel network containing the root-edge of $\mathfrak{b}$. If $\mathfrak{b}$ is endowed with a spanning tree $\tau := \tau(\mathfrak{b})$, then for each edge e of $\mathfrak{c}$, letting N_e be the associated series-parallel network, on N_e the tree τ induces either a spanning tree, or a spanning forest with two trees containing each of the two poles. In the first case, e is declared a tree-edge of $\mathfrak{c}$, and thus τ induces a spanning tree on $\mathfrak{c}$. In terms of generating functions, $Q(w)$ is the counting series for the core, each non-root edge of the core then contributing either $A(y)$ if a tree-edge (crossing case) or contributing $\overline{A}(y)$ otherwise (non-crossing case). Since, by duality we have $A(y) = \overline{A}(y)$, every non-root edge turns out to have the same contribution $A(y)$. On the other hand, the root-edge of the core contributes $A'(y)$ because of the choice of the root-edge. This yields the claimed equation (8), where the added term $2yA(y)$ accounts for 2-connected tree-rooted series-parallel maps.

Now, to get the statement about ρ_A, it is well-known that a series-parallel network is either reduced to a single edge, or made of at least two series-parallel networks connected in series, or made of at least two series-parallel networks connected in parallel. The series-parallel decomposition then yields the following equation-system:

$$A(y) = y + S(y) + P(y), \qquad\qquad \overline{A}(y) = y + \overline{S}(y) + \overline{P}(y),$$

$$S(y) = \frac{(y + P(y))^2}{1 - y - P(y)}, \qquad\qquad \overline{S}(y) = (y + \overline{P}(y))\left(\frac{1}{(1 - y - P(y))^2} - 1\right),$$

$$P(y) = (y + S(y))\left(\frac{1}{(1 - y - \overline{S}(y))^2} - 1\right), \qquad \overline{P}(y) = \frac{(y + \overline{S}(y))^2}{1 - y - \overline{S}(y)}.$$

By symmetry, one has $S(y) = \overline{P}(y)$ and $P(y) = \overline{S}(y)$, which yields $A(y) = \overline{A}(y)$ (this is also clear by duality). Hence, the function $A(y)$ is algebraic and satisfies[1]

$$A(y)^3 + (y + 1)\, A(y)^2 + (2y - 1)\, A(y) + y = 0$$

and its radius of convergence is the smallest positive root of $4y^3 - 24y^2 + 48y - 5$ which gives $\rho_A = 2 - 3 \cdot 2^{-2/3} \approx 0.11$ and is larger than ρ_B.   ◀

From Proposition 4, by applying transfer theorems [13, Chap.VI] one directly obtains:

▶ **Corollary 7.** *When* $n \to \infty$,

$$b_n \sim \frac{4\,(3\pi - 8)^3}{27\pi\,(4 - \pi)^3} \cdot \rho_B^{-n} \cdot n^{-3}. \tag{9}$$

3.2 Enumerative phase transition for block-weighted tree-rooted maps

In this section, we investigate the singular expansion of $z \mapsto M(z, u)$ around its radius of convergence. We prove that this expansion exhibits three possible behaviours depending on the value of u.

---

[1] The series $A(y)$ is represented in Sloane's OEIS by the sequence A121873, which enumerates non-crossing plants in the $(n + 1)$-sided regular polygon [8].
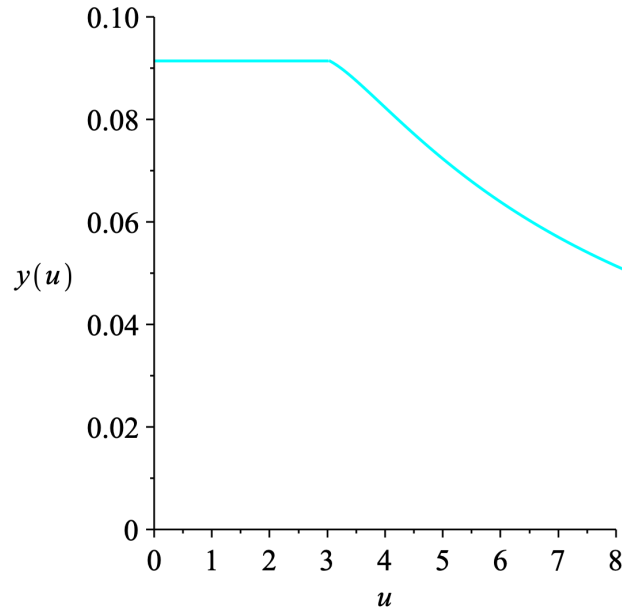
■ **Figure 2** Plot of $y(u)$ defined in Proposition 8. At u_C the tangent from the right side is also horizontal.

▶ **Proposition 8** (Definitions of u_C and of $y(u)$). *Recall that ρ_B is the radius of convergence of $B(y)$. For $u \geq 0$, the equation*

$$\frac{2yuB'(y)}{uB(y) + 1 - u} = 1 \tag{10}$$

has a unique solution in $[0, \rho_B]$, denoted by $y(u)$, if and only if $u \geq u_C$, where

$$u_C := \frac{9\pi(4-\pi)}{420\pi - 81\pi^2 - 512} \simeq 3.02. \tag{11}$$

Moreover, we set $y(u) := \rho_B$ for $u \leq u_C$.

▶ Remark 9. The function $u \mapsto y(u)$ is non-increasing. It is plotted in Figure 2.

The value of u_C defined above is the *critical point* of the model, and permits to identify three regimes for which the singular behavior of $M(z, u)$ differs:

▶ **Proposition 10.** *For $u > 0$, let $\rho(u)$ be the radius of convergence of $z \mapsto M(z, u)$. Then, one has*

$$\rho(u) = \frac{y(u)}{(uB(y(u)) + 1 - u)^2}, \tag{12}$$

and the following singular expansions hold in a Δ-neighbourhood of $\rho(u)$, with $Z = 1 - z/\rho(u)$.

■ *When $u < u_C$ (subcritical case),*

$$M(z, u) = q(u) - r(u)Z - s(u)\ln(Z)Z^2 + O(Z^2), \tag{13}$$

where

$$q(u) = 1 + u\left(7 - \frac{64}{3\pi}\right), \quad r(u) = \frac{8u(3\pi - 8)(10 - 3\pi)(21\pi u + 3\pi - 64u)}{(243u - 27)\pi^3 - (1260u - 108)\pi^2 + 1536\pi u},$$

$$s(u) = \frac{2u(21\pi u + 3\pi - 64u)^3(3\pi - 8)^3}{\pi(81\pi^2 u + 512u + 36\pi - 420\pi u - 9\pi^2)^3}.$$

- *When $u = u_c$ (critical case),*

$$M(z, u) = q_C + s_C \ln(Z)^{-1/2} Z^{1/2} + O(Z), \tag{14}$$

where

$$q_C = q(u_C) = \frac{864\pi - 144\pi^2 - 1280}{420\pi - 81\pi^2 - 512}, \quad s_C = \frac{16\sqrt{6} \, (10 - 3\pi)^{\frac{3}{2}} (4 - \pi)}{420\pi - 81\pi^2 - 512}.$$

- *When $u > u_C$ (supercritical case),*

$$M(z, u) = q(u) - s(u) \, Z^{1/2} + O(Z), \tag{15}$$

where

$$q(u) = uB(y(u)) + 1 - u, \quad s(u) = \frac{uB(y(u)) + 1 - u}{\sqrt{1 + 2y(u) \frac{B''(y(u))}{B'(y(u))}}}.$$

Moreover, $\rho(u)$ is the unique dominant singularity of $z \mapsto M(z, u)$ for every $u > 0$.

Proof. Let $H^{(u)}(z) := zM(z, u)^2$, and $B^{(u)}(y) := uB(y) + 1 - u$. Squaring the equation $M(z, u) = B^{(u)}(H^{(u)}(z))$ and multiplying both sides by z, one gets the functional equation

$$H^{(u)}(z) = zB^{(u)}(H^{(u)}(z))^2,$$

which is of Lagrangean type. The functional inverse of $H^{(u)}(z)$ is thus $\Psi^{(u)}(y) := y/B^{(u)}(y)^2$, and the singular expansion of $H^{(u)}(z)$ (and hence of $M(z, u)$) depends on whether $\frac{\mathrm{d}}{\mathrm{d}y} \Psi^{(u)}(y) = 0$ – which is equivalent to (10) – admits a solution in $(0, \rho_B)$.

More precisely, for $u > u_c$, $H^{(u)}(z)$ has a dominant singularity of square-root type at $\rho(u) = \Psi^{(u)}(y(u))$ (see [13, Thm VI.6, p. 404-405],), and the same holds for $M(z, u) = (H^{(u)}(z)/z)^{1/2}$, with the constants in (15).

In the limit case $u = u_C$, one has $\frac{\mathrm{d}}{\mathrm{d}y} \Psi^{(u)}(y) = 0$ at $y = \rho_B$, where one gets (with $Y = 1 - y/\rho_B$) the expansion $z = \Psi^{(u)}(y) = \rho(u_C) + \xi Y^2 \ln(Y) + O(Y^2)$ for some explicit $\xi > 0$. By inversion and bootstrapping, one gets $y = H^{(u)}(z) = \rho_B - \rho_B \sqrt{2\rho(u_C)/\xi} \sqrt{Z/\ln(Z)} + O(Z)$, and a similar expansion holds for $M(z, u) = (H^{(u)}(z)/z)^{1/2}$, with the explicit constants in (14).

For $u < u_C$, one has $\frac{2\rho_B uB'(\rho_B)}{uB(\rho_B)+1-u} < 1$, and there is no solution to $\frac{\mathrm{d}}{\mathrm{d}y} \Psi^{(u)}(y) = 0$ on $[0, \rho_B]$. At ρ_B, one gets the expansion $z = \Psi^{(u)}(y) = \rho(u) - \kappa Y + \xi Y^2 \ln(Y) + O(Y^2)$ for some explicit $\kappa, \xi > 0$, and writing $Y = 1 - y/\rho_B$. By inversion and bootstrapping, one gets $y = H^{(u)}(z) = \rho_B - \frac{\rho(u)\rho_B}{\kappa} Z - \frac{\rho(u)^2 \rho_B}{\kappa^3} Z^2 \ln(Z) + O(Z^2)$, and a similar expansion holds for $M(z, u) = (H^{(u)}(z)/z)^{1/2}$, with the explicit constants in (13).

Finally, for every fixed $u > 0$, the equation for $H^{(u)}(z)$ is of (non-periodic) Lagrangean type, hence $\rho(u)$ is the unique dominant singularity of $H^{(u)}(z)$, and the same holds for $z \mapsto M(z, u) = B^{(u)}(H^{(u)}(z))$. ◀

Applying transfer theorems to the expansions in Proposition 10 then gives:

▶ **Corollary 11.** *Let $u > 0$. Then, with the notation of Proposition 10, one has the following asymptotic estimates as $n \to \infty$.*
- *When $u < u_C$,*

$$[z^n]M(z, u) \sim 2s(u) \, \rho(u)^{-n} \, n^{-3}. \tag{16}$$

- *When $u = u_C$,*

$$[z^n]M(z, u_C) \sim \frac{s_C}{2\sqrt{\pi}} \, \rho(u)^{-n} \, n^{-3/2} \ln(n)^{-1/2}. \tag{17}$$

- *When $u > u_C$,*

$$[z^n]M(z, u) \sim \frac{s(u)}{2\sqrt{\pi}} \, \rho(u)^{-n} \, n^{-3/2}. \tag{18}$$

▶ **Remark 12.** Instead of analytic combinatorics methods, one can also obtain these results using probabilistic methods. In the subcritical case, they require having an estimate of the size of the largest block, which is provided in Section 4.2.

4 Probabilistic study of tree-rooted maps

The purpose of this section is to study the phase transition undergone by a random tree-rooted maps weighted by their number of 2-connected blocks. Following [1, 14], this study is based on a interpretation of the block tree as a Bienaymé–Galton–Watson process (Section 4.1), to obtain asymptotic estimates on the size of the largest blocks (Section 4.2), and scaling limit results in the critical and supercritical cases (Section 4.3).

4.1 Definition of the probabilistic model and Bienaymé–Galton–Watson trees

We consider the following probability distribution on the class $\mathcal{M}$ of tree-rooted maps, indexed by a parameter $u > 0$: for any integer $n \geq 0$, we define

$$\mathbb{P}_n^{(u)}(\mathfrak{m}) = \frac{u^{b(\mathfrak{m})}}{[z^n]M(z, u)} \quad \text{for any } \mathfrak{m} \in \mathcal{M}_n. \tag{19}$$

We denote by $\mathrm{M}_n^{(u)}$ a tree-rooted map sampled from $\mathbb{P}_n^{(u)}$, by $\mathrm{T}_n^{(u)}$ the block tree associated to it, and by $(\mathrm{B}_v, v \in \mathrm{T}_n^{(u)})$ its corresponding decorations.

For $n \in \mathbb{Z}_{\geq 0}$ and μ a probability distribution on $\mathbb{Z}_{\geq 0}$, $GW(\mu, n)$ denotes the law of a Bienaymé–Galton–Watson tree with offspring distribution μ conditioned to have n edges. For $u > 0$ and $y \in [0, \rho_B]$, let $\mu^{y,u}$ be the probability distribution on $\mathbb{Z}_{\geq 0}$ defined by setting, for $j \geq 0$,

$$\mu^{y,u}(2j) := \frac{b_j y^j u^{1_{j \neq 0}}}{uB(y) + 1 - u}, \qquad \text{so that} \qquad \mathbb{E}\left[\mu^{y,u}\right] = \frac{2uyB'(y)}{uB(y) + 1 - u}. \tag{20}$$

Moreover, we set $\mu^u := \mu^{y(u),u}$, where $y(u)$ is defined in Proposition 8. Then, the following proposition is the tree-rooted analogue of [14, Proposition 3.1] (itself an extension of [1, Proposition 3.1]).

▶ **Proposition 13.** *For every $u > 0$ and any $n \geq 0$, under $\mathbb{P}_n^{(u)}$, the law of the tree of blocks $(\mathrm{T}_n^{(u)}, (\mathrm{B}_v, v \in \mathrm{T}_n^{(u)}))$ can be described as follows.*

- *$\mathrm{T}_n^{(u)}$ follows the law $GW(\mu^u, 2n)$;*
- *Conditionally given $\mathrm{T}_n^{(u)} = \mathfrak{t}$, the blocks $(\mathrm{B}_v, v \in \mathfrak{t})$ are independent random variables, and, for $v \in \mathfrak{t}$, B_v follows a uniform distribution on the set of blocks of size $k_v(\mathfrak{t})/2$, where $k_v(\mathfrak{t})$ is the number of children of v in $\mathfrak{t}$.*

Therefore the behavior of $\mathrm{T}_n^{(u)}$ will be driven by the properties of μ^u. It follows from Corollary 7 and Proposition 8 that μ^u exhibits the following phase transition:

▶ **Lemma 14.** *For any $u > 0$, define*

$$c(u) = \frac{4\,(3\pi - 8)^3}{9\,(4 - \pi)^3} \frac{u}{(21\pi - 64)u + 3\pi}. \tag{21}$$

Then, one has:
Subcritical case. *For $u < u_C$,*

$$E(u) := \mathbb{E}\,[\mu^u] = \frac{16\,(3\pi - 8)\,(10 - 3\pi)}{3\,(4 - \pi)} \frac{u}{(21\pi - 64)u + 3\pi} < 1 \quad and \quad \mu^u(\{2j\}) \underset{j \to \infty}{\sim} c(u)j^{-3}; \tag{22}$$

Critical case. *For $u = u_C$,*

$$\mathbb{E}\,[\mu^u] = 1 \qquad and \qquad \mu^{u_C}\,(\{2j\}) \underset{j \to \infty}{\sim} c(u_C)j^{-3} = \frac{(3\pi - 8)^2}{12\,(10 - 3\pi)\,(4 - \pi)^2}j^{-3} \simeq 0.40\,j^{-3};$$

Supercritical case. *For $u > u_C$,*

$$\mathbb{E}\,[\mu^u] = 1 \qquad and \qquad \mu^u(\{2j\}) \sim c(u) \left(\frac{y(u)}{\rho_B}\right)^j j^{-3},$$

where $y(u) < \rho_B$ so μ^u has exponential moments.

4.2   Phase transition for the sizes of the largest blocks

This section puts into light a phase transition for the block sizes of random tree-rooted maps drawn according to $\mathbb{P}_n^{(u)}$. We use probabilistic techniques to obtain the results, but an analysis using a saddle-point method could also be carried out.

For $\mathfrak{m}$ a tree-rooted map, denote by $\mathrm{LB}_1(\mathfrak{m}) \geq \cdots \geq \mathrm{LB}_{b(\mathfrak{m})}(\mathfrak{m})$ the sizes of its blocks in decreasing order. By convention, we set $\mathrm{LB}_k(\mathfrak{m}) = 0$ if $k > b(\mathfrak{m})$. For a random variable X_n and a positive sequence a_n, recall that $X_n = O_\mathbb{P}(a_n)$ (resp. $X_n = \Theta_\mathbb{P}(a_n)$) means that $(\mathbb{P}(X_n \leq a_n u_n))_{n \geq 0}$ (resp. $(\mathbb{P}(a_n/u_n \leq X_n \leq a_n u_n))_{n \geq 0}$) tends to 1 for any positive u_n tending to $+\infty$.

▶ **Theorem 15.** *The random tree-rooted map $\mathrm{M}_n^{(u)}$, drawn according to $\mathbb{P}_n^{(u)}$, exhibits the following behaviours when n tends to infinity.*
Subcritical case. *For $u < u_c$, the largest bloc is macroscopic, and more precisely one has:*

$$\frac{\mathrm{LB}_1(\mathrm{M}_n^{(u)}) - (1 - E(u))n}{\sqrt{c(u)n\ln(n)}} \xrightarrow[n \to \infty]{(d)} \mathcal{N}(0, 1). \tag{23}$$

Furthermore, for any fixed $j \geq 2$, it holds that $\mathrm{LB}_j(\mathrm{M}_n^{(u)}) = \Theta_\mathbb{P}(n^{1/2})$ and for $x > 0$:

$$\mathbb{P}\left(\mathrm{LB}_j(\mathrm{M}_n^{(u)}) \leq x\sqrt{n}\right) \xrightarrow[n \to \infty]{} e^{-\lambda(x)} \sum_{p=0}^{j-2} \frac{\lambda(x)^p}{p!}, \qquad where \ \lambda(x) := \frac{c(u)}{2x^2}. \tag{24}$$

Critical case. *For $u = u_C$, for any fixed $j \geq 1$, it holds that $\mathrm{LB}_j(\mathrm{M}_n^{(u)}) = \Theta_\mathbb{P}(n^{1/2})$. More precisely, up to a shift of indices, the sizes of the blocks exhibit a similar behavior as the sizes of non-macroscopic blocks in the subcritical regime, namely, for $x > 0$:*

$$\mathbb{P}\left(\mathrm{LB}_j(\mathrm{M}_n^{(u)}) \leq x\sqrt{n}\right) \xrightarrow[n \to \infty]{} e^{-\lambda(x)} \sum_{p=0}^{j-1} \frac{\lambda(x)^p}{p!}, \qquad where \ \lambda(x) := \frac{c(u_C)}{2x^2}. \tag{25}$$

Supercritical case. *For $u > u_C$, for all fixed $j \geq 1$, it holds as $n \to \infty$ that*

$$\mathrm{LB}_j(\mathrm{M}_n^{(u)}) = \frac{\ln(n)}{\ln\left(\frac{\rho_B}{y(u)}\right)} - \frac{3\ln(\ln(n))}{\ln\left(\frac{\rho_B}{y(u)}\right)} + O_{\mathbb{P}}(1).$$

Proof. In all three cases, we make extensive use of Janson's survey [19]. In the supercritical case, one can proceed as in the non–tree-rooted case [14] and use the survey's Theorem 19.16. In the critical case, its Example 19.29 can be applied. The subcritical case is a bit more involved, Janson's Theorem 19.34 can be applied for the size of the j-th largest block for $j \geq 2$. Results from Kortchemski [20, Theorem 1] allow to conclude for the largest block. ◄

▶ **Remark 16.** One can get a local limit theorem for $\mathrm{LB}_1(\mathrm{M}_n^{(u)})$ in the subcritical case as in [24] (up to the technicality that nodes of the block tree have only even numbers of children). Furthermore, one can state a joint limit law for the sizes $\mathrm{LB}_j(\mathrm{M}_n^{(u)})$. For any fixed $r \geq 1$,

$$\left(\frac{c(u)}{2} \left(\frac{\mathrm{LB}_j(\mathrm{M}_n^{(u)})}{\sqrt{n}} \right)^{-2}, \ 2 \leq j \leq r+1 \right) \xrightarrow[n\to\infty]{(d)} (A_1, \ldots, A_r),$$

where the A_i are the decreasingly ordered atoms of a Poisson Point Process of rate 1 on $\mathbb{R}_+$. The same joint limit law holds at u_C (with j from 1 to r).

▶ **Remark 17.** In contrast to the case of non–tree-rooted maps [14], here in the subcritical case $u < u_C$ the size of the second block is negligible compared to the order of fluctuation of the size of largest block. Moreover, for $u < u_C$ and fixed $j \geq 1$, the j-th largest critical block has the same limit law (up to constant rescaling) as the $j + 1$-th largest block in the subcritical regime, which did not hold in the non–tree-rooted case. Informally, the conditioning that a random walk (subcritical case) is an excursion (critical case) has negligible effect on the law of the largest steps, so subcritical blocks (for $j \geq 2$) behave like critical blocks.

4.3 Scaling limit in the critical and supercritical cases

In the critical and the supercritical cases, we can establish the following convergence result:

▶ **Theorem 18.** *For any fixed $u \geq u_c$, there exist some constants α_u, β_u and γ_u such that:*
- *If $u > u_c$, it holds that:*

$$\frac{\gamma_u}{\sqrt{n}} \cdot \left(\mathrm{M}_n^{(u)}, \tau(\mathrm{M}_n^{(u)}), \mathrm{T}_n^{(u)} \right) \xrightarrow[n\to\infty]{(d)} (\alpha_u \cdot \mathcal{T}_\mathrm{e}, \beta_u \cdot \mathcal{T}_\mathrm{e}, \mathcal{T}_\mathrm{e}); \tag{26}$$

- *If $u = u_c$, it holds that:*

$$\frac{\gamma_{u_C}\sqrt{\log(n)}}{\sqrt{n}} \left(\mathrm{M}_n^{(u_C)}, \tau(\mathrm{M}_n^{(u_C)}), \mathrm{T}_n^{(u_C)} \right) \xrightarrow[n\to\infty]{(d)} (\alpha_{u_C} \cdot \mathcal{T}_\mathrm{e}, \beta_{u_C} \cdot \mathcal{T}_\mathrm{e}, \mathcal{T}_\mathrm{e}); \tag{27}$$

where, in both cases, each $\mathcal{T}_\mathrm{e}$ is a copy of the same realization of the Brownian Continuum Random Tree (CRT), and the convergence holds in the Gromov-Hausdorff-Prokhorov sense.

Convergence towards the CRT in the supercritical case was previously obtained by Stufler [23, Theorem 6.63], who considers a framework where a block-weighted map is sampled, and *afterwards* one of its spanning trees is uniformly sampled. Block-weighted models of random tree-rooted maps fall in this model upon tweaking the weights [23, Remark 6.65]. Our contribution lies in showing scaling limit in the critical case (and having a unified proof for both the supercritical and critical cases), and finding the value of u_C.

Proof of Theorem 18. The convergence of the sequence of tree of blocks $(T_n^{(u)})$ follows from classical results about the scaling limit of Bienaymé–Galton–Watson trees towards the CRT [21, 10], with an offspring distribution with a finite second moment in the $u > u_c$ case or with finite moments of order $2 + \varepsilon$ in the critical case. The values of the constant γ_u follows from general results. In the case $u > u_c$, $\gamma_u = \sigma_u/2$, where σ_u is the standard deviation of μ^u, given explicitely by

$$\sigma_u^2 = 1 + 2y(u)\frac{B''(y(u))}{B'(y(u))}.$$

For $u = u_C$, the variance of μ^{u_C} is infinite, and we get $\gamma_{u_c} = \sqrt{2c(u_C)}$ (see *e.g.* [18, Ex. 7.10]).

To establish the scaling limit of $M_n^{(u)}$ and of $\tau(M_n^{(u)})$, we proceed as in [23, 14] and prove that the distances in the map and in the spanning tree are – up to a linear factor – equivalent in the limit to the distances in the block tree. The proof extends effortlessly, and only requires (for the critical case) to have a control on the diameter of the 2-connected blocks and of their spanning trees, that we state in Lemma 19. The Prokhorov part of the convergence can be established along the same lines as [14, Lemma 5.13]. ◀

A sequence $(p_n)_{n\geq 0}$ of nonnegative real numbers is called *stretched-exponential* if there is $\gamma > 0$ such that $p_n \leq \exp(-n^\gamma)$ for n large enough.

▶ **Lemma 19** (Bound on the diameter of random 2-connected tree-rooted maps). *Let B_n be a uniformly random 2-connected tree-rooted map of size n, and recall that $\tau(B_n)$ denotes its distinguished spanning tree. Then, for any $\varepsilon > 0$, the sequences $\big(\mathbb{P}(\mathrm{Diam}(B_n) \geq n^{1/2+\varepsilon})\big)_{n\geq 0}$ and $\big(\mathbb{P}(\mathrm{Diam}(\tau(B_n)) \geq n^{1/2+\varepsilon})\big)_{n\geq 0}$ are stretched-exponential[2].*

Proof. Note first that it is enough to establish the result for $\mathrm{Diam}(\tau(B_n))$ since $\mathrm{Diam}(B_n) \leq \mathrm{Diam}(\tau(B_n))$ deterministically.

By Mullin's bijection [22], for M_n a uniform element of $\mathcal{M}_n$, the height of $\tau(M_n)$ is distributed as the maximal abscissa X_n in a random walk (with steps in $\{W, E, S, N\}$) of length $2n$, ending at the origin and staying in the right-hand upper quadrant. It is easy to establish, *e.g.* using Chernoff's bound and a union bound, that the maximal abscissa $\widetilde{X_n}$ in a random walk of length $2n$ in $\mathbb{Z}^2$ is such that, for any $\varepsilon > 0$, the sequence $\left(\mathbb{P}(\widetilde{X_n} \geq n^{1/2+\varepsilon})\right)_{n\geq 0}$ is stretched-exponential.

Since the random walk has probability $\Theta(n^{-3})$ to end at the origin and to stay in the quadrant, the sequence $\big(\mathbb{P}(X_n \geq n^{1/2+\varepsilon})\big)_{n\geq 0}$ is also stretched-exponential. And so is the sequence $\big(\mathbb{P}(\mathrm{Diam}(\tau(M_n)) \geq n^{1/2+\varepsilon})\big)_{n\geq 0}$ since the diameter of a tree is at most twice its height. Let $\alpha = 1 - E(1)$. By the results of the previous section, one gets

$$\mathbb{P}(\mathrm{LB}_1(M_{\lfloor n/\alpha\rfloor}) = n) = \Theta(1/\sqrt{n\ln(n)}),$$

and in that case a block of maximal size is distributed as B_n. Hence, the sequence $\big(\mathbb{P}(\mathrm{Diam}(\tau(B_n)) \geq n^{1/2+\varepsilon})\big)_{n\geq 0}$ is also stretched-exponential. ◀

---

[2] To prove Theorem 18, it is actually enough to have the result for an exponent strictly smaller than 1.

5 Perspectives

It has been recently shown [9, 16, 17] that for M_n a random tree-rooted map of size n, the volume-growth exponent (whose inverse should give the exponent for the order of magnitude of the diameter) is with high probability in the interval $[3.550408, 3.63299]$. It would be interesting to verify whether these bounds also hold for the random 2-connected tree-rooted map B_n, and more generally for the random tree-rooted map $M_n^{(u)}$ in the subcritical regime.

Regarding extensions of the model, one could consider maps endowed with a spanning forest, with weight v per tree in the forest, which were studied by Bousquet-Mélou and Courtiel [7], and one could additionally have a weight $u > 0$ per 2-connected block. They showed that, for $v > 0$, one gets the asymptotic behaviour $n^{-5/2}$ as in pure maps [7]. The phase transition should thus be of the same nature than for the non–tree-rooted case [14], and we expect the scaling limit to be a stable tree of parameter $3/2$ at the critical weight $u_C(v)$. The case $v = 0$ corresponds to tree-rooted maps as studied here. Interestingly, their model still has a combinatorial interpretation for $v \in [-1, 0)$, with asymptotic behaviour $n^{-3} \ln(n)^{-2}$ [7]. From this behaviour it can be expected that, at the critical weight $u_C(v)$, the asymptotic enumeration has a correcting term $n^{-3/2}$ and the scaling limit is the CRT with distances rescaled by $n^{1/2}$ (same order of magnitude as in the supercritical case). To have a continuous range of asymptotic exponents, one could more generally consider random maps weighted by a Potts model, and additionally weighted at blocks (a method to derive the singular exponents of general maps weighted by a Potts model has been developed in [6]; see also [11, 4]).

Finally, one could also consider other kinds of block-decompositions in the context of decorated maps, such as 3-oriented triangulations or 2-oriented quadrangulations decomposed into irreducible components, having a weight u per such component. The asymptotic exponents at $u = 1$ are n^{-5} and n^{-4} [12], respectively. This suggests that, as in the above mentioned model, at the critical weight u_C, the model exhibits a tree-behaviour: the asymptotic enumeration has polynomial correction $n^{-3/2}$ and rescaling the distances by $n^{1/2}$ gives convergence towards the CRT.

References

1 Louigi Addario-Berry. A probabilistic approach to block sizes in random maps. *ALEA - Latin American Journal of Probability and Mathematical Statistics*, XVI:1–13, 2019.

2 Olivier Bernardi. Bijective counting of tree-rooted maps and shuffles of parenthesis systems. *Electron. J. Comb.*, 14(1), 2007.

3 Olivier Bernardi. Tutte polynomial, subgraphs, orientations and sandpile model: New connections via embeddings. *Electron. J. Comb.*, 15(1), 2008.

4 Olivier Bernardi and Mireille Bousquet-Mélou. Counting colored planar maps: algebraicity results. *Journal of Combinatorial Theory, Series B*, 101(5):315–377, 2011.

5 Valentin Bonzom. Large N limits in tensor models: Towards more universality classes of colored triangulations in dimension $d \geq 2$. *Symmetry, Integrability and Geometry: Methods and Applications*, 12(073):39, 2016.

6 Gaëtan Borot, Jérémie Bouttier, and Emmanuel Guitter. Loop models on random maps via nested loops: the case of domain symmetry breaking and application to the Potts model. *Journal of Physics A: Mathematical and Theoretical*, 45(49):494017, 2012.

7 Mireille Bousquet-Mélou and Julien Courtiel. Spanning forests in regular planar maps. *Journal of Combinatorial Theory, Series A*, 135:1–59, 2015.

8 Frédéric Chapoton. The Anticyclic Operad of Moulds. *International Mathematics Research Notices*, 2007.

9   Jian Ding and Ewain Gwynne. The fractal dimension of Liouville quantum gravity: universality, monotonicity, and bounds. *Communications in Mathematical Physics*, 374(3):1877–1934, 2020.

10   Thomas Duquesne and Jean-François Le Gall. *Random trees, Lévy processes and spatial branching processes*, volume 281. Société mathématique de France Paris, France, 2002.

11   B. Eynard and G. Bonnet. The Potts-q random matrix model: loop equations, critical exponents, and rational case. *Physics Letters B*, 463(2-4):273–279, 1999.

12   Stefan Felsner, Éric Fusy, and Marc Noy. Asymptotic enumeration of orientations. *Discrete Mathematics & Theoretical Computer Science*, 12, 2010.

13   Philippe Flajolet and Robert Sedgewick. *Analytic combinatorics*. Cambridge University press, 2009.

14   William Fleurat and Zéphyr Salvy. A phase transition in block-weighted random maps. *Electron. J. Probab.*, 29:1–61, 2024.

15   Xavier Gourdon. Largest component in random combinatorial structures. *Discret. Math.*, 180(1-3):185–209, 1998.

16   Ewain Gwynne, Nina Holden, and Xin Sun. A mating-of-trees approach for graph distances in random planar maps. *Probability Theory and Related Fields*, 177(3-4):1043–1102, 2020.

17   Ewain Gwynne and Joshua Pfeffer. Bounds for distances and geodesic dimension in Liouville first passage percolation. *Electron. Commun. Probab.*, 24:1–12, 2019.

18   Svante Janson. Stable distributions. *arXiv preprint arXiv:1112.0220*, 2011.

19   Svante Janson. Simply generated trees, conditioned Galton–Watson, random allocations and condensation. *Probability Surveys*, 9:103–252, 2012.

20   Igor Kortchemski. Limit theorems for conditioned non-generic Galton–Watson trees. *Ann. Inst. Henri Poincaré Probab. Stat.*, 51(2):489–511, May 2015.

21   Jean-François Le Gall. Random trees and applications. *Probability Surveys*, 2:245, 2005.

22   R. C. Mullin. On the enumeration of tree-rooted maps. *Canadian Journal of Mathematics*, 19:174–183, 1967.

23   Benedikt Stufler. Limits of random tree-like discrete structures. *Probability Surveys*, pages 318–477, 2020.

24   Benedikt Stufler. On the maximal offspring in a subcritical branching process. *Electronic Journal of Probability*, 25:1–62, 2020.

25   William T. Tutte. A census of planar maps. *Canadian Journal of Mathematics*, 15:249–271, 1963.

Composition Schemes: q-Enumerations and Phase Transitions in Gibbs Models

Cyril Banderier 🏠 🆔
Laboratoire d'Informatique de Paris Nord, Université Sorbonne Paris Nord, Villetaneuse, France

Markus Kuba 🏠 🆔
Department Applied Mathematics & Physics, University of Applied Sciences - Technikum Wien, Austria

Stephan Wagner 🏠 🆔
Institute of Discrete Mathematics, TU Graz, Austria
Department of Mathematics, Uppsala Universitet, Sweden

Michael Wallner 🏠 🆔
Institut für Diskrete Mathematik und Geometrie, TU Wien, Austria

── Abstract ──

Composition schemes are ubiquitous in combinatorics, statistical mechanics and probability theory. We give a unifying explanation to various phenomena observed in the combinatorial and statistical physics literature in the context of q-enumeration (this is a model where objects with a parameter of value k have a Gibbs measure/Boltzmann weight q^k). For structures enumerated by a composition scheme, we prove a phase transition for any parameter having such a Gibbs measure: for a critical value $q = q_c$, the limit law of the parameter is a two-parameter Mittag-Leffler distribution, while it is Gaussian in the supercritical regime ($q > q_c$), and it is a Boltzmann distribution in the subcritical regime ($0 < q < q_c$). We apply our results to fundamental statistics of lattice paths and quarter-plane walks. We also explain previously observed limit laws for pattern-restricted permutations, and a phenomenon uncovered by Krattenthaler for the wall contacts in watermelons.

1 Introduction

1.1 q-enumeration and Gibbs distributions

Let $\mathcal{T}$ be a family of combinatorial objects, let $|\cdot|$ denote the size of objects, and let $\mathcal{X}\colon \mathcal{T} \to \mathbb{N}$ be a statistic defined on $\mathcal{T}$. The statistic $\mathcal{X}$ on objects of $\mathcal{T}$ of size n can be encoded by the sum

$$f_n(q) = \sum_{T \in \mathcal{T} : |T| = n} q^{\mathcal{X}(T)}. \tag{1}$$

This sum reduces for $q = 1$ to the total number $f_n(1) = f_n$ of objects of size n. In combinatorics, for any given $q \in \mathbb{R}$, it is called the *q-enumeration* of $\mathcal{T}$ of size n with respect to $\mathcal{X}$ (see, e.g., [1, 31]). In the language of statistical mechanics, $f_n(q)$ is a partition function with Boltzmann weight q. The bivariate generating function $F(z, q)$ is then defined as

$$F(z, q) = \sum_{T \in \mathcal{T}} z^{|T|} q^{\mathcal{X}(T)} = \sum_{n \geqslant 0} f_n(q) z^n = \sum_{n \geqslant 0} \sum_{k \geqslant 0} f_{n,k} z^n q^k. \tag{2}$$

Here $f_{n,k}$ denotes the number of objects of $\mathcal{T}$ of size n for which $\mathcal{X}$ equals k. It is usual to associate with the statistic $\mathcal{X}$ the random variables X_n, $n \geqslant 1$, defined as

$$\mathbb{P}(X_n = k) = \frac{f_{n,k}}{f_n} = \frac{[q^k] f_n(q)}{f_n(1)} = \frac{[z^n q^k] F(z, q)}{[z^n] F(z, 1)}, \tag{3}$$

such that each object from $\mathcal{T}$ of size n is equally likely. The associated probability generating function is given by $\mathbb{E}(q^{X_n}) = \frac{[z^n] F(z, q)}{[z^n] F(z, 1)}$. In Equation (3), the reader is probably used to consider q as a formal variable, but in this work, like in statistical mechanics, we shall consider q as an adjustable parameter (weight $\in \mathbb{R}_+$) of the underlying combinatorial and physical structures. This is also the spirit of the Boltzmann sampling method [18], where q is tuned to minimize the number of rejection steps in the sampling algorithm.

More precisely, in this article, we put a Gibbs measure on the statistic $\mathcal{X}$; that is, one has the following probabilistic model.

▶ **Definition 1** (Gibbs distribution). *Let a family $\mathcal{T}$ of combinatorial objects and a statistic $\mathcal{X}\colon \mathcal{T} \to \mathbb{N}$ be given. For real $q > 0$, the Gibbs distribution of this statistic is the law of the random variable $X_n(q)$ with probability mass function*

$$\mathbb{P}(X_n(q) = k) = \frac{f_{n,k} q^k}{f_n(q)}, \quad k \geqslant 0.$$

In terms of the probability generating function $p(v) = \mathbb{E}(v^{X_n(1)})$, we have $\mathbb{E}(v^{X_n(q)}) = \frac{p(vq)}{p(q)}$.

A well-known example in probability theory is the Mallows distribution [36] on permutations with respect to the inversion statistic.

In many applications, one is interested in the limit distribution of $X_n(q)$, which depends on the value of $q > 0$; see, e.g., [9, 11, 12, 14, 33, 34, 37, 40]. Let us pinpoint the result of Krattenthaler [33], who uncovered a phase transition in the normalized *mean number* of wall contacts at $q = 2$ in watermelons. Using methods from analytic combinatorics, we will show that similar phase transitions naturally occur in a great many instances, not only with respect to the expectation but also for the *limit laws*. We will use the framework of composition schemes, which often provide a direct and unifying way to explain why phase transitions occur [5, 6, 25]. In Section 2, we establish in which way the phase transitions in the Gibbs model depend on the value of q. Particular instances of similar phenomena have been observed in [13, 43]. We give further examples in Sections 3 and 4.

1.2 Composition schemes and Gibbs distributions

Functional composition schemes such as $F(z) = G\big(H(z)\big)$ are of great importance in combinatorics [5, 6, 25] and probability theory [39]. The main focus is to analyse probabilistic properties of compositions like

$$F(z, u) = G\big(uH(z)\big) \tag{4}$$

as a multitude of parameters $\mathcal{X}$ can be modelled in this way. Here u marks the so-called size of the core, i.e., the involved G-component; see [5, 6, 25]. The distribution of the corresponding random variable X_n is then readily defined by

$$\mathbb{P}(X_n = k) = \frac{[z^n u^k] F(z, u)}{[z^n] F(z, 1)}.$$

Structurally, such schemes are at the heart of many fascinating phase transition phenomena (analytically corresponding, e.g., to coalescing saddle points or to confluence of singularities), related to the Gibbs measure in statistical physics and probability theory [39].

First, we relate composition schemes to q-enumeration.

▶ **Lemma 2** (Composition schemes and Gibbs distributions). *Let a combinatorial statistic $\mathcal{X}$ with bivariate generating function $F(z, u)$ be given. Then, for real $q > 0$ the Gibbs distribution of $\mathcal{X}$ has a probability mass function given in terms of F by*

$$\mathbb{P}(X_n(q) = k) = \frac{[z^n u^k] F(z, qu)}{[z^n] F(z, q)}.$$
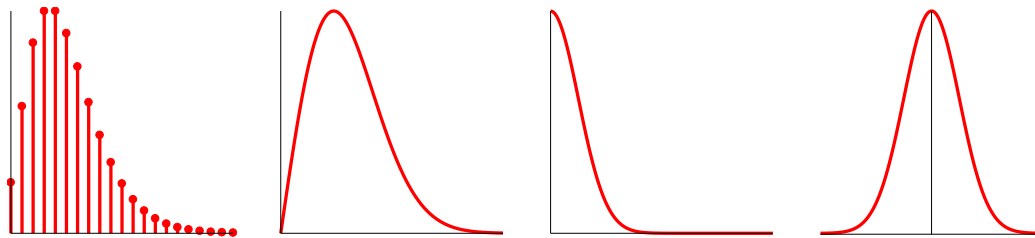
■ **Figure 1** Under the Gibbs model, we get different limit distributions for the random variable $X_n(q)$ associated with statistics counted by a composition scheme: from left to right, Boltzmann (e.g., the negative binomial), Mittag-Leffler (e.g., Rayleigh), chi (e.g., half-normal), Gaussian. We establish their universality in the next sections.

In many (combinatorial) applications of composition schemes, the following assumptions hold [6]: all involved generating functions $G(z)$, $H(z)$, and $F(z)$ have nonnegative coefficients and aperiodic support, are analytic in a Δ-domain with a finite radius of convergence ρ_F and possess singular expansions of the form

$$F(z) = P\left(1 - \frac{z}{\rho_F}\right) + c_F \cdot \left(1 - \frac{z}{\rho_F}\right)^{\lambda_F} (1 + o(1)), \tag{5}$$

where $\lambda_F \in \mathbb{R} \backslash \{0, 1, 2, \dots\}$ is called the *singular exponent* (at $z = \rho_F$), and where $P(x) \in \mathbb{C}[x]$ is a polynomial (of degree $\geqslant 1$ for $\lambda_F > 1$, of degree 0 for $0 < \lambda_F < 1$, and $P = 0$ for $\lambda_F < 0$). Then, singularity analysis [25] can often be used to compute the asymptotics of the coefficients. Note that the sign of the constant c_F depends on the Puiseux exponent λ_F; see [6, Equation (8)].

For the reader's convenience, we recall the classical terminology for composition schemes in the following definition.

▶ **Definition 3** (Classification of composition schemes). *Let $\tau_H = H(\rho_H)$. A composition scheme $F(z) = G\big(H(z)\big)$ is called* subcritical *if it satisfies $\tau_H < \rho_G$,* critical *if it satisfies $\tau_H = \rho_G$, and* supercritical *if it satisfies $\tau_H > \rho_G$.*

We note that each individual class of critical schemes leads to very diverse combinatorial and probabilistic phenomena [3, 5, 6, 25, 39].

One typical example of a combinatorial construction of shape $F(z, u) = G\big(uH(z)\big)$ is given by the sequence construction.

▶ **Example 4** (Sequence of objects). Given a combinatorial structure $\mathcal{H}$, let $\mathcal{F} = \text{SEQ}(\mathcal{H})$. Using the variable u to encode the core size (i.e., the number of $\mathcal{H}$ components), one has $F(z, u) = 1/(1 - uH(z))$. This is a composition scheme with $G(z) = 1/(1 - z)$, with $\rho_G = 1$ and $\lambda_G = -1$.

We observe that the parameter q can thus directly influence the nature of the underlying singular structure when the total mass (obtained by $u = 1$) changes from $[z^n]F(z, 1)$ to $[z^n]F(z, q)$. Let us make this more precise for the following class of functions that includes the sequence construction.

▶ **Lemma 5** (Nature and asymptotics of q-enumerated composition schemes). *Let a composition scheme $F(z, u) = G\big(uH(z)\big)$ with singular exponents $\lambda_G < 0$ and $0 < \lambda_H < 1$ be given. Let $q_c := \frac{\rho_G}{\tau_H} = \frac{\rho_G}{H(\rho_H)} > 0$. The nature of the scheme then splits into three different regimes:*

- *for $0 < q < q_c$, the scheme is subcritical;*
- *for $q = q_c$, the scheme is critical;*
- *for $q > q_c$, the scheme is supercritical.*

Accordingly, if one imposes a Gibbs measure on the number of $\mathcal{H}$-components, this impacts the asymptotics of their q-enumeration $f_n(q)$ as follows:

$$
f_n(q) \sim
\begin{cases}
\frac{c_H q G'(q\tau_H)}{\Gamma(-\lambda_H)} \rho_H^{-n} n^{-\lambda_H - 1}, & \text{for} \quad 0 < q < q_c, \\[2mm]
c_G \left(-\frac{c_H}{\tau_H}\right)^{\lambda_G} \frac{1}{\Gamma(-\lambda_H \lambda_G)} \rho_H^{-n} n^{-\lambda_H \lambda_G - 1}, & \text{for} \quad q = q_c, \\[2mm]
c_G \left(\frac{q\rho H'(\rho)}{\rho_G}\right)^{\lambda_G} \frac{1}{\Gamma(-\lambda_G)} \rho^{-n} n^{-\lambda_G - 1}, & \text{for} \quad q > q_c,
\end{cases}
$$

where, in the last case, ρ is the unique solution of $qH(\rho) = \rho_G$ in the interval $(0, \rho_H)$.

Proof of Lemma 5. First, we turn to the nature of the scheme $F_q(z) = G(qH(z))$. Since $F_q(z)$ has nonnegative coefficients, Pringsheim's Theorem [25] implies that there is a singularity on the real axis. Further, $H(z)$ is monotonically increasing on the real axis from 0 to ρ_H, where it attains the value $\tau_H = H(\rho_H) < \infty$. Thus, the nature of $F_q(z)$ depends on the relation between the singularity ρ_G of $G(z)$ and $q\tau_H$ as claimed. Next, we look at the singular expansions. We start with those of $G(z)$ and $H(z)$. By the assumptions on λ_G and λ_H we have

$$
G(z) \sim c_G \left(1 - \frac{z}{\rho_G}\right)^{\lambda_G} \qquad \text{and} \qquad H(z) \sim \tau_H + c_H \left(1 - \frac{z}{\rho_H}\right)^{\lambda_H}. \tag{6}
$$

In the subcritical regime $0 < q < \frac{\rho_G}{\tau_H}$, the outer function $G(z)$ is analytic at $q\tau_H$ and we combine its expansion with the singular expansion of $H(z)$ around $z = \rho_H$ to obtain

$$F_q(z) = G(q\tau_H) + G'(q\tau_H)\big(qH(z) - q\tau_H\big)(1 + o(1))$$
$$\sim G(q\tau_H) + c_H q G'(q\tau_H)\big(1 - \frac{z}{\rho_H}\big)^{\lambda_H}.$$

Basic singularity analysis [25] provides the stated expansion.

For the critical regime $q = \frac{\rho_G}{\tau_H}$ we obtain

$$F_q(z) \sim c_G \left(1 - q\frac{\tau_H}{\rho_G} - q\frac{c_H}{\rho_G}\left(1 - \frac{z}{\rho_H}\right)^{\lambda_H}\right)^{\lambda_G} = c_G \left(-\frac{c_H}{\tau_H}\right)^{\lambda_G}\left(1 - \frac{z}{\rho_H}\right)^{\lambda_H\lambda_G},$$

which leads to the desired result.

Finally, in the supercritical regime $q > \frac{\rho_G}{\tau_H}$, there exists a unique $0 < \rho < \rho_H$ such that $qH(\rho) = \rho_G$. As $\rho < \rho_G$, we may expand $H(z)$ around ρ. This leads to

$$qH(z) = qH(\rho) + qH'(\rho)(z - \rho) + \mathcal{O}\left((z - \rho)^2\right).$$

Note that by the positivity of the coefficients of H one has $H'(\rho) \neq 0$. Thus, we have

$$F_q(z) \sim c_G \left(\frac{q\rho H'(\rho)}{\rho_G}\right)^{\lambda_G}\left(1 - \frac{z}{\rho}\right)^{\lambda_G},$$

and the asymptotic formula for $f_n(q)$ follows by singularity analysis. ◀

2　Main theorem: Gibbs models and phase transitions with respect to q

The following result is our main theorem. It describes the dependency of the limit law on the parameter q. In this extended abstract, we present only the case of sequence-like schemes.

▶ **Theorem 6** (Gibbs distribution and phase transitions of sequence-like schemes). *Let a composition scheme $F(z, u) = G\big(uH(z)\big)$, with singular exponents $\lambda_G < 0$ and $0 < \lambda_H < 1$, be given. Then, the Gibbs distribution of $X_n = X_n(q)$ associated with $F(z, qv)$ has the following limit laws and phase transition diagram that depend on $q_c = \frac{\rho_G}{\tau_H}$:*

Parameter q	$0 < q < q_c$	$q = q_c$	$q > q_c$
Regime	subcritical	critical	supercritical
Singularity	ρ_H	ρ_H	$\rho < \rho_H$
Singular exponent	Z^{λ_H}	$Z^{\lambda_G\lambda_H}$	Z^{λ_G}
Limit law	discrete	continuous	continuous
	(Boltzmann)	(Mittag-Leffler)	(Gaussian)

▬ *In the subcritical regime $0 < q < q_c$, the random variable $X_n - 1$ converges to a discrete distribution, a Boltzmann distribution $\mathcal{B}_{G'}(q\tau_h)$ with explicit probability generating function given by:*

$$\mathbb{P}(X_n - 1 = k) \to [v^k]\frac{G'(vq\tau_H)}{G'(q\tau_H)}.$$

In particular, if $G(z) = \frac{1}{(1-z)^m}$, the limit law of $X_n - 1$ is a negative binomial distribution $NegBin(m+1, 1-q\tau_H)$, where $X \sim NegBin(r, p)$ is defined by $\mathbb{P}(X = k) = \binom{k+r-1}{k}p^r(1-p)^k$ for $k \geqslant 0$.

- *In the critical regime $q = q_c$, the random variable X_n/n^{λ_H} converges in distribution:*

$$\frac{-c_H X_n}{\tau_H \, n^{\lambda_H}} \xrightarrow{\mathcal{L}} X,$$

  where the random variable X follows the two-parameter Mittag-Leffler distribution $\mathrm{ML}(\alpha, \beta)$ (with $\alpha := \lambda_H$ and $\beta := -\lambda_G \lambda_H$) that is associated with the density $f_X(x) = \frac{\Gamma(\beta+1)}{\alpha \Gamma(\frac{\beta}{\alpha}+1)} \sum_{n=1}^{\infty} \frac{(-1)^n}{n! \Gamma(-n\alpha)} x^{n+\beta/\alpha-1}$ determined by its moments $\mathbb{E}(X^r) = \frac{\Gamma(\beta)\Gamma(r+\beta/\alpha)}{\Gamma(\beta/\alpha)\Gamma(\alpha r+\beta)}$. In particular, for $\lambda_G = -1$ and $\lambda_H = \frac{1}{2}$, X follows the Rayleigh distribution $\mathcal{R}(\sqrt{2})$, where $X \sim \mathcal{R}(\sigma)$ is defined by the density $\frac{x}{\sigma^2} e^{-x^2/(2\sigma^2)}$ for $x \geqslant 0$.

- *In the supercritical regime $q > q_c$, the centred and normalized random variable $(X_n - \mu_n)/\sigma_n$ converges in distribution to a standard normal distribution $\mathcal{N}(0,1)$, where mean μ_n and variance σ_n^2 are both of order n: we have, with $\rho \equiv \rho(q)$ given by $qH(\rho) = \rho_G$,*

$$\mu_n \sim \frac{\rho_G}{q\rho H'(\rho)} \cdot n, \quad \sigma_n^2 \sim \Big(\frac{\rho_G^2}{q^2 \rho^2 H'(\rho)^2} - \frac{\rho_G}{q\rho H'(\rho)} + \frac{\rho_G^2 H''(\rho)}{q^2 \rho H'(\rho)^3} \Big) \cdot n.$$

In particular, the expected value of X_n is for $n \to \infty$ asymptotically equivalent to

$$\mathbb{E}(X_n) \sim \begin{cases} 1 + \frac{q\tau_H G''(q\tau_H)}{G'(q\tau_H)}, & \text{for} \quad 0 < q < q_c, \\ \frac{\lambda_G \tau_H \Gamma(-\lambda_G \lambda_H)}{c_H \Gamma((1-\lambda_G)\lambda_H)} \cdot n^{\lambda_H}, & \text{for} \quad q = q_c, \\ \frac{\rho_G}{q\rho H'(\rho)} \cdot n, & \text{for} \quad q > q_c. \end{cases}$$

Proof of Theorem 6 (Sketch). For $q < q_c$ we are in the subcritical regime and follow the proof of Lemma 5. We build on the results of [6, 25]. We expand $F(z, qv)$ for $0 < v < 1$ to obtain $F(z, qv) \sim G(qv\tau_H) + c_H qvG'(qv\tau_H)\left(1 - \frac{z}{\rho_H}\right)^{\lambda_H}$. This implies that the probability generating function satisfies

$$\lim_{n\to\infty} \mathbb{E}(v^{X_n(q)}) = \lim_{n\to\infty} \frac{[z^n]F(z, qv)}{f_n(q)} = \frac{vG'(qv\tau_H)}{G'(q\tau_H)},$$

leading to a Boltzmann distribution $\mathcal{B}_{G'}(q\tau_H)$. In particular, for $G(z) = 1/(1-z)$ we have $G'(z) = 1/(1-z)^2$, leading to a negative binomial distribution.

For $q = q_c$ we are at the critical value and we thus apply [6, Theorem 4.1] with $\lambda_G < 0$ and $0 < \lambda_H < 1$. This yields the stated limit law, as discussed in [6, Remark 4.2].

In the supercritical regime for $q > q_c$, our claim results from the approach of Bender [25, Propositions IX.6 and IX.7]. The singularity $\rho = \rho(qv)$ becomes an analytic function of v while the nature of the singularity remains unchanged for v in a sufficiently small neighbourhood of 1. In particular, the expected value, the variance, and the normal limit law follow by an application of Hwang's quasi-power theorem [25, 32]. ◀

It is worth pointing out that the constant in the asymptotics of σ_n^2 in the supercritical case is always strictly positive. Thus degenerate limit laws are not possible. This is shown in the following. The proof follows the lines of [35, Proposition 23].

▶ **Proposition 7** (Positivity of the variance constant). *In the supercritical regime $q > q_c$ of Theorem 6, we have*

$$\frac{\rho_G^2}{q^2 \rho^2 H'(\rho)^2} - \frac{\rho_G}{q\rho H'(\rho)} + \frac{\rho_G^2 H''(\rho)}{q^2 \rho H'(\rho)^3} > 0,$$

thus $\frac{\sigma_n^2}{n}$ converges to a constant that is strictly positive.

$$q = 1 \qquad\qquad q = 1.5 \qquad\qquad q = 3$$
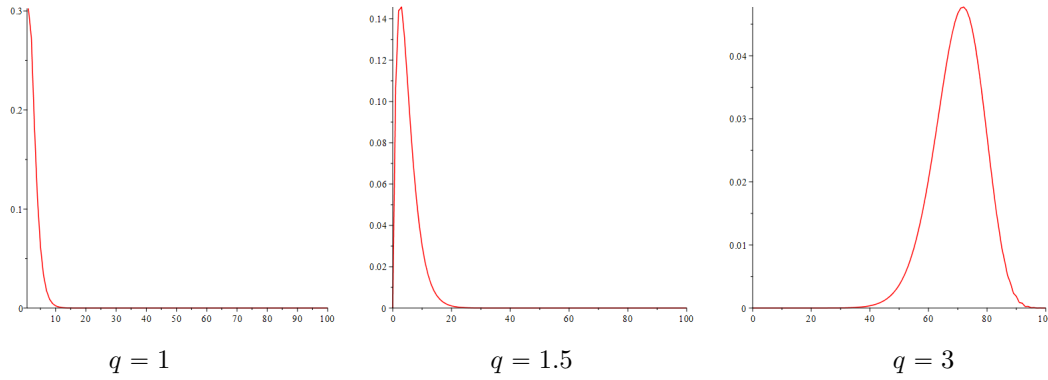
Figure 2 The distribution (with the histogram interpolated to a curve) of returns to 0 in Motzkin excursions (as analysed in Section 3), for walks of length $n = 100$, under the Gibbs measure of weight q. While the distribution for a finite n evolves continuously when one increases q, it exhibits a phase transition when one increases n (at $q_c = 3/2$ for this example). As stated in Theorem 6, it converges to a negative binomial distribution for $q < q_c$, to a Rayleigh distribution for $q = q_c$, and to a Gaussian distribution for $q > q_c$.

Proof. For v in a suitably small neighbourhood of 1, the singularity $\rho = \rho(qv)$ of $F(z, qv)$ becomes an analytic function of v as mentioned in the proof of Theorem 6. It is implicitly determined by $qvH(\rho(qv)) = \rho_G$. Singularity analysis gives us the asymptotic formula

$$f_n(qv) = [z^n]F(z, qv) \sim c_G\Big(\frac{qv\rho(qv)H'(\rho(qv))}{\rho_G}\Big)^{\lambda_G} \frac{1}{\Gamma(-\lambda_G)}\rho(qv)^{-n}n^{-\lambda_G-1};$$

see Lemma 5. For the moment generating function of the variable $X_n(q)$, this means that

$$\mathbb{E}(e^{tX_n(q)}) = \frac{f_n(qe^t)}{f_n(q)} \sim \exp\big(a(t)n + b(t)\big), \tag{7}$$

uniformly for t in a suitable neighbourhood of 0, where

$$a(t) = \log\rho(q) - \log\rho(qe^t), \qquad b(t) = \lambda_G\Big(t + \log\frac{\rho(qe^t)H'(\rho(qe^t))}{\rho(q)H'(\rho(q))}\Big).$$

Note that $a(0) = b(0) = 0$ and $a'(0) = -\frac{q\rho'(q)}{\rho(q)} = \frac{\rho_G}{q\rho(q)H'(\rho(q))}$ (the latter identity follows by implicit differentiation from $qH(\rho(q)) = \rho_G$). Now let k be the smallest positive integer greater than or equal to 2 such that $a^{(k)}(0) \neq 0$, provided that such an integer exists. Substitute $t = sn^{-1/k}$ in (7) and apply the Taylor expansion of $a(t)$ and $b(t)$ around 0 to obtain

$$\mathbb{E}\Big(\exp\Big(\frac{s(X_n(q) - a'(0)n)}{n^{1/k}}\Big)\Big) = \exp\Big(\frac{a^{(k)}(0)}{k!}s^k + o(1)\Big).$$

By Lévy's continuity theorem, this would mean that $\frac{X_n(q) - a'(0)n}{n^{1/k}}$ converges in distribution to a random variable X with moment generating function

$$M(s) = \mathbb{E}(e^{sX}) = \exp\Big(\frac{a^{(k)}(0)}{k!}s^k\Big).$$

However, such a random variable can only exist for $k = 2$: otherwise, it would have second moment 0, thus be almost surely equal to 0 and have moment generating function $M(s) = 1$.

If $k = 2$, then

$$
\begin{aligned}
0 \neq a''(0) &= -\frac{q^2 \rho''(q)}{\rho(q)} + \frac{q^2 \rho'(q)^2}{\rho(q)^2} - \frac{q\rho'(q)}{\rho(q)} \\
&= \frac{\rho_G^2}{q^2 \rho(q)^2 H'(\rho(q))^2} - \frac{\rho_G}{q\rho(q)H'(\rho(q))} + \frac{\rho_G^2 H''(\rho(q))}{q^2 \rho(q)H'(\rho(q))^3},
\end{aligned}
$$

and we are done (here, implicit differentiation of $qH(\rho(q)) = \rho_G$ is used again in the second step). It only remains to exclude the possibility that there is no $k \geqslant 2$ such that $a^{(k)}(0) \neq 0$. Then $a(t)$ must be a linear function: $a(t) = \kappa t$ for some constant κ, thus $\rho(qe^t) = \rho(q)e^{-\kappa t}$. This would have to hold for t in a neighbourhood of 0. But in view of the implicit equation $qe^t H(\rho(qe^t)) = \rho_G$, this would imply that the function H is given by $H(z) = \frac{\rho_G}{q}(z/\rho(q))^{1/k}$, contradicting our assumptions. ◀

3 Applications: phase transitions from negative binomial to Rayleigh to Gaussian

We start our list of applications with the case of fixed-point-biased permutations avoiding a pattern of length three, whose asymptotic behaviour was recently considered in [13]. Then we consider several instructive examples from the theory of lattice paths, namely the returns to zero in Dyck bridges and Dyck excursions, as well as in Motzkin bridges and Motzkin excursions. Thereby, we add to the existing classical results (that is, for the uniform distribution) the phase transitions stemming from q-enumeration (that is, for the Gibbs distribution on the parameter). We also consider the number of boundary interactions in some quarter-plane walk models, and the number of contacts between two paths in some watermelon models.

3.1 Fixed-point-biased permutations avoiding a pattern of length three

Let $\mathcal{S}_n(p)$ be the set of permutations of $1, 2, \ldots, n$ that avoid a given pattern p (where, as usual, the elements of the pattern p need not be contiguous in the permutation; see Figure 3). The generating function counting the statistic $\mathrm{fp}(\sigma)$ (number of fixed points) of such permutations for the pattern 321 was obtained by Vella [41, Theorem 2.13] and for the other two patterns 132 and 213 by Elizalde [19, Theorem 3.5]. It is in all three cases equal to

$$
F(z, u) = 1 + \sum_{n=1}^{\infty} \sum_{\sigma \in \mathcal{S}_n(p)} u^{\mathrm{fp}(\sigma)} z^n = \frac{2}{1 + 2(1-u)z + \sqrt{1-4z}}.
$$
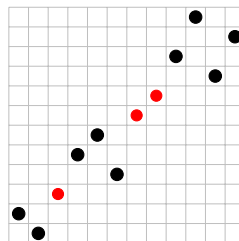
Figure 3 An example of a 321-avoiding permutation of length 12 with 3 fixed points marked by red dots. (A permutation $\pi = (\pi_1, \ldots, \pi_n)$ avoids the pattern $p = 321$ if there is no triplet $1 \leqslant i < j < k \leqslant n$ such that $\pi_k < \pi_j < \pi_i$.)

Later, Chelikavada and Panzo [13] used this generating function to establish a phase transition which we rederive now.

▶ **Theorem 8** (Phase transition for fixed-point-biased permutations)**.** *The limit Gibbs distribution of the fixed-point statistic in permutations avoiding any given pattern $p \in \{132, 321, 213\}$ has a phase transition with critical value $q_c = 3$:*

Parameter q	$0 < q < 3$	$q = 3$	$q > 3$
Limit law of $X_n(q)$	Negative binomial $NegBin(2, 1 - q/3)$	Rayleigh $\mathcal{R}(\sqrt{2})$	Gaussian $\mathcal{N}(0, 1)$

Proof. In order to deduce the phase transitions from Theorem 6, we write $F(z, u)$ as a sequence of components $H(z)$ marked with u

$$F(z, u) = \frac{H(z)}{z} \cdot \frac{1}{1 - uH(z)} = \frac{1}{uz} \cdot \frac{1}{1 - uH(z)} - \frac{1}{uz}, \quad \text{where} \quad H(z) = \frac{2z}{1 + 2z + \sqrt{1 - 4z}}.$$

Here, $F(z, u)$ is not in the shape of the composition scheme (4), but very close. For the limit law, these perturbing factors are irrelevant, since by Lemma 2 one has for $n, k \geqslant 0$

$$\mathbb{P}(X_n(q) = k) = \frac{[z^n u^k]F(z, qu)}{[z^n]F(z, qu)} = \frac{[z^{n+1}u^{k+1}]G(uqH(z))}{[z^{n+1}]G(uqH(z))}, \quad \text{where} \quad G(z) = \frac{1}{1 - z}.$$

Thus, one has $\rho_G = 1$, $\rho_H = \frac{1}{4}$ and $\tau_H = H(\rho_H) = \frac{1}{3}$. Theorem 6 then implies the phase transition at $q_c = \rho_G/\tau_H = 3$ with a Boltzmann distribution (simplifying here to a NegBin), a Mittag-Leffler distribution (simplifying here to a Rayleigh distribution since $\lambda_H = \frac{1}{2}$), and a Gaussian distribution. ◀

3.2 Returns to zero in Dyck and Motzkin paths

We consider two classical directed models: Dyck and Motzkin paths. Dyck paths consist of the steps up $(1, 1)$ and down $(1, -1)$, while Motzkin paths allow additionally some horizontal steps $(1, 0)$. They are called *bridges* if they start at $(0, 0)$ and end at $(2n, 0)$, and *excursions* if it is additionally required that they never cross the x-axis.

The random variable X_n, counting the number of returns to the x-axis in a random excursion and bridge of size $2n$, is a well-studied object [4, 25], leading for excursions to a negative binomial limit law for X_n and leading for bridges to a Rayleigh limit law for $X_n/\sqrt{n}$. Note that the root degree in plane trees behaves the same due to a classical bijection between plane trees and Dyck excursions.

▶ **Theorem 9** (q-enumerations: limit laws for returns to zero)**.** *The Gibbs distribution of the number $X_n = X_n(q)$ of returns to zero in Dyck excursions and bridges of length $2n$ as well as Motzkin excursions and bridges of length n, has a phase transition at q_c and follows, after suitable rescaling for $n \to +\infty$, either a negative binomial, Rayleigh, or Gaussian distribution.*

Parameter q	Limit law
$0 < q < q_c$	$X_n - 1 \xrightarrow{\mathcal{L}} NegBin(2, 1 - q\tau_H)$
$q = q_c$	$\frac{-c_H}{\tau_H} \frac{X_n}{\sqrt{n}} \xrightarrow{\mathcal{L}} Rayleigh(\sqrt{2})$
$q > q_c$	$\frac{X_n - \mu \cdot n}{\sigma \cdot \sqrt{n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$

$$q_c = \begin{cases} 2 & \text{for Dyck excursions,} \\ 1 & \text{for Dyck bridges,} \\ \frac{3}{2} & \text{for Motzkin excursions,} \\ 1 & \text{for Motzkin bridges.} \end{cases}$$

Here, τ_H is $\frac{1}{2}, 1, \frac{2}{3}, 1$, and c_H is $-\frac{1}{2}, -1, -\frac{1}{\sqrt{3}}, -\frac{2}{\sqrt{3}}$ for Dyck excursions, Dyck bridges, Motzkin excursions and Motzkin bridges, respectively.

Proof. Cutting each time the path returns to the x-axis [25, p. 636], one directly sees that the generating functions $D(z)$ and $B_D(z)$ of Dyck excursions and bridges, respectively, satisfy the relations

$$D(z) = \frac{1}{1 - z^2 D(z)} = \frac{1 - \sqrt{1 - 4z^2}}{2z^2} \qquad \text{and} \qquad B_D(z) = \frac{1}{1 - 2z^2 D(z)} = \frac{1}{\sqrt{1 - 4z^2}}.$$

The generating functions $D(z, u)$ and $B_D(z, u)$ of Dyck excursions and bridges marking the number of returns are

$$D(z, u) = \frac{1}{1 - z^2 u D(z)} \qquad \text{and} \qquad B_D(z, u) = \frac{1}{1 - 2z^2 u D(z)}. \tag{8}$$

In both cases we recognize a composition scheme (4) with $G(z) = \frac{1}{1-z}$ and $H(z) = z^2 D(z)$ for excursions and $H(z) = 2z^2 D(z)$ for bridges. Therefore, we can readily apply Theorem 6.

Motzkin excursions have the generating function

$$M(z) = \frac{1 - z - \sqrt{(1 + z)(1 - 3z)}}{2z^2},$$

with dominant singularity $\rho = \frac{1}{3}$ [25]. With the same ideas as for Dyck paths, we directly get the bivariate generating functions $M(z, u)$ and $B_M(z, u)$ of excursions and bridges, respectively, as

$$M(z, u) = \frac{1}{1 - zu(1 + zM(z))} \qquad \text{and} \qquad B_M(z, u) = \frac{1}{1 - zu(1 + 2zM(z))}. \tag{9}$$

Again, we recognize the composition scheme (4) with $G(z) = \frac{1}{1-z}$ and $H(z) = z(1 + zM(z))$ for excursions and $H(z) = z(1 + 2zM(z))$ for bridges. This leads again to similar phase transitions. ◀

▶ Remark 10 (Weighted paths). Let $p_{-1}, p_0, p_1 \geqslant 0$ be the weights of the steps $(1, -1)$, $(1, 0)$, $(1, 1)$, respectively. The weight of a path is the product of its weights. Weighted Dyck excursions and bridges behave exactly as unweighted ones, as each path of length $2n$ has a weight $(p_{-1}p_1)^n$. However, weighted Motzkin excursions and bridges behave differently. With the same techniques it is easy to show that for weighted Motzkin excursions, the phase transition occurs at

$$q_c = \frac{p_0 + 2\sqrt{p_{-1}p_1}}{p_0 + \sqrt{p_{-1}p_1}} = 1 + \frac{1}{1 + \frac{p_0}{\sqrt{p_{-1}p_1}}} \in (1, 2].$$

Finally, for weighted Motzkin bridges the phase transition again always occurs at $q_c = 1$, because $\tau_H = 1$ is independent of the weights.

3.3 Boundary contacts for quarter-plane walks

A direct byproduct of our results are phase transitions for Hadamard models of quarter-plane walks. Beaton, Owczarek, and Rechnitzer [7] initiated the study of quarter-plane walks with wall interactions. We are interested in walks restricted to the quarter plane, starting and ending at the origin, and their interaction with the walls (that is, their number of contacts with x- or y-axis).

It turns out that in many models the generating functions are rather complicated [7], but for three families of walks called *Hadamard models*, the analysis of contacts can be done in a fairly simple way. Such models are enumerated by a *Hadamard product* of generating functions

$$A(z) \odot B(z) := \sum_{n \geqslant 0} a_n b_n z^n, \qquad \text{where } A(z) = \sum_{n \geqslant 0} a_n z^n \quad \text{and} \quad B(z) = \sum_{n \geqslant 0} b_n z^n.$$

They correspond to the diagonal, diabolo, and king walk models with stepsets

$$\boxtimes, \; \boxtimes, \; \boxtimes,$$

respectively (for the last two models, one has in fact a slight variant of the usual diabolo or king models: here we allow additionally the step $(0,0)$, as indicated by the centre dot in the stepset representation). Walks of length $2n$ in the quarter plane that start and end at $(0,0)$ decompose into two independent directed excursions of length n. Therefore, these models are in bijection with pairs of Dyck and Motzkin excursions. This is summarized in the following table, where $D(z,u)$ and $M(z,u)$ are the generating functions from (8) and (9), respectively.

Model	Steps	Generating function $Q(z, u_1, u_2)$	Sequence Q_{2n}	OEIS
Diagonal	$\boxtimes$	$D(z, u_1) \odot D(z, u_2)$	$C_n \cdot C_n$	A001246
Diabolo	$\boxtimes$	$D(z, u_1) \odot M(z, u_2)$	$C_n \cdot M_n$	A151362
King	$\boxtimes$	$M(z, u_1) \odot M(z, u_2)$	$M_n \cdot M_n$	A133053

The limit laws for the number of wall interactions with the x-axis or y-axis depend on the particular values of the q-enumerations; compare with the results in Theorem 9. More precisely, we get the following proposition. By symmetry it also translates to y-axis contacts in the missing cases.

▶ **Theorem 11** (Boundary interactions for some quarter-plane walks). *The number of x-axis contacts of diagonal walks and the number of y-axis contacts of diabolo walks follows the law of the q-enumeration of Dyck returns to 0. The number of x-axis contacts of diagonal, diabolo, and king walks follows the law of the q-enumeration of Motzkin returns to 0. Accordingly, the phase transitions are the same as in Theorem 9.*

Proof (sketch). In the diagonal model, the generating function of the x-axis contacts is equal to $D(z, qu) \odot D(z)$. Therefore, by Lemma 2 we have

$$\mathbb{P}(X_n(q) = k) = \frac{[z^n u^k] D(z, qu) \odot D(z)}{[z^n] D(z, q) \odot D(z)} = \frac{[z^n u^k] D(z, qu)}{[z^n] D(z, q)}.$$

Thus, the result follows directly from Theorem 9. The other models follow in the same fashion.   ◀

▶ Remark 12 (x-axis plus y-axis contacts). The law of the number of x-axis plus y-axis contacts is more involved, as it requires the study of the analytic behaviour of Hadamard products of the shape $(1-z)^a \odot (1-z)^b$. These products were studied by Fill, Flajolet, and Kapur, who gave the corresponding Puiseux expansions; see [23, Proposition 8]. Note that, as pinpointed in [23], the case $a + b$ integer requires the use of additional hypergeometric identities.

3.4 Friendly two-watermelons without wall: contacts and returns

We consider a pair of directed walkers with Dyck steps $(1, -1)$ and $(1, 1)$. The walkers start and end at the same point, may meet and share edges but not cross. Such walker configurations $\mathcal{W}$ are also called friendly two-watermelons (see the related work of Roitner [38] and Krattenthaler, Guttmann, and Viennot [34]). We are interested in walks of length n and the number of contacts C of the two walkers. A contact in a two-watermelon is a point (not counting the starting point) where both paths occupy the same lattice point. (See Figure 4.)

▶ **Theorem 13** (Phase transition for contacts in friendly two-watermelons). *With the renormalizations of Theorem 6, the limit Gibbs distribution of the number of contacts in friendly two-watermelons has a phase transition with critical value $q_c = 4/3$:*

Parameter q	$0 < q < \frac{4}{3}$	$q = \frac{4}{3}$	$q > \frac{4}{3}$
Limit law of $X_n(q)$	Negative binomial	Rayleigh	Gaussian
	$NegBin(2, 1 - \frac{3}{4}q)$	$\mathcal{R}(\sqrt{2})$	$\mathcal{N}(0, 1)$

Proof. Let $F(z, u)$ denote the bivariate generating function of friendly two-watermelons with respect to the number of contacts:

$$F(z, u) = \sum_{w \in \mathcal{W}} z^{|w|} u^{C(w)}.$$

This generating function was determined by Roitner [38], using a reduction to weighted Motzkin paths. It is given by

$$F(z, u) = \frac{1}{1 - u(z^2 W(z) + 2z)}, \qquad W(z) = \frac{1 - 2z - \sqrt{1 - 4z}}{2z^2}. \tag{10}$$

Under the uniform distribution model, Roitner also obtained a discrete limit law for the parameter (number of contacts). We note in passing that closely related problems in families of osculating walkers have been considered before by Bousquet-Mélou [10]. Under the Gibbs distribution model, we recognize a composition scheme with $G(z) = 1/(1 - z)$, $\rho_G = 1$, and $H(z) = z(zW(z) + 2)$ so that $\tau_H = \frac{3}{4}$. Therefore, Theorem 6 applies: we get the 3 phases, with the critical value $q_c = \frac{\rho_G}{\tau_H} = \frac{4}{3}$. ◀
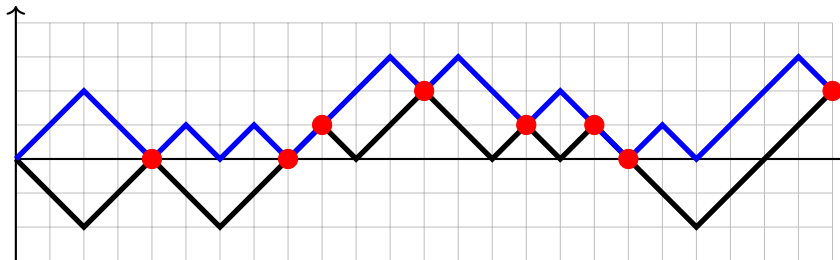
🟨 **Figure 4** An example of a friendly two-watermelon without wall of length 24 with 8 contacts marked by red dots. Under the Gibbs model where such an object is given the weight q^8, the distribution of the number of contacts then depends on the value of q, according to the phase transition given in Theorem 13.

4 Extensions to other constructions: new phase transitions from negative binomial to chi to Gaussian

It is appealing to extend Theorem 6 to even more general composition schemes such as

$$F(z, u) = M(z) \cdot G\big(uH(z)\big), \tag{11}$$

where an additional factor $M(z)$ appears. This extended scheme is of interest as it captures many classical combinatorial structures: some families of trees or lattice paths (meanders), Pólya urns, and other probabilistic processes like the Chinese restaurant model. In its critical phase, this scheme was recently analysed in [6] under the uniform distribution model for the associated combinatorial structures. This extends the work of pioneers like Bender, Flajolet, Soria, Drmota, Hwang, Gourdon [8, 16, 17, 26, 28, 32], later synthesized in [25].

In the long version of this article, we analyse this extended scheme under the Gibbs measure model, and we show that the phase transition for Gibbs distributions leads, in some cases, to the 3-parameter Mittag-Leffler distribution introduced in [6]. Further examples of combinatorial problems involving an extended scheme are the root degree in two-connected outerplanar graphs (see Drmota, Giménez, and Noy [15, Theorem 3.2]), the returns to zero in coloured walks, and the number of wall interactions in watermelons. We now analyse the phase transitions for these last two models.

4.1 Number of wall contacts in watermelons

An m-watermelon of length $2n$ consists of m walkers moving from $(0, 2i - 2)$ to $(2n, 2i - 2)$, $1 \leqslant i \leqslant m$, where every walker may either take an up step $(1, 1)$ or a down step $(1, -1)$, but walkers are not allowed to occupy the same position (they are thus called *vicious*). One says that the watermelon has *a wall* if the x-axis serves as a barrier which the lowest walker may touch but not cross. (See Figure 5.) Watermelons were introduced by Fisher [24] for modelling wetting and melting. We refer the reader to the work of Krattenthaler, Guttmann, and Viennot [33, 34] or Feierl [20–22] for more results on watermelons and related problems.
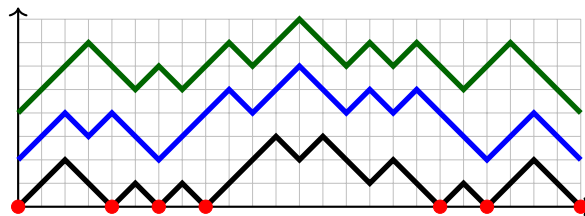
Figure 5 A 3-watermelon (with a wall) of length 24 with 7 x-axis contacts marked by red dots. Under the Gibbs model where such an object is given the weight q^7, the distribution of the number of contacts then depends on the value of q, according to the phase transition given in Theorem 14.

We are interested in the number of x-axis contacts of the lowest walker. For $m = 1$ this reduces to Theorem 9 on Dyck excursions. For $m \geqslant 1$, we get the following theorem.

▶ **Theorem 14** (Phase transition for wall contacts). *The limit Gibbs distribution of the number of x-axis contacts in m-watermelons with a wall has a phase transition at $q_c = 2$:*

Parameter q	$0 < q < 2$	$q = 2$	$q > 2$
Limit law of $X_n(q)$	Negative binomial $NegBin(2m, 1 - \frac{q}{2})$	Chi distribution $\frac{X_n}{\sqrt{n}} \xrightarrow{\mathcal{L}} \chi(2m)$	Gaussian $\mathcal{N}(0, 1)$

Proof (sketch). The q-enumeration of the number of contacts in m-watermelons of length $2n$ ($n > 0$) with a wall was given by Krattenthaler [33, Theorem 4]:

$$f_n(q) = \frac{(n-1)! \prod_{i=0}^{m-1}(2i+1)! \prod_{i=0}^{m-2}(2n+2i)!}{\prod_{i=0}^{2m-2}(n+i)!} \sum_{\ell=2}^{n+1} \binom{2n-\ell}{n-1}\binom{\ell+2m-3}{\ell-2} q^\ell. \tag{12}$$

In the following, we denote by X_n the random variable counting the number of wall contacts, where we drop the dependence on m. Its probability generating function $\mathbb{E}(q^{X_n})$ satisfies

$$\mathbb{E}(q^{X_n}) = \frac{f_n(q)}{f_n(1)}. \tag{13}$$

Let $F(z, q)$ denote the generating function of the numerator of the reduced fraction (13), i.e.,

$$F(z, q) = \sum_{n>0} z^n \sum_{\ell=2}^{n+1} \binom{2n-\ell}{n-1}\binom{\ell+2m-3}{\ell-2} q^\ell. \tag{14}$$

We change the order of summation and shift the index to get

$$F(z, q) = \sum_{\ell \geqslant 2} \binom{\ell+2m-3}{\ell-2} q^\ell \sum_{n \geqslant \ell-1} \binom{2n-\ell}{n-1} z^n = q^2 z \sum_{\ell \geqslant 0} \binom{\ell+2m-1}{\ell} q^\ell z^\ell \sum_{n \geqslant 0} \binom{2n+\ell}{n} z^n.$$

Introducing the Catalan generating function $C(z) = \frac{1-\sqrt{1-4z}}{2z}$, one has

$$\sum_{n \geqslant 0} \binom{2n+\ell}{n} z^n = \frac{C(z)^\ell}{\sqrt{1-4z}}.$$

While such a formula can be proven by convolution identities [29, Eq. (5.72)], it is pleasant to give a bijective proof; we invite the reader to pause here and find it before reading on.

The bijection consists in taking a walk of length $2n + \ell$ ending at altitude ℓ, cutting it at the initial longest bridge, and after this, at the last passage at each altitude. This gives $\sum_{n \geqslant 0} \binom{2n+\ell}{n} z^{2n+\ell} = \frac{1}{\sqrt{1-4z}}(zC(z^2))^\ell$.

Going back to the quest for simplifying $F(z, q)$, we thus obtain

$$F(z, q) = \frac{q^2 z}{\sqrt{1-4z}} \cdot \frac{1}{\left(1 - qzC(z)\right)^{2m}}, \tag{15}$$

where in the last step we used the generating function identity $G(z) = \frac{1}{(1-z)^{2m}} = \sum_{\ell \geqslant 0} \binom{\ell+2m-1}{\ell} z^\ell$. Note that Eq. (15) suggests that there may be a bijective proof of Formula (12) using links with bridges and arches (instead of Krattenthaler's tour de force relying on determinants and jeu de taquin).

Now, in order to get the limit laws, our key observation is that Eq. (15) is a composition scheme of shape $F(z, q) = M(z)G\big(qH(z)\big)$, where $H(z) = zC(z)$, and the probability generating function (under the Gibbs measure) of the number of contacts is given by

$$\mathbb{E}(v^{X_n(q)}) = \frac{[z^{n-1}v^k]F(z, qv)}{[z^{n-1}]F(z, q)}.$$

Next, we just apply singularity analysis with values $\lambda_G = -2m$, $\rho_G = 1$, $\lambda_H = \frac{1}{2}$. The scheme is critical for $q_c = 2$, where one gets the Mittag-Leffler distribution $\frac{1}{\sqrt{2}} \mathrm{ML}(\frac{1}{2}, 2m - \frac{1}{2})$, which can be seen (from its moments) to be the same as the chi distribution $\chi(2m)$. This gives the theorem. ◀

4.2   Returns to zero in coloured walks

Let $m > 0$ be an integer. An m-coloured bridge is an m-tuple $(B_1, \ldots, B_m)$ of (possibly empty) bridges B_i. As a visual representation, we think of them appended one after the other, B_i is coloured in colour i. Note that not all colours need to be present. Let an m-coloured walk be an m-coloured bridge to which a final walk is appended that never returns to the x-axis. See Andrews [2] and [27, 30] for some combinatorial properties of these walks, and links with multicompositions.

We now prove that their number of returns to zero follows a $\chi(m)$ distribution. In particular, this gives the half-normal distribution for $m = 1$ (which extends Theorem 9 to unconstrained walks, see [42]), the Rayleigh distribution for $m = 2$, and the Maxwell distribution for $m = 3$.
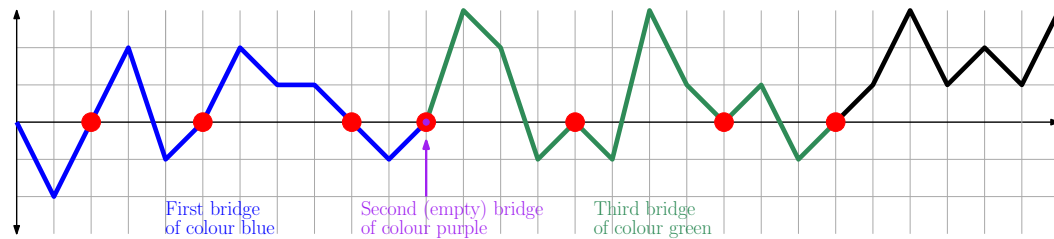
Figure 6 A 3-coloured walk (only 2 of the available 3 colours are present) with 7 returns to zero (red dots).

▶ **Theorem 15** (Phase transitions for returns to zero). *The Gibbs distribution of the number $X_n(q)$ of returns to zero in m-coloured walks of length n has a phase transition at $q_c = 1$ and follows, after suitable rescaling for $n \to +\infty$, either a negative binomial, chi, or Gaussian distribution.*

Parameter q	$0 < q < 1$	$q = 1$	$q > 1$
Limit law of $X_n(q)$	Negative binomial $NegBin(m, 1-q)$	Chi distribution $\chi(m)$	Gaussian $\mathcal{N}(0,1)$

Proof (Sketch). We start by showing that returns to zero in m-coloured walks satisfy the composition scheme $F(z, q) = M(z)G(qA(z))$. First, let $B(z)$ be the generating function of bridges. Each bridge can be decomposed into a sequence of minimal bridges (or generalized arches), which have no return to zero between their extremities (yet jumps might cross the x-axis). Hence, $A(z) = 1 - \frac{1}{B(z)}$. Second, the m-tuple corresponding to an m-coloured bridge has the generating function $G(A(z))$, with $G(z) = \frac{1}{(1-z)^m}$. Third, the final part of the walk (that never returns to the x-axis) corresponds to $M(z) = W(z)/B(z)$, since each walk can be factored into an initial bridge and this final part.

Then, as before, the probability generating function (under the Gibbs measure) is

$$\mathbb{E}(v^{X_n(q)}) = \frac{[z^n v^k]F(z, qv)}{[z^n]F(z, q)}.$$

We get the values $\lambda_G = -m$, $\rho_G = 1$, and $\lambda_M = -1/2$. Furthermore, for any directed walk model [4], $B(z)$ has a critical exponent $\lambda_B = -1/2$, and therefore $\lambda_A = 1/2$. This also means that $B(z)$ diverges at the singularity, and thus $A(z)$ is equal to 1. Hence, the scheme is universally critical at $q_c = 1$, where we get $\frac{1}{\sqrt{2}} \mathrm{ML}(\frac{1}{2}, \frac{m-1}{2})$, which can be seen (from its moments) to be the same as the chi distribution $\chi(m)$. Other cases follow by singularity analysis.   ◀

5   Conclusion

We unified the recent results of [6] with the classical results of [8,25] to obtain phase transitions for Gibbs models under the umbrella of composition schemes. This allows us to obtain a variety of limit laws, leading to new results, as well as summarizing and generalizing classical results in analytic combinatorics. It also explains the universality hidden behind some phase transitions up to now sporadically observed in the literature. In the full version of this article, we give several extensions of Theorem 6, thus treating many other examples of more general combinatorial constructions.

It is interesting to have an informal physicist look at our results: in statistical mechanics, the Gibbs measure can be seen as $q^k = \frac{\exp(-k/T)}{Z(1/T)}$, where T is the temperature of the model. Accordingly, $T \to 0$ (i.e., q is very small) gives a frozen "solid" phase (typically leading to a discrete distribution), while $T \to +\infty$ gives a "gaseous" phase (typically leading to a Gaussian distribution), and, around (or at) a critical temperature T_c, this gives a "liquid" phase (where the wild things are: one often observes at this location an unexpected fancy distribution). Our article is one more illustration of this informal paradigm.

References

1   Florian Aigner. A new determinant for the q-enumeration of alternating sign matrices. *Journal of Combinatorial Theory, Series A*, 180:27 pp., 2021. URL: `https://arxiv.org/abs/1810.08022`.

2   George E. Andrews. The theory of compositions. IV: Multicompositions. *The Mathematics Student*, 2007:25–31, 2007. URL: `https://georgeandrews1.github.io/publications.html`.

3   Cyril Banderier and Michael Drmota. Formulae and asymptotics for coefficients of algebraic functions. *Combinatorics, Probability and Computing*, 24(1):1–53, 2015. URL: `http://lipn.univ-paris13.fr/~banderier/Papers/Alg.pdf`.

4   Cyril Banderier and Philippe Flajolet. Basic analytic combinatorics of directed lattice paths. *Theoretical Computer Science*, 281(1-2):37–80, 2002. `doi:10.1016/S0304-3975(02)00007-5`.

5   Cyril Banderier, Philippe Flajolet, Gilles Schaeffer, and Michèle Soria. Random maps, coalescing saddles, singularity analysis, and Airy phenomena. *Random Structures & Algorithms*, 19(3-4):194–246, 2001. URL: `https://lipn.univ-paris13.fr/~banderier/Papers/rsa.pdf`.

6   Cyril Banderier, Markus Kuba, and Michael Wallner. Phase transitions of composition schemes: Mittag-Leffler and mixed-Poisson distributions. *Annals of Applied probability*, 2024. 53 pp., to appear. `doi:10.48550/arXiv.2103.03751`.

7   Nicholas R. Beaton, A. L. Owczarek, and Andrew Rechnitzer. Exact solution of some quarter plane walks with interacting boundaries. *Electronic Journal of Combinatorics*, 26(3 (P53)), 2019. `doi:10.37236/8024`.

8   Edward A. Bender. Central and local limit theorems applied to asymptotic enumeration. *Journal of Combinatorial Theory, Series A*, 15:91–111, 1973. `doi:10.1016/0097-3165(73)90038-1`.

9   Alexei Borodin, Vadim Gorim, and Eric M. Rains. q-distributions on boxed plane partitions. *Selecta Mathematica*, 16(4):731–789, 2010. `doi:10.1007/s00029-010-0034-y`.

10 Mireille Bousquet-Mélou. Three osculating walkers. *Journal of Physics: Conference Series*, 42:35–46, 2006. `doi:10.1088/1742-6596/42/1/005`.

11 Richard Brak, John W. Essam, and Aleksander L. Owczarek. New results for directed vesicles and chains near an attractive wall. *Journal of Statistical Physics*, 93:155–192, 1998. `doi:10.1023/B:JOSS.0000026731.35385.93`.

12 Richard Brak, John W. Essam, and Aleksander L. Owczarek. Scaling analysis for the adsorption transition in a watermelon network of n directed non-intersecting walks. *Journal of Statistical Physics*, 102:997–1017, 2001. `doi:10.1023/A:1004819507352`.

13 Aksheytha Chelikavada and Hugo Panzo. Limit theorems for fixed point biased permutations avoiding a pattern of length three. *arXiv*, 2023. URL: `https://arxiv.org/abs/2311.04623`.

14 Mihai Ciucu and Christian Krattenthaler. Enumeration of lozenge tilings of hexagons with cut off corners. *Journal of Combinatorial Theory, Series A*, 100:201–231, 2002. `doi:10.1006/jcta.2002.3288`.

15 Michael Drmota, Omer Giménez, and Marc Noy. Degree distribution in random planar graphs. *Journal of Combinatorial Theory. Series A*, 118(7):2102–2130, 2011. `doi:10.1016/j.jcta.2011.04.010`.

16 Michael Drmota and Michèle Soria. Marking in combinatorial constructions: generating functions and limiting distributions. *Theoretical Computer Science*, 144(1-2):67–99, 1995. `doi:10.1016/0304-3975(94)00294-S`.

17 Michael Drmota and Michèle Soria. Images and preimages in random mappings. *SIAM Journal on Discrete Mathematics*, 10(2):246–269, 1997. `doi:10.1137/S0895480194268421`.

18 Philippe Duchon, Philippe Flajolet, Guy Louchard, and Gilles Schaeffer. Boltzmann samplers for the random generation of combinatorial structures. *Combinatorics, Probability and Computing*, 13(4-5):577–625, 2004. `doi:10.1017/S0963548304006315`.

19 Sergi Elizalde. Multiple pattern avoidance with respect to fixed points and excedances. *Electronic Journal of Combinatorics*, 11(1):40 pp., 2004. `doi:10.37236/1804`.

20 Thomas Feierl. The height of watermelons with wall. *Journal of Physics A: Mathematical and Theoretical*, 45(9), 2012. `doi:10.1088/1751-8113/45/9/095003`.

21 Thomas Feierl. The height and range of watermelons without wall. *European Journal of Combinatorics*, 34(1):138–154, 2013. `doi:10.1016/j.ejc.2012.07.021`.

22 Thomas Feierl. Asymptotics for the number of walks in a Weyl chamber of type B. *Random Structures & Algorithms*, 45(2):261–305, 2014. `doi:10.1002/rsa.20467`.

23 James Allen Fill, Philippe Flajolet, and Nevin Kapur. Singularity analysis, Hadamard products, and tree recurrences. *Journal of Computational and Applied Mathematics*, 174(2):271–313, 2005. `doi:10.1016/j.cam.2004.04.014`.

24 Michael E. Fisher. Walks, walls, wetting, and melting. *Journal of Statistical Physics*, 34:667–729, 1984. `doi:10.1007/BF01009436`.

25 Philippe Flajolet and Robert Sedgewick. *Analytic Combinatorics*. Cambridge University Press, 2009. URL: `http://algo.inria.fr/flajolet/Publications/book.pdf`.

26 Philippe Flajolet and Michèle Soria. General combinatorial schemas: Gaussian limit distributions and exponential tails. *Discrete Mathematics*, 114(1-3):159–180, 1993. `doi:10.1016/0012-365X(93)90364-Y`.

27 Manosij Ghosh Dastidar and Michael Wallner. Bijections and congruences involving lattice paths and integer compositions. *arXiv*, 2024. A short version of this work was published in the proceedings of GASCOM 2024, `https://cgi.cse.unsw.edu.au/~eptcs/paper.cgi?GASCom2024.22`. URL: `https://arxiv.org/pdf/2402.17849`.

28 Xavier Gourdon. Largest component in random combinatorial structures. *Discrete Mathematics*, 180(1-3):185–209, 1998. `doi:10.1016/S0012-365X(97)00115-5`.

29 Ronald L. Graham, Donald E. Knuth, and Oren Patashnik. *Concrete Mathematics*. Addison-Wesley Publishing Company, second edition, 1994. URL: `https://www-cs-faculty.stanford.edu/~knuth/gkp.html`.

30 Brian Hopkins and Stéphane Ouvry. Combinatorics of multicompositions. In *Combinatorial and additive number theory IV*, pages 307–321. Springer, 2021. `doi:10.1007/978-3-030-67996-5_16`.

31 Sam Hopkins, Alexander Lazar, and Svante Linusson. On the q-enumeration of barely set-valued tableaux and plane partitions. *European Journal of Combinatorics*, 2021. `doi:10.1016/j.ejc.2023.103760`.

32 Hsien-Kuei Hwang. On convergence rates in the central limit theorems for combinatorial structures. *European Journal of Combinatorics*, 19:329–343, 1998. `doi:10.1006/eujc.1997.0179`.

33   Christian Krattenthaler. Watermelon configurations with wall interaction: exact and asymptotic results. *Journal of Physics: Conference Series*, 42:179–212, 2006. `doi:10.1088/1742-6596/42/1/017`.

34   Christian Krattenthaler, Anthony J. Guttmann, and Xavier G. Viennot. Vicious walkers, friendly walkers and Young tableaux II: with a wall. *Journal of Physics A: Mathematical and General*, 33:8835–8866, 2000. `doi:10.1088/0305-4470/33/48/318`.

35   Sara Kropf and Stephan Wagner. On q-quasiadditive and q-quasimultiplicative functions. *Electronic Journal of Combinatorics*, 24(1):Paper No. 1.60, 22 pages, 2017. `doi:10.37236/6373`.

36   Colin L. Mallows. Non-null ranking models I. *Biometrika*, 44:114–130, 1957. `doi:10.2307/2333244`.

37   Aleksander L. Owczarek and Thomas Prellberg. Exact solution of pulled, directed vesicles with sticky walls in two dimensions. *Journal of Mathematical Physics*, 60:8 pp., 2019. `doi:10.1063/1.5083149`.

38   Valerie Roitner. Contacts and returns in 2-watermelons without wall. *Bulletin of the Institute of Combinatorics and its Applications*, 89, 2020. URL: `http://bica.the-ica.org/Volumes/89/Reprints/BICA2020-01-Reprint.pdf`.

39   Benedikt Stufler. Gibbs partitions: a comprehensive phase diagram. *Annales de l'Institut Henri Poincaré - Probabilités et Statistiques*, 2023. To appear. URL: `https://arxiv.org/abs/2204.06982`.

40   Rami Tabbara, Aleksander L. Owczarek, and Andrew Rechnitzer. An exact solution of two friendly interacting directed walks near a sticky wall. *Journal of Physics A: Mathematical and Theoretical*, 47(1), 2014. `doi:10.1088/1751-8113/47/1/015202`.

41   Antoine Vella. Pattern avoidance in permutations: Linear and cyclic orders. *Electronic Journal of Combinatorics*, 9(2):43 pp., 2003. `doi:10.37236/1690`.

42   Michael Wallner. A half-normal distribution scheme for generating functions. *European Journal of Combinatorics*, 87:103138, 2020. `doi:10.1016/j.ejc.2020.103138`.

43   Shuang Wu. Sachdev–Ye–Kitaev model with an extra diagonal perturbation: phase transition in the eigenvalue spectrum. *Journal of Physics A: Mathematical and Theoretical*, 55(41):415207, 2022. `doi:10.1088/1751-8121/ac93cd`.

Galled Tree-Child Networks

Yu-Sheng Chang ✉ 🆔
Department of Mathematical Sciences, National Chengchi University, Taipei, Taiwan

Michael Fuchs ✉ 🆔
Department of Mathematical Sciences, National Chengchi University, Taipei, Taiwan

Guan-Ru Yu ✉ 🆔
Department of Applied Mathematics, National Sun Yat-sen University, Kaohsiung, Taiwan

---- **Abstract** ---

We propose the class of *galled tree-child networks* which is obtained as intersection of the classes of galled networks and tree-child networks. For the latter two classes, (asymptotic) counting results and stochastic results have been proved with very different methods. We show that a counting result for the class of galled tree-child networks follows with similar tools as used for galled networks, however, the result has a similar pattern as the one for tree-child networks. In addition, we also consider the (suitably scaled) numbers of reticulation nodes of random galled tree-child networks and show that they are asymptotically normal distributed. This is in contrast to the limit laws of the corresponding quantities for galled networks and tree-child networks which have been both shown to be discrete.

1 Introduction

Phylogenetic networks are used to visualize, model, and analyze the ancestor relationship of taxa in reticulate evolution. To make them more relevant for biological applications as well as devise algorithms for them, many subclasses of the class of phylogenetic networks have been proposed; see the comprehensive survey [14]. A lot of recent research work was concerned with fundamental questions such as counting them and understanding the shape of a network drawn uniformly at random from a given class; see, e.g., [2, 3, 4, 8, 9, 11, 12, 10, 13, 15, 16]. Despite this, even counting results are still missing for most of the major classes of phylogenetic networks. Two notable exceptions are tree-child networks and galled networks for which such results have been proved in [11, 12]. In this work, we consider the intersection of these two network classes. We start with some basic definitions and then explain why we find this class interesting.

First, a phylogenetic network is defined as follows.
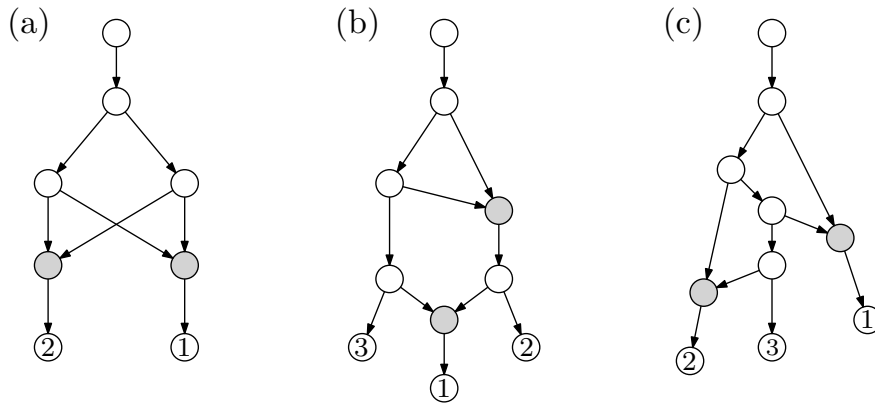
■ **Figure 1** (a) A galled network which is not tree-child; (b) A tree-child network which is not galled; (c) A galled tree-child network.

▶ **Definition 1** (Phylogenetic Network). *A (rooted) phylogenetic network of size n is a rooted, simple, directed, acyclic graph whose nodes fall into the following three (disjoint) categories:*
(a) *A unique root which has indegree 0 and outdegree 1;*
(b) *Leaves which have indegree 1 and outdegree 0 and are bijectively labeled with labels from the set $\{1, \ldots, n\}$;*
(c) *Internal nodes which have indegree and outdegree at least 1 and total degree at least 3. Moreover, a phylogenetic network is called binary if all internal nodes have either indegree 1 and outdegree 2 (tree nodes) or indegree 2 and outdegree 1 (reticulation nodes).*

▶ Remark 2.
 (i) Phylogenetic networks with all internal nodes having indegree equal to 1 are called *phylogentic trees.* They have been used as visualization tool in evolutionary biology at least since Darwin.
 (ii) If not explicitly mentioned, phylogenetic networks are always binary in the sequel.

We next define galled networks and tree-child networks which are two of the major classes of phylogenetic networks. (The former has been introduced for computational reasons, the latter because of its biological relevance; see [14].) For the definition, we need the notion of a *tree cycle* which is a pair of edge-disjoint paths in a phylogenetic network that start at a common tree node and end at a common reticulation node with all other nodes being tree nodes.

▶ **Definition 3.**
(a) *A phylogenetic network is called a tree-child network if every non-leaf node has at least one child which is either a tree node or a leaf.*
(b) *A phylogenetic network is called a galled network if every reticulation node is in a (necessarily unique) tree cycle.*

▶ Remark 4. Note that neither the class of tree-child networks is contained in the class of galled networks nor vice versa; see Figure 1.

Let $\mathrm{TC}_{n,k}$ and $\mathrm{GN}_{n,k}$ denote the number of tree-child networks and galled networks of size n with k reticulation nodes, respectively. It is not hard to see that $k \leq n-1$ for tree-child networks and $k \leq 2n-2$ for galled networks where both bounds are sharp; see, e.g., [11, 12]. Thus, the total numbers are given by:

$$\mathrm{TC}_n := \sum_{k=0}^{n-1} \mathrm{TC}_{n,k} \qquad \text{and} \qquad \mathrm{GN}_n := \sum_{k=0}^{2n-2} \mathrm{GN}_{n,k}. \tag{1}$$

The asymptotic growth of both of these sequences is known. First, in [11], it was proved that for the number of tree-child networks, as $n \to \infty$,

$$\mathrm{TC}_n = \Theta\left(n^{-2/3}e^{a_1(3n)^{1/3}}\left(\frac{12}{e^2}\right)^n n^{2n}\right), \tag{2}$$

where a_1 is the largest root of the Airy function of the first kind. The surprise here was the presence of a *stretched exponential* in the asymptotic growth term. On the other hand, no stretched exponential is contained in the asymptotics of the number of galled networks. More precisely, it was proved in [12] that, as $n \to \infty$,

$$\mathrm{GN}_n \sim \frac{\sqrt{2e\sqrt[4]{e}}}{4}n^{-1}\left(\frac{8}{e^2}\right)^n n^{2n}. \tag{3}$$

The tools used to establish (2) and (3) were very different: for (2), a bijection to a class of words was proved and a recurrence for these word was found which could be (asymptotically) analyzed with the approach from [6]; for (3), the component graph method introduced in [13] together with the Laplace method and a result from [1] was used.

Another difference was the location in (1) of the terms which dominate the two sums. For tree-child networks, the main contribution comes from networks with k close to $n-1$ (the maximally reticulated networks), whereas for galled networks, the main contributions comes from networks with $k \approx n$. In fact, the limit law of the number of reticulation nodes, say R_n, was derived in [5, 12] for both network classes if a network of size n is sampled uniformly at random. More precisely, for tree-child networks, it was shown in [5] that, as $n \to \infty$,

$$n - 1 - R_n \xrightarrow{d} \mathrm{Poisson}(1/2),$$

where $\xrightarrow{d}$ denotes convergence in distribution and $\mathrm{Poisson}(\lambda)$ is a Poisson law with parameter λ. A similar discrete limit law was proved in [12] for galled networks. More precisely, it was shown that, as $n \to \infty$,

$$\mathbb{E}(R_n) = n - \frac{3}{8} + o(1)$$

and that the limit law of $n - R_n$ is not Poisson but a mixture of Poisson laws; see Theorem 2 in [12] for more details.

Due to the above results and differences, one wonders how the intersection of the class of tree-child networks and galled networks behaves?

▶ **Definition 5** (Galled Tree-Child Network). *A galled tree-child network is a network which is both a galled network and a tree-child network.*

Let $\mathrm{GTC}_{n,k}$ denote the number of galled tree-child networks of size n with k reticulation nodes. We show below that again k has the sharp upper bound $n-1$. (See Lemma 19 in Section 3.) Set:

$$\mathrm{GTC}_n := \sum_{k=0}^{n-1} \mathrm{GTC}_{n,k}.$$

Then, this sequence has the following first-order asymptotics.

▶ **Theorem 6.** *For the number of galled tree-child networks, we have, as $n \to \infty$,*

$$\mathrm{GTC}_n \sim \frac{1}{2\sqrt[4]{e}}n^{-5/4}e^{2\sqrt{n}}\left(\frac{2}{e^2}\right)^n n^{2n}.$$

▶ **Remark 7.** Note that the asymptotic expansion contains a stretched exponential as does the expansion (2) for tree-child networks, however, the proof will use the tools which were developed in [12] to derive (3) for galled networks.

We next consider the number of reticulation nodes R_n of a *random galled tree-child network* which is a galled tree-child network of size n that is sampled uniformly at random from the set of all galled tree-child networks of size n. In contrast to tree-child networks and galled networks, the limit law of R_n (suitably scaled) is continuous.

▶ **Theorem 8.** *The number of reticulation nodes R_n of a random galled tree-child networks satisfies, as $n \to \infty$,*

$$\frac{R_n - \mathbb{E}(R_n)}{\sqrt{\text{Var}(R_n)}} \xrightarrow{d} N(0,1),$$

where $N(0,1)$ denotes the standard normal distribution. Moreover, as $n \to \infty$,

$$\mathbb{E}(R_n) = n - \sqrt{n} + o(\sqrt{n}) \qquad and \qquad \text{Var}(R_n) \sim \sqrt{n}/2.$$

The above results show that galled tree-child networks behave quite different from both tree-child networks and galled networks. That is one reason why we find them interesting.

Another reason stems from a recent result which was proved in [4]. In the latter paper, the asymptotics of $\text{GN}_{n,k}$ for fixed k was derived. Let $\text{PN}_{n,k}$ denote the number of phylogenetic networks of size n and k reticulation nodes. (Note that this number is finite, whereas it becomes infinite when summing over k.) Then, one of the main results from [4] implies that for fixed k, as $n \to \infty$,

$$\text{PN}_{n,k} \sim \text{TC}_{n,k} \sim \text{GN}_{n,k} \sim \frac{2^{k-1}\sqrt{2}}{k!} \left(\frac{2}{e}\right)^n n^{n+2k-1}. \tag{4}$$

(The first two asymptotic equivalences were proved in [10, 15].) That $\text{TC}_{n,k}$ and $\text{GN}_{n,k}$ have the same first-order asymptotics for fixed k was a surprise since the classes of tree-child networks and galled networks are quite different, e.g., neither contains the other; see Remark 4. However, the above result can be explained via the class of galled tree-child networks as will be seen in Section 3 below.

We conclude the introduction with a short sketch of the paper. The proofs of Theorem 6 and Theorem 8 follow with a similar approach as used for galled networks in [11]. This approach is based on the component graph method from [13] which we recall in the next section. Then, in Section 3, we consider $\text{GTC}_{n,k}$ for small and large values of k. Finally, Section 4 contains the proofs of our main results (Theorem 6 and Theorem 8). We conclude the paper with some final remarks in Section 5.

2 The Component Graph Method

The component graph method for galled networks was introduced in [13] and used in [4, 12] to prove asymptotic results. It is explained in detail in all these papers. However, to make the current paper more self-contained, we briefly recall it.

Let N be a galled network. Then, by removing all the edges leading to reticulation vertices (these are the so-called *reticulation edges*), we obtain a forest whose trees are called the *tree-components* of N.

The *component graph* of N, denoted by $C(N)$, is now a directed, acyclic graph which has a vertex for every tree-component. Moreover, the vertices are connected by the removed reticulation edges in the same way as the tree-components have been connected by them.
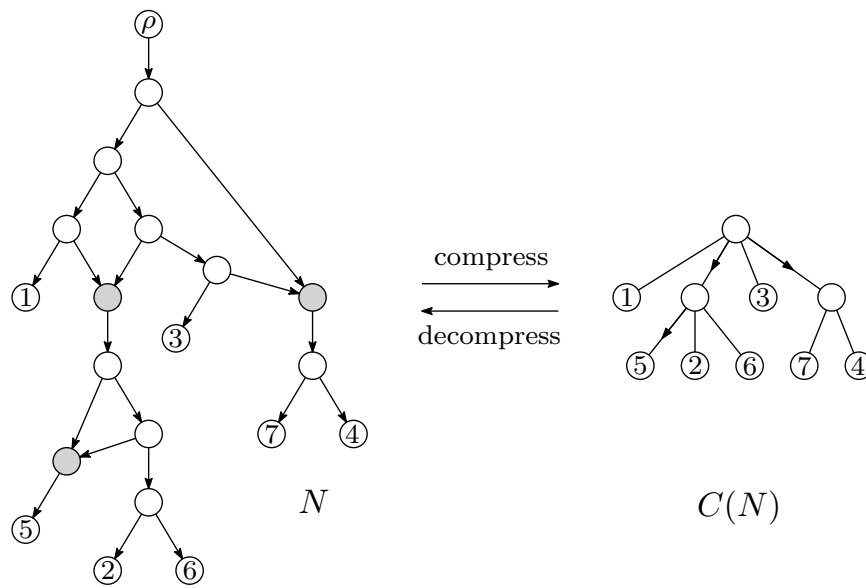
Figure 2 A galled network N and its component graph $C(N)$ which is a phylogenetic tree.

Finally, we attach the leaves in the tree-components to the corresponding vertices in $C(N)$ unless a vertex v of $C(N)$ is a terminal vertex and its corresponding tree-component has exactly one leaf, in which case we use the label of that leaf to label v. Note that $C(N)$ may contain double edges. We replace such a double edge by a single edge and indicate that it was a double edge by placing an arrow on it; see Figure 2 for a galled network together with its component graph. Also, denote by $\tilde{C}(N)$ the component graph of $C(N)$ with all arrows on edges removed. Then, the authors of [13] made the following important observation.

▶ **Proposition 9** ([13]). *N is a galled network if and only if $\tilde{C}(N)$ is a (not necessarily binary) phylogenetic tree.*

▶ Remark 10. By this result, for a galled network N, $C(N)$ must have arrows on all internal edges (i.e., all edges whose two endpoints are both internal nodes).

The component graph can be seen as a kind of compression of N that retains some but not all structural properties of N. Indeed, different networks N might share the same component graph. However, we can generate all galled networks of size n from a list of all component graphs (i.e., phylogenetic trees) with n labeled leaves by a decompression procedure which is explained below.

First, we need the notion of *one-component networks*.

▶ **Definition 11** (One-component Network). *A phylogenetic network is called a one-component network if every reticulation node has a leaf as its child.*

▶ Remark 12. The name comes from the fact that one-component networks only have one non-trivial tree-component.

Now, let a component graph C of a galled tree-child network be given. We do a breadth-first traversal of the internal vertices of C and replace these vertices v by a one-component galled network O_v whose leaves below reticulation vertices are labeled with the first k labels, where k is the number of outgoing edges of v in C that have an arrow on them, and whose size is equal to the outdegree $c(v)$ of v. (In order to avoid confusion, the labels of O_v are

subsequently assumed to be from the set $\{\overline{1}, \ldots, \overline{c(v)}\}$.) Then, attach the subtrees rooted at the children of v which are connected to v by edges with arrows on them to the leaves of O_v with labels $\{\overline{1}, \ldots, \overline{k}\}$, where the subtree with the smallest label is attached to $\overline{1}$, the subtree with the second smallest label is attached to $\overline{2}$, etc. Moreover, relabel the remaining leaves of O_v, namely the ones with the labels $\{\overline{k+1}, \ldots, \overline{c(v)}\}$, by the remaining labels of the subtrees of v (which are all of size 1, i.e., they are leaves in C) in an order-consistent way. By using all possible one-component galled networks in every step, this gives all possible galled networks with C as component graph. Moreover, if we start from $\tilde{C}$, then we first have to place arrows on all edges whose heads are internal nodes of $\tilde{C}$ (see Remark 10) and for all remaining edges, we can freely decide if we want to place an arrow on them or not. Overall, this gives the following result which was one of the main results in [13].

▶ **Proposition 13** ([13]). *We have,*

$$\mathrm{GN}_n = \sum_{\mathcal{T}} \prod_v \sum_{j=0}^{c_{\mathrm{lf}}(v)} \binom{c_{\mathrm{lf}}(v)}{j} M_{c(v), c(v) - c_{\mathrm{lf}}(v) + j},$$

where the first sum runs over all (not necessarily binary) phylogenetic trees $\mathcal{T}$ of size n, the product runs over all internal nodes of $\mathcal{T}$, $c(v)$ is the outdegree of v, $c_{\mathrm{lf}}(v)$ is the number of children of v which are leaves, and $M_{n,k}$ denotes the number of one-component galled networks of size n with k reticulation vertices, where the leaves below the reticulation vertices are labeled with labels from the set $\{1, \ldots, k\}$.

For galled tree-child networks, it is now clear that the same formula holds with the only difference that $M_{n,k}$ has to be replaced by the corresponding number of one-component galled tree-child networks. However, this number is the same as the number of one-component tree-child networks.

▶ **Lemma 14.** *Every one-component tree-child network is a one-component galled tree-child network.*

Proof. Let v be a reticulation vertex and consider a pair of edge-disjoint paths from a common tree vertex to v. (Note that such a pair trivially exists.) Then, no internal vertex can be a reticulation vertex because such a reticulation vertex would not be followed by a leaf. Thus, v is in a tree cycle which shows that the network is indeed galled. ◀

Denote by $B_{n,k}$ the number of one-component tree-child networks of size n and k reticulation vertices, where the labels of the leaves below the reticulation vertices are $\{1, \ldots, k\}$. Then, we have the following analogous result to Proposition 13.

▶ **Proposition 15.** *We have,*

$$\mathrm{GTC}_n = \sum_{\mathcal{T}} \prod_v \sum_{j=0}^{c_{\mathrm{lf}}(v)} \binom{c_{\mathrm{lf}}(v)}{j} B_{c(v), c(v) - c_{\mathrm{lf}}(v) + j}, \tag{5}$$

where notation is as in Proposition 13 and $B_{n,k}$ was defined above.

▶ Remark 16. Using this result, by systematically generating all (not necessarily binary) phylogenetic trees of size n and computing $B_{n,k}$ with the closed-form expression below, we obtain the following table for small values of n:

■ **Table 1** The values of GTC_n for $1 \leq n \leq 10$.

n	GTC_n
1	1
2	3
3	48
4	1,611
5	87,660
6	6,891,615
7	734,112,540
8	101,717,195,895
9	17,813,516,259,420
10	3,857,230,509,496,875

We will deduce all our results from (5). In addition, we make use of the following results for $B_{n,k}$ which were proved in [3] and [11]. To state them, denote by $\mathrm{OTC}_{n,k}$ the number of one-component tree-child networks of size n with k reticulation vertices and by OTC_n the (total) number of one-component tree-child networks of size n. Then,

$$\mathrm{OTC}_{n,k} = \binom{n}{k} B_{n,k} \tag{6}$$

and

$$\mathrm{OTC}_n = \sum_{k=0}^{n-1} \mathrm{OTC}_{n,k}.$$

(Note that the tree-child property implies the $k \leq n-1$ and this bound is sharp.)

▶ **Proposition 17** ([3, 11]).

(i) *We have,*

$$\mathrm{OTC}_{n,k} = \binom{n}{k} \frac{(2n-2)!}{2^{n-1}(n-k-1)!}.$$

(ii) *As $n \to \infty$,*

$$\mathrm{OTC}_{n,k} = \frac{1}{2\sqrt{e}\pi} n^{-3/2} e^{2\sqrt{n}} \left(\frac{2}{e^2}\right)^n n^{2n} e^{-x^2/\sqrt{n}} \left(1 + \mathcal{O}\left(\frac{1 + |x|^3}{n} + \frac{|x|}{\sqrt{n}}\right)\right),$$

where $k = n - \sqrt{n} + x$ and $x = o(n^{1/3})$.

The second result above gives a local limit theorem (see, e.g., Section IX.9 in [7]) for the (random) number of reticulation vertices of a one-component tree-child network of size n which is picked uniformly at random from all one-component tree-child networks of size n. It implies the following (asymptotic) counting result for OTC_n.

▶ **Corollary 18** ([11]). *As $n \to \infty$,*

$$\mathrm{OTC}_n \sim \frac{1}{2\sqrt{e}} n^{-5/4} e^{2\sqrt{n}} \left(\frac{2}{e^2}\right)^n n^{2n}.$$

3   Networks with Few and Many Reticulation Nodes

In this section, we consider $\text{GTC}_{n,k}$ for small and large k. We start with large k.

As mentioned in the last section, for tree-child networks, we have that $k \leq n-1$ and this bound is sharp. Clearly, this implies that $k \leq n-1$ also holds for galled tree-child networks. Again this bound is sharp. We summarize this in the following lemma.

▶ **Lemma 19.** *The number of reticulation vertices of a galled tree-child network of size n is at most $n-1$ where this bound is sharp.*

Proof. Let $\tilde{C}$ be the component graph of a galled tree-child network of size n which by Proposition 9 is a phylogenetic tree. The maximal number of reticulation vertices of a network decompressed from $\tilde{C}$ is achieved by placing the maximal number of arrows at all outgoing edges of internal vertices v of $\tilde{C}$. Note that this number is $c(v) - 1$, where $c(v)$ denotes the degree of v, since placing arrows on all outgoing edges is not possible because $B_{c(v),c(v)} = 0$ (as $B_{n,k}$ denotes the number of certain one-component tree-child networks and $k \leq n-1$). Thus, the maximal number of reticulation vertices equals

$$\sum_v (c(v) - 1) = \sum_v c(v) - (\# \text{ internal nodes of } \tilde{C}), \tag{7}$$

where the sums run over all internal vertices of $\tilde{C}$. By the handshake lemma,

$$\sum_v c(v) = (\# \text{ internal nodes of } \tilde{C} - 1) + n$$

which, by plugging into (7), gives the claimed result.   ◀

The proof of the last lemma also reveals the structure of maximally reticulated galled tree-child networks of size n: They are obtained by decompressing component graphs $\tilde{C}$ that are phylogenetic trees of size n with at least one leaf ℓ attached to every internal vertex v by placing arrows on all outgoing edges of v except the one leading to ℓ. This can be translated into generating functions. Set:

$$M(z) := \sum_{n \geq 1} \text{GTC}_{n,n-1} \frac{z^n}{n!}, \qquad B(z) := \sum_{n \geq 1} B_{n,n-1} \frac{z^n}{n!} = \sum_{n \geq 1} \frac{(2n-2)!}{2^{n-1} n!} z^n,$$

where the last line follows from (6) and Proposition 17-(i). Then, we have the following result.

▶ **Lemma 20.** *We have,*

$$M(z) = z + zB'(M(z)). \tag{8}$$

Proof. According to the explanation in the paragraph preceding the lemma, a maximally reticulated galled tree-child network is either a leaf or obtained from a maximally reticulated one-component tree-child network with the leaves below the reticulation vertices replaced by maximally reticulated galled tree-child networks. This translates into

$$M(z) = z + \sum_{n \geq 1} B_{n,n-1} \frac{z M(z)^{n-1}}{(n-1)!},$$

where the z inside the sum counts the leaf which is not below the reticulation vertex and the factor $1/(n-1)!$ discards the order of the maximally reticulated galled tree-child networks (counted by $M(z)^{n-1}$) which are attached to the children below the reticulation vertices. The claimed result follows from this.   ◀

Note that (8) is of *Lagrangian type*. Thus, we can obtain the asymptotics of $\text{GTC}_{n,n-1}$ by applying Lagrange's inversion formula and the following result from [1].

▶ **Theorem 21** ([1]). *Let $S(z)$ be a formal power series with $s_0 = 0, s_1 \neq 0$ and $ns_{n-1} \sim \gamma s_n$. Then, for $\alpha \neq 0$ and β real numbers,*

$$[z^n](1 + S(z))^{\alpha n + \beta} \sim \alpha e^{\alpha s_1 \gamma} n s_n.$$

▶ **Theorem 22.** *The number of maximally reticulated galled tree-child networks $\text{GTC}_{n,n-1}$ satisfies, as $n \to \infty$,*

$$\text{GTC}_{n,n-1} \sim \sqrt{e\pi}n^{-1/2}\left(\frac{2}{e^2}\right)^n n^{2n}.$$

▶ **Remark 23.** For tree-child networks, it was proved in [11] that $\text{TC}_n = \Theta(\text{TC}_{n,n-1})$. (This was a main step in the proof of (2).) The above result together with Theorem 6 shows that the same is not true for galled tree-child networks.

Proof. Applying the Lagrange inversion formula to (8) gives

$$\text{GTC}_{n,n-1} = n![z^n]M(z) = (n-1)![\omega^{n-1}](1 + B'(\omega))^n. \tag{9}$$

Next, by Stirling's formula, as $n \to \infty$,

$$[z^n]B'(z) = \frac{B_{n+1,n}}{n!} = \frac{(2n)!}{2^n n!} \sim \sqrt{2}\left(\frac{2}{e}\right)^n n^n.$$

Thus, we can apply Theorem 21 to (9) with $\gamma = 1/2$ and obtain that, as $n \to \infty$,

$$\text{GTC}_{n,n-1} \sim \sqrt{e}nB_{n,n-1} = \sqrt{e}n\frac{(2n-2)!}{2^{n-1}} \sim \sqrt{e\pi}n^{-1/2}\left(\frac{2}{e^2}\right)^n n^{2n}.$$

This is the claimed result. ◀

We next consider $\text{GTC}_{n,k}$ with k small, i.e., the other extreme case of the number of reticulation vertices. Here, we have the following result which shows that the distribution of a uniformly chosen phylogenetic network with n leaves and k reticulation nodes concentrates on the set of galled tree-child networks. This explains why the asymptotic expansions of $\text{TC}_{n,k}$ and $\text{GN}_{n,k}$ in (4) are the same. (It would be interesting to know whether or not this distribution concentrates on an even smaller set.)

▶ **Theorem 24.** *For fixed k, as $n \to \infty$,*

$$\text{GTC}_{n,k} \sim \frac{2^{k-1}\sqrt{2}}{k!}\left(\frac{2}{e}\right)^n n^{n+2k-1}. \tag{10}$$

The proof of this result uses ideas from [10].

Proof. First consider galled tree-child networks of size n which are obtained by decompressing phylogenetic trees of size n which have all k arrows on the edges from the root, i.e., the root has at least one leaf and all other children are either internal nodes or leaves (with at most k internal nodes) and all internal nodes have just leaves as children. By Proposition 8 in [10], the number of these galled tree-child network has the same asymptotics as the one on the right-hand side of (10). Moreover, these networks also dominate the asymptotics in the case of tree-child networks. Thus, the remaining galled tree-child networks are asymptotically negligible as their number is bounded above by the number of the remaining tree-child networks. ◀

▶ Remark 25. Note that this re-proves the (surprising) asymptotic result for $\mathrm{GN}_{n,k}$ in (4) from [4]. On the other hand, the above asymptotic result could be also deduced from (4). In order to explain this, denote by $\mathcal{PN}_{n,k}$ (resp. $\mathcal{TC}_{n,k}/\mathcal{GN}_{n,k}/\mathcal{GTC}_{n,k}$) the set of all phylogenetic networks (resp. tree-child networks/galled networks/galled tree-child networks) with n leaves and k reticulation nodes. Then,

$$
\begin{aligned}
|\mathcal{TC}_{n,k} \cup \mathcal{GN}_{n,k}| &= |\mathcal{TC}_{n,k}| + |\mathcal{GN}_{n,k}| - |\mathcal{TC}_{n,k} \cap \mathcal{GN}_{n,k}| \\
&= \mathrm{TC}_{n,k} + \mathrm{GN}_{n,k} - \mathrm{GTC}_{n,k}
\end{aligned}
$$

and $|\mathcal{TC}_{n,k} \cup \mathcal{GN}_{n,k}| \leq \mathrm{PN}_{n,k}$. From this the asymptotic result for $\mathrm{GTC}_{n,k}$ follows from those of (4). (We are thankful to one of the reviewers for this remark.)

4 Proof of the Main Results

In this section, we first prove Theorem 6 and then state a result which implies Theorem 8.

For the proof of Theorem 6, we closely follow the method of proof of (3) from [12]. The main idea is to use (5) to find asymptotic matching upper and lower bounds for GTC_n.

First, for an upper bound, we pick a (not necessarily binary) phylogenetic tree $\mathcal{T}$ of size n (which is considered to be a component graph of a galled tree-child network of size n) and decompress it by picking for internal vertices v of $\mathcal{T}$ *any* one-component tree-child network of size $c(v)$ (where the notation is as in Proposition 13). Since, as explained in Section 2, actually only certain one-component tree-child networks are permissible, this modified decompression procedure overcounts the number of galled tree-child networks of size n. More precisely, we consider

$$
U_n := \sum_{\mathcal{T}} \prod_v \mathrm{OTC}_{c(v)},
$$

where the first sum runs over all phylogenetic trees $\mathcal{T}$ of size n and the product runs over internal vertices of $\mathcal{T}$. Then, we have $\mathrm{GTC}_n \leq U_n$. Next, set

$$
U(z) := \sum_{n \geq 1} U_n \frac{z^n}{n!}, \qquad A(z) := \sum_{n \geq 1} \mathrm{OTC}_{n+1} \frac{z^n}{(n+1)!}.
$$

Then, the definition of U_n implies the following result.

▶ **Lemma 26.** *We have,*

$$
U(z) = z + U(z)A(U(z)).
$$

Proof. The networks counted by U_n are either a leaf or a one-component tree-child network with n leaves which are replaced by an unordered sequence of networks of the same type. This gives

$$
U(z) = z + \sum_{n \geq 2} \mathrm{OTC}_n \frac{U(z)^n}{n!}
$$

from which the claimed result follows. ◀

Now, we can proceed as in the proof of Theorem 22 to obtain the following asymptotic result for U_n.

▶ **Proposition 27.** *As $n \to \infty$,*

$$U_n \sim \frac{1}{2\sqrt[4]{e}} n^{-5/4} e^{2\sqrt{n}} \left(\frac{2}{e^2}\right)^n n^{2n}.$$

Proof. From Lemma 26 and the Lagrange inversion formula,

$$U_n = n![z^n]U(z) = (n-1)![\omega^{n-1}](1 - A(\omega))^{-n}.$$

The result follows from this by applying Theorem 21 and Corollary 18.   ◀

Next, we need a matching lower bound. Therefore, we consider (5) with the first sum restricted to phylogenetic trees of the shape (where we have removed the leaf labels):
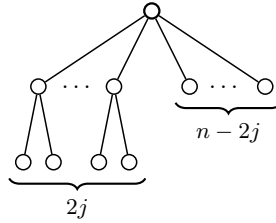
We denote the resulting term by L_n. The decompression procedure from Section 2 then gives the following result.

▶ **Lemma 28.** *We have,*

$$L_n = \sum_{j=0}^{\lfloor n/2 \rfloor} \binom{n}{2j} \frac{(2j)!}{j!2^j} \sum_{\ell=0}^{n-2j} \binom{n-2j}{\ell} L_{n-j,j+\ell}$$

$$= \sum_{j=0}^{\lfloor n/2 \rfloor} \binom{n}{2j} \frac{(2j)!}{j!2^j} \sum_{\ell=0}^{n-2j} \binom{n-2j}{\ell} \frac{(2n-2j-2)!}{2^{n-j-1}(n-2j-\ell-1)!}. \tag{11}$$

Proof. The first equality is explained as in the proof of Lemma 9 in [12] and the second equality follows from (6) and Proposition 17-(i).   ◀

From this result, we can deduce (matching) first-order asymptotics for L_n which then together with the asymptotics of the upper bound (Proposition 27) concludes the proof of Theorem 6.

▶ **Proposition 29.** *As $n \to \infty$,*

$$L_n \sim \frac{1}{2\sqrt[4]{e}} n^{-5/4} e^{2\sqrt{n}} \left(\frac{2}{e^2}\right)^n n^{2n}.$$

Sketch of the proof. From Stirling's formula (similar to the proof of Proposition 17-(ii)),

$$\binom{n-2j}{\ell} \frac{(2n-2j-2)!}{2^{n-j-1}(n-2j-\ell-1)!} \sim \frac{1}{2^{j+1}\sqrt{e\pi}} n^{-3/2} e^{2\sqrt{n}} \left(\frac{2}{e^2}\right)^j n^{2n-2j} e^{-x^2/\sqrt{n}},$$

where $k = n - \sqrt{n} + x$ and this holds uniformly for $|x| \leq n^{1/2+\epsilon}$ and $j \leq n^\epsilon$ with $\epsilon > 0$ arbitrarily small. Using the Laplace method then gives,

$$\sum_{\ell=0}^{n-2j} \binom{n-2j}{\ell} \frac{(2n-2j-2)!}{2^{n-j-1}(n-2j-\ell-1)!} \sim \frac{1}{2^{j+1}\sqrt{e}} n^{-5/4} e^{2\sqrt{n}} \left(\frac{2}{e^2}\right)^n n^{2n-2j}$$

uniformly for $j \leq n^\epsilon$ for arbitrarily small $\epsilon > 0$. Finally, by plugging the last relation into (11),

$$L_n \sim \frac{1}{2\sqrt{e}} \left(\sum_{j \geq 0} \frac{1}{j!4^j} \right) n^{-5/4} e^{2\sqrt{n}} \left(\frac{2}{e^2} \right)^n n^{2n}$$

which gives the claimed result.   ◀

▶ **Remark 30.** Note that this proposition shows that a "typical" galled tree-child network of size n is obtained by decompressing component graphs of the form given before Lemma 28. This implies, e.g., that the Sackin index defined in [17] of a galled tree-child network has the unusual expected order $n^{7/4}$.

Finally, by refining the above method (see Section 6 of [12] where the same was done for galled networks), we obtain the following result which implies our second main result (Theorem 8).

▶ **Theorem 31.** *Let I_n be the number of reticulation vertices of a random galled tree-child network of size n which are not followed by a leaf and R_n be the total number of reticulation vertices. Then, as $n \to \infty$,*

$$\left(I_n, \frac{R_n - n + \sqrt{n}}{\sqrt[4]{n/4}} \right) \xrightarrow{d} (I, R),$$

where I and R are independent with $I \overset{d}{=} \mathrm{Poisson}(1/4)$ and $R \overset{d}{=} N(0,1)$.

5   Conclusion

In this paper, we introduced the class of *galled tree-child networks* which is obtained as intersection of the classes of galled networks and tree-child networks. Our reason for doing so was two-fold: (i) Different tools have been used to prove results for galled networks and tree-child networks [11, 12]; consequently, we were curious about which tools apply to the combination of these classes? (ii) It was recently proved that the number of galled networks and tree-child networks have the same first-order asymptotics when the number of reticulation vertices is fixed [4, 10]. Why is that the case?

As for (i), we showed that an asymptotic counting result for galled tree-child networks (Theorem 6) can be obtained with the methods for galled networks, however, the result contains a stretched exponential as does the asymptotic result for tree-child networks. In addition, we showed that the number of reticulation vertices for a random galled tree-child networks is asymptotically normal (Theorem 8), whereas the limit laws of the same quantities for galled networks and tree-child networks were discrete. As for (ii), we showed that the number of galled tree-child networks also satisfies the same first order asymptotics when the number of reticulation vertices is fixed. This explains the previous results from [4, 10].

Overall, the class of galled tree-child networks is interesting and thus merits further examination. In particular, due to Remark 30, studying the shape of random galled tree-child networks seems to be more feasible than studying the shape of random networks from other network classes because such a study boils down to the easier task of studying the shape of one-component tree-child networks which have a straightforward recursive decomposition that, e.g., resulted in a closed-form expression for their numbers; see [17]. The latter paper, where one-component tree-child networks are called *simplex networks*, e.g., asks for properties of the height and such results would immediately entail corresponding results for random galled tree-child networks. (Studying the height is an open problem for most classes of phylogenetic networks.) We may come back to this question in future work.

References

1 Edward A. Bender and L. Bruce Richmond. An asymptotic expansion for the coefficients of some power series. II. Lagrange inversion. *Discrete Math.*, 50(2-3):135–141, 1984. `doi: 10.1016/0012-365X(84)90043-8`.

2 Mathilde Bouvel, Philippe Gambette, and Marefatollah Mansouri. Counting phylogenetic networks of level 1 and 2. *J. Math. Biol.*, 81(6-7):1357–1395, 2020. `doi:10.1007/ s00285-020-01543-5`.

3 Gabriel Cardona and Louxin Zhang. Counting and enumerating tree-child networks and their subclasses. *J. Comput. System Sci.*, 114:84–104, 2020. `doi:10.1016/j.jcss.2020.06.001`.

4 Yu-Sheng Chang and Michael Fuchs. Counting phylogenetic networks with few reticulation vertices: Galled and recticulation-visible networks. *Bull. Math. Biol.*, 86:Paper No. 76, 2024. `doi:10.1007/s11538-024-01309-w`.

5 Yu-Sheng Chang, Michael Fuchs, Hexuan Liu, Michael Wallner, and Guan-Ru Yu. Enumerative and distributional results for *d*-combining tree-child networks. *Adv. in Appl. Math.*, 157:Paper No. 102704, 2024. `doi:10.1016/j.aam.2024.102704`.

6 Andrew Elvey Price, Wenjie Fang, and Michael Wallner. Compacted binary trees admit a stretched exponential. *J. Combin. Theory Ser. A*, 177:Paper No. 105306, 40, 2021. `doi: 10.1016/j.jcta.2020.105306`.

7 Philipp Flajolet and Robert Sedgewick. *Analytic Combinatorics*. Cambridge University Press, 2009.

8 Michael Fuchs, Bernhard Gittenberger, and Marefatollah Mansouri. Counting phylogenetic networks with few reticulation vertices: tree-child and normal networks. *Australas. J. Combin.*, 73:385–423, 2019.

9 Michael Fuchs, Bernhard Gittenberger, and Marefatollah Mansouri. Counting phylogenetic networks with few reticulation vertices: exact enumeration and corrections. *Australas. J. Combin.*, 81:257–282, 2021.

10 Michael Fuchs, En-Yu Huang, and Guan-Ru Yu. Counting phylogenetic networks with few reticulation vertices: a second approach. *Discrete Appl. Math.*, 320:140–149, 2022. `doi:10.1016/j.dam.2022.03.026`.

11 Michael Fuchs, Guan-Ru Yu, and Louxin Zhang. On the asymptotic growth of the number of tree-child networks. *European J. Combin.*, 93:Paper No. 103278, 20, 2021. `doi:10.1016/j. ejc.2020.103278`.

12 Michael Fuchs, Guan-Ru Yu, and Louxin Zhang. Asymptotic enumeration and distributional properties of galled networks. *J. Combin. Theory Ser. A*, 189:Paper No. 105599, 28, 2022. `doi:10.1016/j.jcta.2022.105599`.

13 Andreas D. M. Gunawan, Jeyaram Rathin, and Louxin Zhang. Counting and enumerating galled networks. *Discrete Appl. Math.*, 283:644–654, 2020. `doi:10.1016/j.dam.2020.03.005`.

14 Sungsik Kong, Joan Carles Pons, Laura Kubatko, and Kristina Wicke. Classes of explicit phylogenetic networks and their biological and mathematical significance. *J. Math. Biol.*, 84(6):Paper No. 47, 44, 2022. `doi:10.1007/s00285-022-01746-y`.

15 Marefatollah Mansouri. Counting general phylogenetic networks. *Australas. J. Combin.*, 83:40–86, 2022.

16 Miquel Pons and Josep Batle. Combinatorial characterization of a certain class of words and a conjectured connection with general subclasses of phylogenetic tree-child networks. *Sci. Rep.*, 11:Article number: 21875, 2021.

17 Louxin Zhang. The Sackin index of simplex networks. In *Comparative genomics*, volume 13234 of *Lecture Notes in Comput. Sci.*, pages 52–67. Springer, Cham, 2022. `doi:10.1007/ 978-3-031-06220-9_4`.

On Fluctuations of Complexity Measures for the FIND Algorithm

Jasper Ischebeck[1] ✉ 🏠 iD
Institut für Mathematik, Goethe University Frankfurt, Germany

Ralph Neininger ✉ 🏠 iD
Institut für Mathematik, Goethe University Frankfurt, Germany

── **Abstract** ──────────────────

The FIND algorithm (also called Quickselect) is a fundamental algorithm to select ranks or quantiles within a set of data. It was shown by Grübel and Rösler that the number of key comparisons required by FIND as a process of the quantiles $\alpha \in [0, 1]$ in a natural probabilistic model converges after normalization in distribution within the càdlàg space $D[0, 1]$ endowed with the Skorokhod metric. We show that the process of the residuals in the latter convergence after normalization converges in distribution to a mixture of Gaussian processes in $D[0, 1]$ and identify the limit's conditional covariance functions. A similar result holds for the related algorithm QuickVal. Our method extends to other cost measures such as the number of swaps (key exchanges) required by FIND or cost measures which are based on key comparisons but take into account that the cost of a comparison between two keys may depend on their values, an example being the number of bit comparisons needed to compare keys given by their bit expansions.

1 Introduction

In 1961, Hoare [11] introduced the algorithm FIND, also called Quickselect, to select a key (an element) of a given rank from a linearly ordered finite set of data. We assume that the data are distinct real numbers. To be definite a simple version of the FIND algorithm is given as follows: $\mathrm{FIND}(S, k)$ has as input a set $S = \{s_1, \ldots, s_n\}$ of distinct real numbers of size n and an integer $1 \leq k \leq n$. The algorithm FIND operates recursively as follows: If $n = 1$ we have $k = 1$ and FIND returns the single element of S. If $n \geq 2$ and $S = \{s_1, \ldots, s_n\}$ the algorithm first chooses an element from S, say s_j, called pivot, and generates the sets

$$S_< := \{s_i \,|\, s_i < s_j, i \in \{1, \ldots, n\} \setminus \{j\}\}, \quad S_\geq := \{s_i \,|\, s_i \geq s_j, i \in \{1, \ldots, n\} \setminus \{j\}\}.$$

If $k = |S_<| + 1$, the algorithm returns s_j. If $k \leq |S_<|$, recursively $\mathrm{FIND}(S_<, k)$ is applied. If $k \geq |S_<| + 2$, recursively $\mathrm{FIND}(S_\geq, k - |S_<| - 1)$ is applied. Note that $\mathrm{FIND}(S, k)$ returns the element of rank k from S. There are various variants of the algorithm, in particular regarding how the pivot element is chosen and how S is partitioned into the subsets $S_<$ and $S_\geq$.

In a standard probabilistic model one assumes that the data are ordered, i.e. given as a vector $(s_1, \ldots, s_n)$, and are randomly permuted, all permutations being equally likely. This can be achieved assuming that the data are given as $(U_1, \ldots, U_n)$ where $(U_j)_{j \in \mathbb{N}}$ is a sequence

---

[1] corresponding author

of i.i.d. random variables with distribution unif$[0, 1]$, the uniform distribution over the unit interval $[0, 1]$. This is the probabilistic model considered below. Note that the randomness is within the data, while the algorithm is deterministic.

Various cost measures have been considered for FIND, mainly the number of key comparisons required which we analyze in detail below. At the end of this extended abstract we state related results for the number of swaps (key exchanges) required and for cost measures which are based on key comparisons, where the cost of a comparison may depend on the values of the keys s_i, s_j, the number of bit comparisons required to decide whether $s_i < s_j$ or not being a prominent example.

For analysis purposes a related process, called QuickVal, has been considered, see [19, 8]. Informally, QuickVal for an $\alpha \in [0, 1]$ mimics FIND to select (or to try to select) the value α from the set of data, which, in our probabilistic model for large n, comes close to FIND selecting rank $\lfloor \alpha n \rfloor$. To be definite, QuickVal$((U_1, \ldots, U_n), \alpha)$ compares the U_i with U_1 to generate sublists

$$S_< := (U_{j_1}, \ldots, U_{j_{m-1}}), \quad S_\geq := (U_{j_{m+1}}, \ldots, U_{j_n}),$$

with $U_{j_i} < U_1$ for $i = 1, \ldots, m-1$ and $2 \leq j_1 < \cdots < j_{m-1}$ and $U_{j_i} \geq U_1$ for $i = m+1, \ldots, n$ and $2 \leq j_{m+1} < \cdots < j_n$. Hence, $m - 1 \in \{0, \ldots, n-1\}$ is the number of the U_i, $2 \leq i \leq n$, being smaller than U_1. The algorithm recursively calls QuickVal$(S_<, \alpha)$ if $\alpha < U_1$ and $|S_<| > 0$. If $\alpha \geq U_1$ and $|S_\geq| > 0$ recursively QuickVal$(S_\geq, \alpha - U_1)$ is called. The number of key comparisons required by QuickVal$((U_1, \ldots, U_n), \alpha)$ is denoted by $S_{\alpha,n}$.

To describe the processes $(S_{\alpha,n})_{\alpha \in [0,1]}$ and their limit (after scaling) conveniently we also consider the binary search tree constructed from the the data $(U_i)_{i \in \mathbb{N}}$. Part of the following definitions are depicted in Figure 1. The data are inserted into the rooted infinite binary tree, where we denote its nodes by the elements of $\{0, 1\}^* := \cup_{n=0}^\infty \{0, 1\}^n$ as follows. Its root is denoted by the empty word ϵ and for each node $\phi \in \{0, 1\}^*$ we denote by $\phi 0$ and $\phi 1$ (the word ϕ appended with a 0 resp. 1) its left and right child respectively. Moreover $|\phi|$ denotes the length of the word ϕ, which is the depth of the corresponding node in the tree. To construct the binary search tree for $(U_1, \ldots, U_n)$ the first key U_1 is inserted into the root and occupies the root. Then, successively the following keys are inserted, where each key traverses the already occupied nodes starting at the root as follows: Whenever the key traversing is less than the occupying key at a node it moves on to the left child of that node, otherwise to its right child. The first empty node found is occupied by the key.

To describe the costs of the algorithms we organize, using notation of Fill and Nakama [8], the sub-intervals $([L_\phi, R_\phi))_{\phi \in \{0,1\}^*}$ implicitly generated starting with $[0, 1) =: [L_\epsilon, R_\epsilon)$ and recursively setting

$$\tau_\phi := \inf\{i \in \mathbb{N} \mid L_\phi < U_i < R_\phi\},$$
$$L_{\phi 0} := L_\phi, \quad R_{\phi 1} := R_\phi, \quad L_{\phi 1} := R_{\phi 0} := U_{\tau_\phi}, \quad I_\phi := R_\phi - L_\phi. \tag{1}$$

Now, if a sublist starting with pivot U_{τ_ϕ} has to be split by QuickVal, the keys which are inserted in the subtree rooted at U_{τ_ϕ} need to be compared with U_{τ_ϕ}. Hence, we get a contribution of key comparisons of

$$S_{\phi,n} = \sum_{\tau_\phi < k \leq n} \mathbf{1}_{[L_\phi, R_\phi)}(U_k). \tag{2}$$
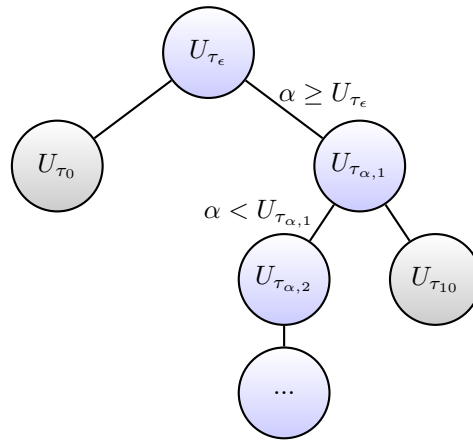
Figure 1 Part of the binary search tree. The pivots of sublists split by QuickVal$((U_1, \ldots, U_n), \alpha)$ for some $\alpha \in [0, 1]$ are on the path indicated. Note that we have $\tau_\epsilon = \tau_{\phi(\alpha, 0)} = \tau_{\alpha, 0} = 1$ and in this example $\alpha \geq U_1$ and $\alpha < U_{\tau_{\alpha, 1}}$ so that $\phi(\alpha, 2) = 10 \in \{0, 1\}^2$.

Now, for $\alpha \in [0, 1]$, QuickVal$((U_1, \ldots, U_n), \alpha)$ generates and splits sublists encoded by $\phi(\alpha, k)$ for $k = 0, 1, \ldots$ for which we obtain by $\phi(\alpha, 0) = \epsilon$ and

$$\phi(\alpha, k+1) = \begin{cases} \phi(\alpha, k)0, & \text{if } \alpha < U_{\tau_{\phi(\alpha, k)}}, \\ \phi(\alpha, k)1, & \text{if } \alpha \geq U_{\tau_{\phi(\alpha, k)}}. \end{cases} \tag{3}$$

When using the variables defined in (1) or (2), we abbreviate the notation $\phi(\alpha, k)$ by α, k, such as writing $I_{\alpha, k} := I_{\phi(\alpha, k)}$ or $S_{\alpha, k, n} := S_{\phi(\alpha, k), n}$.

The number of key comparisons required by QuickVal$((U_1, \ldots, U_n), \alpha)$ is then given by the (finite) sum

$$S_{\alpha, n} = \sum_{k=1}^{\infty} S_{\alpha, k, n}.$$

Fill and Nakama [8, Theorem 3.2] showed (considering more general complexity measures) that for each $\alpha \in [0, 1]$ almost surely

$$\frac{1}{n} S_{\alpha, n} \to S_\alpha := \sum_{k=0}^{\infty} I_{\alpha, k}, \quad (n \to \infty). \tag{4}$$

The latter convergence also holds in L_p, see Fill and Matterer [7, Proposition 6.1].

We take the point of view that such an almost sure asymptotic result may be considered a strong law of large numbers (SLLN). The subject of the present extended abstract is to study the fluctuations in such SLLN, sometimes called a central limit analogue. We study these fluctuations as processes in the metric space $(D[0, 1], d_{\text{SK}})$ of càdlàg functions endowed with the Skorokhod metric; see Section 3 for the definitions and Billingsley [2] for background on weak convergence of probability measures on metric spaces in general and on $(D[0, 1], d_{\text{SK}})$ in particular. Note that, by definition, $(S_{\alpha, n})_{\alpha \in [0,1]}$ and $(S_\alpha)_{\alpha \in [0,1]}$ have càdlàg paths almost surely. As the normalized process of fluctuations we define

$$G_n := (G_{\alpha, n})_{\alpha \in [0,1]} := \left(\frac{S_{\alpha, n} - n S_\alpha}{\sqrt{n}} \right)_{\alpha \in [0,1]}. \tag{5}$$

Then we have the following result:

▶ **Theorem 1.** *Let $S_{\alpha,n}$ be the number of key comparisons required by* QuickVal$((U_1,\ldots,U_n),\alpha)$ *and* $(S_\alpha)_{\alpha\in[0,1]}$ *as in (4). Then for the fluctuation process* G_n *defined in (5) we have*

$$G_n \xrightarrow{\;d\;} G_\infty \;\; in \; (D[0,1], d_{\mathrm{SK}}) \qquad (n\to\infty),$$

where G_∞ is a mixture of centered Gaussian processes with random covariance function given by

$$\Sigma_{\infty,\alpha,\beta} := \sum_{k=0}^{J}\sum_{j=0}^{\infty} I_{\alpha,j\vee k} + \mathbf{1}_{\{\alpha\neq\beta\}}(J+1)\sum_{j=J+1}^{\infty}(I_{\beta,j}) - S_\alpha S_\beta, \quad \alpha,\beta\in[0,1], \tag{6}$$

where $J = J(\alpha,\beta) := \max\{k\in\mathbb{N}_0 \mid \tau_{\alpha,k} = \tau_{\beta,k}\} \in \mathbb{N}_0 \cup \{\infty\}$.

▶ Remark 2. Note that a strength of a functional limit theorem such as Theorem 1 is its versatility implied by the (continuous) mapping theorem [2, Theorem 2.7]: For any metric space (M,ϱ) and any continuous function $h : D[0,1] \to M$ we obtain the convergence $h(G_n) \to h(G_\infty)$ in distribution. This even holds for discontinuous (measurable) functions h if the set D_h of discontinuities of h satisfies $\mathbb{P}(G_\infty \in D_h) = 0$. Examples include the maximum (or minimum) of the process, i.e., we have for the worst case fluctuation $\max_\alpha G_{\alpha,n} \to \max_\alpha G_{\alpha,\infty}$ in distribution. Also projections to one (or multiple) points, i.e., $G_{\alpha,n} \to G_{\alpha,\infty}$ for $\alpha \in [0,1]$, cf. also Lemma 13. Furthermore, for a random index V with arbitrary probability distribution on $[0,1]$ we obtain $G_{V,n} \to G_{V,\infty}$ in distribution.

▶ Remark 3. In his PhD thesis, Matterer [15, Theorem 6.4] showed the convergence of the one-dimensional marginals for the functional limit law in Theorem 1.

▶ Remark 4. An alternative representation of the random covariance function in (6) is as follows: With an independent random variable V uniformly distributed over $[0,1]$, we have

$$\Sigma_{\infty,\alpha,\beta} = \mathrm{Cov}\big(J(V,\alpha), J(V,\beta)\,\big|\,\mathcal{F}_\infty\big), \tag{7}$$

with the σ-algebra

$$\mathcal{F}_\infty := \sigma\big\{I_\phi \mid \phi \in \{0,1\}^*\big\}. \tag{8}$$

▶ Remark 5. A related functional limit law for the complexity of Radix Selection, an algorithm to select ranks based on the bit expansions of the data, with a limiting Gaussian process with a covariance function related to (7) can be found in [13, Theorem 1.2]. See [18, Theorem 1.1] for another related functional limit law.

The analysis of QuickVal is usually considered an intermediate step to analyze the original FIND algorithm. Grübel and Rösler [9] already pointed out that a version of FIND such as stated above with $C_n^*(k)$ denoting the number of key comparisons for finding rank k within $(U_1,\ldots,U_n)$ does not lead to convergence within $(D[0,1], d_{\mathrm{SK}})$ after the normalization $\alpha \mapsto \frac{1}{n}C_n^*(\lfloor\alpha n\rfloor + 1)$, where here and below the convention $C_n^*(n+1) := C_n^*(n)$ is used. To overcome this problem they propose a version that does not stop in case a pivot turns out to be the rank to be selected by including the pivot in the list $S_<$ and proceeding until a list of size 1 is generated. Moreover, their pivots are chosen uniformly at random. The number of key comparisons $C_n'(k)$ for Grübel and Rösler's FIND-version has the property that

$$\left(\frac{1}{n}C_n'\big(\lfloor\alpha n\rfloor+1\big)\right)_{\alpha\in[0,1]} \xrightarrow{\;d\;} (S_\alpha)_{\alpha\in[0,1]}, \quad \text{in } (D[0,1], d_{\mathrm{SK}}), \tag{9}$$

see [9, Theorem 4]. Without using random pivots we may also obtain right-continuous limits by just recursively calling FIND($S_{\geq}, 0$) in case the pivot turns out to be the rank sought. We denote the number of key comparisons for this version by $C_n(k)$, which is close to Grübel and Rösler's FIND-version and also satisfies (9).

The convergence in (9) could only be stated weakly (not almost surely) since Grübel und Rösler's FIND-version due to randomization within the algorithm does not have a natural embedding on a probability space. Note that the formulation of the QuickVal complexity does have such an embedding which, e.g., makes the almost sure convergence in (4) possible. However, it is easy to see that we have the distributional equality

$$\left(C_n\left(|\{U_i \leq \alpha : 1 \leq i \leq n\}|\right)\right)_{\alpha \in [0,1]} \overset{d}{=} (S_{\alpha,n})_{\alpha \in [0,1]}. \tag{10}$$

This allows to naturally couple the complexities on one probability space, which we call its *natural coupling*. See [7, page 807] for a related discussion of natural couplings.

To transfer Theorem 1 to FIND we need to align jumps to come up with a suitable fluctuation process. The conventions $C_n(0) := C_n(1)$ and $C_n(n+1) := C_n(n)$ are used.

▶ **Corollary 6.** *Let $C_n(k)$ be the number of key comparisons required to select rank $1 \leq k \leq n$ within a set of n data by* FIND *with the natural coupling (10). Let $\Lambda_n : [0,1] \to [0,1]$, $n \in \mathbb{N}$, be any (random) monotone increasing bijective function such that $\Lambda_n(\frac{k}{n+1})$ is equal to the element of rank k within $\{U_1, \ldots, U_n\}$. Then we have*

$$\left(\frac{C_n(\lfloor t(n+1) \rfloor) - n S_{\Lambda_n(t)}}{\sqrt{n}}\right)_{t \in [0,1]} \overset{d}{\longrightarrow} G_\infty \quad in \ (D[0,1], d_{\mathrm{SK}}),$$

where G_∞ is the process defined in Theorem 1.

The extended abstract is organized as follows: In Section 2 we introduce a novel perturbation argument which is the basis of our analysis. Section 3 contains a criterion for weak convergence of probability measures on $(D[0,1], d_{\mathrm{SK}})$, which is applied in Section 4 to proof Theorem 1 and Corollary 6. In Section 5 further functional fluctuation results are stated for the number of swaps (key exchanges) required by QuickVal (depending on the specific algorithm used to partition S into the sublists $S_<$ and $S_{\geq}$) as well as functional fluctuation results for cost measures which are based on key comparisons, where the cost of a comparison may depend on the values of the keys.

For proofs omitted in the present extended abstract see the full paper version in preparation, which will be available as a future version of [12].

2   Perturbation of the data

QuickVal splits an interval $[L_\phi, R_\phi)$ by the first value falling into $[L_\phi, R_\phi)$ denoted by U_{τ_ϕ}. Obviously, this implies dependencies between the data U_i and the lengths I_ϕ of the intervals $[L_\phi, R_\phi)$. In the present section we construct a perturbed sequence $(\widetilde{U}_i)_{i \in \mathbb{N}}$ to the data $(U_i)_{i \in \mathbb{N}}$ such that we gain independence of $(\widetilde{U}_i)_{i \in \mathbb{N}}$ from the σ-algebra $\mathcal{F}_\infty$ generated by the interval lengths defined in (8). In particular, we aim that conditional on $\mathcal{F}_\infty$ the number of data $(\widetilde{U}_1, \ldots, \widetilde{U}_n)$ falling into an interval $[L_\phi, R_\phi)$ is binomial $B(n, I_\phi)$ distributed, see Lemma 8 below.

Every value U_i, $i \in \mathbb{N}$, falls successively into subintervals generated by QuickVal until becoming a pivot element. These subintervals correspond to the path between the root of the corresponding binary search tree and the node where U_i is inserted. Let $\phi_i \in \{0,1\}^*$ denote the node where U_i is inserted. Hence, we have $\tau_{\phi_i} = i$ and $U_i = L_{\phi_i} + I_{\phi_i 0}$.

Let $(V_i)_{i \in \mathbb{N}}$ be a sequence of i.i.d. unif$[0,1]$ random variables being independent of $(U_i)_{i \in \mathbb{N}}$. We define

$$\widetilde{U}_i := L_{\phi_i} + I_{\phi_i} V_i. \tag{11}$$

▶ **Lemma 7.** *The sequence $(\widetilde{U}_i)_{i \in \mathbb{N}}$ defined in (11) consists of i.i.d.* unif$[0,1]$ *distributed random variables and is independent of $\mathcal{F}_\infty$.*

Proof. It suffices to show that $\widetilde{U}_i$ conditional on $\mathcal{F}_\infty$ and $\widetilde{U}_1, \ldots, \widetilde{U}_{i-1}$ is uniformly distributed on $[0,1]$ for all $i \in \mathbb{N}$. We use infinitesimal notation to denote this claim by

$$\mathbb{P}\Big(\widetilde{U}_i \in \mathrm{d}u \,\Big|\, \mathcal{F}_\infty, \widetilde{U}_1, \ldots, \widetilde{U}_{i-1}\Big) = \mathbf{1}_{[0,1]}(u)\mathrm{d}u, \quad i \in \mathbb{N}.$$

For each $i \in \mathbb{N}$ the random variables $\widetilde{U}_i$ and U_i fall into the same interval $[L_{\phi_i}, R_{\phi_i})$, hence $\phi_1, \ldots, \phi_{i-1}$ are determined by $\widetilde{U}_1, \ldots, \widetilde{U}_{i-1}$. Let us additionally condition on ϕ_i, then, by definition,

$$\mathbb{P}\Big(\widetilde{U}_i \in \mathrm{d}u \,\Big|\, \mathcal{F}_\infty, \widetilde{U}_1, \ldots, \widetilde{U}_{i-1}, \phi_i\Big) = \frac{1}{I_{\phi_i}} \mathbf{1}_{[L_{\phi_i}, R_{\phi_i})}(u)\mathrm{d}u.$$

Note that ϕ_i denotes one of the i external nodes of the binary search tree with internal nodes denoted by $\phi_1, \ldots, \phi_{i-1}$. We denote by Ext_{i-1} the set of the labels of these external nodes. Hence, conditional on $\mathcal{F}_\infty, \phi_1, \ldots, \phi_{i-1}$ the label ϕ_i is chosen from Ext_{i-1} with probability given by the length of the corresponding interval, i.e., $\mathbb{P}(\phi_i = \phi \,|\, \mathcal{F}_\infty, \phi_1, \ldots, \phi_{i-1}) = I_\phi$ for all $\phi \in \mathrm{Ext}_{i-1}$. Thus, by the law of total probability we obtain

$$\mathbb{P}\Big(\widetilde{U}_i \in \mathrm{d}u \,\Big|\, \mathcal{F}_\infty, \widetilde{U}_1, \ldots, \widetilde{U}_{i-1}\Big) = \sum_{\phi \in \mathrm{Ext}_{i-1}} I_\phi \frac{1}{I_\phi} \mathbf{1}_{[L_\phi, R_\phi)}(u)\mathrm{d}u = \mathbf{1}_{[0,1]}(u)\mathrm{d}u.$$

This implies the assertion.   ◀

The $\widetilde{U}_i$ are now coupled with the U_i but independent of the I_ϕ. To compare with the number of key comparisons required by QuickVal$((U_1, \ldots, U_n), \alpha)$ we define

$$\widetilde{S}_{\alpha,k,n} := \sum_{i=1}^n \mathbf{1}_{[L_{\alpha,k}, R_{\alpha,k})}(\widetilde{U}_i).$$

▶ **Lemma 8.** *Conditional on $I_{\alpha,k}$ we have that $\widetilde{S}_{\alpha,k,n}$ has the binomial $B(n, I_{\alpha,k})$ distribution. Moreover, for all $\alpha \in [0,1]$, $n \in \mathbb{N}$ and $0 \le k \le n$ we have*

$$S_{\alpha,k,n} \le (\widetilde{S}_{\alpha,k,n} - 1)^+ \le S_{\alpha,k,n} + k - 1. \tag{12}$$

Proof. The conditional distribution of $\widetilde{S}_{\alpha,k,n}$ follows from Lemma 7. Recall that $S_{\alpha,k,n}$ is defined as $\sum_{i=\tau_{\alpha,k}}^n \mathbf{1}\{L_{\alpha,k-1} \le U_i < R_{\alpha,k-1}\}$. By definition, U_i and $\widetilde{U}_i$ are in the interval $[L_{\phi_i}, R_{\phi_i})$ for all $i \in \mathbb{N}$. If $U_i \in (L_{\alpha,k}, R_{\alpha,k})$, then U_i appears as a pivot after the k-th pivot. Hence, its interval $[L_{\phi_i}, R_{\phi_i})$ and thus also $\widetilde{U}_i$ are contained in $(L_{\alpha,k}, R_{\alpha,k})$. The k-th pivot $U_{\tau_{\alpha,k}}$ itself does not contribute to $S_{\alpha,k,n}$, which implies the left inequality stated in the present lemma.

For the right inequality, assume for some $i \in \mathbb{N}$ that the perturbed value $\widetilde{U}_i$ is in $(L_{\alpha,k}, R_{\alpha,k})$, but U_i is not. Then the corresponding interval (L_{ϕ_i}, R_{ϕ_i}) must contain $(L_{\alpha,k}, R_{\alpha,k})$, thus making U_i a pivot that appears before the k-th pivot. Since there are only k such pivots, the right inequality follows.   ◀

3 On Weak Convergence in $D[0,1]$

The space $D[0,1]$ consists of all functions $f : [0,1] \to \mathbb{R}$ having left limits and being right-continuous, i.e., with

$$\lim_{s \uparrow t} f(s) \text{ exists for all } t \in [0,1) \qquad \text{and} \qquad \lim_{s \downarrow t} f(s) = f(t) \text{ for all } t \in [0,1).$$

These two properties are abbreviated as càdlàg (continue à droite, limites à gauche). Càdlàg functions are continuous almost everywhere, but may have right-continuous jumps. Measuring closeness of functions $f, g \in D[0,1]$ in the Skorokhod metric is more flexible than just considering the supremum norm $\|f - g\|_\infty$: The Skorokhod metric allows aligning jumps before comparing them in the supremum norm by setting

$$d_{\mathrm{SK}}(f,g) := \inf_\lambda \max\{\|f \circ \lambda - g\|_\infty, \|\lambda - \mathrm{id}\|_\infty\}, \tag{13}$$

where the infimum is taken over all increasing bijections $\lambda : [0,1] \to [0,1]$ and id denotes identity.

To prove the convergence in distribution in Theorem 1 within the metric space $(D[0,1], d_{\mathrm{SK}})$, we use the following Proposition 9. It can be proved by classical tools of weak convergence theory based on a study of the modulus of continuity and the Arzelà–Ascoli theorem in form of a general theorem of Billingsley [2, Theorem 13.2].

▶ **Proposition 9.** *Let $X_1, X_2, \ldots$ be a sequence of random variables in $(D[0,1], d_{SK})$. Suppose that for every $K \in \mathbb{N}$, there exist random càdlàg step functions $X_1^K, X_2^K, \ldots$ with all jumps contained in $\{U_\phi \mid \phi \in \{0,1\}^*, |\phi| < K\}$. If*
   (i) *for all $r \in \mathbb{N}$ and $\alpha_1, \ldots, \alpha_r \in [0,1]$, the marginals $\mathcal{L}(X_n(\alpha_1), \ldots, X_n(\alpha_r))$ converge weakly to some distribution $\mu_{\alpha_1, \ldots, \alpha_r}$,*
   (ii) *for all $\varepsilon > 0$,*

$$\lim_{K \to \infty} \limsup_{n \to \infty} \mathbb{P}\Big(\big\|X_n - X_n^K\big\|_\infty > \varepsilon\Big) \to 0, \tag{14}$$

then $(X_n)_{n \in \mathbb{N}}$ converges in distribution to a random variable X on $(D[0,1], d_{SK})$, and for all $r \in \mathbb{N}$ and $\alpha_1, \ldots, \alpha_r \in [0,1]$ we have

$$\mathcal{L}(X(\alpha_1), \ldots, X(\alpha_r)) = \mu_{\alpha_1, \ldots, \alpha_r}. \tag{15}$$

4 Proof of Theorem 1

To split the contributions to the process G_n into costs resulting from above and below a level $K \in \mathbb{N}$ we define

$$G_{\alpha,k,n} := \frac{S_{\alpha,k,n} - nI_{\alpha,k}}{\sqrt{n}} \tag{16}$$

as the normalized fluctuations of the contribution at level k, and set

$$G_{\alpha,n}^{\leq K} := \sum_{k=0}^{K} G_{\alpha,k,n}, \qquad G_n^{\leq K} := \big(G_{\alpha,n}^{\leq K}\big)_{\alpha \in [0,1]}, \qquad G_{\alpha,n}^{>K} := \sum_{k=K+1}^{\infty} G_{\alpha,k,n}. \tag{17}$$

Hence, $G_{\alpha,n} = G_{\alpha,n}^{\leq K} + G_{\alpha,n}^{>K}$. Analogously, for the perturbed values $\widetilde{S}_{k,n}$ we define

$$W_{\alpha,k,n} := \frac{\widetilde{S}_{\alpha,k,n} - nI_{\alpha,k}}{\sqrt{n}}, \qquad W_{\alpha,n}^{\leq K} := \sum_{k=0}^{K} W_{\alpha,k,n} \qquad W_n^{\leq K} := \big(W_{\alpha,n}^{\leq K}\big)_{\alpha \in [0,1]}. \tag{18}$$

▶ **Lemma 10.** *For all $K \in \mathbb{N}$ we have convergence in distribution of $(G_{\overline{n}}^{\leq K})_{n \in \mathbb{N}}$ towards a mixture $G_\infty^{\leq K} = (G_{\alpha,\infty}^{\leq K})_{\alpha \in [0,1]}$ of centered Gaussian processes within $\| \cdot \|_\infty$. Conditional on $\mathcal{F}_\infty$, the limit $G_\infty^{\leq K}$ is a centered Gaussian process with covariance function given, for $\alpha, \beta \in [0,1]$ by*

$$\mathrm{Cov}\big(G_{\alpha,\infty}^{\leq K}, G_{\beta,\infty}^{\leq K} \mid \mathcal{F}_\infty\big) = \sum_{k=0}^{K \wedge J} \sum_{j=0}^{K} I_{\alpha, j \vee k} + \big(1 + (K \wedge J)\big) \sum_{j=J+1}^{K} I_{\beta, j} - S_\alpha^{\leq K} S_\beta^{\leq K}, \quad (19)$$

where $J = J(\alpha, \beta)$ is as in Theorem 1 and $S_\alpha^{\leq K} := \sum_{k=0}^{K} I_{\alpha,k}$. The stated convergence in distribution also holds conditionally in $\mathcal{F}_\infty$, i.e., we have almost surely that $\mathcal{L}(G_{\overline{n}}^{\leq K} \mid \mathcal{F}_\infty)$ converges weakly towards $\mathcal{L}(G_\infty^{\leq K} \mid \mathcal{F}_\infty)$.

Proof. First note that by Lemma 8 we have $\big\|G_{\overline{n}}^{\leq K} - W_{\overline{n}}^{\leq K}\big\|_\infty < K^2/\sqrt{n}$, so it suffices to show the lemma for $W_{\overline{n}}^{\leq K}$. Conditional on $\mathcal{F}_\infty$, the value of $W_{\alpha,n}^{\leq K}$ is given by

$$W_{\alpha,n}^{\leq K} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{k=0}^{K} \mathbf{1}\{L_{\alpha,k} \leq \widetilde{U}_i < R_{\alpha,k}\} - (R_{\alpha,k} - L_{\alpha,k}), \quad (20)$$

thus the 2^k different values of the process $W_{\overline{n}}^{\leq K}$ can be expressed as the sum of n centered, bounded i.i.d. random vectors, scaled by $1/\sqrt{n}$. By the multivariate central limit theorem, these converge towards a multivariate, centered normal variable. As the positions of the jumps, still conditional on $\mathcal{F}_\infty$, are fixed, we have convergence of $W_{\overline{n}}^{\leq K}$ and thus also of $G_{\overline{n}}^{\leq K}$ towards a Gaussian process. Define $X_{\alpha,k} := \mathbf{1}\{L_{\alpha,k-1} \leq \widetilde{U}_1 < R_{\alpha,k-1}\}$. The covariance function then is given by

$$\begin{aligned}
\mathrm{Cov}\big(G_{\alpha,\infty}^{\leq K}, G_{\beta,\infty}^{\leq K} \mid \mathcal{F}_\infty\big) &= \mathrm{Cov}\left(\sum_{k=0}^{K} X_{\alpha,k}, \sum_{j=0}^{K} X_{\beta,j} \,\bigg|\, \mathcal{F}_\infty \right) \\
&= \sum_{k=0}^{K} \sum_{j=0}^{K} \mathbb{E}[X_{\alpha,k} X_{\beta,j} \mid \mathcal{F}_\infty] - S_\alpha^{\leq K} S_\beta^{\leq K} \\
&= \sum_{k=0}^{K \wedge J} \sum_{j=0}^{K} I_{\alpha, k \vee j} + \big(1 + (K \wedge J)\big) \sum_{j=J+1}^{K} I_{\beta,j} - S_\alpha^{\leq K} S_\beta^{\leq K}. \quad (21)
\end{aligned}$$

The assertion follows.   ◀

To see that the covariance functions in (21) converge towards the covariance function of G_∞ stated in Theorem 1 we restate a Lemma of Grübel and Rösler [9, Lemma 1] that the maximal length of the intervals at a level is decreasing geometrically with increasing levels. It is obtained observing that $\mathbb{E}\big[\sum_{|\phi|=k} I_\phi^2\big] = \big(\frac{2}{3}\big)^k$ and states:

▶ **Lemma 11.** *There exists an a.s. finite random variable K_1 such that for all $k \geq K_1$:*

$$\max_{\alpha \in [0,1]} I_{\alpha,k} \leq k \left(\frac{2}{3}\right)^{k/2}. \quad (22)$$

Lemma 11 implies that the covariance functions of $G_\infty^{\leq K}$ from (19) converge a.s. to the covariance function of G_∞ from (6).

For the costs from levels $k > K$ we find:

▶ **Proposition 12.** *For all $\varepsilon, \eta > 0$ there are constants $K, N \in \mathbb{N}$ such that for all $n \geq N$*

$$\mathbb{P}\big(\|G_n^{>K}\|_\infty > \eta\big) < \varepsilon. \quad (23)$$

We postpone the proof of the latter proposition and first use Proposition 12 and Lemma 10 to show convergence of the finite-dimensional distributions, denoted fdd-convergence.

▶ **Lemma 13.** *We have fdd-convergence of G_n towards G_∞.*

Proof. For any K, we can split $G_n = G_n^{\leq K} + G_n^{>K}$. By Lemma 10, we have

$$G_n^{\leq K} \xrightarrow{\text{fdd}} G_\infty^{\leq K} \qquad (n \to \infty).$$

Furthermore, because the covariance functions of the $G_\infty^{\leq K}$ converge a.s., we obtain

$$G_\infty^{\leq K} \xrightarrow{\text{fdd}} G_\infty \qquad (K \to \infty)$$

by Lévy's continuity theorem. Hence, for all $\alpha_1, \ldots, \alpha_\ell \in [0,1]$ and all $t_1, \ldots, t_\ell \in \mathbb{R}$ we find a sequence $(K_n)_{n \in \mathbb{N}}$ in $\mathbb{N}$ such that

$$\mathbb{P}\big(G_{\alpha_1,n}^{\leq K_n} < t_1, \ldots, G_{\alpha_\ell,n}^{\leq K_n} < t_\ell\big) \longrightarrow \mathbb{P}\big(G_{\alpha_1,\infty} < t_1, \ldots, G_{\alpha_\ell,\infty} < t_\ell\big) \quad (n \to \infty).$$

Now, since $\|G_n^{>K_n}\|_\infty \to 0$ in probability by Proposition 12 the claim of Lemma 13 follows by Slutzky's theorem. ◀

To prepare for the proof of Proposition 12, we show that the fluctuations on each level are also at least geometrically decreasing. Recall K_1 from Lemma 11.

▶ **Lemma 14.** *There exists a constant $a > 1$ such that for all $k, n \in \mathbb{N}$*

$$\mathbb{P}\left(\max_{\alpha \in [0,1]} |W_{\alpha,k,n}| > a^{-k}, K_1 \leq k\right) \leq b(k) + c(k,n)$$

with $b(k), c(k,n) \geq 0$ such that

$$\sum_{k=1}^\infty b(k) < \infty \quad and \quad \sum_{k=1}^{\lfloor (9/2)\log n \rfloor} c(k,n) \to 0 \quad (n \to \infty). \tag{24}$$

For the proof of Lemma 14 we require the following Chernoff bound:

▶ **Lemma 15.** *Let S_n be binomial $B(n,p)$ distributed for some $p \in [0,1]$ and $n \in \mathbb{N}$ and let $\mu := \mathbb{E}[S_n]$, $\varepsilon \geq 0$. Then*

$$\mathbb{P}\left(S_n \notin \big((1-\varepsilon)\mu, (1+\varepsilon)\mu\big)\right) \leq 2\exp\left(-\frac{\varepsilon^2 \mu}{2+\varepsilon}\right).$$

Proof. Combine upper and lower bound in McDiarmid [16, Theorem 2.3]. ◀

Proof of Lemma 14. Fix some $\alpha \in [0,1]$. Conditionally on $I_{k,\alpha}$, the costs $\widetilde{S}_{\alpha,k,n}$ are $B(n, I_{k,\alpha})$-distributed by Lemma 7. The Chernoff bound in Lemma 15 implies

$$\mathbb{P}\big(|W_{\alpha,k,n}| > a^{-k} \mid I_{\alpha,k}\big) = \mathbb{P}\left(\big|\widetilde{S}_{\alpha,k,n} - nI_{\alpha,k}\big| > \sqrt{n}a^{-k} \mid I_{\alpha,k}\right)$$

$$\leq 2\exp\left(-\frac{na^{-2k}}{nI_{\alpha,k}(2 + \sqrt{n}a^{-k}/(nI_{\alpha,k}))}\right)$$

$$= 2\exp\left(-\big(2a^{2k}I_{\alpha,k} + a^k/\sqrt{n}\big)^{-1}\right). \tag{25}$$

For the two summands in the exponent in (25) we have the following behavior: Summand $2a^{2k}I_{k,\alpha}$ is falling geometrically with k for sufficiently small $a > 1$. Summand $a^k/\sqrt{n}$ is falling with n, but growing with k. To separate these two contributions, note that $\exp(x^{-1}) \geq \frac{1}{m!}x^{-m}$ and thus $\exp(-x^{-1}) \leq m!x^m$ for all $m \in \mathbb{N}$ and $x \geq 0$. Choosing $m = 7$, we obtain

$$\mathbb{P}\big(|W_{\alpha,k,n}| > a^{-k} \mid I_{\alpha,k}\big) \leq 2 \cdot 7!\big(2a^{2k}I_{\alpha,k} + a^k/\sqrt{n}\big)^7$$
$$\leq 2^{14}7!a^{14k}I_{\alpha,k}^7 + 2^77!a^{7k}n^{-7/2}$$

by convexity of $x \mapsto x^7$. Note that the 2^k intervals at level k have lengths $I_{\alpha,k}$ summing to 1. Hence,

$$\mathbb{P}\Big(\max_{\alpha \in [0,1]}|W_{\alpha,k,n}| > a^{-k} \mid I_{\alpha,k}\Big) \leq 2^{14}7!a^{14k}\max_{\alpha \in [0,1]}I_{\alpha,k}^6 + 2^77!(2a^7)^k n^{-7/2}.$$

When furthermore $K_1 \leq k$, by Lemma 11 the length $I_{\alpha,k}$ is bounded by $k\big(\frac{2}{3}\big)^{k/2}$, hence

$$\mathbb{P}\Big(\max_{\alpha \in [0,1]}|W_{\alpha,k,n}| > a^{-k}, K_1 \leq k\Big) \leq 2^{14}7!ka^{14k}\Big(\frac{2}{3}\Big)^{3k} + 2^77!(2a^7)^k n^{-7/2}.$$

Define the first summand on the right hand side of the latter inequality by $b(k)$ and the second summand by $c(k,n)$. For all $1 < a < (3/2)^{3/14}$ the $b(k)$ form a convergent series. To also show the second statement in (24) note that

$$\sum_{k=1}^{\lfloor (9/2)\log n \rfloor} c(k,n) = O\Big(\big(2a^7\big)^{(9/2)\log n}n^{-7/2}\Big) = O\Big(n^{(9/2)\log 2 + (9/2)\cdot 7\log a - 7/2}\Big).$$

The latter O-term converges to 0 for $(63/2)\log a < 7/2 - (9/2)\log 2 \approx 0.381\ldots$, thus we may choose a as required.   ◀

We are now prepared for the proof of Proposition 12.

Proof of Proposition 12. Let $\varepsilon, \eta > 0$. To K_1 from Lemma 11 and a and $b(k)$ from Lemma 14 we choose K sufficiently large such that

$$\mathbb{P}(K_1 > K) \leq \frac{\varepsilon}{4}, \qquad \sum_{k=K}^{\infty} a^{-k} \leq \frac{\eta}{4} \qquad \text{and} \qquad \sum_{k=K}^{\infty} b(k) \leq \frac{\varepsilon}{4}. \tag{26}$$

Let H_n be the maximum amount of steps needed by QuickVal$((U_1, \ldots, U_n), \alpha)$ for any α. Thus, H_n is also the height of the binary search tree built from $U_1, \ldots, U_n$. Devroye [4] showed that the height has expectation $\mathbb{E}[H_n] = \gamma \log n + o(\log n)$ with $\gamma = 4.311\ldots$ Reed [17] further showed that $\text{Var}(H_n) = O(1)$. Hence, we can choose N sufficiently large such that

$$\mathbb{P}\big(H_n > \lfloor (9/2)\log n \rfloor\big) < \frac{\varepsilon}{4}, \quad \sum_{k=1}^{\lfloor (9/2)\log n \rfloor} c(k,n) \leq \frac{\varepsilon}{4}, \quad \text{and} \quad \frac{((9/2)\log n)^2}{\sqrt{n}} \leq \frac{\eta}{4} \tag{27}$$

for all $n \geq N$. Subsequently we use the decomposition

$$G_{\alpha,n}^{>K} = \sum_{k=K+1}^{\lfloor (9/2)\log n \rfloor} \frac{S_{\alpha,k,n} - nI_{\alpha,k}}{\sqrt{n}} + \sum_{\lfloor (9/2)\log n \rfloor + 1}^{\infty} \frac{S_{\alpha,k,n} - nI_{\alpha,k}}{\sqrt{n}} =: \Gamma_n + G_{\alpha,n}^{> \lfloor (9/2)\log n \rfloor}$$

and consider the event

$$A_n := \{K_1 > K\} \cup \{H_n > \lfloor (9/2)\log n \rfloor\}.$$

We have $\mathbb{P}(A_n) < \varepsilon/2$ for all $n \geq N$. Note that on A_n^c (the complement of A_n) we have $S_{\alpha,k,n} = 0$ for all $k > \lfloor (9/2) \log n \rfloor$ and also the bound on $I_{\alpha,k}$ from Lemma 11 applies, hence

$$\left| G_{\alpha,n}^{> \lfloor (9/2) \log n \rfloor} \right| \leq \sum_{\lfloor (9/2) \log n \rfloor + 1}^{\infty} \sqrt{n} I_{\alpha,k} \leq \sum_{\lfloor (9/2) \log n \rfloor + 1}^{\infty} \sqrt{n} k \left(\frac{2}{3} \right)^{k/2}$$

$$= \mathrm{O}\left(n^{1/2 - (9/4) \log(3/2)} \log n \right) = \mathrm{o}(1)$$

since $(9/4) \log(3/2) = 0.912\ldots$ Hence, we can enlarge N so that on A_n^c we have $\left| G_{\alpha,n}^{> \lfloor (9/2) \log n \rfloor} \right| < \eta/2$ for all $n \geq N$. This implies the bound

$$\mathbb{P}\left(|G_{\alpha,n}^{>K}| > \eta \right) \leq \mathbb{P}(A_n) + \mathbb{P}\left(\left\{ |G_{\alpha,n}^{>K}| > \eta \right\} \cap A_n^c \right)$$

$$\leq \frac{\varepsilon}{2} + \mathbb{P}\left(\left\{ |\Gamma_n| > \frac{\eta}{2} \right\} \cap A_n^c \right) + \mathbb{P}\left(\left\{ |G_{\alpha,n}^{> \lfloor (9/2) \log n \rfloor}| > \frac{\eta}{2} \right\} \cap A_n^c \right). \quad (28)$$

Note that the third summand in (28) is 0. Hence, it remains to bound the second summand in (28). To this end note that

$$|\Gamma_n| \leq \sup_{\alpha \in [0,1]} \sum_{k=K+1}^{\lfloor (9/2) \log n \rfloor} \left(\left| \frac{\widetilde{S}_{\alpha,k,n} - n I_{\alpha,k}}{\sqrt{n}} \right| + \left| \frac{S_{\alpha,k,n} - \widetilde{S}_{\alpha,k,n}}{\sqrt{n}} \right| \right)$$

$$\leq \left(\sum_{k=K+1}^{\lfloor (9/2) \log n \rfloor} \max_{\alpha \in [0,1]} |W_{\alpha,k,n}| \right) + \frac{\lfloor (9/2) \log n \rfloor^2}{\sqrt{n}}, \quad (29)$$

where Lemma 8 is used. The third relation in (27) assures that the second term in (29) is smaller than $\eta/4$. In view of the second relation in (26) and (29), we have

$$\left\{ |\Gamma_n| > \frac{\eta}{2} \right\} \cap A_n^c \subset \bigcup_{k=K}^{\lfloor (9/2) \log n \rfloor} \left\{ \max_{\alpha \in [0,1]} |W_{\alpha,k,n}| > a^{-k}, K_1 \leq k \right\}.$$

Thus, Lemma 14 together with (26) and (27) imply that the second summand in (28) is bounded by $\varepsilon/2$. This implies the assertion. ◀

Proof of Theorem 1. We apply Proposition 9 to G_n and $G_n^{\leq K}$. The first condition, fdd convergence, is Lemma 13, the second condition is Proposition 12. ◀

We now transfer the fluctuation result for QuickVal in Theorem 1 to the original FIND process in Corollary 6.

Proof of Corollary 6. Let $\widetilde{F}_n$ be the inverse of Λ_n in the statement of Corollary 6. By definition of Λ_n, the value of the element $U_{(k)}$ of rank k within $U_1, \ldots, U_n$ is given by $\frac{k}{n+1}$, so

$$\lfloor (n+1) \widetilde{F}_n(\alpha) \rfloor = \left| \{ U_i \leq \alpha \mid 1 \leq i \leq n \} \right| \quad (30)$$

for all $\alpha \in [0,1)$. Thus, $C_n\left(\lfloor (n+1) \widetilde{F}_n(\alpha) \rfloor \right) = S_{\alpha,n}$ a.s. for all $\alpha \in [0,1]$, see (10). For $\alpha = 1$ note that $\widetilde{F}_n(\alpha) = 1$ and $C_n(n+1) = C_n(n)$ by definition. The Skorokhod distance d_{SK} is then bounded by

$$d_{\mathrm{SK}}\left(G_n, \left(\frac{C_n(\lfloor t(n+1) \rfloor) - n S_{\Lambda_n(t)}}{\sqrt{n}} \right)_{t \in [0,1]} \right) = d_{\mathrm{SK}}\left(G_n, \left(\frac{S_{\Lambda_n(t),n} - n S_{\Lambda_n(t)}}{\sqrt{n}} \right)_{t \in [0,1]} \right)$$

$$= d_{\mathrm{SK}}\left(G_n, (G_{\Lambda_n(t),n})_{t \in [0,1]} \right)$$

$$\leq \| \widetilde{F}_n - \mathrm{id} \|_\infty.$$

By (30), $\widetilde{F}_n$ is close to the empirical distribution function and thus converges a.s. uniformly to the identity id by the Glivenko–Cantelli theorem. The statement of Corollary 6 then follows from Slutzky's theorem. ◀

5   Further cost measures

In this section we sketch results analogous to Theorem 1 for other cost measures than the number of key comparisons. We consider the number of swaps required by QuickVal, which however depends on the implementation of the procedure to partition the input $(U_1, \ldots, U_n)$ into the sublists $S_<$ and $S_\geq$. We consider two such procedures, the one originally proposed by Hoare [10] and one that is attributed to Lomuto, see [1, 3, 14]. Our results are stated in Subsection 5.1.

As a further cost measure, we consider the model where the costs to compare two keys may depend on their values, e.g. the number of bit comparisons required to compare them when they are given by their binary expansions. The total cost for all key comparisons required by QuickVal($(U_1, \ldots, U_n), \alpha$) or FIND($(U_1, \ldots, U_n), k$) is no longer determined by the fact that the ranks of $(U_1, \ldots, U_n)$ form an uniformly random permutation. Here, the distribution of the U_i matters. We only consider the uniform distribution as in the previous sections and hope to report on other distributions in the full paper version of this extended abstract. Our results are stated in Subsection 5.2. A probabilistic analysis for the number of bit comparisons of the related Quicksort algorithms was given in [6, 5].

5.1   Number of swaps

Usually, QuickSelect is implemented in-place, meaning that it only requires the memory for the list S of values and a bounded amount of additional memory. This is achieved by swapping values within S so that the elements of $S_<$ and $S_\geq$ are contained in contiguous parts of the list. Such a procedure is called *partition*. There are various procedures of *partition*.

The original procedure by Hoare [10] searches the list S from both ends at once: It repeatedly finds the index $i = \min\{2 \leq i \leq n \mid U_i > U_1\}$ of the leftmost element bigger than the pivot and the index $j := \max\{2 \leq j \leq n \mid U_j < U_1\}$ of the rightmost element smaller than the pivot. If $i < j$, it swaps U_i and U_j. Else the algorithm terminates.

A simpler, but less efficient implementation is the so-called Lomuto partition scheme [1, 3, 14] that only searches from one end of S. It keeps track of the amount i of elements at the start of the list it has already swapped. In every step, it finds the index $j := \max\{2 \leq j \leq n \mid U_j < U_1\}$ of the rightmost element smaller than the pivot. If $i + 1 < j$, it swaps U_{i+1} and U_j and increases i by one. Otherwise the algorithm terminates.

Both partition schemes only compare elements to the pivot, so the model of randomness is preserved within the sublists $S_<$ and $S_\geq$. However, their original order is not preserved, so QuickSelect run on $U_1, \ldots, U_n$ will usually not select the same pivots as QuickSelect on $U_1, \ldots, U_{n+1}$. For convenience, we assume that the pivot to split a sublist S' of S is the element of S' that came first in the original list S. We call this choice of the pivots a *suitable embedding*.

5.1.1 Hoare's partition

For Hoare's partition, via a hypergeometric distribution the expected number of swaps in step k given $\mathcal{F}_\infty$ is approximately $nI_{\alpha,k+1}(I_\alpha - I_{\alpha,k})/I_{\alpha,k}$, which leads to the limit process $L = (L_\alpha)_{\alpha \in [0,1]}$ given by

$$L_\alpha := \sum_{k=0}^\infty \frac{I_{\alpha,k+1}(I_\alpha - I_{\alpha,k})}{I_{\alpha,k}}, \quad \alpha \in [0,1]. \tag{31}$$

It is now possible to study the fluctuations by their contributions from the individual levels and combine them as for the number of key comparisons above. Since we are still in the range of the central limit theorem we again obtain a mixture of centered Gaussian processes. To be explicit, first denote by Z_ϕ the limit of the $G_{\alpha,k,n}$ as $n \to \infty$ where $\phi = \phi(\alpha, k) \in \{0,1\}^k$. Further, denote by $\{Y_\phi \,|\, \phi \in \{0,1\}^*\}$ a set of i.i.d. $\mathcal{N}(0,1)$ random variables being independent of $\{Z_\phi \,|\, \phi \in \{0,1\}^*\}$ and of $\mathcal{F}_\infty$. Then the limiting process $G^{\mathrm{swap}} = (G_\alpha^{\mathrm{swap}})_{\alpha \in [0,1]}$ is given by

$$G_\alpha^{\mathrm{swap}} := \sum_{\phi \in \{\phi(k,\alpha) \,|\, k \in \mathbb{N}_0\}} Y_\phi \frac{I_{\phi 0} I_{\phi 1}}{I_\phi^{3/2}} + Z_{\phi 0} \frac{I_{\phi 1}}{I_\phi} + Z_{\phi 1} \frac{I_{\phi 0}}{I_\phi} - Z_\phi \frac{I_{\phi 0} I_{\phi 1}}{I_\phi^2}, \quad \alpha \in [0,1]. \tag{32}$$

The Y_ϕ represent fluctuations caused by the hypergeometric distribution connected to *partition*, while the terms with Z represent fluctuations around the limit (31). Then we have the following result for key exchanges corresponding to Theorem 1:

▶ **Theorem 16.** *Let $K_{\alpha,n}$ be the number of key exchanges required by* QuickVal$((U_1, \ldots, U_n), \alpha)$ *with Hoare's partition algorithm in a suitable embedding. Then, as $n \to \infty$, we have*

$$\left(\frac{K_{\alpha,n} - nL_\alpha}{\sqrt{n}} \right)_{\alpha \in [0,1]} \xrightarrow{d} G^{\mathrm{swap}} \quad in \ (D[0,1], d_{\mathrm{SK}}).$$

5.1.2 Lomuto's partition

The Lomuto partition is simpler to implement and much easier to analyze. The Lomuto partition swaps every element smaller than the pivot, so the amount of swaps at some path $\phi \in \{0,1\}^*$ is given by $S_{\phi 0} + 1$. With the Z_ϕ introduced in Subsection 5.1.1 we find that $(Z_\phi)_{\phi \in \{0,1\}^*}$ is a mixture of centered Gaussian processes with conditional covariance function given by

$$\mathrm{Cov}(Z_\phi, Z_\psi \,|\, \mathcal{F}_\infty) = I_{\phi \vee \psi} - I_\phi I_\psi, \quad \phi, \psi \in \{0,1\}^*,$$

where $I_{\phi \vee \psi}$ is the length of the interval $[L_\phi, R_\phi) \cap [L_\psi, R_\psi)$, thus $I_{\phi \vee \psi}$ is only nonzero if one of ψ and ϕ is a prefix of the other. Then the limiting process $G^{\mathrm{Lo}} = (G_\alpha^{\mathrm{Lo}})_{\alpha \in [0,1]}$ is given by

$$G_\alpha^{\mathrm{Lo}} = \sum_{k=0}^\infty Z_{\phi(\alpha,k)0}, \quad \alpha \in [0,1].$$

We can directly apply Lemma 14 and our proof for the number of key comparisons can be straightforwardly transferred.

▶ **Theorem 17.** *Let $K_{\alpha,n}^{\mathrm{Lo}}$ be the number of key exchanges required by* QuickVal$((U_1, \ldots, U_n), \alpha)$ *with Lomuto's partition procedure in a suitable embedding. Then, as $n \to \infty$, we have*

$$\left(\frac{K_{\alpha,n}^{\mathrm{Lo}} - n\sum_{k=0}^\infty I_{\phi(\alpha,k)0}}{\sqrt{n}} \right)_{\alpha \in [0,1]} \xrightarrow{d} G^{\mathrm{Lo}} \quad in \ (D[0,1], d_{\mathrm{SK}}).$$

5.2   Number of bit comparisons

We now consider the model where the cost to compare two keys depends on their values. These costs are described by a measurable *cost function* $\beta : [0,1]^2 \to [0,\infty)$, and we require that they have a polynomial tail, that is: There are constants $c, \varepsilon > 0$ such that for all $u \in [0,1], x \in \mathbb{N}$ and for V being unif$[0,1]$ distributed

$$\mathbb{P}(\beta(u, V) \geq x) \leq cx^{-1/\varepsilon}.$$

This condition is called (c, ε)-*tameness*, see Matterer [15], and β is called to be ε-*tame* if it is (c, ε)-tame for some $c > 0$. Note that, e.g., β counting the number of bit comparisons is ε-tame for all $\varepsilon > 0$. The costs of QuickVal$((U_1, \ldots, U_n), \alpha)$ in this model are given by

$$S_{\alpha,n}^{\beta} := \sum_{k=0}^{\infty} \sum_{\tau_{\alpha,k} < i \leq n} \mathbf{1}_{[L_{\alpha,k}, R_{\alpha,k})}(U_i)\beta(U_{\tau_{\alpha,k}}, U_i)$$

and the limit is, with V being unif$[0,1]$ distributed and independent of the $U_1, \ldots, U_n$, given as

$$S_{\alpha,\infty}^{\beta} := \sum_{k=0}^{\infty} \mathbb{E}\big[\mathbf{1}_{[L_{\alpha,k}, R_{\alpha,k})}(V)\beta(U_{\tau_{\alpha,k}}, V) \,\big|\, \mathcal{F}_{\infty}\big].$$

Matterer [15, Theorem 6.4 and Theorem 6.14] shows for $\varepsilon < \frac{1}{2}$ that for fixed $\alpha \in [0,1]$ the resulting residual

$$G_{\alpha,n}^{\beta} := \frac{S_{\alpha,n}^{\beta} - nS_{\alpha,\infty}^{\beta}}{\sqrt{n}}$$

converges to a mixed centered Gaussian random variable $G_{\alpha,\infty}^{\beta}$ in distribution and with all moments. It is possible to combine them to a mixture of centered Gaussian processes

$$G_{\infty}^{\beta} = (G_{\alpha,\infty}^{\beta})_{\alpha \in [0,1]}, \tag{33}$$

defined by the conditional covariance functions given, with $X_{\alpha,k}^{\beta} := \mathbf{1}_{[L_{\alpha,k}, R_{\alpha,k})}(V) \cdot \beta(U_{\tau_{\alpha,k,n}}, V)$, by

$$\mathrm{Cov}\big(G_{\alpha,\infty}^{\beta}, G_{\gamma,\infty}^{\beta} \mid \mathcal{F}_{\infty}\big) = \mathrm{Cov}\left(\sum_{k=0}^{\infty} X_{\alpha,k}^{\beta}, \sum_{k=0}^{\infty} X_{\gamma,k}^{\beta} \mid \mathcal{F}_{\infty}\right), \quad \alpha, \gamma \in [0,1]. \tag{34}$$

The latter expression is well-defined due to the following lemma using ε-tameness:

▶ **Lemma 18.** *Let $[L, R) \subseteq [0,1]$ be an interval of length $I = R - L > 0$ and $u \in [L, R)$. For V being unif$[0,1]$ distributed set $X := \mathbf{1}_{[L,R)}(V)\beta(u, V)$. Then, for every $s \in (0, \varepsilon^{-1})$, uniformly in L, R, we have*

$$\mathbb{E}[X^s] = I \cdot \mathbb{E}[X^s \mid V \in [L, R)] = \mathrm{O}\Big(I^{1-\varepsilon s}\Big).$$

We have the following result corresponding to Theorem 1.

▶ **Theorem 19.** *Let β be an ε-tame cost function with $\varepsilon < \frac{1}{4}$. Then we have*

$$\left(\frac{S_{\alpha,n}^{\beta} - nS_{\alpha,\infty}^{\beta}}{\sqrt{n}}\right)_{\alpha \in [0,1]} \xrightarrow{d} G_{\infty}^{\beta} \quad in \ (D[0,1], d_{\mathrm{SK}}),$$

where G_{∞}^{β} is the mixture of centered Gaussian processes defined in (33).

─── **References** ───

1 Jon Bentley. *Programming pearls.* Addison-Wesley, Reading, Mass. URL: `https://www.pearson.de/programming-pearls-9780134498027`.

2 Patrick Billingsley. *Convergence of probability measures.* Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons Inc., New York, 1999. A Wiley-Interscience Publication. `doi:10.1002/9780470316962`.

3 Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, and Clifford Stein. *Introduction to algorithms.* MIT Press, Cambridge, MA, third edition, 2009.

4 Luc Devroye. A note on the height of binary search trees. *J. ACM*, 33(3):489–498, May 1986. `doi:10.1145/5925.5930`.

5 James Allen Fill. Distributional convergence for the number of symbol comparisons used by QuickSort. *Ann. Appl. Probab.*, 23(3):1129–1147, 2013. `doi:10.1214/12-AAP866`.

6 James Allen Fill and Svante Janson. The number of bit comparisons used by Quicksort: an average-case analysis. *Electron. J. Probab.*, 17:no. 43, 22, 2012. `doi:10.1214/EJP.v17-1812`.

7 James Allen Fill and Jason Matterer. QuickSelect tree process convergence, with an application to distributional convergence for the number of symbol comparisons used by worst-case find. *Combin. Probab. Comput.*, 23(5):805–828, 2014. `doi:10.1017/S0963548314000121`.

8 James Allen Fill and Takehiko Nakama. Distributional convergence for the number of symbol comparisons used by quickselect. *Advances in Applied Probability*, 45(2):425–450, 2013. `doi:10.1239/aap/1370870125`.

9 Rudolf Grübel and Uwe Rösler. Asymptotic distribution theory for Hoare's selection algorithm. *Adv. in Appl. Probab.*, 28(1):252–269, 1996. `doi:10.2307/1427920`.

10 C. A. R. Hoare. Algorithm 63: Partition. *Commun. ACM*, 4(7):321, July 1961. `doi:10.1145/366622.366642`.

11 C. A. R. Hoare. Algorithm 65: Find. *Commun. ACM*, 4(7):321–322, July 1961. `doi:10.1145/366622.366647`.

12 Jasper Ischebeck and Ralph Neininger. On fluctuations of complexity measures for the find algorithm, 2024. `arXiv:2403.07685`.

13 Kevin Leckey, Ralph Neininger, and Henning Sulzbach. Process convergence for the complexity of radix selection on markov sources. *Stochastic Processes and their Applications*, 129(2):507–538, 2019. `doi:10.1016/j.spa.2018.03.009`.

14 Hosam M. Mahmoud. Distributional analysis of swaps in quick select. *Theoretical Computer Science*, 411(16):1763–1769, 2010. `doi:10.1016/j.tcs.2010.01.029`.

15 Jason Matterer. *Quickselect Process and QuickVal Residual Convergence.* PhD thesis, The Johns Hopkins University, Baltimore, Maryland, 2015.

16 Colin McDiarmid. Concentration. In Michel Habib, Colin McDiarmid, Jorge Ramirez-Alfonsin, and Bruce Reed, editors, *Probabilistic Methods for Algorithmic Discrete Mathematics*, pages 195–248. Springer Berlin Heidelberg, Berlin, Heidelberg, 1998. `doi:10.1007/978-3-662-12788-9_6`.

17 Bruce Reed. The height of a random binary search tree. *J. ACM*, 50(3):306–332, May 2003. `doi:10.1145/765568.765571`.

18 Henning Sulzbach, Ralph Neininger, and Michael Drmota. A Gaussian limit process for optimal FIND algorithms. *Electronic Journal of Probability*, 19(none):1–28, 2014. `doi:10.1214/EJP.v19-2933`.

19 Brigitte Vallée, Julien Clément, James Allen Fill, and Philippe Flajolet. The number of symbol comparisons in QuickSort and QuickSelect. In *Automata, languages and programming. Part I*, volume 5555 of *Lecture Notes in Comput. Sci.*, pages 750–763. Springer, Berlin, 2009. `doi:10.1007/978-3-642-02927-1_62`.

A Bijection for the Evolution of B-Trees

Fabian Burghart ✉ 🄸

Department of Mathematics and Computer Science, Eindhoven University of Technology, The Netherlands

Stephan Wagner ✉ 🄸

Institute of Discrete Mathematics, TU Graz, Austria

Department of Mathematics, Uppsala University, Sweden

---- **Abstract** --

A B-tree is a type of search tree where every node (except possibly for the root) contains between m and $2m$ keys for some positive integer m, and all leaves have the same distance to the root. We study sequences of B-trees that can arise from successively inserting keys, and in particular present a bijection between such sequences (which we call histories) and a special type of increasing trees. We describe the set of permutations for the keys that belong to a given history, and also show how to use this bijection to analyse statistics associated with B-trees.

1 Introduction and main results

B-trees, since their inception in [4], have become a popular data structure. Regarding their mathematical analysis, there were some early results by Yao [13] and Odlyzko [11] for the special case of 2-3-trees, but despite Knuth posing a natural open question in [7], progress has been scarce. Perhaps most notable is the approach using Pólya urns as in [1–3,6], which yielded results especially for the fringe analysis of B-trees. In this paper, we propose a novel way of investigating B-trees, by focussing on what we call histories.

1.1 B-trees and their insertion algorithm

By a search tree, we mean a rooted plane tree whose nodes contain *keys*, which we think of as pairwise distinct real numbers, in such a way that (1) the keys are stored in increasing order from left to right (including within a single node), and (2) every non-leaf node containing k keys has exactly $k + 1$ children, where we think of the i-th child as being attached between the $(i-1)$-th and the i-th key of its parent node. For $i = 1$ we interpret this as being attached to the left of the first key, and analogously for $i = k + 1$, the child is attached to the right of the last key in the node. Note that we explicitly allow leaves to contain keys, and will refer to the intervals between consecutive keys in a leaf as *gaps*; thus we do not follow the convention of [10] where the leaves really take the place of our gaps, and therefore cannot contain keys.

Let $m \geq 1$. Following Knuth [10, Section 6.2.4], a *B-tree* of order $2m + 1$ is a search tree satisfying the following properties: Every node contains at least m and at most $2m$ keys, except for the root which contains at least 1 and at most $2m$ keys. Moreover, the tree is balanced in the sense that all leaves have the same distance to the root. We remark that some authors (e.g. [4]) refer to such a tree as a *B*-tree of order m instead.

B-trees can be constructed via the following *insertion algorithm*: Given a *B*-tree and a key that is not already stored in the tree, place the key in the appropriate leaf and the appropriate position within the keys of the leaf. If, after this placement, the leaf still contains at most $2m$ keys, then we are done. Otherwise, we split the node containing $2m + 1$ keys by moving the median key up into the parent node and grouping the lowest m keys and the largest m keys each in their own node. By doing this, it might now happen that the parent node contains $2m + 1$ keys, in which case we again split it into two nodes of m vertices and move the median key (of the parent node) up. This process may propagate all the way along the path from the leaf where we inserted the key to the root vertex, in which case we create a new root vertex above the old root, containing only a single key (the one that was the median among the $2m + 1$ keys of the old root), and split the old root in two. Note that the latter case of splitting the root is the only situation in which the height of the *B*-tree can increase.

For the purpose of this article, we are interested in *B*-trees up to isomorphism of rooted plane trees. Equivalently, we can represent an isomorphism class by replacing all keys by dots, as in Figure 1(left). For brevity's sake, we will henceforth use *B*-tree to mean such an isomorphism class. An alternative way to think about these isomorphism classes is to fix the keys instead, e.g. by saying the keys are the set $\{1, \ldots, n\}$ – the disadvantage of this approach being that inserting another key means having to reassign the values of some of the old keys. Nonetheless, we will make use of both of these representations.

1.2   Main results

Let T_n be a *B*-tree of order $2m + 1$ containing n keys. A *history* of T_n is a finite sequence $(T_1, \ldots, T_n)$ of *B*-trees of order $2m + 1$ such that for all $i = 2, \ldots, n$, the tree T_i is obtained from T_{i-1} through inserting a single key using the insertion algorithm outlined above. In particular, T_i contains i keys. We denote by $\mathscr{H}_m(T_n)$ the set of all histories of T_n, and by $\mathscr{H}_m(n)$ the set of all histories of any *B*-tree of order $2m + 1$ with n keys. In other words, $\mathscr{H}_m(n) = \bigcup_{T_n} \mathscr{H}_m(T_n)$, where the union is taken over all (non-isomorphic) *B*-trees of order $2m + 1$ with n keys.

We can now state our main result:

▶ **Theorem 1.** *Let $n, m \geq 1$. There is a bijection between $\mathscr{H}_m(n)$ and the set of all trees H_n satisfying the following properties:*
   (i) *H_n is a rooted plane tree on n vertices, labelled by $\{1, \ldots, n\}$, such that along each path from the root to a leaf, the labels are increasing.*
   (ii) *The vertices of H_n at heights $2m, 3m + 1, 4m + 2, \ldots$ have up to two children, all other vertices have at most one child.*

We will call trees H_n satisfying properties (i) and (ii) in the theorem $(2m + 1)$-*historic* (or just *historic*, if it is not ambiguous) in the interest of brevity. Given a historic tree H on n vertices, it will be useful throughout to consider all potential positions for attaching a vertex $n + 1$ that lead to another historic tree. We think of these positions as *external vertices*, and call the vertices in H *internal* to tell them apart. We also write $\bar{H}$ to denote H together with the external vertices. Furthermore, we call the internal vertices at height $2m, 3m + 1, 4m + 2, \ldots$ *branchings* (irrespective of how many internal children they have).

▶ **Proposition 2.** *Let H_n be the historic tree corresponding to a history $(T_1, \ldots, T_n)$ of B-trees of order $2m + 1$ under the bijection in Theorem 1. Then, the following holds:*

(i) *For any $n \geq 1$, the number of external vertices of H_n equals the number of leaves of T_n.*

(ii) *For any $n \geq 1$, the number of branchings in H_n equals the number of keys in T_n that are not stored in leaves.*

(iii) *Let $n \geq 2m + 1$. Consider the i-th external vertex v of H_n from the left, and let s be the number of internal vertices in H_n strictly between v and the closest branching above v. Then, the i-th leaf of T_n from the left contains exactly $m + s$ keys.*

We dedicate Section 2 to the proof of Theorem 1 and Proposition 2. That section will also contain the description of the bijection. In Section 3, we exhibit a recursive construction of $\underline{\pi}(H_n)$, the set of all permutations $\pi \in S_n$ that, when used as key sequence for a B-tree, lead to the history described by the historic tree H_n. As part of this description, we obtain the following result:

▶ **Proposition 3.** *Let H_n be a $(2m+1)$-historic tree having $b \geq 1$ branchings. Let $s_1, \ldots, s_{b+1}$ be the number of internal vertices in H_n strictly between the i-th external vertex and its closest branching. Then*

$$|\underline{\pi}(H_n)| = \left(\frac{(2m+1)!}{(m!)^2} \right)^b \cdot \prod_{i=1}^{b+1} (m + s_i)!. \tag{1}$$

This formula is somewhat reminiscent of the classical hook length formula, see e.g. [10, Section 5.1.4, Exercise 20]: the number of increasing labellings of a tree with n vertices is given by

$$n! \prod_v \frac{1}{N_v},$$

where the product is over all vertices and N_v is the number of vertices in the subtree consisting of v and all its descendants.

▶ Remark 4. It is possible to consider B-trees of order $2m$ as well, where a node splits whenever it is assigned $2m$ keys. In that case, the smallest $m - 1$ keys end up in the left node, the m-th key is pushed into the parent node, and the largest m keys end up in the right node. It is still possible to define suitable $(2m)$-historic trees, but the distance between a branching and the next branching below will depend on whether we go to the left or to the right in H_n.

2 The bijection

The purpose of this section is to prove Theorem 1.

We begin by describing the bijection; see Figure 1 for an example. If $n = 1$, there is only one B-tree and only one corresponding H_1. For an arbitrary history $(T_1, \ldots, T_n) \in \mathscr{H}_m(n)$, construct the corresponding H_n as follows: Assume we already constructed H_k corresponding to the history $(T_1, \ldots, T_k) \in \mathscr{H}_m(k)$ for some $1 \leq k < n$. Then T_{k+1} is obtained from T_k by inserting a single new key. If this insertion takes place in the i-th leaf (counted from left to right) of T_k before accounting for possible splits, then we attach the vertex labelled $k + 1$ to H_k at the i-th external vertex of H_k (counted from left to right). This gives H_{k+1}, and inductively, H_n.
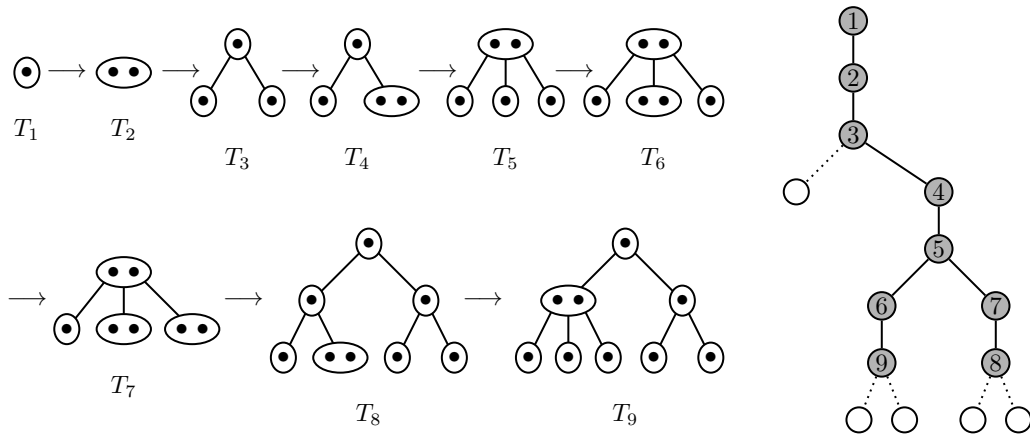
Figure 1 A history of B-trees of order $2m + 1 = 3$ on the left, with the corresponding historic tree H_9 shown on the right. The external vertices of H_9 are shown in white, and are connected by dotted lines. The vertices $3, 5, 8,$ and 9 are the branchings of H_9.

Conversely, given some historic H_n, we can construct trees $H_1, \ldots, H_{n-1}$ such that H_k is the subtree consisting of the vertices with label $\leq k$. Suppose that the vertex $k + 1$ is attached to $H_k \subseteq H_{k+1}$ in the i-th external vertex of H_k, and suppose we have already constructed the history $(T_1, \ldots, T_k)$ corresponding to H_k. Then we can extend this history to the one corresponding to H_{k+1} by inserting a key into the i-th leaf from the left of T_k, and let T_{k+1} be the B-tree obtained by this (possibly after performing the necessary splits).

It is clear from the description that this gives inverse maps between $\mathscr{H}_m(n)$ and $(2m+1)$-historic trees on n vertices, provided the constructions are at all well-defined. This is the case if the number of external vertices on H_k equals the number of leaves of T_k, which is exactly claim (i) in Proposition 2. Thus we proceed by proving Proposition 2, which will imply Theorem 1.

Proof of Proposition 2. We first note that (i) is equivalent to (ii). Indeed, since all non-branchings have outdegree 1 in $\overline{H_n}$, and the branchings have outdegree exactly 2, the number of branchings is one less than the number of external vertices. Similarly, it is a simple consequence of the insertion algorithm for B-trees that every key that gets moved out of a leaf by a split increases the number of leaves by one, so that the number of keys not stored in leaves is one less than the number of leaves.

Next, we observe that (i) holds for $n \leq 2m$. This is the case since any B-tree of order $2m + 1$ for those values of n only has a single node (which is simultaneously root and leaf), and all vertices with these labels in H_n necessarily have outdegree 1. We proceed by induction on n.

For $n = 2m + 1$, we see the first split in T_n, leading to a root node containing a single key, and two leaves containing m keys each. For H_n, we have now reached height $2m$, and thus have two external vertices – these are the children of a branching in $\overline{H_n}$, thus there are no internal vertices between them and the branching. This establishes both (i) and (iii) for $n = 2m + 1$.

Now assume that (i) and (iii) hold for some $n \geq 2m + 1$, and that we obtain T_{n+1} from T_n by adding a key to the i-th leaf, which held $m + s$ keys in T_n. We distinguish two cases for s:

For $0 \leq s \leq m - 1$, we end up with $m + s + 1$ keys in the i-th leaf of T_{n+1}, and the number of leaves does not change. For H_n, we need to append the vertex labelled $n + 1$ in place of the i-th external vertex. Denote by w the closest branching to $n + 1$ (i.e., the most recent

predecessor that is a branching; such a vertex exists since $n \geq 2m + 1$). By assumption, there are exactly s vertices strictly between w and $n + 1$, so $n + 1$ is not a branching, and only has a single external child (and the path from that external vertex to w is one vertex longer). Thus, properties (i) and (iii) hold for $n + 1$.

If, on the other hand, $s = m$, then adding the key splits the i-th leaf; producing two leaves in its stead that each hold m keys. For H_n, we denote again by w the closest branching to the i-th external vertex which becomes the position of the new vertex $n + 1$. Invoking (iii) for H_n shows that there are m vertices between $n + 1$ and w, so $n + 1$ is another branching and therefore has two new external vertices as children, replacing the old one. The closest branching to the new external vertices is now $n + 1$, and there are no internal vertices strictly between them, which again corresponds to the number of keys in the new vertices. This shows that splits in the B-tree correspond to branchings in the historic tree and asserts (i) and (iii) for $n + 1$, finishing the induction argument. ◀

3 The permutations associated with a history

Let T be a B-tree of order $2m + 1$, containing n keys. We denote by $\underline{\pi}(T)$ the set of all permutations $\pi \in S_n$ that, when used as a key sequence for the insertion algorithm, yield the tree T. The aim of this section is to give a recursive description of $\underline{\pi}(T)$ in terms of the "trimmed" tree $T^{(1)}$ that is obtained from T by deleting all leaves. For this purpose, write n_1 for the number of keys stored in $T^{(1)}$. We will rely on the following observation:

Consider a history $(T_1, \ldots, T_n = T)$. Let $i_1 < i_2 < \cdots < i_{n_1}$ be those i where T_i was obtained from T_{i-1} by inserting a key that led to a split (note that this is consistent with the indexing). Then $\left(T_{i_1}^{(1)}, \ldots, T_{i_{n_1}}^{(1)}\right)$ is a valid history of $T^{(1)}$. We remark that this is a purely combinatorial statement: If we instead looked at i.i.d. keys sampled from some continuous probability distribution, then the processes $(T_n)_{n \geq 1}$ and $\left(T_{i_n}^{(1)}\right)_{n \geq 1}$ would be quite different!

As a consequence, suppose we are given a permutation $\pi \in \underline{\pi}(T)$. This π produces a history $(T_1, \ldots, T_n)$. Moreover, keeping track of the actual keys, we obtain a sequence $K_{i_1}, \ldots, K_{i_{n_1}}$ of those keys that ascend above the leaves at times $i_1, \ldots, i_{n_1}$. Forgetting about their actual values and only keeping track of the relative size of the K_{i_j} then produces a new permutation $\pi^{(1)} \in \underline{\pi}(T^{(1)})$, where, moreover, $\pi^{(1)}$ produces the history $\left(T_{i_1}^{(1)}, \ldots, T_{i_{n_1}}^{(1)}\right)$. This defines a map $\Psi_T : \underline{\pi}(T) \to \underline{\pi}(T^{(1)}), \pi \mapsto \pi^{(1)}$, and our goal will be to invert this: Given a $\pi^{(1)}$, we want to find all $\pi \in S_n$ that lead to such $\pi^{(1)}$.

This inversion will come in the form of a 3-step algorithm, described in detail below, in Section 3.2. However, we will give a high-level overview now:

1. In the first step, we start from a given $\pi^{(1)} \in \underline{\pi}(T^{(1)})$, and lift it to a sequence $(K_{i_1}, \ldots, K_{i_{n_1}})$ as above.

2. In the second step, we use $\pi^{(1)}$ and Proposition 2 to construct an acyclic digraph $G = G(T, \pi^{(1)})$. Lemma 7 states that the set of topological labellings of G corresponds bijectively to the set of historic trees of T that produce $\pi^{(1)}$.

3. Therefore, in the third step, we can fix a historic tree H obtained from step 2, and restrict our attention to $\underline{\pi}(H)$. The algorithm given will produce an arbitrary element of $\underline{\pi}(H)$ after making a sequence of choices; different choices will lead to different permutations, and going over all permitted choices produces the entire set, see Lemma 8. In more concrete terms, we start step 3 with an "empty" permutation consisting of n blank symbols, and by recursively comparing it against $K_{i_1}, \ldots, K_{i_{n_1}}$ and H we will replace the blanks by entries from $\{1, \ldots, n\}$.

3.1   Preparatory lemmas

To ensure well-definedness at a later point (Lemma 7), we need the following lemma:

▶ **Lemma 5.** *There is a well-defined map* $\hat{\Psi}_T : \mathscr{H}_m(T) \to \underline{\pi}(T^{(1)})$ *that assigns to a history* $(T_1, \ldots, T_n = T)$ *the* $\pi^{(1)}$ *constructed above, where* $\pi \in \underline{\pi}(T)$ *is any permutation producing the history.*

Proof. We will assume that the keys in T_j are exactly $1, \ldots, j$ (labelled from left to right, since T_j is a search tree) for $1 \leq j \leq n$, and relabel them accordingly whenever we insert a new key. We show inductively that we can (a) determine uniquely which key moved up from the leaves at the times $i_1, i_2, \ldots$ and (b) keep track of how the keys in $T_{j-1}^{(1)}$ change as we go to $T_j^{(1)}$. For $T_1, \ldots, T_{2m}$, there is nothing to show. In T_{2m+1}, we know that the unique key in the root node has label $m + 1$. Suppose we have verified (a) and (b) for some $j \geq 2m + 1$. If $j + 1$ is not one of $i_1, i_2, \ldots, i_{n_1}$ then no splits happen, and comparing T_{j+1} with T_j reveals which leaf grew by one. All keys in $T_j^{(1)}$ to the right of that leaf are increased by 1 for $T_{j+1}^{(1)}$, all other keys in $T_j^{(1)}$ remain the same. If on the other hand $j + 1$ is one of $i_1, i_2, \ldots, i_{n_1}$ then comparing T_{j+1} and T_j reveals which leaf of T_j split. As before, all keys in $T_j^{(1)}$ to the right of that leaf are increased by 1 for $T_{j+1}^{(1)}$, all other keys in $T_j^{(1)}$ remain the same. Moreover, let K be the largest key in $T_j^{(1)}$ to the left of the splitting leaf. Then the new key introduced to $T_{j+1}^{(1)}$ will be $K + m + 1$, and it will be the unique key in T_{j+1} that is placed between the two leaves coming from the split leaf.

Thus, only from the history of T_n we can keep track of which keys were introduced to $T_j^{(1)}$ in which order, which yields $K_{i_1}, \ldots, K_{i_{n_1}}$ after updating all the keys and thus $\pi^{(1)}$. ◀

We also note the following simple fact about the bijection from Theorem 1:

▶ **Lemma 6.** *Let H_n be the historic tree for $(T_1, \ldots, T_n)$. Suppose that vertex i of a historic tree H is a branching, and suppose that the key that is pushed upwards from the splitting leaf at that time is $K_i \in \{1, \ldots, n\}$ in T_n. Let $j \in \{i+1, \ldots, n\}$ be another vertex of H, and let k_j denote the key added at time j in the history. Then, $k_j > K_i$ if and only if j is positioned to the right of i in H (not necessarily as a descendant of i), and $k_j < K_i$ otherwise. Moreover, if $j > i$ is another branching of H, then also $K_j > K_i$ if and only if j is to the right of i, and $K_j < K_i$ otherwise.*

Proof. This follows from the observation that after the split pushes K_i upwards, the leaves of the B-tree can be partitioned into those containing keys $< K_i$, which are therefore to the left, and those containing keys $> K_i$, which are further to the right, as well as from the description of the bijection given in Section 2. ◀

3.2   The algorithm

We now turn our attention to the promised "inverse" of Ψ_T. Denote by h the height of T.

For $h = 0$, the tree T only consists of the root node, and then $\underline{\pi}(T) = S_n$. For $h > 0$, suppose we know $\underline{\pi}(T^{(1)})$.

Step 1. By performing an in-order traversal of the keys in T, we can see which of the numbers $1, \ldots, n$ correspond to the keys in $T^{(1)}$. In other words, in-order traversal gives a monotone injection $\iota : \{1, \ldots, n_1\} \hookrightarrow \{1, \ldots, n\}$, by sending i to the j that is the i-th key from the left among those not in a leaf node of T. This injection in turn allows us to write any $\pi^{(1)} \in \underline{\pi}(T^{(1)})$ as a sequence $\pi_\iota = (K_{i_1}, \ldots, K_{i_{n_1}})$.

Step 2. We construct a rooted digraph $G = G(T, \pi^{(1)})$ in the following way: First, construct a binary search tree from $\pi^{(1)}$. Then, subdivide the edges (and move the root up) in such a fashion that the nodes of the binary search tree become the branchings of a historic tree and append extra vertices to match with the leaves of T, according to Proposition 2(iii). We then direct all edges away from the root, and consider the directed path $\pi^{(1)}(1) \longrightarrow \cdots \longrightarrow \pi^{(1)}(n_1)$. Merge this path into the (mostly empty) historic tree by identifying the vertex $\pi^{(1)}(i)$ in the path with the vertex containing $\pi^{(1)}(i)$ in the tree, for all $i = 1, \ldots, n_1$. For bookkeeping, we colour the edges coming from the path red, and the edges from the tree black. Finally, delete all labels/keys from the resulting digraph G.

▶ **Lemma 7.** *The digraph $G = G(T, \pi^{(1)})$ constructed in this fashion is acyclic. Furthermore, any topological labelling of G (that is, any labelling such that all edges point towards the higher label) induces a historic tree H for T on the black edges. Such H corresponds to those histories of T that are obtained by $\pi \in S_n$ such that $\pi^{(1)}$ is the associated history of $T^{(1)}$. In other words, we have*

$$\{\text{topological labellings of } G(T, \pi^{(1)})\} \xleftrightarrow{1:1} \hat{\Psi}^{-1}(\pi^{(1)}),$$

where the bijection is the one from Theorem 1 after removing the red edges from G.

Step 3. It remains to give the actual description of $\underline{\pi}(T)$. Specifically, writing $\underline{\pi}(H)$ for the set of $\pi \in S_n$ that produce the history encoded by the historic tree H, we can pick an H coming from a topological labelling of $G(T, \pi^{(1)})$ and describe the corresponding $\underline{\pi}(H)$. We are given the key sequence $\pi_\iota = (K_{i_1}, \ldots, K_{i_{n_1}})$ from Step 1, as well as a fixed topological labelling of G, with the induced historic tree H.

In what follows, p will be a sequence of distinct integers which is to be determined, thought of as a map onto some range $\mathcal{R}$. Furthermore $\mathcal{H}$ will be a historic tree on $|\mathcal{R}|$ vertices, and $\mathcal{K}$ is a subsequence of π_ι containing (in the same order) all those K_{i_j} that appear in $\mathcal{R}$. Moreover, we demand that the length of the sequence $\mathcal{K}$ equals the number of branchings in $\mathcal{H}$. Then, the following recursive procedure constructs all desired π:

Step 3.0. Initialize $p = \pi$ as a yet undetermined permutation in S_n, thus $\mathcal{R} = \{1, \ldots, n\}$. Further, set $\mathcal{H} = H$ and $\mathcal{K} = \pi_\iota$.

Step 3.1. If $|\mathcal{R}| \le 2m$, let p be an arbitrary bijection onto $\mathcal{R}$. Otherwise, choose an arbitrary position $1 \le j_1 \le 2m+1$ to place $\mathcal{K}_1$, the first element of $\mathcal{K}$ (i.e., fix $p(j_1) = \mathcal{K}_1$), mark m additional positions among the first $2m+1$ of p as *small*, and the remaining m as *large*. For $j > 2m+1$, mark the j-th entry of the permutation as *small* if the vertex labelled j is positioned to the left of the topmost branching in H, and as *large* otherwise.

Step 3.2. Define new undetermined bijections $p^\pm$, where p^+ contains all the large positions of p, and p^- all the small ones. These bijections will have the ranges $\mathcal{R}^+ := \mathcal{R} \cap \{\mathcal{K}_1 + 1, \ldots, n\}$ and $\mathcal{R}^- := \mathcal{R} \cap \{1, \ldots, \mathcal{K}_1 - 1\}$, respectively. Moreover, let $\mathcal{K}^\pm$ be the subsequences of $\mathcal{K}$ containing, in the same order, the entries strictly larger/smaller than $\mathcal{K}_1$. Also split $\mathcal{H}$ into $\mathcal{H}^\pm$ such that $\mathcal{H}^-$ contains the vertices labelling the small positions in π and such that $\mathcal{H}^-$ below the m-th vertex is equal to the left subtree of $\mathcal{H}$ from Step 3.1. Construct $\mathcal{H}^+$ analogously, then relabel $\mathcal{H}^\pm$ with integers from $1, \ldots, |\mathcal{H}^\pm|$ while maintaining the relative order.

Step 3.3: Repeat steps 3.1–3.3 for both $(p^\pm, \mathcal{R}^\pm, \mathcal{K}^\pm, \mathcal{H}^\pm)$.

▶ **Lemma 8.** *If H comes from a topological ordering of $G(T, \pi^{(1)})$ and π_ι is constructed from T and $\pi^{(1)}$ as in step 1, then the following holds for step 3.2:*

(i) *$\mathcal{K}^\pm$ consists of those entries of π_ι that are contained in $\mathcal{R}^\pm$.*

(ii) *$|\mathcal{H}^\pm| = |\mathcal{R}^\pm| \geq m$, and each of $\mathcal{R}^\pm$ is a set of consecutive integers.*

(iii) *The lengths of $\mathcal{K}^\pm$ are equal to the number of branchings in $\mathcal{H}^\pm$.*

Moreover, the set of permutations constructible with step 3 is $\underline{\pi}(H)$.

▶ **Remark 9.** Since we have $\underline{\pi}(T) = \bigcup_H \underline{\pi}(H)$, where the union is disjoint and to be taken over all histories leading to T, this means we can construct $\underline{\pi}(T)$ out of $\underline{\pi}(T^{(1)})$ by performing steps 1–3 for all $\pi^{(1)}$ in $\underline{\pi}(T^{(1)})$.

3.3   An example

We give an example to illustrate the procedure: Suppose $m = 1$, $n = 9$, and consider the permutation $\pi = (6, 1, 2, 4, 7, 5, 9, 8, 3)$. This permutation produces a B-tree T of the form given in Figure 2a – in fact, this permutation gives the history shown in Figure 1. Thus $T^{(1)}$ contains 4 keys, and $\pi^{(1)} = (1, 3, 4, 2) \in \underline{\pi}(T^{(1)})$.

Step 1.   The in-order traversal of T reveals that the keys in $T^{(1)}$ correspond to the keys $2, 4, 6, 8$ in T. Then the injection on the keys is given by $\{1, 2, 3, 4\} \ni i \mapsto 2i \in \{1, \ldots, 9\}$, and applying this to the entries of π gives $\pi_\iota = (2, 6, 8, 4)$.

Step 2.   Constructing a binary search tree from $\pi^{(1)}$ gives the one shown in Figure 2b which is then turned into the DAG $G = G(T, \pi^{(1)})$ shown in Figure 2c (the remaining labels are there to indicate how it connects to the binary search tree and to $\pi^{(1)}$). This graph has three distinct topological labellings, one of which induces the H depicted in Figure 2d.

Step 3.   We initialize $p = (\_, \_, \_, \_, \_, \_, \_, \_, \_)$, $\mathcal{R} = \{1, \ldots, 9\}$, $\mathcal{K} = (2, 6, 8, 4)$ and $\mathcal{H} = H$. After step 3.1, we have e.g. $p = (\ell, s, 2, \ell, \ell, \ell, \ell, \ell, \ell)$, where we write s for a small position, and ℓ for a large. Here, the assignment of $\ell, s, 2$ to the first 3 positions can be done arbitrarily (but we choose the options that will reconstruct π from above), the remainder is given by comparing it to H: the vertices labelled $4, 5, \ldots$ are all positioned to the right of the top-most branching in H. This leads to $p^- = (\_)$ with $\mathcal{R}^- = \{1\}$, which in the next round of the recursion simply becomes $p = (1)$, and to $p^+ = (\_, \_, \_, \_, \_, \_, \_)$ with $\mathcal{R}^+ = \{3, \ldots, 9\}$, $\mathcal{K}^+ = \{6, 8, 4\}$, and an $\mathcal{H}^+ = H'$ given by Figure 2e.

In the second round of the recursion (using the "+"-branch, as the other one is trivial), we have e.g. $p = (6, s, \ell, s, \ell, \ell, s)$, where the assignment of $\ell, s, 6$ to the first 3 positions can again be done arbitrarily, and the rest is governed by H'. This gives $p^\pm = (\_, \_, \_)$ with $\mathcal{R}^- = \{3, 4, 5\}, \mathcal{K}^- = (4)$ and $\mathcal{R}^+ = \{7, 8, 9\}, \mathcal{K}^+ = (8)$, respectively. Due to the small number of entries, both $\mathcal{H}^\pm$ are given by the unique 3-historic tree on 3 vertices. In the next two rounds of the recursion, the $p^\pm$ will then be filled by arbitrary assignments of the numbers in their range, say $p^+ = (7, 9, 8)$ and $p^- = (4, 5, 3)$.

Finally, we can put everything back together by embedding a pair of $p^\pm$ into the previous p according to the assignment of s and ℓ. Thus, $p^+ = (7, 9, 8)$ and $p^- = (4, 5, 3)$ together yield $(6, 4, 7, 5, 9, 8, 3)$. This in turn was p^+ from the first iteration of step 3, and together with the corresponding $p^- = (1)$, we regain $\pi = (6, 1, 2, 4, 7, 5, 9, 8, 3)$.

(a) The B-tree T.

(b) Binary search tree.

(c) $G(T, \pi^{(1)})$.

(d) H.

(e) H'.
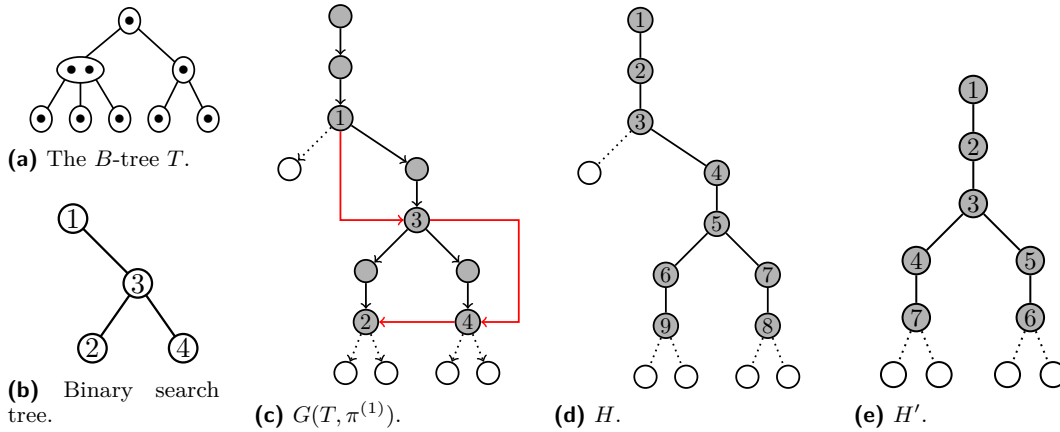
Figure 2 Steps in the algorithm of Section 3, as performed in Subsection 3.3.

3.4 Proofs

Proof of Lemma 7. For G to contain a directed cycle, we need two vertices $v_i = \pi^{(1)}(i)$ and $v_j = \pi^{(1)}(j)$ such that v_i is a descendant of v_j in the tree (i.e., according to the black edges), but v_j is a descendant of v_i according to the red edges. However, the latter only means that $i < j$. Accordingly, $\pi^{(1)}(i)$ was the first to be used for the binary search tree's construction, and hence v_i cannot be below v_j in the tree. Thus G is acyclic.

Trivially, any topological ordering of G yields an increasing labelling for the tree, so the induced H is historic. By construction of H, the final tree in the corresponding history $(T_1, \ldots, T_n)$ will have the same leaves as T (according to Proposition 2(iii)), and thus the same set of keys in $T_n^{(1)}$ as in $T^{(1)}$. Moreover, by Lemma 6, these keys are moved upwards from the leaves in the relative order described by $\pi^{(1)}$, hence $T_n^{(1)} = T^{(1)}$ and H is a history of T that belongs to $\hat{\Psi}^{-1}(\pi^{(1)})$.

Conversely, consider now a history $(T_1, \ldots, T_n) \in \hat{\Psi}^{-1}(\pi^{(1)})$ with the associated historic tree H_n. As in step 1, we obtain from $\pi^{(1)}$ and T_n a sequence $\pi_\iota = (K_{i_1}, \ldots, K_{i_{n_1}})$, where the i_j are precisely the times in the history when a leaf was split. The historic tree H_n therefore must have its branchings labelled by the i_j (and this forces the labelling to be increasing along the red edges in G), and to be consistent with Lemma 6, the left/right-positioning of the branchings has to correspond to the one in a binary search tree obtained from π_ι or equivalently $\pi^{(1)}$. Thus, such an H_n is of the type constructed in Step 2, and its labelling is a topological labelling of G. ◀

Proof of Lemma 8. Claim (i) is evident from the construction. For (ii), the equality $|\mathcal{H}^\pm| = |\mathcal{R}^\pm|$ follows from Lemma 6, we have $|\mathcal{R}^\pm| \geq m$ since $|\mathcal{R}| \geq 2m+1$, and consecutivity is again immediate from the construction. Claim (iii) again follows from Lemma 6. Taken together, these three claims ensure that the recursion in step 3 is well-defined whenever we initialize as in step 3.0.

For the final assertion, we first define $\pi^\pm$ to be permutations obtained from $p^\pm$ in step 3.2 by mapping $\mathcal{R}^\pm$ to $\{1, \ldots, |\mathcal{R}^\pm|\}$ in an order-preserving fashion. We now use strong induction on n, where $T, \pi^{(1)}$, and H are arbitrary but coherent in the sense of Lemma 7. For $n \leq 2m$, there is nothing to show, as step 3.1 gives $\underline{\pi}(H) = S_n$.

For all larger n, observe that if $\pi \in \underline{\pi}(H)$ then K_{i_1} is the median of $\pi(1), \ldots, \pi(2m+1)$ and $\pi^\pm \in \underline{\pi}(\mathcal{H}^\pm)$. The first property is equivalent to K_{i_1} being moved upwards from the leaves at the first branching of H and is ensured by step 3.1. The second property comes from

Lemma 6: The large entries in π are precisely (except for the first m) those corresponding to the right descendants of the first branching in H. This holds by step 3.2, and $\pi^{\pm}$ are constructible by the induction hypothesis since $|\mathcal{R}^{\pm}| < n$. Thus, such π is constructible.

Conversely, suppose π is constructible. Then, by the recursion, $\pi^{\pm}$ are constructible, and thus $\pi^{\pm} \in \underline{\pi}(\mathcal{H}^{\pm})$ by the induction hypothesis. The i-th entry of $\pi^{\pm}$ is simultaneously the i-th large/small entry (according to step 3.1) of π. For $i < m$, this is chosen arbitrarily among all possible configurations for the first $2m + 1$ entries of π. For $i > m$, the entry in π is dictated by H, but is not among the first $2m + 1$. Thus, it corresponds to a descendant of the first branching, and it follows from Lemma 6 that filling the large/small entries with the entries from $\pi^{\pm}$ in order produces a $\pi \in \underline{\pi}(H)$. ◀

Proof of Proposition 3. This lemma follows from an analysis of step 3. Indeed, whenever we are placing a $\mathcal{K}_1$ into the permutation in step 3.1, we have $2m + 1$ choices for the exact position, and then $\binom{2m}{m}$ choices for the location of the small positions within the first $2m + 1$ slots of p. Placing such a $\mathcal{K}_1$ corresponds exactly to the branchings in H, and it is clear that every possible choice will lead to a different π in the end. Whenever we have $|\mathcal{R}| \leq 2m$, we have $|\mathcal{R}|!$ choices. Moreover, invoking Lemma 8, the corresponding $\mathcal{K}$ is the empty sequence, hence the entries of $\mathcal{R}$ are the keys that end up in a joint leaf – say, the i-th leaf – of T. Thus by Proposition 2(iii), $|\mathcal{R}| = m + s_i$, and (1) follows. ◀

4   The number of histories

In this section, we will be interested in the number of possible histories that can arise, and in particular the asymptotic behaviour of this number. We focus on the case $m = 1$. In this case, a historic tree is a binary increasing tree where only vertices at even heights can have two children. Vertex 2 is always the only child of vertex 1, and vertex 3 is always the only child of vertex 2. It will be advantageous later to remove vertex 1 (and decrease all other labels by 1); we call the result a *reduced historic tree*. In such a tree, only vertices at odd heights can have two children.

If we remove the top two vertices (vertices 1 and 2, referred to in the following as the *stem*) from a reduced historic tree with n vertices, then it decomposes into two smaller reduced historic trees, each possibly only consisting of a single external vertex. On the level of generating functions, this translates to a second-order differential equation for the exponential generating function $H(x) = \sum_{n \geq 0} \frac{h_n}{n!} x^n$, where h_n is the number of reduced historic trees with n internal vertices (equivalently, the number of histories of length $n + 1$). We have

$$H''(x) = H(x)^2, \qquad H(0) = H'(0) = 1. \tag{2}$$

One can compare this to the well-known differential equation $T'(x) = T(x)^2$ for the exponential generating function associated with arbitrary binary increasing trees, see for example [8, Lemma 6.4]. We remark here that the tree consisting only of a single external vertex is often not counted, in which case the equation becomes $T'(x) = (1 + T(x))^2$ instead. The sequence h_n (see [12, A007558]) and the associated differential equation (2) were analysed in a different context in [5]: the differential equation has an explicit solution that can be expressed in terms of the Weierstrass elliptic function. It has a dominant singularity at $\rho \approx 2.3758705509$ where

$$H(x) \sim \frac{c}{(1 - x/\rho)^2}$$

for a constant $c = 6\rho^{-2} \approx 1.0629325375$. This leads to the following asymptotic behaviour:

$$\frac{h_n}{n!} \sim cn\rho^{-n} = 6n\rho^{-n-2}.$$

We would now like to generalise the differential equation to arbitrary $m \geq 1$. We remove the first m vertices from a $(2m+1)$-historic tree to obtain a reduced $(2m+1)$-historic tree, which is now a binary increasing tree where only vertices at heights $\equiv -1 \mod m+1$ can have two children. Removing the stem consisting of $m+1$ vertices decomposes such a tree into two smaller trees with the same property (each of them can also be a single external vertex). So in analogy to (2), we obtain a differential equation of order $m+1$, namely

$$H^{(m+1)}(x) = H(x)^2, \qquad H(0) = H'(0) = \cdots = H^{(m)}(0) = 1. \tag{3}$$

This higher-order differential equation can no longer be solved in an explicit fashion, as it was the case for $m = 1$. If we assume that there is a dominant singularity ρ_m where the behaviour of H is of the form $c_m(1 - x/\rho_m)^{-a_m}$, then comparing the two sides of the equation gives us

$$c_m \frac{a_m(a_m + 1) \cdots (a_m + m)}{\rho_m^{m+1}} (1 - x/\rho_m)^{-a_m-m-1} = c_m^2 (1 - x/\rho_m)^{-2a_m},$$

thus $a_m = m + 1$ and $c_m = \frac{(2m+1)!}{m!} \rho_m^{-m-1}$. Applying singularity analysis would then yield

$$[x^n]H(x) \sim \frac{c_m}{m!} n^m \rho_m^{-n} = \frac{(2m+1)!}{(m!)^2} n^m \rho_m^{-n-m-1}.$$

This leads us to the following conjecture:

▶ **Conjecture 10.** *For every $m \geq 1$, the number of reduced $(2m+1)$-historic trees with n vertices (corresponding to histories of length $n+m$) is asymptotically equal to*

$$n! \cdot \frac{(2m+1)!}{(m!)^2} n^m \rho_m^{-n-m-1}$$

for some positive constant ρ_m.

Numerical evidence for small values of m seems to support this conjecture, as the fit of the asymptotic formula with the actual coefficients is excellent. Experimental values of the exponential growth rate ρ_m^{-1} are given in Table 1.

▮ **Table 1** Experimental values of ρ_m^{-1} for $2 \leq m \leq 6$.

m	2	3	4	5	6
ρ_m^{-1}	3.7746	5.1792	6.5857	7.9928	9.3999

5 Statistics of B-trees via historic trees

Let us now study B-trees that are constructed by successive insertion of n random numbers. Equivalently, we can think of them as being constructed from a random permutation of $1, 2, \ldots, n$. In order to apply the connection to historic trees, we need to take the number of permutations associated with a specific history into account.

Again, we focus on the special case $m = 1$. Proposition 3 tells us that the number of permutations corresponding to a specific historic tree T is in this case $6^{b(T)} 2^{i(T)}$, where $b(T)$ is the number of branchings and $i(T)$ the number of internal vertices that lie directly between

a branching and an external vertex. This remains true if we consider reduced historic trees. We associate this number as a weight $w(T) = 6^{\mathsf{b}(T)}2^{\mathsf{i}(T)}$ with every reduced historic tree T and consider the weighted exponential generating function (rather than the unweighted one that was analysed in the previous section). In the recursive decomposition of a reduced historic tree into its stem and two smaller trees T_1 and T_2, we have $\mathsf{b}(T) = \mathsf{b}(T_1) + \mathsf{b}(T_2) + 1$ and $\mathsf{i}(T) = \mathsf{i}(T_1) + \mathsf{i}(T_2)$. This is even true if T_1 or T_2 (or both) only consist of a single external vertex. Thus we obtain

$$w(T) = 6^{\mathsf{b}(T_1)+\mathsf{b}(T_2)+1} \cdot 2^{\mathsf{i}(T_1)+\mathsf{i}(T_2)} = 6w(T_1)w(T_2). \tag{4}$$

On the level of the weighted exponential generating function $W(x)$, (2) becomes

$$W''(x) = 6W(x)^2, \qquad W(0) = 1, W'(0) = 2. \tag{5}$$

Unlike (2), however, there is now a very simple explicit solution, namely $W(x) = \frac{1}{(1-x)^2}$. This is not unexpected, since the total weight of all reduced historic trees with n internal vertices must be equal to the number of permutations of $1, 2, \ldots, n+1$. Thus

$$W(x) = \sum_{n \geq 0} \frac{(n+1)!}{n!} x^n = \frac{1}{(1-x)^2}.$$

Recall that the number of external vertices in n-vertex reduced 3-historic trees is in bijection with the number of leaves in 2-3-trees built from $n+1$ keys. Thus, as a next step, we incorporate the number of external vertices $\mathsf{e}(T)$ as an additional statistic in our generating function in order to prove the following theorem:

▶ **Theorem 11.** *Let L_n be the number of leaves in a 2-3-tree built from n random keys. Then we have $\mathbb{E}(L_n) = \frac{3}{7}(n+1)$ and $\mathbb{V}(L_n) = \frac{12}{637}(n+1)$ for $n > 11$. Moreover, the central limit theorem*

$$\frac{L_n - \mathbb{E}(L_n)}{\sqrt{\mathbb{V}(L_n)}} \xrightarrow{\mathrm{d}} N(0, 1)$$

holds.

Proof. Let us consider the bivariate generating function in which the second variable u marks the number of external vertices $\mathsf{e}(T)$:

$$W(x, u) = \sum_T \frac{1}{|T|!} x^{|T|} u^{\mathsf{e}(T)}.$$

Since $\mathsf{e}(T) = \mathsf{e}(T_1) + \mathsf{e}(T_2)$, the differential equation (5) is actually unaffected by the additional variable; the only change concerns the initial values. We have (where derivatives are taken with respect to x)

$$W''(x, u) = 6W(x, u)^2, \qquad W(0, u) = u, W'(0, u) = 2u, \tag{6}$$

which no longer has an equally simple explicit solution. Using the method described in [5], it can, however, be expressed as the inverse function to

$$X(w, u) = \int_u^w \frac{1}{\sqrt{4t^3 + 4u^2(1-u)}} \, dt.$$

It follows that $W(x, u)$ has a dominant singularity at $\rho(u) = \int_u^\infty \frac{1}{\sqrt{4t^3 + 4u^2(1-u)}} \, dt$: as $w \to \infty$, we have

$$X(w, u) = \rho(u) - \frac{1}{\sqrt{w}} + O(w^{-7/2}),$$

thus

$$W(x, u) \sim \frac{1}{(\rho(u) - x)^2}$$

at the singularity. An application of the quasi-power theorem [9, Theorem IX.8] yields a central limit theorem for the number of external vertices. Moreover, one can obtain explicit expressions for the moments. Differentiating (6) with respect to u and plugging in $u = 1$, we obtain the following differential equation for $W_1(x) = \frac{\partial}{\partial u} W(x, u)\big|_{u=1}$:

$$W_1''(x) = 12W(x, 1)W_1(x) = \frac{12}{(1-x)^2} W_1(x), \qquad W_1(0) = 1, W_1'(0) = 2,$$

since we already know that $W(x, 1) = W(x) = (1-x)^{-2}$. This linear differential equation has the two linearly independent solutions $(1-x)^{-3}$ and $(1-x)^4$, and one obtains

$$W_1(x) = \frac{6}{7(1-x)^3} + \frac{(1-x)^4}{7}.$$

Thus for $n > 4$, we have $[x^n]W_1(x) = \frac{6}{7}[x^n](1-x)^{-3} = \frac{6}{7}\binom{n+2}{2}$. Consequently, the average number of external vertices is

$$\frac{[x^n]W_1(x)}{[x^n]W(x)} = \frac{\frac{6}{7}\binom{n+2}{2}}{n+1} = \frac{3(n+2)}{7}.$$

In the same way, one can treat the second moment: to this end, we consider $W_2(x) = \left(\frac{\partial}{\partial u}\right)^2 W(x, u)\big|_{u=1}$. Differentiating (6) twice with respect to u and plugging in $u = 1$ now gives us

$$W_2''(x) = 12W(x)W_2(x) + 12W_1(x)^2$$

$$= \frac{12}{(1-x)^2} W_2(x) + 12\left(\frac{6}{7(1-x)^3} + \frac{(1-x)^4}{7}\right)^2, \qquad W_2(0) = W_2'(0) = 0.$$

The solution to this differential equation is given by

$$W_2(x) = \frac{54}{49(1-x)^4} - \frac{108}{91(1-x)^3} - \frac{24}{49}(1-x)^3 + \frac{4}{7}(1-x)^4 + \frac{2}{637}(1-x)^{10}.$$

So for $n > 10$, $[x^n]W_2(x) = \frac{54}{49}\binom{n+3}{3} - \frac{108}{91}\binom{n+2}{2} = \frac{9(n+1)(n+2)(13n-3)}{637}$. It follows that the variance of the number of external vertices is

$$\frac{[x^n](W_1(x) + W_2(x))}{[x^n]W(x)} - \left(\frac{[x^n]W_1(x)}{[x^n]W(x)}\right)^2 = \frac{12(n+2)}{637}.$$

This completes the proof. ◀

The approach in our proof provides an alternative to the analysis via Pólya urns, see [1–3, 13] (in particular, the mean was first determined by Yao [13] by explicitly solving a recursion). Here, one can think of the leaves in a B-tree as balls in an urn of different types

depending on the number of keys they hold. Adding a new key then corresponds to picking a ball from the urn and replacing it by a new ball (of different type), or two new balls in the case of a node split.

The same calculations for the moments as in Theorem 11 can also be carried out for higher values of m, though the expressions become more complicated. For general $m \geq 1$, the differential equation becomes

$$W^{(m+1)}(x, u) = \frac{(2m + 1)!}{m!^2} W(x, u)^2,$$

with initial values

$$W^{(i)}(0, u) = (m + i)! u, \qquad i = 0, 1, \ldots, m.$$

In particular, we have $W(x) = W(x, 1) = m!(1 - x)^{-m-1}$, and $W_1(x) = \frac{\partial}{\partial u} W(x, u)\big|_{u=1}$ satisfies the linear differential equation

$$W_1^{(m+1)}(x) = \frac{2(2m + 1)!}{m!}(1 - x)^{-m-1} W_1(x).$$

Up to a trivial change of variables (substituting for $1 - x$), this is a linear differential equation of Cauchy–Euler type that can be solved with standard tools. In fact, setting $e^{-t} = 1 - x$ turns it into a linear differential equation with constant coefficients. Functions of the form $f(x) = (1 - x)^{-b}$ with

$$b^{\overline{m+1}} = b(b + 1) \cdots (b + m) = \frac{2(2m + 1)!}{m!} = \frac{(2m + 2)!}{(m + 1)!} \tag{7}$$

are particular solutions to this differential equation. Note that $b = m + 2$ is always a solution to (7). The general solution can be determined as linear combination of particular solutions, taking the initial values into account. The term of the form $c(1 - x)^{-m-2}$ in $W_1(x)$ dominates asymptotically.

To give one more concrete example, for $m = 2$ we have the differential equation

$$W'''(x, u) = 30W(x, u)^2, \qquad W(0, u) = 2u, W'(0, u) = 6u, W''(0, u) = 24u.$$

Thus $W(x, 1) = 2(1 - x)^{-3}$. The solutions to (7) are now $b = 4$ and $b = \frac{-7 \pm \sqrt{71}i}{2}$. Taking the initial values into account, we obtain

$$W_1(x) = \frac{60}{37(1 - x)^4} + \frac{7\sqrt{71} + 31i}{37\sqrt{71}}(1 - x)^{(7+\sqrt{71}i)/2} + \frac{7\sqrt{71} - 31i}{37\sqrt{71}}(1 - x)^{(7-\sqrt{71}i)/2}.$$

Thus the average number of external vertices in reduced 5-historic trees with n vertices is $\frac{10(n+3)}{37} + O(n^{-13/2})$.

For general m, one finds from the differential equation that the function $\ell(t) = W_1(1 - e^{-t})$ has Laplace transform

$$L(s) = \frac{m!((m + 1)^{\overline{m+1}} - s^{\overline{m+1}})}{(s - m - 1)((m + 2)^{\overline{m+1}} - s^{\overline{m+1}})},$$

where $s^{\overline{h}} = s(s + 1) \cdots (s + h - 1)$ is a rising factorial as in (7). The term $\frac{\kappa_m}{s-m-2}$ in the partial fraction decomposition corresponds to the dominant term

$$W_1(x) \sim \frac{\kappa_m}{(1 - x)^{m+2}}.$$

Here we have, with $H_k = 1 + \frac{1}{2} + \cdots + \frac{1}{k}$ denoting a harmonic number,

$$\kappa_m = \frac{m!}{2(H_{2m+2} - H_{m+1})}.$$

Consequently, the average number of external vertices in reduced $(2m+1)$-historic trees with n vertices is asymptotically equal to $\frac{\kappa_m}{(m+1)!} \cdot n = \frac{1}{2(m+1)(H_{2m+2} - H_{m+1})} \cdot n$. Some explicit values of the constant $\frac{\kappa_m}{(m+1)!}$ are given in Table 2.

■ Table 2 Values of the constant $\frac{\kappa_m}{(m+1)!} = \frac{1}{2(m+1)(H_{2m+2} - H_{m+1})}$ for $1 \leq m \leq 10$.

m	1	2	3	4	5	6	7	8	9	10
$\frac{\kappa_m}{(m+1)!}$	$\frac{3}{7}$	$\frac{10}{37}$	$\frac{105}{533}$	$\frac{252}{1627}$	$\frac{2310}{18107}$	$\frac{25740}{237371}$	$\frac{9009}{95549}$	$\frac{136136}{1632341}$	$\frac{11639628}{155685007}$	$\frac{10581480}{156188887}$

6 Conclusion and perspective

The connection between B-tree histories and historic trees provides us with a novel way to analyse B-trees and their evolution. Possible future directions include studying further statistics of B-trees and historic trees and considering higher values of m. In particular, a proof of Conjecture 10 would be desirable. It might even be interesting, at least from a purely mathematical perspective, to allow m to grow with n.

References

1 D. Aldous, B. Flannery, and J.L. Palacios. Two applications of urn processes. *Probability in the Engineering and Informational Sciences*, 2(3):293–307, 1988.

2 Ricardo A. Baeza-Yates. Fringe analysis revisited. *ACM Comput. Surv.*, 27(1):111–119, 1995. `doi:10.1145/214037.214103`.

3 A. Bagchi and A.K. Pal. Asymptotic normality in the generalized Pólya–Eggenberger urn model, with an application to computer data structures. *SIAM Journal on Algebraic Discrete Methods*, 6(3):394–405, 1985.

4 R. Bayer and E. McCreight. Organization and maintenance of large ordered indexes. *Acta Informatica*, 1:173–189, 1972.

5 Olivier Bodini, Matthieu Dien, Xavier Fontaine, Antoine Genitrini, and Hsien-Kuei Hwang. Increasing diamonds. In *LATIN 2016: theoretical informatics*, volume 9644 of *Lecture Notes in Comput. Sci.*, pages 207–219. Springer, Berlin, 2016. `doi:10.1007/978-3-662-49529-2_16`.

6 B. Chauvin, D. Gardy, N. Pouyanne, and D.-H. Ton-That. B-urns. *ALEA. Latin American Journal of Probability & Mathematical Statistics*, 13:605–634, 2016.

7 V. Chvátal, D.A. Klarner, and D.E. Knuth. Selected combinatorial research problems. *Computer Science Department, Stanford University*, STAN-CS-72-292, 1972.

8 Michael Drmota. *Random trees*. SpringerWienNewYork, Vienna, 2009. `doi:10.1007/978-3-211-75357-6`.

9 Philippe Flajolet and Robert Sedgewick. *Analytic Combinatorics*. Cambridge University Press, 2009.

10 D.E. Knuth. *The Art of Computer Programming: Volume 3. Sorting and Searching*. Addison-Wesley, 2nd edition, 1998.

11 A.M. Odlyzko. Periodic oscillations of coefficients of power series that satisfy functional equations. *Advances in Mathematics*, 44:180–205, 1982.

12 OEIS Foundation Inc. The On-Line Encyclopedia of Integer Sequences, 2024. Published electronically at `https://oeis.org`.

13 A.C.C. Yao. On random 2–3 trees. *Acta informatica*, 9(2):159–170, 1978.

Tree Walks and the Spectrum of Random Graphs

Eva-Maria Hainzl ⓘ
TU Wien, Austria

Élie de Panafieu ⓘ
Nokia Bell Labs, Nozay, France

─── **Abstract** ───

It is a classic result in spectral theory that the limit distribution of the spectral measure of random graphs $G(n, p)$ converges to the semicircle law in case np tends to infinity with n. The spectral measure for random graphs $G(n, c/n)$ however is less understood. In this work, we combine and extend two combinatorial approaches by Bauer and Golinelli (2001) and Enriquez and Menard (2016) and approximate the moments of the spectral measure by counting walks that span trees.

1 Introduction

Random matrix theory studies the spectrum of random matrices and has found many applications, including in physics [22], wireless communication [19] and numerical analysis [7]. A fundamental result of this field is that the limit distribution of the spectral measure of so-called Wigner matrices converges to the semicircle law [20, 21] and it is worth mentioning that a common proof of this theorem by the moment method relies on counting closed walks on trees (*e.g.* [10]). This universal law has been extended to several other classes, such as adjacency matrices of random regular graphs [13, 18] and Erdős-Rényi random graphs $G(n, p)$ when $pn \to \infty$. In particular, Bauer and Golinelli [1] pointed out the importance of the spectral measure of adjacency matrices of random graphs and explained how to compute the moments by counting walks on trees. Zakharevich [24] picked up on the approach and showed further that the spectral distribution of $G(n, c/n)$ converges to a limit distribution μ^c which has infinite support. However, for $p = c/n$, several technical conditions of classic theorems in probability theory are not met such that one could apply standard techniques and despite recent progress [3, 15, 4, 16, 6], μ^c remains an enigma. In [8], Enriquez and Ménard returned to combinatorial methods and computed several terms of the asymptotic expansion, as c tends to infinity, of the moments of the normalized spectral measures

$$\mu_n^c = \frac{1}{n} \sum_{\lambda \in Sp(c^{-1/2} A(G(n, c/n)))} \delta_\lambda$$

35th International Conference on Probabilistic, Combinatorial and Asymptotic Methods for the Analysis of Algorithms (AofA 2024).
Editors: Cécile Mailler and Sebastian Wild; Article No. 11; pp. 11:1–11:15
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

where $A(G(n, c/n))$ is the adjacency matrix of a random graph $G(n, c/n)$. We go along the steps in the computation of moments of this measure for clarity and start with

$$m_\ell(\mu_n^c) = \sum_G \mathbb{P}\left[G\left(n, \frac{c}{n}\right) = G\right] \cdot \frac{1}{n} \sum_{\lambda \in Sp\left(c^{-1/2}A(G)\right)} \lambda^\ell.$$

This formulation reduces to counting closed walks in G, since the sum of the eigenvalues to the power ℓ is just the trace of the matrix to the power ℓ, and a value $\left(A^\ell\right)_{i,i}$ on the diagonal of this matrix is the number of closed walks of length ℓ starting at the vertex i. That is,

$$\sum_{\lambda \in Sp\left(c^{-1/2}A(G)\right)} \lambda^\ell = \left(\frac{1}{c}\right)^{\ell/2} \text{tr}\left(A(G)^\ell\right) = \sum_{\text{closed walk } (v_1, v_2, \ldots v_\ell, v_1) \in G} \left(\frac{1}{c}\right)^{\ell/2}.$$

Thus, the moment equals

$$m_\ell(\mu_n^c) = \frac{1}{n}\frac{1}{c^{\ell/2}} \sum_{(v_1, v_2, \ldots v_\ell) \in [n]^\ell} \mathbb{E}\left[X_{v_1, v_2} X_{v_2, v_3} \cdots X_{v_\ell, v_1}\right],$$

where X_{v_i, v_j} is the random variable taking the value 1 if the edge (v_i, v_j) is in the graph and 0 otherwise. Observe that if a closed walk $(v_1, \ldots, v_\ell, v_1)$ contains e distinct edges, then $\mathbb{E}\left[X_{v_1, v_2} X_{v_2, v_3} \cdots X_{v_\ell, v_1}\right] = (c/n)^e$. The number of closed walks on $[n]$ of length ℓ with m vertices is bounded by $n^m m^\ell$. Since the total number of vertices is bounded by the length, we have $n^m m^\ell \leq n^m \ell^\ell$. The contribution to the moment of all such closed walks containing e distinct edges is bounded by

$$\frac{1}{n}\frac{1}{c^{\ell/2}} n^m \ell^\ell \left(\frac{c}{n}\right)^e = c^{e-\ell/2} \ell^\ell n^{m-e-1}.$$

We are considering a fixed moment ℓ, so this tends to 0 with n whenever $m < e + 1$, that is, whenever the graph (necessarily connected) induced by the closed walk is not a tree. In particular, when ℓ is odd, the induced graph cannot be a tree, so the moment of order ℓ tends to 0.

Let $w_{m, 2\ell}$ denote the number of closed walks of length 2ℓ spanning a tree with m vertices. We now consider the even moment of order 2ℓ and split the sum according to the number m of distinct vertices in the closed walk

$$m_{2\ell}(\mu_n^c) = \frac{1}{n}\frac{1}{c^\ell} \sum_{m=1}^{\ell+1} \binom{n}{m}\left(\frac{c}{n}\right)^{m-1} w_{m, 2\ell}.$$

Let us define the limit distribution $\mu^c = \lim_{n \to +\infty} \mu_n^c$. Then its odd moments are zero and its moment of order 2ℓ is

$$m_{2\ell}(\mu^c) = \lim_{n \to +\infty} \frac{1}{n}\frac{1}{c^\ell} \sum_{m=1}^{\ell+1} \binom{n}{m}\left(\frac{c}{n}\right)^{m-1} w_{m, 2\ell} = \sum_{m=1}^{\ell+1} \frac{1}{c^{\ell-m+1}} \frac{w_{m, 2\ell}}{m!}.$$

By identifying the generating functions of $(w_{m, 2\ell})_{\ell \geq 0}$, for $m = \ell + 1$ and $m = \ell$, as the Stieltjes transform of a specific measure, Enriquez and Ménard were able to derive an approximation of the moments of the limit law and computational experiments showed that even the density of this measure approximated the shape of the histograms of eigenvalues of sampled matrices quite well. An extension of this approximation to the order c^{-2} took considerable effort on several sides, including the combinatorics of closed walks on trees.

The aim of this paper is to provide further insight into what we call *tree walks*, and consequently an efficient way to compute the numbers $w_{m,2\ell}$, for all $2\ell \geq 0$ and $0 \leq m \leq \ell+1$, and their generating functions. But as we delve further into their connection with the spectral measure, we come across surprising and beautiful identities involving the generating function of the Catalan numbers.

Section 2 presents the formal definition of various tree walk families and our main results, which are Theorem 3 and Theorem 4. Theorem 3 expresses the generating function of tree walks as a rational function of the Catalan generating function. Theorem 4 gives several error terms for an asymptotic approximation of μ^c as c tends to infinity. We also presents Conjecture 5, which states that this asymptotic approximation could be extended to an arbitrary order, turning it into a form of asymptotic expansion. This paper contains only the main steps of the proofs, a complete version being available on arxiv [12]. The main steps of the proof of Theorem 3 and Theorem 4 are given respectively in Sections 3 and 4. Numerical experiments are provided in Section 5.

2 Main results

Before we state our main results, let us clarify some definitions.

▶ **Definition 1** (Tree walks). *A tree walk of size m is a walk on the complete labeled graph of size m that visits every node, starts and ends at the same node, and induces a tree. More formally, a tree walk $W = (v_1, v_2, \ldots, v_\ell)$ is a sequence of $v_i \in [m]$ such that*

$$V := \bigcup_{j \in [\ell]} \{v_j\}, \quad E := \bigcup_{j \in [\ell-1]} \{(v_j, v_{j+1})\} \cup \{(v_\ell, v_1)\}$$

define a labelled tree $T(W)$ with vertex set $V = [m]$ and edge set E. Further, we define v_1 to be the root of the induced tree $T(W)$. Thus, we talk freely about the root and leaves of W, when referring to the root and leaves of $T(W)$. We further stick to the convention that if the root has degree 1 it is also a leaf of $T(W)$.
In the following, we study the number $w_{m,2\ell}$ of tree walks of length 2ℓ that span a tree of size m and the generating function

$$W(v,z) = \sum_{\ell,m \geq 0} w_{m,2\ell} \frac{v^m}{m!} z^\ell.$$

Since a walk of length 2ℓ spans a tree with at most $\ell+1$ vertices, we have $w_{m,2\ell} = 0$ if $\ell < m-1$ and for $\ell \geq 1$, we define $w_{0,2\ell} = w_{1,2\ell} = 0$ and $w_{1,0} = 1$, $w_{0,0} = 0$.

The ordinary generating function of the moments of μ^c is therefore given by

$$M_{\mu^c}(z) = \sum_{\ell \geq 0} m_{2\ell}(\mu^c) z^{2\ell} = \sum_{\ell \geq 0} \sum_{m=0}^{\ell+1} w_{m,2\ell} \frac{c^m}{m!} c^{-\ell-1} z^{2\ell} = \frac{1}{c} W\left(c, \frac{z^2}{c}\right), \tag{1}$$

where $m_0(\mu^c) = 1$ as always. However, we could have restructured $M_{\mu^c}(z)$ like Enriquez and Ménard in [8] as well. We just sum over the negative exponent $\xi = \ell + 1 - m$ of c such that

$$M_{\mu^c}(z) = \sum_{\ell \geq 0} m_{2\ell}(\mu^c) z^{2\ell} = \sum_{\ell \geq 0} \sum_{m \geq 0} w_{m,2\ell} \frac{c^m}{m!} c^{-\ell-1} z^{2\ell} = \sum_{\ell \geq 0} \sum_{\xi \geq 0} \frac{w_{\ell-\xi+1,2\ell}}{(\ell-\xi+1)!} \frac{z^{2\ell}}{c^\xi}. \tag{2}$$

This expansion in turn motivates the following definition.

▶ **Definition 2** (Excess of a tree walk). *If an edge is traversed $2k$ times in a tree walk, then the* excess *of the edge e is defined as $\xi(e) = k - 1$. The* excess *of a tree walk W is the sum over the excess of all edges in the induced tree $T(W) = (V, E)$. Hence, it is half its length minus the number of edges of the tree, $\xi(W) = \ell - |E|$. An edge with positive excess is called an* excess edge *and an edge without excess a* simple edge*. We denote the generating function of tree walks with excess ξ by*

$$W_\xi(z) = \sum_{\ell \geq 0} \frac{w_{\ell - \xi + 1, 2\ell}}{(\ell - \xi + 1)!} z^\ell,$$

where $w_{m,2\ell}$ is the number of tree walks of length 2ℓ that span a tree of size m.

Thus, the relation between the generating functions we have defined so far is

$$M_{\mu^c}(z) = \frac{1}{c} W\left(c, \frac{z^2}{c}\right) = \sum_{\xi \geq 0} \frac{1}{c^\xi} W_\xi\left(z^2\right).$$

Bauer and Golinelli [1] introduced in the sequence $w_{m,2\ell}$ an additional parameter d counting the number of times the walk leaves the root. This approach allowed them to compute the values $w_{m,2\ell}$ for 2ℓ and m up to 120 [17], and they conjectured a particular form for $w_{m,2\ell}$ that we prove in the next theorem. When we translate this decomposition in generating functions, an equation for $W(x, v, z)$ is obtained, where x marks the parameter d. Unfortunately, this equation is not particularly amenable to classic analysis with complex analytic methods as it involves a Laplace transform. Our approach on the other hand is reminiscent of the decomposition of graphs with given excess by Wright [23]. Not only do we prove a well founded recursion in z and v, but we provide more insight into the structure of tree walks and their generating function. Most importantly, we compute closed expressions for $w_{m,2\ell}, \ell \geq 0$ for fixed (small) m and prove a conjecture from [1].

▶ **Theorem 3.** *Let $C(z) = \frac{1 - \sqrt{1 - 4z}}{2z}$ denote the generating function of the Catalan numbers, and $W_\xi(z)$ denote the generating function of tree walks of excess ξ from Definition 2. Then $W_0(z) = C(z)$ and for any $\xi \geq 1$, there are polynomials $(K_{\xi,s}(x))_{0 \leq s \leq 2\xi - 2}$ with non-negative coefficients of degree $2\xi + s$ such that*

$$W_\xi(z) = C(z) \sum_{s=0}^{2\xi - 2} \frac{K_{\xi,s}\left(zC(z)^2\right)}{\left(1 - zC(z)^2\right)^{s+1}}.$$

In particular, denoting by $\mathrm{Cat}(n)$ the n-th Catalan number, we have

$$K_{\xi, 2\xi - 2}(x) = \mathrm{Cat}(\xi - 1) x^{4\xi - 2} \qquad and \qquad K_{\xi, 2\xi - 3}(x) = (3\xi - 1) \mathrm{Cat}(\xi - 1) x^{4\xi - 3}.$$

We establish a recursion for the polynomials $K_{\xi,s}(x)$ in Section 3. This enables the successive computation of three quantities. First, the generating function $W_\xi(z)$ for any ξ, then the series $M_{\mu^c}(z)$ up to an arbitrary degree in c, given sufficient computational power, and finally the moments $m_{2\ell}$.

Our next theorem significantly extends Theorem 3 from [8]. It approximates μ^c for large c. There are many notions of convergence for measures. The one we consider here is the convergence of all moments (restricting to the even ones since the odd ones vanish). Further, when looking at the limit of a sequence of random variables, it is common to rescale them by their mean and standard deviation. Here, the rescaling takes the form of a dilation operator Λ_α, for $\alpha > 0$. This operator transforms a measure μ into the measure $\Lambda_\alpha(\mu)$ satisfying for every Borel set A

$$\Lambda_\alpha(\mu)(A) = \mu(A/\alpha).$$

▶ **Theorem 4.** *Let $m_\ell(\mu)$ denote the ℓ-th moment of a measure μ and Λ_α the dilation operator defined above. Then as $c \to \infty$, it holds for all $\ell \geq 0$ that*

$$m_{2\ell}(\mu^c) = m_{2\ell}\left(\Lambda_{f(1/c)}\left(\sigma + \sum_{i=1}^{5} \frac{1}{c^i}\sigma_i\right)\right) + \mathcal{O}\left(\frac{1}{c^6}\right)$$

where $f(1/c) = 1 + \frac{1}{2c} + \frac{3}{8c^2} + \frac{29}{16c^3} + \frac{1987}{128c^4} + \frac{47247}{256c^5}$, σ is the semicircle law and all $\sigma_1, \sigma_2, \ldots \sigma_5$ are signed measures explicitly given in Section 5, with total mass 0.

This approximation entails some curious identities concerning the generating functions $W_\xi(z)$ and prompts us to state the following conjecture which we discuss in more detail in Section 4.

▶ **Conjecture 5.** *Let $M_{\mu^c}(z)$ be the ordinary moment generating function of μ^c as defined in (1). Then there exists a unique power series $P(x)$ with non-negative integer coefficients such that all $V_i(z)$ which are given by*

$$V_i(z) := \left[c^{-i}\right] M_{\mu^c}\left(\sqrt{\frac{z}{P(1/c)}}\right), \quad i \geq 0$$

are the product of $C(z)$ and a polynomial in $zC(z)^2$.

Let us denote by $f_k(x)$ the truncation of order k of $\sqrt{P(x)}$. If the previous conjecture holds, there exist signed measures $\sigma_1, \ldots, \sigma_k$ explicitly computable from $V_1(z), \ldots, V_k(z)$ such that for any ℓ, the moment of order 2ℓ of μ^c is

$$m_{2\ell}(\mu^c) = m_{2\ell}\left(\Lambda_{f_k(1/c)}\left(\sigma + \sum_{i=1}^{k} c^{-i}\sigma_i\right)\right) + \mathcal{O}(c^{-k-1}).$$

Thus, Conjecture 5 provides a form of asymptotic expansion for μ^c as c tends to infinity.

3 Decomposition of tree walks

Our proof of Theorem 3 involves reducing a tree walk with excess ξ by most of its simple edges to its *kernel walk* and subsequently reversing the contraction by blowing it up to an arbitrary tree walk with excess ξ. The following subsection is focused on this decomposition process and the subsequent subsection on the proof of Theorem 3 and a recursion enumerating kernel walks.

3.1 Kernel walks

Recall that an edge of a tree walk W is *simple* if it is traversed exactly twice, and is an *excess edge* otherwise.

▶ **Definition 6** (Kernel walks). *Given a tree walk W, we define the* kernel *of the tree walk or simply the* kernel walk W_K *as the tree walk we obtain by the following procedure.*
1. *Set $W' = W$ and let $T(W')$ be its induced tree.*
2. *While there exists a simple edge e incident to a leaf in $T(W')$ which is not the root, delete both occurrences of e in W'.*
3. *While the root u of the tree is a leaf and incident to a simple edge $\{u, v\}$, delete this edge in W' and choose v as the root of $T(W')$.*

4. *While there exists a vertex v in $T(W')$ that is not the root and only incident to two simple edges $e_i = e_{j+1} = \{u, v\}$ and $e_{i+1} = e_j = \{v, w\}$, replace both consecutive pairs e_i, e_{i+1} and e_j, e_{j+1} with $\{u, w\}$ in W'.*

5. *Set $W_K = W'$.*

Naturally, a tree walk W with kernel $W_K = W$ is itself called a kernel walk. Further, we define $k_{\xi, s, 2\ell}$ to be the number of kernel walks of length 2ℓ with excess ξ, where the induced tree has s simple edges and we define the corresponding generating function

$$K_\xi(u, v, z) = \sum_{s, \ell \geq 0} k_{\xi, s, 2\ell}\, u^s \frac{v^{\ell - \xi + 1}}{(\ell - \xi + 1)!} z^\ell,$$

where u counts the number of simple edges, v the number of vertices in the induced tree and z the half-length of the walk.

This procedure is illustrated below. Note that the variable v in the generating function of kernel walks is superfluous since its exponent is fully determined by the length and the excess of the walk. However, we choose to keep it to explain the factorial in the denominator. If we consider the generating function $K(u, v, z) = \sum_{\xi \geq 0} K_\xi(u, v, z)$, we can reconstruct the individual generating functions by

$$K_\xi(u, v, z) = \left[y^{\xi - 1}\right] K\left(u, \frac{v}{y}, yz\right).$$

Example. Reducing a tree walk W and its induced tree $T(W)$ to its kernel.

(a) Step 1: Set $T(W')$. Excess edges and the root in $T(W')$ are marked red.

(b) Step 2: Identify all leaves which are not incident to an excess edge.

(c) Step 2: Remove blue vertices and update $T(W')$ by relabeling the vertices.

(d) Repeat Step 2.

(e) Step 3: The root is a leaf.

(f) Step 3: Choose new root and relabel vertices.

(g) Step 4: Find adjacent simple edges.

(h) Step 4: Update $T(W')$ by deleting vertex 4 and relabeling the vertices.

Tree walks of a given excess ξ can be arbitrarily large. However, our next result establishes that there are only finitely many kernel walks of excess ξ. This is reminiscent of the result of Wright [23] on the enumeration of connected graphs.

▶ **Lemma 7.** *Let W_K be a kernel walk with excess ξ. Then its induced tree $T(W_K) = (V, E)$ satisfies $|V| \leq 3\xi - 1$ and the number of its simple edges is at most $2\xi - 2$. These bounds are tight. Thus, $K_\xi(u, v, z)$ is a polynomial of degree $2\xi - 2$ in u, $3\xi - 1$ in v and $4\xi - 2$ in z.*

Proof. Consider a kernel walk W_K of excess ξ, with m vertices, ℓ_1 leaves, ℓ_2 vertices of degree 2 that are not the root, and outdegree sequence $(d_1, \ldots, d_m)$. Each leaf is incident to an excess edge, so $\ell_1 \leq \xi$. Each vertex of degree 2 is incident to an excess edge, so $\ell_2 \leq 2\xi - \ell_1$. The sum of the outdegrees is $m - 1$, so

$$m - 1 = \sum_j d_j \geq 2(m - \ell_1 - \ell_2) + \ell_2$$

which implies $m \leq 3\xi - 1$. The number of simple edges is bounded by $m - 1 - \xi \leq 2\xi - 2$. The kernel walk has at most $2\xi - 2$ half-steps along the simple edges, and 2ξ half-steps along the excess edges, so the half-length is bounded by $4\xi - 2$. Any binary tree on $2\xi - 1$ vertices, with additional edges of excess 1 attached to each leaf, reaches those bounds. ◀

Although expressing $K_\xi(u, v, z)$ directly is challenging, some subfamilies of kernel walks have a simple expression. A kernel walk of excess ξ is said to be *optimal* if it contains $2\xi - 2$ simple edges, and *near-optimal* if it contains $2\xi - 3$ simple edges.

▶ **Lemma 8.** *Let $\mathrm{Cat}(n)$ denote the n-th Catalan number. There are $(3\xi - 1)! \, \mathrm{Cat}(\xi - 1)$ optimal kernels of excess ξ for $\xi \geq 1$, and $(3\xi - 1)! \, \mathrm{Cat}(\xi - 1)$ near-optimal kernels of excess ξ for $\xi \geq 2$. Let $K_{\xi,s}(z)$ denote the generating function of kernel walks with excess ξ and s simple edges in the induced tree, where z marks the half-length of the walk. This implies*
(a) $K_{\xi, 2\xi-2}(z) = \mathrm{Cat}(\xi - 1)z^{4\xi-2}$,   *for $\xi \geq 1$*
(b) $K_{\xi, 2\xi-3}(z) = (3\xi - 1)\,\mathrm{Cat}(\xi - 1)z^{4\xi-3}$,   *for $\xi \geq 2$.*

Given a kernel walk W_K with excess ξ, we reconstruct a tree walk W by substituting every simple edge by a sequence of back and forth steps, adding a sequence of steps at the root of $T(W_K)$, moving the root to the leaf of this attached path and adding a tree walk without excess at the beginning and after each step in this extension of W_K.

▶ **Lemma 9.** *Let $W_\xi(z)$ be the generating function of the number of tree walks with excess $\xi \geq 1$ and $K_\xi(u, v, z)$ the generating function of kernel walks with excess ξ and where u marks the number of simple edges, v the number of vertices and z the half-length of the walk. Then*

$$W_\xi(z) = \frac{C(z)}{1 - zC(z)^2} K_\xi\left(\frac{1}{1 - zC(z)^2}, 1, zC(z)^2\right).$$

The proof of Theorem 3 is now straightforward.

3.2 A recursion for the generating function of tree walks of excess ξ

Theorem 3 raises the question of the computation of the generating function $K_{\xi,s}(z)$ of kernel walks of excess ξ with s simple edges, where z marks the half-length. There exists a recurrence, but we prefer to decompose the tree walks further and enumerate simpler objects. This path also alleviates the work of the computer algebra system when computing $K_{\xi,s}(z)$.

▶ **Definition 10.** *A superreduced walk is a tree walk where no edge is simple. Denoting by $s_{m,2\ell}$ the number of such walks of length 2ℓ on m vertices, their generating function is*

$$S(v, z) = \sum_{\substack{\ell \geq 0 \\ m \geq 0}} s_{m,2\ell} \frac{v^m}{m!} z^\ell.$$
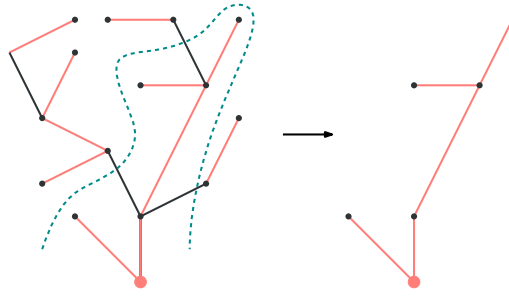
■ **Figure 2** Decomposing kernel walks by isolating the superreduced component including the root (red).

The following lemma reduces the enumeration of kernels to the enumeration of super-reduced walks. The main idea is to consider the induced tree of a kernel walk and isolate the component which contains the root after deleting all simple edges (see Figure 2). The restriction of the kernel walk to this component is a superreduced walk and the restriction of the kernel walk to all of the other components are kernel walks again.

▶ **Lemma 11.** *Let $S(v, z)$ be the generating function of superreduced walks, that is, kernel walks without simple edges, where v counts the number of vertices in the induced tree and z the half-length of the walk. Then for the generating function of kernel walks $K(u, v, z) = \sum_{\xi \geq 0} K_\xi(u, v, z)$ it holds that*

$$K(u, v, z) = \frac{1}{\left(1 - uz(K(u, v, z) - v)\right)} S\left(v, \frac{z}{\left(1 - uz(K(u, v, z) - v)\right)^2}\right) - uvz(K(u, v, z) - v).$$

Once given the generating function $S(v, z)$ of superreduced walks, we compute $K(u, v, z)$ by Lagrange inversion (see *e.g.* [11]). Our next lemma provides an equation characterizing $S(v, z)$. The proof relies on the idea from [1] to mark the number of times the walk leaves the root (see Figure 3). Applying the symbolic method [2, 9] to translate it into generating functions results in a series $S(x, v, z)$ for superreduced walks, where a new auxiliary variable x marks how often the walk leaves the root.

▶ **Lemma 12.** *Let $s_{j,m,2\ell}$ denote the number of superreduced walks on m vertices, length 2ℓ and leaving the root j times. Let*

$$S(x, v, z) = \sum_{j,m,\ell \geq 0} s_{j,m,2\ell} \frac{x^j}{j!} \frac{v^m}{m!} z^\ell$$

denote the generating function of superreduced kernel walks, where z marks the half-length of the walk, v the number of vertices in the induced tree and x how often the walk leaves the root. Then

$$S(x, v, z) = v \exp\left(\mathcal{L}_{t=1}\Big(D\left(t, xz\right) S(t, v, z)\Big)\right)$$

where $D(t, x) = \sum_{k \geq 1} \frac{x^{k+1}}{(k+1)!} \frac{t^k}{k!}$ and $\mathcal{L}_{t=1}(A(t)) = \sum_{k \geq 0} k! [t^k] A(t)$.

By implementing this well founded recursion in v *and* z it is easy to compute $S(v, z)$ up to order $\xi + 1$ in v and 2ξ in z, then we compute $[u^s]K(u, v, z)$ for $s \in [1, \xi]$ by Lagrange
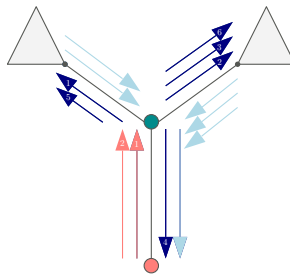
■ Figure 3 Decomposition of a superreduced walk.

inversion from Lemma 11, and finally $W_\xi(z)$. For example for $\xi = 1, 2, 3$, we obtain the generating functions

$$W_1(z) = \frac{z^2 C(z)^5}{1 - zC(z)^2}, \quad W_2(z) = C(z) \frac{z^3 C(z)^6 + 4z^4 C(z)^8 - 6z^5 C(z)^{10} + 2z^6 C(z)^{12}}{\left(1 - zC(z)^2\right)^3}$$

$$W_3(z) = z^4 C(z)^9 \frac{1 + 16zC(z)^2 + 11z^6 C(z)^{12} + 95z^4 C(z)^8 - 54z^5 C(z)^{10} - 62z^3 C(z)^6 - 5z^2 C(z)^4}{\left(1 - zC(z)^2\right)^5},$$

recovering and extending the results of [1] and of [8] (except for $W_2(z)$ where our calculation differs from [8] and agree with [1]).

4 A refined normalisation of the spectral measure and some curious identities

In this section, we return to our initial motivation to describe the moments of the spectral measure μ^c by the identity

$$M_{\mu^c}(z) = \frac{1}{c} W\left(c, \frac{z^2}{c}\right) = \sum_{\xi \geq 0} \frac{1}{c^\xi} W_\xi\left(z^2\right).$$

As Zacharevich [24] pointed out, μ^c is fully determined by its moments and if μ^c were a continuous measure, we could compute its density by the inversion formula of Stieltjes-Perron. This is not the case (μ^c has a dense set of atoms [5, 4]), but nonetheless a better understanding of the Stieltjes transform of μ^c would entail a better understanding of the measure itself.

In combinatorial terms, the Stieltjes transform $S_\mu(z)$ of a measure μ with finite moments is simply the ordinary generating function of moments evaluated at z^{-1} multiplied by z^{-1}. That is,

$$S_\mu(z) = \sum_{\ell \geq 0} m_\ell(\mu) z^{-(\ell+1)}.$$

In turn, under some conditions, the Stieltjes-Perron formula expresses the density ρ of the measure μ by

$$\rho(z) = \lim_{\varepsilon \to 0} -\frac{1}{\pi} \mathrm{Im}\left(S_\mu(z + i\varepsilon)\right). \tag{3}$$

For example, the Stieltjes transform of the limit law μ of the normalized spectral measure of $G(n, p)$ with p constant, and its density, are respectively

$$S_\mu(z) = \frac{1}{z} C\left(\frac{1}{z^2}\right), \qquad \lim_{\varepsilon \to 0} -\frac{1}{\pi} \mathrm{Im}\left(S_\mu(z + i\varepsilon)\right) = \frac{\sqrt{4 - z^2}}{2\pi} \mathbb{1}_{(-2,2)}(z).$$

The distribution given by this density is called after its shape, the *semicircle distribution*. The Stieltjes transform of μ^c equals

$$S_{\mu^c}(z) = \frac{1}{z} M_{\mu^c}\left(\frac{1}{z}\right) = \sum_{\xi \geq 0} \frac{1}{zc^\xi} W_\xi\left(\frac{1}{z^2}\right).$$

Given the structure of $W_\xi(z)$ from Theorem 3, $S_{\mu^c}(z)$ is a sum of rational functions in $S_\mu(z)$

$$S_{\mu^c}(z) = S_\mu(z) + \frac{1}{c} \cdot \frac{S_\mu(z)^5}{1 - S_\mu(z)^2} + \frac{1}{c^2} \cdot \frac{S_\mu(z)^7 + 4S_\mu(z)^9 - 6S_\mu(z)^{11} + 2S_\mu(z)^{13}}{(1 - S_\mu(z)^2)^3} + \ldots$$

Now one could hope that the inversion formula applied to each of the $z^{-1} W_\xi(z^{-1})$ would yield a density of a measure and the density of μ_c would turn out to be a weighted sum of them. This hope is certainly too far fetched, as μ^c has a dense set of atoms. But Enriquez and Ménard [8] found a way to still make use of this expansion by using a dilation operator in their Theorem 3. The main idea is to scale the spectral measure and evaluate $M_{\mu^c}(z)$ at $z/(1 + \frac{1}{2c})$ instead. This scaling entails a perturbation on the level of coefficients of $1/c$. In particular, $\sum_{\ell \geq 0} m_{2\ell}\left(\frac{z}{1 + \frac{1}{2c}}\right)^{2\ell}$ is equal to

$$\sum_{\ell \geq 0}\left(w_{0,2\ell} z^{2\ell} + \frac{1}{c}(w_{1,2\ell} - \ell w_{0,2\ell}) z^{2\ell} + \frac{1}{c^2}\left(w_{2,2\ell} - \ell w_{1,2\ell} + \left(\frac{\ell^2}{2} + \frac{\ell}{4}\right) w_{0,2\ell}\right) z^{2\ell} + \ldots\right).$$

Now the generating functions at c^{-1} and c^{-2} are polynomials in $z^2 C(z^2)^2$ multiplied by $C(z^2)$, and the corresponding densities can be computed with the inversion formula. We expand their calculation to order 5 instead of 2.

Instead of using the dilation operator, we can rescale the adjacency matrix $A(G(n, c/n))$ of $G(n, c/n)$ by $\frac{1}{\sqrt{c\,p(1/c)}}$ instead of $\frac{1}{\sqrt{c}}$. We define

$$\mu_n^p = \frac{1}{n} \sum_{\lambda \in Sp\left((c\,p(1/c))^{-1/2} A(G(n,c/n))\right)} \delta_\lambda,$$

where $p(x)$ is a polynomial in x with constant term 1 which is yet to be determined, and μ^p for the limit as n tends to infinity. This implies

$$\mu_n^p = \Lambda_{p(1/c)^{-1/2}}(\mu_n^c) \quad \text{and} \quad M_{\mu^p}(z) = M_{\mu^c}\left(\frac{z}{\sqrt{p(1/c)}}\right).$$

The original scaling factor $1/\sqrt{c}$ derives from the classical scaling of Wigner matrices, where one scales the matrix by $1/\sqrt{n\mathbb{V}(X)}$, where X is distributed as the individual matrix entries. In the case of adjacency matrices of $G(n, c/n)$ the variance of Bernoulli variables determining the entries of the matrix is of course $c/n(1 - c/n)$ such that we obtain the scaling factor $1/\sqrt{c}$ in the limit. We do not have a similar interpretation for our proposed alternative scaling.

▶ **Proposition 13.** *Let $p_5(x) = 1 + x + x^2 + 4x^3 + 33x^4 + 386x^5$ and $M_{\mu^c}(z)$ be the ordinary moment generating function of μ^c as defined in (1). Then for $V_i(z) = \left[c^{-i}\right] M_{\mu^p}\left(\sqrt{z}\right)$ we have*

$$V_i(z) = C(z)Q_i\left(zC(z)^2\right), \quad i = 0, 1, 2, \ldots, 5,$$

where

$$Q_0(x) = 1, \qquad Q_1(x) = -x, \qquad Q_2(x) = -2x^3,$$
$$Q_3(x) = -\left(11x^5 + x^4 - 2x^3 + 2x^2 + 3x\right),$$
$$Q_4(x) = -\left(90x^7 + 27x^6 - 19x^5 + 17x^4 + 23x^3 + 20x^2 + 26x\right),$$
$$Q_5(x) = -\left(931x^9 + 529x^8 - 163x^7 + 166x^6 + 301x^5 + 239x^4 + 249x^3 + 266x^2 + 324x\right).$$

The proof consists in computing the coefficients of $p_5(x)$ one by one, starting with $[x^0]p(x) = 1$. Then, for any k, let us assume the first $k-1$ coefficients have been computed and set $[x^k]p(x)$ as a variable. We observe in our computations that for the first few values of k,

$$[c^{-k}]M_{\mu^c}\left(\sqrt{\frac{z}{\sum_{j=0}^{k}[x^j]p(x)c^{-j}}}\right)$$

is a fraction with denominator a power of $1 - zC(z)^2$, and the coefficient $[x^k]p(x)$ can be chosen so that this fraction reduces to a polynomial.

Proof of Theorem 4. The generating function of the moments of $\Lambda_{f(1/c)}(\mu^c)$ is given by $M_{\mu^c}\left(\frac{z}{f(1/c)}\right)$. Note that

$$f(x)^2 = 1 + x + x^2 + 4x^3 + 33x^4 + 386x^5 + O(x^6)$$

such that we can expand

$$M_{\mu^c}\left(\frac{z}{f(1/c)}\right) = M_{\mu^{f^2}}(z) = \sum_{i=0}^{5}\frac{1}{c^i}V_i\left(z^2\right) + \sum_{i\geq 6}\frac{1}{c^i}[c^{-i}]M_{\mu^{f^2}}(z)$$

where the $V_i(z)$ are given by Proposition 13. Applying the inversion formula to these functions yield densities of signed measures with null mass. ◄

To illustrate why the existence of $p_5(x)$ is surprising, we observe that if

$$\tilde{W}_2(z) := W_2(z) + \frac{z^3C(z)^7}{(1 - zC(z)^2)^3}$$

is given instead of $W_2(z)$, then there is no choice for $[x^2]p(x)$ allowing this magical simplification between numerator and denominator and the reduction to a polynomial.

This example highlights the difficulty of proving the existence of $P(x)$ in Conjecture 5. A combinatorial approach seems reasonable, but we are not aware of any combinatorial meaning of the generating functions $V_i(z)$, nor do we have a combinatorial interpretation of the differential equations which are satisfied by the generating functions $W_\xi(z)$, except for the equation of $V_1(z)$. Nevertheless the next theorem sheds partial light on *why* the scaling by $p_5(x)$ results in Proposition 13. It shows that keeping the same first two coefficients as in $p_5(x)$ but changing the others gives fractions, in the expansion in c^{-1}, with denominators that are powers of $1 - zC(z)^2$ that are two less than expected.

▶ **Theorem 14.** *Let $p(x) = \sum_{i\geq 0} x^i$. Then $\hat{V}_i(z) := [c^{-i}]M_{\mu^p}(\sqrt{z})$ is a polynomial in $zC(z)^2$ multiplied by $C(z)$ for $i = 0, 1, 2$ and for $i \geq 3$ there exist polynomials $\hat{Q}_i(x)$ such that*

$$\hat{V}_i(z) = C(z)\frac{\hat{Q}_i(zC(z)^2)}{(1 - zC(z)^2)^{2i-3}}.$$

5   Computational experiments

As curious as Conjecture 5 is from a purely mathematical perspective, the alternative scaling of the matrices of the spectral measure seems to have advantages in the approximation of the limit measure $\bar{\mu}^c$. There are certain important details to take into account though.

Since the $V_i(z)$ in Corollary 13 are polynomials in $zC(z)^2$ multiplied by $C(z)$, the evaluation $\frac{1}{z}V_i\left(\frac{1}{z}\right)$ is a polynomial in the Stieltjes transform of the semicircle law. The inversion formula (3) therefore always yields densities of signed measures with zero mass for these Stieltjes transforms. In particular, we obtain a sequence of densities $f_i(z)$ from the Stieltjes transforms $\frac{1}{z}V_i\left(\frac{1}{z}\right)$ for $1 \le i \le 5$ which are given by

$$f_0(z) = \frac{1}{2\pi}\sqrt{4-z^2}\,\mathbb{1}_{(-2,2)}(z),$$

$$f_1(z) = \frac{1}{2\pi}\left(1-z^2\right)\sqrt{4-z^2}\,\mathbb{1}_{(-2,2)}(z),$$

$$f_2(z) = \frac{1}{2\pi}\left(1-6z^2+5z^4-z^6\right)\sqrt{4-z^2}\,\mathbb{1}_{(-2,2)}(z),$$

$$f_3(z) = \frac{1}{2\pi}\left(9-140z^2+358z^4-299z^6+98z^8-11z^{10}\right)\sqrt{4-z^2}\,\mathbb{1}_{(-2,2)}(z),$$

$$\begin{aligned}f_4(z) = \frac{1}{2\pi}\Big(&56+1602z^2-8625z^4+16004z^6\\&-13447z^8+5624z^{10}-1143z^{12}+90z^{14}\Big)\sqrt{4-z^2}\,\mathbb{1}_{(-2,2)}(z),\end{aligned}$$

$$\begin{aligned}f_5(z) = \frac{1}{2\pi}\Big(&442-17946z^2+171911z^4-574676z^6+904447z^8\\&-768354z^{10}+373181z^{12}-103622z^{14}+15298z^{16}-931z^{18}\Big)\sqrt{4-z^2}\,\mathbb{1}_{(-2,2)}(z).\end{aligned}$$

Now, it is easy to see that the coefficients of the polynomial factors of the $f_i(z)$ grow rapidly and that these functions oscillate quite heavily. Hence, there exists a largest integer $t(c)$ depending on c such that

$$\sum_{\xi=0}^{t(c)}\frac{1}{c^\xi}f_\xi(z)$$

takes non-negative values on the interval $(-2,2)$ and is therefore the density of a probability measure. Experiments for $c=5,10,20$ show that this $t(c)$ seems to be the right scaling for $\bar{\mu}^c$ such that most of the eigenvalues are exactly in the interval $(-2,2)$. This is reminiscent of divergent asymptotic expansions (see *e.g.* the introduction of [14]). For example, consider Stirling's asymptotic expansion $n! \approx n^n e^{-n}\sqrt{2\pi n}(s_0+s_1 n^{-1}+s_2 n^{-2}+\cdots)$ where $(s_0,s_1,s_2,\ldots)=(1,\frac{1}{12},\frac{1}{288},\ldots)$. For any n, there exists $t(n)$ such that the accuracy of the approximation of order k improves for k from 0 to $t(n)$, then decreases with k.

Further, the densities seem to approximate the histograms of eigenvalues of sampled matrices quite well. In Table 1, we can see histograms of random matrices with $p=5/n$. In each row, we sampled N matrices of size $n \times n$ such that we always obtained 100000 eigenvalues. They were scaled by $\sqrt{c(1+1/c)}$ such that we would expect a reasonable approximation by the density $f_0(z)+1/cf_1(z)+1/c^2 f_2(z)$. Indeed, in the columns we see the histograms of the eigenvalues in green and the densities given by the approximations of $f_0(z)$, $f_0(z)+c^{-1}f_1(z)$ and $f_0(z)+c^{-1}f_1(z)+c^{-2}f_2(z)$. As n grows, the curve of the latter fits the histogram best. Another example is illustrated in Table 2. In this case, $c=10$ and $t(c)=3$ such that we consider the densities $f_0(z)+c^{-1}f_1(z)$, $f_0(z)+c^{-1}f_1(z)+c^{-2}f_2(z)$ and $f_0(z)+c^{-1}f_1(z)+c^{-2}f_2(z)+c^{-3}f_3(z)$.

Table 1 Histograms (100 bins) of eigenvalues of N random adjacency matrices of $G(n, 5/n)$ compared to the densities $f_0(z), f_1(z)$ and $f_2(z)$.

sample size	$f_0(z)$	$f_0(z) + \frac{1}{5}f_1(z)$	$f_0(z) + \frac{1}{5}f_1(z) + \frac{1}{25}f_2(z)$
n=40 N=2500			
n=200 N=500			
n=1000 N=100			
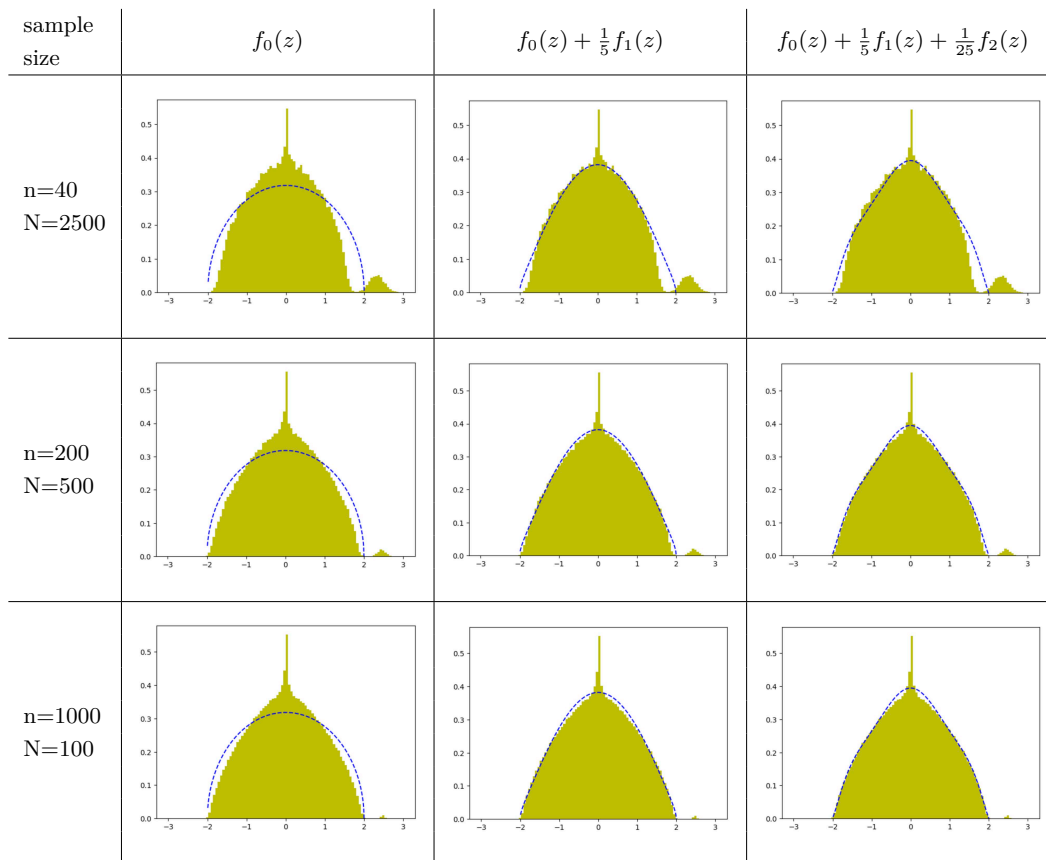
References

1. Michel Bauer and Olivier Golinelli. Random incidence matrices: Moments of the spectral density. *Journal of Statistical Physics*, 103:301–337, 2001. `doi:10.1023/A:1004879905284`.

2. François Bergeron, Gilbert Labelle, and Pierre Leroux. *Combinatorial Species and Tree-like Structures*. Cambridge University Press, 1997.

3. Charles Bordenave, Marc Lelarge, and Justin Salez. The rank of diluted random graphs. *Annals of Probability*, 39(3):1097–1121, 2011. `doi:10.1214/10-AOP567`.

4. Charles Bordenave, Arnab Sen, and Bálint Virág. Mean quantum percolation. *Journal of the European Mathematical Society*, 19(12):3679–3707, 2017. `doi:10.4171/JEMS/750`.

5. JT Chayes, L Chayes, Judy R Franz, James P Sethna, and SA Trugman. On the density of state for the quantum percolation problem. *Journal of Physics A: Mathematical and General*, 19(18):L1173, 1986. `doi:10.1088/0305-4470/19/18/011`.

6. Simon Coste and Justin Salez. Emergence of extended states at zero in the spectrum of sparse random graphs. *Annals of Probability*, 49(4), 2021. `doi:10.1214/20-AOP1499`.

7. Alan Edelman and N Raj Rao. Random matrix theory. *Acta numerica*, 14:233–297, 2005. `doi:10.1017/S0962492904000236`.

8. Nathanaël Enriquez and Laurent Ménard. Spectra of large diluted but bushy random graphs. *Random Structures & Algorithms*, 49(1):160–184, 2016. `doi:10.1002/rsa.20618`.

9. Philippe Flajolet and Robert Sedgewick. *Analytic Combinatorics*. Cambridge University Press, 2009. URL: `http://www.cambridge.org/uk/catalogue/catalogue.asp?isbn=9780521898065`.

■ **Table 2** Histograms (200 bins) of eigenvalues of N random adjacency matrices of $G(n, 10/n)$ compared to the densities $\tilde{f}_1(z)$, $\tilde{f}_2(z)$ and $\tilde{f}_3(z)$.
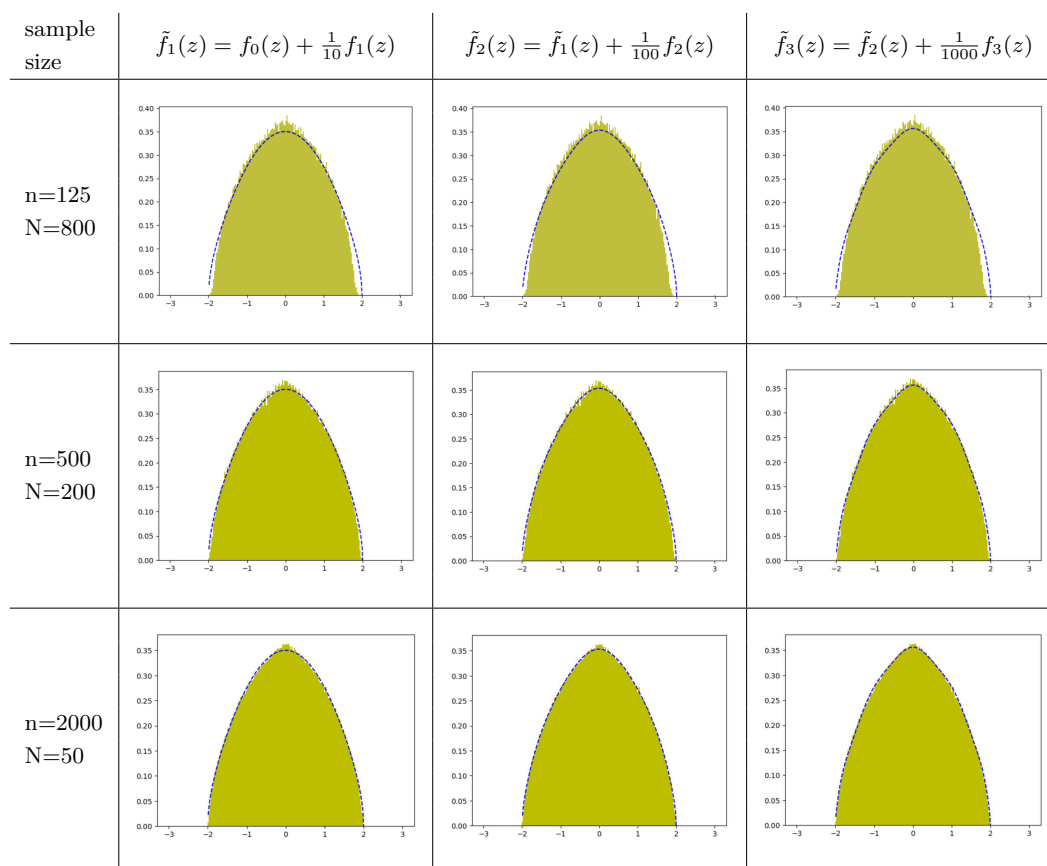
sample size	$\tilde{f}_1(z) = f_0(z) + \frac{1}{10}f_1(z)$	$\tilde{f}_2(z) = \tilde{f}_1(z) + \frac{1}{100}f_2(z)$	$\tilde{f}_3(z) = \tilde{f}_2(z) + \frac{1}{1000}f_3(z)$
n=125 N=800			
n=500 N=200			
n=2000 N=50			

10   Zoltán Füredi and János Komlós. The eigenvalues of random symmetric matrices. *Combinatorica*, 1:233–241, 1981. `doi:10.1007/BF02579329`.

11   Ira M Gessel. Lagrange inversion. *Journal of Combinatorial Theory, Series A*, 144:212–249, 2016. `doi:10.1016/j.jcta.2016.06.018`.

12   Eva-Maria Hainzl and Élie de Panafieu. Tree walks and the spectrum of random graphs. *arXiv*, arXiv:2405.08347:26, 2024. `doi:10.48550/arXiv.2405.08347`.

13   Brendan D McKay. The expected eigenvalue distribution of a large regular graph. *Linear Algebra and its applications*, 40:203–216, 1981. `doi:10.1016/0024-3795(81)90150-6`.

14   Frank Olver. *Asymptotics and special functions*. CRC Press, 1997. `doi:10.1201/9781439864548`.

15   Justin Salez. Every totally real algebraic integer is a tree eigenvalue. *Journal of Combinatorial Theory, Series B*, 111:249–256, 2015. `doi:10.1016/j.jctb.2014.09.001`.

16   Justin Salez. Spectral atoms of unimodular random trees. *Journal of the European Mathematical Society (EMS Publishing)*, 22(2), 2020. `doi:10.4171/JEMS/923`.

17   N.J.A. Sloane. sequence a294439. *The on-line encyclopedia of integer sequences*, 2017.

18   Linh V Tran, Van H Vu, and Ke Wang. Sparse random graphs: Eigenvalues and eigenvectors. *Random Structures & Algorithms*, 42(1):110–134, 2013. `doi:10.1002/rsa.20406`.

19   Antonia M Tulino, Sergio Verdú, et al. Random matrix theory and wireless communications. *Foundations and Trends® in Communications and Information Theory*, 1(1):1–182, 2004. `doi:10.1561/0100000001`.

20 Eugene P Wigner. Characteristic vectors of bordered matrices with infinite dimensions. *Annals of Mathematics*, pages 548–564, 1955. `doi:10.2307/1970079`.

21 Eugene P Wigner. On the distribution of the roots of certain symmetric matrices. *Annals of Mathematics*, 67(2):325–327, 1958. `doi:10.2307/1970008`.

22 Eugene P Wigner. Random matrices in physics. *SIAM review*, 9(1):1–23, 1967. `doi:10.1137/1009001`.

23 Edward M Wright. The number of connected sparsely edged graphs I-III. *Journal of Graph Theory*, 1977-80.

24 Inna Zakharevich. A generalization of wigner's law. *Communications in Mathematical Physics*, 268:403–414, 2006. `doi:10.1007/s00220-006-0074-5`.

Asymptotics of Weighted Reflectable Walks in A_2

Torin Greenwood ✉ 🏠 ⓘ
Department of Mathematics, North Dakota State University, Fargo, ND, USA

Samuel Simon ✉ 🏠
Department of Mathematics, Simon Fraser University, Burnaby, Canada

──── **Abstract** ──

Lattice walks are used to model various physical phenomena. In particular, walks within Weyl chambers connect directly to representation theory via the Littelmann path model. We derive asymptotics for centrally weighted lattice walks within the Weyl chamber corresponding to A_2 by using tools from analytic combinatorics in several variables (ACSV). We find universality classes depending on the weights of the walks, in line with prior results on the weighted Gouyou-Beauchamps model. Along the way, we identify a type of singularity within a multivariate rational generating function that is not yet covered by the theory of ACSV. We conjecture asymptotics for this type of singularity.

1   Lattice walks

Lattice walks have a rich history both as a model of phenomena throughout math and science, and as a driving force for the development of new analytic techniques to extract asymptotics from general combinatorial problems. For example, lattice walks have modeled melting phenomena in statistical mechanics [13], diffusion and Brownian motion [1], queueing systems [10], and Young diagrams [17, 22]. Additionally, lattice walks have pushed forward the techniques of analytic combinatorics in several variables (ACSV), as the categorization of increasingly many families of lattice walks has continually stretched the limits of generating functions one can analyze [6, 7, 28].

This work continues the tradition, studying the asymptotics of reflectable weighted lattice walks within a Weyl chamber. While this family of walks has direct connections to the Littelmann path model and representation theory [26], the analysis here also reveals a type of singularity within a generating function previously unseen in applications. Our main results include leading term asymptotics for weighted walks in the Tandem and Double Tandem models for almost all choices of *central weightings*, as defined in Section 1.2. Additionally, Conjecture 8 predicts asymptotics generally for generating functions in the new singularity regime we identified, based on merging the results on several related types of singularities.

A *lattice model* in d dimensions is defined by a finite stepset $\mathcal{S} \subset \mathbb{Z}^d$. A *lattice walk* of length n, or *lattice path* of length n, is a sequence $w = (w_1, w_2, \cdots, w_n)$ of steps $w_j \in \mathcal{S}$. After m steps, the walk is at the point given by $\sum_{i=1}^{m} w_i$. We consider counting the number

of walks restricted to the Weyl chamber A_2, defined in Section 1.3 below. As in Figure 1, we will find that the walks we study could also be viewed as walks in the positive quarter plane, although the Weyl chamber interpretation allows us to use the generalized reflection principle [14] to derive a generating function encoding the walks.

1.1   Walks in restricted regions

Dyck paths form a prototypical one-dimensional lattice path enumeration problem with a domain restriction: Dyck paths of length $2n$ start and end at 0, take $2n$ steps from $\{1, -1\}$, and always remain at or above the point 0. One way to show Dyck paths are enumerated by the Catalan numbers is to use the *reflection principle*, where paths that do cross below 0 are mapped bijectively to paths that are easier to count.

Natural extensions include counting walks in higher dimensions, with different stepsets, or in other restricted regions. For one-dimensional walks, [2] provides a generating function and asymptotic formula for restricted walks with general weighted stepsets, which assign a positive weight to each step. Moving up one dimension, walks in the half plane $\mathbb{Z} \times \mathbb{N}$ can sometimes be reduced to pairs of one-dimensional weighted walks by treating the horizontal and vertical coordinates as independent walks.

When walks are otherwise restricted in multiple dimensions, the analysis is substantially more involved. For walks in the quarter plane, [10, 6] provided a systematic approach for deriving a generating function for broad classes of stepsets, instead of developing ad-hoc methods for individual stepsets. Symmetry plays a major role in computing generating functions, which we explore in Section 1.3. Many additional works have contributed to the study of walks in the positive quadrant, including [25, 4, 24, 32].

In [27], asymptotics are found for walks in the positive d-dimensional orthant with highly symmetric nontrivial stepsets using the kernel method. The authors of [27] express the generating function as the diagonal of a multivariate rational function. They give asymptotics for such unweighted walks as a function of the stepset and number of dimensions. By adding one degree of freedom, work in [28] generalized these results and determined asymptotics for stepset models which are symmetric over all but one axis.

Considering other domain restrictions, [8] gives asymptotic behavior of a multidimensional random walk in a general cone, including in Weyl chambers. In this work, Denisov and Wachtel provide a formula for counting the number of walks of length n between two specified points in d-dimensional space. They show that such walks have asymptotics of the form $K \cdot \rho^n \cdot n^{-\ell - d/2}$. The value of ℓ is given as a function of the smallest eigenvalue of the Laplace-Beltrami operator, which can add a barrier to directly applying their theorem. Furthermore, their approach can not give an explicit expression for the constant factor in the asymptotics. The work of [9] extends these results to additional cases, where a parameter of the weighted walks called the *drift* no longer needs to be zero.

Bostan, Raschel, and Salvy make explicit the results of Denisov and Wachtel in the case $d = 2$ with the cone $\mathcal{R} = \mathbb{N}^2$. They determine asymptotic formulas for excursions for all 79 small step models in the quarter plane [5]. Bogosel et al. [3] further extract results from Denisov and Wachtel and make explicit the cases $\mathcal{S} \subset \{-1, 0, 1\}^3 \setminus \{\mathbf{0}\}$ with the cone $\mathcal{R} = \mathbb{N}^3$. They study three-dimensional excursions by associating a spherical triangle to each model.

1.2   Weighted walks

Many discrete models require non-uniform probabilities on the steps. Assigning weights to steps in a given model allows for a more detailed analysis of the asymptotic counting function. Through asymptotic analyses with weights, we can discover relations between aspects of the model and the asymptotic formula for the number of walks.

If each step w_i in a walk $(w_1, w_2, \cdots, w_n)$ has associated weight a_i, define the weight of the walk as $\prod_{i=1}^{n} a_i$. If the weights are positive integers, we can interpret the weighted model as allowing colors or multisets of steps. Weights could represent probabilities when they sum to 1. We restrict our attention to *central weights*, which are defined by the property that two walks having same length and endpoints must have the same weight. Central weights can equivalently be defined by assigning a weight to each orthogonal axis. We write $\boldsymbol{\alpha} = (a, b)$ for two-dimensional central weights.

One goal of the work here is to provide an explicit connection between the weights of the steps in a walk and the subexponential asymptotic behavior of the walks. This relationship is depicted in Figure 2, illustrating the transitions between various subexponential regimes. Because this description may be difficult to extract from the general results of [8], we prove the results directly.

Most similar to our results, a weighted version of the Gouyou-Beauchamps (GB) model was studied in [7], following the work in [6, 4] on the unweighted model. Here, the stepset is $\mathcal{S} = \{(1,0), (-1,0), (-1,1), (1,-1)\}$, and the coordinates of the steps are centrally weighted with $a, b > 0$. In [7, Theorem 1], the authors showed asymptotics are always of the form $\kappa V^{[n]}(i,j)\rho^n n^{-r}$ for constants ρ and r that depend on the weights a and b, and a harmonic function $V^{[n]}(i,j)$ depending on the weights and parity of n. In particular, the exponential growth ρ is a continuous function of a and b across boundaries, while r is not. We observe this same behavior in Theorem 3 below.

In [7], the authors also give a diagram of the subexponential regimes for the Tandem stepset without proof that matches our subexponential regimes in Theorem 3 below, but we provide a complete description of the asymptotics with constant terms and additionally note a particularly challenging regime and a possible solution in Conjecture 8 below.

Finally, in [30], the second author and a collaborator found results for weighted walks in A_1^d for arbitrary d. Much of the work there provides a scaffold for the asymptotic analyses here, although the case of A_2 turns out to be more complicated for several reasons. In particular, when using the asymptotic integral estimate described in Theorem 9 below, the leading term for A_2 is sometimes difficult to find because many of the initial terms in the asymptotic expansion are zero. The complexity in finding leading term asymptotics implies that it would be even more challenging to find full asymptotic expansions in these cases.

1.3 Weyl chambers

Weyl groups allow us to generalize the notion of symmetric stepsets. For a broad treatment of Weyl groups, see [21]. Some core results on walks in Weyl chambers appear in [14].

▶ **Definition 1** (Reduced Root System). *For vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$, let $\sigma_{\mathbf{x}}(\mathbf{y})$ be the reflection of $\mathbf{y}$ through the hyperplane perpendicular to $\mathbf{x}$. A* reduced root system *is a finite set of vectors $\Phi \subset \mathbb{R}^d$ such that for any $\mathbf{x}, \mathbf{y} \in \Phi$: $\sigma_{\mathbf{x}}(\mathbf{y}) \in \Phi$; $\mathbf{y} - \sigma_{\mathbf{x}}(\mathbf{y})$ is an integer multiple of $\mathbf{x}$; and the only nontrivial scalar multiple of $\mathbf{x}$ in Φ is $-\mathbf{x}$.*

Root systems appear throughout math, especially in relation to Lie groups, and they capture important symmetry. Given a root system Φ, a special subset of *positive roots* Φ^+ can be chosen, where for each $\alpha \in \Phi$, exactly one of $\pm\alpha$ is in Φ^+, and also if $\alpha, \beta \in \Phi^+$ and $\alpha + \beta \in \Phi$, then $\alpha + \beta \in \Phi^+$. Then, as one more refinement, the elements of Φ^+ which cannot be decomposed into sums of elements from Φ^+ form a *base* for Φ.

The isometries defined by $\{\sigma_{\mathbf{x}} : \mathbf{x} \in \Phi\}$ form a group under composition, called a *Weyl group*. Additionally, the collection of hyperplanes associated to all of the isometries of the Weyl group partition $\mathbb{R}^d$ into regions called *Weyl chambers*, as illustrated on the left in
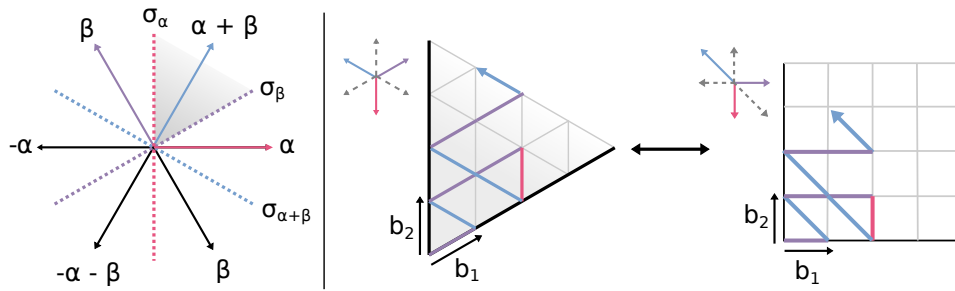
■ **Figure 1** The root system $\Phi = \{\pm\alpha, \pm\beta, \pm(\alpha+\beta)\} \subset \mathbb{R}^2$ appears on the left with a colored choice of positive roots. The dotted lines illustrate the hyperplanes defining the Weyl group of reflections, A_2. The fundamental Weyl chamber is shaded. On the right, a walk in the chamber using the Tandem model stepset (colored), and the corresponding walk in the positive quadrant of $\mathbb{Z}^2$. The Double Tandem stepset additionally includes the dashed lines.

Figure 1. One of the chambers consists of points $v \in \mathbb{R}^d$ such that $\langle \gamma, v \rangle > 0$ for all $\gamma \in \Phi$, and this chamber is called the *fundamental* or *principal Weyl chamber*. The root system, group of isometries, and principal Weyl chamber for A_2 are shown on the left in Figure 1. Finally, we define a reflectable stepset with respect to the Weyl group.

▶ **Definition 2.** *Let $\mathcal{W}$ be a Weyl group acting on a real inner product space V with a distinguished basis $\mathcal{B} = (\mathbf{b}_1, \ldots, \mathbf{b}_d)$ and Weyl chamber C. We say that a nonempty set of vectors $\mathcal{S}$ is a $(\mathcal{W}, \mathcal{B})$-reflectable stepset if for all $g \in \mathcal{W}$ and $s \in \mathcal{S}$, we have $g(s) \in \mathcal{S}$, and for all $s \in \mathcal{S}$ and $1 \leq i \leq d$, there is an integer c_i such that the dot product $\langle s, \mathbf{b}_i \rangle \in \{-c_i, 0, c_i\}$.*

For A_2, there are exactly two non-equivalent reflectable stepsets up to change of basis: the Tandem and Double Tandem stepsets illustrated in the middle in Figure 1. If the basis $\{b_1, b_2\}$ is chosen as unit vectors along the edge of the cone, then we can stretch the cone to a quadrant with axes corresponding to these basis vectors. In this way, we can identify the walks within A_2 as walks in the positive quadrant of $\mathbb{Z}^2$. Define $\mathcal{S}_T = \{(1,0),(-1,1),(0,-1)\}$ for the Tandem model, and $\mathcal{S}_{DT} = \{(1,0),(-1,1),(0,-1),(0,1),(1,-1),(-1,0)\}$ for the Double Tandem model.

Crucially, while $\mathcal{S}_T$ and $\mathcal{S}_{DT}$ do not appear symmetrical in the quarter plane, they are reflectable when considered within A_2. Thus, the *generalized reflection principle* can be used to analyze the number of walks within the chamber [14, Theorem 1].

Grabiner and Magyar gave exact results for walks in Weyl Chambers [18]. Their formulas are for walks between two points staying within the designated chamber. They obtain these formulas using determinants. A number of their formulas include the hyperbolic Bessel function of the first kind of order m.

Grabiner later gave asymptotics for a number of Weyl Chambers including the region defined by $x_1 \geq x_2 \geq \cdots \geq x_d$, which corresponds to the d-candidate ballot problem [15, 16, 17]. Here, the problem was interpreted as distributions of subtableaux in order to appeal to known formula for computing and manipulating Young tableaux.

Krattenthaler [23] completed the asymptotic analysis for the number of random walks in a Weyl chamber and random walks on a circle, noting that computing the multiplicative constants remains a challenge. Feierl extends this work by giving asymptotics for the zero drift reflectable walks in type A Weyl chambers [12]. This work uses Taylor approximations and the saddle-point method to obtain asymptotics from known determinant formulas. Here, we derive results without using determinants, which leads to asymptotics of a simpler form.

2 Results

Here, we state our results for our asymptotic counts of weighted walks within A_2. For the Tandem model, we recover the universality classes as found in [7], while also computing the asymptotic constants for almost all classes. We extend this to the Double Tandem model. In the exceptional cases when $a = 1, b < 1$ or $a < 1, b = 1$, we offer conjectured asymptotics and Conjecture 8, a prediction for general asymptotics in such a regime.

▶ **Theorem 3.** *Let $\mathcal{R} = \mathbb{N}^2$ and let $\boldsymbol{\alpha} = (a, b)$. For $\mathcal{S}_T = \{(1,0),(-1,1),(0,-1)\}$ (the Tandem model) and $\mathcal{S}_{DT} = \{(1,0),(-1,1),(0,-1),(0,1),(1,-1),(-1,0)\}$ (the Double Tandem model), the number of weighted walks of length n which stay in $\mathcal{R}$ satisfies*

$$q_{(a,b)}(n) \sim \gamma \rho^n n^{-r}$$

where the exponential growth ρ and subexponential growth r for each of $\mathcal{S}_T$ and $\mathcal{S}_{DT}$ are given in Figure 2, with the starred case conjectured. The constant terms are given in [34, Tables 5.3, 5.4].
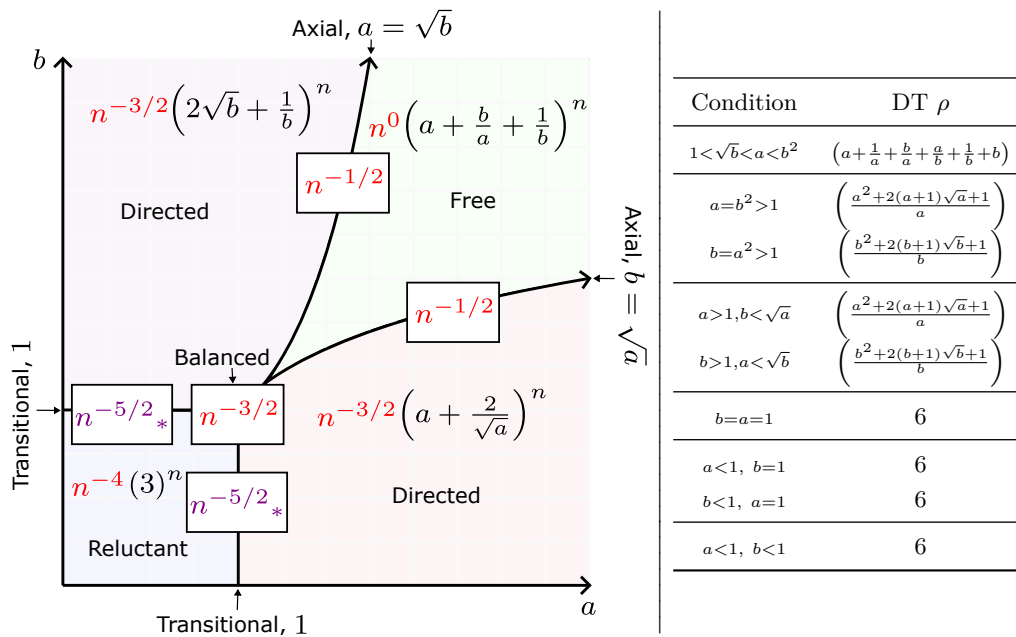
Figure 2 The Tandem and Double Tandem model have the same growth rate regimes with different exponential growth rates. The regimes for the Tandem model are pictured on the left, with the subexponential growth (in red) and exponential growth (in black). The exponential growth is continuous across boundaries, and is unmarked on the boundaries. On the right, the same regimes are listed with the corresponding exponential growth rates for the Double Tandem model. Starred cases are conjectured.

We verified the results given in Theorem 3 numerically by computing $q_{(a,b)}(n)$ exactly for specific choices of (a, b) in each regime and some large values of n (see [34, Table 5.4]). In particular, we use the gfun Maple package provided by Salvy and Zimmermann [33].

3   Extracting asymptotics

In many instances, analytic combinatorics in several variables (ACSV) provides a quick pipeline from a combinatorial description of a problem to asymptotics. Once a generating function is obtained, the singularities of the generating function can be classified. Existing libraries of results (as in [31]) describe the asymptotics of the array for many of the most common types of singularities.

We represent a d-variate multivariate rational GF as $F(\mathbf{z}) := G(\mathbf{z})/H(\mathbf{z}) = \sum a_{\mathbf{n}}\mathbf{z}^{\mathbf{n}}$, where $\mathbf{z} = (z_1, \ldots, z_d)$ and $\mathbf{z}^{\mathbf{n}} = z_1^{n_1} \cdots z_d^{n_d}$. The zero set $\mathcal{V} := \{\mathbf{z} : H(\mathbf{z}) = 0\}$ determines the singular variety of F. We seek asymptotics for $[\mathbf{z}^{\mathbf{n}}]F(\mathbf{z})$ as $\mathbf{n} \to \infty$ in a prescribed *direction* $\hat{\mathbf{r}} \in \mathbb{R}^d_{>0}$, so that $\mathbf{n} \approx \hat{\mathbf{r}}n$ with $n \to \infty$. In most combinatorial cases, finitely many *critical points* determine the asymptotics of a generating function. To find the critical points, consider representing the coefficients via the Cauchy integral formula,

$$[\mathbf{z}^{\mathbf{n}}]F(\mathbf{z}) = \left(\frac{1}{2\pi i}\right)^d \int_T F(\mathbf{z})\mathbf{z}^{-\mathbf{n}-1}d\mathbf{z}, \tag{1}$$

where T is a d-dimensional torus enclosing the origin but no singularities of F. Heuristically, the critical points are determined by expanding T until it reaches points on the singular variety closest to the origin that minimize the exponential growth $\mathbf{z}^{-\mathbf{n}}$ within the integrand.

A critical point $\mathbf{p}$ is called *smooth* if $\mathcal{V}$ is a smooth manifold in a neighborhood of $\mathbf{p}$. This means that if $\mathcal{V}$ is d-dimensional, then in a neighborhood of $\mathbf{p}$ there is a smooth parameterization of $\mathcal{V}$ using only $d-1$ variables. For rational generating functions, smoothness is easily checked using the implicit function theorem (see [31, Lemma 7.6]).

For many lattice path enumeration problems, there are also *transverse multiple points*, where $\mathcal{V}$ can locally be smoothly deformed into the intersection of perpendicular hyperplanes. For rational GFs, these types of critical points satisfy systems of polynomial equations in terms of the denominator H and its partial derivatives (see Section 4.2).

Call a critical point $\mathbf{p} = (p_1, \ldots, p_d)$ *minimal* when there are no other points $\mathbf{q} \in \mathcal{V}$ where $|q_i| \leq |p_i|$ for each coordinate with at least one inequality strict. Smooth minimal critical points always contribute to asymptotics. However, for transverse minimal critical points, an additional technical condition must be met (Definition 6). A highlight of the analysis of weighted walks in A_2 is that there is a case where the technical condition is *almost* met. Conjecture 8 predicts this halves the contribution of the critical point to the asymptotics.

4   Proof sketch

We obtain the asymptotics in Theorem 3 with the following steps:

1. **Encoding as a diagonal.** Using the symmetry group corresponding to the stepset, represent the generating function as a diagonal of a rational function.
2. **Computing critical points.** Find the solutions to the critical point equations.
3. **Finding contributing critical points.** Determine which critical points are contributing as a function of the weights.
4. **Evaluating the Cauchy integral.** Simplify the Cauchy integral (Equation (1)) to a Fourier-Laplace integral and then use existing results.

4.1 Encoding as a diagonal

For both $\mathcal{S}_T$ and $\mathcal{S}_{DT}$, the reflection group is generated by the involutions $\Psi(x, y) = (y/x, y)$ and $\Phi(x, y) = (x, x/y)$. Using either the generalized reflection principle (as in [27]) or evaluations of the unweighted generating functions in [28, Examples 6.5.1 and 6.5.2], we find that weighted walks starting at the origin of length n are encoded as coefficients of $x^n y^n t^n$ in the following functions.

$$F_T(x, y, t) = \frac{G_T(x, y)}{H_T(x, y, t)} = \frac{(b^2 x - ay^2)(bx^2 - a^2 y)(xy - ab)}{(1 - txy(\frac{a}{x} + \frac{bx}{ay} + \frac{y}{b}))a^3 b^3 (1 - x)x(1 - y)y}, \tag{2}$$

$$F_{DT}(x, y, t) = \frac{G_{DT}(x, y)}{H_{DT}(x, y, t)}$$

$$= \frac{(b^2 x - ay^2)(bx^2 - a^2 y)(xy - ab)}{(1 - txy(\frac{a}{x} + \frac{x}{a} + \frac{bx}{ay} + \frac{ay}{bx} + \frac{y}{b} + \frac{b}{y}))a^3 b^3 (1 - x)x(1 - y)y}. \tag{3}$$

4.2 Computing critical points

First, we compute all possible critical points for all values of the weights (a, b). Then, in Section 4.3, we determine which critical points contribute to asymptotics. We focus on the Tandem case here, as the Double Tandem case follows a similar analysis.

Weighted walks are encoded as the main diagonal of the functions in Equation (2), so we search for critical points in the $\mathbf{1} = (1, 1, 1)$ direction. By definition, smooth critical points satisfy $\{H = 0, xH_x = yH_y = tH_t\}$, where $H = H_T$. Next, to rule out non-smooth, non-transverse points, we verify that the factorization of H_T given in Equation (2) is a *transverse polynomial factorization* (as in [29, Definition 9.3]): define the *inventory* $S(x, y) = ax + by/ax + 1/by$, and label the factors $H_0 = (1 - txyS(1/x, 1/y))$, $H_1 = (1 - x)$, $H_2 = (1 - y)$. At any point $\mathbf{w}$ where a factor $H_i(\mathbf{w}) = 0$, its gradient $\nabla H_i(\mathbf{w})$ is nonzero, and also at any point where the factors are simultaneously zero, their gradients are linearly independent. (In fact, this applies broadly to GFs encoding other types of walks.) This implies there are no non-smooth, non-transverse points.

To find the transverse multiple points, we must consider all 7 combinations of whether H_0, H_1, and H_2 are zero, and use [29, Definition 9.7] to compute the transverse critical points for each such *stratum* individually. Conveniently, the technical definition of transverse critical points simplifies greatly in these lattice walk cases where all but one of the factors are of the form $1 - x$ and $1 - y$. For example, to compute the transverse critical points for $\mathcal{V}_{0,1}$ (where $H_0, H_1 = 0$ and $H_2 \neq 0$), the equations simplify to using the smooth critical point equations on $H_0(1, y, t)$ to compute the y and t critical point coordinates. Ultimately, we obtain the critical points in Table 1 for each stratum.

4.3 Finding contributing critical points

We now refine to contributing critical points, starting by checking minimality. The form of the generating function here is close enough to the Gouyou-Beauchamps generating function that we can reuse a result from [7].

▶ **Lemma 4** (Lemma 3 of [7]). *For the rational function $F(x, y, t)$ described by (2), when G and H are coprime the point $(x, y, t) \in \mathcal{V}$ is minimal if and only if*

$$|x| \leq 1, \qquad |y| \leq 1, \qquad |t| \leq \frac{1}{|xy|S(|\frac{1}{x}|, |\frac{1}{y}|)},$$

where the three strict inequalities do not occur simultaneously.

■ **Table 1** Critical points for each stratum, corresponding to every possible non-trivial choice of setting some of the factors $\{H_0, H_1, H_2\}$ to zero.

Stratum	x	y	t
$\mathcal{V}_0$	a	b	$\frac{1}{3ab}$
	$ae^{i(2\pi/3)}$	$be^{i(4\pi/3)}$	$\frac{e^{i(4\pi/3)}}{3ab}$
	$ae^{i(4\pi/3)}$	$be^{i(2\pi/3)}$	$\frac{e^{i(2\pi/3)}}{3ab}$
$\mathcal{V}_{0,1}$	1	$b/\sqrt{a}$	$\frac{a^{5/2}-2a}{b(a^3-4)}$
	1	$-b/\sqrt{a}$	$-\frac{a^{5/2}+2a}{b(a^3-4)}$
$\mathcal{V}_{0,2}$	$a/\sqrt{b}$	1	$\frac{2b^3-b^{3/2}}{(4b^3-1)a}$
	$-a/\sqrt{b}$	1	$\frac{2b^3+b^{3/2}}{(4b^3-1)a}$
$\mathcal{V}_{0,1,2}$	1	1	$\frac{1}{a+\frac{b}{a}+\frac{1}{b}}$

Next, we filter to minimal points minimizing the height function $|xyt|^{-1}$.

▶ **Lemma 5.** *For each value of a, b, the unique positive minimal point that minimizes the height function $|xyt|^{-1}$ is given in Table 2.*

■ **Table 2** Positive minimal critical points for choices of the weights a and b.

CP	Conditions on weights	Positive minimal critical point	Exponential growth
1	$1 < \sqrt{b} < a < b^2$	$x = 1, y = 1, t = \frac{1}{b+a/b+1/a}$	$a + \frac{b}{a} + \frac{1}{b}$
2	$a > 1, b \leq \sqrt{a}$	$x = 1, y = \frac{b}{\sqrt{a}}, t = \frac{a^{5/2}-2a}{b(a^3-4)}$	$a + \frac{2}{\sqrt{a}}$
3	$b > 1, a \leq \sqrt{b}$	$x = \frac{a}{\sqrt{b}}, y = 1, t = \frac{2b^3-b^{3/2}}{4ab^3-a}$	$2\sqrt{b} + \frac{1}{b}$
4	$a \leq 1, b \leq 1$	$x = a, y = b, t = \frac{1}{3ab}$	3

Proof (sketch). Minimizing the height $|xyt|^{-1}$ is equivalent to minimizing $|S(1/x, 1/y)|$, which we can accomplish using calculus. Both here and in arbitrary dimension, the contributing critical points display a non-obvious boolean lattice structure in the following sense. Any given critical point is minimal when each of its coordinates (except the t coordinate) is at most 1. If there is a minimal critical point with coordinate $x_i \neq 1$ and another minimal critical point with coordinate $x_j \neq 1$, then there must also be a minimal critical point where $x_i \neq 1$ and $x_j \neq 1$. This greatly simplifies the problem of finding contributing critical points because it is easy to show that the more coordinates are equal to one in a critical point, the less the corresponding exponential growth is. Then, from the boolean structure, there is never a need to compare the contributions of two different critical points with the same number of non-one coordinates.   ◀

When there are only finitely many smooth minimal critical points, we can use existing results to compute asymptotics, but we need an additional definition and criterion in the presence of transverse multiple points.

▶ **Definition 6** (Definition 9.8 of [29]). *Let $H(\mathbf{z}) = H_1(\mathbf{z}) \cdots H_m(\mathbf{z})$ be a square-free factorization of H. Fix $K = \{k_1, \cdots, k_q\} \subseteq \{1, \ldots, m\}$, and let $\mathbf{w} \in \mathbb{C}^d$ be a solution to the critical point equations for the stratum where $H_i = 0$ if and only if $i \in K$. For each $1 \leq j \leq q$ let $b_j \in \{1, \cdots, d\}$ be an index such that the partial derivative $(\partial H_{k_j}/\partial z_{b_j})(\mathbf{w}) \neq 0$. The vector*

$$\mathbf{v}_{k_j} = \frac{(\nabla_{\log} H_{k_j})(\mathbf{w})}{w_{b_j}(\partial H_j/\partial z_{b_j})(\mathbf{w})} = \left(\frac{w_1(\partial H_{k_j}/\partial z_1)(\mathbf{w})}{w_{b_j}(\partial H_j/\partial z_{b_j})(\mathbf{w})}, \dots, \frac{w_d(\partial H_{k_j}/\partial z_d)(\mathbf{w})}{w_{b_j}(\partial H_j/\partial z_{b_j})(\mathbf{w})} \right)$$

has real coordinates. The normal cone *of H at* $\mathbf{w}$ *is the set*

$$N(\mathbf{w}) = \left\{ \sum_{j=1}^{q} a_j \mathbf{v}_{k_j} : a_j > 0 \right\} \subset \mathbb{R}^d. \tag{4}$$

The point $\mathbf{w}$ *is called a* contributing point *if* $\mathbf{w}$ *is minimal,* $\mathbf{w}$ *minimizes* $|\mathbf{z}|^{-1}$ *among all minimal points, and* $\mathbf{1} \in N(\mathbf{w})$.

In some regimes below, it turns out that $\mathbf{1}$ is in the boundary of $N(\mathbf{w})$ (i.e. some a_j must be 0). Although $\mathbf{w}$ then does not meet the requirements to be a contributing point, it may still determine asymptotics. The following lemma applies to both the critical points in Table 2, and also to the critical points more generally for reflectable walks in A_2^d.

▶ **Lemma 7.** *Let* $\mathbf{w}$ *be a minimal critical point. If* $\mathbf{w}$ *has a coordinate of* 1 *and* $\mathbf{w}$ *satisfies the smooth critical point equations for* $H_0 = 1 - txyS(\frac{a}{x}, \frac{b}{y})$ *in the direction* $\mathbf{1}$, *then* $\mathbf{1}$ *is on the boundary of the normal cone* $N(\mathbf{w})$ *(see Definition 6). Otherwise,* $\mathbf{1}$ *is on interior of* $N(\mathbf{w})$.

Proof (sketch). For H_T factored as in Equation (2), we can compute $\nabla_{\log} H_i(\mathbf{w})$ for $i = 0, 1, 2$ explicitly. When $i = 1$ or 2, the logarithmic gradient is a basis vector. For $i = 0$,

$$\nabla_{\log} H_0(\mathbf{w}) = \left(-1 - xyt \left(-\frac{a}{x} + \frac{bx}{ay} \right), -1 - xyt \left(-\frac{bx}{ya} + \frac{y}{b} \right), -1 \right).$$

In cases where $H_1(\mathbf{w}) = 0$, $\mathbf{1}$ is in the interior cone if and only if $-a/x + bx/ay < 0$, and it is on the boundary if $-a/x + bx/ay = 0$. A similar statement can be made for H_2. It is then a matter of algebra to show that equality occurs exactly when the critical point equations for $H_0 = 1 - txyS(1/x, 1/y)$ are met. ◀

From Lemma 7, we find that the critical points from Table 2 always contribute except perhaps when $a = 1$ or $b = 1$. In these exceptional cases, we note that [31, Theorem 10.65] indicates that when the numerator of a GF is nonzero at the critical point, the direction being on a facet of $N(\mathbf{w})$ cuts the asymptotic contribution in half. Here, the numerator is zero, but we conjecture the idea is still true regardless.

▶ **Conjecture 8.** *When a direction* $\mathbf{r}$ *is on a facet of the normal cone* $N(\mathbf{w})$ *defined by a minimal transverse critical point* $\mathbf{w}$, *then* $\mathbf{w}$ *contributes half as much to the asymptotics as when* $\mathbf{r}$ *is in the interior.*

As with all of the other regimes for the Tandem and Double Tandem model, we have verified this conjecture numerically when $a = 1$ or $b = 1$, and our conjectured subexponential growth aligns with Figure 7 of [7]. In particular, we looked at the weights $a = 1/8$ and $b = 1$ and found that for walks of length 2000, the error between the asymptotic estimate and the exact number of walks is less than 1%. For weights $a = 1$ and $b = 1/4$ and walks of length 2000, the error was approximately 1.2%.

Note that this situation does not occur in the analysis of the Gouyou-Beauchamps walks in [7]. This is because in the transitional cases for the Gouyou-Beauchamps walks, the corresponding generating function has a factor of $1 - y$ in the numerator and denominator that cancels and makes these cases among the easier cases to analyze. This is notable in

particular because the factor of $1 - y$ in the numerator is independent of the weights in this regime. Although we too find cancellation of factors in the numerator and denominator for some regimes (see Section 4.5 below), there is no cancellation in the transitional cases for the Tandem or Double Tandem models, and indeed there is no factor in the numerator that is independent of the weights.

4.4   Evaluating the Cauchy integral

The final step is to set up the integral to compute asymptotics. Note that the textbook [31] includes results for transverse critical points that could be applied directly at this point, but for a more complete and elementary viewpoint, we include a residue approach. Beginning with the Cauchy integral equation (Equation (1)), we expand the torus T until it nears the minimal critical points in Table 2. When different minimal points from Table 2 end up being equal at certain weight values, the analysis differs in these cases because it causes cancellation between factors of G and H. Ultimately, we are left with the 9 cases as described in Figure 2. We outline here an overview of the process of extracting asymptotics. The details for each of the 9 cases can be found in [34], with an example in Section 4.5 below. We also include `SageMath` code at the following URL illustrating how to compute asymptotics in each of these cases.

<div align="center">https://github.com/TorinGreenwood/AofA-A2Walks</div>

The overall goal is to simplify the integral until it is a Fourier-Laplace type integral where the following result applies:

▶ **Theorem 9** (Theorem 7.7.3 of [20]; Lemma 5.16 of [31]). *Suppose that the functions $A(\boldsymbol{\theta})$ and $\phi(\boldsymbol{\theta})$ in r variables are smooth in a neighborhood $\mathcal{N}$ of the origin and that the gradient $\nabla\phi(\mathbf{0}) = \mathbf{0}$; the Hessian $\mathcal{H}$ of ϕ at $\mathbf{0}$ is non-singular; $\phi(\mathbf{0}) = \mathbf{0}$; and the real part of $\phi(\boldsymbol{\theta})$ is non-negative on $\mathcal{N}$. Then for each $M > 0$ there are complex constants $C_0, \ldots, C_M$ such that*

$$\int_{\mathcal{N}} A(\boldsymbol{\theta}) e^{-n\phi(\boldsymbol{\theta})} d\boldsymbol{\theta} = \left(\frac{2\pi}{n}\right)^{r/2} \det{(\mathcal{H})}^{-1/2} \cdot \sum_{j=0}^{M} C_j n^{-j} + O(n^{-M-1}). \tag{5}$$

The constants C_j are given by the formula:

$$C_j = (-1)^j \sum_{\ell \leq 2j} \frac{\mathcal{D}^{\ell+j}(A\underline{\phi}^{\ell})(\mathbf{0})}{2^{\ell+j}\ell!(\ell+j)!}, \quad with \quad \underline{\phi} := \phi - \langle \boldsymbol{\theta}, \mathcal{H}\boldsymbol{\theta} \rangle \tag{6}$$

where $\mathcal{D}$ is the differential operator $\mathcal{D} := \sum_{u,v} (\mathcal{H}^{-1})_{u,v} \frac{\partial}{\partial\theta_u} \frac{\partial}{\partial\theta_v}$.

The computational obstacle in using Theorem 9 is determining the first j for which C_j is nonzero, as this gives the subexponential growth. If G vanishes to order k at the critical point, then $C_j = 0$ for $0 \leq j \leq \lceil k/2 \rceil - 1$. Whenever the critical point is not smooth, we first take residues to reduce the number of variables in the integral and also make the singular variety smooth. Because the non-smoothness comes from factors of the form $(1 - x)$ or $(1 - y)$, it is typically straightforward to compute residues.

For example, when a critical point has x coordinate equal to 1, we can compare the value of the integral over the circle $|x| = 1 - \epsilon$ to the integral at $|x| = 1 + \epsilon$ and add a term which has smaller exponential growth, so it does not contribute to the dominant asymptotics. Then we compute the difference of the two integrals using the residue theorem, which corresponds to removing the factor of $(1 - x)$ in the denominator and evaluating the remaining function

at $x = 1$. After applying the residue, we check to see if factors between G and H now cancel, which can impact the order to which G vanishes. Then, we do a change of variables to set the integral to use Theorem 9. Lastly, we compute the C_j to obtain the asymptotics, which is completed using code. We incorporate portions of the code available in the online supplement to the textbook, [29].

4.5   Example analysis: axial regime

Here, we compute the asymptotics in the case where $a = b^2 > 1$. Equivalently, by expanding the generating function in Equation (2) as a geometric series in t, we aim for an asymptotic expression for the following:

$$q_{(a,b)}(n) := [x^0][y^0] \left(\frac{a(x - y^2)(a^{1/2}x^2 - a^2 y)(a^{3/2} - xy)}{a^{9/2}(x - 1)x(y - 1)y} \left(\frac{a}{x} + \frac{x}{a^{1/2}y} + \frac{y}{a^{1/2}} \right)^n \right).$$

The critical point that is contributing is $\left(1, \frac{b}{\sqrt{a}} \right) = (1, 1)$. However, we calculate that the direction $(1, 1)$ is not in the normal cone at this point, and is instead on the boundary. To get around this, we take the term $(x - y^2)$ in the numerator and express it at $(x - 1) - (y^2 - 1)$. Since coefficient extraction is linear, we have the following

$$q_{(a,b)}(n) = [x^0][y^0] \left(\frac{a(a^{1/2}x^2 - a^2 y)(a^{3/2} - xy)}{a^{9/2}x(y - 1)y} \left(\frac{a}{x} + \frac{x}{a^{1/2}y} + \frac{y}{a^{1/2}} \right)^n \right)$$
$$+ [x^0][y^0] \left(\frac{a(y + 1)(a^{1/2}x^2 - a^2 y)(a^{3/2} - xy)}{a^{9/2}(x - 1)xy} \left(\frac{a}{x} + \frac{x}{a^{1/2}y} + \frac{y}{a^{1/2}} \right)^n \right).$$

The first function has critical point at $(a, 1)$. The second function has critical point $(1, 1)$. Thus, the first function does not contribute to the asymptotics. The cancellation of factors here is similar to [7].

In order to obtain asymptotics from the second function, we start by taking a residue at $x = 1$. The next step is to do a change of variables to make it of Fourier-Laplace type so we can use Theorem 9. We apply the change of variables $y = e^{i\theta}, dy = ie^{i\theta}d\theta$, so the region of integration is over $[-\pi/2, 3\pi/2]$. With this transformation the integral becomes

$$\int_{[-\pi/2, 3\pi/2)} A(\theta)e^{-n\phi(\theta)}d\theta,$$

where

$$A(\theta) = \frac{(a^2 e^{i\theta} - a^{1/2})(a^{3/2} - e^{i\theta})(e^{i\theta} + 1)e^{-i\theta}}{a^{7/2}}$$

and

$$\phi(\theta) = \log \left(\frac{a + \frac{2}{\sqrt{a}}}{a^{-1/2}e^{2i\theta} + ae^{i\theta} + a^{-1/2}} \right).$$

Applying Theorem 9 gives the formula

$$q_{(a,b)}(n) \sim (a + 2a^{-1/2})^n \cdot n^{-1/2} \cdot \frac{(a^3 - 2a^{3/2} + 1)\sqrt{a^{3/2} + 2}}{\sqrt{\pi}a^3}.$$

For the Double Tandem stepset we compute

$$q_{(a,b)}(n) \sim \left(\frac{a^2 + 2(a+1)\sqrt{a} + 1}{a} \right)^n \cdot n^{-1/2}$$

$$\cdot \frac{(a^{7/2} - 2a^2 + \sqrt{a})\sqrt{2a^2 + (a^2+1)\sqrt{a} + 2a}}{\sqrt{\pi} a^4 \sqrt{a+1}}.$$

5   Next steps

The results here merely scratch the surface of possible questions about walks within Weyl chambers. An obvious next step would be to analyze the d-dimensional Tandem and Double Tandem stepsets. For example, the d-dimensional Tandem stepset has steps given by

$$\mathcal{S}_{T_d} = \{e_i - e_{i-1} : 2 \leq i \leq d\} \cup \{e_1\} \cup \{-e_d\}$$

where e_i is the ith elementary basis vector with a one in the ith coordinate and zeroes elsewhere. The first steps in computing the asymptotics are not the main obstructions. We can express the generating function for these walks as the diagonal of a rational function, and solve the critical point equations in d dimensions. We additionally find a similar structure to the contributing critical points as in the 2-dimensional case. However, there are more cases where Conjecture 8 may apply and computing constants becomes increasingly difficult.

These difficulties appear largely because applying Theorem 9 involves solving for the first nonzero C_j in Equation (5). This is in contrast to existing results for A_1^d, where the functional form of the group sum in the A_1^d case allowed the authors in [30] work through the calculations in general. In particular, the first nonzero C_j was always the first term where there are nonzero derivatives of order $2j$. For A_d it is straightforward to determine the degree to which the function vanishes at a critical point, but this is not sufficient. For A_2 when $a < 1$ and $b < 1$, the function vanishes to degree three but the constant C_2 is still zero at the critical point. It is possible that there are aspects of the governing function, coming from the Weyl denominator, that must be exploited in order to give a general statement. Even for A_3, computations can include taking 90 different mixed partial derivatives of order 24. Certainly, there are simplifications that can be made to obtain this, but it presents a barrier to quickly getting results in higher dimensions to find a pattern.

While current work has focused on the Weyl chambers of A_1^d and A_d, there are other families of interest. In particular, there may still be room to use the approach here to derive explicit asymptotic results for weighted reflectable walks for the family of Weyl groups B_d for $d > 2$. In [11] Feierl counted weighted walks in B_d using determinants, while the case of weighted reflectable walks in B_2 has been covered in [7].

More generally, one goal is to have results for walks in the product of any Weyl chambers. This would be the culmination of multiple projects, as there are not general results for all Weyl chambers. This is a plausible project as the product of the chambers should decompose in the same sense as the reflectable walks.

References

1   F. Baccelli and G. Fayolle. Analysis of models reducible to a class of diffusion processes in the positive quarter plane. *SIAM Journal on Applied Mathematics*, 47(6):1367–1385, 1987. `doi:10.1137/0147090`.

2   C. Banderier and P. Flajolet. Basic analytic combinatorics of directed lattice paths. *Theoretical Computer Science*, 281(1-2):37–80, 2002. `doi:10.1016/S0304-3975(02)00007-5`.

3 B. Bogosel, V. Perrollaz, K. Raschel, and A. Trotignon. 3d positive lattice walks and spherical triangles. *Journal of Combinatorial Theory, Series A*, 172:105189, 2020. `doi:10.1016/j.jcta.2019.105189`.

4 A. Bostan, F. Chyzak, M. van Hoeij, M. Kauers, and L. Pech. Hypergeometric expressions for generating functions of walks with small steps in the quarter plane. *European Journal of Combinatorics*, 61:242–275, 2017. `doi:10.1016/j.ejc.2016.10.010`.

5 A. Bostan, K. Raschel, and B. Salvy. Non-D-finite excursions in the quarter plane. *Journal of Combinatorial Theory, Series A*, 121(0):45–63, 2014. `doi:10.1016/j.jcta.2013.09.005`.

6 M. Bousquet-Mélou and M. Mishna. Walks with small steps in the quarter plane. *Contemporary Mathematics*, 520:1–40, 2010. `doi:10.1090/conm/520`.

7 J. Courtiel, S. Melczer, M. Mishna, and K. Raschel. Weighted lattice walks and universality classes. *Journal of Combinatorial Theory, Series A*, 152:255–302, 2017. `doi:10.1016/j.jcta.2017.06.008`.

8 D. Denisov and V. Wachtel. Random walks in cones. *The Annals of Probability*, 43(3):992–1044, 2015. URL: `http://www.jstor.org/stable/24519214`.

9 Jetlir Duraj. Random walks in cones: The case of nonzero drift. *Stochastic Processes and their Applications*, 124(4):1503–1518, 2014. `doi:10.1016/j.spa.2013.12.003`.

10 G. Fayolle, R. Iasnogorodski, and V. Malyshev. *Random walks in the quarter-plane*, volume 40 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, 1999. `doi:10.1007/978-3-319-50930-3`.

11 T. Feierl. Asymptotics for the number of walks in a Weyl chamber of type *B*. *Random Structures Algorithms*, 45(2):261–305, 2014. `doi:10.1002/rsa.20467`.

12 T. Feierl. Asymptotics for the number of zero drift reflectable walks in a Weyl chamber of type A. *Preprint arXiv:1806.05998*, 2018. `doi:10.48550/arXiv.1806.05998`.

13 M. E. Fisher. Walks, walls, wetting, and melting. *Journal of Statistical Physics*, 34(5):667–729, 1984. `doi:10.1007/BF01009436`.

14 I. M. Gessel and D. Zeilberger. Random walks in a Weyl chamber. *Proceedings of the American Mathematical Society*, 115(1):27–31, 1992. `doi:10.1090/S0002-9939-1992-1092920-8`.

15 D. J. Grabiner. Brownian motion in a Weyl chamber, non-colliding particles, and random matrices. *Annales de l'Institut Henri Poincaré Probabilités et Statistiques*, 35(2):177–204, 1999. `doi:10.1016/S0246-0203(99)80010-7`.

16 D. J. Grabiner. Random walk in an alcove of an affine Weyl group, and non-colliding random walks on an interval. *Journal of Combinatorial Theory, Series A*, 97(2):285–306, 2002. `doi:10.1006/jcta.2001.3216`.

17 D. J. Grabiner. Asymptotics for the distributions of subtableaux in Young and up-down tableaux. *Electronic Journal of Combinatorics*, 11(2):R29, 2006. `doi:10.37236/1886`.

18 D. J. Grabiner and P. Magyar. Random walks in Weyl chambers and the decomposition of tensor powers. *J. Algebraic Combin.*, 2(3):239–260, 1993. `doi:10.1023/A:1022499531492`.

19 Torin Greenwood and Samuel Simon. AofA-A2Walks. Software, swhId: `swh:1:dir:f547652d646a07bd8b230ec1a3e1c796f57bec10` (visited on 2024-07-12). URL: `https://github.com/TorinGreenwood/AofA-A2Walks`.

20 L. Hörmander. *The analysis of linear partial differential operators I: Distribution theory and Fourier analysis*. Springer, 2015. `doi:10.1007/978-3-642-61497-2`.

21 J. E. Humphreys. *Introduction to Lie algebras and representation theory*, volume 9. Springer Science & Business Media, 2012. `doi:10.1007/978-1-4612-6398-2`.

22 W. König. Orthogonal polynomial ensembles in probability theory. *Probability Surveys*, 2:385–447, 2005. `doi:10.1214/154957805100000177`.

23 C. Krattenthaler. Asymptotics for random walks in alcoves of affine Weyl groups. *Séminaire Lotharingien de Combinatoire*, 52:B52i, 2007. URL: `https://www.emis.de/journals/SLC/wpapers/s52kratt.html`.

24 C. Krattenthaler. Lattice path enumeration. *Handbook of enumerative combinatorics*, pages 589–678, 2015. `doi:10.1201/b18255`.

25   I. Kurkova and K. Raschel. On the functions counting walks with small steps in the quarter plane. *Publ. Math. Inst. Hautes Études Sci.*, 116:69–114, 2012. `doi:10.1007/s10240-012-0045-7`.

26   P. Littelmann. Characters of representations and paths in $\mathfrak{H}^*_{\mathbb{R}}$. *Representation theory and automorphic forms (Edinburgh, 1996)*, 61:29–49, 1997. `doi:10.1090/pspum/061`.

27   S. Melczer and M. Mishna. Asymptotic lattice path enumeration using diagonals. *Algorithmica*, 75(4):782–811, 2016. `doi:10.1007/s00453-015-0063-1`.

28   S. Melczer and M. C. Wilson. Higher dimensional lattice walks: Connecting combinatorial and analytic behavior. *SIAM Journal on Discrete Mathematics*, 33(4):2140–2174, 2019. `doi:10.1137/18M1220856`.

29   Stephen Melczer. *An Invitation to Analytic Combinatorics: From One to Several Variables*. Springer Nature, 2021. `doi:10.1007/978-3-030-67080-1`.

30   Marni Mishna and Samuel Simon. The asymptotics of reflectable weighted walks in arbitrary dimension. *Advances in Applied Mathematics*, 118:102043, 2020. `doi:10.1016/j.aam.2020.102043`.

31   Robin Pemantle, Mark C. Wilson, and Stephen Melczer. *Analytic Combinatorics in Several Variables*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2 edition, 2024. `doi:10.1017/9781108874144`.

32   K. Raschel. Counting walks in a quadrant: a unified approach via boundary value problems. *Journal of the European Mathematical Society*, 14(3):749–777, 2012. `doi:10.4171/JEMS/317`.

33   B. Salvy and P. Zimmermann. GFUN: a maple package for the manipulation of generating and holonomic functions in one variable. *ACM Transactions on Mathematical Software*, 20(2):163–177, 1994. `doi:10.1145/178365.178368`.

34   Samuel Lee Krumm Simon. *Walks 'n' Blocks: Asymptotic Enumeration of Weighted Reflectable Walks in A_1^d and A_2 and Exploration of Balanced Splittable Hadamard Matrices*. PhD thesis, Simon Fraser University, 2023. URL: `https://summit.sfu.ca/item/36175`.

On the Number of Distinct Fringe Subtrees in Binary Search Trees

Stephan Wagner ✉ 🄳

Institute of Discrete Mathematics, TU Graz, Austria
Department of Mathematics, Uppsala University, Sweden

⸺ Abstract ⸺

A fringe subtree of a rooted tree is a subtree that consists of a vertex and all its descendants. The number of distinct fringe subtrees in random trees has been studied by several authors, notably because of its connection to tree compaction algorithms. Here, we obtain a very precise result for binary search trees: it is shown that the number of distinct fringe subtrees in a binary search tree with n leaves is asymptotically equal to $\frac{c_1 n}{\log n}$ for a constant $c_1 \approx 2.4071298335$, both in expectation and with high probability. This was previously shown to be a lower bound, our main contribution is to prove a matching upper bound. The method is quite general and can also be applied to similar problems for other tree models.

1 Introduction

A *fringe subtree* of a rooted tree is a subtree that consists of a vertex and all its descendants, see for instance Figure 1. Fringe subtrees of random trees have been studied quite thoroughly under different models of randomness. Typical results include limit theorems for the number of fringe subtrees of a given size or shape (we will use those as an auxiliary tool in this paper as well), see for example [12, 14]. Fringe subtrees are intrinsically related to *additive functionals* of rooted trees [14–16, 19, 24], which can in fact be seen as linear combinations of fringe subtree counts. There are general limit theorems for additive functionals under different assumptions, and many relevant quantities associated with trees can be expressed as additive functionals.
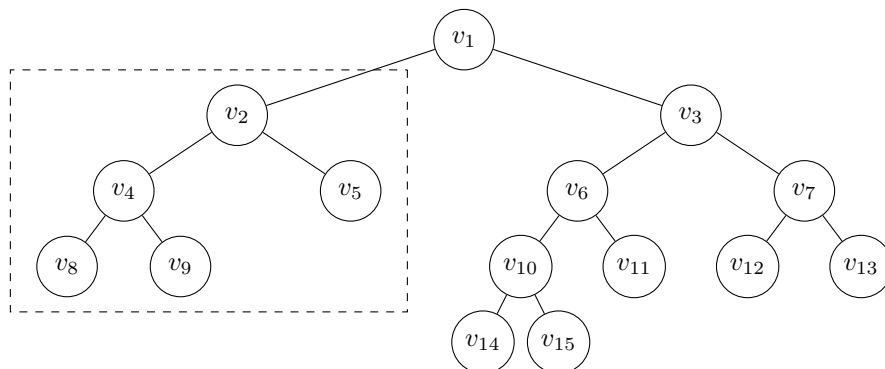
Figure 1 A binary tree. The fringe subtree rooted at v_2 is indicated by the dashed rectangle.

It is clear that an n-vertex tree has n fringe subtrees, one corresponding to each of its vertices. Usually, some of these will be identical/isomorphic as rooted trees, so the number of *distinct fringe subtrees* is generally smaller. In most of this paper, fringe subtrees will be considered identical if they are the same as plane trees (i.e., the order of the children of a vertex matters). The vertex labels are ignored. Otherwise, we regard them as distinct. There are however also other possible notions of distinctness that will be mentioned briefly in the final section.

The number of distinct fringe subtrees is connected to *tree compression*: in a fundamental algorithm to compress trees, vertices whose associated fringe subtrees have the same shape are merged to form what is called the *minimal directed acyclic graph (DAG)*. The precise shape of the tree can be recovered from the minimal DAG. Consider the tree in Figure 1 for a simple example: note that the fringe subtrees rooted at v_2 and v_6 are identical, so they are merged. For the same reason, v_4, v_7, v_{10} are merged as their fringe subtrees are identical in shape. Figure 2 shows the minimal DAG associated with the tree in Figure 1. Observe that the number of vertices of the minimal DAG is precisely the number of distinct fringe subtrees.

There are various applications of this compression technique by means of minimal DAGs. Let us mention XML compression and querying [5, 11], symbolic model checking [4] and compiler construction [1, Chapter 6.1 and 8.5] as notable examples. It is therefore of natural interest in computer science to analyse the extent to which the number of vertices is reduced by compressing a tree to its minimal DAG.
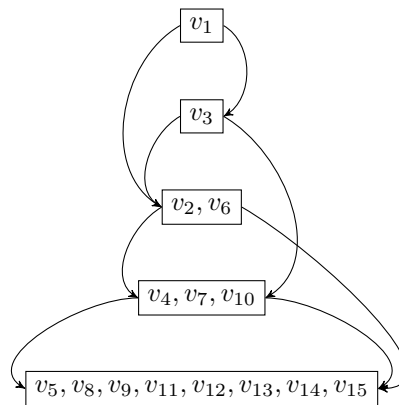
Figure 2 The minimal DAG associated with the tree in Figure 1. The vertices of the original tree that are compressed to a single vertex are listed.

For simply generated trees, it was shown by Flajolet, Sipala and Steyaert [10] that the expected size of the minimal DAG is of order $\frac{n}{\sqrt{\log n}}$. For instance, the average number of vertices in the minimal DAG associated with a uniformly random binary tree (a tree in which every internal vertex has precisely two children) with n leaves is asymptotically equal to $\frac{2n}{\sqrt{\pi \log_4 n}}$. It was also proven (see [22]) that this does not only hold in expectation, but also with high probability: in other words, with probability tending to 1, the size of the minimal DAG lies in an interval of the form $[(1 - o(1))\frac{2n}{\sqrt{\pi \log_4 n}}, (1 + o(1))\frac{2n}{\sqrt{\pi \log_4 n}}]$. The result of Flajolet, Sipala and Steyaert was further extended to Σ-labelled unranked trees in [3]. Moreover, an extension to the number of fringe subtrees that occur more than once

or generally at least a fixed number of times was considered in [20]. Interestingly, periodic fluctuations start to occur in the asymptotics: the average number of trees that occur at least r times ($r \geq 2$) as a fringe subtree is asymptotically

$$\psi_r(\log n)\frac{n}{(\log n)^{3/2}} + O\Big(\frac{n}{(\log n)^{5/2}}\Big) \tag{1}$$

for a positive periodic function ψ_r (see [20, Theorem 5.1] for the precise statement).

In this paper, we will be concerned with the model of random *binary search trees*. We consider binary trees where all internal vertices have two children: a left child and a right child. In the following, the *size* of a binary tree will always be the number of leaves; the number of internal vertices is always one less. In the probabilistic model that we study, a binary search tree is built from a random permutation of the numbers $1, 2, \ldots, n$. These numbers are stored in the internal vertices of the tree in such a way that all numbers less than the root label are in the left branch, while all numbers greater than the root label are in the right branch. See Figure 3 for an example.
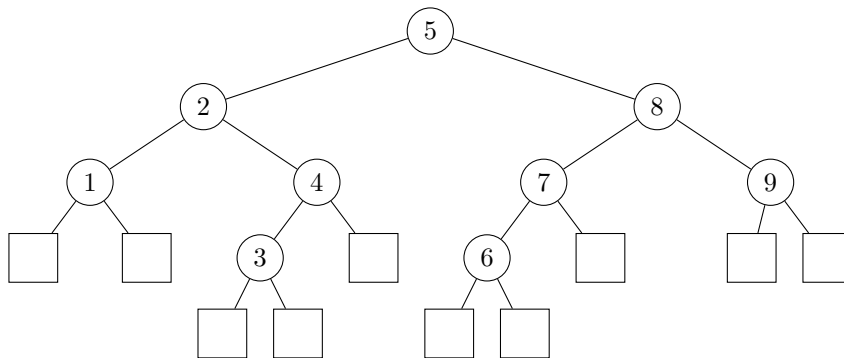
Figure 3 The binary search tree resulting from the permutation $(5, 2, 8, 4, 1, 7, 9, 3, 6)$. Internal vertices are indicated by circles, leaves by squares.

It is well known that this model is also essentially equivalent to that of *binary increasing trees* (binary trees with vertex labels that are increasing from the root to the leaves), see [7, Section 1.4.1]. For this and other types of increasing trees, the typical number of distinct fringe subtrees is of the order $\frac{n}{\log n}$ rather than $\frac{n}{\sqrt{\log n}}$. The main reason for this difference is the fact that the number of fringe subtrees with k vertices in an n-vertex tree is on average $\frac{n}{k^{3/2}}$ (asymptotically, up to a constant factor) for simply generated trees and $\frac{n}{k^2}$ for increasing trees.

The first result on binary search trees is due to Flajolet, Gourdon, and Martínez [9]. Letting F_n be the number of distinct fringe subtrees in a random binary search tree of size n, they proved that

$$\mathbb{E}(F_n) \leq \frac{(4\log 2)n}{\log n} + O\Big(\frac{n\log\log n}{(\log n)^2}\Big).$$

Devroye [6] provided a lower bound of the same order of magnitude (and also reproved the upper bound of Flajolet, Gourdon, and Martínez), showing that

$$\mathbb{E}(F_n) \geq \frac{(\log 3)n}{2\log n}(1 + o(1)).$$

The constant in the lower bound (i.e., $\frac{\log 3}{2} \approx 0.5493061443$) was improved to 0.6017824584 by Seelbach Benkner and Lohrey [21]. Seelbach Benkner and the present author [22] presented a general approach to proving results of this form. Specifically, it is shown in [22] that the number of distinct fringe subtrees is of order $\frac{n}{\sqrt{\log n}}$, both in expectation and with high probability, for simply generated trees/conditioned Bienaymé–Galton–Watson trees under various notions of what "distinct" means (e.g., distinct as plane trees, nonisomorphic as rooted trees). An analogous result for increasing trees holds with an order of magnitude of $\frac{n}{\log n}$ rather than $\frac{n}{\sqrt{\log n}}$. As a special case of the general approach, one obtains the following bounds with $c_1 \approx 2.4071298335$ and $c_2 = 4 \log 2 \approx 2.7725887222$:

$$\frac{c_1 n}{\log n}(1 + o(1)) \leq \mathbb{E}(F_n) \leq \frac{c_2 n}{\log n}(1 + o(1)),$$

which further improves the lower bound (the upper bound is identical with that of Flajolet, Sipala and Steyaert). These inequalities hold not only for the expected value, but also with high probability. Even though upper and lower bound are of the same order of magnitude and the constants c_1 and c_2 in the upper and lower bound are fairly close to each other, it is clear that there is still a gap. The aim of this paper is to close the gap and show that the constant $c_1 = 4 \sum_{k \geq 1} \frac{\log k}{(k+1)(k+2)}$ in the lower bound is in fact best possible. We will specifically prove the following theorem.

▶ **Theorem 1.** *For the constant $c_1 = 4 \sum_{k \geq 1} \frac{\log k}{(k+1)(k+2)}$, we have*

$$\mathbb{E}(F_n) \sim \frac{c_1 n}{\log n}$$

as $n \to \infty$. Moreover, we also have convergence in probability:

$$\frac{F_n}{n/\log n} \xrightarrow{p} c_1.$$

The approach taken in [22, 23] leading to the lower bound will be briefly described in Section 3. The proof of the upper bound that is required for Theorem 1 will be presented afterwards in Section 4. Before that, we require some technical results on fringe subtrees in binary search trees as well as an important invariant that is called the *shape functional*. These auxiliary results will be outlined in the following section. The paper concludes with a brief discussion and an outlook to other problems to which the same method applies.

2   Preliminaries

Let us first fix some notation. We let $\mathfrak{B}_n$ be the set of binary trees of size n (for instance, Figure 4 shows the set $\mathfrak{B}_4$), and let $\mathcal{T}_n$ be a random binary tree of size n constructed according to the random binary search tree model. In this section, we gather some results on the distribution of different random variables associated with $\mathcal{T}_n$.

2.1   The binary search tree distribution and the shape functional

We first need some auxiliary results related to the probability distribution of the shape of binary search trees. Let T be a binary tree of size n, and let N_v be the number of *internal vertices* in the fringe subtree rooted at v. We say that a binary search tree has *shape T* if the binary tree obtained by ignoring all labels is T. It is well known that the probability that the shape of a random binary search tree of size n is exactly T can be expressed as

$$p(T) = \prod_v \frac{1}{N_v},$$

the product being over all internal vertices, see for example Fill [8]. The quantity

$$(n-1)! \prod_v \frac{1}{N_v}$$

is also the number of ways to label the internal vertices with labels $1, 2, \ldots, n-1$ in an increasing fashion, i.e., in such a way that each vertex other than the root has a greater label than its parent [18, Section 5.1.4, Exercise 20].

Consider for example Figure 4: there are five possible shapes for binary search trees of size 4, occurring respectively with probability $\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}$ and $\frac{1}{3}$. The negative logarithm of $p(T)$, which can be expressed as

$$-\log p(T) = \sum_v \log N_v,$$

is called the *shape functional* of T [8] – to be more precise, it is the shape functional of the tree formed by the internal vertices.
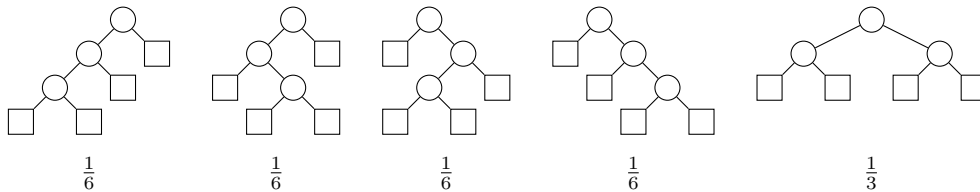
Figure 4 The five different binary trees with four leaves and their respective probabilities.

The distribution of the shape functional in random binary search trees was first studied by Fill in [8]. One can also obtain the following central limit theorem from an application of a general theorem on additive functionals due to Holmgren and Janson [14].

▶ **Lemma 2.** *Let the random variable L_n be defined by $L_n = -\log p(\mathcal{T}_n)$. We have*

$$\mathbb{E}(L_n) = \mu n + O(\log n),$$

where $\mu = \sum_{k=1}^{\infty} \frac{2 \log k}{(k+1)(k+2)}$. Moreover, $\mathbb{V}(L_n) = \sigma^2 n + O(1)$ for a constant $\sigma^2 > 0$, and the centred and normalised random variable $\frac{L_n - \mu n}{\sigma \sqrt{n}}$ converges in distribution to a standard normal distribution.

For our purposes, the asymptotic formulas for mean and variance will already be sufficient, since all we actually need is that the random variable L_n is concentrated around its mean.

2.2 The total number of fringe subtrees of a given shape or size

The second key ingredient concerns fringe subtrees that belong to a specific set. As mentioned earlier, there are many results on the number of fringe subtrees of a specific shape or size. The following lemma, which is specifically geared towards our needs, was proven (in greater generality) in [22], see also [23, Lemma 2].

▶ **Lemma 3.** *Let a, ε be two fixed positive real numbers with $\varepsilon < \frac{1}{2}$. For every positive integer k, let $\mathfrak{S}_k \subseteq \mathfrak{B}_k$ be a set whose elements are binary trees of size k. Let $p_k = \sum_{B \in \mathfrak{S}_k} p(B)$ be the probability that a random binary search tree $\mathcal{T}_k$ of size k has a shape that belongs to $\mathfrak{S}_k$.*

Now let $Z_{n,k}$ denote the (random) number of fringe subtrees of size k in a random binary search tree $\mathcal{T}_n$ of size n whose shape belongs to $\mathfrak{S}_k$. Moreover, let $Y_{n,\varepsilon}$ denote the total number of arbitrary fringe subtrees of size greater than n^{ε}. Then

(a) $\mathbb{E}(Z_{n,k}) = \frac{2np_k}{k(k+1)}$ *for all* $k < n$,

(b) $\mathbb{V}(Z_{n,k}) = O(p_k n/k^2)$ *for all* k *with* $a \log n \le k \le n^\varepsilon$,

(c) $\mathbb{E}(Y_{n,\varepsilon}) = O(n^{1-\varepsilon})$, *and*

(d) *with high probability, the following statements hold simultaneously:*

    (i) $|Z_{n,k} - \mathbb{E}(Z_{n,k})| \le p_k^{1/2} k^{-1} n^{1/2+\varepsilon}$ *for all* k *with* $a \log n \le k \le n^\varepsilon$,

    (ii) $Y_{n,\varepsilon} \le n^{1-\varepsilon/2}$.

Equipped with this and the previous lemma, we now have the necessary tools to prove both a lower bound and an upper bound that will ultimately yield Theorem 1. The interval from $a \log n$ to n^ε in Lemma 3 is such that it covers the asymptotically relevant range in the proof of Theorem 1, where we split into several parts according to the fringe subtree size.

3 The lower bound

In this section, we give a brief account of the proof of the lower bound, see [22, 23], slightly adapted to our specific situation to provide more explicit error terms than in those papers.

The key idea to bound the number of distinct fringe subtrees from below is to only consider trees that are relatively "large". Specifically, we set $k_0 := \frac{1}{\mu}(\log n + (\log n)^{3/4})$, with μ as defined in Lemma 2, and only count fringe subtrees whose size is at least k_0, while all smaller fringe subtrees are ignored. It is clear that this will give us a lower bound on the total number of distinct fringe subtrees. It turns out that for this particular choice of k_0, most fringe subtrees of size $k \ge k_0$ only occur once in the tree.

In the setting of Lemma 3, let us choose $\mathfrak{S}_k$ to be the subset of $\mathfrak{B}_k$ consisting of those trees B for which $p(B) \le \exp(-\mu k + k^{2/3})$, or equivalently $-\log p(B) \ge \mu k - k^{2/3}$. We can apply Lemma 2 to show that this condition is satisfied with high probability for random binary trees. Indeed, the Chebyshev inequality yields

$$\mathbb{P}(L_k \le \mu k - k^{2/3}) \le \frac{\mathbb{V}(L_k)}{(\mathbb{E}(L_k) - \mu k + k^{2/3})^2},$$

which by Lemma 2 becomes

$$\mathbb{P}(L_k \le \mu k - k^{2/3}) = O(k^{-1/3}).$$

Thus we can conclude that p_k in Lemma 3 is $1 - O(k^{-1/3})$ for our specific choice of $\mathfrak{S}_k$.

So the expected contribution of trees in $\mathfrak{S}_k$ for $k \ge k_0$ to the total fringe subtree count is, by part (a) of Lemma 3,

$$\sum_{k \ge k_0} \mathbb{E}(Z_{n,k}) = \sum_{k \ge k_0} \frac{2np_k}{k(k+1)} = 2n \sum_{k \ge k_0} k^{-2}\left(1 - O(k^{-1/3})\right)$$

$$= \frac{2n}{k_0}\left(1 - O(k_0^{-1/3})\right) = \frac{2\mu n}{\log n}\left(1 - O\left((\log n)^{-1/4}\right)\right). \tag{2}$$

Moreover, part (d.i) of Lemma 3 guarantees that this is also valid not just in expectation, but also with high probability.

Now we show that there are very few duplicates (identical fringe subtrees) among these. For every $k \ge k_0$, let $Z_{n,k}^{(2)}$ be the number of pairs of identical fringe subtrees in a random binary search tree of size n whose shape is in $\mathfrak{S}_k$. We condition on the total number of fringe subtrees of size k, which we denote by $X_{n,k}$. Since every fringe subtree follows, conditioned on its size k, the probability distribution of a random binary search tree $\mathcal{T}_k$, we have

$$\mathbb{E}(Z_{n,k}^{(2)} \mid X_{n,k} = N) = \binom{N}{2} \sum_{B \in \mathfrak{S}_k} p(B)^2.$$

By the definition of $\mathfrak{S}_k$, this gives us

$$\mathbb{E}(Z_{n,k}^{(2)} \mid X_{n,k} = N) \leq \binom{N}{2} e^{-\mu k + k^{2/3}} \sum_{B \in \mathfrak{S}_k} p(B) \leq \binom{N}{2} e^{-\mu k + k^{2/3}}.$$

Clearly, $X_{n,k} \leq n$, so the law of total expectation gives us

$$\mathbb{E}(Z_{n,k}^{(2)}) \leq \binom{n}{2} e^{-\mu k + k^{2/3}}.$$

Summing over all $k \geq k_0$, we finally find that

$$\sum_{k \geq k_0} \mathbb{E}(Z_{n,k}^{(2)}) \leq \binom{n}{2} \sum_{k \geq k_0} e^{-\mu k + k^{2/3}} = O\left(n^2 e^{-\mu k_0 + k_0^{2/3}}\right) = O\left(n e^{-(\log n)^{3/4} + O((\log n)^{2/3})}\right).$$

This shows that $\sum_{k \geq k_0} Z_{n,k}^{(2)}$ is (in expectation) negligible compared to $\sum_{k \geq k_0} Z_{n,k}$ (see (2)). By a standard application of the Markov inequality, this also applies with high probability.

Note that $Z_{n,k} - Z_{n,k}^{(2)}$ is a lower bound on the number of distinct fringe subtrees whose shape is in $\mathfrak{S}_k$: a shape that occurs r times contributes $r - \binom{r}{2} = \frac{r(3-r)}{2} \leq 1$ to this quantity. Moreover, the number of distinct fringe subtrees whose shape belongs to $\mathfrak{S}_k$ for some $k \geq k_0$ clearly provides a lower bound on the overall number of distinct fringe subtrees F_n, so we can conclude that

$$F_n \geq \sum_{k \geq k_0} \left(Z_{n,k} - Z_{n,k}^{(2)}\right) = \frac{2\mu n}{\log n}\left(1 - O\left((\log n)^{-1/4}\right)\right),$$

both in expectation and with high probability.

4 The upper bound

Let us now move on to the upper bound. We can express the number of distinct fringe subtrees as a sum of indicators. For every binary tree B, let $I_n(B)$ be the indicator random variable for the event that a random binary search tree of size n has a fringe subtree whose shape is B. With this definition, it is clear that

$$F_n = \sum_{k \geq 1} \sum_{B \in \mathfrak{B}_k} I_n(B).$$

The key to proving the upper bound that yields Theorem 1 is to split this sum into several parts and analyse their contributions. Specifically, the three regions are defined as follows:

- Small: $k \leq k_1 := \frac{1}{2} \log_4 n$;
- Medium: $k_1 < k \leq k_2 := \frac{1}{\mu}(\log n - (\log n)^{3/4})$, with μ as defined in Lemma 2;
- Large: $k_2 < k$.

This cutting technique is also the main idea behind many of the previously mentioned results on the quantity F_n. The novel contribution of this paper lies mainly in the middle region and its precise analysis.

4.1 Bounding the contribution of small fringe subtrees

This part is the easiest: clearly the contribution of trees whose size is at most $k_1 = \frac{1}{2} \log_4 n$ to the random variable F_n is no greater than the total number of distinct binary trees whose size is at most k_1. Since the number of possible trees for every given size k is a Catalan number (thus $|\mathfrak{B}_k| = \frac{1}{k}\binom{2k-2}{k-1} = O(4^k)$), we immediately obtain the (deterministic) upper bound

$$\sum_{k \leq k_1} \sum_{B \in \mathfrak{B}_k} I_n(B) \leq \sum_{k \leq k_1} |\mathfrak{B}_k| = O(4^{k_1}) = O(\sqrt{n}),$$

which renders all these trees negligible.

4.2   Bounding the contribution of medium-sized fringe subtrees

In the medium region, we have to perform a more careful analysis, separating trees not only by their size but also the value of their shape functional. We will split into trees with "large" shape functional and thus (relatively) low probability to occur as a fringe subtree, and trees with "small" shape functional, which have a comparatively high probability to occur. For the former, we show that the expected total number of occurrences is too low to make a significant contribution, while for the latter we prove that there are not enough distinct trees with sufficiently small shape functional to contribute to the main term of the asymptotics.

Let us now make this precise. For an integer k in the range $k_1 < k \leq k_2$, let us define a partition of $\mathfrak{B}_k$ into two subsets (depending on n) as follows:

- $\mathfrak{B}_k^1$ contains all trees $B \in \mathfrak{B}_k$ with the property that $p(B) \leq \frac{k^3}{n}$,
- $\mathfrak{B}_k^2$ contains all remaining trees in $\mathfrak{B}_k$.

Lemma 2 can be used to show that it is unlikely for the shape of a random binary search tree $\mathcal{T}_k$ to be in $\mathfrak{B}_k^1$: the inequality $p(\mathcal{T}_k) \leq \frac{k^3}{n}$ can be rewritten as $e^{-L_k} \leq \frac{k^3}{n}$, or $L_k \geq \log n - 3 \log k$. This time, the Chebyshev inequality yields

$$\mathbb{P}(L_k \geq \log n - 3 \log k) \leq \frac{\mathbb{V}(L_k)}{(\log n - 3 \log k - \mathbb{E}(L_k))^2}.$$

For $k \leq k_2$, we have $\log n - 3 \log k - \mathbb{E}(L_k) = \log n - \mu k + O(\log \log n) \geq (\log n)^{3/4} + O(\log \log n)$, thus (by Lemma 2)

$$\mathbb{P}(L_k \geq \log n - 3 \log k) = O\Big(\frac{k}{(\log n)^{3/2}}\Big).$$

So if we set $\mathfrak{S}_k = \mathfrak{B}_k^1$ in Lemma 3, then it follows that

$$p_k = \sum_{B \in \mathfrak{B}_k^1} p(B) = O\Big(\frac{k}{(\log n)^{3/2}}\Big).$$

Consequently, by part (d.i) of Lemma 3, we have, with high probability,

$$\sum_{k_1 < k \leq k_2} \sum_{B \in \mathfrak{B}_k^1} I_n(B) \leq \sum_{k_1 < k \leq k_2} Z_{n,k}$$
$$\leq \sum_{k_1 < k \leq k_2} \Big(\frac{2np_k}{k(k+1)} + \frac{p_k^{1/2} n^{1/2+\varepsilon}}{k}\Big)$$
$$= O\Big(\frac{n}{(\log n)^{3/2}}\Big).$$

Observe that this also holds in expectation (even without the term $\frac{p_k^{1/2} n^{1/2+\varepsilon}}{k}$) by part (a) of Lemma 3.

For the remaining part, we prove that there are comparatively few trees in the set $\mathfrak{B}_k^2$ as compared to $\mathfrak{B}_k^1$, even though the majority of the probability mass lies with $\mathfrak{B}_k^2$. Specifically, we bound the contribution as follows: for every $B \in \mathfrak{B}_k^2$, we have $p(B) \geq \frac{k^3}{n}$ by definition and thus

$$\sum_{B \in \mathfrak{B}_k^2} I_n(B) \leq \sum_{B \in \mathfrak{B}_k^2} 1 \leq \sum_{B \in \mathfrak{B}_k^2} \frac{np(B)}{k^3}.$$

Now by definition of $p(B)$, we have $\sum_{B \in \mathfrak{B}_k^2} p(B) \leq \sum_{B \in \mathfrak{B}_k} p(B) = 1$, thus

$$\sum_{B \in \mathfrak{B}_k^2} I_n(B) \leq \frac{n}{k^3}.$$

This inequality even holds deterministically. Finally, summing over all k in our range yields

$$\sum_{k_1 < k \leq k_2} \sum_{B \in \mathfrak{B}_k^2} I_n(B) \leq \sum_{k_1 < k \leq k_2} \frac{n}{k^3} = O\Big(\frac{n}{(\log n)^2}\Big).$$

Both this and the previous error bound that we found for $\mathfrak{B}_k^1$ are negligible compared to the term of order $\frac{n}{\log n}$ that we will obtain in the final case.

4.3 Bounding the contribution of large fringe subtrees

Finally, we look at large fringe subtrees whose size is greater than $k_2 = \frac{1}{\mu}(\log n - (\log n)^{3/4})$. Here, we apply Lemma 3 with $\mathfrak{S}_k = \mathfrak{B}_k$ for all $k > k_2$ to show that the total number of such subtrees (regardless of whether they are distinct or not) is equal to

$$\sum_{k_2 < n \leq n^\varepsilon} Z_{n,k} + Y_{n,\varepsilon} = \sum_{k_2 < n \leq n^\varepsilon} \frac{2n}{k(k+1)} + O(n^{1-\varepsilon/2}) = \frac{2n}{k_2}(1 + o(1)) = \frac{2\mu n}{\log n}(1 + o(1)),$$

both in expectation and with high probability. This term dominates the contribution of the two other cases, so we end up with

$$F_n = \sum_B I_n(B) \leq \frac{c_1 n}{\log n}(1 + o(1)), \tag{3}$$

both in expectation and with high probability. Since the matching lower bound was already provided (see Section 3), this completes the proof of Theorem 1. ◀

5 Discussion and outlook

As the proof shows, the dominant contribution to the number of distinct fringe subtrees comes from those fringe subtrees that are "large" – specifically, whose size is at least approximately $\frac{1}{\mu} \log n$. The significance of this value is as follows: above this threshold, a typical binary search tree B of size k satisfies $p(B) = o(1/n)$; as a result, the number of duplicates among the fringe subtrees of size k in $\mathcal{T}_n$ becomes insignificant, and the contribution to the number of distinct fringe subtrees is essentially just the number of fringe subtrees. Below the threshold of $\frac{1}{\mu} \log n$, it is precisely the opposite: we have $p(B) = \omega(1/n)$ (i.e., $np(B) \to \infty$) for a typical binary search tree B of size k, which ultimately leads to a negligible contribution.

Further examples of the same kind are presented in [22]: in all these examples, there are upper and lower bounds of the same order of magnitude, namely $\frac{n}{\sqrt{\log n}}$ or $\frac{n}{\log n}$. However, in most of them the constants in the bounds do not quite match.

The same technique as presented in this paper can be applied to other examples of this kind to determine the precise asymptotic behaviour of many similar quantities. To this end, one needs sufficient information on the behaviour of the analogue of the quantity $p(B)$ – specifically, a result of the same type as Lemma 2 is required.

Let us give one concrete example: the number of nonisomorphic fringe subtrees in recursive trees was studied recently by Bodini, Genitrini, Gittenberger, Larcher and Naima [2]. For this quantity, the analogue of $p(B)$ is the probability that a recursive tree of a given size

is isomorphic to a fixed unlabelled tree. The general central limit theorem for additive functionals of recursive trees due to Holmgren and Janson [14] can be applied to show that the analogue of Lemma 2 does indeed hold. At the end of the procedure, we have the following result:

▶ **Theorem 4.** *The number of nonisomorphic fringe subtrees in a random recursive tree with n vertices is asymptotically equal to $\frac{c_3 n}{\log n}$, where the constant c_3 is approximately equal to* 0.9136401430*, both in expectation and with high probability.*

The constant c_3 already appears in the lower bound in [22, Theorem 16]. The numerical computation of this constant is discussed there as well. This and further examples will be considered in the full version of this paper in a broader context.

Let us finally mention an interesting connection to the concept of *entropy* for random tree models (compare [13, 17]): recall that the constant μ in our main theorem stems from the mean of the quantity L_n (the shape functional of a random binary search tree) as given in Lemma 2. Note that we have

$$\mathbb{E}(L_n) = \mathbb{E}(-\log p(\mathcal{T}_n)) = -\sum_{B \in \mathfrak{B}_n} p(B) \log p(B),$$

which can be interpreted as the entropy of the random variable $\mathcal{T}_n$. Thus the growth constant for the number of distinct fringe subtrees is directly connected to the growth constant for the entropy. A similar interpretation is possible in other examples, such as Theorem 4.

References

1   Alfred V. Aho, Ravi Sethi, and Jeffrey D. Ullman. *Compilers: Principles, Techniques, and Tools.* Addison-Wesley series in computer science / World student series edition. Addison-Wesley, 1986.

2   Olivier Bodini, Antoine Genitrini, Bernhard Gittenberger, Isabella Larcher, and Mehdi Naima. Compaction for two models of logarithmic-depth trees: analysis and experiments. *Random Structures Algorithms*, 61(1):31–61, 2022. `doi:10.1002/rsa.21056`.

3   Mireille Bousquet-Mélou, Markus Lohrey, Sebastian Maneth, and Eric Noeth. XML compression via DAGs. *Theory of Computing Systems*, 57(4):1322–1371, 2015. `doi:10.1145/2448496.2448506`.

4   Randal E. Bryant. Symbolic boolean manipulation with ordered binary-decision diagrams. *ACM Computing Surveys*, 24(3):293–318, 1992. `doi:10.1145/136035.136043`.

5   Peter Buneman, Martin Grohe, and Christoph Koch. Path queries on compressed XML. In Johann Christoph Freytag et al., editors, *Proceedings of the 29th Conference on Very Large Data Bases, VLDB 2003*, pages 141–152. Morgan Kaufmann, 2003. `doi:10.1016/B978-012722442-8/50021-5`.

6   Luc Devroye. On the richness of the collection of subtrees in random binary search trees. *Information Processing Letters*, 65(4):195–199, 1998. `doi:10.1016/S0020-0190(97)00206-8`.

7   Michael Drmota. *Random Trees: An Interplay Between Combinatorics and Probability.* Springer, 1st edition, 2009.

8   James Allen Fill. On the distribution of binary search trees under the random permutation model. *Random Structures & Algorithms*, 8(1):1–25, 1996. `doi:10.1002/(SICI)1098-2418(199601)8:1<1::AID-RSA1>3.0.CO;2-1`.

9   Philippe Flajolet, Xavier Gourdon, and Conrado Martínez. Patterns in random binary search trees. *Random Structures & Algorithms*, 11(3):223–244, 1997. `doi:10.1002/(SICI)1098-2418(199710)11:3<223::AID-RSA2>3.0.CO;2-2`.

10   Philippe Flajolet, Paolo Sipala, and Jean-Marc Steyaert. Analytic variations on the common subexpression problem. In *Proceedings of the 17th International Colloquium on Automata, Languages and Programming, ICALP 1990*, volume 443 of *Lecture Notes in Computer Science*, pages 220–234. Springer, 1990. `doi:10.1007/BFb0032034`.

11   Markus Frick, Martin Grohe, and Christoph Koch. Query evaluation on compressed trees. In *Proceedings of the 18th Annual IEEE Symposium on Logic in Computer Science, LICS 2003*, pages 188–197. IEEE Computer Society Press, 2003. `doi:10.1109/LICS.2003.1210058`.

12   Michael Fuchs. Limit theorems for subtree size profiles of increasing trees. *Combinatorics, Probability and Computing*, 21(3):412–441, 2012. `doi:10.1017/S096354831100071X`.

13   Zbigniew Gołębiewski, Abram Magner, and Wojciech Szpankowski. Entropy and optimal compression of some general plane trees. *ACM Trans. Algorithms*, 15(1):Art. 3, 23, 2019. `doi:10.1145/3275444`.

14   Cecilia Holmgren and Svante Janson. Limit laws for functions of fringe trees for binary search trees and random recursive trees. *Electronic Journal of Probability*, 20:1–51, 2015. `doi:10.1214/EJP.v20-3627`.

15   Cecilia Holmgren, Svante Janson, and Matas Šileikis. Multivariate normal limit laws for the numbers of fringe subtrees in m-ary search trees and preferential attachment trees. *Electron. J. Combin.*, 24(2):Paper No. 2.51, 49 pp., 2017. `doi:10.37236/6374`.

16   Svante Janson. Asymptotic normality of fringe subtrees and additive functionals in conditioned Galton-Watson trees. *Random Struct. Algorithms*, 48(1):57–101, 2016. `doi:10.1002/rsa.20568`.

17   John C. Kieffer, En-Hui Yang, and Wojciech Szpankowski. Structural complexity of random binary trees. In *Proceedings of the 2009 IEEE International Symposium on Information Theory, ISIT 2009*, pages 635–639. IEEE, 2009. `doi:10.1109/ISIT.2009.5205704`.

18   Donald E. Knuth. *The art of computer programming. Vol. 3*. Addison-Wesley, Reading, MA, 1998.

19   Dimbinaina Ralaivaosaona and Stephan Wagner. A central limit theorem for additive functionals of increasing trees. *Combin. Probab. Comput.*, 28(4):618–637, 2019. `doi:10.1017/s0963548318000585`.

20   Dimbinaina Ralaivaosaona and Stephan G. Wagner. Repeated fringe subtrees in random rooted trees. In *Proceedings of the Twelfth Workshop on Analytic Algorithmics and Combinatorics, ANALCO 2015*, pages 78–88. SIAM, 2015. `doi:10.1137/1.9781611973761.7`.

21   Louisa Seelbach Benkner and Markus Lohrey. Average case analysis of leaf-centric binary tree sources. In *43rd International Symposium on Mathematical Foundations of Computer Science, MFCS 2018, August 27-31, 2018, Liverpool, UK*, pages 16:1–16:15, 2018. `doi:10.4230/LIPIcs.MFCS.2018.16`.

22   Louisa Seelbach Benkner and Stephan Wagner. Distinct fringe subtrees in random trees. *Algorithmica*, 84(12):3686–3728, 2022. `doi:10.1007/s00453-022-01013-y`.

23   Louisa Seelbach Benkner and Stephan G. Wagner. On the collection of fringe subtrees in random binary trees. In Yoshiharu Kohayakawa and Flávio Keidi Miyazawa, editors, *LATIN 2020: Theoretical Informatics - 14th Latin American Symposium, São Paulo, Brazil, January 5-8, 2021, Proceedings*, volume 12118 of *Lecture Notes in Computer Science*, pages 546–558. Springer, 2020. `doi:10.1007/978-3-030-61792-9_43`.

24   Stephan Wagner. Central limit theorems for additive tree parameters with small toll functions. *Combin. Probab. Comput.*, 24(1):329–353, 2015. `doi:10.1017/S0963548314000443`.

Early Typical Vertices in Subcritical Random Graphs of Preferential Attachment Type

Peter Mörters ✉ 🏠 ⬤
University of Cologne, Germany

Nick Schleicher[1] ✉ ⬤
University of Cologne, Germany

Abstract

We study the size of the connected component of early typical vertices in a subcritical inhomogeneous random graph with a kernel of preferential attachment type. The principal tools in our analysis are, first, a coupling of the neighbourhood of a typical vertex in the graph to a killed branching random walk and, second, an asymptotic result for the number of particles absorbed at the killing barrier in this branching random walk.

1 Introduction and statement of results

There is currently a huge demand for models of scale-free networks coming from a variety of application areas, ranging from social sciences, telecommunications to power grids. These applications lead to competing demands on the network models: On the one hand they should be amenable to mathematical and statistical analysis; models like stochastic block models or, more generally, inhomogeneous random graphs have shown to be useful here. On the other hand models should also incorporate network features beyond the scale-free distribution of the degrees. A sensible approach here is to go beyond phenomenological modelling of a scale-free network and observe which network features emerge from basic building principles. Preferential attachment, popularized by Barabási and Albert [2], has shown to be a particularly natural and interesting principle. In this paper we study, from a mathematical point of view, an inhomogeneous random graph model with a kernel that mimics the connection probabilities of preferential attachment models. We show that this model, while having many features of more complicated preferential attachment networks, allows a very fine analysis even in the difficult subcritical case, when despite the scale-free degree distribution the network exhibits only weak connectivity.

---

[1] Corresponding author

For the general *inhomogeneous graph model* [4] we take a symmetric kernel

$$\kappa : (0,1]^2 \to (0,\infty)$$

and for each $n \in \mathbb{N}$ we build the graph G_n with vertex set $V_n = \{1, \ldots, n\}$ by connecting two distinct vertices $i, j \in V_n$ independently with probability

$$p_{ij} := \frac{1}{n} \left(\kappa\left(\tfrac{i}{n}, \tfrac{j}{n}\right) \wedge n \right).$$

For example, in the *stochastic block model* the interval $(0,1]$ is partitioned into finitely many blocks and κ chosen to be constant on the cartesian product of any pair of blocks. In order to get scale-free networks, however, one uses a kernel κ with a singularity at the origin.

In preferential attachment models vertices arrive one-by-one and attach themselves to existing vertices with a preference for powerful vertices, specifically those which already have a large degree. There are various ways to turn this idea into a proper definition, but they all have in common that the expected degree of a fixed vertex i in a network of n vertices grows, as $n \to \infty$, like $\approx c(n/i)^\gamma$, for some constant c and exponent $\gamma \in (0,1)$. Choosing a connection probability of the nth vertex to each earlier vertex $i < n$ which is proportional to its expected degree at time n and picking the proportionality factor such that the expected number of connections remains bounded from zero and infinity leads to a connection probability $p_{i,n} = \beta n^{\gamma-1} i^{-\gamma}$, for some constant $\beta > 0$, which makes the model an inhomogeneous random graph with kernel

$$\kappa(x, y) = \beta(x \vee y)^{\gamma-1}(x \wedge y)^{-\gamma},$$

where $0 < \gamma < 1$ parametrizes the strength of the preferences of early vertices and $\beta > 0$ is an edge density parameter. We call this model the *inhomogeneous random graph of preferential attachment type* and explore some of its properties here.

The graph has a phase transition in the sense that if and only if the parameters γ and β are big enough, there exists a component of the graph of macroscopic size. More precisely (G_n) has a giant component if the size S_n of the largest connected component in G_n satisfies

$$\frac{S_n}{n} \to \theta > 0 \text{ in probability.}$$

For the inhomogeneous random graph of preferential attachment type we have:

▶ **Theorem 1.** *A giant component exists if and only if*

$$\gamma \geq \frac{1}{2} \quad or \quad \beta > \beta_c := \frac{1}{4} - \frac{\gamma}{2}.$$

This is a simplification of the main result in [5]. The proof can be based on taking a weak local limit in the graph, a sketch of the argument can be found in [11] .

In this paper we are primarily interested in the *subcritical regime*, i.e. when $\gamma < \frac{1}{2}$ and $0 < \beta < \beta_c$. In this case all component sizes are of smaller order than n. Our main result, Theorem 2 below, identifies the component sizes of vertices in a moving observation window, which we call early typical vertices. More precisely, a sequence of vertices $o_n \in V_n$ is called *typical* if $o_n/n \to u$ for some $u > 0$ and our observation window comprises typical vertices with small u, which are the *early typical vertices*. We show that these vertices have a connected component of asymptotic size $Y u^{-\rho_-}$ independent of n, where ρ_- is an explicitly given exponent and Y a positive random variable, whose tail behaviour we also identify.

▶ **Theorem 2.** *Let $S_n(i)$ be the size of the connected component of vertex $i \in V_n$ in the inhomogeneous random graph of preferential attachment type in the subcritical regime. If $o_n \in V_n$ is such that $\frac{o_n}{n} \to u \in (0,1]$, then*

$$\lim_{u \downarrow 0} \lim_{n \to \infty} \mathbb{P}\left(S_n(o_n) \geq u^{-\rho_-} x\right) = \mathbb{P}\left(Y \geq x\right),$$

for all $x > 0$, where

$$\rho_\pm = \tfrac{1}{2} \pm \sqrt{(\gamma - \tfrac{1}{2})^2 + \beta(2\gamma - 1)}.$$

and Y is a positive random variable satisfying

$$\mathbb{P}\left(Y \geq x\right) = x^{-(\rho_+/\rho_-)+o(1)} \text{ as } x \to \infty.$$

The remainder of the paper explains the ideas behind the proof of Theorem 2. We first look at the inner limit, when $n \to \infty$, which we investigate using a coupling of the neighbourhood of vertex o_n to a killed branching random walk. This is done in Section 2. In Section 3 we study the number of particles absorbed at the killing boundary of the branching random walk, from which our result follows.

2 Local coupling

The main object in this section is a branching random walk on the real line with a killing barrier at the origin. The branching random walk is started with a particle located in $\log u < 0$ and the displacements of the children of a vertex are given by an independent Poisson point process with intensity

$$\pi(\mathrm{d}y) = \beta(\mathrm{e}^{\gamma y}\mathbb{1}_{y>0} + \mathrm{e}^{(1-\gamma)y}\mathbb{1}_{y<0})\,\mathrm{d}y.$$

As π is an infinite measure initially every particle has infinitely many children, but we kill all particles located to the right of the killing barrier together with their offspring. As a result the killed branching process lives entirely on the negative half axis and it turns out that, for parameters $\gamma < \frac{1}{2}$ and $\beta < \beta_c$, the killed branching process becomes extinct after a finite number of generations and its genealogical tree is therefore finite. We denote this marked tree (with the vertex locations as marks) by $\mathcal{T}(u)$ and by $T(u)$ the number of vertices in this tree. The main result of this section is the following proposition.

▶ **Proposition 3.** *If $o_n \in V_n$ is such that $\frac{o_n}{n} \to u \in (0,1]$ and $x > 0$, then*

$$\lim_{n \to \infty} \mathbb{P}\left(S_n(o_n) \geq u^{-\rho_-} x\right) = \mathbb{P}\left(T(u) \geq u^{-\rho_-} x\right).$$

The proof is an adaptation of arguments in [5] to our model. It is based on a coupling of the neighbourhood of vertex o_n in the graph G_n to the killed branching random walk starting with a particle in location $\log u_n$, for a suitable sequence (u_n) with $u_n \to u$, which we will carry out in two steps in the following sections.

2.1 First step: Coupling to a random labelled tree

We first couple our graph to a tree, which we call the *random labelled tree.* Each vertex of this tree carries a label from the set $\{1, \cdots, n\}$, we denote by $\mathbb{T}_n(o)$ the tree with the root labelled by $o \in \{1, \cdots, n\}$. Every vertex with label $i \in \{1, \ldots, n\}$ produces independently for every $j \in \{1, \ldots, i-1, i+1, \ldots, n\}$ exactly one offspring of label j with probability $p_{ij} := \beta(i \vee j)^{\gamma-1}(i \wedge j)^{-\gamma} \wedge 1$. Note that different vertices in $\mathbb{T}_n(o)$ may carry the same label.

We now use a depth first search on the graph G_n to couple the connected component of o_n to the random labelled tree $\mathbb{T}_n(o_n)$. Sequences (b_n) and (c_n) with $b_n, c_n \in \{1, \ldots, n\}$, which we specify later, are used to stop the coupling when certain bad events occur. The coupling of the random labelled tree $\mathbb{T}_n(o_n)$ and the connected component of o_n in the graph G_n is defined on a probability space of Bernoulli variables. For every unordered pair $\{i, j\}$ of distinct labels in $\{1, \ldots, n\}$ we generate a sequence $P_{ij}^{(1)}, P_{ij}^{(2)}, \ldots,$ of independent Bernoulli variables $(P_{ij}^{(k)})_k$ with parameter p_{ij}. We classify all labels into one of the following categories:

Unseen labels that have not been seen in the construction,

Active labels that have been seen but not yet explored,

Passive labels that have been seen and explored.

Initially, we set $k(\{i, j\}) = 1$ for every unordered pair $\{i, j\}$ of distinct labels. We declare o_n active and all other labels unseen. In every further step, if there are no active labels left we stop and declare the coupling as *successful*. Otherwise we pick the smallest active label, say i, declare it as passive and explore it. This means that, for every $j \in \{1, \ldots, i-1, i+1, \ldots, n\}$,

- if j is unseen then $k(\{i, j\}) = 1$. We form an edge between i and j in G_n and simultaneously create a child of i with label j in $\mathbb{T}_n(o_n)$ if and only if $P_{ij}^{(1)} = 1$. If we formed an edge in this way we declare the label j as active;

- otherwise, if j is active or passive, and $P_{ij}^{(k(\{i,j\}))} = 1$ we stop and declare the coupling *unsuccessful*, if $P_{ij}^{(k(\{i,j\}))} = 0$ we change neither graph nor tree;

- we increase $k(\{i, j\})$ by one.

If after this step

- one of the active labels has $j \leq b_n$, or

- the total number of active or passive labels exceeds c_n,

then we stop and declare the coupling *unsuccessful*. If we have not stopped we continue the exploration, again with the smallest active label (if there is any).

Observe that this procedure couples the connected component of o_n in the graph G_n based on the variables $(P_{ij}^{(1)} : \{i, j\} \subset \{1, \ldots, n\})$ and the random labelled tree $\mathbb{T}_n(o_n)$ based on the variables $(P_{ij}^{(k)} : \{i, j\} \subset \{1, \ldots, n\}, k \in \mathbb{N})$ in such a way that for a successful coupling the rooted graph given as the connected component of o_n in G_n coincides with the coupled labelled tree $\mathbb{T}_n(o_n)$. As we are in the subcritical regime we do not expect to see too many labels or very small labels. Hence, for a suitable choice of the sequences (b_n) and (c_n), we expect unsuccessful coupling to be unlikely.

The main technical result of this section confirms this intuition.

▶ **Proposition 4.** *Suppose that $u \in (0, 1]$ and $\dfrac{o_n}{n} \to u$. If*

$$\lim_{n \to \infty} c_n = \infty \;\; and \;\; \lim_{n \to \infty} \frac{c_n^2}{b_n} = 0 \;\; and \;\; \lim_{n \to \infty} \frac{c_n b_n^\gamma}{n^\gamma} = 0 \,,$$

then with high probability the coupling is successful.

The simple proof is omitted.

2.2 Second step: Coupling to the killed branching random walk

The second step is to couple $\mathbb{T}_n(o_n)$ to the marked tree $\mathcal{T}(u_n)$ of the killed branching random walk started in position $\log u_n$, for a suitable $u_n \to u$. For this purpose, we have to map labels from $\{1, \cdots, n\}$ to positions in $(-\infty, 0]$. We do this using the following map

$$\phi_n \colon \{1, \ldots, n\} \to (-\infty, 0] \quad , \quad i \mapsto -\sum_{j=i+1}^{n} \frac{1}{j} \ .$$

Note that the youngest vertex is mapped to the origin, and older vertices are placed to the left with decreasing intensity. Conversely, we define the projection

$$\pi_n \colon (-\infty, 0] \to \{1, \ldots, n\}$$
$$t \mapsto \min\{m \colon t \le \phi_n(m)\}.$$

We pick $u_n = \exp(\phi_n(o_n))$ and note that $\frac{o_n}{n} \to u$ implies $u_n \to u$. We now couple $\mathcal{T}(u_n)$ to the random labelled tree $\mathbb{T}_n(o_n)$. We obtain from $\mathcal{T}(u_n)$ the *projection with labels in* $\{1, \ldots, n\}$ by taking the particles of $\mathcal{T}(u_n)$ and give each of them the label obtained by applying π_n to its position. However, the process thus obtained is not equal to $\mathbb{T}_n(o_n)$ in law. For example, a particle could have several children with the same label. A more careful coupling is therefore required.

▶ **Proposition 5.** *Suppose that $\frac{o_n}{n} \to u \in (0, 1]$ and $u_n = \exp(\phi_n(o_n))$. If $\mathcal{T}(u_n)$ contains no more than c_n vertices and no vertex in location to the left of $\phi_n(b_n)$ where*

$$\lim_{n\to\infty} \frac{c_n}{b_n^{1-\gamma}} = 0 \ and \ \lim_{n\to\infty} \frac{c_n n^\gamma}{b_n^{\gamma+1}} = 0,$$

then it can be coupled to $\mathbb{T}_n(o_n)$ so that with high probability the projection of $\mathcal{T}(u_n)$ with labels in $\{1, \ldots, n\}$ agrees with $\mathbb{T}_n(o_n)$.

We sketch the proof of Proposition 5. We start with the root, which is positioned at $\log(u_n)$. By assumption it will be projected to o_n. We now go through the particles in the lexicographical order of $\mathcal{T}(u_n)$. A vertex at location $t \in (-\infty, 0]$ is projected to $i = \pi_n(t)$ if

$$-\sum_{k=i}^{n} \frac{1}{k} < t \le -\sum_{k=i+1}^{n} \frac{1}{k}.$$

When t branches the number of children with label in $j \in \{1, \ldots, n\}$ in the projection is Poisson distributed with parameter

$$\pi\Big(\Big(-t - \sum_{k=j}^{n} \frac{1}{k}, -t - \sum_{k=j+1}^{n} \frac{1}{k}\Big)\Big).$$

If $i < j$ this is roughly

$$\pi\Big(\Big(\sum_{k=i}^{j-1} \frac{1}{k}, \sum_{k=i}^{j} \frac{1}{k}\Big)\Big) \approx \beta \int_{\log \frac{j-1}{i}}^{\log \frac{j}{i}} e^{\gamma y} \ dy = (\beta/\gamma)\big(\big(\tfrac{j}{i}\big)^\gamma - \big(\tfrac{j-1}{i}\big)^\gamma\big) \approx \beta\big(\tfrac{1}{i}\big)\big(\tfrac{j}{i}\big)^{\gamma-1} = p_{ij},$$

and if $i > j$ this is roughly

$$\pi\Big(\Big(-\sum_{k=j}^{i} \frac{1}{k}, -\sum_{k=j+1}^{i} \frac{1}{k}\Big)\Big) \approx \beta \int_{\log \frac{j}{i}}^{\log \frac{j+1}{i}} e^{(1-\gamma)y} \ dy = (\beta/1 - \gamma)\big(\big(\tfrac{j+1}{i}\big)^{1-\gamma} - \big(\tfrac{j}{i}\big)^{1-\gamma}\big)$$
$$\approx \beta\big(\tfrac{1}{i}\big)\big(\tfrac{j}{i}\big)^{-\gamma} = p_{ij}.$$

Making this more precise, if $i \geq b_n$ one can couple the Poisson random variable to a Bernoulli random variable with parameter p_{ij} with an error probability that can be summed to at most $n^\gamma b_n^{-\gamma-1}$ over all $j \in \{i+1, \ldots, n\}$, resp. $b_n^{\gamma-1}$ over all $j \in \{b_n, \ldots, i-1\}$. Summing these error probabilities over the at most c_n vertices i projected gives the result.

Proof of Proposition 3. Observe that it is possible to satisfy all the conditions on c_n and b_n imposed in Proposition 4 and 5. If the coupling of the labelled tree $\mathbb{T}_n(o_n)$ and the connected component of o_n in G_n and simultaneously with the branching process $\mathcal{T}(u_n)$ is successful we have $S_n(o_n) = T(u_n)$ and the result follows because

$$\lim_{n \to \infty} \mathbb{P}\left(S_n(o_n) \geq u^{-\rho_-}x\right) = \lim_{n \to \infty} \mathbb{P}\left(T(u_n) \geq u^{-\rho_-}x\right) = \mathbb{P}\left(T(u) \geq u^{-\rho_-}x\right),$$

using stochastic continuity of the family $(T(u)\colon u \in (0,1])$ in the last step.   ◀

3   The killed branching random walk

In this section we complete the proof of Theorem 2 by showing the following result about the killed branching random walk.

▶ **Proposition 6.** *Under the conditions of Theorem 2, for every $x > 0$,*

$$\lim_{u \downarrow 0} \mathbb{P}\left(T(u) \geq xu^{-\rho_-}\right) = \mathbb{P}\left(Y \geq x\right),$$

where Y is a positive random variable satisfying

$$\mathbb{P}\left(Y \geq x\right) = x^{-(\rho_+/\rho_-)+o(1)} \text{ as } x \to \infty.$$

The proof uses arguments from Aidekon et al. [1] in our setup. The role of the exponents $\rho_\pm$ will become clear in Section 3.1, while in Section 3.2 we will use a famous law of large number for general branching processes due to Nerman [12] to obtain the desired asymptotic.

3.1   Background on branching random walks without killing

Consider the marked tree we get from the branching random walk (without killing) with displacements given by a Poisson process Π with intensity π, where the mark of a particle x (identified, for example, by its Ulam-Harris label) corresponds to its location τ_x.

We define

$$\psi(\alpha) = \mathbb{E}\Big[\sum_{x \in \Pi} \mathrm{e}^{-\alpha\tau_x}\Big].$$

We can calculate $\psi(\alpha)$ for $\gamma < \alpha < 1 - \gamma$ exactly with Campbell's formula [8],

$$\psi(\alpha) = \mathbb{E}\Big[\sum_{x \in \Pi} \mathrm{e}^{-\alpha\tau_x}\Big] = \int \mathrm{e}^{-\alpha t}\pi(\mathrm{d}t) = \beta \int_0^\infty \mathrm{e}^{(\gamma-\alpha)t}\,\mathrm{d}t + \beta \int_{-\infty}^0 \mathrm{e}^{(1-\gamma-\alpha)t}\,\mathrm{d}t$$

$$= \frac{\beta}{\alpha-\gamma} + \frac{\beta}{1-\gamma-\alpha},$$

otherwise, for $\alpha \notin (\gamma, 1-\gamma)$, we have $\psi(\alpha) = \infty$. There exists α with $\psi(\alpha) < 1$ if and only if $\gamma < \frac{1}{2}$ and $\beta < \frac{1}{4} - \frac{\gamma}{2}$, i.e. in the subcritical regime for the inhomogeneous random graph. This is also the exact condition for the branching random walk with killing barrier at the origin to suffer extinction in finite time almost surely.

If there exists α with $\psi(\alpha) < 1$, by continuity, there exist two real numbers $\gamma < \rho_- < \rho_+ < 1 - \gamma$ with $\psi(\rho_-) = \psi(\rho_+) = 1$. We can calculate both values explicitly

$$\rho_\pm = \tfrac{1}{2} \pm \sqrt{(\gamma - \tfrac{1}{2})^2 + \beta(2\gamma - 1)}.$$

Because $\psi(\rho_-) = 1$ we obtain a nonnegative martingale (W_n) by letting W_n be the sum of $e^{-\rho_- \tau_x}$ over all the particles x (with position denoted τ_x) in the nth generation of the branching random walk. By Biggins' theorem for branching random walks, see e.g. [3, 10], the martingale limit W is strictly positive if and only if the following two conditions hold,

(i) $\log(\psi(\rho_-)) - \frac{\rho_- \psi'(\rho_-)}{\psi(\rho_-)} > 0$,

(ii) $\mathbb{E}[W_1 \log W_1] < \infty$.

The first one holds as $\psi(\rho_-) = 1$ and $\psi'(\rho_-) < 0$. For the second condition it suffices to check $\mathbb{E}[W_1^\alpha] < \infty$ for some $\alpha > 1$. For this we define

$$f(x, \Pi) = e^{-\rho_- \tau_x} \big(\sum_{y \in \Pi} e^{-\rho_- \tau_y} \big)^{\alpha - 1} .$$

Then $\mathbb{E}[W_1^\alpha] = \mathbb{E}[\int f(x, \Pi) \, \Pi(dx)]$ and by Mecke's equation [8, Theorem 4.1] we get

$$\mathbb{E}[W_1^\alpha] = \int \mathbb{E}[f(x, \Pi + \delta_x)] \, \pi(dx) = \int e^{-\rho_- x} \mathbb{E}\big[(e^{-\rho_- x} + \int e^{-\rho_- t} \, \Pi(dt))^{\alpha - 1} \big] \pi(dx)$$

$$\leq 2^{\alpha - 1} \Big(\int e^{-\alpha \rho_- x} \pi(dx) + \mathbb{E}\big[(\int e^{-\rho_- t} \, \Pi(dt))^{\alpha - 1} \big] \psi(\rho_-) \Big).$$

The right summand is finite if $1 < \alpha \leq 2$ because in this case, by Jensen's inequality, the expectation is bounded by one. The left summand is equal to $\psi(\alpha \rho_-)$ which is finite for $1 < \alpha < \frac{1 - \gamma}{\rho_-}$. Hence W is strictly positive.

3.2 Convergence of the total number of particles

We now introduce the setting of general branching processes as used in Nerman [12]. Let ξ be a point process on $[0, \infty)$. The points represent the ages at which an individual gives birth to another particle. We denote by $\mu = \mathbb{E}[\xi]$ the intensity measure of the point process. The following conditions have to be met:

(i) μ is not concentrated on any lattice,

(ii) there exists an $\alpha \in (0, \infty)$ such that $\int_0^\infty e^{-\alpha t} \mu(dt) = 1$ and

(iii) we have $\int_0^\infty t e^{-\alpha t} \mu(dt) < \infty$.

α is called the *Malthusian parameter*. A continuous-time branching process where every individual x (identified, again, by its Ulam-Harris label) gives birth to a single new individual at the times given by adding to its own birth time σ_x the points of an independent copy ξ_x of a point process as above, is called a *Crump-Mode-Jagers* or *general branching process*. We denote by $\mathcal{T}$ the set of all particles that exist in the general branching process.

The set-up of [12] allows to also include a time dependent characteristic for each particle x, but in our case it suffices to consider a random variable X_x, which may depend on ξ_x but is independent for each particle and distributed like some X. We sum X_x over all particles born before time t,

$$Z_t^X := \sum_{x \in \mathcal{T}, \sigma_x < t} X_x.$$

The following result is [12, Theorem 3.1] formulated in our set-up.

▶ **Proposition 7.** *Suppose that* $\mathbb{E}[X] < \infty$, *then*

$$\mathrm{e}^{-\alpha t} Z_t^X \to Y \text{ in probability, as } t \to \infty,$$

where Y *is a finite non-negative random variable.*

We now have all the tools to prove Proposition 6. In order to use Proposition 7 we need to derive a suitable ξ and X from π. For this purpose we take a branching random walk started with a particle at the origin, with offspring displacements given by a Poisson point process with intensity π. We do not kill particles, but we only allow particles at locations in $(-\infty, 0]$ to branch, leaving the particles in $(0, \infty)$ frozen. We let $X \geq 1$ be the total number of branching particles including the particle at the origin, which is finite because the branching random walk with killing barrier at 0 becomes extinct almost surely. We let ξ be the point process of locations of the frozen (non-branching) particles, see Figure 1.
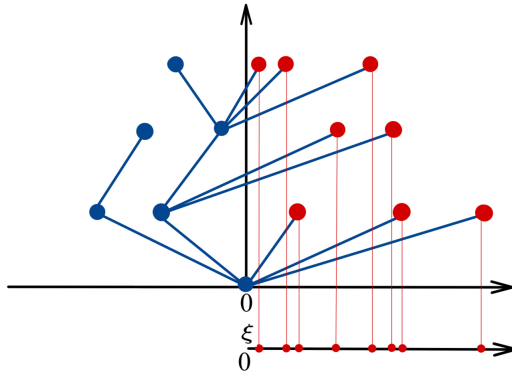
Figure 1 Branching particles are marked in blue, there are $X = 6$ in total. The positions on $[0, \infty)$ of the frozen particles, which are marked in red, yield the point process ξ.

▶ **Proposition 8.** *We have* $\mathbb{E}[X] < \infty$ *and* ξ *satisfies the conditions above for the Malthusian parameter* $\alpha = \rho_-$. *Moreover, for* $t = -\log u$, *we have*

$$Z_t^X \overset{d}{=} T(u).$$

Proof. Shifting all particle positions by $t = -\log u$ the killed branching random walk $\mathcal{T}(u)$ becomes a branching random walk $\mathcal{T}'(u)$ started with a particle at the origin, with displacements given by a Poisson point process with intensity π, with a killing barrier at $t = -\log u$. We now construct a coupling of $\mathcal{T}'(u)$ and the general branching process with ξ and X, so that the identity $Z_t^X = T'(u)$ holds, where $T'(u)$ is the number of particles of $\mathcal{T}'(u)$, which has the same law as $T(u)$.

To construct the coupling, we divide the descendants of a particle $x \in \mathcal{T}'(u)$ into branching particles to its left and frozen particles to its right, just as above. The positions of the frozen particles are the birth times of its children in the general branching process, the number of branching particles is the characteristic X_x. In this way the total progeny $T'(u)$ of the killed branching random walk equals Z_t^X, see Figure 2.

We now check that $\mathbb{E}X < \infty$. We pick $\alpha > 0$ with $\psi(\alpha) < 1$ and give a branching particle x in position $\tau_x \leq 0$ the weight $\mathrm{e}^{-\alpha \tau_x} \geq 1$. Then the expected sum over all weights of branching particles in generation n is bounded by $\psi(\alpha)^n$. Hence the total weight summed over all branching particles, and in particular the total number X of such particles, has an expectation which is bounded by $\frac{1}{1-\psi(\alpha)}$.
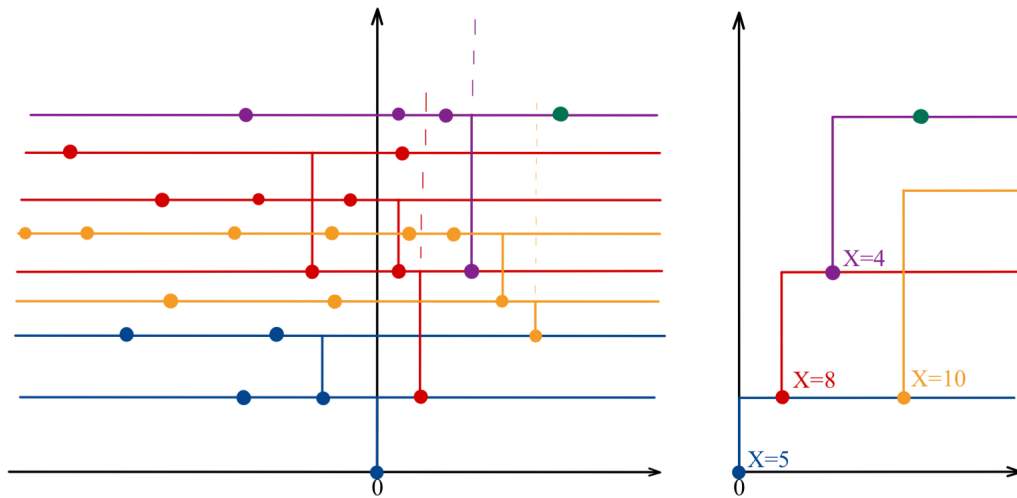
It remains to check that ρ_- is the Malthusian parameter associated to ξ. To this end we construct a martingale (M_n) as follows: We start with a particle at the origin and $M_0 = 1$. In every step we replace the leftmost particle in $(-\infty, 0]$ by its offspring chosen with displacements according to a Poisson process of intensity π and leave all other particles alive. Particles in $(0, \infty)$ never branch and remain alive but frozen. If there is no particle in $(-\infty, 0]$ the process stops and the positions of the frozen particles make up ξ. The random variable M_n is obtained as the sum of all particles x alive after the nth step weighted with $e^{-\rho_- \tau_x}$, where τ_x is the position of particle x. Because $\psi(\rho_-) = 1$ the process (M_n) is indeed a martingale, and it clearly converges almost surely to $\int_0^\infty e^{-\rho_- t}\xi(\mathrm{d}t)$. Now take $\alpha > \rho_-$ with $\psi(\alpha) < 1$. The martingale (M_n) is dominated by the random variable given as the sum over all branching particles x (with nonpositive position τ_x) weighted with $e^{-\alpha\tau_x}$ and all frozen particles x (with positive position τ_x) weighted with $e^{-\rho_- \tau_x}$. This random variable is integrable, as the sum of weights of frozen particles born from a single particle x in position $\tau_x < 0$ is independent with expectation bounded by $e^{-\alpha\tau_x}$ and the expected sum over these bounds for all branching particles is itself bounded by $\frac{1}{1-\psi(\alpha)}$, as above. We thus get (ii) from dominated convergence and hence ρ_- is the Malthusian parameter. Condition (i) is obvious and (iii) is easy to check.   ◀

To complete the proof of Proposition 6 we combine Proposition 7 and 8 to obtain

$$u^{\rho_-} T(u) \to Y \text{ in distribution, as } u \downarrow 0.$$

By [12, Theorem 6.3] the ratios of two cumulative characteristics of the same general branching process converges to a constant. Hence we get, as in [1, Lemma 21], that the limit Y is a constant multiple of the positive martingale limit W. In particular, W and Y share the same tail behaviour at infinity, which by [9, Theorem 2.2] applied to $\chi = \rho_+/\rho_-$ is given by

$$\mathbb{P}\left(W \geq x\right) = x^{-(\rho_+/\rho_-)+o(1)}.$$

4   Outlook

One of our principal aims is to find the size of the largest component in the subcritical inhomogeneous random graph of preferential attachment type. In rank one models like the configuration model or the inhomogeneous random graph with product kernel this component is known to have a size of the order of the largest degree in the graph, in our language n^γ, see [7, 6]. However, for the inhomogeneous random graph of preferential attachment type we expect this to be considerably larger because in this model powerful vertices are less well connected so that exploration beyond the first generation is still relevant. We heuristically derive a conjecture from our Theorem 2: Suppose we were allowed to let $n \to \infty$ and $u \to 0$ simultaneously. At best we could be allowed $u \approx \frac{c}{n}$. Then our hypothetic result would give that the most powerful vertices (with index o_n independent of n) would have a connected component of size $n^{\rho-}$. Our conjecture is therefore that this is the right order for the size of the largest component. Verifying this conjecture is subject of ongoing work of the authors.

───── **References** ─────

1   Elie Aïdékon, Yueyun Hu, and Olivier Zindy. The precise tail behavior of the total progeny of a killed branching random walk. *The Annals of Probability*, 41(6), November 2013. `doi:10.1214/13-aop842`.

2   Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, October 1999. `doi:10.1126/science.286.5439.509`.

3   John D. Biggins. Martingale convergence in the branching random walk. *J. Appl. Probab.*, 14:25–37, 1977. `doi:10.2307/3213258`.

4   Béla Bollobás, Svante Janson, and Oliver Riordan. The phase transition in inhomogeneous random graphs. *Random Structures and Algorithms*, 31(1):3–122, June 2007. `doi:10.1002/rsa.20168`.

5   Steffen Dereich and Peter Mörters. Random networks with sublinear preferential attachment: The giant component. *The Annals of Probability*, 41(1), January 2013. `doi:10.1214/11-aop697`.

6   Aolan Ding and Rick Durrett. The phase transition in Chung-Lu graphs, 2023. `arXiv:2305.08203`.

7   Svante Janson. The largest component in a subcritical random graph with a power law degree distribution. *The Annals of Applied Probability*, 18(4):1651–1668, 2008. `doi:10.1214/07-AAP490`.

8   Günter Last and Mathew Penrose. *Lectures on the Poisson process*, volume 7 of *IMS Textb.* Cambridge: Cambridge University Press, 2018. `doi:10.1017/9781316104477`.

9   Quansheng Liu. On generalized multiplicative cascades. *Stochastic Processes and their Applications*, 86(2):263–286, 2000. `doi:10.1016/S0304-4149(99)00097-6`.

10   Russell Lyons. A simple path to Biggins' martingale convergence for branching random walk, 1998. `arXiv:math/9803100`.

11   Peter Mörters. Tangent graphs. *Pure and Applied Functional Analysis*, 8, 2023. URL: `http://yokohamapublishers.jp/online2/oppafa/vol8/p1767.html`.

12   Olle Nerman. On the convergence of supercritical general (C-M-J) branching processes. *Z. Wahrscheinlichkeitstheor. Verw. Geb.*, 57:365–395, 1981. `doi:10.1007/BF00534830`.

Asymptotics of Relaxed k-Ary Trees

Manosij Ghosh Dastidar 🄳
Institut für Diskrete Mathematik und Geometrie, TU Wien, Austria

Michael Wallner ⌂ 🄳
Institut für Diskrete Mathematik und Geometrie, TU Wien, Austria

─── **Abstract** ───

A relaxed k-ary tree is an ordered directed acyclic graph with a unique source and sink in which every node has out-degree k. These objects arise in the compression of trees in which some repeated subtrees are factored and repeated appearances are replaced by pointers. We prove an asymptotic theta-result for the number of relaxed k-ary tree with n nodes for $n \to \infty$. This generalizes the previously proved binary case to arbitrary finite arity, and shows that the seldom observed phenomenon of a stretched exponential term $e^{cn^{1/3}}$ appears in all these cases. We also derive the recurrences for compacted k-ary trees in which all subtrees are unique and minimal deterministic finite automata accepting a finite language over a finite alphabet.

1 Introduction and Main Result

The enumeration of directed acyclic graphs (DAGs) is an important and timely topic in computer science [3], mathematics [2,13,16], and many related areas such as phylogenetics [15] and theoretical physics [11,12]. Several problems in this area have remained open for a long time, with bounds sometimes differing by an exponential factor. One of those problems is the enumeration of minimal deterministic finite automata (DFAs) with n states recognizing a finite language over a finite alphabet [14]. In [6] Elvey Price, Fang, and Wallner solved the corresponding asymptotic counting problem for a binary alphabet, and uncovered the remarkable phenomenon of a stretched exponential term $e^{c\,n^{1/3}}$. This term provides an explanation for the previously encountered difficulties. For example, the associated generating function cannot be algebraic, and it can only be D-finite (satisfy a linear differential equation with polynomial coefficients) if it has an irregular singularity.

This phenomenon was first observed by the above-mentioned authors in [7] in the asymptotic number of another class of DAGs: compacted binary trees of size n. These arise in the compression of XML documents [3], in the common subexpression problem in, e.g., compiling [5], and in data structures of, e.g., computer algebra software [8]. Since then, this phenomenon has also been shown in many classes of DAGs and related objects, such as phylogenetic networks [4], permutation patterns [18], and Young tableaux [1]. In this paper,

we show that the examples of compacted binary trees and minimal DFAs are just single cases of infinite families admitting a stretched exponential. Our main result is the following asymptotics of a super-class of the latter two.

▶ **Theorem 1.** *Let $k \geq 2$ be an integer. The number of relaxed k-ary trees satisfies for $n \to \infty$*

$$\Theta\left(n!^{k-1}\left(\frac{k^k}{(k-1)^{k-1}}\right)^n e^{3\left(\frac{k(k-1)}{2}\right)^{1/3}a_1 n^{1/3}} n^{\frac{2k-1}{3}}\right),$$

where $a_1 \approx -2.338$ is the largest root of the Airy function $\mathrm{Ai}(x)$ defined as the unique function satisfying $\mathrm{Ai}''(x) = x\mathrm{Ai}(x)$ and $\lim_{x\to\infty} \mathrm{Ai}(x) = 0$.
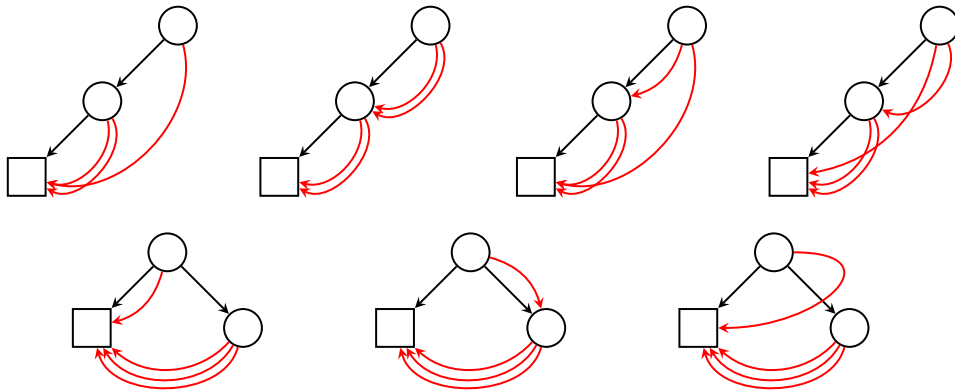
Figure 1 All 7 ternary relaxed ternary trees with 2 internal nodes.

2   Bijections and recurrences

In this paper, we consider a special class of DAGs, in which the outgoing edges (equivalently, the children) are equipped with an order.

▶ **Definition 2.** *An* ordered DAG *is a directed acyclic graph where there is a left to right ordering among the children.*

This brings us to the main object of this paper: the class of relaxed trees, a subclass of ordered DAGs, defined as follows; see Figure 1 and 2. The word relaxed signifies that they are a super-class of compacted trees, which are in bijection with trees after a compression procedure; see [7, 8].

▶ **Definition 3.** *A* relaxed k-ary tree *is an ordered DAG consisting of a unique source and a unique sink such that every node except the sink has out degree exactly k. Its* spine *is the spanning subtree created by the depth first search. The edges of the spine are called* internal edges, *while the other edges are called* external edges *or* pointers. *All nodes except the unique sink, are called* internal nodes.

▶ Remark 4. Note that in the binary case, the previous definition is equivalent to the one given in [7], in which relaxed trees are defined as trees enriched by pointers. Whereas here we characterize relaxed trees as a subclass of ordered DAGs.
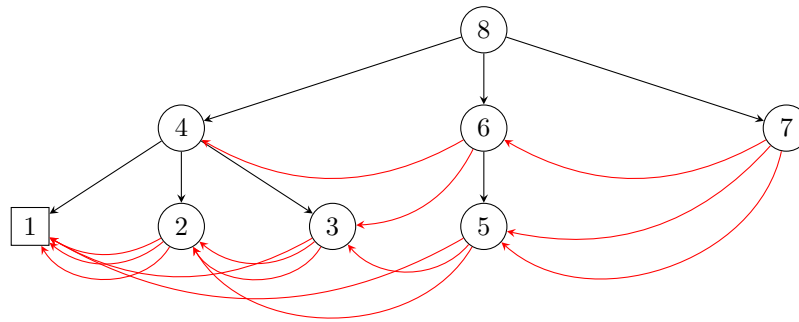
Figure 2 Example of relaxed ternary tree with 7 internal nodes (circles) labelled in postorder. The unique sink is depicted by a square. The black edges belong to the spine, the red ones are so-called pointers.

We start as in [7] by drawing a bijection between relaxed k-ary trees and Dyck paths with weights on their horizontal steps. This is done so that we can convert internal nodes into vertical steps and pointers into horizontal steps.

Every internal node u in the spine of a relaxed tree has a k-ary tree $T(u)$ associated with it where the nodes are traversed in postorder.

▶ **Definition 5.** *A* compacted *k-ary tree is a special case of a relaxed k-ary tree where for any arbitrary two nodes u, v in the spine, the k-ary trees associated with them $T(u), T(v)$ are not identical.*

In order to count relaxed k-ary trees, we will now describe a bijection to a class of paths, which are easier to enumerate. Let us first define the specific paths.

▶ **Definition 6.** *A horizontally k-decorated path P is defined as a lattice path consisting of up steps $U = (0, 1)$ and horizontal steps $H = (1, 0)$ from $(0, -1)$ with decorations such that:*

- *The first step is a U step, and its removal leaves a path never crossing the diagonal $y = \frac{x}{k-1}$.*
- *Below each H step, there is exactly one cross in one of the unit boxes below this H step and $y = -1$.*

The following lemma describes the bijection between relaxed k-ary trees and horizontally k-decorated paths. An example is shown in Figure 3.

▶ **Lemma 7.** *There is a bijection between horizontally k-decorated paths ending at $((k-1)n, n)$ and relaxed k-ary trees of size $n + 1$.*

Proof. Let R be a relaxed k-ary tree of size $n + 1$. We transform this tree uniquely into a k-decorated path ending at $((k - 1)n, n)$. We traverse the spine of R in post-order and label nodes from 1 to $n + 1$. During the traversal when we move up (i.e., traverse a spine edge the second time) we add a U step and when crossing a pointer for the last time (i.e., circling around its parent) we add a H step. Thereby, we associate each U step with the node we leave, and each H step with the parent of its pointer. Moreover, below each H step we draw a cross in the unit box that intersects the column of this H step with the row of the U step that is associated with the target of the pointer of this H step.

Observe now, that the first step is always a U step, as the unique sink is processed first in post-order and has no pointers. Thus, after removing this step, we start at the origin $(0, 0)$. Furthermore, note that R consists of n spine edges, and $(k - 1)n$ pointers. Hence, the
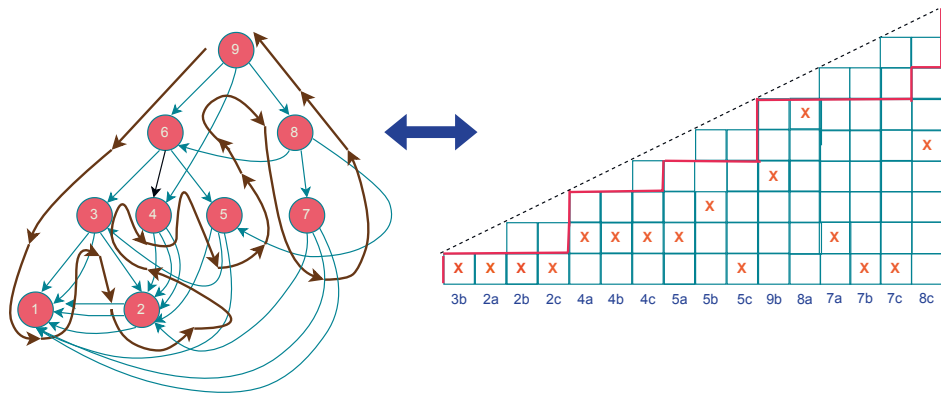
■ **Figure 3** Example of the bijection described in Lemma 7 between relaxed ternary trees and horizontally 3-decorated paths.

path consists of n U steps and $(k-1)n$ H steps, and after attaching a final U step, the path ends at $((k-1)n, n)$. It remains to show that the path never crosses $y = \frac{x}{k-1}$. Note that before a U step is added, the relaxed subgraph of which the associated node is the root has been traversed. The subgraph consisting only of the node 1 is treated by the fact that we start at $(0, -1)$. Now we proceed by induction on the depth of the subgraph. We need to distinguish, whether the node is on the left-branch from the root or not. First, we assume it is not. The minimal cases is a single pointer, which leads to a step H and increases the distance to the diagonal $y = \frac{x}{k-1}$ by one unit. Now assume that a subgraph with $i > 0$ nodes is given. The root has k children, which all have size less than i. Thus by induction, each part does not cross $y = \frac{x}{k-1}$ and moves one unit further to the right from the diagonal. Therefore, processing the k children moves the path k steps to the right from $\frac{x}{k-1}$, while (after that) processing the root, moves the path one unit up. Thus the path stays below $y = \frac{x}{k-1}$ and is one unit further to the right. Second, we assume that the node is on the diagonal. Then, the first child does not move the path one unit to the right, but the other $k-1$ do. By induction, the path still does not cross the diagonal $\frac{x}{k-1}$ but the distance now also stays the same.

Now in the reverse direction, let us take a horizontally k-decorated path ending at $((k-1)n, n)$. We want to recover a relaxed k-ary tree from this path. We start by noting that the number of up steps in the path is n and therefore the number of nodes is $n + 1$. Our first step is a horizontal step.

We start by adding a H step which corresponds to the left most leaf labelled 1. Then along with the path P, we create the spine of the the relaxed k-ary tree. Thus along the up steps we create nodes of the spine as the ith up step creates the $(i+1)$th node in post-order. Along the horizontal steps we say that we add a pointer from the existing node that we are in to a node $j + 1$ which is indicated by the cross placed on the level j below the path (counting from the bottom up). Completing this process we get a relaxed k-ary tree with its nodes labelled in post-order. ◀

This bijection allows directly to derive the following recurrence relations, following the step-by-step construction of the paths. In Table 1 we give the initial terms for $k = 2, \ldots, 5$.

▶ **Proposition 8.** *Let $r_{n,m}$ be the number of horizontally decorated paths ending at (n, m) and $\forall k \in \mathbb{Z}$, $k > 1$.*

$$
\begin{aligned}
r_{n,m} &= r_{n,m-1} + (m+1)r_{n-1,m}, & &\text{for } 1 \le m \le \frac{n}{k-1}, & (1) \\
r_{n,m} &= 0 & &\text{for } m > \frac{n}{k-1}, \\
r_{n,0} &= 1 & &\text{for } n \ge 0.
\end{aligned}
$$

Thus, by Lemma 7, the number of relaxed k-ary trees with n internal nodes is equal to $r_{(k-1)n,n}$.

🟨 **Table 1** Number of relaxed k-ary trees. Also the number of unlabeled acyclic single-source automata with n transient states on a k-letter input alphabet. The matrix consisting of these rows is given by A128249.

k	Relaxed k-ary trees (r_k)	OEIS
2	$(1, 3, 16, 127, 1363, 18628, 311250, 6173791, \dots)$	A082161
3	$(1, 1, 7, 139, 5711, 408354, 45605881, 7390305396, \dots)$	A082162
4	$(1, 1, 15, 1000, 189035, 79278446, 63263422646, 86493299281972, \dots)$	A102102
5	$(1, 1, 31, 6631, 5470431, 12703473581, 68149976969707, \dots)$	—

In a next step, we will also state the respective recurrences for compacted k-ary trees and minimal DFAs accepting a finite language on an alphabet of size k. These results generalize [7, Proposition 2.11] and [6, Proposition 5], respectively.

In a relaxed k-ary tree an internal node will be called a *cherry* if all its children are pointers.

▶ **Theorem 9.** *A relaxed k-ary tree C is a compacted k-ary tree if and only if no internal nodes u, v have the same children in the same order. Moreover, if C is not a compacted k-ary tree there exists a pair (u, v) with identical children and v is a cherry.*

Proof. Let us take C to be a compacted k-ary tree. Let two internal nodes u, v have the same children in the same order. This implies that the k-ary trees associated with u and v are isomorphic, which violates the definition of a compacted tree.

Conversely, let us assume that C is not a compacted tree. Therefore from the definition of a compacted tree it follows that we can find at least one pair of internal nodes (u, v) such that the k-ary trees associated with them are the same. At this point if v is a cherry then we are done. But if v is not a cherry then we take the first child of both nodes, say u_1 and v_1, respectively, and note that $T(u_1) = T(v_1)$. Now, if v_1 is not a cherry then we can continue this process. Therefore by infinite descent we see that there has to be two nodes u, v such that v is a cherry. ◀

In order to continue, we need the following concept. First, recall that in each horizontally k-decorated path, each step H is decorated by a marked box below. Further, each up step U corresponds to an internal node in the k-ary tree. Thus, we assign to each step U a k-tuple $(u_1, u_2, \dots, u_k)$ corresponding to the nodes of its children. (This will help us to construct the compacted trees from the relaxed ones.) Let us define a C_k-*decorated path* as a horizontally decorated path with the restriction that for consecutive steps $H^k U$ (k steps H followed by U) the k-tuple $(h_1, h_2, \dots h_k)$ corresponding to the labels below the H steps, we have $(h_1, h_2, \dots h_k) \neq (u_1, u_2, \dots u_k)$ for all preceding steps U.

▶ **Proposition 10.** *There is a bijection between the number of compacted k-ary trees of size n and C_k-decorated paths of length kn.*

Proof. Note that the C_k-decorated paths are a subset of the horizontally k-decorated paths and the previous bijection sends the relaxed k-ary trees to horizontally k-decorated paths. The C_k-decorated paths have been constructed in this way to reinforce the condition that the relaxed trees corresponding to these decorated paths are such that the k-ary trees associated with any two arbitrary nodes can never be identical. ◀

This allows us, again by a direct step-by-step construction, to derive the following bivariate recurrence for C_k-decorated paths; see Table 2.

▶ **Proposition 11.** *Let $c_{n,m}$ denote the number of C_k-decorated paths ending at (n, m). Then*

$$c_{n,m} = c_{n,m-1} + (m+1)c_{n-1,m} - (m-1)c_{n-k,m-1} \qquad \text{for } 1 \le m \le \frac{n}{k-1},$$

$$c_{n,m} = 0 \qquad \text{for } m > \frac{n}{k-1},$$

$$c_{n,0} = 1 \qquad \text{for } n \ge 0.$$

The number of compacted k-ary trees with n internal nodes is $c_{(k-1)n,n}$.

■ **Table 2** Number of compacted k-ary trees, which are defined as relaxed k-ary trees with the additional constraint that each fringe subtree is unique.

k	Compacted k-ary trees (c_k)	OEIS
2	$(1, 1, 3, 15, 111, 1119, 14487, 230943, 4395855, \ldots)$	A254789
3	$(1, 1, 7, 133, 5299, 371329, 40898599, 6561293893, \ldots)$	—
4	$(1, 1, 15, 975, 182175, 75961695, 60422966655, 82450320955455, \ldots)$	—
5	$(1, 1, 31, 6541, 5373571, 12458850121, 66790559866471, \ldots)$	—

We have a similar recursion for the minimal DFAs over a k-ary alphabet. We use the same scheme for the bijection with the H steps having a decoration. Since we have accepting and rejecting states, the U steps are colored either red or green. We call such paths 2-coloured. Again, we summarize the first values of Table 3.

▶ **Proposition 12.** *Let $b_{n,m}$ denote the number of 2-coloured horizontally decorated paths corresponding to DFAs ending in (n, m). Then*

$$b_{n,m} = 2b_{n,m-1} + (m+1)b_{n-1,m} - m\, b_{n-k,m-1} \qquad \text{for } 1 \le m \le \frac{n}{k-1},$$

$$b_{n,m} = 0 \qquad \text{for } m > \frac{n}{k-1},$$

$$b_{n,0} = 1 \qquad \text{for } n \ge 0.$$

The number of minimal DFAs accepting a accepting a finite language on an alphabet of size k with $n + 1$ states is $b_{(k-1)n,n}$.

Table 3 Number of minimal deterministic finite automata recognizing a finite k-letter language.

k	DFA (m_k)	OEIS
2	$(1, 1, 6, 60, 900, 18480, 487560, \ldots)$	A331120
3	$(1, 1, 14, 532, 42644, 6011320, 1330452032, \ldots)$	—
4	$(1, 1, 30, 3900, 1460700, 1220162880, \ldots)$	—
5	$(1, 1, 62, 26164, 43023908, 199596500056, \ldots)$	—

3 Asymptotics of relaxed k-ary trees

The goal of this section is to prove our main Theorem 1 on relaxed k-ary trees. As the proof is rather technical and complex, we will begin with an overview of the main steps.

First, in Section 3.1 we transform the recurrence (1) into a new recurrence $(d_{i,j})_{i,j \geq 0}$ by changing the used steps and extracting the super-exponential and certain exponential and polynomial contributions. Second, in Section 3.2 we perform a heuristic analysis of the asymptotics of $(d_{i,j})_{i,j \geq 0}$, and guess the shape for a rigorous proof. Using this shape, we build and prove in Section 3.3 explicit sequences $(\tilde{X}_{i,j})_{i,j \geq 0}$ and $(\hat{X}_{i,j})_{i,j \geq 0}$ that satisfy the same recurrence as $(d_{i,j})_{i,j \geq 0}$ but with the equality sign replaced by the inequality signs $\leq$ and $\geq$, respectively. This is the most technical part in which we heavily rely on computer algebra. Finally, in Section 3.4 we use these explicit sequences to prove inductively asymptotically tight upper and lower bounds, which yield our main theorem.

3.1 Transformation into a Dyck-like recurrence

We start from recurrence (1) and our goal is to determine the asymptotics of $r_{(k-1)n,n}$. First, we observe that the path closest to the diagonal given by $(H^{k-1}V)^n$ has a weight $(n!)^{k-1}$ as there are n H steps at height $0, 1, \ldots, n-1$. All other paths will get smaller weights. Therefore, after rescaling with this weight, all paths have a weight bounded by one. Note that it is technically easier to work with the rescaling $((k-1)n)!$, which has by Stirling's formula the same super-exponential growth. The difference in the exponential growth is a factor $(k-1)^{(k-1)n}$, which we will take into account later in (5). Thus, we set $\tilde{r}_{n,m} = \frac{r_{n,m}}{((k-1)n)!}$ and get the new recurrence

$$\tilde{r}_{n,m} = \frac{m+1}{n}\tilde{r}_{n-1,m} + \tilde{r}_{n,m-1}, \qquad\qquad m \leq \frac{n}{k-1}.$$

Next, as we are interested in the asymptotics of $\tilde{r}_{(k-1)n,n}$, we will transform this recurrence from North and East steps, to Dyck-like up and down steps. For this purpose, we define the new variables

$$\begin{cases} i = n + m, \\ j = n - (k-1)m, \end{cases} \quad \text{equivalently} \quad \begin{cases} n = \frac{(k-1)i+j}{k}, \\ m = \frac{i-j}{k}. \end{cases}$$

The idea behind this choice, is that i tracks the length of the walk, i.e., the number of steps, and j the distance to the diagonal. By this choice, we have $i, j \geq 0$. This gives the following generalized Dyck-like recurrence:

$$\tilde{d}_{i,j} = \tilde{U}(i,j)\tilde{d}_{i-1,j-1} + \tilde{d}_{i-1,j+k-1}, \qquad\qquad i > 0, j \geq 0, \qquad\qquad (2)$$

with the initial condition $\tilde{d}_{0,0} = 1$ and the weight $\tilde{U}(i,j) = \frac{i-j+k}{(k-1)i+j}$. Thus, we have the new steps $(1,1)$ and $(1,-k+1)$. This is a simple directed lattice path model with space-dependent

weights. Note that as the change in the x-direction is one unit per step, a path of length n consists of n steps, and it suffices therefore to track the current altitude. Therefore, from now on we consider only the changes in the y-direction, which we call jumps $+1$ and $-(k-1)$.

Let us now look at the drift of this new model. For now, we assume that i is large and that $j = o(i)$. The *drift* $\tilde{\delta}(i,j)$ at a point (i,j) is defined as the expected next jump size when leaving (i,j). Therefore, at (i,j) the jump $+1$ gets weight $\tilde{U}(i+1,j+1)$ and the jump $-(k-1)$ weight 1 for $j \geq k-1$ and 0 otherwise. Hence, we get that $\tilde{\delta}(i,j) = 1$ for $0 \leq j \leq k-2$ and

$$\tilde{\delta}(i,j) = \frac{i-j+k}{(k-1)(i+1)+j+1} - (k-1) = -\frac{k(k-2)}{k-1} + \mathcal{O}\left(\frac{1}{i}\right), \qquad \text{for } j > k-2.$$

Note that in the binary case ($k = 2$), the first term is zero and therefore the drift is converging to zero for large i. However, in the general case the drift is negative for large i. But we can define the following transformation to achieve the same behavior: $d_{i,j} = (k-1)^{2n}\tilde{d}_{i,j}$ where $n = \frac{(k-1)i+j}{k}$. This gives the final Dyck-like recurrence

$$d_{i,j} = U(i,j)d_{i-1,j-1} + d_{i-1,j+k-1}, \qquad\qquad i > 0, j \geq 0, \qquad\qquad (3)$$

with the initial condition $d_{0,0} = 1$ and the following weight for the up step

$$U(i,j) = \frac{(k-1)^2(i-j+k)}{(k-1)i+j} = (k-1)\left(1 - \frac{k(j-k+1)}{(k-1)i+j}\right). \qquad\qquad (4)$$

The drift $\delta(i,j)$ in this model is again $\delta(i,j) = 1$ for $0 \leq j \leq k-2$ but now we have

$$\delta(i,j) = -\frac{k(k-1)(j-k+2)}{k(i+1)-i+j} = -\frac{k(j-k+2)}{i} + \mathcal{O}\left(\frac{1}{i^2}\right), \qquad \text{for } j > k-2,$$

and therefore converges to zero. In this final model (3), we are interested in $d_{kn,0}$ since

$$r_{(k-1)n,n} = ((k-1)n)!\tilde{r}_{(k-1)n,n} = ((k-1)n)!\tilde{d}_{kn,0} = \frac{((k-1)n)!}{(k-1)^{2(k-1)n}}d_{kn,0}$$

$$\sim \sqrt{k-1}(2\pi)^{1-k/2}\frac{(n!)^{k-1}}{(k-1)^{(k-1)n}}n^{1-k/2}d_{kn,0}, \qquad\qquad (5)$$

where the last equality follows by Stirling's formula.

3.2 Heuristic analysis

In the next step, we will now heuristically analyze the recurrence (3). Inspired by the binary case and numerical experiments, we use the ansatz

$$d_{i,j} = h(i)f\left(\frac{j+1}{i^{1/3}}\right). \qquad\qquad (6)$$

The function $h(i)$ captures a macroscopic amplitude that is independent of j, while the function $f(x)$ captures the local behavior around the origin. The rescaling by $i^{1/3}$ is motivated by an analogy to pushed Dyck path and the binary case [7].

Substituting this ansatz into the recurrence (3) and reordering we get

$$\frac{h(i)}{h(i-1)} = \frac{U(i,j)f\left(\frac{j}{(i-1)^{1/3}}\right) + f\left(\frac{j+k}{(i-1)^{1/3}}\right)}{f\left(\frac{j+1}{i^{1/3}}\right)}.$$

As the left-hand side is independent of j, for large i both sides should have an expansion in i with whose coefficients are independent of j. Let us now zoom into the region $i^{1/3}$ by binding the variables i and j using the transformation $j = x\, i^{1/3} + 1$. This gives

$$\frac{h(i)}{h(i-1)} = \frac{U(i, x\, i^{1/3} + 1) f\left(\frac{x\, i^{1/3}+1}{(i-1)^{1/3}}\right) + f\left(\frac{x\, i^{1/3}+k+1}{(i-1)^{1/3}}\right)}{f(x)}. \tag{7}$$

Assuming now that $f(x)$ has a convergent Taylor series expansion around x, we get

$$\frac{h(i)}{h(i-1)} = k + \frac{k}{2}\frac{(k-1)f''(x) - 2xf(x)}{f(x)}i^{-2/3} + \mathcal{O}(i^{-1}).$$

Hence, in order to be consistent, the left-hand side needs to have an expansion in decreasing powers of $i^{1/3}$. In particular, it starts like

$$\frac{h(i)}{h(i-1)} = k + c\, i^{-2/3} + \mathcal{O}(i^{-1}),$$

for a constant $c \in \mathbb{R}$. Comparing coefficients, we see that $f(x)$ has to satisfy the differential equation

$$f''(x) = \frac{2(kx + c)}{k(k-1)} f(x).$$

This differential equation is, as in the binary case, solved by the Airy functions of the first and second kind. Due to the combinatorial nature of the problem, we require $f(x) \geq 0$ for $x \geq 0$ as well as $\lim_{x\to\infty} f(x) = 0$. Therefore, we get up to a multiplicative constant the following solution

$$f(x) = \mathrm{Ai}\left(\frac{2^{1/3}(kx + c)}{k(k-1)^{1/3}}\right).$$

Now, note that due the boundary conditions $d_{n,-1} = 0$ we must have $f(0) = 0$. Together, with $f(x) \geq 0$ for $x \geq 0$, this means that the argument of the Airy function must evaluate to the largest zero $a_1 \approx -2.338$ of $\mathrm{Ai}(x)$. Therefore

$$c = \frac{k(k-1)^{1/3}}{2^{1/3}}a_1$$

and

$$f(x) = \mathrm{Ai}\left(a_1 + B\,x\right), \qquad \text{where} \qquad B = \left(\frac{2}{k-1}\right)^{1/3}.$$

In order to capture the polynomial term, we need to use a more general ansatz than (6) that includes a second function $g(x)$ capturing the scale of order $i^{-1/3}$. In particular, we use

$$d_{i,j} = h(i)\left(f\left(\frac{j+1}{i^{1/3}}\right) + \frac{g\left(\frac{j+1}{i^{1/3}}\right)}{i^{1/3}}\right). \tag{8}$$

Note that the function $g(x)$ will influence the terms starting from order i^{-1} in (7). Analogous computations as performed above for $f(x)$ lead us to the ansatz

$$g(x) = -\left(\frac{(k+2)x^2}{6(k-1)} + \frac{2^{2/3}a_1(k-2)x}{6(k-1)^{2/3}}\right) f(x).$$

3.3   Explicit bounds

All these heuristic arguments above guide us to the following two results. These generalize [7, Lemmas 4.2 and 4.4], whose results are recovered by setting $k = 2$. The proofs are analogous to [4, 7]; for the details we refer to the accompanying Maple worksheet. In particular, this shows that the method developed in [7] is powerful enough to analyze bivariate recurrences that include parameters.

▶ **Lemma 13.** *Let $k \geq 2$ be an integer and $B = (\frac{2}{k-1})^{1/3}$. For all $i, j \geq 0$ let*

$$\tilde{X}_{i,j} := \left(1 - \frac{2^{2/3}a_1(k-2)}{6(k-1)^{2/3}} \frac{j}{i^{2/3}} - \frac{k+2}{6(k-1)} \frac{j^2}{i} + \frac{7k-11}{6(k-1)} \frac{j}{i} \right.$$
$$\left. + \frac{a_1^2(k-2)^2 B^4}{72} \frac{j^2}{i^{4/3}} + \frac{a_1(k^2-4)^2 B^5}{72} \frac{j^3}{i^{5/3}} \right) \mathrm{Ai}\left(a_1 + \frac{B(j+1)}{i^{1/3}}\right),$$
$$\tilde{s}_i := k\left(1 + \frac{a_1}{Bi^{2/3}} + \frac{7k-6}{6i} - \frac{1}{i^{7/6}}\right).$$

Then, for any $\varepsilon > 0$, there exists an $\tilde{i}_0$ such that

$$\tilde{X}_{i,j}\tilde{s}_i \leq U(i,j)\tilde{X}_{i-1,j+1} + \tilde{X}_{i-1,j-1}$$

for all $i \geq \tilde{i}_0$ and for all $0 \leq j < i^{2/3-\varepsilon}$, where $U(i,j)$ is defined in (4).

▶ **Lemma 14.** *Let $k \geq 2$ be an integer and $B = (\frac{2}{k-1})^{1/3}$. Choose $\eta > \frac{(k+2)^2}{72(k-1)^2}$ fixed and for all $i, j \geq 0$ let*

$$\hat{X}_{i,j} := \left(1 - \frac{2^{2/3}a_1(k-2)}{6(k-1)^{2/3}} \frac{j}{i^{2/3}} - \frac{k+2}{6(k-1)} \frac{j^2}{i} + \frac{7k-11}{6(k-1)} \frac{j}{i} \right.$$
$$\left. + \frac{a_1^2(k-2)^2 B^4}{72} \frac{j^2}{i^{4/3}} + \frac{a_1(k^2-4)^2 B^5}{72} \frac{j^3}{i^{5/3}} + \eta\frac{m^4}{n^2} \right) \mathrm{Ai}\left(a_1 + \frac{B(j+1)}{i^{1/3}}\right),$$
$$\hat{s}_i := k\left(1 + \frac{a_1}{Bi^{2/3}} + \frac{7k-6}{6i} + \frac{1}{i^{7/6}}\right).$$

Then, for any $\varepsilon > 0$, there exists an $\hat{i}_0$ such that

$$\hat{X}_{i,j}\hat{s}_i \geq U(i,j)\hat{X}_{i-1,j+1} + \hat{X}_{i-1,j-1}$$

for all $i \geq \hat{i}_0$ and for all $0 \leq j < i^{1-\varepsilon}$, where $U(i,j)$ is defined in (4).

Note that in the binary case $k = 2$, many terms in the previous Lemmas 13 and 14 are zero. These terms do not affect the final asymptotics, but they are needed for the proof using generalized Newton polygons. In particular, they allow to set certain points on the convex hull to zero, which then leads to the same behavior as in the binary case. For this reason, the technical proofs follow nearly verbatim the binary ones and we omit them in this extended abstract and refer to [7] and the accompanying Maple worksheet.

These explicit recurrences that satisfy the recurrence (3) with = replaced by ≤ and ≥, are the key ingredient to prove our main theorem.

3.4 Proof of Theorem 1 on relaxed k-ary trees

We start with the lower bound. First, we define a sequence $X_{i,j} := \max\{\tilde{X}_{i,j}, 0\}$ which satisfies the inequality of Lemma 13 for *all* $j \leq \frac{i}{k-1}$. Note that the factor of the Airy function Ai becomes negative for large j. Then, we define an explicit sequence $\tilde{h}_i := \tilde{s}_i \tilde{h}_{i-1}$ for $i > 0$ and $\tilde{h}_0 = \tilde{s}_0$. Using this we can prove by induction that $d_{i,j} \geq C_0 \tilde{h}_i X_{i,j}$ for some constant $C_0 > 0$ and all $i \geq \tilde{i}_0$ and all $0 \leq j \leq \frac{i}{k-1}$. Therefore,

$$d_{kn,0} \geq C_0 \tilde{h}_{kn} X_{kn,0}$$

$$\geq C_0 \prod_{i=1}^{kn} k \left(1 + \frac{a_1}{Bi^{2/3}} + \frac{7k-6}{6i} - \frac{1}{i^{7/6}}\right) \mathrm{Ai}\left(a_1 + \frac{B}{(kn)^{1/3}}\right)$$

$$\geq C_1 k^{kn} e^{3a_1(kn)^{1/3}/B} n^{\frac{7k-8}{6}}. \tag{9}$$

Finally, combining this with (5) we get the lower bound.

We continue with the upper bound, whose proof is similar, yet more technical. The starting point is Proposition 14 and, as in the lower bound, a function $X_{i,j}$ that is valid for *all* $0 \leq j \leq \frac{i}{k-1}$. For this purpose we define as in the binary case [7] a sequence $\tilde{d}_{i,j}$ depending on some large integer parameter $I > 0$ such that

$$\tilde{d}_{i,j} := \begin{cases} d_{i,j} & \text{for } 0 \leq j \leq i^{3/4} \text{ and } i > I, \\ 0 & \text{otherwise.} \end{cases} \tag{10}$$

The missing key step is now to show that $d_{kn,0} = \mathcal{O}(\tilde{d}_{kn,0})$. Combining this with the analogous computations performed for the lower bound above we get

$$\tilde{d}_{kn,0} \leq \hat{C}_1 k^{kn} e^{3a_1(kn)^{1/3}/B} n^{\frac{7k-8}{6}}.$$

To complete the prove we show $d_{kn,0} \leq 2\tilde{d}_{kn,0}$ using lattice path theory and computer algebra. We start from Equation (3) of $d_{i,j}$, which we interpret as a recurrence counting lattice paths. They are composed of steps $(1, 1)$ weighted by $U(i, j)$ and $(1, -k+1)$ weighted by 1 when the respective step ends at (i, j). The total weight of a path is the product of its weights. Now, we are interested in the paths never crossing $y = 0$ and ending at $(kn, 0)$. Let now $p_{\ell,k,kn}$ be the number of such paths starting at (r, s) and ending at $(kn, 0)$. From (3) we directly get

$$p_{r,s,kn} = \frac{(k-1)^2(r-s+k)}{(k-1)(r+1)+s+1} p_{r+1,s+1,kn} + p_{r+1,s-k+1,kn},$$

with $p_{\ell,-1,kn} = 0$ and $p_{kn,s,kn} = \delta_{s,0}$, where the factor is $U(r+1, s+1)$.

Now, we are able to show that

$$\frac{p_{r,s_1,kn}}{s_1 + 1} \geq \frac{p_{r,s_2,kn}}{s_2 + 1}, \tag{11}$$

for integers $0 \leq s_1 < s_2 \leq r \leq kn$ such that $k \mid s_2 - s_1$. The proof is completely elementary and uses reverse induction on r starting from $r = kn$. We refer to our accompanying Maple worksheet.

Finally, from (11) we directly get

$$p_{kx,ky,kn} \leq (ky + 1) p_{kx,0,kn}. \tag{12}$$

This allows to prove the following generalization of [7, Lemma 4.6] whose proof follows exactly the same lines and we omit it here. For the following statement, recall that $d_{i,j}$ is the weighted number of paths ending at (i, j). Let $\tilde{d}_{i,j}$ be the number of these paths such that no intermediate point (kx, ky) on the path satisfies $x > I_\varepsilon$ and $y > x^{3/4}$.

▶ **Lemma 15.** *For all $\varepsilon > 0$ there exists a constant $I_\varepsilon > 0$ that acts as the parameter I in the definition of $\tilde{d}_{i,j}$ in (10), such that $d_{kn,0} \leq (1+\varepsilon)\tilde{d}_{kn,0}$ for all $n > 0$.*

Proof. As in the binary case, we first rewrite the claimed inequality into $1 - \frac{\tilde{d}_{kn,0}}{d_{kn,0}} \leq \frac{\varepsilon}{1-\varepsilon}$. The left-hand side here represents the proportion of walks that pass through at least one point (kx, ky) such that $x > I_\varepsilon$ and $y > x^{3/4}$. Let $s_{x,y,n}$ be the proportion of such walks that pass through one fixed such point:

$$s_{x,y,n} = \frac{d_{kx,ky}p_{kx,ky,kn}}{d_{kn,0}}.$$

The idea is that the sum of these values over all violating points x, y is of course an upper bound for our claim. So we want to prove that this sum is very small for large x.

By (12) combined with the fact that $s_{x,y,n} \leq 1$, we get

$$s_{x,y,n} \leq \frac{(ky+1)p_{kx,0,kn}d_{kx,ky}}{d_{kn,0}s_{x,0,n}} = (ky+1)\frac{d_{kx,ky}}{d_{kx,0}}$$

$$\leq \frac{ky+1}{C_1 k^{kx}e^{3a_1(kx)^{1/3}/B}x^{\frac{7k-8}{6}}}(k-1)^{(k-1)x+y}\binom{kx}{x-y}$$

where we also used the lower bound (9) for $d_{kx,0}$ and the crude bound $d_{kx,ky} \leq (k-1)^{(k-1)x+y}\binom{kx}{x-y}$. The latter holds, as we may bound non-negative paths ending at (kx, ky) by unconstrained paths with weights $(k-1)$ for the up step $(1,1)$ and 1 for down step $(1, -k+1)$, since $U(i,j) \leq k-1$ for all $i, j \geq 0$.

Now this last expression is completely explicit, and for large x one can see that it is of order $\Theta(e^{-x^{1/2}})$. Now, the proof of the binary case follows verbatim and the claim follows. ◀

Finally, this proves $d_{kn,0} \leq 2\tilde{d}_{kn,0}$ and ends the proof of our main Theorem 1.

4 Conclusion and Outlook

The aim of this paper was to show that the method from [7] developed for the asymptotics of compacted binary trees can be applied to more general recurrences. Previously, in [4] we showed how to handle more general weights including an integer parameter, while in this paper we generalized the used steps. Instead of steps $(1,1)$ and $(1,-1)$ of Dyck type we studied a recurrence with larger steps given by $(1, k-1)$ and $(1,-1)$, for an arbitrary fixed positive integer k. In particular, we proved in Theorem 1 that in the class of relaxed k-ary trees the phenomenon of a stretched exponential appears for any integer $k \geq 2$.

In the long version, we will also prove the generalizations from the binary to the k-ary case for compacted trees and minimal deterministic finite automata accepting a finite language on an alphabet of size k for which we showed in Propositions 11 and 12 that their respective recurrences have similar shapes. We expect that the phenomenon of a stretched exponential persists also in these cases. In general, the presented method can be applied to more general cases including several parameters in the weights and also recurrences with more than two steps.

Another research direction is to generalize the previously studied relaxed binary trees of bounded right height [9] to the k-ary case. In the binary case, this class has been proved to be D-finite using exponential generating function methods. In the k-ary case, a similar rescaling, as we used in Section 3.1 by $n!^{k-1}$ should allow to analyze this class using generalized exponential functions of the type $\sum_{n \geq 0} a_n \frac{z^n}{(n!)^{k-1}}$. This class is also interesting, since the binary case has rich combinatorial properties, such as bijection and closed-form enumeration formulas [17], which are also worth to investigate in the k-ary case.

───── **References** ─────

1 Cyril Banderier and Michael Wallner. Young Tableaux with Periodic Walls: Counting with the Density Method. *Sém. Lothar. Combin.*, 85B, Article 85B.47:12 pp., 2021. URL: `https://www.mat.univie.ac.at/~slc/wpapers/FPSAC2021/47.html`.

2 Béla Bollobás. *Random graphs*, volume 73 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, second edition, 2001. `doi:10.1017/CBO9780511814068`.

3 Mireille Bousquet-Mélou, Markus Lohrey, Sebastian Maneth, and Eric Noeth. XML compression via directed acyclic graphs. *Theory Comput. Syst.*, 57(4):1322–1371, 2015. `doi:10.1007/s00224-014-9544-x`.

4 Yu-Sheng Chang, Michael Fuchs, Hexuan Liu, Michael Wallner, and Guan-Ru Yu. Enumerative and distributional results for d-combining tree-child networks. *Adv. in Appl. Math.*, 157:Paper No. 102704, 2024. `doi:10.1016/j.aam.2024.102704`.

5 Peter J. Downey, Ravi Sethi, and Robert Endre Tarjan. Variations on the common subexpression problem. *J. Assoc. Comput. Mach.*, 27(4):758–771, 1980. `doi:10.1145/322217.322228`.

6 Andrew Elvey Price, Wenjie Fang, and Michael Wallner. Asymptotics of Minimal Deterministic Finite Automata Recognizing a Finite Binary Language. In *AofA 2020*, volume 159 of *LIPIcs. Leibniz Int. Proc. Inform.*, pages 11:1–11:13, 2020. `doi:10.4230/LIPIcs.AofA.2020.11`.

7 Andrew Elvey Price, Wenjie Fang, and Michael Wallner. Compacted binary trees admit a stretched exponential. *J. Combin. Theory Ser. A*, 177:105306, 40, 2021. `doi:10.1016/j.jcta.2020.105306`.

8 Philippe Flajolet, Paolo Sipala, and Jean-Marc Steyaert. Analytic variations on the common subexpression problem. In *Automata, languages and programming*, pages 220–234. Springer, 1990. `doi:10.1007/BFb0032034`.

9 Antoine Genitrini, Bernhard Gittenberger, Manuel Kauers, and Michael Wallner. Asymptotic enumeration of compacted binary trees of bounded right height. *J. Combin. Theory Ser. A*, 172:105177, 2020. `doi:10.1016/j.jcta.2019.105177`.

10 Manosij Ghosh Dastidar and Michael Wallner. Maple worksheet - asymptotics of relaxed k-ary trees. Software, FWF P 34142, swhId: `swh:1:dir:f9b3ed0c7c098f24ee8860808959b84e402dc276` (visited on 2024-07-09). URL: `https://gitlab.tuwien.ac.at/michael.wallner/relaxed-k-ary-trees`.

11 A. J. Guttmann and S. G. Whittington. Two-dimensional lattice embeddings of connected graphs of cyclomatic index two. *J. Phys. A*, 11(4):721–729, 1978. URL: `http://stacks.iop.org/0305-4470/11/721`.

12 Anthony J. Guttmann. Analysis of series expansions for non-algebraic singularities. *Journal of Physics A: Mathematical and Theoretical*, 48(4):045209, 2015. `doi:10.1088/1751-8113/48/4/045209`.

13 Svante Janson, Tomasz Łuczak, and Andrzej Rucinski. *Random graphs*. Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley-Interscience, New York, 2000. `doi:10.1002/9781118032718`.

14 Valery A. Liskovets. Exact enumeration of acyclic deterministic automata. *Discrete Applied Mathematics*, 154(3):537–551, 2006. `doi:10.1016/j.dam.2005.06.009`.

15 Colin McDiarmid, Charles Semple, and Dominic Welsh. Counting phylogenetic networks. *Ann. Comb.*, 19(1):205–224, 2015. `doi:10.1007/s00026-015-0260-2`.

16 Remco van der Hofstad. *Random graphs and complex networks. Vol. 1.* Cambridge Series in Statistical and Probabilistic Mathematics, [43]. Cambridge University Press, Cambridge, 2017. `doi:10.1017/9781316779422`.

17 Michael Wallner. A bijection of plane increasing trees with relaxed binary trees of right height at most one. *Theoretical Computer Science*, 755:1–12, 2019. `doi:10.1016/j.tcs.2018.06.053`.

18 Michael Wallner. Dyck paths and inversion tables. *Permutation Patterns 2023*, pages 142–144, 2023. URL: `https://2023.permutationpatterns.com/booklet/bookletco.pdf#page=142`.

Matching Algorithms in the Sparse Stochastic Block Model

Anna Brandenberger (ORCID)
Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA, USA

Byron Chin (ORCID)
Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA, USA

Nathan S. Sheffield ✉ ⌂ (ORCID)
Massachusetts Institute of Technology, Cambridge, MA, USA

Divya Shyamal (ORCID)
Massachusetts Institute of Technology, Cambridge, MA, USA

───── **Abstract** ─────

In sparse Erdős–Rényi graphs, it is known that a linear-time algorithm of Karp and Sipser achieves near-optimal matching sizes asymptotically almost surely, giving a law-of-large numbers for the matching numbers of such graphs in terms of solutions to an ODE [9]. We provide an extension of this analysis, identifying broad ranges of stochastic block model parameters for which the Karp–Sipser algorithm achieves near-optimal matching sizes, but demonstrating that it cannot perform optimally on general stochastic block model instances. We also consider the problem of constructing a matching *online*, in which the vertices of one half of a bipartite stochastic block model arrive one-at-a-time, and must be matched as they arrive. We show that, when the expected degrees in all communities are equal, the competitive ratio lower bound of 0.837 found by Mastin and Jaillet for the Erdős–Rényi case [14] is achieved by a simple greedy algorithm, and this competitive ratio is optimal. We then propose and analyze a linear-time online matching algorithm with better performance in general stochastic block models.

1 Introduction

Real-life allocation problems, particularly those related to the display of ads in search engine results, have motivated a substantial line of research investigating the problem of online matching in random graphs. Most prior work on this problem assumes the vertices on one side of a bipartite graph are drawn i.i.d. from an adversarially-chosen distribution. In that setting, upper bounds on the competitive ratio are known [13]. However, in Erdős–Rényi graphs, it is possible to exceed these bounds [14]. One might therefore hope that graphs arising in nature tend to permit better online matching algorithms than adversarial distributions. In this work, we consider matching problems in the ***stochastic block model***, in which vertices belong to one of a constant number of classes, and the probability of an edge between two vertices depends only on their classes. This is a broad class of structured distributions on graphs, which includes the Erdős–Rényi model as a special case. In stochastic block models,

neither the optimal online matching algorithm nor the (offline) matching number are known. We make progress on both of these questions, finding expressions for the matching number in a number of regimes, and proposing and analyzing several heuristics for online matching.

1.1 Preliminaries

The idea of a population divided into a fixed number of distinct but internally-homogeneous groups is captured by the **stochastic block model**, first proposed by Holland et. al. to model social networks [6].

▶ **Definition 1** (Stochastic block model). *For fixed q, consider q disjoint sets of vertices (classes) $S_1, \ldots, S_q$, and a symmetric probability matrix associating a value $p_{ij} \in [0, 1]$ to each pair i, j of classes. Given these parameters, the associated **stochastic block model** is the distribution over graphs obtained by including each edge (u, v) independently with probability $p_{\sigma_u \sigma_v}$, where σ_u and σ_v are the classes to which u and v, respectively, belong.*

If $q = 1$, we recover the Erdős–Rényi graph $G(n, p)$ on n vertices with edge probability p. If $q = 2$, where each class has n vertices and $p_{12} = p$, $p_{11} = p_{22} = 0$, we recover the bipartite Erdős–Rényi graph $G(n, n, p)$.

We are interested in the number of vertices in a maximum-cardinality matching on a graph drawn from this distribution, which we call the **matching number**. We are also interested in the problem of constructing a matching online.

▶ **Definition 2** (Online bipartite matching problem). *Given a bipartite stochastic block model, we can define an associated online matching problem. An algorithm is given the classes of n right vertices, and knows a distribution over the left classes. For each of n time-steps, a new left-node is revealed (a left label is drawn from the distribution, and coins are flipped according to the block model probabilities to determine which edges it has to the right vertices) – the algorithm then decides which, if any, of these edges to add to M. Once it has made this decision, it is never allowed to revisit that vertex.*

The most interesting range of stochastic block model parameters turns out to be the **sparse regime**, when all probabilities p_{ij} are $\Theta(1/n)$. When p_{ij} grows faster than $1/n$, as n grows to infinity the graph becomes dense enough such that with high probability, there is a near perfect matching between S_i and S_j. On the other hand, when p_{ij} grows slower than $1/n$, the graph becomes so sparse that we can include almost every edge in M. Therefore, we consider the regime in which $p_{ij} = c_{ij}/n$ for some constants c_{ij}.

1.2 Background

In 1981, Karp and Sipser demonstrated that a simple linear-time heuristic achieves matchings within $o(n)$ of the true matching number of Erdős–Rényi graphs with high probability [9]. By associating the performance of that algorithm with a Markov chain and examining the limiting differential equation, they were able to prove a law of large numbers on the matching number of such graphs. In 1998, Frieze, Pittel, and Aronson improved the error estimate of the Karp–Sipser algorithm in Erdős–Rényi graphs from $o(n)$ to $n^{1/5+o(1)}$ [1]. In 2011, Bohman and Frieze extended analysis of the Karp–Sipser algorithm to the model of graphs drawn uniformly over a fixed degree sequence, showing that a log concavity condition is sufficient for the algorithm to find near-perfect matchings in such graphs [2]. Because of its simplicity, the Karp–Sipser algorithm has also received attention as a practical method for data reduction; some recent work investigates efficient implementations [11, 12].

The online bipartite matching problem was first introduced by Karp, Vazirani and Vazirani in 1990; they achieved a tight $1 - 1/e$ competitive ratio on worst-case inputs [10]. In 2009, Feldman et. al. showed when the left vertices are instead drawn from an arbitrary known distribution, with integral expected arrival rates, it is possible to get a competitive ratio strictly better than $1-1/e$ [5]. For arbitrary distributions, the best known algorithm achieves competitive ratio 0.716, and there is an upper bound of 0.823 [8, 13]. There has also been work considering algorithms for specific left vertex distributions. Mastin and Jaillet found that in $G(n, n, p)$, the random bipartite graph where all edges are independent and equally likely to exist, all greedy algorithms achieve competitive ratio at least 0.837 – they conjectured that this lower bound is optimal, but were unable to compute the matching number [14]. Sentenac et. al. studied the problem in the 1-dimensional geometric model, where they found expressions for both the matching number and the performance of a particular online heuristic [16]. To the best of our knowledge, the only previous work that considers the stochastic block model is by Soprano-Loto et. al., who consider the regime where the graph is dense (i.e. all probabilities are constants not depending on n), and characterize when it is possible to achieve an asymptotically near-perfect matching [17].

1.3 Main Contributions

We show that the Karp–Sipser algorithm achieves near-optimal matchings for probability matrices p_{ij} satisfying any of the following conditions. However, the algorithm does not achieve near-optimal matchings in general stochastic block models.

- **Equitable:** We call a stochastic block model "equitable" if each vertex has the same expected degree, i.e., for every i, $\sum_j p_{ij}|S_j| = \sum_j c_{ij}|S_j|/n = c$ for some constant c. We show that the asymptotic matching number for any such graph is $\alpha n + o(n)$ where α is an explicit constant. See Theorem 5 for the full statement.
- **Sub-Critical:** When $\sum_j c_{ij}|S_j|/n < e$ for all i, we show that the model is in a sub-critical regime similar to the one found for Erdős–Rényi graphs. We show that the asymptotic matching number converges to the solution of an explicit ODE, see Theorem 9.
- **Bipartite Erdős–Rényi:** We also determine, in terms of the solution of an explicit ODE, the asymptotic matching number of $G(kn, n, c/n)$, the bipartite graph with part sizes kn and n and independent edge probability c/n; see Lemma 13. This case is of particular interest as a simple example for which usual arguments about the Karp–Sipser algorithm fail.

With regards to the online matching problem, we will first show that a simple strategy, GREEDY, of matching to uniform random neighbours achieves the optimal 0.837 competitive ratio in any equitable model. We then propose the more sophisticated heuristic, SHORT-SIGHTED, of preferring the available class that maximizes the probability of finding a match on the next step. While SHORTSIGHTED performs extremely close to the online optimal in experiments, we are able to show analytically that it is not an asymptotically optimal algorithm in general.

2 Analysis of the Karp–Sipser Algorithm for Offline Matching

2.1 Karp–Sipser Algorithm and Outline of Analysis

Karp and Sipser proposed the following greedy algorithm: whenever there exists a vertex of degree 1 in the graph, choose one uniformly at random, add its edge to the matching, and remove the vertex and its neighbor from the graph. When there are no degree 1 vertices, instead choose a uniform random edge to add to the matching, removing both of its incident vertices from the graph.

When there exist degree 1 vertices it is always "safe" to add their edges in the sense that there exists an optimal matching which includes these edges. Therefore, any "mistakes" the algorithm makes can only happen after the first time the graph no longer contains any degree 1 vertices (we call the steps before this **Phase 1**, and the steps after this **Phase 2**). The analysis of this algorithm in the Erdős–Rényi graph setting proceeds as follows [9, 1]:

1. Conditioned on the state of a Markov process on small tuples of integers, the graph maintains a simple distribution law even after several steps of the algorithm.

2. Estimating transition probabilities from the degree distribution, we can appeal to known approximation theorems to conclude that the Markov process stays close to the solution of a corresponding ODE as $n \to \infty$ with high probability.

3. In Phase 2 of the algorithm, it is very likely that the algorithm finds a near-perfect matching on the remaining graph. Since the algorithm makes only optimal decisions in Phase 1, this means that overall it finds a matching within $o(n)$ vertices of the true optimal value, and so the matching number of the graph is described by solutions to the ODE from step 2.

We apply steps analogous to 1 and 2 in the general stochastic block model. On the other hand, we find that step 3 does not always hold – that is, there exist stochastic block models in which Phase 2 does not find a near-perfect matching. We list some interesting cases in which we can show Phase 2 does find a near-perfect matching, and give an example where a precise analysis is possible even when it does not.

2.2   Convergence of Phase 1 Transitions to Continuous Approximation

In Appendix A, we show that to an observer tracking only the following values for all class labels i, j: the number E_{ij} of j-type half-edges (i.e. edges whose other endpoint lies in class j) incident to label-i vertices, the number T_i of label-i vertices of degree exactly 1 (the "thin" vertices), and the number F_i of label-i vertices of degree at least 2 (the "fat" vertices), the system follows a Markov chain, whose transition probabilities we determine from the limiting distribution of degrees in the block model. Letting $\bar{E}_{ij} = \frac{E_{ij}}{n}$, $\bar{F}_i = \frac{F_i}{n}$, and $\bar{T}_i = \frac{T_i}{n}$, where n is the total number of vertices in the original graph, this analysis gives

$$\mathbb{E}[\Delta \bar{E}_{ij}] = \frac{-\omega_{ij} - \omega_{ji} - \sum_l \omega_{li}\delta_{ij} - \sum_l \omega_{lj}\delta_{ji}}{n}$$

$$\mathbb{E}[\Delta \bar{F}_i] = -\frac{\left(\sum_l \omega_{li}\right)\left(\frac{\bar{h}_i - \bar{T}_i}{\bar{h}_i}\right) - \sum_j \sum_l \omega_{jl}\delta_{li}\theta_i}{n}$$

$$\mathbb{E}[\Delta \bar{T}_i] = -\frac{\left(\sum_l \omega_{il}\right) - \left(\sum_l \omega_{li}\right)\left(\frac{\bar{T}_i}{\bar{h}_i}\right) - \sum_j \sum_l \omega_{jl}\delta_{li}\left(\frac{\bar{T}_i}{\bar{h}_i} - \theta_i\right)}{n},$$

where $h_i = \sum_l E_{il}$, $\omega_{ij} = \frac{T_i}{\sum_l T_l} \cdot \frac{E_{ij}}{h_i}$, $\delta_{ij} = \frac{h_i - T_i}{h_i} \cdot \frac{\lambda_i}{1 - e^{-\lambda_i}} \cdot \frac{E_{ij}}{h_i}$, and $\theta_i = \frac{h_i - T_i}{h_i} \cdot \frac{\lambda_i}{e^{\lambda_i} - 1}$, for λ_i a solution to $\frac{\lambda_i(e^{\lambda_i} - 1)}{e^{\lambda_i} - 1 - \lambda_i} = \frac{h_i - T_i}{F_i}$. These values come from our degree distribution estimates in Appendix A, where h_i is the total number of half-edges attached to class i, ω_{ij} is the probability that a random degree-1 vertex is in class i and has neighbour in class j, δ_{ij} is the expected number of (additional) j-type half edges shared by the endpoint of a random class-i half-edge, and θ_i is the probability that a random class-i half-edge is incident to a degree-2 vertex.

This is a process that takes order n time steps, and where the expected change at each time step scales like $\frac{1}{n}$. Informally, we can observe that in the limit of n, the many small steps should average out and produce a process evolving according to their expectations; this suggests looking at the following system of equations:

$$\frac{d}{dt}\bar{E}_{ij}(t) = -\omega_{ij}(t) - \omega_{ji}(t) - \sum_l \omega_{li}(t)\delta_{ij}(t) - \sum_l \omega_{lj}(t)\delta_{ji}(t)$$

$$\frac{d}{dt}\bar{F}_i(t) = -\left(\sum_l \omega_{li}(t)\right)\left(\frac{\bar{h}_i(t) - \bar{T}_i(t)}{\bar{h}_i(t)}\right) - \sum_j \sum_l \omega_{jl}(t)\delta_{li}(t)\theta_i(t)$$

$$\frac{d}{dt}\bar{T}_i(t) = -\left(\sum_l \omega_{il}(t)\right) - \left(\sum_l \omega_{li}(t)\right)\left(\frac{\bar{T}_i(t)}{\bar{h}_i(t)}\right) - \sum_j \sum_l \omega_{jl}(t)\delta_{li}(t)\left(\frac{\bar{T}_i(t)}{\bar{h}_i(t)} - \theta_i(t)\right)$$

with initial conditions $\bar{E}_{ij}(0) = c_{ij}\bar{S}_i\bar{S}_j$, $\bar{F}_i(0) = \bar{S}_i\left(1 - \left(1 + \sum_j c_{ij}\bar{S}_j\right)e^{-\sum_j c_{ij}\bar{S}_j}\right)$, and $\bar{T}_i(0) = \bar{S}_i\left(\sum_j c_{ij}\bar{S}_j\right)e^{-\sum_j c_{ij}\bar{S}_j}$, where c_{ij} and $\bar{S}_i = |S_i|/n$ are the connection probabilities and label class sizes, respectively, of the stochastic block model instance. We justify this passage to differential equations by Wormald's theorem [19], which guarantees that, in the limit of n, the evolution of Phase 1 stays close to the unique solution of this ODE with probability approaching 1 (this is formally justified in Appendix B). This implies the following:

▶ **Lemma 3.** *If $\mathcal{Y}(t) = \{\bar{\mathcal{E}}_{ij}, \bar{\mathcal{F}}_i, \bar{\mathcal{T}}_i\}$ is a solution to the above ODE, with high probability the total number of unmatched isolated vertices created in Phase 1 of the Karp–Sipser algorithm is $n\left(1 - 2\tau - \sum_i \bar{\mathcal{F}}_i(\tau)\right) + o(n)$, where τ is the first time such that $\bar{\mathcal{T}}_i(\tau) = 0$ for all i.*

In the Erdős–Rényi case, studying Phase 2 of the Karp–Sipser algorithm reveals that with high probability at most $o(n)$ unmatched isolated vertices are created, meaning that the algorithm is asymptotically optimal, and that the matching number is $n\left(2\tau + \bar{\mathcal{F}}_i(\tau)\right) + o(n)$ with high probability. However, this analysis turns out not to work for general stochastic block model instances. We first illustrate a few examples (namely the equitable and sub-critical cases) where, with a little bit of work, we can prove similar results; then, we examine where the algorithm fails.

2.3 Equitable Case

▶ **Definition 4.** *We call stochastic block model parameters **equitable** if there is some constant c such that for all classes i, $\sum_j \frac{c_{ij}|S_j|}{n} = c$.*

In other words, although the edge density in some parts of the graph may be higher than other parts, the expected degree of every vertex is c regardless of what label class it belongs to. In these cases, we show that not only does the Karp–Sipser algorithm construct an asymptotically-optimal matching, but that the matching size it constructs is asymptotically the same as the matching number of the Erdős–Rényi graph $G(n, c/n)$. The intuition behind this claim is that, despite the nontrivial correlation between the edges of the graph, we expect the degree distributions to look the same everywhere. The crucial point we need to justify to make this intuition precise is that the degree distributions necessarily *remain* close to equal across classes, given that they start that way.

▶ **Theorem 5.** *With high probability, the matching number of an equitable stochastic block model is $\left(1 - \frac{x + ce^{-x} + xce^{-x}}{2c}\right)n + o(n)$, where x is the smallest solution to $x = ce^{-ce^{-x}}$, and the Karp–Sipser algorithm achieves within $o(n)$ of this value.*

The theorem is proven with the help of the following three lemmas.

▶ **Lemma 6.** *Given an equitable stochastic block model, after Phase 1 of the Karp–Sipser algorithm, with high probability for every i, F_i (number of vertices of degree at least 2; note that this is the same as total number of vertices because at the end of Phase 1 there are no vertices of degree 1) and h_i (total number of incident edges) differ by at most $o(n)$ from the corresponding values in $G(n, \frac{c}{n})$.*

Proof. This follows directly from our application of Wormald's theorem. Observe that $\left(\frac{h_i(t)}{S_i}, \frac{F_i(t)}{S_i}, \frac{T_i(t)}{S_i} \right)$ satisfy a symmetric system of equations, and have a symmetric initial condition, so the symmetry will be preserved by a solution. Since Wormald's theorem guarantees that with high probability the result of Phase 1 differs by at most $o(n)$ from a solution to this ODE, we have the desired statement. ◀

▶ **Lemma 7.** *Fix a constant $\epsilon > 0$, and suppose, at the start of Phase 2 of the Karp–Sipser algorithm, the values of $F_i/\bar{S}_i$ and $h_i/\bar{S}_i$ each differ by at most $o(n)$ between classes. If the Karp–Sipser algorithm is run until some class has average degree less than $2 + \epsilon$, with high probability at most $o(n)$ isolated vertices will have been produced since the start of Phase 2, and all classes will have average degree $2 + \epsilon + o(1)$*

Proof. In the regime where the average degree $h_i/F_i > 2 + \epsilon$ for all classes i, we want to argue that the number of steps between the times when the graph is free of thin vertices is small, so that by bunching those runs of degree-1 stripping together we can again control the evolution by an ODE. The following is a high-level treatment of the argument; formal justification from Wormald's theorem is given in Appendix B:

When we first remove a random edge, this may create some vertices of degree 1. In removing those, we may create more vertices of degree 1. In general, the expected number of new degree-1 vertices created when a degree-1 vertex of class i is removed is $\sum_j \frac{E_{ij}}{h_i} \delta_i \theta_j \leq (\max_i \delta_i)(\max_j \theta_j)$, where δ_i is the expected number of half-edges sharing a vertex with a random class-i half-edge, and θ_i is the probability that a random class-i half-edge is attached to a degree-2 vertex. Now, when we have that the average degree in each class is at least $2 + \epsilon$, and that the difference in average degree between any pair of classes i and j is very small (say, less than γ), then we know by our degree estimates (see Appendix A) that $\delta_i \theta_j = \frac{\lambda_i}{1 - e^{-\lambda_i}} \cdot \frac{\lambda_j}{e^{\lambda_j} - 1} < 1 - \eta$ for any pair i, j of classes, where η depends on γ and ϵ (the existence of some $\gamma > 0$, $\eta > 0$ in terms of ϵ with this property is guaranteed by a continuity argument, see Appendix B). So, the expected number of degree-1 vertices created for each degree-1 vertex removed is at most a constant, $1 - \eta$, that is bounded away from 1. The size of a subcritical Galton–Watson tree with $1 - \eta$ expected offspring is very unlikely to exceed $\mathcal{O}(1/\eta)$ – by a Chernoff bound, we can argue that it is exponentially unlikely in n to achieve size $n^{\Omega(1)}$. Thus, while we are within the region where average degrees are γ-close and greater than $2 + \epsilon$, the duration of a "run" of degree-1 stripping is w.h.p. much smaller than n, and we can appeal to the law of large numbers to claim the process evolves like its expectation. Whenever the average degree is the same in all classes, the expected change in the number of edges of a given type is proportional to the number of edges currently of that type (all edges in the graph are equally likely to be chosen as the first edge removed, equally likely to be the edge chosen one step into the run, etc). So, since average degrees start out $o(1)$ away from each other, we expect them to remain that way up until one of them drops below $2 + \epsilon$.

Now, note that the rate of creation of isolated vertices on a given step is always proportional to the fraction of thin vertices in the graph: when we remove a vertex (the neighbour of a thin vertex), it has δ_i other neighbours in expectation, and by the Markov property shown

in Appendix A, we know that those neighbour edges are equally likely to be any of the edges leaving the class. So, the number of vertices isolated at each step is in expectation at most $\frac{\delta_i}{v}$ times the total number of degree-1 vertices, where v is the total number of vertices remaining (i.e. not matched or isolated) in the graph. For a run of length $n^{o(1)}$, there can only ever be $n^{o(1)}$ thin vertices present at any given moment, so we expect to create $o(1)$ isolated vertices in any given run. Thus, the total number created while we're in this regime is $o(n)$ with high probability. ◄

▶ **Lemma 8.** *If the Karp–Sipser algorithm is run to completion, starting from average degree $2 + \epsilon + o(1)$ in every class, with high probability at most $2\epsilon n \left(1 - 2\log(2\epsilon)\right)$ isolated vertices are produced.*

Proof. We're no longer guaranteed that individual runs are short, so can't argue that each run isolates only $o(1)$ vertices in expectation. Instead, we will bound the number of isolated vertices by controlling the number of degree-1 vertices that are ever produced in this regime. The only way to remove a thin vertex and create more than one in its place is to have its neighbour have degree greater than 2. Since no class has average degree more than $2 + 2\epsilon$, we know that the entire graph has at most $2n\epsilon$ edges associated with vertices of degree greater than 2. Those edges are the only places we can branch out and create more degree-1 things than we consume, so, throughout the course of the rest of the algorithm, there can never be more than $2n\epsilon$ thin vertices in the graph at once.

As in Lemma 7, we observe that the expected number of vertices isolated in a single step is at most $\frac{\delta_i}{v} \leq \frac{2}{v}$ times the total number of degree-1 vertices in the graph, where v is the total number of remaining remaining. Using our bound on the number of degree-1 vertices, and a law-of-large numbers argument, we can upper bound the number of isolated vertices created in all remaining steps by $2\epsilon n + \int_{2\epsilon n}^{n} \frac{2}{v} 2\epsilon n \, dv = 2\epsilon n \left(1 - 2\log(2\epsilon)\right)$. ◄

Proof of Theorem 5. In Phase 1, the Karp–Sipser algorithm is guaranteed to perform optimally, and Lemmas 6, 7, and 8 gurantee that, for any constant choice of ϵ, with high probability Phase 2 isolates only $o(n) + 2\epsilon n \left(1 - 2\log(2\epsilon)\right)$ vertices. We can make this arbitrarily small by choice of ϵ; so, with high probability the total number of vertices isolated in Phase 2 is $o(n)$, meaning that overall the algorithm is within $o(n)$ of optimal. Since the ODE determining the evolution of Phase 1 evolves the same as in an Erdős–Rényi graph with parameter c/n, the total number of lost vertices must be within $o(n)$ of the number that would be lost in $G(n, c/n)$. The matching number of $G(n, c/n)$ is known to be $\left(1 - \frac{x + ce^{-x} + xce^{-x}}{2c}\right) n + o(n)$, where x is the smallest solution to $x = ce^{-ce^{-x}}$ [9, 1], so that must also be the matching number in this case. ◄

2.4 Sub-Critical case

In the Erdős–Rényi case, Karp and Sipser proved that the number of unmatched non-isolated vertices remaining in the graph after Phase 1 (which we follow recent literature in calling the "Karp–Sipser core" [3]) is $o(n)$ with high probability if $c < e$, and $\Theta(n)$ with high probability if $c > e$ [9]. Our above analysis implies that any equitable stochastic block model also follows this critical transition at $c = e$. In this section, we examine criticality in the non-equitable case. We prove the following theorem:

▶ **Theorem 9.** *If $\sum_j c_{ij} \bar{S}_j < e$ for all i, then the Karp–Sipser core has size $o(n)$ with high probability.*

Interestingly, however, we find that the converse is not true – in fact, it's possible for the model to be subcritical even when $\sum_j c_{ij}\bar{S}_j > e$ for all i. We note the following facts:

▶ **Fact 10** (Karp and Sipser [9])**.** *The set of vertices removed by Phase 1 is fixed, regardless of the order in which degree-1 vertices are stripped. So, if there exists some valid sequence of degree-1 strippings that removes a given vertex v from the graph (either matching or isolating it), that vertex is not in the Karp–Sipser core.*

▶ **Fact 11** (Implied by a result of Mossel, Neeman and Sly [15] for 2 classes; Sly and Chin [4] for greater than 2 classes)**.** *For any constant d, in the limit of n the d-neighbourhood of any vertex of G (i.e., the subgraph obtained by a BFS of depth d from the vertex) converges in distribution to the first d levels of a multitype branching process, where nodes of type i have independently $Pois(c_{ij}\bar{S}_j)$ children of type j, and the root class corresponds to the class of the vertex in G.*

If we can show that, under certain conditions, with probability tending to 1 with d, there exists a sequence of valid Karp–Sipser vertex removals in this tree, all of which are at depth at most d, and which result in the root being removed, this then implies that the Karp–Sipser algorithm is subcritical. This is because we know that, in the limit of n, any structure that appears in the first d levels of the tree is equally likely to appear in the d-neighbourhood of a given vertex in G; so, if with probability at least 1 - ϵ there is a way to remove the root of such a tree for any root class, the expected number of vertices remaining in G after Phase 1 is at most ϵn. The following characterizes these conditions:

▶ **Lemma 12.** *The probability of removing the root of this branching process tends to 1 in the limit of d whenever the system $x_i = e^{-\left(\sum_j c_{ij}\bar{S}_j e^{-\left(\sum_k c_{jk}\bar{S}_k x_k\right)}\right)}$ has no more than one solution $(x_1,\ldots,x_q)$ in $[0,1]^q$.*

The proof of Lemma 12 is given in the extended version of the paper, where it is derived as a corollary of results about winning probabilities of games on multitype branching processes (those results are an extension of work of Holroyd and Martin on Galton–Watson trees [7], and may be of independent interest). We now prove the forward (true) direction of our claim.

Proof of Theorem 9. By the above analysis, it suffices to show that whenever $\sum_j c_{ij}\bar{S}_j < e$ for all classes i, the function
$$\begin{bmatrix} x_1 \\ \ldots \\ x_q \end{bmatrix} \mapsto \begin{bmatrix} e^{-\left(\sum_j c_{1j}\bar{S}_j e^{-\left(\sum_k c_{jk}\bar{S}_k x_k\right)}\right)} \\ \ldots \\ e^{-\left(\sum_j c_{qj}\bar{S}_j e^{-\left(\sum_k c_{jk}\bar{S}_k x_k\right)}\right)} \end{bmatrix} - \begin{bmatrix} x_1 \\ \ldots \\ x_q \end{bmatrix}$$
has only one root on $[0,1]^q$. Denoting $e^{-\left(\sum_j c_{ij}\bar{S}_j e^{-\left(\sum_k c_{jk}\bar{S}_k x_k\right)}\right)}$ as f_i, the Jacobian of this function looks like
$$\begin{bmatrix} f_1\left(\sum_j \left(c_{1j}\bar{S}_j\right)\left(c_{j1}\bar{S}_1\right)e^{-\left(\sum_k c_{jk}\bar{S}_k x_k\right)}\right) - 1 & \ldots & f_1\left(\sum_j \left(c_{1j}\bar{S}_j\right)\left(c_{jq}\bar{S}_q\right)e^{-\left(\sum_k c_{jk}\bar{S}_k x_k\right)}\right) \\ \ldots & \ldots & \ldots \\ f_q\left(\sum_j \left(c_{qj}\bar{S}_j\right)\left(c_{j1}\bar{S}_1\right)e^{-\left(\sum_k c_{jk}\bar{S}_k x_k\right)}\right) & \ldots & f_q\left(\sum_j \left(c_{qj}\bar{S}_j\right)\left(c_{jq}\bar{S}_q\right)e^{-\left(\sum_k c_{jk}\bar{S}_k x_k\right)}\right) - 1 \end{bmatrix}.$$
The sum of the entries in the ith row of this matrix is $f_i\left(\sum_j \left(c_{ij}\bar{S}_j e^{-\left(\sum_k c_{jk}\bar{S}_k x_k\right)} \cdot \sum_l \left(c_{jl}\bar{S}_l\right)\right)\right) - 1$. By assumption, we know $\sum_l \left(c_{jl}\bar{S}_l\right) < e$. So, the above expression is strictly less than $ef_i \log f_i - 1$. For any value of f_i, $ef_i \log f_i$ is at

most 1 (taking the derivative, we find a unique maximum at $f_i = 1/e$). So, we have shown that the sum of every row of the Jacobian is negative everywhere. Now, suppose that this function has two distinct roots, $x = (x_1, \ldots, x_q)$ and $y = (y_1, \ldots, y_q)$. Let i be the index where $y_i - x_i$ is maximal. We have increased x_i by $(y_i - x_i)$, and increased all the other coordinates of x by at most $(y_i - x_i)$. We know that the directional derivative of the ith coordinate in the $[1, \ldots, 1]^\top$ direction is negative, and that the partial derivative with respect to every $j \neq i$ is positive; this implies that the ith coordinate of the function at y must be smaller than the ith coordinate of the function at x, so they cannot both be roots. ◀

As a consequence of this, since it is guaranteed to be optimal in Phase 1, we know the Karp–Sipser algorithm gives a near-optimal matching whenever $\sum_j c_{ij} \bar{S}_j < e$ for all i. By examining the system given in Lemma 12, however, we find that this is not a necessary condition for subcriticality – the phase transition boundary is much more complicated.

2.5 Failure of the Karp–Sipser algorithm

In the previous sections we gave two instances where we can guarantee that the Karp–Sipser algorithm achieves a near-optimal matching, both which use essentially the same framework as the Erdős–Rényi case (i.e., showing that it can achieve within $o(n)$ of a perfect matching during Phase 2, either because all degrees are close to 2 in the equitable case, or because the entire remaining graph has $o(n)$ vertices in the subcritical case). However, the algorithm does **not** return a near-optimal matching in general, and even when it does, this analytical framework does not always work. For example, consider a stochastic block model with 4 classes, all of size $n/4$, and the following probability matrix:

$$\begin{bmatrix} 0 & \frac{100}{n} & 0 & 0 \\ \frac{100}{n} & 0 & \frac{10000}{n} & 0 \\ 0 & \frac{10000}{n} & 0 & \frac{100}{n} \\ 0 & 0 & \frac{100}{n} & 0 \end{bmatrix}.$$

The matching number is very close to perfect. Even ignoring the edges between classes 2 and 3 entirely, we are left with two copies of $G(n/4, n/4, \frac{100}{n})$, on which our equitable analysis guarantees a matching containing more than a .99 fraction of vertices. However, directly analyzing the Karp-Sipser algorithm, we find that Phase 1 finishes very quickly, because the graph is dense enough that very few degree 1 vertices are created. Then, in Phase 2, for a long time we are in the regime of short runs as described in our analysis of the equitable case; in this regime, the algorithm chooses many of its edges uniformly at random, and so likely choose many of them from between classes 2 and 3. Every edge choosen between classes 2 and 3, however, effectively decreases the matching number by 2. Formalizing this argument reveals that the Karp–Sipser algorithm on this graph finds a matching containing only slightly more than $1/2$ of the vertices; an illustration is given in Figure 1.

This suggests the need for a modified version of the Karp–Sipser algorithm that is given the stochastic block model parameters, and takes into account the label classes of the vertices. We propose the following: whenever the algorithm must choose a random edge, instead of choosing uniformly over all edges it instead chooses uniformly from the edge type that we estimate will create the fewest degree-1 vertices on this step. This algorithm performed well in our experimental simulations for a range of block model parameters; we conjecture that it performs near-optimally asymptotically. We have been unable to prove this, however. Part of the reason for the difficulty of analysis lies in the fact that we no longer expect a perfect matching to be possible in Phase 2 – the following section illustrates this with a simple example.

(a) Karp–Sipser.   **(b)** Optimal.
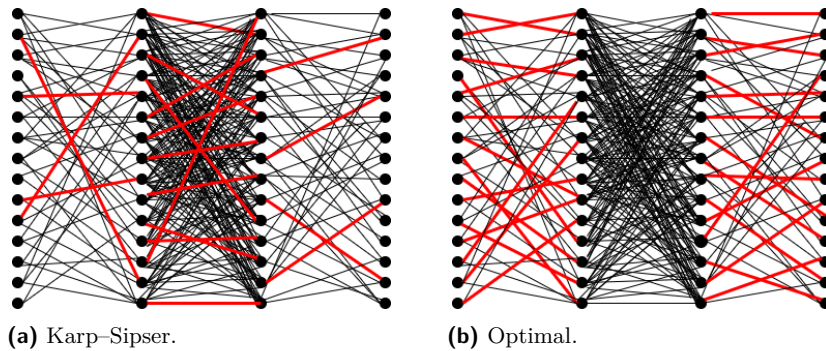
■ **Figure 1** A stochastic block model on which the Karp–Sipser algorithm performs suboptimally.

2.6   Bipartite Erdős–Rényi case

The bipartite graph $G(n, n, c/n)$ with equal part sizes and i.i.d. edges corresponds to an equitable stochastic block model: each vertex on the left and each vertex on the right has c expected neighbours. So, with high probability, the Karp–Sipser algorithm performs optimally and returns a matching the same size as that of $G(2n, c/2n)$. However, the asymmetric case $G(kn, n, c/n)$ where $k \neq 1$ is not equitable, because left vertices have c expected neighbours while right vertices have kc of them. It is this case that we analyze in this section. Note that, as before, the algorithm is guaranteed to be optimal for Phase 1; we need to show that it's also near optimal for Phase 2. However, now even the optimal algorithm on Phase 2 is not guaranteed to find a near-perfect matching. Consider a graph with very unequal part sizes, and very high c; for example, $G(n, 10n, 1000/n)$. This graph is dense enough that we expect at least $5n$ vertices to remain on the right after Phase 1, however we know that the matching number of this graph is at most n – so, we can't actually expect Phase 2 to find a near-perfect matching. Thus, this is an example where the typical analytical framework does not apply [9, 1]. However, we can still use the Karp-Sipser algorithm to determine the matching numbers of such graphs. We sketch the argument.

▶ **Lemma 13.** *Let F_l, F_r and E denote the number of non-isolated unmatched vertices on the left, non-isolated unmatched vertices on the right, and edges in the graph after Phase 1 of the Karp–Sipser algorithm. With high probability, there exists a matching of size $\min(F_l, F_r) - o(n)$ on this graph. As a consequence, the asymptotic behavior of the matching numbers of bipartite Erdős–Rényi graphs can be determined from solutions to the ODEs described in Section 2.2.*

Proof. Assume without loss of generality that $F_l \leq F_r$. We would like to show that there exists a matching of size $F_l - o(n)$. To do so, we recall by the Markov property (see Appendix A) of the Karp-Sipser algorithm that the distribution of the graph G at this point in the process is a uniform random configuration model with E edges on $F_l + F_r$ vertices, conditioned on being bipartite with part sizes F_l and F_r, and having min degree 2. We'll couple G with a graph G' that's drawn from a bipartite configuration model with E edges, min degree 2, and parts both of size F_r. The way the coupling will work is to first generate G', then delete all but the first F_l vertices on the left (where "first" refers to some arbitrary ordering), and redistribute the half-edges connected to the deleted vertices among the remaining vertices on the left. Note there must exist some such randomized redistribution process that causes the resulting graph to be distributed as G, since degree sequences of the first F_l left vertices in a graph distributed as G' are dominated by degree sequences of the left vertices in a graph distributed as G.

To show that G with high probability contains a matching of size $F_l - o(n)$, we draw (G', G) according to this coupling. By our analysis of Phase 2 of the Karp–Sipser algorithm in equitable block models, we know that with high probability G' contains a near-perfect matching. So, in particular, there exists a matching in G' that contains $F_l - o(n)$ of the first F_l vertices on the left. By construction of the coupling, all the edges involved in that matching are also present in G, so this implies the existence of a matching of size $F_l - o(n)$ in G.

This shows a lower bound on the matching number of $G(kn, n, c/n)$. Since Phase 1 of the Karp–Sipser algorithm is optimal, and once Phase 1 ends the matching number is clearly bounded by $\min(F_l, F_r)$, this lower bound is tight (up to $o(n)$).   ◀

This analysis works because, even though the Karp–Sipser core does not admit a perfect matching, we have an upper bound on its true matching number (namely, the size of the smaller of the two parts) that we can show is nearly achieved. This is a natural approach to analyzing the matching number of a random graph: try to show an upper bound on the matching number of the Karp–Sipser core, and then show a variation of the Karp–Sipser algorithm that achieves close to that upper bound on the core. It would be interesting to find other block models where this approach is successful. A simple case we note as an open direction is the bipartite setting where there is only one class on the left (i.e., the class graph is "star shaped"). On such graphs, Label-Aware Karp–Sipser simplifies to "prefer edges to the available right class with minimum average degree" – it seems conceivable that some form of analogous argument could show that this is optimal.

3   Analysis of Online Matching Heuristics

In this section, we consider the problem of ***online matching*** in bipartite stochastic block models with q left and q right classes. We assume that vertices are assigned classes uniformly and independently at random, with classes of the right vertices known to the algorithm ahead of time, and vertices on the left (along with all the edges incident to them) arriving one-at-a-time. Uniformity of the class distribution is roughly without loss of generality; variations could be approximated by further subdividing the classes.

Before discussing algorithms for online matching in stochastic block models, we recall the simpler setting of $G(n, n, c/n)$. First, note that any online algorithm that chooses not to match a left vertex when it is adjacent to at least one available right vertex is sub-optimal. Since, in this setting, all right vertices are indistinguishable, any online algorithm that always matches when possible is optimal. This gives the following:

▶ **Fact 14** (Mastin and Jaillet [14])**.** *In the online setting, the expected size of a matching in* $G(n, n, c/n)$ *produced by an optimal online algorithm is given by* $\left(1 - \frac{\ln(2 - e^{-c})}{c}\right) n$.

The stochastic block model setting allows for more nuance; designing optimal algorithms is nontrivial in general. While it is possible to brute-force compute an optimal online strategy in time $\Theta(n^{q+1})$ using dynamic programming, we leave open the question of whether such an optimum can be achieved asymptotically in constant time per decision. Instead, we first show that in the restricted class of *equitable* stochastic block models, a simple greedy heuristic still achieves asymptotically optimal matchings. We then propose a more sophisticated heuristic with good experimental performance, but demonstrate a case where it can be shown analytically to be asymptotically sub-optimal.

3.1   Analysis of GREEDY

First, consider the simplest possible matching heuristic. For each left vertex that arrives, GREEDY chooses uniformly at random one of the available edges adjacent to that vertex to add to the matching. While GREEDY is sub-optimal in general, we can show that it returns an asymptotically optimal matching in any equitable block model.

▶ **Lemma 15.** *In the equitable case, i.e. when all of the vertices have the same expected degree $c = \frac{1}{q} \sum_i c_{ij}$, GREEDY returns a matching of expected size $\left(1 - \frac{\ln(2-e^{-c})}{c}\right) n$ (as in the bipartite Erdős-Rényi case; see Fact 14).*

Proof. Let R_j denote the set of vertices in right class j, and $R_j^{(t)} \subseteq R_j$ denote the subset that remain unmatched after t steps. If the next left vertex to arrive has class i, the probability that it has no available matches is $\left(\prod_{j=1}^{q}(1 - p_{ij})^{|R_j^{(t)}|}\right)$. Conditional on at least one available match, the probability that GREEDY chooses to match to a vertex of R_j is $\frac{p_{ij}|R_j^{(t)}|}{\sum_{l=1}^{q} p_{il}|R_l^{(t)}|} + \mathcal{O}\left(\frac{1}{\sum_l |R_l^{(t)}|}\right)$ – this can be seen by replacing the selection process with "repeatedly choose a random right vertex and match with probability p_{ij}", which has the same distribution up to $\mathcal{O}\left(\frac{1}{\sum_l |R_l^{(t)}|}\right)$ replacement error. So, we have

$$|R_j^{(t+1)}| \approx \begin{cases} |R_j^{(t)}| - 1 \text{ with probability} & \sum_{i=1}^{q} \frac{1}{q}\left(1 - e^{-\sum_{l=1}^{q} p_{il}|R_l^{(t)}|}\right) \frac{p_{ij}|R_j^{(t)}|}{\sum_{l=1}^{q} p_{il}|R_l^{(t)}|} \\ |R_j^{(t)}| \quad \text{ with probability } 1 - \sum_{i=1}^{q} \frac{1}{q}\left(1 - e^{-\sum_{l=1}^{q} p_{il}|R_l^{(t)}|}\right) \frac{p_{ij}|R_j^{(t)}|}{\sum_{l=1}^{q} p_{il}|R_l^{(t)}|} \end{cases}.$$

Letting $x_j(\tau) = \frac{|R_j^{(\tau n)}|}{n}$, we can appeal to Wormald's theorem to argue that this process is controlled in the limit $n \to \infty$ by the following ODE:

$$x_j'(\tau) = -\sum_{i=1}^{q} \frac{1}{q}\left(1 - e^{-\sum_{l=1}^{q} c_{il}x_l(\tau)}\right) \frac{c_{ij}x_j(\tau)}{\sum_{l=1}^{q} c_{il}x_l(\tau)} \qquad j \in \{1, 2, \ldots q\},$$

with initial conditions $x_j(0) = \frac{1}{q}$ for all $j \in \{1, 2, \ldots q\}$. If we guess that the solution is symmetric – that is, that $x_j(\tau) = \frac{x(\tau)}{q}$ for some x and all j, τ – then these equations simplify to

$$x'(\tau) = -\sum_{i=1}^{q}\left(1 - e^{-\sum_{l=1}^{q} c_{il}x(\tau)/q}\right) \frac{c_{ij}x(\tau)/q}{\sum_{l=1}^{q} c_{il}x(\tau)/q} = -\sum_{i=1}^{q}\left(1 - e^{-cx(\tau)}\right)\frac{c_{ij}}{c} = 1 - e^{-cx(\tau)},$$

since when all the $x_j(\tau)$ are equal, we also have all the $x_j'(\tau)$ equal. Solving this ODE yields $x(1) = \left(\frac{\ln(2-e^{-c})}{c}\right)$, meaning by Wormald's that the expected matching size converges to $\left(1 - \frac{\ln(2-e^{-c})}{c}\right) n$. ◀

Note that, unlike in $G(n, n, c/n)$, it is **not** the case that any strategy for matching available vertices achieves this value – for example, the strategy of always preferring to match class R_1 will perform worse in general. It was important that we chose a strategy that evenly distributed matches across right classes. We'll now show that this lower bound is tight; i.e., that no algorithm can do asymptotically better than GREEDY on an equitable block model.

▶ **Lemma 16.** *In an equitable stochastic block model, if there are a total of $|R^{(t)}|$ unmatched vertices on the right, the probability of matching on the next step is maximized when those $|R^{(t)}|$ vertices are equally distributed among all classes (that is, each right class r has $|R_r^{(t)}| = \frac{|R^{(t)}|}{q}$ unmatched vertices).*

Proof. Let $\rho_r = \frac{|R_r^{(t)}|}{|R^{(t)}|}$ denote the fraction of unmatched right vertices belonging to class r at time t. The probability of there being no available edge in the next step of the algorithm is, as $n \to \infty$, $\frac{1}{q} \sum_l e^{-\sum_r c_{lr}\rho_r}$. By AM-GM inequality, this is at least $\left(\prod_l e^{-\sum_r c_{lr}\rho_r}\right)^{1/q} = e^{-\sum_l \frac{1}{q}\sum_r c_{lr}\rho_r} = e^{-\sum_r \frac{1}{q}(\sum_l c_{lr})\rho_r} = e^{-c}$, which is precisely the value obtained by setting all $\rho_r = \frac{1}{q}$. ◀

From this fact, we see that no algorithm can do asymptotically better than GREEDY.

▶ **Theorem 17.** *The optimal competitive ratio for any online algorithm in an equitable stochastic block model is $matchingsize \frac{c-\ln(2-e^{-c})}{(2c-x+ce^{-x}+xce^{-x})}$, where x is the smallest solution of $x = ce^{-ce^{-x}}$.*

Proof. First, note that this value is precisely what we've shown for the competitive ratio of GREEDY; this can be found by dividing the asymptotic matching size we proved in that case by the offline matching number we proved in Theorem 5. Therefore, it is left to establish that no algorithm can do better than GREEDY. We proceed by contradiction - assume there is a better algorithm, and consider the expected values of ρ_i as in Lemma 16 in this algorithm. If the ρ_i's stay within $o(1)$ of equal over all time, then the algorithm looks asymptotically identical to GREEDY. If at some point they become unequal, then at that time the algorithm must fall behind GREEDY and never catch up, by Lemma 16. ◀

Note that this value is the competitive ratio lower bound conjectured by Mastin and Jaillet to be tight for Erdős–Rényi graphs; as a special case, we have shown that conjecture [14].

3.2 A Less Greedy Algorithm

Although GREEDY is optimal in equitable models, it's not difficult to describe non-equitable models where it's very far from optimal. For example, if $p_{ij} < p_{il}$ for all l, it is always strictly better to match to vertices of class j than class l if possible (i.e. if one class is very rare to match to, it's good to take those matches when the chance arises), but GREEDY does not do this. To correct this, one might consider always matching to the right class with the lowest expected degree – however it's easy to see that this is not optimal, since when all classes have close to the same expected degree, it's better to distribute matches more evenly across classes than to focus on the one with the smallest expected degree. Determining the appropriate extent to prefer rare classes versus trying to keep class sizes balanced is a non-trivial task.

A reasonable heuristic – which we'll term SHORTSIGHTED– is to, at each step, minimize the probability of being unable to match the next vertex (i.e., the probability that the next left vertex to arrive has no available edges). That is, SHORTSIGHTED always chooses to match to the available class l that minimizes $\sum_{i=1}^{q} \frac{1}{q} \prod_{j=1}^{q} (1-p_{ij})^{|R_j^{(t)}|-\mathbb{1}_{j=l}}$.

One might hope that looking a single step into the future is sufficient to determine the optimal class to match to, and so SHORTSIGHTED is an asymptotically optimal approach. Indeed, in a range of computer experiments we found that SHORTSIGHTED found a matching of almost the same size as a brute-force optimal online algorithm. However, it turns out that SHORTSIGHTED is not asymptotically optimal.
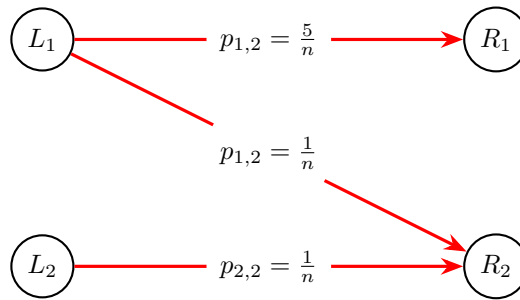
Figure 2 An example of stochastic block model parameters in which BRUTE-FORCE beats SHORTSIGHTED. Here all classes are of size $\frac{n}{2}$, and $p_{2,1} = 0$.

In the model shown in Figure 2, SHORTSIGHTED prefers class 1 whenever the number of unmatched vertices of class R_1 is at least $\frac{2\ln 2}{5}n$. By associating SHORTSIGHTED's behaviour in the limit $n \to \infty$ with an ODE, we obtain that the expected size of the matching is $\approx 0.574946n$ (this analysis is outlined in Appendix B). To show this is suboptimal, consider the following algorithm: prefer class R_1 until $.88n$ vertices have arrived, then thereafter prefer R_2. With essentially the same differential equation analysis as above, we find that the expected size of the matching is $\approx 0.575597n$. This algorithm out-performs SHORTSIGHTED by a linear factor in n (albeit small, $\approx .0006n$), implying that SHORTSIGHTED is asymptotically sub-optimal.

References

1 Jonathan Aronson, Alan M. Frieze, and Boris G. Pittel. Maximum matchings in sparse random graphs: Karp-sipser revisited. *Random Struct. Algorithms*, 12:111–177, 1998.

2 Tom Bohman and Alan M. Frieze. Karp–sipser on random graphs with a fixed degree sequence. *Combinatorics, Probability and Computing*, 20:721–741, 2011.

3 Thomas Budzinski, Alice Contat, and Nicolas Curien. The critical Karp–Sipser core of random graphs, 2022. `arXiv:2212.02463`.

4 Byron Chin and Allan Sly. Optimal reconstruction of general sparse stochastic block models, 2021. `arXiv:2111.00697`.

5 Jon Feldman, Aranyak Mehta, Vahab S. Mirrokni, and S. Muthukrishnan. Online stochastic matching: Beating 1-1/e. *CoRR*, abs/0905.4100, 2009. `arXiv:0905.4100`.

6 Paul W. Holland, Kathryn Blackmond Laskey, and Samuel Leinhardt. Stochastic blockmodels: First steps. *Social Networks*, 5(2):109–137, 1983. `doi:10.1016/0378-8733(83)90021-7`.

7 Alexander E. Holroyd and James B. Martin. Galton–watson games. *Random Structures & Algorithms*, 59(4):495–521, 2021. `doi:10.1002/rsa.21008`.

8 Zhiyi Huang, Xinkai Shu, and Shuyi Yan. The power of multiple choices in online stochastic matching, 2022. `arXiv:2203.02883`.

9 R. M. Karp and M. Sipser. Maximum matching in sparse random graphs. In *22nd Annual Symposium on Foundations of Computer Science (SFCS 1981)*, pages 364–375, 1981. `doi:10.1109/SFCS.1981.21`.

10 Richard M. Karp, Umesh V. Vazirani, and Vijay V. Vazirani. An optimal algorithm for on-line bipartite matching. In *Symposium on the Theory of Computing*, 1990.

11 Kamer Kaya, Johannes Langguth, Ioannis Panagiotas, and Bora Uçar. *Karp-Sipser based kernels for bipartite graph matching*, pages 134–145. Society for Industrial and Applied Mathematics, 2020. `doi:10.1137/1.9781611976007.11`.

12 Johannes Langguth, Ioannis Panagiotas, and Bora Uçar. Shared-memory implementation of the karp-sipser kernelization process. In *2021 IEEE 28th International Conference on High Performance Computing, Data, and Analytics (HiPC)*, pages 71–80, 2021. `doi:10.1109/HiPC53243.2021.00021`.

13 Vahideh H. Manshadi, Shayan Oveis Gharan, and Amin Saberi. Online stochastic matching: Online actions based on offline statistics, 2011. `arXiv:1007.1673`.

14 Andrew Mastin and Patrick Jaillet. Greedy online bipartite matching on random graphs. *CoRR*, abs/1307.2536, 2013. `arXiv:1307.2536`.

15 Elchanan Mossel, Joe Neeman, and Allan Sly. A proof of the block model threshold conjecture. *Combinatorica*, 38(3):665–708, 2018.

16 Flore Sentenac, Nathan Noiry, Matthieu Lerasle, Laurent Ménard, and Vianney Perchet. Online matching in geometric random graphs, 2023. `arXiv:2306.07891`.

17 Nahuel Soprano-Loto, Matthieu Jonckheere, and Pascal Moyal. Online matching for the multiclass stochastic block model, 2023. `arXiv:2303.15374`.

18 Lutz Warnke. On wormald's differential equation method, 2019. `arXiv:1905.08928`.

19 Nicholas C. Wormald. Differential Equations for Random Processes and Random Graphs. *The Annals of Applied Probability*, 5(4):1217–1235, 1995. `doi:10.1214/aoap/1177004612`.

A   Karp–Sipser Transition Probabilities

A.1   Passing to a Configuration Model

For analytical purposes, we consider the Karp–Sipser algorithm on multigraphs drawn from a configuration model, and then justify why the results translate to stochastic block models.

▶ **Definition 18** (Blocked configuration model)**.** *Let G be a graph drawn from a stochastic block model. For each pair of labels i, j, let E_{ij} denote the total number of edges between vertices of label i and vertices of label j in G (when $i = j$, we let E_{ij} denote* twice *the number of edges lying within class $i = j$, so that E_{ij} always refers to the number of j-type half-edges attached to label-i vertices). Construct a random multigraph as follows: if $i \neq j$, distribute E_{ij} half-edges among the i vertices and E_{ij} half-edges among the j vertices uniformly at random (à la balls-in-bins), then define edges by a uniform random pairing between the half edges on the i and j sides. This process defines a distribution over multigraphs, because in the process of "reshuffling" the half-edges of G we introduce the possibility of multiple edges and self-loops.*

Note that any result which holds with high probability in the blocked configuration model also holds with high probability for the stochastic block model, so we are justified in proving our results for this model instead (this is formally justified in the extended version of the paper).

A.2   Markov Property of Karp–Sipser Algorithm

As we run the Karp–Sipser algorithm on a random multigraph, the distribution changes, since the earlier steps of the algorithm have produced some effects conditional on the previous states of the graph. Fortunately, it turns out that these effects have a simple description.

▶ **Lemma 19.** *Suppose we generate a graph from the blocked configuration model and run the Karp–Sipser algorithm for an arbitrary number of steps. Then conditioned on the following, the resulting graph is still distributed according to the blocked configuration model.*

- *For each pair of (i, k) of block model labels, the number E_{ij} of edges between label class i and j.*
- *For each label class i, the number T_i of label-i vertices of degree exactly 1 (the "thin" vertices).*
- *For each label class i, the number F_i of label-i vertices of degree at least 2 (the "fat" vertices).*

Thus, if we collect all of these values into a tuple $Y = (E_{ij}, T_i, F_i)$, the algorithm's progress can be described by a Markov chain on Y.

The proof of this lemma is a relatively straightforward counting argument; it is described in detail in the extended version of the paper.

A.3   Estimates on Degree Distributions

In order to determine transition probabilities of this Markov process, we first note the following straightforward lemma about the degree distribution (whose proof we will again defer to the extended version):

▶ **Lemma 20.** *Fix a description tuple Y. Let nY be the tuple obtained by scaling every element of Y by n, and let G_n be a multigraph drawn from a blocked configuration model conditional on tuple nY. Choose a uniform half-edge incident to class i, and let v be its incident vertex. In the limit as $n \to \infty$, the degree of v converges in distribution to*

$$\mathbb{P}[d(v) = k] = \begin{cases} \frac{T_i}{\sum_l E_{il}} & \text{for } k = 1 \\ \frac{\left(\sum_l E_{il}\right) - T_i}{\sum_l E_{il}} \cdot \frac{\lambda_i^{k-1}}{(k-1)! e^{\lambda_i} \left(1 - e^{-\lambda_i}\right)} & \text{for } k > 1, \end{cases}$$

where λ_i is a solution to $\frac{\lambda_i e^{\lambda_i} - \lambda_i}{e^{\lambda_i} - 1 - \lambda_i} = \frac{\left(\sum_l E_{il}\right) - T_i}{F_i}$.

This allows us to determine the probability that a random edge is adjacent to a degree 2 vertex, which is important for understanding the evolution of the algorithm.

▶ **Corollary 21.** *In particular, as $n \to \infty$, the probability that a random edge into class i connects to a vertex of degree exactly 2 tends to $\frac{\left(\sum_l E_{il}\right) - T_i}{\left(\sum_l E_{il}\right)} \cdot \frac{\lambda_i}{e^{\lambda_i} - 1}$.*

The above lemma also allows us to determine the label distribution in the neighborhood of a randomly chosen edge.

▶ **Corollary 22.** *Let uv be a randomly chosen edge where u has label i and v has label j. As $n \to \infty$, the number of neighbors of u with label k tends to $\frac{\left(\sum_l E_{il}\right) - T_i}{\sum_l E_{il}} \cdot \frac{\lambda_i}{1 - e^{-\lambda_i}} \cdot \frac{E_{ik}}{\left(\sum_l E_{il}\right)}$.*

A.4   Computing Transition Probabilities

From these degree distribution estimates, we produce estimates on the transition probabilities of the Markov process. On each step of the algorithm, if there exist degree 1 vertices, the algorithm chooses one of them and removes it and its neighbour. So, the graph loses one edge from between class i and j whenever that degree 1 vertex has its edge between i and j. When the neighbour vertex is in class i, the graph also loses edges equal to however many neighbours it had in class j. Similar accounting can be made for the number of fat or thin vertices in a class: the graph loses a fat vertex either by having it as the neighbour of the degree 1 vertex we removed, or by having it initially have degree 2 and appear as the

neighbour of a removed neighbour, so that it's then reduced to degree 1. A thin vertex is lost whenever its the removed degree 1 vertex, whenever its the neighbour of the removed vertex, or whenever its a neighbour of the neighbour of a removed vertex, but gain one whenever a degree 2 vertex is a neighbour-of-the-neighbour. To make these expressions explicit, we define the notation:

- Let h_i denote the total number of half-edges in class i, $h_i = \sum_l E_{il}$.
- Let ω_{ij} denote the probability that a randomly selected degree-1 vertex is in class i and has its neighbour in class j, $\omega_{ij} = \frac{T_i}{\sum_l T_l} \cdot \frac{E_{ij}}{h_i}$.
- Let δ_{ij} denote expected number of other j-type half-edges attached to the vertex a random half-edge in class i is attached to. If λ_i is a solution to $\frac{\lambda_i(e^{\lambda_i}-1)}{e^{\lambda_i}-1-\lambda_i} = \frac{h_i-T_i}{F_i}$, then, by our degree estimates, $\delta_{ij} = \frac{h_i-T_i}{h_i} \cdot \frac{\lambda_i}{1-e^{-\lambda_i}} \cdot \frac{E_{ij}}{h_i}$.
- Let θ_i denote the probability that a random half-edge attached to class i is attached to a degree 2 vertex. Again, letting $\frac{\lambda_i(e^{\lambda_i}-1)}{e^{\lambda_i}-1-\lambda_i} = \frac{h_i-T_i}{F_i}$, we have $\theta_i = \frac{h_i-T_i}{h_i} \cdot \frac{\lambda_i}{e^{\lambda_i}-1}$.

Now, while there are degree-1 vertices remaining in the graph, we can write the expected change in the description tuple after one step of the algorithm as

$$\mathbb{E}[\Delta E_{ij}] = -\omega_{ij} - \omega_{ji} - \sum_l \omega_{li}\delta_{ij} - \sum_l \omega_{lj}\delta_{ji}$$

$$\mathbb{E}[\Delta F_i] = -\left(\sum_l \omega_{li}\right)\left(\frac{h_i-T_i}{h_i}\right) - \sum_j \sum_l \omega_{jl}\delta_{li}\theta_i$$

$$\mathbb{E}[\Delta T_i] = -\left(\sum_l \omega_{il}\right) - \left(\sum_l \omega_{li}\right)\left(\frac{T_i}{h_i}\right) - \sum_j \sum_l \omega_{jl}\delta_{li}\left(\frac{T_i}{h_i} - \theta_i\right).$$

On the other hand, when there are no degree 1 vertices remaining, we choose an edge uniformly at random, so, by similar reasoning,

$$\mathbb{E}[\Delta E_{ij}] = -\left(\frac{2E_{ij}}{\sum_k h_k}\right) - \left(\frac{2h_i}{\sum_k h_k}\right)\delta_{ij} - \left(\frac{2h_j}{\sum_k h_k}\right)\delta_{ji}$$

$$\mathbb{E}[\Delta F_i] = -\left(\frac{2h_i}{\sum_k h_k}\right) - \sum_l \left(\frac{2h_l}{\sum_k h_k}\right)\delta_{li}\theta_i$$

$$\mathbb{E}[\Delta T_i] = \sum_l \left(\frac{2h_l}{\sum_k h_k}\right)\delta_{li}\theta_i.$$

An important point to note is that the expected transitions above are "scale-invariant", meaning that they remain the same upon re-scaling all entries in Y by the same amount. So, letting $\bar{E}_{ij} = E_{ij}/n$, $\bar{T}_i = T_i/n$, $\bar{F}_i = F_i/n$, we can write (for Phase 1 – Phase 2 is similar):

$$\mathbb{E}[\Delta \bar{E}_{ij}] = \frac{-\omega_{ij} - \omega_{ji} - \sum_l \omega_{li}\delta_{ij} - \sum_l \omega_{lj}\delta_{ji}}{n}$$

$$\mathbb{E}[\Delta \bar{F}_i] = -\frac{\left(\sum_l \omega_{li}\right)\left(\frac{\bar{h}_i-\bar{T}_i}{\bar{h}_i}\right) - \sum_j \sum_l \omega_{jl}\delta_{li}\theta_i}{n}$$

$$\mathbb{E}[\Delta \bar{T}_i] = -\frac{\left(\sum_l \omega_{il}\right) - \left(\sum_l \omega_{li}\right)\left(\frac{\bar{T}_i}{\bar{h}_i}\right) - \sum_j \sum_l \omega_{jl}\delta_{li}\left(\frac{\bar{T}_i}{\bar{h}_i} - \theta_i\right)}{n}.$$

B   Conditions of Wormald's Theorem

We on several occasions in this paper claim that particular Markov processes remain close to a limiting system of differential equations. In this section, we step through for each of those instances the justification of those claims. The key tool is a theorem of Wormald, restated in its general form below. Here, n indexes a family of discrete time random processes, each of which has "history" sequence $H_n \in S_n^+$. The notation Y_t is shorthand for $y(H_t)$.

▶ **Theorem 23** (Wormald [19]). *Let a be fixed. For $1 \le l \le a$, let $y^{(l)} : \bigcup_n S_n^+ \to \mathbb{R}$ and $f_l : \mathbb{R}^{a+1} \to \mathbb{R}$, such that for some constant C and all l, $|y^{(l)}(h_t)| < Cn$ for all $h_t \in S_n^+$ for all n. Suppose also that for some function $m = m(n)$:*

1. *for some functions $w = w(n)$ and $\lambda = \lambda(n)$ with $\lambda^4 \log n < w < n^{2/3}/\lambda$ and $\lambda \to \infty$ as $n \to \infty$, for all l and uniformly for all $t < m$,*

$$\mathbb{P}\left[|Y_{t+1}^{(l)} - Y_t^{(l)}| > \frac{\sqrt{w}}{\lambda^2 \sqrt{\log n}} \,\Big|\, H_t\right] = o(n^{-3});$$

2. *for all l and uniformly over all $t < m$, we always have*

$$\mathbb{E}(Y_{t+1}^{(l)} - Y_t^{(l)}|H_t) = f_l(t/n, Y_t^{(1)}/n, \dots, Y_t^{(a)}/n) + o(1);$$

3. *for each l, the function f_l is continuous and satisfies a Lipschitz condition on D, where D is some bounded connected open set containing the intersection of $\{(t, z^{(1)}, \dots, z^{(a)}) : t \ge 0\}$ with some neighbourhood of $\{(0, z^{(1)}, \dots, z^{(a)}) : \mathbb{P}(Y_0^{(l)} = z^{(l)}n, 1 \le l \le a) \ne 0$ for some $n\}$.*

Then,

1. *For $(0, \hat{z}^{(1)}, \dots, \hat{z}^{(a)}) \in D$, the system of differential equations*

$$\frac{dz_l}{ds} = f_l(s, z_1, \dots, z_a), \qquad l = 1, \dots, a,$$

has a unique solution in D for $z_l : \mathbb{R} \to \mathbb{R}$ passing through

$$z_l(0) = \hat{z}^{(l)}, \qquad 1 \le l \le a$$

and which extends to points arbitrarily close to the boundary of D.

2. *Almost surely,*

$$Y_t^l = n z_l(t/n) + o(n)$$

uniformly for $0 \le t \le \min\{\sigma n, m\}$ and for each l, where $z_l(t)$ is the solution in (i) with $\hat{z}^{(l)} = Y_0^{(l)}/n$, and $\sigma = \sigma(n)$ is the supremum of those s to which a solution can be extended.

For a simple proof of Wormald's theorem, we refer the reader to Warnke's recent exposition [18].

B.1   Phase 1 of the Karp–Sipser algorithm

The first time we make use of this differential equations method is in the analysis of the first phase of the Karp–Sipser algorithm. We will outline how to apply Wormald's theorem in this case. Here, we take Y as we defined it, f as the derivative we wrote down for the corresponding ODE, and note that with high probability, $y^{(l)}(h_t) < 100(\max_{ij} c_{ij})n$ for all

sufficiently large n (the F_i and T_i components of $y^{(l)}(h_t)$ are clearly bounded by n; w.h.p. the edge counts are all initially within a factor of 100 of their expectations, and once they start that way they never increase). We take the stopping time m to be the first time all of the T_i entries of Y drop below $n^{0.01}$. Now,

1. Take for instance $w = n^{0.5}$ and $\lambda = \log n$. The probability that any vertex in the initial graph has degree polynomial in n decays exponentially in n, and it can be observed that the magnitude of a transition is bounded by twice the maximum degree, so we certainly have the desired condition.

2. This convergence of expected transition size of the Markov process to f is precisely what is guaranteed by the convergence of our degree estimates. Note that this convergence holds as the total number of thin vertices, fat vertices, and edges is going to infinity, so taking our stopping time to be m prevents border cases once these values drop down to constant sizes.

3. To show that f is Lipschitz in the neighbourhood of solutions, it suffices to show that solutions to $z'(t) = f(t, z)$ always maintain constant average degree in each label class. To do so, note that with high probability the initial average degrees are at most $100(\max_{ij} c_{ij})$ in each class, and then observe from the equations that whenever the average degree in a given class i is more than 100 at some time t, $-\sum_j \bar{\mathcal{E}}'_{ij}(t) > \bar{\mathcal{F}}'_i(t) + \bar{\mathcal{T}}'_i(t)$.

Note that we have only shown that the conditions of the theorem hold with high probability over initial graph configurations; clearly, this is sufficient for our desired result.

B.2   Phase 2 of the Karp–Sipser algorithm (equitable case)

We also use Wormald's theorem to justify our analysis of the second phase of the algorithm in the equitable case. Here, Y_i is the state of the tuple after the ith run of the algorithm – i.e., the ith time where there are no thin vertices. We define f_l to be the expected change in the tuple that one of these runs would incur if the transition probability estimates from Appendix A held exactly and did not change throughout the run.

In order for this process to have the desired Lipschitz properties, we will need to define a more restricted domain than the entire possible space of tuples. In particular, we will fix some constants γ and ϵ, and consider the domain of Y to consist only of tuples where the average degree into each class differs by at most γ, and is at least $2 + \epsilon$. The value of ϵ is chosen in the analysis; to determine γ, we observe the following:

- λ_i is defined such that $\frac{\lambda_i(e^{\lambda_i}-1)}{e^{\lambda_i}-1-\lambda_i}$ is equal to the average degree of class i. This is monotonically increasing in the average degree of class i; so there is some constant $\lambda > 0$ such that if the average degree in class i is at least $2 + \epsilon$, then $\lambda_i > \lambda$.

- The function $\delta_i\theta_i = \frac{\lambda_i}{1-e^{-\lambda_i}} \cdot \frac{\lambda_i}{e^{\lambda_i}-1}$ is bounded above by 1 and monotonically decreasing for $\lambda_i > 0$. So, there exists some constant $\eta > 0$ such that $\frac{\lambda}{1-e^{-\lambda}} \cdot \frac{\lambda}{e^{\lambda}-1} < 1 - 2\eta$.

- $\lambda_i, \lambda_j \mapsto \frac{\lambda_i}{1-e^{-\lambda_i}} \cdot \frac{\lambda_j}{e^{\lambda_j}-1}$ is uniformly continuous, so there exists some constant κ such that $|\lambda_i - \lambda_j| < \kappa$ implies that $\left| \left(\frac{\lambda_i}{1-e^{-\lambda_i}} \cdot \frac{\lambda_j}{e^{\lambda_j}-1} \right) - \left(\frac{\lambda_i}{1-e^{-\lambda_i}} \cdot \frac{\lambda_i}{e^{\lambda_i}-1} \right) \right| < \eta$ – which, if $\lambda_i > 2 + \epsilon$, implies $\left(\frac{\lambda_i}{1-e^{-\lambda_i}} \cdot \frac{\lambda_j}{e^{\lambda_j}-1} \right) < 1 - \eta$.

- When defined on the domain $[2 + \epsilon, \infty)$, λ_i is uniformly continuous as a function of the average degree of i. So, we can define γ such that |average degree in class i − average degree in class $j| < \gamma$ implies $|\lambda_i - \lambda_j| < \kappa$ whenever average degrees are greater than $2 + \epsilon$.

We will define the stopping time m of the process to be the first time it leaves this domain. Note that, as in Phase 1, the value of $|Y|$ is bounded by $100c_{\max}n$ with probability $1 - o(1)$, since it is initially and can only decrease. We're now ready to verify the criteria of Wormald's theorem.

1. Take again $w = n^{0.5}$ and $\lambda = \log n$. We can describe a single run of the algorithm as a branching process: each degree-1 vertex created by the algorithm corresponds to a node that has children according to the number of degree-2 vertices adjacent to its neighbour; this corresponds to expected offspring number of $\delta_i \theta_j < 1 - \eta$. After each removal we are uniform over graphs with the remaining statistics; so, as long as the branching process size is $o(n)$, we can treat the treat these offspring distributions roughly independently for each node (in particular, as long as the branching process is $o(n)$ with high probability, all offspring distributions have expectation at most $1 - \frac{\eta}{2}$ regardless of the values for the other nodes). To prove that $\mathbb{P}\left[|Y_{t+1}^{(l)} - Y_t^{(l)}| > \frac{\sqrt{w}}{\lambda^2 \sqrt{\log n}} \Big| H_t\right] = o(n^{-3})$, we therefore just need to show that the probability that a Galton–Watson tree with $\mu < 1 - \frac{\eta}{2}$ reaches size $\frac{n^{0.25}}{(\log n)^{5/2}}$ is $o\left(\frac{1}{n^3}\right)$; this follows from a standard Chernoff bound.

2. The reason that $\mathbb{E}(Y_{t+1}^{(l)} - Y_t^{(l)}|H_t) = f_l(t/n, Y_t^{(1)}/n, \ldots, Y_t^{(a)}/n) + o(1)$ holds is because we expect a constant run duration, and we know our initial degree estimates will hold with small error terms when we remove a constant portion of the graph.

3. Continuity of f_l is clear from continuity of our degree estimates and the fact that small changes in the edge densities of the graph can't bias the branching process too heavily. Similar justification that it's Lipschitz on the given domain can be found by examining the degree estimate functions.

Since Y_0/n is with high probability $o(1)$ from having equal average degrees, and this ODE keeps equal average degrees equal, with high probability the degrees stay within $o(1)$ of equal until the stopping time.

B.3   Analysis of SHORTSIGHTED

Define the 2-D Markov Chain Z_t, where the first coordinate represented the number of unmatched R_0 vertices while the second coordinate represents the number of unmatched R_1 vertices at time t during a run of SHORTSIGHTED. We have that $Z_0 = (|R_0|, |R_1|)$. If we are in the regime where we prefer class 0, we have transition probabilities as follows:

$$\mathbb{P}(Z_{t+1} = (x-1, y)|Z_t = (x,y)) = \frac{1}{2}\left(1 - (1-p_{0,0})^x\right) + \frac{1}{2}\left(1 - (1-p_{1,0})^x\right)$$
$$\rightarrow \frac{1}{2}\left(1 - e^{-c_{0,0}x/n}\right) + \frac{1}{2}\left(1 - e^{-c_{1,0}x/n}\right)$$

$$\mathbb{P}(Z_{t+1} = (x, y-1)|Z_t = (x,y)) = \frac{1}{2}(1-p_{0,0})^x\left(1 - (1-p_{0,1})^y\right) + \frac{1}{2}(1-p_{1,0})^x\left(1 - (1-p_{1,1})^y\right)$$
$$\rightarrow \frac{1}{2}e^{-c_{0,0}x/n}\left(1 - e^{-c_{0,1}y/n}\right) + \frac{1}{2}e^{-c_{1,0}x/n}\left(1 - e^{-c_{1,1}y/n}\right)$$

$$\mathbb{P}(Z_{t+1} = (x, y)|Z_t = (x,y)) = \frac{1}{2}(1-p_{0,0})^x(1-p_{0,1})^y + \frac{1}{2}(1-p_{1,0})^x(1-p_{1,1})^y$$
$$\rightarrow \frac{1}{2}e^{-c_{0,0}x/n}e^{-c_{0,1}y/n} + \frac{1}{2}e^{-c_{1,0}x/n}e^{-c_{1,1}y/n}$$

We now verify the three conditions of Wormald's:

1. This simply comes from the fact that at each time step, each coordinate of Z_t can change by at most 1.

2. We use the following fact from [14]: for $n > 0$, $c \leq n/2$, and $x \in [0, 1]$, we have

$$\left| e^{-cx} - \left(1 - \frac{c}{n}\right)^{nx} \right| \leq \frac{c}{ne}$$

We can apply this term-wise to each of our probabilities, giving us the desired condition.

3. e^x is Lipschitz continuous on $[0, 1]$, therefore we have the third condition as well.

An essentially identical proof follows for the Markov chain where we prefer class 1. Therefore we may apply Wormald's theorem to treat these expect transitions as exact in the continuous limit, and solve to determine the matching size. Verification of these conditions is similar for GREEDY.

Lexicographic Unranking Algorithms for the Twelvefold Way

Amaury Curiel ✉ 🄳
Sorbonne Université, CNRS, LIP6 - UMR 7606, Paris, France

Antoine Genitrini ✉ 🄳
Sorbonne Université, CNRS, LIP6 - UMR 7606, Paris, France

──── **Abstract** ────

The Twelvefold Way represents Rota's classification, addressing the most fundamental enumeration problems and their associated combinatorial counting formulas. These distinct problems are connected to enumerating functions defined from a set of elements denoted by $\mathcal{N}$ into another one $\mathcal{K}$. The counting solutions for the twelve problems are well known. We are interested in unranking algorithms. Such an algorithm is based on an underlying total order on the set of structures we aim at constructing. By taking the rank of an object, i.e. its number according to the total order, the algorithm outputs the structure itself after having built it. One famous total order is the lexicographic order: it is probably the one that is the most used by people when one wants to order things. While the counting solutions for Rota's classification have been known for years it is interesting to note that three among the problems have yet no lexicographic unranking algorithm. In this paper we aim at providing algorithms for the last three cases that remain without such algorithms. After presenting in detail the solution for set partitions associated with the famous Stirling numbers of the second kind, we explicitly explain how to adapt the algorithm for the two remaining cases. Additionally, we propose a detailed and fine-grained complexity analysis based on the number of bitwise arithmetic operations.

1 Introduction

The Twelvefold Way, a classification from the 1960s by Rota, was introduced to address the most fundamental enumeration problems associated with their combinatorial counting formulas. It has been extensively discussed in Stanley's book [23, Section 1.9]. The distinct problems are related to the enumeration of functions defined from a set of elements denoted by $\mathcal{N}$ into another set denoted by $\mathcal{K}$. The respective cardinalities of these sets are denoted as n and k. Each set may consist of either distinguishable or indistinguishable elements, resulting in consideration of four pairs of sets. Additional constraints pertain to the properties of the functions, whether they are injective, surjective, or arbitrary. Consequently, we encounter twelve cases when enumerating these functions. The counting solutions are well-known, as presented in Stanley's book [23, Section 1.9]. In Table 1, we illustrate the classical combinatorial object enumerating each set of functions, in contrast to Stanley, who directly presents the counting solution.

■ **Table 1** The Twelvefold Way[*]

[*]The notation $k^{\underline{n}}$ corresponds to the product $k \cdot (k-1) \cdots (k-n+1)$; $[n \leq k]$ is the Iverson bracket returning 1 when $n \leq k$ and 0 otherwise; $\left\{ {\cdot \atop \cdot} \right\}$ and $\left({\cdot \atop \cdot} \right)$ stand respectively for the Stirling numbers of the second kind and the binomial coefficients; and $p_k(n)$ is the number of integer partitions of n into k positive integers.

elts of $\mathcal{N}$	elts of $\mathcal{K}$	f is arbitrary	f is injective	f is surjective
dist.	dist.	**1.** n-sequence in $\mathcal{K}$	**2.** n-permutation of $\mathcal{K}$	**3.** composition of $\mathcal{N}$ with k subsets
enumeration		k^n	$k^{\underline{n}}$	$k! \cdot \left\{ {n \atop k} \right\}$
lex. unranking		easy	[7, Section 5]	Section 4.1
indist.	dist.	**4.** n-multisubset of $\mathcal{K}$	**5.** n-subset of $\mathcal{K}$	**6.** composition of n with k terms
enumeration		$\binom{k+n-1}{n}$	$\binom{k}{n}$	$\binom{n-1}{n-k}$
lex. unranking		see survey [7] and references therein		
dist.	indist.	**7.** partition of $\mathcal{N}$ into $\leq k$ subsets	**8.** partition of $\mathcal{N}$ into $\leq k$ elements	**9.** partition of $\mathcal{N}$ into k subsets
enumeration		$\sum_{i=0}^{k} \left\{ {n \atop i} \right\}$	$[n \leq k]$	$\left\{ {n \atop k} \right\}$
lex. unranking		Section 4.2	easy	Theorem 11
indist.	indist.	**10.** partition of n into $\leq k$ parts	**11.** partition of n into $\leq k$ parts $\{1\}$	**12.** partition of n into k parts
enumeration		$p_k(n+k)$	$[n \leq k]$	$p_k(n)$
lex. unranking		[20, Section 4.8]	easy	[20, Section 4.8]

In this paper, our focus lies in the generation of these classical combinatorial objects. To initiate our exploration, we arrange each object within a given class in lexicographic order. Subsequently, given the rank of an object, our goal is to construct it directly. This process is referred to as a *lexicographic unranking algorithm*. For instance, among the six permutations of $\{1, 2, 3\}$, the first one (with rank 0 in lexicographic order) is $[1, 2, 3]$, followed by the second one (with rank 1), which is $[1, 3, 2]$, and so forth, culminating with the last one (rank 5) being $[3, 2, 1]$. Consequently, the lexicographic unranking algorithm for the function with rank 4 returns $[3, 1, 2]$. In Table 1, we provide references to such algorithms for 9 out of the 12 cases. However, for cases 3, 7, and 9, no knowledge about lexicographic unranking algorithms seems available in the literature. This paper introduces an approach to unranking in lexicographic order for the *set partitions of an n-set into k blocks* (case 9). Furthermore, we present extensions of this approach to address cases 3 and 7.

The problem of unranking objects emerges as one of the most fundamental challenge in combinatorial generation, as seen in [21], and is applicable in various domains such as software testing [17], optimization [9], or scheduling [24]. In different contexts, it serves as the core problem for generating complex structures, as observed in phylogenetics [2] and bioinformatics [1]. As mentioned earlier, to unrank, one must first establish a total order over the objects in question. The often-utilized order is the *lexicographic* order due to its ease of handling, leading to extensive study in the literature. However, Ruskey notes

in [20, p. 59] that lexicographic generation is typically not the most efficient, thus requiring particular care in lexicographic unranking. Knuth dedicates a section to the lexicographic generation of combinatorial objects in [11], relating it to the special case of Gray codes. Other combinatorial objects are also explored in Ruskey's and Kreher and Stinson's books on combinatorial generation [20, 12]. Skiena focuses on the practical implementation of such algorithms [22].

Usually, the approach for constructing structures using a recursive decomposition schema involves leveraging this decomposition to build a larger object from smaller ones. This method is extensively detailed in the well-known book by Nijenhuis and Wilf [18]. The approach has been systematically applied to decomposable objects in the context of analytic combinatorics, initially for recursive generation [6], and later for unranking methods [16].

Related work. Let us first quickly detail the classical unranking methods for the Twelvefold Way. As indicated in Table 1, cases 1, 8, and 11 are straightforward. In fact, an n-sequence in $\mathcal{K}$ consists of a word of length n over the finite alphabet $\mathcal{K}$, making lexicographic unranking direct. Cases 8 and 11 are extreme situations, both corresponding to the Iverson bracket $[n \leq k]$. As a result, the enumeration problems contain either one function (only when $n \leq k$) or none. The unranking method is trivial.

Cases 4, 5, and 6 are all associated with the enumeration of subsets and are directly related to combination enumerations. Various algorithms to solve such lexicographic unranking problems are relevant in the literature. In [7], we present a survey of the most efficient methods with a modern algorithm complexity analysis. Moreover, we introduce a new algorithm based on the factoradics number system, which is at least as efficient as the others.

Case 12 is associated with integer partition enumerations, and [12] presents an efficient recursive algorithm. This algorithm follows lexicographic order but for the reverse standard form of printing a partition. In standard form, partitions print the components from the largest to the smallest, whereas this algorithm is based on the reverse printing (from the smallest component to the largest one). It appears that, currently, there is no existing lexicographic unranking method specifically designed for the standard form of printing. case 10 can be considered an extension of case 12, much like case 7 is an extension of case 9.

The last three cases pertain to set partition problems. Various combinatorial objects, such as permutations with a specific pattern [4], graph coloring [10], walks in graphs [5], or trees for phylogenetics [2], are enumerated by set partitions. In a recent paper [14] the uniform random generating for set partitions for given n and k is studied, in the context of clustering algorithms. However, as far as we know, there is no lexicographic algorithm that takes arguments n, k and the rank r, returning the r-th partition in lexicographic order. Instead, there exists another classical object called a *restricted growth sequence* that is in bijection with set partitions (see [15, 20]). The unranking approaches presented in these works return such restricted growth sequences in lexicographic order. However, the natural bijection from restricted growth sequences to partitions does not preserve the lexicographic order.

Main results. To develop an efficient unranking generator for set partitions, we first introduce the lexicographic order over set partitions. Some care must be taken since we are dealing with sets of integers. Therefore, we use a standard printing of a set partition to obtain a canonical representation. We then introduce an *ad hoc* combinatorial algorithm to unrank set partitions in lexicographic order. Due to the very large integers manipulated in the algorithms, of order of $n \ln n$ bits, our algorithm computes the necessary ones on-the-fly in a lazy paradigm. The correctness and complexity of the algorithm are managed based

on specific combinatorial properties derived throughout the paper. Finally, we present some experiments using a Go[1] implementation for our algorithm. We leverage the simple and efficient parallelism mechanism provided by this language to significantly reduce space consumption without degrading time consumption for large values of n.

Organization of the paper.   Following the introduction of the paper, Section 2 highlights the combinatorial aspects of set partitions and presents some preliminary properties. In Section 3, we introduce our method for unranking set partitions, providing key insights into proving the correctness and complexity of our approach. Additionally, we present ideas for running calculations in parallel and share experiments that validate our parallel approach. Finally, Section 4 extends our algorithm to address cases 3 and 7 from the Twelvefold Way.

2   Preliminaries

2.1   Context of set partitions

▶ **Definition 1.** *Let $\mathcal{N}$ be a set of n distinguishable elements. A partition π of $\mathcal{N}$ in k blocks is a collection $B_1, B_2, \ldots, B_k$ of disjoint non-empty subsets of $\mathcal{N}$ such that every element from $\mathcal{N}$ belongs to exactly one B_i, for $i \in \{1, \ldots, k\}$.*

As an example let $\mathcal{N}$ be $\{1, \alpha, 2, 3, 4, \beta, 6, 12\}$. The collection $\{2, 3\}, \{4, 6, 12\}, \{\beta, 1, \alpha\}$ is a partition of $\mathcal{N}$ in 3 blocks. In the rest of for paper, the set of positive integers from 1 to n is denoted by $[\![n]\!]$. We can identify a set $\mathcal{N}$ of n elements with $[\![n]\!]$, thus from now we will only be interested in partitions for $[\![n]\!]$.

▶ **Definition 2.** *Let $1 \leq k \leq n$ be two positive integers and $\mathcal{N}$ be $[\![n]\!]$. The set of k-partitions of $\mathcal{N}$ is denoted by $\mathcal{P}_k^n$. The sequential form of a partition of $\mathcal{P}_k^n$ (i.e. a k-partitions of $\mathcal{N}$) is such that for all $i \in [\![k]\!]$, the block B_i contains the smallest integer from $[\![n]\!]$ not present in $\cup_{j<i} B_j$. Furthermore for each block, it is represented in the increasing order of its elements.*

For example $\{1\}, \{2, 3, 5\}, \{4, 6\}$ is a 3-partition of $[\![6]\!]$ represented in its sequential form. The sequential form is a canonical representation of the partition. As a shortcut, we will from now represent a partition simply as $1/235/46$. In the paper we have chosen to use the terminology and the notations from Mansour [15].

▶ **Fact.** *Let $1 \leq k \leq n$ be two positive integers. The number of partitions in $\mathcal{P}_k^n$ is the Stirling number of the second kind denoted by $\left\{{n \atop k}\right\}$. It satisfies the following recurrence:*

$$\left\{{n \atop k}\right\} = \begin{cases} \left\{{n-1 \atop k-1}\right\} + k \cdot \left\{{n-1 \atop k}\right\} & \text{if } 1 < k < n; \\ 1 & \text{otherwise, i.e. when } k = 1 \text{ or } k = n. \end{cases} \tag{1}$$

This sequence is stored in `OEIS A008277`[2]. We now introduce a natural order over k-partitions.

▶ **Fact.** *Let A and B two subsets of positive integers. We say that $A \leq B$ iff either*
- *$A = B$, or*
- *$A \subset B$ and $\max(A) < \min(B \setminus A)$, or*
- *$B \subset A$ and $\min(A \setminus B) < max(B)$, or*
- *$\min(A \setminus B) < min(B \setminus A)$.*

The relation $\leq$ is a total order over subsets of $[\![n]\!]$.

---

[1] The Go language offers routines to manage `concurrency`.
[2] OEIS stands for the On-line Encyclopedia of Integer Sequences.

For example $\{1,3\} \leq \{1,3,4\}$ and $\{1,3\} \leq \{1,4\}$. But we also have $\{1,3,4\} \leq \{1,4\}$.

▶ **Fact.** *Let $1 \leq k \leq n$ be two positive integers. The lexicographic order[3] over partitions from $\mathcal{P}_k^n$, in sequential form, is well defined using the latter order to compare two blocks: in fact a partition in k blocks is a Cartesian product of k subsets of positive integers.*

There is another classical representation for partitions called *canonical form* in [15]. A partition in k blocks is represented as a word over a k-letters alphabet. For example the partition $1/235/46$ is represented by the word 122323. The ith letter is the index of the block containing the integer i. Using this representation we can also define a lexicographic order over partitions, but here we compare partitions that do not necessarily contain the same numbers of blocks. The lexicographic order over the sequential form is not compatible with the lexicographic order used for the sequential form we are interested in. This can be noted in the Table 2.

▶ **Definition 3.** *Using the lexicographic order over the sequential form for partitions in $\mathcal{P}_k^n$, we define a ranking function assigning to each partition its rank corresponding to the number of k-partitions smaller than it in the lexicographic order.*

Table 2 Ranking of the 3-partitions of $[\![5]\!]$.

Rank	Partition	Canonical form [15]	Rank	Partition	Canonical form [15]
0	1/2/345	12333	13	13/2/45	12133
1	1/23/45	12233	14	13/24/5	12123
2	1/234/5	12223	15	13/25/4	12132
3	1/235/4	12232	16	134/2/5	12113
4	1/24/35	12323	17	135/2/4	12131
5	1/245/3	12322	18	14/2/35	12313
6	1/25/34	12332	19	14/23/5	12213
7	12/3/45	11233	20	14/25/3	12312
8	12/34/5	11223	21	145/2/3	12311
9	12/35/4	11232	22	15/2/34	12331
10	123/4/5	11123	23	15/23/4	12231
11	124/3/5	11213	24	15/24/3	12321
12	125/3/4	11231			

▶ **Definition 4.** *Let $1 \leq k \leq n$ be two positive integers. Let P be a partition from $\mathcal{P}_k^n$, represented in the sequential form as $B_1/B_2/\ldots/B_k$. An integer subset p is called* prefix *of P if $p \subset B_1$ and $p \leq B_1$.*

For the partition $12/35/4$, there are three possible prefixes $\emptyset, 1$ and 12. We can further extend the definition of prefixes of a partition by letting $\mathcal{N}$ being any subset of $[\![n]\!]$. Thus removing the first block of the latter partition gives $35/4$, we define prefixes of the 2-partition (of $\{3,4,5\}$) to be $\emptyset, 3$ and 35. Here we formalize this extension.

▶ **Definition 5.** *The definition of a prefix p of a partition is extended to any set $\mathcal{N}$ partitioned in a sequence of blocks (with the first one being denoted by B_1) such that $p \leq B_1$.*

---

[3] The lexicographic order of partitions from $\mathcal{P}_k^n$ in sequential form is a total order.

2.2 Combinatorial properties

We are now interested in counting results for partitions sharing the same prefix. These are the core results for our unranking algorithm.

▶ **Proposition 6.** *Let $1 \leq k \leq n$ be two positive integers. Let ℓ and d be two integers such that either $\ell = d = 1$ or $1 < \ell \leq d$. For a given prefix $1 \; \alpha_2 \; \alpha_3 \ldots \alpha_{\ell-1} \; d$, we define $S_k^n(\ell, d)$ to be the number of partitions in $\mathcal{P}_k^n$ accepting this prefix of length ℓ: We have*

$$S_k^n(\ell, d) = \sum_{u=0}^{\min(n-k-\ell+1,n-d)} \left\{ \begin{matrix} n-\ell-u \\ k-1 \end{matrix} \right\} \binom{n-d}{u}.$$

Let us remark that the notation $S_k^n(\ell, d)$ is a bit confusing in the sense that it is relative to the whole prefix $1 \; \alpha_2 \; \alpha_3 \ldots \alpha_{\ell-1} \; d$. However the specific values of α_2 up to $\alpha_{\ell-1}$ are not modifying the values of $S_k^n(\ell, d)$.

Proof. First if $\ell = d = 1$, then in the sequential form the first block necessarily contains 1. Thus $S_k^n(1, 1) = |\mathcal{P}_k^n| = \left\{ {n \atop k} \right\} = \sum_{u=0}^{n-k} \left\{ {n-1-u \atop k-1} \right\} \binom{n-1}{u}$. The latter equality is given e.g. in [8, p. 251, Table 251].

In the second case when $1 < \ell \leq d$, we aim at counting the number of partitions in $\mathcal{P}_k^n$ accepting $1 \; \alpha_2 \; \alpha_3 \ldots \alpha_{\ell-1} \; d$ as a prefix. In order to exhibit a combinatorial interpretation, we rewrite $S_k^n(\ell, d)$ as

$$S_k^n(\ell, d) = \sum_{u=0}^{\min((n-\ell)-(k-1),n-d)} \left\{ \begin{matrix} n-(\ell+u) \\ k-1 \end{matrix} \right\} \binom{n-d}{u}.$$

Once the prefix is given, it remains to complete the first block B_1 from the partition, and then to calculate how we can further partition the other elements in the next blocks. The variable u in the sum corresponds to the number of elements that are appended to the prefix to complete B_1. Its value ranges from 0 up to the maximal number of elements that we can append i.e. $(n-\ell) - (k-1)$ because at least $k-1$ among the remaining $n-\ell$ elements must be assigned to the other $k-1$ blocks. Obviously the number of possible elements u is also upper bounded by the number of remaining elements, i.e. $n-d$. Once the number u of elements for the completion of B_1 is given, we choose u elements greater than d: the number of possibilities is given by the binomial coefficient. Finally it remains to build the other blocks of the partition: we partition $n - (\ell + u)$ elements into $k-1$ blocks. Hence the formula is proved. ◀

We introduce an example using Table 2 for $\mathcal{P}_3^5$. If we are interested in the prefix 13, then there are 3 partitions without completing block B_1, in the sum, when $u = 0$ we get $\left\{ {3 \atop 2} \right\} = 3$. The other possible value is $u = 1$ with the general term being $\left\{ {2 \atop 2} \right\} \binom{2}{1} = 2$ as it appears in the table.

In order to get a formula that is more efficient to calculate, we observe that the latter numbers $S_k^n(\ell, d)$ depend essentially in three variables instead of four. The proof is direct with some variable renaming.

▶ **Proposition 7.** *Let n, k, d be integers with $0 \leq k \leq n$ and $0 \leq d \leq n$. By defining*

$$\tilde{S}_k^n(d) = \sum_{u=0}^{\min(n-k,n-d)} \left\{ \begin{matrix} n-u \\ k \end{matrix} \right\} \binom{n-d}{u}, \; \text{we get } S_k^n(\ell, d) = \tilde{S}_{k-1}^{n-\ell}(d - \ell).$$

We note that $S_k^n(1,1) = \tilde{S}_{k-1}^{n-1}(0) = \left\{{n \atop k}\right\}$. Note that the 3-dimension sequence $\tilde{S}$ seems not to be stored in OEIS. There exit several generalizations of Stirling numbers, but none of them apparently corresponds to our sequence $\tilde{S}$.

▶ **Corollary 8.** *The numbers $\tilde{S}_k^n(d)$ satisfy the following recurrence:*

$$\tilde{S}_k^n(d) = \begin{cases} \tilde{S}_{k-1}^{n-1}(d-1) + k \cdot \tilde{S}_k^{n-1}(d-1) & \text{if } 1 \le k \le n \text{ and } 1 \le d \le n; \\ \left\{{n+1 \atop k+1}\right\} & \text{if } d = 0 \text{ and } 0 \le k \le n; \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

Note the later recurrence is similar to the one satisfied by Stirling numbers of the second kind (but with here a third variable d giving some kind of level of numbers).

▶ **Proposition 9.** *Let n, k, d be integers with $0 \le k \le n$ and $0 \le d \le n$. The function $\tilde{S}_k^n(d)$ can be represented as a binomial transform:*

$$\tilde{S}_k^n(d) = \sum_{u=0}^{\min(n-k,d)} (-1)^u \left\{{n+1-u \atop k+1}\right\} \binom{d}{u}.$$

The main idea of the proof consists in proving that the two expressions given in Propositions 7 and 9 are satisfying the same recurrence and thus are equal.

Proof. In order to prove this new expression for $\tilde{S}$, we just have to prove that this expression satisfy the recurrence stated in Corollary 8. Substituting d by 0 we get the base case. We now consider the case where the three integers n, k, d satisfy $0 \le k \le n$ and $1 \le d \le n$. Using Proposition 9 in the case where $0 < k < n$ (the cases $k = 0$ or $k = n$ are obvious) we have

$$\tilde{S}_{k-1}^{n-1}(d-1) + k \cdot \tilde{S}_k^{n-1}(d-1) = \sum_{u=0}^{\min(n-k,d-1)} (-1)^u \left\{{n-u \atop k}\right\} \binom{d-1}{u}$$
$$+ k \cdot \sum_{u=0}^{\min(n-1-k,d-1)} (-1)^u \left\{{n-u \atop k+1}\right\} \binom{d-1}{u}$$

By using factorization and Stirling numbers of the second kind recurrence, we obtain:

$$\tilde{S}_{k-1}^{n-1}(d-1) + k \cdot \tilde{S}_k^{n-1}(d-1) = \sum_{u=0}^{\min(n-k,d-1)} (-1)^u \left(\left\{{n+1-u \atop k+1}\right\} - \left\{{n-u \atop k+1}\right\} \right) \binom{d-1}{u}.$$

After having telescoped the two sums we get the stated result. ◀

Finally, given two prefixes, one being smaller than the second one, the next proposition allows to compute how many partitions are in-between the two prefixes. More formally:

▶ **Proposition 10.** *Let $1 \le k \le n$ be two positive integers. Let $d_1 \in [\![n]\!] \setminus \{1\}$, $d_0 \in [\![d_1 - 1]\!]$ and $\ell > 1$ be integers. For a given prefix $1 \; \alpha_2 \; \alpha_3 \ldots \alpha_{\ell-2} \; d_0$, the number of elements of $\mathcal{P}_k^n$ that admit a length-ℓ prefix satisfying $1 \; \alpha_2 \ldots \alpha_{\ell-2} \; d_0 \; \tilde{d}_1$ (for all $\tilde{d}_1$ ranging from $d_0 + 1$ to d_1) is given by*

$$R_k^n(\ell, d_0, d_1) = \tilde{S}_{k-1}^{n-\ell}(d_0 - \ell) - \tilde{S}_{k-1}^{n-\ell}(d_1 + 1 - \ell).$$

<div style="border-left:4px solid gold; padding-left:8px">

3   Methods for unranking set partitions

</div>

Merging the combinatorial properties stated in the previous section, we are now ready to design algorithms to unrank set partitions in the lexicographic order.

3.1   Unranking algorithm design

Our aim consists in constructing the r-th partition related to a pair n, k in sequential form for the lexicographic order. The constructions follow the next main lines. The global idea consists in building the normalization of the partition. So we build together its block pattern and its reversed factoradics (seeing the partition as a size-n permutation).

- The building of the blocks is going from left to right;
- The construction of a block is also from left to right, component by component using a binary search approach;
- Finally once the block pattern and the reversed factoradics are set, a slight adaptation of the lexicographic permutation unranking algorithm gives the result.

The details for the correctness of our approach lies on the ranking function associated to the set partitions.

We first present in detail the main function UNRANKING of Algorithm 1. Using a loop, at each turn it defines the next block of the partition and then refine the value of the rank related to the remaining part of the partition. The result B returned by NEXT_BLOCK contains the indices of the components of the block that has been calculated and acc allows to update the rank so that it is related to the remaining part of the partition that must still be computed. With our previous definition, B is the normalization of the corresponding partition block. At the end of the function a dynamic extraction is executed in an array containing elements from 1 to n according to the indices in Res.

<div style="border-left:4px solid gold; padding-left:8px">

Algorithm 1 Lexicographic unranking of the partition with rank r in $\mathcal{P}_k^n$.

</div>

```
1: function UNRANKING(n, k, r)
2:     n' := n
3:     Res := [ ]
4:     while k > 1 do
5:         (B, acc) := NEXT_BLOCK(n, k, r)
6:         Append(Res, B)
7:         r := r − acc
8:         n := n − len(B)
9:         k := k − 1
10:    Append(Res, [0, 0, . . . , 0])
11:    Res := EXTRACT(n', Res)
12:    return Res
```

```
1: function EXTRACT(n, R)
2:     L := [1, 2, . . . , n]
3:     P := [ ]
4:     for r in R do
5:         p := [ ]
6:         for i in r do
7:             Append(p, L[i])
8:             Remove(L, i)
9:         Append(P, p)
10:    return P
```

Remove(L, i) removes element with index i in L.

```
1: function NEXT_BLOCK(n, k, r)
2:     Block := [0]; acc := { n−1 }
                               { k−1 }
3:     if r < acc then
4:         return (Block, 0)
5:     d0 := 1; index := 2; inf := 2; sup := n
6:     complete := False
7:     while not complete do
8:         while inf < sup do
9:             mid := ⌊(inf + sup)/2⌋
10:            if r >= acc + R_k^n(index−1, d0, mid−1) then
11:                inf := mid + 1
12:            else
13:                sup := mid
14:        mid := inf; threshhold := { n−index }
                                     {  k−1   }
15:        acc := acc + R_k^n(index − 1, d0, mid − 2)
16:        Append(Block, mid − index)
17:        if r < threshhold + acc then
18:            complete := True
19:        else
20:            index := index + 1
21:            d0 := mid; inf := d0 + 1; sup := n
22:            acc := acc + threshhold
23:    return (Block, acc)
```

The function NEXT_BLOCK takes parameters n, k, r and returns essentially the first block of the r-th partition in $\mathcal{P}_k^n$. In fact, using Table 2 the call NEXT_BLOCK$(5, 3, 16)$ returns 0 1 1 (instead of 1 3 4), the latter block being obtained through a dynamic extraction of the element 0 in $[1, 2, 3, 4, 5]$ then the element 1 is extracted in the remaining part $[2, 3, 4, 5]$ and finally the element 1 in $[2, 4, 5]$. Constructing the blocks of indices instead of the blocks of values allows to neglect about the remaining elements for the further blocks construction. Note that obviously the last block of the partition contains only the indices 0 (Line 10 from UNRANKING Algorithm) and the first element of a block is always index 0, both due to the sequential form. Finally while calling UNRANKING$(5, 3, 16)$, at the end of Line 10, *Res* contains $[[0\ 1\ 1][0][0]]$. Reading the components from right to left we get the factoradics 0 0 1 1 0 of the number 8 corresponding to the lexicographic rank of the permutation $[1, 3, 4, 2, 5]$ (*cf.* [7] for details).

▶ **Theorem 11.** UNRANKING$(n, k, \cdot)$ *is a lexicographic unranking algorithm for set partitions from* $\mathcal{P}_k^n$.

Proof key-ideas. The core of the function NEXT_BLOCK relies on the `while` loop (line 7). When it is entered (let say for the i-th time), the variable `Block` contains the length-i prefix of the normalized (final) block. Thus at this evaluation the loop determines (with a binary search) the $i + 1$th element that is appended to `Block`. Then, we calculate if the block is finished (line 17) or if we continue (line 19). ◀

3.2 Complexity analysis and experiments for unranking

In our implementation in Go[4], we offer two approaches for the necessary Stirling numbers calculations: either a precomputation of them or a computation on the fly of those that are needed at each step. We never precompute the 3 dimension table $\tilde{S}_k^n(d)$. In fact, in many bad cases these numbers are of order of $n!$, thus precomputing would be too expensive while only few of the numbers are needed. We compute the necessary numbers $\tilde{S}_k^n(d)$ on the fly.

First let us recall the behavior of the sequence of Stirling numbers of the second kind when k is ranging from 1 to n.

▶ **Fact.** *Let* $1 \leq k \leq n$ *be two positive integers. The sequence* $\left\{ {n \atop k} \right\}$ *is unimodal and its maximum is reached when* $k_n \sim n / \ln n$. *Around this value, we have* $\log \left\{ {n \atop k_n} \right\} = \Theta(n \log n)$[5]. *Furthermore, we have an upper bound valid for all* $1 \leq k \leq n$:

$$\log \left\{ {n \atop k} \right\} \leq (n - k) \log k + \log \binom{n}{k} \leq (n - k) \log k + k \log \left(\frac{n \cdot e}{k} \right).$$

See the fundamental paper of Rennie and Dobson [19] to get a proof for these results.

In the following we propose six distinct implementations of the function $\tilde{S}$ presented in Proposition 7 and underlying the function R from Proposition 10.

S_v1 direct implementation of the formula stated in Proposition 7;

S_v2 implementation of the formula from Proposition 7 taking into account the symmetry of binomial coefficients, thus the sum contains half of the terms in comparison to S_v1 (and thus half of the multiplications);

S_v3 direct implementation of the formula stated in Proposition 9;

---

[4] Go implementation and the material used for repeating the experiments are all available `here`.
[5] In this paper we use the notation log for the logarithm in basis 2.

S_v4 implementation of the formula from Proposition 9 taking into account the symmetry of binomial coefficients, thus the sum contains half of the terms in comparison to S_v3;

S_v5 is our most efficient algorithm without precomputations. The calculation way consists in deciding according whether a call to S_v2 or to S_v4 should be the most efficient, according to the number of terms in the sums interacting with Propositions 7 and 9;

S_v6 same algorithm than S_v5 but with all necessary Stirling numbers of the second kind precomputed.

The integers computed during the unranking algorithm are very large, thus a classical complexity in the number of arithmetical operations is not precise. We hence are interested in the bit-complexity, corresponding the the number of atomic operations on digits.

▶ **Theorem 12.** *For the time complexity, the algorithm* Unranking *based on each of the function S_v· has a bit-complexity bounded by*

$$O\left(\frac{(n-k)^3 M(n)}{n} \ln n \log k + \frac{k(n-k)^2 M(n)}{n} \ln n \log\left(\frac{n \cdot e}{k}\right)\right),$$

where $M(n)$ is the bit complexity for the multiplication of two numbers, each one containing n bits.

The naïve multiplication algorithm satisfies $M(n) = \Theta(n^2)$. But using, for example, Karatsuba algorithm, we obtain $M(n) = \Theta(n^{\log 3})$ for the time complexity. In Go[6], as soon as the integers are greater than 2^{40}, Karatsuba multiplication algorithm is used. In our context, almost all cases are thus based on the latter algorithm.

Proof. We are interested in a worst case complexity analysis when n is large and for k ranging in $[\![n]\!]$. We are using the same kind of analysis in bit complexity as the one presented in [7, Section 4.3]. We compute an upper bound of the complexity in the central range of the Stirling numbers of the second kind. The central range, when n tends to infinity, is observed when $k = \Theta(n/\ln n)$. A detailed similar analysis is presented in the paper [13]. In our context each Stirling number necessitates $\log\left\{{n \atop k}\right\}$ bits to be stored. They are multiplied by binomial coefficients containing at most n bits. Thus Stirling numbers are separated in blocks of n bits in order to use a multiplication of similar sizes numbers, inducing a time complexity bounded by $\log\left\{{n \atop k}\right\}/n \cdot M(n)$. Furthermore the number of calls the the function $\tilde{S}$ is $O((n-k)\ln n)$ induced by the repetitive calls to the binary search algorithm. Compiling all these upper bounds gives the stated bit-complexity.   ◀

For approach S_v6, the following result establishes that the precomputation is negligible in terms of time complexity compared to the unranking itself. However note that the memory complexity is quadratic instead of linear (in n) by using the precomputation step.

▶ **Proposition 13.** *The bit-complexity for the Stirling numbers precomputation is bounded by*

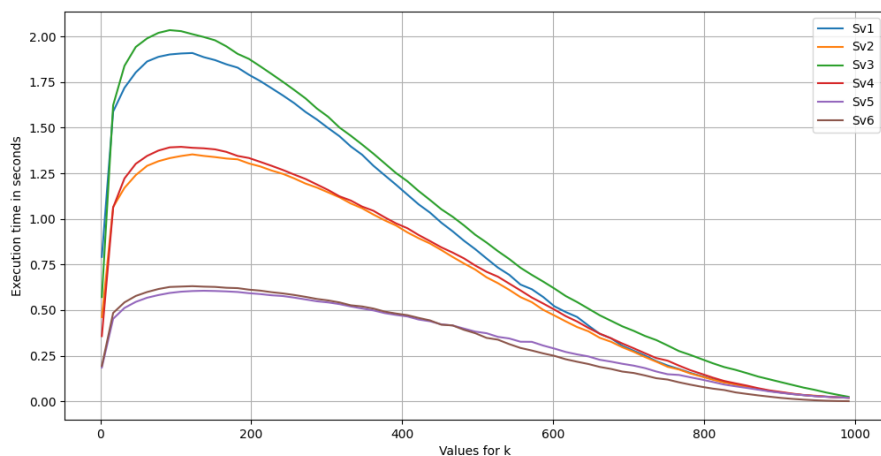$$O\left(k(n-k)^2 M(\log k) + k^2(n-k)\frac{\log n}{\log k} M(\log k)\right).$$

In order to get the Stirling numbers on the fly, we use parallel compuatations. In fact, for each block determination, we observe that only two neighbors columns from the triangle of numbers are needed. Thus during the determination of a block, we compute in parallel the

---

[6] Go documentation for `big integers manipulations`.

Figure 1 Time (in seconds) for unranking a partition in $\mathcal{P}_k^{1000}$ when k is ranging in $[\![1000]\!]$.

next two columns that will be necessary for the next block. Computing column $n-1$ from column n costs less time than the unranking algorithm. Thus, with parallel computation of the Stirling numbers, we achieve the same time complexity as the algorithm where Stirling numbers are precomputed while consuming a $O(n)$ amount of memory thus needing $O(n^2)$ memory size.

In Figure 1, we run experiments[7] by fixing $n = 1000$ and k ranging from 2 to 992 with steps of 15 units. For each value of k, 500 uniform samples are computed and the average time for the building of the partition is drawn for each Algorithm S_v1 up to S_v6. Obviously Algorithms S_v2 and S_v4 are better than their naïve versions respectively S_v1 and S_v2. It is interesting to note that the optimization S_v5, obtained by computing the most efficient formula between Propositions 7 and 9. Finally we remark that the Algorithm S_v5 is almost as efficient as S_v6 where all precomputation of Stirling numbers have been stored before the computation of the partition. Strangely, for the smallest values of k, we note that S_v5 is even faster than S_v6. This is probably due to the RAM accesses: in fact in some preliminary experiments with computers equipped with DDR5 RAM Algorithm S_v6 is always faster than S_v5, and this is what is expected.

4 Extension and conclusion

As we observe in Table 1, both enumeration cases 3 and 7 are some extended version of the enumeration case 9. An adaptation for the RANKING function allows to rank the families counted by cases 3 and 7; then adapting the unranking algorithm solves these cases.

4.1 Ordered set partitions

Recall Stirling numbers of the second kind are counting the numbers of surjective functions from set $\mathcal{N}$ to set $\mathcal{K}$, where the elements of $\mathcal{N}$ are distinguishable and those of $\mathcal{K}$ are indistinguishable. We can represent these functions as set partitions. Now, what happens

---

[7] The experiments provided in this paper are driven by using a PC equipped with an Intel Xeon X5677 processor, 32GB of DDR4-SDRAM and running Debian GNU/Linux 12.

when elements of $\mathcal{K}$ are distinguished? These functions are counted by *ordered Stirling numbers of the second kind.* In addition, they can be represented as *ordered set partitions,* which are similar to set partitions except that the order of the subsets matters. For instance, while in the world of unordered set partitions, elements $14/25/3; 14/3/25; 25/14/3; 25/3/14; 3/14/25$ and $3/25/14$ are equivalent and represented by the partition $14/25/3$ in sequential form, in the world of ordered set partition, the 6 elements are all different.

▶ **Proposition 14.** *Let $1 \le k \le n$, be two integers with n being the cardinality of set $\mathcal{N}$. The number of ordered set partitions of $\mathcal{N}$ in k (non empty disjoint) subsets is $k! \cdot \left\{ {n \atop k} \right\}$. The family of these partitions is denoted by $\mathcal{O}_k^n$.*

The proof is direct: the blocks in the sequential form of a set partition are distinguishable, thus permuting them gives the associated ordered set partitions.

▶ **Fact.** *Let $1 \le k \le n$, be two integers, the lexicographic order on set partitions $\mathcal{P}_k^n$ is easily extended to get the lexicographic order for the ordered set partitions from $\mathcal{O}_k^n$.*

We can now derive the enumeration core result in this new context.

▶ **Proposition 15.** *Let $1 \le k \le n$ be two integers. Let ℓ and d be two integers such that either $\ell = d = 1$ or $1 < \ell \le d$. For a given prefix $1\ \alpha_2\ \alpha_3 \ldots \alpha_{\ell-1}\ d$, we define $T_k^n(\ell, d)$ to be the number of ordered set partitions in $\mathcal{O}_k^n$ accepting this prefix.*

$$T_k^n(\ell, d) = \sum_{u=0}^{\min(n-k-\ell+1, n-d)} k! \left\{ {n - \ell - u \atop k - 1} \right\} \binom{n - d}{u}.$$

This formula is the analogous to $S_k^n(\ell, d)$. Using the same variable changes, we also get a three variable function, like $\tilde{S}_k^n(d)$. Then we can deduce an adaptation of our first algorithm by replacing Stirling numbers of the second kind by ordered Stirling numbers of the second kind and using the latter formula.

4.2 Bell's set partitions

We denote by $\mathcal{F}$ the family of these functions. Such functions can be represented as unordered set partitions with at most k blocks where k is the numbers of elements in $\mathcal{K}$.

Let $\mathcal{K}_i \subset \mathcal{K}$ be a subset of i distinguishable elements and $\mathcal{B}_i$ the functions that are surjective from $\mathcal{N}$ into $\mathcal{K}_i$. We have $\mathcal{B}_k^n = \bigcup_{i=1}^k \mathcal{B}_i$ and for a given $i \in [\![k]\!]$, $|\mathcal{B}_i| = \left\{ {n \atop i} \right\}$. Obviously $|\mathcal{B}_k^n| = \sum_{i=0}^k |\mathcal{B}_i| = \sum_{i=0}^k \left\{ {n \atop i} \right\}$. The cardinality of $\mathcal{B}_k^n$ is counted by the *k-restricted Bell numbers* and finally, when $k = n$, we get the Bell numbers.

▶ **Fact.** *Let $1 \le k \le n$, be two integers, the lexicographic order on set partitions $\mathcal{P}_k^n$ is also lexicographic for $\mathcal{B}_k^n$.*

▶ **Proposition 16.** *Let $1 \le k \le n$ be two integers. Let ℓ and d be two integers such that either $\ell = d = 1$ or $1 < \ell \le d$. For a given prefix $1\ \alpha_2\ \alpha_3 \ldots \alpha_{\ell-1}\ d$, we define $U_k^n(\ell, d)$ to be the number of Bell's set partitions in $\mathcal{B}_k^n$ accepting of this prefix.*

$$U_k^n(\ell, d) = \sum_{u=0}^{\min(n-k-\ell+1, n-d)} \sum_{i=1}^k \left\{ {n - \ell - u \atop i - 1} \right\} \binom{n - d}{u}.$$

As a final remark, the correctness of both previous algorithms is directly hanging to the one for the set partition algorithm. What is remaining is their complexity analysis: it is not difficult, and it will be written in a long version of this paper.

---- **References** ----

1   N. Ali, A. Shamoon, N. Yadav, and T. Sharma. Peptide Combination Generator: a Tool
    for Generating Peptide Combinations. *ACS Omega*, 5(11):5781–5783, 2020. `doi:10.1021/`
    `acsomega.9b03848`.

2   O. Bodini, A. Genitrini, C. Mailler, and M. Naima. Strict monotonic trees arising from
    evolutionary processes: Combinatorial and probabilistic study. *Advances in Applied Math.*,
    133:102284, 2022. `doi:10.1016/J.AAM.2021.102284`.

3   Amaury Curiel and Antoine Genitrini. Set Partition Unranking. Software, swhId:
    `swh:1:dir:b01a69e78b0fd972fdafc0c080421688cd9c9be6` (visited on 2024-07-09). URL:
    `https://github.com/AMAURYCU/setpartition_unrank`.

4   C. Defant. Highly sorted permutations and Bell numbers. *Enumerative Combinatoricas and
    Applications*, 1(1), 2021. `doi:10.48550/arXiv.2012.03869`.

5   A. Dzhumadil'daev and D. Yeliussizov. Walks, partitions, and normal ordering. *Electronic
    Journal of Combinatorics*, 22(4):4, 2015. `doi:10.37236/5181`.

6   P. Flajolet, P. Zimmermann, and B. Van Cutsem. A calculus for the random generation
    of labelled combinatorial structures. *Theoretical Computer Science*, 132(1-2):1–35, 1994.
    `doi:10.1016/0304-3975(94)90226-7`.

7   A. Genitrini and M. Pépin. Lexicographic Unranking of Combinations Revisited. *Algorithms*,
    14(3), 2021. `doi:10.3390/A14030097`.

8   R. L. Graham, D. E. Knuth, and O. Patashnik. *Concrete Mathematics: A Foundation
    for Computer Science*. Addison-Wesley, Reading, MA, second edition, 1989. URL: `https:`
    `//www-cs-faculty.stanford.edu/%7Eknuth/gkp.html`.

9   I. Grebennik and O. Lytvynenko. Developing software for solving some combinatorial generation
    and optimization problems. In *7th Int. Conf. on Application of Information and Communication
    Technology and Statistics in Economy and Education*, pages 135–143, 2017. URL: `http:`
    `//openarchive.nure.ua/handle/document/5498`.

10 A. Hertz and H. Mélot. Counting the number of non-equivalent vertex colorings of a graph.
    Discrete Applied Mathematics, 203:62–71, 2016. `doi:10.1016/J.DAM.2015.09.015`.

11 D. E. Knuth. *The Art of Computer Programming, Volume 4A, Combinatorial Algorithms*.
    Addison-Wesley Professional, 2011. URL: `https://www-cs-faculty.stanford.edu/~knuth/`
    `taocp.html`.

12 D. L. Kreher and D. R. Stinson. *Combinatorial Algorithms: generation, enumeration, and
    search*. CRC Press, 1999. `doi:10.1145/309739.309744`.

13 G. Louchard. Asymptotics of the Stirling numbers of the second kind revisited. *Applicable
    Analysis and Discrete Mathematics*, 7(2):193–210, 2013. `doi:10.46298/dmtcs.501`.

14 Q. Lutz, E. de Panafieu, M. Stein, and A. Scott. Active clustering for la-
    beling training data. In *Advances in Neural Information Processing Systems 34:
    Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021*,
    pages 8469–8480, 2021. URL: `https://proceedings.neurips.cc/paper/2021/hash/`
    `47841cc9e552bd5c40164db7073b817b-Abstract.html`.

15 T. Mansour. *Combinatorics of Set Partitions*. Discrete Mathematics and Its Applications.
    Taylor & Francis, 2012. `doi:10.1201/b12691`.

16 C. Martínez and X. Molinero. A generic approach for the unranking of labeled combinatorial
    classes. *Random Structures & Algorithms*, 19(3-4):472–497, 2001. `doi:10.1002/RSA.10025`.

17 A. F. Myers. *k*-out-of-*n*:G System Reliability With Imperfect Fault Coverage. *IEEE Transac-
    tions on Reliability*, 56(3):464–473, 2007. `doi:10.1109/TR.2007.903229`.

18 A. Nijenhuis and H. S. Wilf. *Combinatorial algorithms*. Computer science and applied
    mathematics. Academic Press, New York, NY, 1975. URL: `https://www2.math.upenn.edu/`
    `~wilf/website/CombAlgDownld.html`.

19 B.C. Rennie and A.J. Dobson. On stirling numbers of the second kind. *Journal of Combinatorial
    Theory*, 7(2):116–121, 1969. `doi:10.1016/S0021-9800(69)80045-1`.

20 F. Ruskey. Combinatorial Generation, 2003.

21   Y. Shablya, D. Kruchinin, and V. Kruchinin. Method for Developing Combinatorial Generation Algorithms Based on AND/OR Trees and Its Application. *Mathematics*, 8(6):962, 2020. `doi:10.3390/math8060962`.

22   S. Skiena. *The Algorithm Design Manual, Third Edition*. Texts in Computer Science. Springer, 2020.

23   R. P. Stanley. *Enumerative Combinatorics: Volume 1*. Cambridge Univ. Press, 2011. `doi:10.1017/CBO9781139058520`.

24   Y. Tamada, S. Imoto, and S. Miyano. Parallel Algorithm for Learning Optimal Bayesian Network Structure. *J. Mach. Learn. Res.*, 12:2437–2459, 2011. `doi:10.5555/1953048.2021080`.

Periodic Behavior of the Minimal Colijn-Plazzotta Rank for Trees with a Fixed Number of Leaves

Michael R. Doboli [1] ✉ iD
Department of Mathematics, Stanford University, CA, USA

Hsien-Kuei Hwang ✉ iD
Institute of Statistical Science, Academia Sinica, Taipei, Taiwan

Noah A. Rosenberg ✉ iD
Department of Biology, Stanford University, CA, USA

───── **Abstract** ─────

The Colijn-Plazzotta ranking is a certain bijection between the unlabeled binary rooted trees and the positive integers, such that the integer associated with a tree is determined from the integers associated with the two immediate subtrees of its root. Letting a_n denote the minimal Colijn-Plazzotta rank among all trees with a specified number of leaves n, the sequence $\{a_n\}$ begins 1, 2, 3, 4, 6, 7, 10, 11, 20, 22, 28, 29, 53, 56, 66, 67 (OEIS A354970). Here we show that $a_n \sim 2[2^{P(\log_2 n)}]^n$, where P varies as a periodic function dependent on $\{\log_2 n\}$ and satisfies $1.24602 < 2^{P(\log_2 n)} < 1.33429$.

1 Introduction

Consider an unlabeled binary rooted tree t with $m(t)$ leaves. Colijn & Plazzotta [2] introduced a ranking for the unlabeled binary rooted trees, according to which the rank $f(t)$ of t is determined from the ranks $\ell(t)$ of its left subtree and $r(t)$ of its right subtree: $f(t) = 1$ for $m(t) = 1$, and $f(t) = f(\ell(t))[f(\ell(t)) - 1]/2 + 1 + f(r(t))$ for $m(t) \geq 2$. To compute the Colijn-Plazzotta rank, or CP rank, of a tree t, the "left" and "right" subtrees of t are arranged in a canonical order, such that $f(\ell(t)) \geq f(r(t))$.

The ranking f bijectively associates positive integers to unlabeled binary rooted trees – which number 1, 1, 2, 3, 6, 11, 23, 46, 98 for trees of $n = 1$ to 10 leaves (the Wedderburn-Etherington numbers, OEIS A001190). Among trees with n leaves, CP ranks vary greatly; for example, the 8-leaf symmetric tree has rank 11 and the 8-leaf caterpillar has rank 2,598,062.

The CP rank has been proposed for various uses in the mathematical study of evolutionary trees [2, 3, 9]. It provides a tree encoding with the property that similar shapes often have nearby ranks, even if they possess different numbers of leaves. As a result, it gives a basis for computing a distance between unlabeled trees of differing size – a useful metric for the evolutionary trees that might be produced from genetic sequences in pathogens and other organisms. Because highly balanced shapes have the smallest rank among trees with a fixed

---

[1] Corresponding author.

number of leaves and highly unbalanced shapes have the largest rank, the CP rank can serve as a measure of the balance of an unlabeled tree – for example, in studies that seek to compare the balance of observed trees to that predicted by models of evolutionary processes.

The minimal and maximal CP ranks across all trees with a fixed number of leaves n can assist in assessing the CP ranks of specific trees, for example by normalizing the CP rank as a measure of tree balance. Rosenberg [9] studied the minimal and maximal CP ranks across trees with n leaves, identifying the trees that give rise to those ranks. The maximal rank, denoted b_n, recursively follows $b_n = b_{n-1}(b_{n-1} - 1)/2 + 2$ for $n \geq 2$, with $b_1 = 1$ [9, Theorem 9]. As a quadratic recursion of a form studied by Aho & Sloane [1], b_n has asymptotic growth $b_n \sim 2\beta^{2^n}$ for a constant $\beta \approx 1.05653$ [9, Corollary 14].

The minimal CP rank, denoted a_n, recursively follows [9, Theorem 6]

$$a_n = \begin{cases} 1, & n = 1 \\ a_{\lceil n/2 \rceil}(a_{\lceil n/2 \rceil} - 1)/2 + 1 + a_{\lfloor n/2 \rfloor}, & n \geq 2. \end{cases} \tag{1}$$

For n equal to a power of 2, Rosenberg [9] showed that the recursion for a_n is related to that for b_n, producing $a_n \sim 2\alpha^n$ for a constant $\alpha = \beta^4$, $\alpha \approx 1.24602$ [9, Theorem 13]. For general n, however, Rosenberg [9] gave only an upper bound, $a_n < (\frac{3}{2})^n$ [9, Proposition 15].

Here we obtain the asymptotic growth of a_n. Informally, our main result, obtained in Theorems 6 and 9 and summarized in Corollary 10, states that the minimal Colijn-Plazzotta rank a_n across trees with n leaves is approximately equal to $2[2^{P(\log_2 n)}]^n$, where P is a bounded periodic function of period 1. Moreover, the minimum and supremum of 2^P are given by constants $c_1 \approx 1.24602$ and $c_2 \approx 1.33429$.

Extremal properties of the non-differentiable periodic functions arising from recursions such as eq. 1 that involve $\lfloor n/2 \rfloor$ and $\lceil n/2 \rceil$ are often difficult to characterize; many examples therefore rely on case-dependent approaches. The computation here uses an inductive method for studying the extrema.

2   An elementary improvement to the bounds on a_n

We begin by providing a refined exponential upper and lower bound on a_n for $n \geq 66$ that improves upon the $(\frac{3}{2})^n$ exponential upper bound in [9].

▶ **Proposition 1.** *For all integers $n \geq 66$, $3(1.2)^n < a_n < (1.34)^n$.*

Proof. We proceed via induction on n. For the base case, we verify computationally from eq. 1 that $3(1.2)^n < a_n < (1.34)^n$ for all integers $66 \leq n \leq 132$. For the inductive hypothesis, assume that $3(1.2)^k < a_k < (1.34)^k$ for all k, $66 \leq k < n$. Because we have already considered $66 \leq n \leq 132$, suppose $n > 132$. Writing $a_n = \frac{1}{2}a_{\lceil n/2 \rceil}(a_{\lceil n/2 \rceil} - 1) + a_{\lfloor n/2 \rfloor} + 1$, by the inductive hypothesis, we have for the lower bound

$$\begin{aligned} a_n &> \frac{3(1.2)^{\lceil n/2 \rceil}[3(1.2)^{\lceil n/2 \rceil} - 1]}{2} + 3(1.2)^{\lfloor n/2 \rfloor} + 1 \\ &= \frac{9}{2}(1.2)^{2\lceil n/2 \rceil} - \frac{3}{2}(1.2)^{\lceil n/2 \rceil} + 3(1.2)^{\lfloor n/2 \rfloor} + 1 \\ &\geq \frac{9}{2}(1.2)^{2\lceil n/2 \rceil} + \left[-\frac{3}{2}(1.2) + 3\right](1.2)^{\lfloor n/2 \rfloor} + 1 > \frac{9}{2}(1.2)^n + 1.2(1.2)^{\lfloor n/2 \rfloor} + 1 > 3(1.2)^n. \end{aligned}$$

For the upper bound, by the recursive formula for a_n and the inductive hypothesis,

$$
\begin{aligned}
a_n &< \frac{(1.34)^{\lceil n/2 \rceil}[(1.34)^{\lceil n/2 \rceil} - 1]}{2} + (1.34)^{\lfloor n/2 \rfloor} + 1 \\
&= \frac{1}{2}(1.34)^{2\lceil n/2 \rceil} - \frac{1}{2}(1.34)^{\lceil n/2 \rceil} + (1.34)^{\lfloor n/2 \rfloor} + 1 \\
&\leq \frac{1}{2}(1.34)^{2\lceil n/2 \rceil} + \left(-\frac{1}{2} + 1 \right)(1.34)^{\lfloor n/2 \rfloor} + 1 \\
&= \frac{1}{2}(1.34)^{2\lceil n/2 \rceil} + \frac{1}{2}(1.34)^{\lfloor n/2 \rfloor} + 1 \leq \frac{1}{2}(1.34)^{n+1} + \frac{1}{2}(1.34)^{n/2} + 1 \\
&= (1.34)^n + \left[-0.33(1.34)^n + 0.5(1.34)^{n/2} + 1 \right] < (1.34)^n,
\end{aligned}
$$

where the last step follows by noting that $-0.33(1.34)^n + 0.5(1.34)^{n/2} + 1 < 0$ for $1.34^{n/2} > (0.5 + \sqrt{1.57})/0.66$, or $n > 2\log[(0.5 + \sqrt{1.57})/0.66]/\log 1.34 \approx 6.6754$. ◀

We continue now with a more precise analysis via the methods of [4].

3 Obtaining the periodically varying exponential order

Hwang et al. [4] studied recurrences with the nth term written in terms of $\lfloor \frac{n}{2} \rfloor$ and $\lceil \frac{n}{2} \rceil$. Such recurrences can arise in tree problems, in which a quantity associated with the root is written in terms of corresponding quantities for subtrees (see also e.g. [6, 8]). The floor and ceiling function give rise to periodicity in the exponential orders of the associated sequences.

Theorem 5 of [4], which considers recurrences that involve $\lceil \frac{n}{2} \rceil$, enables asymptotic evaluation of a_n from eq. 1. Denote $\{t\} = t - \lfloor t \rfloor$, writing $\{t^-\}$ as the left-continuous version of $\{t\}$: $\{t^-\} = 1$ for integer t, and $\{t^-\} = \{t\}$ otherwise. In other words, $\{t^-\} = 1 - \{-t\}$.

▶ **Theorem 2** ([4]). *Suppose $f(n) = 2f(\lceil \frac{n}{2} \rceil) + g(n)$ for $n \geq 2$, where $f(1)$ is given and $g(1) = 0$. Suppose further that the function $G_m(t) = \sum_{k=0}^{m} 2^{-k}g(\lceil 2^k t \rceil)$ converges uniformly to $G(t) = \sum_{k=0}^{\infty} 2^{-k}g(\lceil 2^k t \rceil)$ for $t \in [1, 2]$.*

Then for $n \geq 1$, we have $f(n) = nP(\log_2 n) - Q(n)$, with P and Q defined by

$$
P(t) = 2^{-\{t^-\}}\left[G(2^{\{t^-\}}) + 2f(1) \right]
$$

$$
Q(n) = G(n) - g(n) = \sum_{k=1}^{\infty} 2^{-k}g(2^k n).
$$

The theorem states that for a class of recurrences in which $f(n)$ is expressed in terms of $f(\lceil \frac{n}{2} \rceil)$, $f(n)$ can be written in terms of a periodic function P that varies with the fractional part of $\log_2 n$. We rewrite a_n from eq. 1 in a form suited to the theorem.

Expanding eq. 1, for $n \geq 2$, we have $a_n = \frac{1}{2}a_{\lceil n/2 \rceil}^2 - \frac{1}{2}a_{\lceil n/2 \rceil} + a_{\lfloor n/2 \rfloor} + 1$, with $a_1 = 1$. We augment the definition by writing $a_0 = 0$. Writing $a_n = 2g_n - \frac{1}{2}$, we have $g_0 = \frac{1}{4}$, $g_1 = \frac{3}{4}$, and for $n \geq 2$, $g_n = g_{\lceil n/2 \rceil}^2 + h_n$, where $h_n = g_{\lfloor n/2 \rfloor} - g_{\lceil n/2 \rceil} + \frac{11}{16}$, with $h_0 = \frac{11}{16}$ and $h_1 = \frac{3}{16}$.

Let $\lambda_n = \log_2 g_n$. Then $\lambda_0 = -2$, $\lambda_1 = \log_2 \frac{3}{4}$, and for $n \geq 2$, $\lambda_n = 2\lambda_{\lceil n/2 \rceil} + \mu_n$, where $\mu_n = \log_2(1 + h_n/g_{\lceil n/2 \rceil}^2)$ for $n \geq 2$. We set $\mu_1 = 0$; a value for μ_0 is not needed.

▶ **Proposition 3.** *For $n \geq 2$, the sequence λ_n can be written $\lambda_n = nP(\log_2 n) - Q(n)$, where*

$$
P(t) = 2^{-\{t^-\}}\left(2\lambda_1 + \sum_{k=0}^{\infty} 2^{-k}\mu_{\lceil 2^{k+\{t^-\}} \rceil} \right) \tag{2}
$$

$$
Q(n) = \sum_{k=1}^{\infty} 2^{-k}\mu_{2^k n}. \tag{3}
$$

Proof. First, λ_n has the correct recursive form for the theorem: $\lambda_n = 2\lambda_{\lceil n/2 \rceil} + \mu_n$ for $n \geq 2$, with λ and μ in the roles of f and g. λ_1 is given, equaling $\log_2 \frac{3}{4}$, and $\mu_1 = 0$ by definition.

Note that the μ_n depend on the λ_n, which is not the case for $g(n)$ in Theorem 2 in relation to $f(n)$, so that Theorem 2 does not immediately apply. However, because $f(n)$ here is solved in closed form without error, we can check the conditions of the theorem – which amounts to showing the convergence of an infinite series – and still apply the resulting solution to λ_n.

If we can show uniform convergence of $G_m(t) = \sum_{k=0}^{m} 2^{-k} \mu_{\lceil 2^k t \rceil}$ to $G(t) = \sum_{k=0}^{\infty} 2^{-k} \mu_{\lceil 2^k t \rceil}$ on $t \in [1, 2]$, then the proposition will follow by Theorem 2, with f replaced by λ and g by μ. To prove this uniform convergence result, we first note that

$$\mu_n = \log_2 \left(1 + \frac{g_{\lfloor n/2 \rfloor} - g_{\lceil n/2 \rceil} + \frac{11}{16}}{g_{\lceil n/2 \rceil}^2} \right) = \log_2 \left(1 + \frac{g_{\lfloor n/2 \rfloor}}{g_{\lceil n/2 \rceil}^2} - \frac{1}{g_{\lceil n/2 \rceil}} + \frac{\frac{11}{16}}{g_{\lceil n/2 \rceil}^2} \right)$$

$$\leq \log_2 \left(1 + \frac{1}{g_{\lceil n/2 \rceil}} - \frac{1}{g_{\lceil n/2 \rceil}} + \frac{\frac{11}{16}}{g_{\lceil n/2 \rceil}^2} \right) = \log_2 \left(1 + \frac{\frac{11}{16}}{g_{\lceil n/2 \rceil}^2} \right).$$

The inequality follows from $g_{\lfloor n/2 \rfloor} \leq g_{\lceil n/2 \rceil}$, which holds because $g_n = \frac{1}{2}(a_n + \frac{1}{2})$ and $\{a_n\}_{n=1}^{\infty}$ is strictly increasing [9, Lemma 5]. Then $\{g_n\}_{n=1}^{\infty}$ is also strictly increasing. We conclude that there exists a constant upper bound on $\log_2 \left[1 + (\frac{11}{16})/g_{\lceil n/2 \rceil}^2 \right]$ that is applicable for all $n \geq 1$. Next, notice that

$$\mu_n = \log_2 \left(1 + \frac{g_{\lfloor n/2 \rfloor}}{g_{\lceil n/2 \rceil}^2} - \frac{1}{g_{\lceil n/2 \rceil}} + \frac{\frac{11}{16}}{g_{\lceil n/2 \rceil}^2} \right) \geq \log_2 \left(1 - \frac{1}{g_{\lceil n/2 \rceil}} \right).$$

Because $\{g_n\}_{n=1}^{\infty}$ is strictly increasing and $g_2 = \frac{5}{4}$, we can conclude that for all $n \geq 4$, $\mu_n \geq \log_2 \left(1 - 1/g_{\lceil n/2 \rceil} \right) \geq \log_2 \left(1 - 1/g_2 \right) = \log_2 \left(\frac{1}{5} \right)$. Hence, $\mu_n \geq \min \left(\mu_1, \mu_2, \mu_3, \log_2(\frac{1}{5}) \right)$ for all $n \geq 1$, showing that μ_n is also bounded below by a constant applicable for all $n \geq 1$.

Because μ_n is bounded below and above by constants applicable for all n, there exists a constant M such that $|\mu_n| < M$ for all $n \geq 1$. We use this constant to show that $G_m(t)$ converges uniformly to $G(t) = \sum_{k=0}^{\infty} 2^{-k} \mu_{\lceil 2^k t \rceil}$ for $t \in [1, 2]$. Indeed, if we let $g_k(t) = 2^{-k} \mu_{\lceil 2^k t \rceil}$, we then have that $G_m(t) = \sum_{k=0}^{m} g_k(t)$. Because $|g_k(t)| = |2^{-k} \mu_{\lceil 2^k t \rceil}| \leq 2^{-k} M$ for all $t \in [1, 2]$ and $k \geq 0$ and $\sum_{k=0}^{\infty} 2^{-k} M = 2M < \infty$, it follows by the Weierstrass M-test that $G_m(t)$ converges uniformly to $G(t)$ on $t \in [1, 2]$, as desired.

By Theorem 2, we deduce that $\lambda_n = nP(\log_2 n) - Q(n)$, where P and Q are defined by

$$P(t) := 2^{-\{t^-\}} \left[G(2^{\{t^-\}}) + 2\lambda_1 \right] = 2^{-\{t^-\}} \left(2\lambda_1 + \sum_{k=0}^{\infty} 2^{-k} \mu_{\lceil 2^{k+\{t^-\}} \rceil} \right) \tag{4}$$

$$Q(n) := G(n) - \mu_n = \sum_{k=1}^{\infty} 2^{-k} \mu_{2^k n}. \tag{5}$$

◄

Examples of $a_n, g_n, h_n, \lambda_n, \mu_n, P(\log_2 n)$ and $Q(n)$ for small values of n appear in Table 1. Values for $P(\log_2 n)$ and $Q(n)$ are numerical approximations, and values for λ_n and μ_n are rounded to four decimal places. To find the asymptotic growth of a_n, we use Proposition 3:

$$a_n = 2g_n - \frac{1}{2} = 2(2^{\lambda_n}) - \frac{1}{2} = 2[2^{nP(\log_2 n) - Q(n)}] - \frac{1}{2} = 2[2^{-Q(n)}] [2^{P(\log_2 n)}]^n - \frac{1}{2}.$$

■ **Table 1** Examples of $a_n, g_n, h_n, \lambda_n, \mu_n, P(\log_2 n)$ and $Q(n)$ for $0 \leq n \leq 5$. a_n is calculated recursively using eq. 1, and g_n is evaluated from $a_n = 2g_n - \frac{1}{2}$. h_n is evaluated as $h_n = g_{\lfloor n/2 \rfloor} - g_{\lceil n/2 \rceil} + \frac{11}{16}$, and λ_n as $\lambda_n = \log_2 g_n$. μ_n is defined as $\log_2(1 + h_n/g^2_{\lceil n/2 \rceil})$ for $n \geq 2$ with $\mu_1 = 0$. The values of $P(\log_2 n)$ and $Q(n)$ are approximated via eqs. 2 and 3, using the values of μ_n and λ_1.

n	a_n	g_n	h_n	λ_n	μ_n	$P(\log_2 n)$	$Q(n)$
0	0	0.25	0.6875	-2	-	-	-
1	1	0.75	0.1875	-0.4150	0	0.3173	0.7324
2	2	1.25	0.6875	0.3219	1.1520	0.3173	0.3127
3	3	1.75	0.1875	0.8074	0.1635	0.3237	0.1639
4	4	2.25	0.6875	1.1699	0.5261	0.3173	0.0994
5	6	3.25	0.1875	1.7004	0.0857	0.3496	0.0474
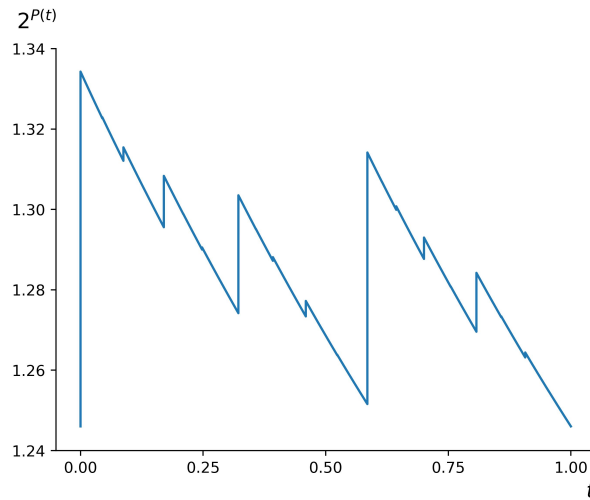
■ **Figure 1** $2^{P(t)}$ as a function of t. $2^{P(t)}$ is a periodic function with period 1. The plot computes all μ_n for $1 \leq n \leq 2056$ using $\mu_n = \log_2(1 + h_n/g^2_{\lceil n/2 \rceil})$ for $n \geq 2$, where $h_n = g_{\lfloor n/2 \rfloor} - g_{\lceil n/2 \rceil} + \frac{11}{16}$ and $g_n = \frac{a_n}{2} + \frac{1}{4}$ for $n \geq 2$ with $g_1 = \frac{3}{4}$ and $h_1 = \frac{3}{16}$; a_n is defined by eq. 1. We use the μ_n to approximate $P(t)$ as in eq. 2, evaluating $P(t)$ for all $t = k/100000$ for integers $0 \leq k < 100000$.

We use a lemma from A to show in B that $Q(n) \to 0$ as $n \to \infty$. It follows that $a_n \sim 2[2^{P(\log_2 n)}]^n$. The asymptotic exponential growth of a_n thus depends only on the value of $\{\log_2 n\}$. Because P is a periodic function with period 1, we have that $P[\log_2(2n)] = P(1 + \log_2 n) = P(\log_2 n)$. The base of the exponent of a_{2n} is the same as that of a_n for any n. A plot of $2^{P(t)}$ as a function of $t \in [0, 1]$ appears in Figure 1.

The function in Figure 1 appears to have many discontinuities, the most visually apparent of which lies at $t = 0$. In the next section, we show that $2^{P(t)}$ has its supremum as $P(t)$ approaches 0 from the right and its minimum at $t = 0$.

4 The upper bound on the exponential order

From Section 3, $a_n \sim 2[2^{P(\log_2 n)}]^n$. Hence, to find upper and lower bounds on the exponential order of a_n, we must find the extreme values of $2^{P(\log_2 n)}$. Because P is a 1-periodic function, it suffices to find the extrema of $2^{P(t)}$ on $t \in [0, 1)$.

We obtain the upper bound in Theorem 6 and the lower bound in Theorem 9. The proof of Theorem 6 requires an inequality that concerns a certain sum involving the μ_n. To prove the inequality, Lemma 4 obtains a term-wise result for terms in the sum that have a sufficiently high index. The term-wise result does not hold for terms with a small index, and Lemma 5 addresses their sum all at once. The lemmas are proven in C.

▶ **Lemma 4.** $\mu_{2^k+1} - 2^{-t}\mu_{\lceil 2^{k+t}\rceil} - \lambda_1(2^{-t}-1) > -[\mu_2 - 2^{-t}\mu_2 - \lambda_1(2^{-t}-1)]$ *for all integers* $k \geq 11$ *and all* $t \in (0,1)$.

▶ **Lemma 5.** *For all* $t \in (0,1)$,

$$\sum_{k=1}^{10}\left[2^{-k}\mu_{2^k+1} - 2^{-k-t}\mu_{\lceil 2^{k+t}\rceil} - 2^{-k}\lambda_1(2^{-t}-1)\right] > -\sum_{k=1}^{10}2^{-k}\left[\mu_2 - 2^{-t}\mu_2 - \lambda_1(2^{-t}-1)\right].$$

▶ **Theorem 6.** $\sup_{t\in(0,1)} 2^{P(t)} = \lim_{t\to0^+} 2^{P(t)}$.

Proof. Because 2^x is a strictly increasing function with respect to x, finding the supremum of $2^{P(t)}$ on $(0,1)$ is equivalent to finding the supremum of $P(t)$. For $t \in (0,1)$, $\{t^-\} = t$. Hence, applying the definition of $P(t)$ from eq. 4,

$$\lim_{t\to0^+} P(t) = \lim_{t\to0^+} 2^{-\{t^-\}}\left(2\lambda_1 + \sum_{k=0}^{\infty}2^{-k}\mu_{\lceil 2^{k+\{t^-\}}\rceil}\right) = 2\lambda_1 + \sum_{k=0}^{\infty}2^{-k}\mu_{2^k+1}.$$

Proving that $P(t) < \lim_{t\to0^+} P(t)$ for $t \in (0,1)$ is equivalent to proving

$$2^{-t}\left(2\lambda_1 + \sum_{k=0}^{\infty}2^{-k}\mu_{\lceil 2^{k+t}\rceil}\right) < 2\lambda_1 + \sum_{k=0}^{\infty}2^{-k}\mu_{2^k+1}. \tag{6}$$

Rearranging eq. 6 and noting $2 = \sum_{k=0}^{\infty}2^{-k}$, we must prove $\sum_{k=0}^{\infty}[2^{-k}\mu_{2^k+1} - 2^{-k-t}\mu_{\lceil 2^{k+t}\rceil} - 2^{-k}\lambda_1(2^{-t}-1)] > 0$, or equivalently, extracting the $k = 0$ term,

$$\mu_2 - 2^{-t}\mu_2 - \lambda_1(2^{-t}-1) + \sum_{k=1}^{\infty}\left[2^{-k}\mu_{2^k+1} - 2^{-k-t}\mu_{\lceil 2^{k+t}\rceil} - 2^{-k}\lambda_1(2^{-t}-1)\right] > 0. \tag{7}$$

By Lemmas 4 and 5, we have the following:

$$\left[\mu_2 - 2^{-t}\mu_2 - \lambda_1(2^{-t}-1)\right] + \sum_{k=1}^{10}\left[2^{-k}\mu_{2^k+1} - 2^{-k-t}\mu_{\lceil 2^{k+t}\rceil} - 2^{-k}\lambda_1(2^{-t}-1)\right]$$

$$+ \sum_{k=11}^{\infty}\left[2^{-k}\left[\mu_{2^k+1} - 2^{-t}\mu_{\lceil 2^{k+t}\rceil} - \lambda_1(2^{-t}-1)\right]\right]$$

$$> \left[\mu_2 - 2^{-t}\mu_2 - \lambda_1(2^{-t}-1)\right]\left[1 - \sum_{k=1}^{10}2^{-k} - \sum_{k=11}^{\infty}2^{-k}\right] = 0.$$

The chain of inequalities verifies eq. 7, proving the theorem. ◀

5 The lower bound on the exponential order

We can use techniques similar to those of Section 4 to find the minimum of $2^{P(t)}$ for $t \in [0,1)$. Again, we need two lemmas, one for terms with a sufficiently large index, and another for terms with small values for the index. The lemmas are proven in D.

▶ **Lemma 7.** *For integers* $k \geq 11$ *and all* $t \in (0,1)$:

$$2^{-t}\mu_{\lceil 2^{k+t}\rceil} - 2^{-1}\mu_{2^k+1} - \lambda_1(2^{-1}-2^{-t}) > -[2^{-t}\mu_2 - 2^{-1}\mu_2 - \lambda_1(2^{-1}-2^{-t})].$$

▶ **Lemma 8.** *For all $t \in (0,1)$,*

$$\sum_{k=1}^{10} \left[2^{-k-t}\mu_{\lceil 2^k+t \rceil} - 2^{-k-1}\mu_{2^{k+1}} - 2^{-k}\lambda_1(2^{-1} - 2^{-t}) \right]$$

$$> -\sum_{k=1}^{10} 2^{-k} \left[2^{-t}\mu_2 - 2^{-1}\mu_2 - \lambda_1(2^{-1} - 2^{-t}) \right].$$

▶ **Theorem 9.** $\min_{t \in [0,1)} 2^{P(t)} = 2^{P(0)}$.

Proof. As before, 2^x is an increasing function in x, so that finding the minimum of the 1-periodic $2^{P(t)}$ is equivalent to finding the minimum of $P(t)$ over $[0,1)$. We must show that

$$P(0) = \frac{1}{2}\left[2\lambda_1 + \sum_{k=0}^{\infty} 2^{-k}\mu_{2^{k+1}} \right] < 2^{-t}\left[2\lambda_1 + \sum_{k=0}^{\infty} 2^{-k}\mu_{\lceil 2^k+t \rceil} \right] = P(t),$$

for all $t \in (0,1)$. Equivalently, replacing 2 by $\sum_{k=0}^{\infty} 2^{-k}$, we must show $\sum_{k=0}^{\infty} [2^{-k-t}\mu_{\lceil 2^k+t \rceil} - 2^{-k-1}\mu_{2^{k+1}} - 2^{-k}\lambda_1(2^{-1} - 2^{-t})] > 0$. Using Lemmas 7 and 8,

$$\sum_{k=0}^{\infty} [2^{-k-t}\mu_{\lceil 2^k+t \rceil} - 2^{-k-1}\mu_{2^{k+1}} - 2^{-k}\lambda_1(2^{-1} - 2^{-t})]$$

$$> 2^{-t}\mu_2 - 2^{-1}\mu_2 - \lambda_1(2^{-1} - 2^{-t}) - \sum_{k=1}^{10} 2^{-k}\left[2^{-t}\mu_2 - 2^{-1}\mu_2 - \lambda_1(2^{-1} - 2^{-t}) \right]$$

$$- \sum_{k=11}^{\infty} 2^{-k}[2^{-t}\mu_2 - 2^{-1}\mu_2 - \lambda_1(2^{-1} - 2^{-t})]$$

$$= [2^{-t}\mu_2 - 2^{-1}\mu_2 - \lambda_1(2^{-1} - 2^{-t})]\left[1 - \sum_{k=1}^{10} 2^{-k} - \sum_{k=11}^{\infty} 2^{-k} \right] = 0. \qquad \blacktriangleleft$$

6 Summary of exponential bounds

Theorems 6 and 9 produce the following corollary. We define two constants, c_1 and c_2:

$$c_1 = 2^{P(0)} = 2^{\frac{1}{2}(2\lambda_1 + \sum_{k=0}^{\infty} 2^{-k}\mu_{2^{k+1}})} \approx 1.2460208329836624$$

$$c_2 = \lim_{\log_2 n \to 0^+} 2^{P(\log_2 n)} = 2^{(2\lambda_1 + \sum_{k=0}^{\infty} 2^{-k}\mu_{2^k+1})} \approx 1.3342827071604892.$$

▶ **Corollary 10.** $\limsup_{n\to\infty} [a_n/(2c_2^n)] = 1$, *and* $\liminf_{n\to\infty} [a_n/(2c_1^n)] = 1$.

Proof. $a_n \sim 2[2^{P(\log_2 n)}]^n$ by Proposition 3, or $\lim_{n\to\infty} [a_n/(2[2^{P(\log_2 n)}]^n)] = 1$. By Theorem 6, the supremum of $2^{P(\log_2 n)}$ on $[0,1)$ is attained as $\{\log_2 n\} \to 0^+$. By definition of c_2, the supremum is $\lim_{\log_2 n \to 0^+} 2^{P(\log_2 n)} = c_2$. Hence, $\limsup_{n\to\infty}[a_n/(2c_2^n)] = 1$.

Similarly, by Theorem 9, the minimum of $2^{P(\log_2 n)}$ is attained at $\{\log_2 n\} = 0$. The minimum is thus $2^{P(0)} = c_1$. We conclude that $\liminf_{n\to\infty}[a_n/(2c_1^n)] = 1$. $\qquad \blacktriangleleft$

Note that the constant c_1 is equal to the value of α in [9, Theorem 13], which finds $a_{2^n} \sim 2\alpha^{(2^n)}$. We can also improve on the upper bound $a_n < (\frac{3}{2})^n$ from [9, Proposition 15], producing a corollary that gives the strictest exponential upper bound possible for a_n.

▶ **Corollary 11.** $a_n < 2c_2^n$ *for all* $n \geq 1$.

Proof. By Proposition 3, $a_n = 2\left[2^{-Q(n)}\right]\left[2^{P(\log_2 n)}\right]^n - \frac{1}{2}$ for all $n \geq 1$. In B, we prove $Q(n) > 0$ for every integer $n \geq 1$. By Theorem 6, we have $a_n = 2\left[2^{-Q(n)}\right]\left[2^{P(\log_2 n)}\right]^n - \frac{1}{2} < 2\left[2^{-Q(n)}\right]\left[2^{P(\log_2 n)}\right]^n < 2\left[2^{P(\log_2 n)}\right]^n < 2c_2^n$. ◀

A result of [4] (see also [5]) enables a computation of $\mathbb{E}[P(t)]$, producing an approximation for the mean of the exponential order $2^{P(t)}$ over the unit interval for t: Theorem 3 of [4] obtains the analogous quantity to $\mathbb{E}[P(t)]$ for recursions $f(n) = f(\lfloor n/2 \rfloor) + f(\lceil n/2 \rceil) + g(n)$. To obtain our next result, we follow its reasoning for a recursion of the form $f(n) = 2f(\lceil n/2 \rceil) + g(n)$.

We wish to compute the mean $\int_0^1 P(t)\, dt$. In the proof of Proposition 3, we showed that $\sum_{k=0}^{\infty} 2^{-k}\mu_{\lceil 2^k t \rceil}$ is bounded above by a constant applicable for all t, so that the 1-periodic $P(t)$ is also bounded on unit intervals for t, say $t \in [0,1]$. To show that the bounded $P(t)$ is integrable on $[0,1]$, it remains to show that $P(t)$ is continuous almost everywhere.

We show $P(t) = 2^{-\{t^-\}}\left(2\lambda_1 + \sum_{k=0}^{\infty} 2^{-k}\mu_{\lceil 2^{k+\{t^-\}} \rceil}\right)$ is continuous outside the countable set $S = \bigcup_{k=1}^{\infty} S_k$, where $S_k = \{t : t \in [0,1] \text{ and } 2^{k+t} \in \mathbb{N}\}$. It suffices to show that $2\lambda_1 + \sum_{k=0}^{\infty} 2^{-k}\mu_{\lceil 2^{k+\{t^-\}} \rceil} = 2\lambda_1 + \sum_{k=0}^{\infty} 2^{-k}\mu_{\lceil 2^{k+t} \rceil}$ is continuous for $t \in (0,1)\backslash S$.

For a positive integer N, recall from the proof of Proposition 3 the uniform convergence of $G_m(t) = \sum_{k=0}^{m} 2^{-k}\mu_{\lceil 2^k t \rceil}$ on $t = [1,2]$ (and uniform boundedness of $|\mu_n|$ by M for all n). Choose $\epsilon_N > 0$ such that $\sum_{k=N+1}^{\infty} |2^{-k}\mu_{\lceil 2^{k+t} \rceil}| < \epsilon_N/2$ for all $t \in (0,1)$. For all $t \in (0,1)\backslash S$, given N, there exists $\delta_N > 0$ such that for all $x \in (t - \delta_N, t + \delta_N)$, $2\lambda_1 + \sum_{k=0}^{N} 2^{-k}\mu_{\lceil 2^{k+x} \rceil} = 2\lambda_1 + \sum_{k=0}^{N} 2^{-k}\mu_{\lceil 2^{k+t} \rceil}$. Therefore, for all $x \in (t - \delta_N, t + \delta_N)$,

$$\left|\left(2\lambda_1 + \sum_{k=0}^{\infty} 2^{-k}\mu_{\lceil 2^{k+t} \rceil}\right) - \left(2\lambda_1 + \sum_{k=0}^{\infty} 2^{-k}\mu_{\lceil 2^{k+x} \rceil}\right)\right| = \left|\sum_{k=N+1}^{\infty} 2^{-k}(\mu_{\lceil 2^{k+t} \rceil} - \mu_{\lceil 2^{k+x} \rceil})\right|.$$

Then $\left|\sum_{k=N+1}^{\infty} 2^{-k}(\mu_{\lceil 2^{k+t} \rceil} - \mu_{\lceil 2^{k+x} \rceil})\right| < \sum_{k=N+1}^{\infty} |2^{-k}\mu_{\lceil 2^{k+t} \rceil}| + \sum_{k=N+1}^{\infty} |2^{-k}\mu_{\lceil 2^{k+x} \rceil}| < \epsilon_N$. We let N grow large, concluding that $P(t)$ is continuous almost everywhere and integrable.

For the integral, with $G(t) = \sum_{k=0}^{\infty} 2^{-k}\mu_{\lceil 2^k t \rceil}$, we have that $\int_0^1 P(t)\, dt = \int_0^1 2^{1-t}\lambda_1\, dt + \int_0^1 2^{-t}G(2^t)\, dt = \lambda_1/\log 2 + \int_0^1 2^{-t}G(2^t)\, dt$. Define $\mu(v) = \mu_{\lceil v \rceil}$. It remains to compute

$$\int_0^1 2^{-t}G(2^t)\, dt = \frac{1}{\log 2}\int_1^2 v^{-2}G(v)\, dv = \frac{1}{\log 2}\int_1^2 v^{-2}\sum_{k=0}^{\infty} 2^{-k}\mu(2^k v)\, dv$$

$$\overset{DCT}{=} \frac{1}{\log 2}\lim_{m\to\infty}\int_1^2 \sum_{k=0}^{m} v^{-2}2^{-k}\mu(2^k v)\, dv = \frac{1}{\log 2}\lim_{m\to\infty}\sum_{k=0}^{m}\int_1^2 v^{-2}2^{-k}\mu(2^k v)\, dv$$

$$= \frac{1}{\log 2}\lim_{m\to\infty}\sum_{k=0}^{m}\int_{2^k}^{2^{k+1}} y^{-2}\mu(y)\, dy = \frac{1}{\log 2}\lim_{m\to\infty}\int_1^{2^{m+1}} y^{-2}\mu(y)\, dy.$$

To justify use of the dominated convergence theorem (DCT), we note that for $v \in [1,2]$, $\left|\sum_{k=0}^{m} v^{-2}2^{-k}\mu(2^k v)\right| \leq \sum_{k=0}^{m} |v^{-2}2^{-k}\mu(2^k v)| \leq \sum_{k=0}^{\infty} |v^{-2}2^{-k}\mu(2^k v)| \leq \sum_{k=0}^{\infty} |2^{-k}\mu(2^k v)|$, a quantity that is uniformly bounded for all v. Next, notice that

$$\int_1^{2^{m+1}} y^{-2}\mu(y)\, dy = \sum_{n=2}^{2^{m+1}}\int_{n-1}^{n} y^{-2}\mu(y)\, dy = \sum_{n=2}^{2^{m+1}}\int_{n-1}^{n} y^{-2}\mu(\lceil y \rceil)\, dy$$

$$= \sum_{n=2}^{2^{m+1}}\mu_n \int_{n-1}^{n} y^{-2}\, dy = \sum_{n=2}^{2^{m+1}}\frac{\mu_n}{n(n-1)}.$$

Therefore,

$$\frac{1}{\log 2}\lim_{m\to\infty}\int_1^{2^{m+1}} y^{-2}\mu(y)\, dy = \frac{1}{\log 2}\lim_{m\to\infty}\sum_{n=2}^{2^{m+1}}\frac{\mu_n}{n(n-1)} = \frac{1}{\log 2}\sum_{n=2}^{\infty}\frac{\mu_n}{n(n-1)}.$$

We have proven the following proposition.

▶ **Proposition 12.** *The mean value of $P(t)$ on the unit interval $[0, 1]$ is*

$$\int_0^1 P(t)\,dt = \frac{\lambda_1}{\log 2} + \frac{1}{\log 2} \sum_{n=2}^{\infty} \frac{\mu_n}{n(n-1)}.$$

By Jensen's inequality on the convex function $\varphi(x) = 2^x$, we obtain a numerical lower bound on the mean of the exponential order $2^{P(t)}$,

$$\int_0^1 2^{P(t)}dt = \mathbb{E}[2^{P(t)}] \geq 2^{\mathbb{E}[P(t)]} = 2^{\int_0^1 P(t)dt} \approx 1.2860382564771475.$$

Note that the mean values of $P(t)$ and $2^{\mathbb{E}[P(t)]}$ represent means for uniformly distributed t; they do not correspond to means over integers n with fixed $\lfloor \log_2 n \rfloor$, as $\log_2 n$ does not have a uniformly distributed fractional part over integers n with fixed $\lfloor \log_2 n \rfloor$.

7   Discussion

We have solved the problem of finding an exact expression for the asymptotic growth of a_n, the minimal Colijn-Plazzotta rank among unlabeled binary rooted trees with n leaves. We find that a_n has periodically varying exponential growth, with exponential order depending on $\{\log_2 n\}$ (Section 3). Its value lies in $[1.246020832983662, 1.3342827071604892)$, where the lower bound is achieved if $\{\log_2 n\} = 0$ and the upper bound is approached as $\{\log_2 n\} \to 0^+$ (Sections 5 and 4). We have obtained the tight upper bound $a_n < 2c_2^n$ for all $n \geq 1$, where $c \approx 1.3342827071604892$ (Corollary 11), improving upon an earlier bound.

The growth of a_n is slowest when n is a power of two and fastest when n is slightly larger than a power of two. This result captures the "jumps" that occur in CP rank near powers of two. For example, in [9, Figure 1] the ratio $a_n/(2\alpha^n)$ for $\alpha \approx 1.24602$ is near 1 if n is a power of two but sharply increases when n is one larger than a power of two. The jumps are visible numerically: $a_{32} = 2279$ yet $a_{33} = 20369$, and $a_{64} = 2598061$ yet $a_{65} = 207440176$. The dependence of the exponential growth of a_n on $\{\log_2 n\}$ reflects these discontinuities.

Maranca & Rosenberg [7] studied an extension of CP rank to strictly and at-most-k-furcating trees, $k \geq 2$, where each internal node of a strictly k-furcating tree has exactly k children, and each internal node of an at-most-k-furcating tree has at least two and at most k children. For such trees, the same questions about the minimal and maximal rank among trees with n leaves can be posed. The work of [4] contains theorems that can potentially be used for asymptotics in these more general cases, whose analyses we defer to future work.

The Colijn-Plazzotta rank has been suggested for use in tree balance indices [3, 9]. Our results characterize the minimal value of the rank across trees with a fixed number of nodes, so that a statistic such as $[f(t) - a_n]/(b_n - a_n)$ or $[\log f(t) - \log a_n]/(\log b_n - \log a_n)$ normalized to lie in $[0, 1]$ can be used as a measure of the balance of a tree.

─── **References** ───────────────────────

1   A. V. Aho and N. J. A. Sloane. Some doubly exponential sequences. *Fibonacci Q.*, 11:429–437, 1973.

2   C. Colijn and G. Plazzotta. A metric on phylogenetic tree shapes. *Syst. Biol.*, 67:113–126, 2018.

3   M. Fischer, L. Herbst, S. Kersting, L. Kühn, and K. Wicke. *Tree balance indices: a comprehensive survey.* Springer, Berlin, 2023.

4   H.-K. Hwang, S. Janson, and T.-H. Tsai. Exact and asymptotic solutions of a divide-and-conquer recurrence dividing at half: theory and applications. *ACM Trans. Alg.*, 13:47, 2017.

5   H.-K. Hwang, S. Janson, and T.-H. Tsai. Identities and periodic oscillations of divide-and-conquer recurrences splitting at half. *Adv. Appl. Math*, 155:102653, 2024.

6   H.-K. Hwang and J.-M. Steyaert. On the number of heaps and the cost of heap construction. In B. Chauvin, P. Flajolet, D. Gardy, and A. Mokkadem, editors, *Mathematics and Computer Science II. Trends in Mathematics*, pages 295–310, Basel, 2002. Birkhäuser.

7   A.R.P Maranca and N.A. Rosenberg. Bijections between the multifurcating unlabeled rooted trees and the positive integers. *Adv. Appl. Math.*, 153:102612, 2024.

8   S.-Y. Oh and J. C. Kieffer. Fractal compression rate curves in lossless compression of balanced trees. In *2010 IEEE International Symposium on Information Theory (ISIT) Proceedings*, pages 106–110, 2010.

9   N. A. Rosenberg. On the Colijn-Plazzotta numbering scheme for unlabeled binary rooted trees. *Discr. Appl. Math.*, 291:88–98, 2021.

Appendix

A   Upper bound on $x|\mu_x|$

We provide an upper bound for $x|\mu_x|$. This bound is utilized in Sections 4 and 5, as well as for a result in Section 3. First, we need an upper bound on $|\log_2 x|$ for all positive reals x.

▶ **Lemma 13.** *For all $x > 0$, $|\log_2 x| \leq \frac{1}{2}(\log_2 e)|x - \frac{1}{x}|$.*

Proof. For $x \geq 1$, we show $\log x \leq \frac{1}{2}(x - \frac{1}{x})$. Writing a function $f(x) = \log(x) - \frac{x}{2} + \frac{1}{2x}$, we have $f(1) = 0$ and $f'(x) = -(1-x)^2/(2x^2) \leq 0$ for $x \geq 1$. As a function that begins at 0 for $x = 1$ and is nonincreasing for $x \geq 1$, $f(x) \leq 0$ for $x \geq 1$. For $0 < x < 1$, we show $\log x \geq \frac{1}{2}(x - \frac{1}{x})$. This statement follows by noting $\frac{1}{x} > 1$ and applying the case of $x \geq 1$.   ◀

Next, we need a uniform lower bound on the expression $1 + g_{\lfloor n/2 \rfloor}/g_{\lceil n/2 \rceil}^2 - 1/g_{\lceil n/2 \rceil} + 11/(16g_{\lceil n/2 \rceil}^2)$ for $n \geq 66$; the use of $n \geq 66$ follows Proposition 1.

▶ **Lemma 14.** *For all integers $n \geq 66$, $1 + g_{\lfloor n/2 \rfloor}/g_{\lceil n/2 \rceil}^2 - 1/g_{\lceil n/2 \rceil} + 11/(16g_{\lceil n/2 \rceil}^2) > \frac{1}{2}$.*

Proof. Because $g_n = (2a_n + 1)/4$ is positive, we have that $1 + g_{\lfloor n/2 \rfloor}/g_{\lceil n/2 \rceil}^2 - 1/g_{\lceil n/2 \rceil} + 11/(16g_{\lceil n/2 \rceil}^2) > 1 - 1/g_{\lceil n/2 \rceil} \geq 1 - 1/g_{33}$, where the last inequality follows from the monotonicity of g_n. We have $1 - 1/g_{33} \approx 0.9999 > \frac{1}{2}$, as desired.   ◀

We are now ready for the main result of this appendix.

▶ **Lemma 15.** *For all integers $x \geq 245$, $x|\mu_x| < (2\lambda_1 + \mu_2)/2$.*

Proof. First note that $1 + g_{\lfloor x/2 \rfloor}/g_{\lceil x/2 \rceil}^2 - 1/g_{\lceil x/2 \rceil} + 11/(16g_{\lceil x/2 \rceil}^2) > 0$ for $x \geq 245$ by the stronger result in Lemma 14. Using Lemma 13,

$$
\begin{aligned}
x|\mu_x| = x\left|\log_2\left(1 + \frac{g_{\lfloor x/2 \rfloor}}{g_{\lceil x/2 \rceil}^2} - \frac{1}{g_{\lceil x/2 \rceil}} + \frac{11}{16g_{\lceil x/2 \rceil}^2}\right)\right| \\
\leq \frac{x(\log_2 e)}{2}\left|\frac{\frac{g_{\lfloor x/2 \rfloor}}{g_{\lceil x/2 \rceil}^2} - \frac{1}{g_{\lceil x/2 \rceil}} + \frac{11}{16g_{\lceil x/2 \rceil}^2}}{1 + \frac{g_{\lfloor x/2 \rfloor}}{g_{\lceil x/2 \rceil}^2} - \frac{1}{g_{\lceil x/2 \rceil}} + \frac{11}{16g_{\lceil x/2 \rceil}^2}} + \frac{g_{\lfloor x/2 \rfloor}}{g_{\lceil x/2 \rceil}^2} - \frac{1}{g_{\lceil x/2 \rceil}} + \frac{11}{16g_{\lceil x/2 \rceil}^2}\right| \\
\leq \frac{x(\log_2 e)}{2}\left[\left|\frac{\frac{g_{\lfloor x/2 \rfloor}}{g_{\lceil x/2 \rceil}^2} - \frac{1}{g_{\lceil x/2 \rceil}} + \frac{11}{16g_{\lceil x/2 \rceil}^2}}{1 + \frac{g_{\lfloor x/2 \rfloor}}{g_{\lceil x/2 \rceil}^2} - \frac{1}{g_{\lceil x/2 \rceil}} + \frac{11}{16g_{\lceil x/2 \rceil}^2}}\right| + \left|\frac{g_{\lfloor x/2 \rfloor}}{g_{\lceil x/2 \rceil}^2} - \frac{1}{g_{\lceil x/2 \rceil}} + \frac{11}{16g_{\lceil x/2 \rceil}^2}\right|\right] \\
\leq \frac{x(\log_2 e)}{2}\left[\frac{\frac{g_{\lfloor x/2 \rfloor}}{g_{\lceil x/2 \rceil}^2} + \frac{1}{g_{\lceil x/2 \rceil}} + \frac{11}{16g_{\lceil x/2 \rceil}^2}}{1 + \frac{g_{\lfloor x/2 \rfloor}}{g_{\lceil x/2 \rceil}^2} - \frac{1}{g_{\lceil x/2 \rceil}} + \frac{11}{16g_{\lceil x/2 \rceil}^2}} + \frac{g_{\lfloor x/2 \rfloor}}{g_{\lceil x/2 \rceil}^2} + \frac{1}{g_{\lceil x/2 \rceil}} + \frac{11}{16g_{\lceil x/2 \rceil}^2}\right].
\end{aligned}
$$

Using Proposition 1 with the fact that $g_n = (2a_n + 1)/4$, we have that for $n \geq 66$,

$$\frac{2}{(1.2)^n} > \frac{1}{g_n} = \frac{4}{2a_n + 1} > \frac{4}{2(1.34)^n + 1} \tag{8}$$

$$\frac{(1.2)^n}{2} + \frac{1}{4} < \frac{a_n}{2} + \frac{1}{4} = g_n < \frac{(1.34)^n}{2} + \frac{1}{4}. \tag{9}$$

The bounds in eqs. 8 and 9 and Lemma 14 then yield

$$x|\mu_x| < \frac{x(\log_2 e)}{2} \left[\left[\frac{\frac{(1.34)^{\lfloor x/2 \rfloor}}{2} + \frac{1}{4}}{\left(\frac{(1.2)^{\lceil x/2 \rceil}}{2} \right)^2} + \frac{2}{(1.2)^{\lceil x/2 \rceil}} + \frac{11}{16} \left(\frac{2}{(1.2)^{\lceil x/2 \rceil}} \right)^2 \right] \middle/ \left(\frac{1}{2} \right) \right.$$

$$\left. + \frac{\frac{(1.34)^{\lfloor x/2 \rfloor}}{2} + \frac{1}{4}}{\left(\frac{(1.2)^{\lceil x/2 \rceil}}{2} \right)^2} + \frac{2}{(1.2)^{\lceil x/2 \rceil}} + \frac{11}{16} \left(\frac{2}{(1.2)^{\lceil x/2 \rceil}} \right)^2 \right]$$

$$= \frac{3x(\log_2 e)}{2} \left[\left(\frac{(1.34)^{\lfloor x/2 \rfloor}}{2} + \frac{1}{4} \right) \left(\frac{2}{(1.2)^{\lceil x/2 \rceil}} \right)^2 + \frac{2}{(1.2)^{\lceil x/2 \rceil}} + \frac{11}{16} \left(\frac{2}{(1.2)^{\lceil x/2 \rceil}} \right)^2 \right]$$

$$\leq \frac{3x(\log_2 e)}{2} \left[\left(\frac{(1.34)^{x/2}}{2} + \frac{1}{4} \right) \left(\frac{2}{(1.2)^{x/2}} \right)^2 + \frac{2}{(1.2)^{x/2}} + \frac{11}{16} \left(\frac{2}{(1.2)^{x/2}} \right)^2 \right]$$

$$= (3\log_2 e) \left[x \left(\frac{\sqrt{1.34}}{1.2} \right)^x + x \left(\frac{1}{\sqrt{1.2}} \right)^x + \frac{15}{8} x \left(\frac{1}{1.2} \right)^x \right]. \tag{10}$$

Eq. 10 sums constant multiples of three terms of the form xa^x, where $a < 1$. For $a < 1$, function $f(x) = xa^x$ attains its maximum at $x_{\max} = -1/\log a$ and is decreasing for $x > x_{\max}$. With $\sqrt{1.34}/1.2$, $1/\sqrt{1.2}$ and $1/1.2$ in the role of a, $x_{\max}$ evaluates to approximately 27.7880, 10.9696, and 5.4848, respectively. The sum of three decreasing functions is also decreasing. It follows that for $x \geq -1/\log(\sqrt{1.34}/1.2) \approx 27.7880$, the quantity in eq. 10 is decreasing.

To show that the quantity in eq. 10 is less than $(2\lambda_1 + \mu_2)/2$ for $x \geq 245$, it suffices to show that if $x = 245$ is inserted into eq. 10, the result is bounded above by $(2\lambda_1 + \mu_2)/2$; indeed, with $x = 245$, we get 0.15718 in eq. 10, while $(2\lambda_1 + \mu_2)/2 \approx 0.1609640474436812$. ◄

B   Properties of $Q(n)$

We give two results about $Q(n)$: Lemma 16 for Section 3, and Lemma 17 for Section 6.

▶ **Lemma 16.** $\lim_{n \to \infty} Q(n) = 0$.

Proof. We apply Lemma 15. For all $n \geq 245$, noting $(2\lambda_1 + \mu_2)/2 < \mu_2/2$ because $\lambda_1 < 0$,

$$Q(n) = \sum_{k=1}^{\infty} 2^{-k} \mu_{2^k n} \leq \sum_{k=1}^{\infty} 2^{-k} |\mu_{2^k n}| < \sum_{k=1}^{\infty} 2^{-k} \frac{\mu_2}{2(2^k n)} \leq \sum_{k=1}^{\infty} 2^{-k} \frac{\mu_2}{2(2^k n)}.$$

For $n \geq 245$, we have $Q(n) < [\mu_2/(2n)] \sum_{k=1}^{\infty} 4^{-k} = \mu_2/(6n)$, so $Q(n) \to 0$ as $n \to \infty$. ◄

▶ **Lemma 17.** *For all positive integers n, $Q(n) > 0$.*

Proof. By definition, $\mu_n = \log_2[1 + (g_{\lfloor n/2 \rfloor} - g_{\lceil n/2 \rceil} + \frac{11}{16})/g_{\lceil n/2 \rceil}^2]$. For any integer $n \geq 1$, $\mu_{2n} = \log_2[1 + (g_n - g_n + \frac{11}{16})/g_n^2] = \log_2[1 + 11/(16g_n^2)] > 0$, noting $g_n > 0$ because $a_n > 0$.

By definition of $Q(n)$, for any positive integer $n \geq 1$, $Q(n) = \sum_{k=1}^{\infty} 2^{-k} \mu_{2^k n} = \sum_{k=1}^{\infty} 2^{-k} \mu_{2^k n} > 0$, where the last inequality follows because $\mu_{2^k n} > 0$ for each $k \geq 1$. ◄

C   Proofs of Lemmas 4 and 5 for Section 4

This appendix proves Lemmas 4 and 5, used in the proof of Theorem 6. First, we prove two additional lemmas, Lemmas 18 and 19, needed for the proof of Lemma 4.

▶ **Lemma 18.** $\mu_{2^k+1} + 2\lambda_1 + \mu_2 > [2^k/(2^k+1)](\mu_{\lceil 2^{k+t}\rceil} + 2\lambda_1 + \mu_2)$ *for all integers* $k \geq 11$ *and all* $t \in \left[\log_2(2^k+1), 1\right)$.

Proof. Because we assume $k \geq 11$ and $t > 0$, $\lceil 2^{k+t}\rceil \geq 245$ and $2^k + 1 \geq 245$. We apply Lemma 15 twice, to $\lceil 2^{k+t}\rceil$ and then to $2^k + 1$. From the first application, we obtain $\lceil 2^{k+t}\rceil |\mu_{\lceil 2^{k+t}\rceil}| < (2\lambda_1 + \mu_2)/2$, from which $2^k |\mu_{\lceil 2^{k+t}\rceil}| < (2\lambda_1 + \mu_2)/2$.

From the second application, we obtain $(2^k + 1) |\mu_{2^k+1}| < (2\lambda_1 + \mu_2)/2$. We then have

$$\frac{2^k}{2^k+1}\mu_{\lceil 2^{k+t}\rceil} - \mu_{2^k+1} \leq \left|\frac{2^k}{2^k+1}\mu_{\lceil 2^{k+t}\rceil}\right| + \left|\mu_{2^k+1}\right|$$

$$< \frac{1}{2^k+1}\frac{2\lambda_1+\mu_2}{2} + \frac{1}{2^k+1}\frac{2\lambda_1+\mu_2}{2} = \frac{1}{2^k+1}(2\lambda_1+\mu_2).$$

Adding $\mu_{2^k+1} + [2^k/(2^k+1)](2\lambda_1+\mu_2)$ to both sides, we obtain the result. ◀

▶ **Lemma 19.** $\mu_x + 2\lambda_1 + \mu_2 > 0$ *for all integers* $x \geq 1$.

Proof. For $1 \leq x \leq 244$, we verify the finite number of cases computationally. For $x \geq 245$, we can use Lemma 15 to obtain $x|\mu_x| < (2\lambda_1 + \mu_2)/2$. Noting that $\lambda_1 < 0$, we have $|\mu_x| < \mu_2/(2x)$, from which $\mu_x > -\mu_2/(2x)$ because $\mu_2 > 0$. We then have $\mu_x + 2\lambda_1 + \mu_2 > -\mu_2/(2x) + 2\lambda_1 + \mu_2 \geq -\mu_2/(2 \cdot 245) + 2\lambda_1 + \mu_2 \approx 0.31957706816604603 \geq 0$. ◀

We are now ready to provide a lower bound on $\mu_{2^k+1} - 2^{-t}\mu_{\lceil 2^{k+t}\rceil}$, applicable for all $k \geq 1$ and all $t \in (0, 1)$, and independent of k. In particular, we prove Lemma 4.

Proof of Lemma 4. The desired inequality is equivalent to

$$\mu_{2^k+1} + 2\lambda_1 + \mu_2 > 2^{-t}(\mu_{\lceil 2^{k+t}\rceil} + 2\lambda_1 + \mu_2). \tag{11}$$

By Lemma 19, $\mu_{\lceil 2^{k+t}\rceil} + 2\lambda_1 + \mu_2 > 0$ for all positive integer values of $\lceil 2^{k+t}\rceil$, and specifically for $k \geq 11$ and $t \in (0, 1)$. Therefore, the right-hand side of eq. 11 is strictly decreasing in t other than at discontinuities: values of t at which $\lceil 2^{k+t}\rceil$ increments by 1. For fixed k, the discontinuities are precisely those values of t at which 2^{k+t} is one of the integers $2^k, 2^k + 1, \ldots, 2^{k+1} - 1$, the values $t = \log_2(2^k + n) - k$ for integers n, $0 \leq n \leq 2^k - 1$.

To verify inequality 11 for all $t \in (0, 1)$, it suffices to check points at which t approaches a discontinuity from the right. For $t \to 0^+$, inequality 11 becomes $\mu_{2^k+1} + 2\lambda_1 + \mu_2 > 2^{-t}(\mu_{2^k+1}+2\lambda_1+\mu_2)$, which holds from the positivity of $\mu_{2^k+1}+2\lambda_1+\mu_2 > 0$ by Lemma 19.

At $t = \log(2^k + n) - k$ for integers $1 \leq n \leq 2^k - 1$, because the discontinuity is approached from the right, $t > \log_2(2^k+1) - k$, so that $2^{-t} < 2^{-[\log_2(2^k+1)-k]} = 2^k/(2^k+1)$. Lemma 18 gives $\mu_{2^k+1} + 2\lambda_1 + \mu_2 > [2^k/(2^k+1)](\mu_{\lceil 2^{k+t}\rceil} + 2\lambda_1 + \mu_2) > 2^{-t}(\mu_{\lceil 2^{k+t}\rceil} + 2\lambda_1 + \mu_2)$. ◀

Proof of Lemma 5. Moving terms with t to one side, we must prove, for all $t \in (0, 1)$,

$$\sum_{k=1}^{10} 2^{-k}(\mu_{2^k+1} + 2\lambda_1 + \mu_2) > 2^{-t}\sum_{k=1}^{10} 2^{-k}(\mu_{\lceil 2^{k+t}\rceil} + 2\lambda_1 + \mu_2). \tag{12}$$

Because $\mu_{\lceil 2^{k+t}\rceil} + 2\lambda_1 + \mu_2 > 0$ by Lemma 19 for all k, $1 \leq k \leq 10$, and $t \in (0, 1)$, the right-hand side of eq. 12 is decreasing except at values of t where $\lceil 2^{k+t}\rceil$ increments by one: set $S = \{\log_2(2^k + n) - k : 1 \leq k \leq 10, 0 \leq n \leq 2^k - 1\}$. Hence, to verify eq. 12 for all $t \in (0, 1)$, it suffices to examine only the limits as t approaches points in S from the right.

First, for $0 \in S$, as $t \to 0^+$, inequality 12 approaches $\sum_{k=1}^{10} 2^{-k}(\mu_{2^k+1} + 2\lambda_1 + \mu_2) > 2^{-t} \sum_{k=1}^{10} 2^{-k}(\mu_{2^k+1} + 2\lambda_1 + \mu_2)$, which holds by Lemma 19, noting $t > 0$.

Next, denote the points in finite set $S' = \{\log_2(2^k + n) - k : 1 \le k \le 10, 1 \le n \le 2^k - 1\}$ by $t_1 < t_2 < \ldots < t_K$, where $K = |S'|$. Notice that if $t \in (t_i, t_{i+1}]$ for some i, then the right-hand side of inequality 12 is maximized as $t \to t_i^+$. Furthermore, for all $t \in (t_i, t_{i+1}]$:

$$2^{-t} \sum_{k=1}^{10} 2^{-k}(\mu_{\lceil 2^{k+t} \rceil} + 2\lambda_1 + \mu_2) < \lim_{t \to t_i^+} 2^{-t} \sum_{k=1}^{10} 2^{-k}(\mu_{\lceil 2^{k+t} \rceil} + 2\lambda_1 + \mu_2)$$

$$< 2^{-t_i} \sum_{k=1}^{10} 2^{-k}(\mu_{\lceil 2^{k+t_i+\epsilon} \rceil} + 2\lambda_1 + \mu_2),$$

where $\epsilon > 0$ satisfies $t_i + \epsilon < t_{i+1}$ for all i, $1 \le i \le K - 1$; we can take $\epsilon = \frac{1}{2} \min_{1 \le i \le K-1}(t_{i+1} - t_i)$. Hence, to prove inequality 12, it suffices to prove the stronger inequality

$$\sum_{k=1}^{10} 2^{-k}(\mu_{2^k+1} + 2\lambda_1 + \mu_2) > 2^{-t_i} \sum_{k=1}^{10} 2^{-k}(\mu_{\lceil 2^{k+t_i+\epsilon} \rceil} + 2\lambda_1 + \mu_2). \tag{13}$$

for each i, $1 \le i \le K$. The advantage of inequality 13 over 12 is that it can be computationally verified by testing a finite number of points. In particular, we consider each $t_i \in S'$, choose an appropriate ϵ ($\epsilon = 10^{-16}$ suffices), and verify inequality 13 with that t_i and ϵ. ◄

D   Proof of Lemmas 7 and 8 for Section 5

This appendix proves two lemmas used for Theorem 9. First, Lemma 20 gives a refined upper bound for $|\mu_x|$ that improves upon Lemma 15. This lemma is needed for Lemma 7.

▶ **Lemma 20.** *For all integers $x \ge 267$, $2^{\lceil \log_2 x \rceil}|\mu_x| < (2\lambda_1 + \mu_2)/2$.*

Proof. From eq. 10, $|\mu_x| \le 3(\log_2 e)\left[(\sqrt{1.34}/1.2)^x + (1/\sqrt{1.2})^x + 15/\left(8(1.2)^x\right)\right]$. Because $2^{\lfloor \log_2 x \rfloor} \le 2^{\log_2 x} = x$, it follows that $2^{\lceil \log_2 x \rceil} \le 2(2^{\lfloor \log_2 x \rfloor}) \le 2x$. Hence, it suffices to prove that $2x|\mu_x| \le (2\lambda_1 + \mu_2)/2$ for integers $x \ge 267$. That is, we can show the stronger inequality

$$2x(3 \log_2 e)\left[\left(\frac{\sqrt{1.34}}{1.2}\right)^x + \frac{1}{(\sqrt{1.2})^x} + \frac{15}{8(1.2)^x}\right] < \frac{2\lambda_1 + \mu_2}{2}.$$

As in A, the left-hand side is the sum of three terms of the form xa^x, where $a < 1$. By the same argument, this sum is decreasing for $x \ge -1/\log(\sqrt{1.34}/1.2) \approx 27.7880$. Hence, it suffices to show that the sum is less than $(2\lambda_1 + \mu_2)/2$ at $x = 267$, which can be accomplished computationally. Therefore, $2^{\lceil \log_2 x \rceil}|\mu_x| < (2\lambda_1 + \mu_2)/2$ for $x \ge 267$. ◄

Proof of Lemma 7. Moving all the terms involving t to one side, we must prove

$$2^{-t}(\mu_{\lceil 2^{k+t} \rceil} + 2\lambda_1 + \mu_2) > 2^{-1}(\mu_{2^{k+1}} + 2\lambda_1 + \mu_2). \tag{14}$$

Lemma 19 finds that $\mu_{\lceil 2^{k+t} \rceil} + 2\lambda_1 + \mu_2 > 0$ for all $k \ge 11$ and $t \in (0, 1)$. Therefore, except at discontinuities where 2^{k+t} is an integer and $\lceil 2^{k+t} \rceil$ increments by 1, the left-hand side strictly decreases as we increase t. It suffices to check inequality 14 at those discontinuities and as $t \to 1^-$. First, as $t \to 1^-$, inequality 14 becomes $2^{-t}(\mu_{2^{k+1}} + 2\lambda_1 + \mu_2) > 2^{-1}(\mu_{2^{k+1}} + 2\lambda_1 + \mu_2)$, a result that follows because $\mu_{2^{k+1}} + 2\lambda_1 + \mu_2 > 0$ and $0 < t < 1$.

Next, because $t \in (0,1)$, 2^{k+t} is an integer if and only if $t = \log_2(2^k + n) - k$ for an integer $1 \leq n \leq 2^k - 1$. Plugging $t = \log_2(2^k + n) - k$ into inequality 14 yields $[2^k/(2^k + n)](\mu_{2^k + n} + 2\lambda_1 + \mu_2) > 2^{-1}(\mu_{2^{k+1}} + 2\lambda_1 + \mu_2)$. Because $\mu_{2^k + n} + 2\lambda_1 + \mu_2 > 0$, it suffices to prove the stronger inequality $[2^k/(2^{k+1} - 1)](\mu_{2^k + n} + 2\lambda_1 + \mu_2) > 2^{-1}(\mu_{2^{k+1}} + 2\lambda_1 + \mu_2)$, or equivalently, $[2^k/(2^{k+1} - 1)]\mu_{2^k + n} - \frac{1}{2}\mu_{2^{k+1}} > -[\frac{1}{2}/(2^{k+1} - 1)](2\lambda_1 + \mu_2)$.

By Lemma 20, which applies for $k \geq 11$ because $2^k + n \geq 267$, we have $2^{\lceil \log_2(2^k + n) \rceil}|\mu_{2^k + n}| < (2\lambda_1 + \mu_2)/2$ and $2^{k+1}|\mu_{2^k + n}| < (2\lambda_1 + \mu_2)/2$. Furthermore, by Lemma 15, we have that $2^{k+1}|\mu_{2^{k+1}}| < (2\lambda_1 + \mu_2)/2$, from which $(2^{k+1} - 1)|\mu_{2^{k+1}}| < (2\lambda_1 + \mu_2)/2$. We then obtain

$$
\frac{2^k}{2^{k+1} - 1}\mu_{2^k + n} - \frac{1}{2}\mu_{2^{k+1}} \geq -\left|\frac{2^k}{2^{k+1} - 1}\mu_{2^k + n}\right| - \left|\frac{1}{2}\mu_{2^{k+1}}\right|
$$

$$
\geq -\frac{1}{2}\frac{1}{2(2^{k+1} - 1)}(2\lambda_1 + \mu_2) - \frac{1}{2}\frac{1}{2(2^{k+1} - 1)}(2\lambda_1 + \mu_2) = -\frac{1}{2(2^{k+1} - 1)}(2\lambda_1 + \mu_2). \blacktriangleleft
$$

Proof of Lemma 8. Moving all the terms involving t to one side, we must prove

$$
2^{-t}\sum_{k=1}^{10} 2^{-k}(\mu_{\lceil 2^{k+t} \rceil} + 2\lambda_1 + \mu_2) > \sum_{k=1}^{10} 2^{-k-1}(\mu_{2^{k+1}} + 2\lambda_1 + \mu_2). \tag{15}
$$

The terms in both summands are positive (Lemma 19). Therefore, except at the discontinuities in the left-hand side, the left-hand side strictly decreases with t. Hence, it suffices to check inequality 15 precisely at the discontinuities of the left-hand side and as $t \to 1^-$.

As $t \to 1^-$, the left-hand side of inequality 15 decreases to $\lim_{t \to 1^-} 2^{-t}\sum_{k=1}^{10} 2^{-k}(\mu_{2^{k+1}} + 2\lambda_1 + \mu_2) = \sum_{k=1}^{10} 2^{-k-1}(\mu_{2^{k+1}} + 2\lambda_1 + \mu_2)$, verifying inequality 15 at $t \to 1^-$.

Hence, it remains to check inequality 15 at the discontinuities of the left-hand side, the points in $S' = \{\log_2(2^k + n) - k : 1 \leq k \leq 10, 1 \leq n \leq 2^k - 1\}$ used in C. We can computationally verify that inequality 15 holds for the finitely many points in S'. $\blacktriangleleft$

Binomial Sums and Mellin Asymptotics with Explicit Error Bounds: A Case Study

Benjamin Hackl ✉ ⓘ
Department of Mathematics and Scientific Computing, University of Graz, Austria

Stephan Wagner ✉ ⓘ
Institute of Discrete Mathematics, TU Graz, Austria
Department of Mathematics, Uppsala University, Sweden

──── **Abstract** ──

Making use of a newly developed package in the computer algebra system SageMath, we show how to perform a full asymptotic analysis by means of the Mellin transform with explicit error bounds. As an application of the method, we answer a question of Bóna and DeJonge on 132-avoiding permutations with a unique longest increasing subsequence that can be translated into an inequality for a certain binomial sum.

1 Introduction

Sums involving binomial coefficients occur frequently in enumerative and analytic combinatorics. For example,

$$\sum_{k=0}^{n} \frac{1}{k+1} \binom{n+k}{n} \binom{n}{k}$$

yields the large Schröder numbers, which count (among other things) many different types of lattice paths and permutations. The sum

$$\sum_{k=0}^{\lfloor n/2 \rfloor} \binom{n}{2k} \frac{(2k)!}{2^k k!}$$

counts involutions, or matchings in complete graphs. There is a well-established toolkit for dealing with such sums, based on techniques such as the *(discrete) Laplace method*, the *Stirling approximation* of factorials and binomial coefficients, and the *Mellin transform*. See [5] for a comprehensive account of these and many other tools. While these methods are well known and in some sense mechanical, it is still not straightforward to implement them in a computer as they often involve ad-hoc estimates and careful splitting into different cases/regions of summation that are analyzed separately. Moreover, while a lot of the complications can often be hidden in O-terms, things become more involved when explicit error bounds are desired.

This paper aims to make a contribution towards building a toolkit for asymptotic analysis in the context of computer algebra, including guaranteed error bounds with explicit constants. The example we use to illustrate the methods is based on a question from a recent paper by Bóna and DeJonge [1]: let a_n be the number of 132-avoiding permutations of length n that have a unique longest increasing subsequence, which is also the number of plane trees with $n + 1$ vertices with a single leaf at maximum distance from the root, or the number of Dyck paths of length $2n$ with a unique peak of maximum height. Moreover, let $p_n = \frac{a_n}{C_n}$, where $C_n = \frac{1}{n+1}\binom{2n}{n}$ is the n-th Catalan number. This can be interpreted as the probability that a 132-avoiding permutation of length n chosen uniformly at random has a unique longest increasing subsequence – equivalently, that a plane tree with $n + 1$ vertices has a single leaf at maximum distance from the root, or that a Dyck path of length $2n$ has a unique peak of maximum height.

▶ **Problem 1** (Bóna and DeJonge [1])**.** *Is it true that the sequence p_n is decreasing for $n \geq 3$?*

While it would obviously be interesting to have a combinatorial proof, it turns out (as we will explain in the following section) that the problem can be translated in a fairly mechanical fashion (using generating functions) into a purely analytic problem: specifically, the inequality

$$F(n) = \sum_{k=1}^{n} k\sigma(k)(k^2 - 3n + 2)(2k^2 - n)\binom{2n}{n-k} < 0 \tag{1}$$

for all $n \geq 5$, where $\sigma(k)$ is the sum of divisors of k. The standard approach to deriving an asymptotic formula for such a sum (cf. [4, Section 5]) involves the following steps:

- Split the sum into "small" and "large" values of k.
- Show that the contribution of large values is negligible.
- Approximate the binomial coefficient $\binom{2n}{n-k}$, e.g. by means of Stirling's formula, for small values of k.
- Turn the sum into an infinite sum, again at the expense of a negligible error term.
- Apply the Mellin transform to obtain an integral representation for the resulting infinite sum.
- Use residue calculus to derive the final asymptotic formula.

As we will see, the problem is complicated in this particular instance by the occurrence of nontrivial cancellations, making precise estimates challenging. The asymptotic formula (that will be proven in this paper)

$$F(n) = \binom{2n}{n}\left(-\frac{n^2}{8} + \frac{n}{24} + o(n)\right) \tag{2}$$

shows that the answer to the question of Bóna and DeJonge is affirmative for *sufficiently large n*. However, the *o*-notation hides an error term that is potentially huge for small values of n, so it is not clear what "sufficiently large" means. In order to show that p_n is increasing for *all $n \geq 3$*, we will have to prove a version of (2) with *explicit error bounds*. To this end, we present a new package for the computer mathematics system SageMath [11] that enhances the core implementation of asymptotic expansions, and in particular the arithmetic with SageMath's analogue of O-terms with explicit error bounds, called B-terms. See Section 3 for a guided tour through the features of our package. We then demonstrate the practical usage of our package in Section 4 in which we derive the desired explicit bounds for $F(n)$.

2   Reducing the problem

One of the possible combinatorial interpretations of the sequence a_n is in terms of lattice paths. Specifically, as it was mentioned before, a_n is the number of Dyck paths of length $2n$ (i.e., lattice paths starting at $(0,0)$ and ending at $(2n,0)$ whose steps are either "up" $(1,1)$ or "down" $(1,-1)$) with a unique peak of maximum height. Such a path can be decomposed into two pieces: before and after the peak. The part before the peak needs to be a path that finishes at its maximum height h (but does not reach it earlier, since the peak is unique), and the path after the peak needs to be a path that starts at its maximum height h and never returns to it (which can also be seen as the reflection of a path that finishes at its maximum height but does not reach it earlier). Such paths were analyzed in [2] and [8]. Specifically, [8, Proposition 2.1] states that the probability that a simple symmetric random walk of length n never drops below 0 and finishes at its maximum height h (which can also be reached earlier) is precisely

$$2[z^{n+1}]\frac{1}{U_{h+1}(1/z)},$$

where U_{h+1} is the Chebyshev polynomial of the second kind of degree $h+1$. A path that finishes at its maximum height h without reaching that height before is obtained from a path that finishes at its maximum height $h-1$ by adding one more step up. Since every path of length n has probability 2^{-n} to occur under the model of a simple symmetric random walk, it follows that the (ordinary) generating function for paths of maximum height h that finish at the maximum and do not reach it earlier is

$$\sum_{n\geq 0} 2^n x^{n+1} \cdot 2[z^{n+1}]\frac{1}{U_h(1/z)} = \frac{1}{U_h(1/(2x))}.$$

For example,

$$\frac{1}{U_3(1/(2x))} = \frac{x^3}{1-2x^2} = x^3 + 2x^5 + 4x^7 + \cdots$$

is the generating function for paths that finish at their maximum height 3 and do not reach this height before the final step. The formula is even true for $h = 0$: $\frac{1}{U_0(1/(2x))} = 1$ is indeed the correct generating function in this case.

Since the paths we are interested in can be seen as pairs of paths that finish at their maximum height and do not reach this height before, we find that the generating function of a_n is

$$A(x) = \sum_{n=0}^{\infty} a_n x^{2n} = \sum_{h=0}^{\infty} \frac{1}{U_h(1/(2x))^2}.$$

We can simplify the expression by means of the substitution $x = \frac{\sqrt{t}}{1+t}$. Note that this yields $\frac{1}{2x} = \frac{1+t}{2\sqrt{t}} = \cosh(\frac{1}{2}\log t)$. Since $U_h(\cosh w) = \frac{\sinh((h+1)w)}{\sinh w}$, this implies that

$$U_h(1/(2x)) = \frac{\sinh(\frac{h+1}{2}\log t)}{\sinh(\frac{1}{2}\log t)} = t^{-h/2} \cdot \frac{1-t^{h+1}}{1-t}.$$

Thus

$$A(x) = \sum_{h=0}^{\infty} \frac{t^h(1-t)^2}{(1-t^{h+1})^2}.$$

Now we can obtain an alternative expression for a_n by applying Cauchy's integral formula to the generating function $A(x)$. For suitable contours $\mathcal{C}$ and $\mathcal{C}'$ around 0, we have

$$a_n = \frac{1}{2\pi i} \oint_{\mathcal{C}} \frac{A(\sqrt{z})}{z^{n+1}} \, dz = \frac{1}{2\pi i} \oint_{\mathcal{C}'} \sum_{h=0}^{\infty} \frac{t^h(1-t)^2}{(1-t^{h+1})^2} \cdot \frac{(1+t)^{2n+2}}{t^{n+1}} \cdot \frac{(1-t)\,dt}{(1+t)^3}$$

$$= \sum_{h=0}^{\infty} \frac{1}{2\pi i} \oint_{\mathcal{C}'} \frac{(1-t)^3(1+t)^{2n-1}}{t^{n+1-h}(1-t^{h+1})^2} \, dt,$$

using the substitution $\sqrt{z} = \frac{\sqrt{t}}{1+t}$ (or equivalently $z = \frac{t}{(1+t)^2}$). It follows that

$$a_n = \sum_{h=0}^{\infty} [t^{n-h}] \frac{(1-t)^3(1+t)^{2n-1}}{(1-t^{h+1})^2} = [t^{n+1}] \sum_{h=0}^{\infty} \frac{t^{h+1}}{(1-t^{h+1})^2}(1-t)^3(1+t)^{2n-1}$$

$$= [t^{n+1}] \sum_{k=1}^{\infty} \sigma(k) t^k (1-t)^3(1+t)^{2n-1} = \sum_{k=1}^{\infty} \sigma(k)[t^{n+1-k}](1-t)^3(1+t)^{2n-1}$$

$$= \sum_{k=1}^{n+1} \sigma(k) \left(\binom{2n-1}{n+1-k} - 3\binom{2n-1}{n-k} + 3\binom{2n-1}{n-1-k} - \binom{2n-1}{n-2-k} \right)$$

$$= \sum_{k=1}^{n+1} \frac{4k\sigma(k)(2k^2-3n-2)(2n-1)!}{(n+1-k)!(n+1+k)!}.$$

We remark here that the identity $[t^a](1+t)^b = \binom{b}{a}$ that we are using even remains true for negative a or for $a > b$ if the binomial coefficient is considered to be 0 then. The manipulation in the final step is consistent with this.

Problem 1 is equivalent to the inequality $C_{n+1}a_n > C_n a_{n+1}$ for $n \geq 3$, and since $C_{n+1} = \frac{4n+2}{n+2}C_n$, it can also be expressed as $(4n+2)a_n > (n+2)a_{n+1}$. Hence we are left to consider the inequality

$$\sum_{k=1}^{n+1} \frac{4k\sigma(k)(2k^2-3n-2)(4n+2)(2n-1)!}{(n+1-k)!(n+1+k)!} > \sum_{k=1}^{n+2} \frac{4k\sigma(k)(2k^2-3n-5)(n+2)(2n+1)!}{(n+2-k)!(n+2+k)!},$$

which reduces to

$$\sum_{k=1}^{n+2} \frac{8k\sigma(k)(k^2-3n-4)(2k^2-n-2)(2n+1)(2n-1)!}{(n+2-k)!(n+2+k)!} < 0$$

after some manipulations. After multiplication by

$$n(n+1)(n+2)(2n+3) = \frac{(2n)(2n+2)(2n+3)(2n+4)}{8},$$

this can be expressed as

$$\sum_{k=1}^{n+2} k\sigma(k)(k^2-3n-4)(2k^2-n-2)\binom{2n+4}{n+2-k} < 0.$$

Finally, replacing $n+2$ by n, what we have to prove in order to settle Problem 1 is that

$$\sum_{k=1}^{n} k\sigma(k)(k^2-3n+2)(2k^2-n)\binom{2n}{n-k} < 0$$

for all $n \geq 5$, which is precisely (1).

3 B-terms and asymptotics with explicit error bounds

In this section, we provide the necessary background on B-terms and their software implementation. We base our work on the module for computing with asymptotic expansions [7] in SageMath [11]. While this module presently also offers some basic support for B-terms, we have extended its capabilities to add support for computations involving an additional monomially bounded variable (e.g., k with $n^\alpha \le k \le n^\beta$ for some $0 \le \alpha < \beta$ where $n \to \infty$), as well as Taylor expansions with explicit error bounds. These improvements are not yet included in the module directly, but can be made available to your local installation of SageMath simply by running

```
sage -pip install dependent_bterms
```

from your terminal. Alternatively, the module can be installed by executing a cell containing

```
!pip install dependent_bterms
```

from within a SageMath Jupyter notebook.

We will now briefly walk through the core functionalities of our toolbox. The central interface is the function

```
AsymptoticRingWithDependentVariable,
```

which generates a suitable parent structure for our desired asymptotic expansions. Listing 1 demonstrates how it is used to instantiate the structure that will be used throughout the following examples. We consider $1 = n^0 \le k \le n^{4/7}$, i.e., $\alpha = 0$ and $\beta = 4/7$.

Listing 1 Setup of the modified `AsymptoticRing`.

```
sage: import dependent_bterms as dbt
sage: AR, n, k = dbt.AsymptoticRingWithDependentVariable(
....:     'n^QQ', 'k', 0, 4/7, bterm_round_to=3, default_prec=5
....: )
sage: AR
Asymptotic Ring <n^QQ> over Symbolic Ring
```

The arguments passed to the interface are, in order,
- `growth_group` – the (univariate) growth group[1] modeling the desired structure of the asymptotic expansions. For example, `'n^QQ'` represents terms like $42n^{9/13}$ or $O(n^{7/42})$.
- `dependent_variable` – a string representation of the symbolic variable being endowed with asymptotic growth bounds, e.g., `'k'`.
- `lower_bound_power` – a real number $\alpha \ge 0$ representing the power to which the ring's independent variable is raised to obtain the lower monomial power.
- `upper_bound_power` – a real number $\beta > \alpha \ge 0$, analogous to `lower_bound_power`, just for the upper bound,
- `bterm_round_to` – a non-negative integer or `None` (the default), determining the number of floating point digits to which the coefficients of B-terms are automatically rounded. If `None`, no rounding is performed.

---

[1] See SageMath's documentation on asymptotic expansions and the `AsymptoticRing` for an introduction to the algebraic terminology used here.

Any other keyword arguments (like `default_prec` in Listing 1 above) are passed to the constructor of `AsymptoticRing`.

In this structure, arithmetic with asymptotic expansions in n can be carried out as usual, see Listing 2. The `default_prec` parameter specified above controls the order of the automatic expansions.

■ **Listing 2** Arithmetic and automatic expansions in `AsymptoticRing`.

```
sage: (1 + 3*n) * (4*n^(-7/3) + 42/n + 1)
3*n + 127 + 42*n^(-1) + 12*n^(-4/3) + 4*n^(-7/3)
sage: prod((1 + n^(-j)) for j in srange(1, 10)) * (1 + O(n^(-10)))
1 + n^(-1) + n^(-2) + 2*n^(-3) + 2*n^(-4) + 3*n^(-5) + 4*n^(-6)
+ 5*n^(-7) + 6*n^(-8) + 8*n^(-9) + O(n^(-10))
sage: n / (n - 1)
1 + n^(-1) + n^(-2) + n^(-3) + n^(-4) + O(n^(-5))
```

In the implementation of the `AsymptoticRing` shipped with SageMath, asymptotic expansions internally rely on ordering their summands with respect to the growth of the independent variable(s), regardless of attached coefficients.

In the extension of our `dependent_bterms` module, expansions are aware of the growth range contributed by the dependent variable appearing in coefficients. In fact, in our modified ring, expansions are ordered with respect to the upper bound of the coefficient growth combined with the growth of the independent variable. This explains the – at first glance counterintuitive – ordering of the summands in Listing 3. The individual growth ranges of the summands are printed at the end of the listing.

■ **Listing 3** Arithmetic involving the dependent variable.

```
sage: k*n^2 + O(n^(3/2)) + k^3*n
k^3*n + k*n^2 + O(n^(3/2))
sage: for summand in expr.summands.elements_topological():
....:       print(f"{summand}␣->␣{summand.dependent_growth_range()}")
O(n^(3/2)) -> (n^(3/2), n^(3/2))
k*n^2 -> (n^2, n^(18/7))
k^3*n -> (n, n^(19/7))
```

Automatic power series expansion (with an O-term error) also works natively in our modified ring, see Listing 4. Observe that the error term $O(n^{-15/7})$ would actually be able to *partially absorb* some of the terms in the automatic expansion like $(k/2 + 1/6)n^{-3}$. This partial absorption is, however, not carried out automatically due to performance reasons. Using the `simplify_expansion` function included in our module expands the symbolic coefficients and enables the error terms to carry out all allowed (partial) absorptions.

■ **Listing 4** Automatic expansions and manual simplifications.

```
sage: auto_expansion = exp((1 + k)/n)
sage: auto_expansion
1 + (k + 1)*n^(-1) + (1/2*(k + 1)^2)*n^(-2) + (1/6*(k + 1)^3)*n^(-3)
+ (1/24*(k + 1)^4)*n^(-4) + O(n^(-15/7))
sage: dbt.simplify_expansion(auto_expansion)
1 + (k + 1)*n^(-1) + (1/2*k^2 + k + 1/2)*n^(-2)
+ (1/6*k^3 + 1/2*k^2)*n^(-3) + 1/24*k^4*n^(-4) + O(n^(-15/7))
```

Now let us turn to the core feature of our extension: B-terms. In a nutshell, B-terms are O-terms that come with an explicitly specified constant and validity point. For example, the term $B_{n \geq 10}(42n^3)$ represents an error term that is bounded by $42n^3$ for $n \geq 10$.

Listing 5 demonstrates basic arithmetic with B-terms. It is worth spending a moment to understand how the resulting constants are determined. In the first example, the B-term $B_{n \geq 10}(5/n)$ absorbs the exact term $3/n^2$ of weaker growth. It does so by automatically estimating $\frac{3}{n^2} \leq \frac{3}{10n}$ (as the term is valid for $n \geq 10$) and then directly absorbing the upper bound; $\frac{53}{10} = 5 + \frac{3}{10}$.

The same mechanism happens in the second example. In order to avoid the (otherwise rapid) accumulation of complicated symbolic expressions in the automatic estimates, we have specified (via the `bterm_round_to`-parameter that we have set to 3) that B-terms should automatically be rounded to three floating point digits. This is why the constant is given as $\lceil (1 + 10^{-1/3}) \cdot 10^3 \rceil \cdot 10^{-3} = \frac{293}{200}$.

■ **Listing 5** Arithmetic with B-terms and the dependent variable.

```
sage: 7*n + AR.B(5/n, valid_from=10) + 3/n^2
7*n + B(53/10*n^(-1), n >= 10)
sage: AR.B(1/n, valid_from=10) + n^(-4/3)
B(293/200*n^(-1), n >= 10)
sage: AR.B(3*k^2/n^3, valid_from=10) + (1 - 2*k + 3*k^2 - 4*k^3)/n^5
B(3373/1000*abs(k^2)*n^(-3), n >= 10)
```

The third example in Listing 5 illustrates arithmetic involving the dependent variable, which requires additional care. With $1 \leq k \leq n^{4/7}$ in place, the growth of the given B-term ranges from $\Theta(n^{-3})$ to $\Theta(n^{-13/7})$. The growth of the explicit term that is added ranges from $\Theta(n^{-5})$ to $\Theta(n^{-23/7})$. In this setting, we consider the explicit term to be of weaker growth, as the lower bound of the B-term is stronger than the lower bound of the explicit term, and likewise for the upper bound. Thus we may let the B-term absorb it. We do so by first estimating

$$\left| \frac{1 - 2k + 3k - 4k^3}{n^5} \right| \leq \frac{(1 + 2 + 3 + 4)k^3}{n^5} = \frac{10k^3}{n^5}.$$

As the power of k in this bound is larger than the maximal power of k in the B-term, we may not yet proceed as above (otherwise we would increase the upper bound of the B-term, which we must avoid). Instead, we use first use $k \leq n^{4/7}$, followed by $n \geq 10$, to obtain

$$\frac{10k^3}{n^5} \leq \frac{10k^2 n^{4/7}}{n^5} = \frac{10k^2}{n^{31/7}} \leq \frac{10k^2}{10^{10/7} \cdot n^3} = 10^{-3/7} \cdot \frac{k^2}{n^3},$$

which the B-term can now absorb directly. Hence the value of the B-term constant is determined by $\lceil (3 + 10^{-3/7}) \cdot 10^3 \rceil \cdot 10^{-3} = \frac{3373}{1000}$.

Finally, our module also provides support for B-term bounded Taylor expansion (again, also involving the dependent variable) in form of the `taylor_with_explicit_error` function. An example is given in Listing 6: we first obtain a Taylor expansion of $f(t) = (1 - t^2)^{-1}$ around $t = (1 + k)/n + B_{n \geq 10}(k^3/n^3)$. Using the `simplify_expansion` function rearranges the terms and lets the B-term absorb coefficients (partially) as far it is able to. Observe that it may happen that the attempted simplification produces summands with a smaller upper growth bound that the implementation cannot absorb ($B_{n \geq 10}(k^3/n^3)$ vs. n^{-2} in this case). The expansion is still correct; just not as compact as it could be. We can also use the `simplify_expansion` function with the `simplify_bterm_growth` parameter set to `True` to collapse the dependent variables in all B-terms by replacing them with their upper bounds, resulting in a single "absolute" B-term.

■ **Listing 6** B-term bounded Taylor expansions.

```
sage: arg = (1 + k)/n + AR.B(k^3/n^3, valid_from=10)
sage: ex = dbt.taylor_with_explicit_error(
....:       lambda t: 1/(1 - t^2), arg,
....:       order=3, valid_from=10)
sage: ex
1 + ((k + 1)^2)*n^(-2)
+ B((abs(7351/250*k^3 + 30*k^2 + 30*k + 10))*n^(-3), n >= 10)
sage: dbt.simplify_expansion(ex)
1 + k^2*n^(-2)
+ B((abs(7351/250*k^3 + 30*k^2 + 30*k + 10))*n^(-3), n >= 10)
+ (2*k + 1)*n^(-2)
sage: dbt.simplify_expansion(ex, simplify_bterm_growth=True)
1 + k^2*n^(-2) + B(41441/1000*n^(-9/7), n >= 10)
```

4   Asymptotic analysis

In the following, we provide the steps of the analysis of the sum $F(n)$, aided by the software package that was presented in the previous section. We will verify (1) for $n \geq N = 10000$ by means of an asymptotic analysis with explicit error terms. For $n < N$, one can verify the inequality with a computer by determining $F(n)$ explicitly in all cases.

All computations carried out in this section can be found in the SageMath notebook located at

$$\texttt{https://arxiv.org/src/2403.09408/anc/2024-bona-dejonge.ipynb},$$

and a corresponding static version (containing computations and results) is available at

$$\texttt{https://arxiv.org/src/2403.09408/anc/2024-bona-dejonge.html}.$$

4.1   Approximating the binomial coefficients

It is useful to divide the entire sum by $\binom{2n}{n}$ and to approximate the quotient. Note that we have

$$\frac{\binom{2n}{n-k}}{\binom{2n}{n}} = \frac{n(n-1)\cdots(n-k+1)}{(n+1)(n+2)\cdots(n+k)} = \frac{n}{n-k}\prod_{j=1}^{k}\frac{n-j}{n+j} = \frac{n}{n-k}\prod_{j=1}^{k}\frac{1-j/n}{1+j/n}.$$

This can be rewritten as

$$\frac{\binom{2n}{n-k}}{\binom{2n}{n}} = \frac{n}{n-k}\exp\Big(\sum_{j=1}^{k}\log(1-j/n)-\log(1+j/n)\Big) = \frac{n}{n-k}\exp\Big(-\sum_{j=1}^{k}\sum_{\substack{r=1\\r\text{ odd}}}^{\infty}\frac{2j^r}{rn^r}\Big), \quad (3)$$

an expression that will also be used later. It follows from it that

$$\frac{\binom{2n}{n-k}}{\binom{2n}{n}} \leq \frac{n}{n-k}\exp\Big(-\sum_{j=1}^{k}\frac{2j}{n}\Big) \leq \frac{n}{n-k}\exp\Big(-\frac{k^2}{n}\Big). \quad (4)$$

For small enough k, we can also obtain an asymptotic expansion. This will be discussed later.

4.2 The tails

In order to replace the binomial coefficient by a simpler expression that is amenable to a Mellin analysis, we first have to handle the tails of the sum. For this purpose, we require an explicit bound for the divisor function $\sigma(k)$ in form of a constant $A > 0$ such that

$$\sigma(k) \leq A \cdot k \cdot \log\log n \tag{5}$$

for $1 \leq k \leq n$ when $n \geq N$. Assume temporarily that $N \leq k \leq n$. Then, using an inequality due to Robin [10],

$$\frac{\sigma(k)}{k} \leq e^\gamma \log\log k + \frac{0.6483}{\log\log k} \leq e^\gamma \log\log n + \frac{0.6483}{\log\log N} \leq \left(e^\gamma + \frac{0.6483}{\log\log N}\right) \log\log n.$$

For $N = 10000$, we can choose $A = 52/25 \geq e^\gamma + 0.6483/\log\log N$, and we can let the computer verify that (5) also holds for $1 \leq k \leq n = N$.

For $k > \frac{n}{2}$, (4) combined with the fact that the binomial coefficients $\binom{2n}{n-k}$ are decreasing in k gives us

$$\frac{\binom{2n}{n-k}}{\binom{2n}{n}} \leq 2e^{-n/4}.$$

So we have

$$0 \leq \frac{1}{\binom{2n}{n}} \sum_{n/2 < k \leq n} k\sigma(k)(k^2 - 3n + 2)(2k^2 - n)\binom{2n}{n-k}$$

$$\leq \sum_{n/2 < k \leq n} 4Ak^6(\log\log n)e^{-n/4}$$

$$\leq (A\log\log n)n^7 e^{-n/4},$$

since it is easily verified that $\sum_{n/2 < k \leq n} k^6 \leq \frac{n^7}{4}$ for $n > 5$. Thus, the contribution of the sum in this range is

$$\frac{1}{\binom{2n}{n}} \sum_{n/2 < k \leq n} k\sigma(k)(k^2 - 3n + 2)(2k^2 - n)\binom{2n}{n-k} = B_{n \geq N}\left(\frac{52}{25}e^{-n/4}n^7 \log\log n\right). \tag{6}$$

Next, fix a constant $\alpha \in (\frac{1}{2}, \frac{3}{4})$; the precise value is in principle irrelevant if one is only interested in an asymptotic formula. However, for our computations with explicit error bounds it is advantageous to take a value close to $\frac{3}{4}$, so we choose $\alpha = \frac{7}{10}$. We bound the sum over all $k \in [n^\alpha, n/2]$. Here, we have

$$\frac{\binom{2n}{n-k}}{\binom{2n}{n}} \leq 2e^{-k^2/n}$$

by (4), thus (assuming that N is large enough that $k^2 \geq n^{2\alpha} \geq 3n$ whenever $n \geq N$, which we can easily verify for $N = 10000$ and $\alpha = 7/10$)

$$0 \leq \frac{1}{\binom{2n}{n}} \sum_{n^\alpha \leq k \leq n/2} k\sigma(k)(k^2 - 3n + 2)(2k^2 - n)\binom{2n}{n-k}$$

$$\leq \sum_{n^\alpha \leq k \leq n/2} 4Ak^6(\log\log k)e^{-k^2/n}$$

$$\leq (4A\log\log n) \sum_{n^\alpha \leq k \leq n/2} k^6 e^{-k^2/n}.$$

The function $t \mapsto t^6 e^{-t^2/n}$ is decreasing for $t \geq \sqrt{3n}$, thus in particular for $t \geq n^\alpha$ under our assumptions. This implies that (by a standard estimate for sums in terms of integrals)

$$\sum_{n^\alpha \leq k \leq n/2} k^6 e^{-k^2/n} \leq n^{6\alpha} e^{-n^{2\alpha-1}} + \int_{n^\alpha}^\infty t^6 e^{-t^2/n}\, dt.$$

The integral can be estimated by elementary means: for $T = n^\alpha$,

$$\int_T^\infty t^6 e^{-t^2/n}\, dt \leq \frac{1}{T} \int_T^\infty t^7 e^{-t^2/n}\, dt = \frac{n}{2T}\big(6n^3 + 6n^2 T^2 + 3nT^4 + T^6\big)e^{-T^2/n}.$$

For large enough $n \geq N$, this is negligibly small. This can be quantified with the help of some explicit computations with B-terms. We find that

$$\frac{1}{\binom{2n}{n}} \sum_{n^\alpha \leq k \leq n/2} \binom{2n}{n-k} k\sigma(k)(k^2-3n+2)(2k^2-n) = B_{n \geq N}\Big(\frac{25073}{5000}e^{-n^{2/5}} n^{\frac{9}{2}} \log\log n\Big). \quad (7)$$

4.3 Approximating the summands

So we are left with the sum over $k < n^\alpha$. Here, we can expand the exact expression in (3): this can be done by cutting the sum over r at some point (we choose the cutoff at $R = 9$) and estimating

$$0 \leq \sum_{j=1}^k \sum_{\substack{r=R \\ r \text{ odd}}}^\infty \frac{2j^r}{rn^r} \leq \sum_{j=1}^k \frac{2j^R}{Rn^R(1 - j^2/n^2)}$$

by means of a geometric series (observe that the factor $1 - j^2/n^2$ in the denominator stems from the fact that we are only summing over odd r), and then further

$$\sum_{j=1}^k \frac{2j^R}{Rn^R(1 - j^2/n^2)} \leq \frac{2}{Rn^R(1 - k^2/n^2)}\Big(k^R + \int_0^k t^R\, dt\Big)$$

$$= \frac{2}{Rn^R(1 - k^2/n^2)}\Big(k^R + \frac{k^{R+1}}{R+1}\Big)$$

$$\overset{R=9}{=} B_{n \geq N}\Big(\Big(\frac{239}{10000}k^{10} + \frac{2223}{10000}k^9\Big)n^{-9}\Big),$$

followed by a Taylor expansion of the exponential multiplied with the expansion of $\frac{n}{n-k}$, cf. (3). The full and sufficiently precise asymptotic expansion can be found in our auxiliary SageMath notebook. It reads

$$\frac{\binom{2n}{n-k}}{\binom{2n}{n}} = e^{-k^2/n}\Big(1 - \frac{k^4 + k^2}{6n^3} + \frac{k^8}{72n^6} + \frac{k^2}{n^2} - \frac{k^{12}}{1296n^9} - \frac{3k^6}{20n^5} + \frac{k^{16}}{31104n^{12}}$$

$$+ \cdots + B_{n \geq N}\Big(\frac{k^{24}}{10000n^{18}}\Big) + B_{n \geq N}\Big(\frac{9k^{21}}{10000n^{16}}\Big) + \cdots \Big),$$

where the summands are ordered based on their individual upper growth bound (found from substituting $k = n^\alpha$). The ellipses $\cdots$ indicate terms that are left out as the expression would otherwise be very long. If it were required, this expansion could also be made more precise. Let us now split the expression inside the brackets: let $S(n,k)$ denote the sum of all "exact" terms, and $S_B(n,k)$ the sum of all B-terms. We want to evaluate

$$\sum_{1 \leq k < n^\alpha} k\sigma(k)\big(S(n,k) + S_B(n,k)\big)(k^2 - 3n + 2)(2k^2 - n)e^{-k^2/n}.$$

Let us first deal with the error estimate: since $|(k^2 - 3n + 2)(2k^2 - n)| \leq 2k^4 + 3n^2$ for all k and n, it suffices to bound

$$\sum_{1 \leq k < n^\alpha} k S_B(n, k) \sigma(k)(2k^4 + 3n^2)e^{-k^2/n}$$

$$\leq (A \log \log n) \sum_{1 \leq k < n^\alpha} S_B(n, k)(2k^6 + 3n^2k^2)e^{-k^2/n}$$

$$\leq (A \log \log n) \sum_{k \geq 1} S_B(n, k)(2k^6 + 3n^2k^2)e^{-k^2/n}.$$

For a positive function $f(t)$ that is increasing up to some maximum t_0 and decreasing thereafter, it is well known that $\sum_{k \geq 1} f(k) \leq f(t_0) + \int_0^\infty f(t)\,dt$. This can now be applied to $t \mapsto t^j e^{-t^2/2}$ to find, with the help of computer algebra,

$$\sum_{1 \leq k < n^\alpha} k S_B(n, k)\sigma(k)(2k^4 + 3n^2)e^{-k^2/n} = B_{n \geq N}\left(\frac{146718899}{10000}\sqrt{n}\log\log n\right). \tag{8}$$

While this error is not quite as small as those collected so far, for $n = 10000$ it is still only about 26.1% of the eventual main term.

Now we can consider the remaining sum

$$\sum_{1 \leq k < n^\alpha} k\sigma(k)S(n, k)(k^2 - 3n + 2)(2k^2 - n)e^{-k^2/n}.$$

To this end, we first add back the terms with $k \geq n^\alpha$ and estimate their sum.

For $k \leq n^{3/4}$ the expansion in $S(n, k)$ can be bounded above, $S(n, k) \leq c_1 \approx 4.372$, and for $k \geq n^{3/4}$ we have $S(n, k) \leq k^{20}/(10000n^{15})$. As for the other factors in our summands, we can bound $(k^2 - 3n + 2)(2k^2 - n)$ from above by $2k^4$. For an estimate of $\sigma(k)$ we use (5) in the range $k < n^{3/4}$, and for the remaining case of $k \geq n^{3/4}$ we use the well-known weaker bound $\sigma(k) \leq k^2$. This leaves us with

$$\sum_{n^\alpha \leq k < n^{3/4}} k\sigma(k)S(n, k)(k^2 - 3n + 2)(2k^2 - n)e^{-k^2/n}$$

$$\leq 2Ac_1 \log\log n \sum_{n^\alpha \leq k < n^{3/4}} k^6 e^{-k^2/n} = B_{n \geq N}\left(\frac{12553}{5000}e^{-n^{2/5}}n^{19/4}\log\log n\right), \tag{9}$$

and

$$\sum_{k \geq n^{3/4}} k\sigma(k)S(n, k)(k^2 - 3n + 2)(2k^2 - n)e^{-k^2/n}$$

$$\leq \frac{2}{10000\,n^{15}} \sum_{k \geq n^{3/4}} k^{27} e^{-k^2/n} = B_{n \geq N}\left(\frac{3}{2000}e^{-n^{1/2}}n^{11/2}\right), \tag{10}$$

where the sums have been bounded using the same integral estimate as before.

4.4 Mellin transform

Having estimated all error terms related to pruning and completing the tails of the sum, we now want to evaluate

$$\sum_{k \geq 1} k\sigma(k)S(n, k)(k^2 - 3n + 2)(2k^2 - n)e^{-k^2/n}. \tag{11}$$

This sum is a linear combination of sums of the form

$$\sum_{k \geq 1} k^a n^b \sigma(k) e^{-k^2/n}. \tag{12}$$

In the precision chosen in our accompanying SageMath worksheet, there are 121 such summands, to be precise. Set $t = n^{-1}$, refer to the sum in (12) as $g_{a,b}(t)$ and let $d_{a,b}$ denote the coefficients such that the sum in (11) can be written as $\sum_{a,b} d_{a,b} g_{a,b}(t)$. The Mellin transform (see [4] for a general reference) of $g_{a,b}(t)$ is given by

$$g^*_{a,b}(s) = \int_0^\infty t^{s-1} \sum_{k \geq 1} k^a t^{-b} \sigma(k) e^{-k^2 t} \, dt = \zeta(2s - 2b - a - 1)\zeta(2s - 2b - a)\Gamma(s - b).$$

By the Mellin inversion formula, the original function $g_{a,b}(t)$ can be recovered from its transform via

$$g_{a,b}(t) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \zeta(2s - 2b - a - 1)\zeta(2s - 2b - a)\Gamma(s - b)t^{-s} \, ds$$

for $c > \frac{a}{2} + b + 1$. We may shift the line of integration further left as long as we collect all corresponding residues. In a first step, we shift the line of integration to $c = 3/4$. While in some summands poles occur as far right as $s = 7/2$, a straightforward computation reveals that, as mentioned in the introduction of this article, nontrivial cancellations take place: after summing all contributions, non-zero residues in the half-plane where $\Re(s) \geq 3/4$ can only be found for $s = 1$ and $s = 2$, where we collect a contribution of

$$\sum_{s_0 \in \{1,2\}} \sum_{a,b} d_{a,b} \operatorname{Res}(g_{a,b}(s), s = s_0) = -\frac{1}{8t^2} + \frac{1}{24t} = -\frac{n^2}{8} + \frac{n}{24}, \tag{13}$$

which proves the asymptotic main term given in (2).

We now need to determine an explicit error bound for these shifted integrals. To do so, we investigate, individually for each summand, how far we can shift the line of integration to the left (in half-integer units) until $\Re(s) = c_{a,b}$ without collecting any further residues.

In a central region of $c_{a,b} + iw$ for $|w| \leq 100$ we use rigorous integration via interval arithmetic to determine the value of the shifted integrals. Outside, for $|w| > 100$, we determine a suitable upper bound of the integrand. For $\Gamma(c_{a,b} + iw)$ where $c_{a,b} > 0$ we use [3, (5.6.9)], and when $c_{a,b} < 0$ we first shift the argument to the right via the functional equation $\Gamma(s) = \frac{1}{s}\Gamma(s + 1)$ and then proceed as before. For $\zeta(c_{a,b} + iw)$ we bound the modulus from above by $\zeta(3/2)$ if $c_{a,b} \geq 3/2$. When $c_{a,b} \leq -1/2$ we first apply the reflection formula [3, (25.4.1)]; the resulting factors can all be estimated directly. For the special case of $c_{a,b} = 1/2$ we use the bound proved by Hiary, Patel, and Yang in [9, Theorem 1.1] to obtain

$$|\zeta(1/2 + iw)| \leq 0.618 \, t^{1/6} \log t \leq 0.618 \, t^{1/2}$$

for $t \geq 100$. Letting a computer collect and combine all of these contributions then yields

$$\left| \sum_{a,b} \frac{1}{2\pi i} \int_{3/4-i\infty}^{3/4+i\infty} g^*_{a,b}(s)t^{-s} \, ds \right|$$

$$\leq \frac{1}{2\pi} \sum_{a,b} n^{c_{a,b}} \int_{-\infty}^{\infty} |g^*_{a,b}(c_{a,b} + iw)| \, dw = B_{n \geq N}\left(\frac{406531}{100} n^{3/4}\right). \tag{14}$$
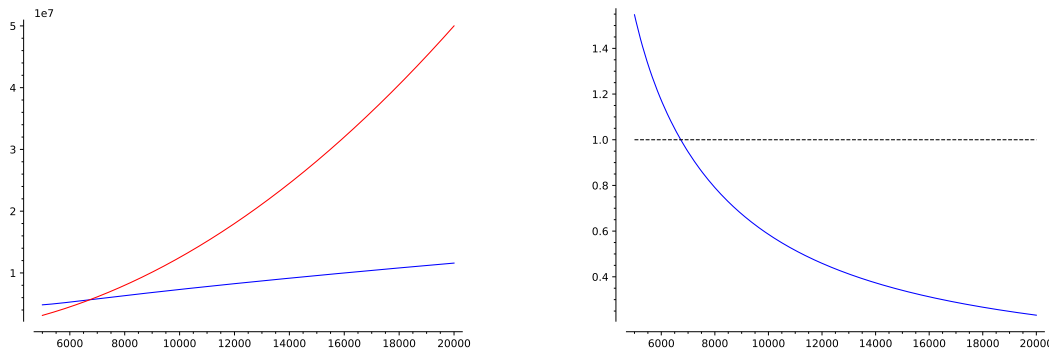
Figure 1 Comparison of the absolute value of the asymptotic main term $-n^2/8 + n/24$ (red) against the accumulated total error (blue) on the left. The right plot depicts the ratio of the error bound to the main term.

5   Conclusion

Throughout Section 4 we have accumulated several explicit error terms. They are given in (6), (7), (8), (9), (10), and (14). Combining them using crude estimates such as $\log \log n \leq n^{1/10}$ for $n \geq N$ proves the following theorem.

▶ **Theorem 2.** *For $n \geq 10000$, the binomial sum $F(n)$ satisfies the asymptotic formula*

$$F(n) = \binom{2n}{n}\left(-\frac{n^2}{8} + \frac{n}{24} + B_{n \geq N}\left(\frac{38755553}{5000}n^{3/4} \right) \right).$$

Observe that for $n = 10000$ the certified error is already only approximately 62.1% of the absolute value of the exact main term. Together with the direct verification for $5 \leq n < N$ this settles Problem 1. See Figure 1 for an illustration of the behavior of the total error compared to the main term.

To conclude this paper, we briefly discuss an alternative approach that was kindly pointed out to us by a referee. Recall that the task is to prove the inequality (1), i.e.,

$$F(n) = \sum_{k=1}^{n} k\sigma(k)(k^2 - 3n + 2)(2k^2 - n)\binom{2n}{n-k} < 0.$$

Now one can use the well-known generating function identity

$$\sum_{n=k}^{\infty} \binom{2n}{n-k}x^n = x^k \sum_{m=0}^{\infty} \binom{2m+2k}{m}x^m = x^k \frac{1}{\sqrt{1-4x}}C(x)^{2k},$$

where $C(x) = \frac{1-\sqrt{4x}}{2x}$ is the generating function for the Catalan numbers, see e.g. [6, (5.72)]. This gives an expression for $F(n)$ in terms of coefficients of functions involving $C(x)$. Specifically, we have

$$F(n) = 3n^2 \sum_{k=1}^{n} k\sigma(k)\binom{2n}{n-k} - n\sum_{k=1}^{n} k(7k^2+2)\sigma(k)\binom{2n}{n-k}$$

$$+ \sum_{k=1}^{n} 2k^3(k^2+2)\sigma(k)\binom{2n}{n-k}$$

$$= 3n^2[x^n]\frac{1}{\sqrt{1-4x}}\sum_{k=1}^{\infty}k\sigma(k)(xC(x)^2)^k$$

$$- n[x^n]\frac{1}{\sqrt{1-4x}}\sum_{k=1}^{\infty}k(7k^2+2)\sigma(k)(xC(x)^2)^k$$

$$+ [x^n]\frac{1}{\sqrt{1-4x}}\sum_{k=1}^{\infty}2k^3(k^2+2)\sigma(k)(xC(x)^2)^k$$

$$= 3n^2[x^n]\frac{1}{\sqrt{1-4x}}f_1(H(x)) - n[x^n]\frac{1}{\sqrt{1-4x}}f_2(H(x)) + [x^n]\frac{1}{\sqrt{1-4x}}f_3(H(x)),$$

where $H(x) = xC(x)^2 = \frac{1-2x-\sqrt{1-4x}}{2x}$, and f_1, f_2, f_3 are given by the series

$$f_1(z) = \sum_{k=1}^{\infty}k\sigma(k)z^k, \quad f_2(z) = \sum_{k=1}^{\infty}k(7k^2+2)\sigma(k)z^k, \quad f_3(z) = \sum_{k=1}^{\infty}2k^3(k^2+2)\sigma(k)z^k.$$

At the singularity $x = \frac{1}{4}$, $H(x)$ has the expansion

$$1 - 2\sqrt{1-4x} + 2(1-4x) + \cdots,$$

so we need the behavior of $f_1(z), f_2(z), f_3(z)$ around $z = 1$. This can be determined by means of the Mellin transform: setting $z = e^{-t}$, we obtain for instance

$$f_1(e^{-t}) = \sum_{k=1}^{\infty}k\sigma(k)e^{-kt},$$

whose Mellin transform is $\Gamma(s)\zeta(s-1)\zeta(s-2)$. Applying the inverse Mellin transform in the same way as in Section 4.4 (though now with complex parameter t) yields

$$f_1(e^{-t}) = \frac{\pi^2}{3t^3} - \frac{1}{2t^2} + O(t^K)$$

for any positive real K. This and analogous asymptotic formulas for f_2 and f_3 give us the behavior of $f_1(H(x))$, $f_2(H(x))$ and $f_3(H(x))$ at the dominant singularity $\frac{1}{4}$, from which the asymptotic formula (2) can be obtained by means of contour integration and singularity analysis. Carrying all of this out with explicit error terms comes with its own challenges, though, as one now has to deal with complex asymptotics.

References

1 Miklós Bóna and Elijah DeJonge. Pattern avoiding permutations with a unique longest increasing subsequence. *Electron. J. Combin.*, 27(4):Paper No. 4.44, 11, 2020. `doi:10.37236/9506`.

2 Mireille Bousquet-Mélou and Yann Ponty. Culminating paths. *Discrete Math. Theor. Comput. Sci.*, 10(2):125–152, 2008.

3 *NIST Digital Library of Mathematical Functions.* `https://dlmf.nist.gov/`, Release 1.1.12 of 2023-12-15. F. W. J. Olver, A. B. Olde Daalhuis, D. W. Lozier, B. I. Schneider, R. F. Boisvert, C. W. Clark, B. R. Miller, B. V. Saunders, H. S. Cohl, and M. A. McClain, eds. URL: `https://dlmf.nist.gov/`.

4 Philippe Flajolet, Xavier Gourdon, and Philippe Dumas. Mellin transforms and asymptotics: Harmonic sums. *Theor. Comput. Sci.*, 144(1-2):3–58, 1995. `doi:10.1016/0304-3975(95)00002-E`.

5 Philippe Flajolet and Robert Sedgewick. *Analytic combinatorics.* Cambridge University Press, Cambridge, 2009. `doi:10.1017/CBO9780511801655`.

6 Ronald L. Graham, Donald E. Knuth, and Oren Patashnik. *Concrete mathematics: a foundation for computer science.* Amsterdam: Addison-Wesley Publishing Group, 2nd ed. edition, 1994.

7 Benjamin Hackl, Clemens Heuberger, and Daniel Krenn. Asymptotic expansions in SageMath. `http://trac.sagemath.org/17601`, 2015. module in SageMath 6.10.

8 Benjamin Hackl, Clemens Heuberger, Helmut Prodinger, and Stephan Wagner. Analysis of bidirectional ballot sequences and random walks ending in their maximum. *Ann. Comb.*, 20(4):775–797, 2016. `doi:10.1007/s00026-016-0330-0`.

9 Ghaith A. Hiary, Dhir Patel, and Andrew Yang. An improved explicit estimate for $\zeta(1/2 + it)$. *J. Number Theory*, 256:195–217, 2024. `doi:10.1016/j.jnt.2023.09.003`.

10 G. Robin. Grandes valeurs de la fonction somme des diviseurs et hypothèse de Riemann. *J. Math. Pures Appl. (9)*, 63:187–213, 1984.

11 The SageMath Developers. SageMath, March 2023. `doi:10.5281/zenodo.593563`.

Multicoloured Hardcore Model: Fast Mixing and Its Applications as a Scheduling Algorithm

Sam Olesker-Taylor ✉ 🏠 ⬤

Department of Statistics, University of Warwick, UK

―――― **Abstract** ――――――――――――――――――――――――――――――――――――――

In the hardcore model, certain vertices in a graph are *active*: the active vertices must form an *independent set*. We extend this to a *multicoloured* version: instead of simply being active or not, the active vertices are assigned a colour; active vertices *of the same colour* must not be adjacent.

This models a scenario in which two neighbouring resources may *interfere* when active – eg, short-range radio communication. However, there are multiple *channels* (colours) available; they only interfere if both use the *same* channel. Other applications include routing in fibreoptic networks.

We analyse Glauber dynamics. Vertices update their status at random times, at which a uniform colour is proposed: the vertex is assigned that colour if it is available; otherwise, it is set inactive.

We find conditions for *fast mixing* of these dynamics. We also use them to model a queueing system: vertices only serve customers in their queue whilst active. The mixing estimates are applied to establish positive recurrence of the queue lengths, and bound their expectation in equilibrium.

1 Introduction and Main Results

We extend the hardcore model, used for sampling independent sets, to a multicoloured version. Given a graph $G = (V, E)$, our objective is to colour a subset $U \subseteq V$ of the vertices such that if $u, u' \in U$ satisfy $\{u, u'\} \in E$, then u and u' are painted with different colours. If there is only one colour, then this condition requires that there is no pair of mutually adjacent vertices. This is the definition of an *independent set*, so we recover the usual *hardcore model*.

We allow an arbitrary number $K \in \mathbb{N}$ of possible colours. If we required *all* vertices to be selected – ie, $U = V$ – then the condition is that no edge in the graph is *monochromatic*: the endpoints must receive different colours. We thus recover the *proper colouring* model. Our model models these two, sampling a properly coloured subset of vertices, or subgraph.

The motivation for this model comes from a desire for a decentralised (and randomised) algorithm for resource sharing. Two examples of this are short-range radio communication, where nearby agents on the same frequency interfere, and routing in fibreoptic networks. Both K and G are given parameters, depending on the particular engineering set-up.

A popular method for sampling proper colourings or independent sets is via *Glauber dynamics*. Our main result is on the *mixing time* of Glauber dynamics for the multicoloured hardcore model, defined precisely below. We then use the system to model a *queueing system*.

- Customers (eg, data packets) arrive to vertices at some (vertex-dependent) rate.
- Coloured vertices are *active*: they serve their customers at some (vertex-dependent) rate.
- Uncoloured vertices are *inactive*: they do not serve, but their queue can still grow.

We apply the mixing-time result to control the queue lengths, under certain conditions.

A   Glauber Dynamics

Let $G = (V, E)$ be a graph and $K \in \mathbb{N}$. Let $n := |V|$; write $[K]_0 := \{0, 1, ..., K\}$. The state space Ω of the system is a subset of configurations $[K]_0^V = \{(\omega_v)_{v \in V} \mid \omega_v \in [K]_0 \, \forall \, v \in V\}$.

▶ **Definition A.1** (State Space). *Let $\Omega := \{\omega \in [K]_0^V \mid \omega \text{ is proper}\}$, where $\omega \in [K]_0^V$ is proper if*

$$\omega_u \neq \omega_v \quad \text{whenever} \quad \{u, v\} \in E \quad \text{and} \quad \omega_u + \omega_v > 0.$$

In a proper configuration, the colour of one vertex must be different to that of all its neighbours, except that colour 0 is exempt from this condition. We view colour 0 as inactive. A configuration is proper if the subgraph induced by its active vertices is properly coloured.

▶ **Definition A.2** (Glauber-Type Dynamics). *Let $\boldsymbol{\lambda} = (\lambda_v) \in (0, \infty)^V$ and $\boldsymbol{p} = (p_v) \in [0, 1]^V$. We analyse the following continuous-time Markov chain, which we denote $\mathsf{MCH}_\Omega(\boldsymbol{\lambda}, \boldsymbol{p})$.*
- *Choose vertex $v \in V$ to update at rate λ_v, simultaneously over all vertices.*
- *Once vertex $v \in V$ is chosen, toss a p_v-biased coin: $C \sim \mathrm{Bern}(p_v)$.*
  - *If $C = 1$, then choose a (non-zero) colour $k \in [K]$ uniformly at random. If colour k is available for v – ie, no neighbour of v has colour k – then paint v with colour k.*
  - *Otherwise, deactivate v – ie, colour 0 – whether or not it was active before.*

Denote the equilibrium distribution by π. The equilibrium active time, or service rate, is

$$s_v := \sum_{\omega \in \Omega : \omega_v \neq 0} \pi(\omega) \quad \text{for} \quad v \in V.$$

The usual Glauber dynamics for proper colourings proposes a colour chosen uniformly *amongst available colours.* However, this requires whoever is making the colour choice to know which colours are available for that vertex. This is unreasonable in the context of routing algorithms in fibreoptic networks, for example. It is often much faster to check if a single proposed colour is available than to find out which colours are available.

Our main theorem establishes *fast mixing.* First, we define mixing times precisely.

▶ **Definition A.3** (Mixing Times). *The total-variation distance between distributions μ and π is*

$$\|\mu - \pi\|_{\mathsf{TV}} := \max_{A \subseteq \Omega} |\mu(A) - \pi(A)| = \tfrac{1}{2} \sum_{\omega \in \Omega} |\mu(\omega) - \pi(\omega)|.$$

The mixing time of a Markov chain $X = (X^t)_{t \geq 0}$ on Ω with invariant distribution π is

$$t_{\mathrm{mix}}(\varepsilon) := \inf\{t \geq 0 \mid \max_{x \in \Omega} \|\mathbb{P}_x[X^t \in \cdot] - \pi\|_{\mathsf{TV}} \leq \varepsilon\} \quad \text{for} \quad \varepsilon \in (0, 1).$$

▶ **Theorem A** (Fast Mixing). *Suppose that there exists $\beta > 0$ such that*

$$\tfrac{1}{K} \sum_{u \in V : \{u, v\} \in E} p_u \lambda_u / \lambda_v \leq 1 - \beta \quad \text{for all} \quad v \in V.$$

Let $\lambda_{\min} := \min_{v \in V} \lambda_v$. If $X, Y \sim \mathsf{MCH}_\Omega(\boldsymbol{\lambda}, \boldsymbol{p})$, then

$$\max_{x, y \in \Omega} \left\| \mathbb{P}_x[X^t \in \cdot] - \mathbb{P}_y[Y^t \in \cdot] \right\|_{\mathsf{TV}} \leq \min\{2n e^{-\beta \lambda_{\min} t}, \, 1\}.$$

In particular,

$$t_{\mathrm{mix}}(\varepsilon) \leq (\beta \lambda_{\min})^{-1} \log(2n/\varepsilon) \quad \text{for all} \quad \varepsilon \in (0, 1).$$

Remark A (Fast-Mixing Condition). The condition in Theorem A arises from requiring the Wasserstein distance between X and Y to contract in a single step, uniformly. Distance is measured vertex-wise: $d(x, y) \coloneqq \sum_{v \in V} \mathbf{1}\{x_v \neq y_v\}$ for $x, y \in \Omega$. Namely, we prove that if configurations x and y differ only in that vertex v is active in one but not the other, then

$$\tfrac{d}{dt}\mathbb{E}_{(x,y)}[d(X^t, Y^t)]\big|_{t=0} \leq \lambda_v\big(\tfrac{1}{K}\sum_{u \in V:\{u,v\}\in E} p_u \lambda_u/\lambda_v - 1\big)$$

under some coupling. The given condition ensures this is negative, uniformly in $x, y \in \Omega$. A standard application of *path coupling* [6] extends this uniform contraction to all $x, y \in \Omega$. △

The graph G and number K of colours are given by the application. In contrast, the parameters $(\lambda_v, p_v)_{v \in V}$ may be chosen by the operator. There are good heuristics for taking

$$\lambda_v \propto d_v \quad \text{and} \quad p_v \propto (K/d_v) \wedge 1.$$

In short, high-degree nodes have more impact on their neighbours, and hence should be updated faster: so, take $\lambda_v \propto d_v$. Further, if v is active with probability p'_v, then it remove a total of $p'_v d_v$ colour choices in expectation (from its neighbours). There are K colours, so vertices shouldn't remove more than K in expectation: hence, $s_v d_v \propto K$; so, take $p_v \propto K/d_v$.

We work in continuous time, so scaling all the rates λ inversely scales the mixing time. We choose the normalisation $\sum_{v \in V} \lambda_v = n$; so, vertices each update at rate 1 on average.

▶ **Corollary A** (Heuristic-Driven Choice). *Suppose that $\lambda_v = d_v/\bar{d}$ and $p_v \leq \tfrac{2}{3}K/d_v$ for all $v \in V$, where $\bar{d} \coloneqq \tfrac{1}{n}\sum_{v \in V} d_v$ is the average degree. Let $X, Y \sim \mathsf{MCH}_\Omega(\boldsymbol{\lambda}, \boldsymbol{p})$. Then,*

$$\max_{x,y \in \Omega}\big\|\mathbb{P}_x[X^t \in \cdot] - \mathbb{P}_y[Y^t \in \cdot]\big\|_{\mathsf{TV}} \leq \min\big\{2ne^{-(\delta/\bar{d})t/3}, 1\big\},$$

where $\delta \coloneqq \min_{v \in V} d_v$. In particular,

$$t_{\mathrm{mix}}(\varepsilon) \leq 3(\bar{d}/\delta)\log(2n/\varepsilon) \quad \text{for all} \quad \varepsilon \in (0, 1).$$

It is standard, or, at least, very common, in the hardcore-model ($K = 1$) literature to require $p_v = p < 1/\Delta$ for all v, where $\Delta \coloneqq \max_{v \in V} d_v$ is the maximum degree; see, eg, [2, 15, 7] or [14, Theorem 5.9]. We take more care, requiring only $p_v < K/d_v$ for each v; [10] have a similar improvement, but restricted to the usual hardcore model ($K = 1$).

A consequence of requiring $p_v = p < 1/\Delta$ is that the mixing time is often proportional to Δ. Ours is proportional to $\bar{d}/d_{\min}$, which is often significantly smaller.

The bound $p < 1/\Delta$ is natural, up to a factor e. Indeed, for the (usual) hardcore model, it has been known since Kelly [13] that the infinite Δ-regular tree has a critical threshold at $p_c(\Delta) \approx e/\Delta$, for large Δ: the corresponding Gibbs distribution is unique if and only if $p < p_c(\Delta)$. When $p < p_c(\Delta)$, known as the *uniqueness regime* to physicists, the "influence" of one vertex on another decays exponentially in their relative distance. On the other hand, long-rage correlations persist when $p > p_c(\Delta)$. See [1, §1.2] for more discussion on this.

Based on this, it appears that we should be able to only require $p_v \leq (1 - \eta)eK/d_v$ and still obtain fast mixing. This would be a natural extension of the critical threshold: $p_c(\Delta, K) \coloneqq eK/\Delta$. We demonstrate this via some simulations at the end of the paper.

We also investigate the proportion of time that vertices are active in equilibrium.

▶ **Proposition A** (Equilibrium Service Rates). *Suppose that $p_v \leq \tfrac{1}{3}K/\tilde{d}_v$ for all $v \in V$, where $\tilde{d}_v \coloneqq \max\{d_u \mid u \sim v \text{ or } u = v\}$ is the maximal degree in the neighbourhood of $v \in V$. Then,*

$$\tfrac{1}{3}p_v \leq s_v \leq p_v \quad \text{for all} \quad v \in V.$$

Our proof is quite flexible, allowing more general p_v. We discuss how to generalise it, and tighten the bounds, after its proof. Again, we expect that we only need $p_v \leq (1 - \eta)eK/d_v$.

B Queueing Network

Our next results concerns queue length in a network. The proof relies on fast mixing.

▶ **Definition B** (Queueing Network). *Let $\boldsymbol{\lambda}, \boldsymbol{\nu}, \boldsymbol{\mu} \in (0, \infty)^V$ and $\boldsymbol{p} \in [0,1]^V$. Let $X \sim \mathsf{MCH}_\Omega(\boldsymbol{\lambda}, \boldsymbol{p})$. The state space of the queueing network is $\mathbb{N}^V$. For $q \in \mathbb{N}^V$ and $v \in V$, let*

$$q_u^{v,\pm} := q_u \pm \mathbf{1}\{u = v\} \quad for \quad u \in V;$$

that is, $q^{v,\pm}$ adds/removes one from the v-th queue. The transition rates given $X = x$ are

$$q \to \begin{cases} q^{v,+} & at \ rate \quad \nu_v \\ q^{v,-} & at \ rate \quad \mu_v \mathbf{1}\{x_v \neq 0\} \end{cases} \quad for \ each \quad v \in V;$$

that is, the v-th queue has arrivals at rate ν_v always and services at rate μ_v provided v is active. We denote the law of this queueing network by $\mathsf{QMCH}_\Omega(\boldsymbol{\lambda}, \boldsymbol{p}; \boldsymbol{\nu})$.

We show that the queues are jointly *positive recurrent* – ie, the expected time until all queues are simultaneously empty is finite – under the fast-mixing conditions of Theorem A and the assumption that the arrival rate ν_v is smaller than the equilibrium service rate s_v.

▶ **Theorem B** (Stable Queues). *Suppose that there exists $\beta > 0$ such that*

$$\tfrac{1}{K} \sum_{u \in V : \{u,v\} \in E} p_u \lambda_u / \lambda_v \leq 1 - \beta \quad for \ all \quad v \in V.$$

Suppose also that $\nu_v < s_v$ for all $v \in V$. If $Q \sim \mathsf{QMCH}_\Omega(\boldsymbol{\lambda}, \boldsymbol{p}; \boldsymbol{\nu})$, then Q is positive recurrent:

$$\tau := \inf\{t \geq 0 \mid Q^t = 0\} \quad satisfies \quad \mathbb{E}_q[\tau] < \infty \quad for \ all \quad q \in \mathbb{N}^V.$$

Moreover, if Q^0 is in equilibrium, then, writing $\lambda_{\min} := \min_{v \in V} \lambda_v$,

$$\mathbb{E}[Q_v^0] \leq \frac{6n \log(2n/e)}{\beta \lambda_{\min}(s_v - \nu_v)^2} \quad for \ all \quad v \in V.$$

We now evaluate this under the heuristic-driven choice from Corollary A.

▶ **Corollary B** (Heuristic-Driven Choice). *Suppose that $\lambda_v = d_v / \bar{d}$ and $p_v \leq \frac{2}{3} K / d_v$ for all $v \in V$, where $\bar{d} := \frac{1}{n} \sum_{v \in V} d_v$ is the average degree. Let $\delta := \min_{v \in V} d_v$. Suppose also that $\nu_v < s_v$ for all $v \in V$. Let $Q \sim \mathsf{QMCH}_\Omega(\boldsymbol{\lambda}, \boldsymbol{p}; \boldsymbol{\nu})$. Then, in equilibrium,*

$$\mathbb{E}[Q_v^0] \leq \frac{18 \bar{d} n \log(2n/e)}{\delta (s_v - \nu_v)^2} \quad for \ all \quad v \in V.$$

A related result was proved by Jiang et al [10] for the usual hardcore model (one colour). Also, they restrict to the special case $p_v = p < 1/\Delta$, where Δ is the maximum degree.

2 Motivation and Related Work

Fibreoptic Routing Application

Our original motivation was to create a *fully decentralised* random access scheme for resource sharing in fibreoptic routing networks. There, nodes are connected by *links*, and they communicate with each other along *routes*, which are sequences of links. Multiple routes may share a subset of links; such routes *interfere*. Each link has a collection of *frequencies* available.

A naive approach has the source node send the data to the first intermediary node on the route, along with instructions of where to send on. That intermediary node processes the data and sends it onto the next node. This continues until the data reaches its target destination.

It is possible for different frequencies to be used along the route, due to the intermediary processing. When checking whether it is possible for a certain collection of routes to be active simultaneously, it is enough to check that no individual link is overloaded. However, the intermediary processing adds overhead. If the time it takes to transmit the data along the link is larger than the processing time, then the overhead is unimportant. However, in fibreoptic networks, data is sent along links extremely quickly, and the processing overhead becomes the performance bottleneck.

Instead of processing and resending the data at an intermediary node, an *optical switch* is configured. This switch is like a prism: light coming from a single source is sent in different directions, depending on its colour. This allows a *light path* to be set up, removing the processing overhead; however, the *same* frequency must be used throughout the entire route.

The difficulty is in choosing the frequency (colour) of the light path. Now, it is not enough to simply check that each link is not overloaded marginally, as the colours are correlated. In the set-up of the multicoloured hardcore model, the vertices correspond to routes, and two routes (vertices) are adjacent, forming an edge, if they interfere – ie, share a link. Certainly, not all routes will be able to be on simultaneously; an access scheme must be devised.

I originally learnt of this model from a talk by Walker [22] at the *Algorithms and Software for Quantum Computers* event at the Isaac Netwon Institute. There, the speaker was looking to quantum computation for solutions. I, as a probabilist, took a randomised approach.

The multicoloured hardcore model has the significant benefit of *decentralisation*. All decisions made can be made by the individual vertices, without any need for synchronisation or knowledge of the state of the other routes. A vertex can even request a light path blindly [11]: the path is set up if it does not conflict with any other already-active paths; otherwise, an error is returned to the initiator. Moreover, optical-switch reconfiguration is fast and easy.

The hardcore model is a popular and well-studied model for random access schemes where there is only a single frequency: ON or OFF. A toy model for this is local radio communication: vertices represent pairs of agents who wish to communicate; nearby pairs of agents cannot communicate simultaneously. Quite separately, Glauber dynamics are used to sample proper colourings on a graph. It seems natural to combine these two, yielding a multicoloured hardcore model which can model more complex interference situations, such as when multiple independent radio frequencies are available. However, to the best of my knowledge, this multicoloured hardcore set-up has not been studied before in the context of routing.

Multihop Wireless Networks

Another application of this type of random routing scheme is to *multihop wireless networks*. In cellular and wireless local area networks, wireless communication only occurs on the last link between a base station and the wireless end system. In *multihop* wireless networks, there are one or more intermediate nodes along the path; these receive and forward packets via the wireless links. There are several benefits to the multihop approach, including extneded coverage and improved connectivity, higher transfer rates and the avoidance of wide deployment of cables. Unfortunately, protocols, particularly those for routing, developed for fixed or cellular networks, or the Internet, are not optimal for these, more complicated, multihop wireless networks; see, eg, [5].

A highly prominent example of multihop wireless networks is in the development and deployment of 5G cellular networks [21]. Conventional cellular networks employ well-planned deployment of tower-mounted base stations. They are undergoing a fundamental change to deployment of smaller base stations. Multihop relaying can be instrumental for tation. See [8, §4.1] for more details, from which part of this paragraph was paraphrased.

A multihop network with a single transmission frequency falls precisely into the framework of the (usual) hardcore model. Glauber dynamics is a powerful tool used to generated randomised, approximate solutions to combinatorially difficult problems. Moreover, it often has natural decentralised implementation. It has already been used in the past to design and analyse distributed scheduling algorithms for multihop wireless networks; see, particularly, [10, 4], from which this paragraph was paraphrased, as well as [16, 17, 3, 9, 19].

Multihop wireless networks with *multiple* transmission frequencies correspond precisely to our model. To the best of our knowledge, it has received little attention. However, with technological and engineering advances, it may become an important extension in the future.

(A)synchronicity

One aspect to point out is our lack of synchronicity: we use *continuous time*, so sites update one at a time. In practice, engineering implementations often prefer synchronised updates. This is the case in [10], where the (usual) hardcore model is analysed and an *independent set* of vertices – ie, a set of vertices with no edges between them – is updated simultaneously. It is crucial that it is an *independent* set: the changes to one vertex in the set do not affect the other vertices, and the updates can be done independently, in a parallel, distributed manner.

The (independent) set of vertices still needs to be chosen in each step. In [10], the authors simply prescribe a distribution q over the collection of all independent sets; no comment is made on *how* to sample one. In principle, this distribution is very complicated, and perhaps even needs approximating – eg, via Glauber dynamics for the (usual) hardcore model.

The path coupling technique that we use, and is used in [10], is robust to parallel updates, provided one update does not affect the others – as for updating an independent set of vertices. If N is the expected size of the independent set chosen – ie, $N := \mathbb{E}_{S \sim q}[|S|]$ – then the mixing bound behaves as if time is sped up by a factor N. We consider single-site, continuous-time updates for simplicity; but, our analysis extends to the parallel set-up, too.

Spin Systems in Statistical Mechanics

Spin systems are widely studied in statistical mechanics, crossing combinatorics, probability and physics: these involve a graph $G = (V, E)$ and a discrete set $\mathcal{K}$ of *spins*; each vertex $v \in V$ is assigned a spin $k \in \mathcal{K}$. Adjacent vertices interact with each other. A zoo of examples of spin systems is discussed extensively in the very recent paper by Peled and Spinka [18].

- In proper colourings, $\mathcal{K} = \{1, ..., K\}$ and the constraint is hard: adjacent vertices *must not* have the same colour. The hardcore model is similar with $\mathcal{K} = \{0, 1\}$.
- In the Ising model, $\mathcal{K} = \{\pm 1\}$ and the constraint is soft: vertices prefer to be aligned with their neighbours, with strength controlled by the *inverse temperature* $\beta \geq 0$.

The multicoloured hardcore model is discussed in [18, §3.2.2]. It was originally introduced by Runnels and Lebowitz [20] in the context of lattice gases.

The results of [20, 18] are specialised to $\mathbb{Z}^d$. The latter is most interested in the case where the dimension d is much larger than the number K of colours. The motivating example for this paper is the fibreoptic routing, for which the lattice $\mathbb{Z}^d$ – particularly in high dimensions – is not an appropriate model. Our results appear to be the first on general graphs.

Notation

We briefly recall some notation which is used throughout the paper.

- The underlying graph is $G = (V, E)$. Let $n := |V|$ denote its number of vertices, and write $u \sim v$ if $\{u, v\} \in E$. The degree of $v \in V$ is $d_v := \sum_{u \in V} \mathbf{1}\{u \sim v\} = |\{u \in V \mid u \sim v\}|$.
- There are $K \in \mathbb{N}$ colours, and we abbreviate $[K]_0 := \{0, 1, ..., K\}$.
- The update rates and probabilities are $\boldsymbol{\lambda} \in (0, \infty)^V$ and $\boldsymbol{p} \in [0, 1]^V$, respectively.
- The state space is $\Omega := \{\omega \in [K]_0^V \mid \omega \text{ is proper}\}$, where $\omega \in [K]_0^V$ is *proper* if

$$\omega_u \neq \omega_v \quad \text{whenever} \quad \{u, v\} \in E \quad \text{and} \quad \omega_u + \omega_v > 0.$$

- The multicoloured hardcore model is denoted $\mathsf{MCH}_\Omega(\boldsymbol{\lambda}, \boldsymbol{p})$; its equilibrium distribution π.
- For $\mathsf{QMCH}_\Omega(\boldsymbol{\lambda}, \boldsymbol{p}; \boldsymbol{\nu})$, the arrival rates are $\boldsymbol{\nu} \in (0, \infty)^V$ and equilibrium service rates

$$s_v := \sum_{\omega \in \Omega : \omega_v \neq 0} \pi(\omega) \quad \text{for} \quad v \in V.$$

3 Proofs of Main Theorems

A Mixing

In this section, we use the classical path coupling argument of Bubley and Dyer [6] to upper bound the mixing time. Throughout, $X, Y \sim \mathsf{MCH}_\Omega(\boldsymbol{\lambda}, \boldsymbol{p})$, under the "natural" coupling:

- the vertex-update clocks are coupled, so the same vertex is chosen at the same time;
- the subsequent coin toss and colour selection are also coupled.

This coupling is clearly *coalescent*:

$$X^t = Y^t \quad \text{implies} \quad X^s = Y^s \quad \text{for all} \quad s \geq t.$$

Proof of Theorem A. We use path coupling, so must define a path space. We say that $x, y \in [K]_0^V$ are *adjacent* if there is a unique $v \in V$ such that $x_v \neq y_v$ *and* $0 \in \{x_v, y_v\}$. In other words, our path space is generated by activating an inactive vertex or deactivating an active vertex; changing the colour of an already active vertex *is not* permitted. This space is connected: let $d(x, y)$ denote the distance between two configurations $x, y \in [K]_0^V$; then, $\mathbf{1}\{x \neq y'\} \leq d(x, y) \leq 2n$ for all $x, y \in [K]_0^V$, going via the empty configuration $(0, ..., 0) \in \Omega$.

For $v \in V$ and $x \in [K]_0^V$, denote the *available colours at v in x* by

$$\mathcal{A}_v(x) := \{1, ..., K\} \setminus \cup_{u \in V : \{u, v\} \in E} \{x_u\} = \{k \in \{1, ..., K\} \mid x_u \neq k \, \forall u \sim v\}.$$

Suppose that $(X^0, Y^0) = (x, y) \in \Omega^2$ with $d(x, y) = 1$; say, $0 = x_v \neq y_v$. Consider the first step of the process from these states. Suppose that vertex $u \in V$ updates.

- Suppose that $u \not\sim v$. Then, $\mathcal{A}_x(u) = \mathcal{A}_y(u)$, since $x_w = y_w$ for all $w \sim u$. Hence, we can perform the same update in both X and Y. The relative distance is unchanged, unless $u = v$, in which case the two coalesce.
- Suppose that $u \sim v$; in particular, $u \neq v$. We may not have $\mathcal{A}_u(x) = \mathcal{A}_u(y)$, but always

$$\mathcal{A}_u(x) \cup \{x_u\} = \mathcal{A}_u(y) \cup \{y_u\}.$$

Hence, $|\mathcal{A}_u(x) \triangle \mathcal{A}_u(y)| \leq 1$. So, the probability that a proposed colour is valid for one and not the other is at most $1/K$. If this is the case, then the relative distance increases by 1; otherwise, it remains unchanged. The probability *some* colour is proposed is p_u.

It is in this last step that the assumption $0 \in \{x_v, y_v\}$ is used: without it, the symmetric difference could be of size 2, giving a probability $2/K$. Summing over $u \in V$, the relative distance increases by 1 at rate at most $\frac{1}{K} \sum_{u:u \sim v} p_u \lambda_u$ and decreases by 1 at rate λ_v. Hence,

$$\tfrac{d}{dt} \mathbb{E}_{x,y}[d(X^t, Y^t)]\big|_{t=0} \le \lambda_v \big(\tfrac{1}{K} \textstyle\sum_{u:u \sim v} p_u \lambda_u / \lambda_v - 1 \big) \le -\beta \lambda_v,$$

with the last inequality using the (main) assumption of the theorem. This can be extended to general $x, y \in [K]_0^V$ – ie, not requiring $d(x,y) = 1$ – by looking at contraction along geodesics, in the usual manner for path coupling. Hence, recalling that $\lambda_{\min} = \min_v \lambda_v$,

$$\max_{x,y \in [K]_0^V} \tfrac{d}{dt} \mathbb{E}_{x,y}\big[d(X^t, Y^t)\big]\big|_{t=0} \le -\beta \lambda_{\min}.$$

By the Grönwall inequality, integrating this and using $\mathbf{1}\{x \ne y\} \le d(x,y) \le 2n$, we obtain

$$\max_{x,y \in [K]_0^V} \mathbb{P}_{x,y}[X^t \ne Y^t] \le \max_{x,y \in [K]_0^V} \mathbb{E}_{x,y}[d(X^t, Y^t)] \le 2n e^{-\beta t}.$$

Finally, the coupling representation of total-variation distance implies that

$$\max_{x,y \in \Omega} \|\mathbb{P}_x[X^t \in \cdot] - \mathbb{P}_y[Y^t \in \cdot]\|_{\mathsf{TV}} \le \min\{2n e^{-\beta \lambda_{\min} t}, \, 1\}. \qquad \blacktriangleleft$$

Remark. If preferred, instead of using a continuous-time version of path coupling, discretise time: let $\tilde{X}^\ell := X^{\delta \ell}$ and $\tilde{Y}^\ell := Y^{\delta \ell}$, where δ is some very small real number. Then,

$$\mathbb{E}_{x,y}[d(\tilde{X}^1, \tilde{Y}^1)] \le \big(1 - \beta \lambda_{\min} \delta + o(\delta)\big) d(x,y) \quad \text{uniformly,}$$

using the fact that the diameter is finite to obtain a uniform $o(\delta)$ term. Path coupling gives

$$\mathbb{E}_{x,y}[d(\tilde{X}^\ell, Y^\ell)] \le 2n \big(1 - \beta \lambda_{\min} \delta + o(\delta)\big)^\ell \le 2n e^{-\beta \lambda_{\min} \delta \ell + o(\delta \ell)} n.$$

Given $t \ge 0$, let $\ell := \lfloor t/\delta \rfloor \ge t/\delta - 1$. Then,

$$\mathbb{E}_{x,y}[d(X^t, Y^t)] \le \mathbb{E}_{x,y}[d(\tilde{X}^\ell, \tilde{Y}^\ell)] \le 2n e^{-\beta t + o(1)}.$$

Finally, taking $\delta \downarrow 0$, we deduce the same bound as before.

We close this section with a discussion of the equilibrium service rates. Here, we assume

$$p_v \le \tfrac{1}{3} K / \tilde{d}_v \quad \text{where} \quad \tilde{d}_v := \max\{d_u \mid u \sim v \text{ or } u = v\} \quad \text{for} \quad v \in V.$$

Proof of Proposition A. The quantity we estimate is the proportion of colours available at a vertex. This allows estimation of the probability an attempted colouring is successful.

Clearly, in equilibrium, each neighbour u of v is active with probability at most $p_u = \frac{1}{3} K / \tilde{d}_u$; in particular, $s_v \le p_v$. Hence, if N_v is the number of colours available at v, then

$$N_v \lesssim \mathrm{Bin}(d_v, \tfrac{1}{3} K / d_v) \quad \text{in equilibrium.}$$

It can be shown that $\mathbb{P}[\mathrm{Bin}(d, \tfrac{1}{3} k/d) \ge \tfrac{1}{2} k] \le \tfrac{1}{3}$ whenever $k \le 3d$. This implies that

$$\mathbb{P}[N_v \ge \tfrac{1}{2} K] \le \tfrac{1}{3}.$$

Hence, upon refreshing, at least $\frac{1}{2}$ of the colours are available with probability at least $\frac{2}{3}$. So, the probability that the proposed colour is accepted is at least $\frac{1}{3}$. Thus, $s_v \ge \frac{1}{3} p_v$. $\qquad \blacktriangleleft$

We discuss briefly extensions of this proof, including heuristics for an upper bound on s_v.

Remark. If we require $p_v \leq (1 - \delta)K/\tilde{d}_v$, then the above argument says that at least a proportion δ of the colours are free in expectation. If K (and $\tilde{d}_v$) are large, then the Binomial concentrates. There is then a probability δ that a uniformly proposed colour is available.

We can extend this, heuristically at least. If $u, u' \sim v$, then the colours at u and u' should be approximately independent if K is large and the graph has few triangles. If $k_1, ..., k_K \sim^{\text{iid}} \text{Unif}([K])$, then $\frac{1}{K} |\{k_1, ..., k_K\}| \approx 1/e$, suggesting that, in fact, a proportion $1/e$ are available after K choices. This would suggest $s_v \geq p_v/e$.

We can also try to iterate this argument. Instead of upper bounding the expected number of colours taken by $\sum_{u:u\sim v} p_u$, we can bound by $\sum_{u:u\sim v} s_u$. Suppose that s_v does not vary much over the vertices: $s_v \approx \bar{s} := \frac{1}{n}\sum_u s_u$, the average of s; see, eg, Figure 2 later. Also, assume graph regularity: $d_v = d$, and $p_v = p$, for all v. Then, $\sum_{u:u\sim v} s_u \approx d\bar{s}$. This imposes

$$\bar{s} \leq p(1 - d\bar{s}/K); \quad \text{ie,} \quad \bar{s} \leq p/(1 + pd/K).$$

Including the factor $1/e$ from the previous heuristic improves this to $\bar{s} \approx p/(1 + e^{-1}pd/K)$. $\triangle$

B Queues

Next, we investigate the *stability* of the queueing network: ie, its positive recurrence (or lack thereof) and expected queue length in equilibrium. The end goal is Theorem B. Similar properties for a related model are established in [10, §V], using the usual *Lyapunov function*

$$L^t := \sum_{v \in V}(Q_v^t)^2 \quad \text{for} \quad t \geq 0 \quad \text{where} \quad Q = (Q^t)_{t\geq 0} \sim \mathsf{QMCH}_\Omega(\boldsymbol{\lambda}, \boldsymbol{p}; \boldsymbol{\nu}).$$

There, the model is slightly simpler, with unit service times, rather than Exponentials. Moreover, they require $p_v = p \leq 1/\Delta$ for all $v \in V$, where $\Delta := \max_v d_v$ is the maximum degree of the graph $G = (V, E)$, and treat Δ as a constant, which is absorbed into a final, unquantified constant. For a sequence $(G_n)_{n\in\mathbb{N}}$ of graphs, this implicitly assumes bounded degrees: $\sup_{n\in\mathbb{N}} \Delta_n < \infty$. We allow much greater generality, both in G and in $\boldsymbol{p}$.

We denote by τ the first time the queue is empty:

$$\tau := \inf\{t \geq 0 \mid Q^t = 0, \cup_{s\in[0,t]} Q^s \neq \{0\}\}.$$

Positive recurrence is equivalent to having $\mathbb{E}_q[\tau] < \infty$ for some, and hence all, $q \neq 0$.

Proof of Theorem B. We establish negative drift for an appropriate Lyapunov function L:

$$L^t := \tfrac{1}{2} \sum_{v \in V}(Q_v^t)^2 \quad \text{for} \quad t \geq 0. \tag{1}$$

We fix some notation and conventions. By the memoryless property of the service times, we may assume that the vertices are *always* providing service, but that a service attempt is rejected if the vertex is inactive at the time of the attempt. Then, the arrivals and attempted services form Poisson processes, independent of each other and the underlying MCH process.

Fix $v \in V$ and $t, T \geq 0$. Write $\hat{S}_v[T, T + t)$ for the number of attempted services by vertex v between times T and $T + t$, and write $\hat{s}_v := \hat{s}_v[T, T + t) := \frac{1}{t}\hat{S}_v[T, T + t)$ for the average (attempted) service rate in this interval. Similarly, write $\hat{A}_v[T, T + t)$ and $\hat{a}_v := \hat{a}_v[T, T + t) := \frac{1}{t}\hat{A}_v[T, T + t)$ for the number of arrivals and average service rate, respectively, between T and $T + t$.

Using these definitions, we have the following simple inequality:

$$Q_v^{T+t} \leq \left[Q_v^t - \hat{S}_v[T, T + t)\right]_+ + \hat{A}_v[T, T + t) = [Q_v^t - t\hat{s}_v]_+ + t\hat{a}_v,$$

where $[\alpha]_+ := \max\{\alpha, 0\}$ for $\alpha \in \mathbb{R}$. Hence, using $[Q_v^t - t\hat{s}_v]_+ \le Q_v^t$,

$$
\begin{aligned}
(Q_v^{T+t})^2 &\le (Q_v^T - t\hat{s}_v)^2 + 2t[Q_v^T - t\hat{s}_v]_+ \hat{a}_v + t^2 \hat{a}_v^2 \\
&\le (Q_v^T)^2 + 2tQ_v^T(\hat{a}_v - \hat{s}_v) + t^2(\hat{a}_v^2 + \hat{s}_v)^2.
\end{aligned}
\tag{2}
$$

Plugging this into the definition (1) of L bounds its random increment:

$$
L^{T+t} - L^T \le t \sum_{v \in V} Q_v^T(\hat{a}_v - \hat{s}_v) + \tfrac{1}{2}t^2 \sum_{v \in V}(\hat{a}_v^2 + \hat{s}_v^2). \tag{3}
$$

Now, if $\hat{\tau}_v$ is the proportion of time during $[T, T+t)$ that vertex v is active, then

$$
t\hat{a}_v = \hat{A}_v[T, T+t) \sim \text{Pois}(t\nu_v) \quad \text{and} \quad t\hat{s}_v = \hat{S}_v[T, T+t) \sim \text{Pois}(t\hat{\tau}_v).
$$

To emphasise, the implicit Poisson variables are independent of the MCH process. Recall that

$$
\text{if} \quad P \sim \text{Pois}(\mu), \quad \text{then} \quad \mathbb{E}[P] = \mu \quad \text{and} \quad \mathbb{E}[P^2] = \mu + \mu^2.
$$

Now, $\nu_v < s_v$, by assumption, and $s_v \le p_v \le 1$; also, $\hat{\tau}_v \le 1$. Hence,

$$
\mathbb{E}[\hat{a}_v] = \nu_v, \quad \mathbb{E}[\hat{a}_v^2] \le 2, \quad \mathbb{E}[\hat{s}_v] \le 1 \quad \text{and} \quad \mathbb{E}[\hat{s}_v^2] \le 2.
$$

Plugging these into (3) bounds the (expected) drift:

$$
\mathbb{E}[L^{T+t} - L^T \mid (X^T, Q^T)] \le t \sum_{v \in V} Q_v^T(\nu_v - \mathbb{E}[\hat{s}_v \mid X^T]) + \tfrac{3}{2}nt^2; \tag{4}
$$

the (attempted) service rate $\hat{s}_v[T, T+t)$ depends only on X^T, not Q^T.

It remains to handle $\mathbb{E}[\hat{s}_v \mid X^T]$. The attempted services are a thinned Poisson process. So,

$$
\mathbb{E}[\hat{s}_v \mid X^T] = \mathbb{E}[\hat{\tau}_v \mid X^T] \quad \text{and} \quad \tau_v = \tfrac{1}{t} \sum_T^{T+t} \mathbf{1}\{X_v^s \ne 0\}ds.
$$

So, if we write $\mu_{x,s}$ for the law of X^s given $X^0 = x$, then

$$
\mathbb{E}[\hat{s}_v \mid X^T] = \tfrac{1}{t} \int_T^{T+t} \mathbb{P}[X^s \ne 0 \mid X^T]ds \tfrac{1}{t} \int_0^t \mu_{X^T, s}(\{\omega \in \Omega \mid \omega_v \ne 0\})ds.
$$

This is very similar to the equilibrium (attempted) service rate

$$
s_v = \sum_{\omega \in \Omega: \omega_v \ne 0} \pi(\omega) = \pi(\{\omega \in \Omega \mid \omega_v \ne 0\});
$$

in fact, by the ergodic theorem, $\hat{s}_v[T, T+t) \to s_v$ as $t \to \infty$. Quantitatively,

$$
\begin{aligned}
|\mathbb{E}[\hat{s}_v \mid X^T] - s_v| &= \left| \tfrac{1}{t} \int_0^t \mu_{X^T, s}(\{\omega \in \Omega \mid \omega_v \ne 0\})ds - \pi(\{\omega \in \Omega \mid \omega_v \ne 0\}) \right| \\
&\le \tfrac{1}{t} \int_0^t |\mu_{X^T, s}(\{\omega \in \Omega \mid \omega_v \ne 0\}) - \pi(\{\omega \in \Omega \mid \omega_v \ne 0\})| \\
&\le \tfrac{1}{t} \int_0^t \|\mu_{X^T, s} - \pi\|_{\text{TV}}ds.
\end{aligned}
$$

It is here that we apply the mixing result, Theorem A: for any $x \in \Omega$ and $s \ge 0$,

$$
\|\mu_{x,s} - \pi\|_{\text{TV}} \le \min\{2ne^{-\beta\lambda_{\min}s}, 1\};
$$

note that the first hypothesis of Theorem B is precisely that required for Theorem A. Then,

$$
\begin{aligned}
\int_0^t \|\mu_{X^T, s} - \pi\|_{\text{TV}}ds &\le t_0 + n \int_{t_0}^{t \vee t_0} e^{-\beta\lambda_{\min}s}ds \\
&\le t_0 + (\beta\lambda_{\min})^{-1} =: t_1 \quad \text{where} \quad t_0 := (\beta\lambda_{\min})^{-1}\log(2n).
\end{aligned}
$$

In particular, this is independent of t, so vanishes once divided by t and $t \to \infty$:

$$|\mathbb{E}[\hat{s}_v \mid X^T] - s_v| \le t_1/t \to 0 \quad \text{as} \quad t \to \infty.$$

We want to plug this bound into (4). Let $\varepsilon_v := \frac{1}{2}(s_v - \nu_v)$ and $t_v := t_1/\varepsilon_v$. Then,

$$|\mathbb{E}[\hat{s}_v \mid X^T] - s_v| \le \varepsilon_v \quad \text{whenever} \quad t \ge t_v \quad \text{for all} \quad v \in V.$$

Set $t_\star := \max_v t_v$, so $t_\star \ge t_v$ Plugging this into (4),

$$\begin{aligned}
\mathbb{E}[L^{T+t} - L^T \mid (X^T, Q^T)] &\le -t \sum_{v \in V} Q_v^T(s_v - \nu_v - \varepsilon_v) + \tfrac{3}{2}nt^2 \\
&\le -\tfrac{1}{2}t \sum_{v \in V} Q_v^T(s_v - \nu_v) + \tfrac{3}{2}nt^2 \quad \text{whenever} \quad t \ge t_\star. \quad (5)
\end{aligned}$$

This expression is negative for large enough $\|Q^T\|$. This establishes negative drift of L. Hence, by the Foster–Lyapunov criterion (eg, [12, Proposition D.1]), $(Q^t)_{t \ge 0}$ is positive recurrent.

It remains to control the expected queue length in equilibrium. We start in equilibrium and take the expectation of the increment $(Q_v^{t_v})^2 - (Q_v^0)^2$. By stationarity and (2),

$$0 = \mathbb{E}[(Q_v^{t_v})^2 - (Q_v^0)^2] \le -t_v \mathbb{E}[Q_v^0](s_v - \nu_v - \varepsilon_v) + \tfrac{3}{2}nt_v^2,$$

using the same manipulations as before. Rearranging,

$$\mathbb{E}[Q_v^0] \le \tfrac{3}{2}nt_v/(s_v - \nu_v - \varepsilon_v) \le 6nt_1/(s_v - \nu_v)^2.$$

Finally, $t_1 = (\beta\lambda_{\min})^{-1}(\log(2n) + 1) = (\beta\lambda_{\min})^{-1}\log(2n/e)$. ◄

4 Simulations: Queue Lengths and Equilibrium Service Rate

We close the paper with a short discussion of some simulations. Specifically, we investigate the queue lengths and the proportion of time that a vertex is active as a rolling average – namely,

$$\hat{Q}_v^t := \tfrac{1}{t} \sum_{s=0}^{t-1} Q_v^s \quad \text{and} \quad \hat{s}_v^t := \tfrac{1}{t} \sum_{s=0}^{t-1} \mathbf{1}\{X_v^s \ne 0\} \quad \text{for} \quad t \ge 0.$$

Then, $\hat{Q}_v^t \to \mathbb{E}_\pi[Q_v^0]$ and $\hat{s}_v^t \to s_v$, the expected equilibrium queue length and service rate.

Our choice of parameters is driven by the same heuristics as for Corollaries A and B:

$$\lambda_v := d_v/\bar{d}, \quad p_v := \min\{\tfrac{4}{5}eK/d_v, \tfrac{3}{4}\} \quad \text{and} \quad \nu_v := \tfrac{1}{3}p_v \quad \text{for} \quad v \in V.$$

Notice the prefactor in p_v: it is $\frac{4}{5}e > 2$, rather than $\frac{1}{3}$ or $\frac{2}{3}$. This is to emphasise the fact that we really can take p_v close to eK/d_v, yet still get high, and stable, service rates s_v.

Figure 1 show the time-averaged queue lengths and service rates when the underlying graph is an Erdős–Rényi graph. Figure 2 show the same for a random regular graph. The average degree is 30 and $K = 10$ colours are used; so, almost all vertices satisfy $p_v = \frac{4}{5}eK/d_v \approx 0.5$. A collection of 10 vertices with typical degrees to be displayed are chosen randomly. Time is scaled so that the average vertex update-rate is 1 – ie, scaled by $\frac{1}{n}\sum_v(\lambda_v + \nu_v + 1)$.

We see that the empirical service rates settle down really quite quickly, and appear to be remain stable. Moreover, the values s_v to which they converge appear to be on the same order as the proposal probabilities p_v. This suggests many proposals are accepted, but not *too many*: if $s_v \approx p_v$, then perhaps a higher proposal probability p_v could have been used. In particular, we found that the normalised difference $|s_v - p_v|/p_v$ averaged around 60%.
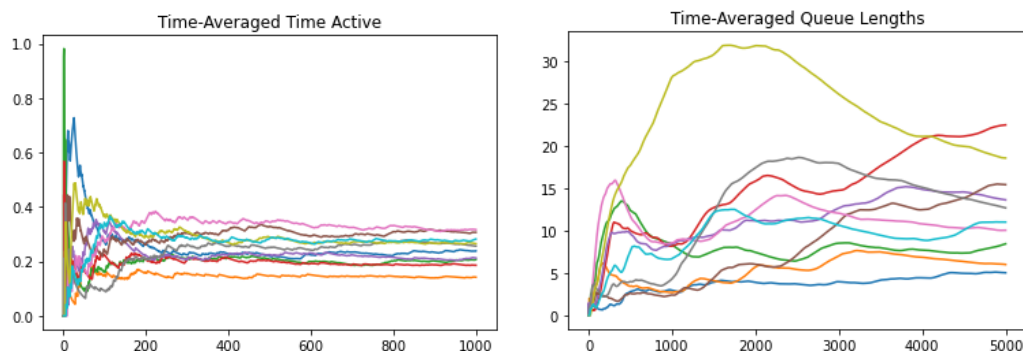
Figure 1 The underlying graph is Erdős–Rényi with $n = 500$ vertices and edge probability $40/n$.
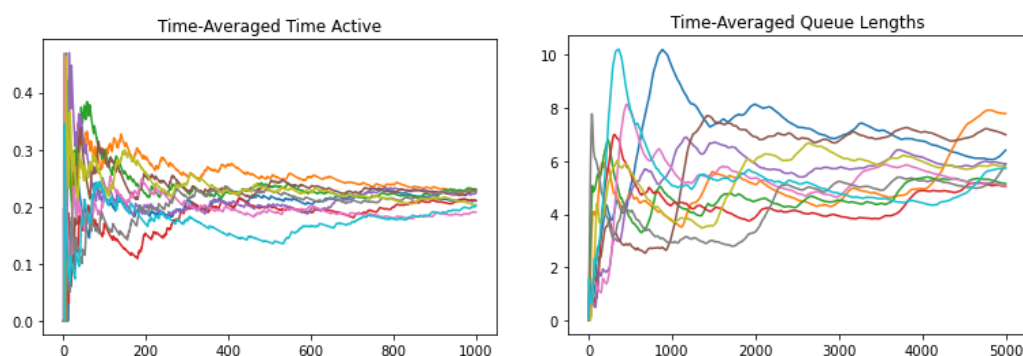
Figure 2 The underlying graph is drawn uniformly over 40-regular graphs on $n = 500$ vertices.

The queue lengths, on the other hand, fluctuate a more. They are a bit more stable in the random regular graph (Figure 2) compared with the Erdős–Rényi graph (Figure 1), perhaps due to inhomogeneities. It is not even completely clear what they are converging to.

We suggest that this is likely caused by the inhomogeneities in the graph along with the fact that we take $\nu_v = p_v/3 \approx 0.33p_v$, which is pretty close to $s_v \approx 0.4p_v$. Indeed, the same calculations (not shown) with $\nu_v = 0.2p_v$ result in much more stable queues.

The primary objective is to get as large an equilibrium service rate s_v as possible, or at least its average $\bar{s} = \frac{1}{n}\sum_v s_v$. Since the 60% above is still quite a large rejection rate, we also tested a slightly smaller value of p_v: namely, we used $p_v = \frac{2}{3}eK/d_v \approx 0.45$. However, we found that $\bar{s}$ was about 10% smaller for these parameters, for both random graph models.

A random d-regular graph locally looks like a d-regular tree, so it is not reasonable to expect better than $eK/d = p_c(\Delta, K)$, the earlier critical threshold. Similarly, a sparse Erdős–Rényi graph locally looks like a Bienaymé–Galton–Watson tree with $\mathrm{Pois}(\bar{d})$ degrees.

References

1 Nima Anari, Kuikui Liu, and Shayan Oveis Gharan. Spectral independence in high-dimensional expanders and applications to the hardcore model. In *2020 IEEE 61st Annual Symposium on Foundations of Computer Science*, pages 1319–1330. IEEE Computer Soc., Los Alamitos, CA, 2020. `doi:10.1109/FOCS46700.2020.00125`.

2 Nayantara Bhatnagar, Allan Sly, and Prasad Tetali. Decay of correlations for the hardcore model on the d-regular random graph. *Electron. J. Probab.*, 21:Paper No. 9, 42 pp., 2016. `doi:10.1214/16-EJP3552`.

3 R. Boorstyn, A. Kershenbaum, B. Maglaris, and V. Sahin. Throughput analysis in multihop csma packet radio networks. *IEEE Transactions on Communications*, 35(3):267–274, March 1987. `doi:10.1109/TCOM.1987.1096769`.

4 Torsten Braun, Andreas Kassler, Maria Kihl, Veselin Rakocevic, Vasilios Siris, and Geert Heijenk. Multihop wireless networks. In Vasilios Siris, Torsten Braun, Francisco Barcelo-Arroyo, Dirk Staehle, Giovanni Giambene, and Yevgeni Koucheryavy, editors, *Traffic and QoS Management in Wireless Multimedia Networks: COST 290 Final Report*, pages 201–265. Springer US, Boston, MA, 2009.

5 Raffaele Bruno and Maddalena Nurchis. Survey on diversity-based routing in wireless mesh networks: Challenges and solutions. *Computer Communications*, 33(3):269–282, February 2010. `doi:10.1016/j.comcom.2009.09.003`.

6 Ross Bubley and Martin Dyer. Path coupling: A technique for proving rapid mixing in markov chains. In *Proceedings of the 38th Annual Symposium on Foundations of Computer Science*, FOCS '97, pages 223–, Washington, DC, USA, 1997. IEEE Computer Society. `doi:10.1109/SFCS.1997.646111`.

7 Andreas Galanis, Qi Ge, Daniel Štefankovič, Eric Vigoda, and Linji Yang. Improved inapproximability results for counting independent sets in the hard-core model. *Random Structures Algorithms*, 45(1):78–110, 2014. `doi:10.1002/rsa.20479`.

8 Song-Nam Hong and Ivana Marić. Multihop wireless backhaul for 5g. In Ivana Marić, Shlomo Shamai (Shitz), and Osvaldo Simeone, editors, *Information Theoretic Perspectives on 5G Systems and Beyond*, pages 131–165. Cambridge University Press, Cambridge, 2022. `doi:10.1017/9781108241267.004`.

9 L. Jiang and J. Walrand. A distributed csma algorithm for throughput and utility maximization in wireless networks. *IEEE/ACM Transactions on Networking*, 18(3):960–972, June 2010. `doi:10.1109/TNET.2009.2035046`.

10 Libin Jiang, Mathieu Leconte, Jian Ni, Rayadurgam Srikant, and Jean Walrand. Fast mixing of parallel glauber dynamics and low-delay csma scheduling. *IEEE Transactions on Information Theory*, 58(10):6541–6555, October 2012. `doi:10.1109/TIT.2012.2204032`.

11 Frank Kelly. Private communication, 2018.

12 Frank Kelly and Elena Yudovina. *Stochastic Networks*, volume 2 of *Institute of Mathematical Statistics Textbooks*. Cambridge University Press, Cambridge, 2014. `doi:10.1017/CBO9781139565363`.

13 Frank P. Kelly. Stochastic models of computer communication systems. *Journal of the Royal Statistical Society*, 47(3):379–395, 415–428, 1985. `arXiv:2345773`.

14 David A. Levin, Yuval Peres, and Elizabeth L. Wilmer. *Markov Chains and Mixing Times*. American Mathematical Society, Providence, RI, USA, second edition, 2017. `doi:10.1090/mbk/107`.

15 Pinyan Lu, Kuan Yang, and Chihao Zhang. Fptas for hardcore and ising models on hypergraphs. In *33rd Symposium on Theoretical Aspects of Computer Science*, volume 47 of *LIPIcs. Leibniz Int. Proc. Inform.*, pages Art. No. 51, 14. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2016.

16 J. Ni and R. Srikant. Distributed csma/ca algorithms for achieving maximum throughput in wireless networks. In *2009 Information Theory and Applications Workshop*, pages 250–250, 2009. `doi:10.1109/ITA.2009.5044953`.

17 Jian Ni, Bo Tan, and R. Srikant. Q-csma: Queue-length based csma/ca algorithms for achieving maximum throughput and low delay in wireless networks. In *Proceedings of the 29th Conference on Information Communications*, INFOCOM'10, pages 271–275, Piscataway, NJ, USA, 2010. IEEE Press.

18 Ron Peled and Yinon Spinka. Long-range order in discrete spin systems. *arXiv:2010.03177 [math-ph]*, January 2024. `doi:10.48550/arXiv.2010.03177`.

19   Shreevatsa Rajagopalan, Devavrat Shah, and Jinwoo Shin. Network adiabatic theorem: An efficient randomized protocol for contention resolution. *SIGMETRICS Perform. Eval. Rev.*, 37(1):133–144, June 2009. `doi:10.1145/2492101.1555365`.

20   L. K. Runnels and J. L. Lebowitz. Analyticity of a hard-core multicomponent lattice gas. *Journal of Statistical Physics*, 14(6):525–533, 1976. `doi:10.1007/BF01012851`.

21   Pol Torres Compta, Frank H. P. Fitzek, and Daniel E. Lucani. Network coding is the 5g key enabling technology: Effects and strategies to manage heterogeneous packet lengths. *Transactions on Emerging Telecommunications Technologies*, 26(1):46–55, January 2015. `doi:10.1002/ett.2899`.

22   Nigel Walker. Design challenges in BT's network. `www.newton.ac.uk/seminar/23532`, 2018.

Binary Search Trees of Permuton Samples

Benoît Corsini ✉
Department of Mathematics and Computer Science,
Eindhoven University of Technology, The Netherlands

Victor Dubach ✉ ⓘ
Université de Lorraine, CNRS, IECL, F-54000 Nancy, France

Valentin Féray ✉ ⓘ
Université de Lorraine, CNRS, IECL, F-54000 Nancy, France

──── **Abstract** ────

Binary search trees (BST) are a popular type of structure when dealing with ordered data. They allow efficient access and modification of data, with their height corresponding to the worst retrieval time. From a probabilistic point of view, BSTs associated with data arriving in a uniform random order are well understood, but less is known when the input is a non-uniform permutation.

We consider here the case where the input comes from i.i.d. random points in the plane with law μ, a model which we refer to as a *permuton sample*. Our results show that the asymptotic proportion of nodes in each subtree only depends on the behavior of the measure μ at its left boundary, while the height of the BST has a universal asymptotic behavior for a large family of measures μ. Our approach involves a mix of combinatorial and probabilistic tools, namely combinatorial properties of binary search trees, coupling arguments, and deviation estimates.

1 Introduction

1.1 Context and informal description of our results

A *binary search tree* (BST) is a rooted binary tree where nodes carry labels – which are real numbers – and where, for each vertex v, all labels of vertices in the left-subtree (resp. right-subtree) attached to v are smaller (resp. bigger) than that of v. Binary search trees are a popular type of data structure for storing ordered data. One key feature is that the worst-case complexity of basic operations (lookup, addition or removal of data) is proportional to the height of the tree.

Given a BST $\mathcal{T}$ and a real number x distinct from the labels of $\mathcal{T}$, there is a unique way to insert x into $\mathcal{T}$, i.e. there is a unique BST $\mathcal{T}^{+x}$ obtained from $\mathcal{T}$ by adding a new node with label x. Iterating this operation starting from the empty tree and a sequence $y = (y_1, \ldots, y_n)$ of distinct values, we get a BST $\mathcal{T}\langle y \rangle$ with n nodes. An example can be found in Figure 1. The shape of $\mathcal{T}\langle y \rangle$ (i.e. the underlying binary tree without node labels)

depends only on the relative order of the numbers $y_1, \ldots, y_n$, and not on their actual values. We can thus assume without loss of generality that the sequence y is a permutation σ of the integers from 1 to n, and we write $\mathcal{T}\langle\sigma\rangle = \mathcal{T}\langle\sigma_1, \ldots, \sigma_n\rangle$ in this case.
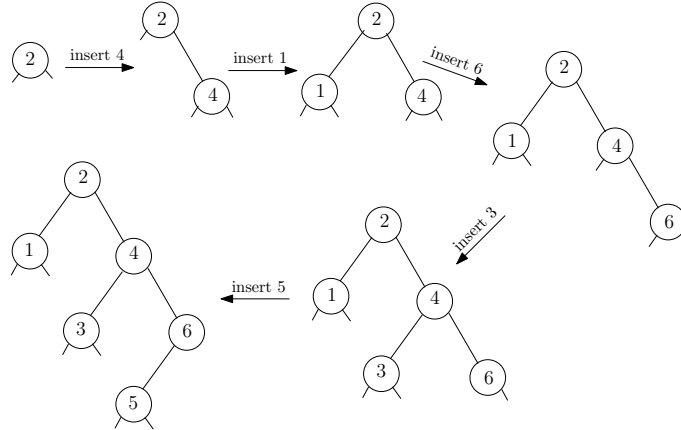
Figure 1 Iterative construction of the BST associated with the sequence $y = (2, 4, 1, 6, 3, 5)$.

In the worst case, the tree $\mathcal{T}\langle\sigma\rangle$ has height $n - 1$ and further operations will have a linear complexity, which is far from optimal. However it has been proven by Devroye [10] that, if σ is a uniformly random permutation of $\{1, \ldots, n\}$, then the height $h(\mathcal{T}\langle\sigma\rangle)$ is asymptotically equivalent to $c^* \log n$ for some constant c^*. Assuming that σ is uniformly distributed means that the data used to construct our BST arrived in a completely random order, which is in general unrealistic. It seems therefore natural to study BSTs associated with non-uniform random permutations, and in particular to see how Devroye's result is modified when changing the distribution of σ.

A first step in this direction has been performed in the papers [1, 7], where the BSTs associated with random Mallows and record-biased permutations are studied, showing interesting phase transition phenomena. In the current paper, we will consider some geometric models of random permutations, sampled via i.i.d. random points in the plane with some common distribution μ. These models will be referred to here as *permuton samples*, and denoted by σ_μ^n; they appear naturally in a recently developed theory of limiting objects for large permutations, called permutons [14]. The goal of studying such models is twofold. First, it is a much larger but still tractable family of models than those considered before (permuton samples are indexed by probability measures on the square, while Mallows and record-biased permutations are one-parameter families of models). Second, since permutons describe the "large-scale shape" of permutations, it enlightens the connection between this "large-scale shape" and the associated BST.

Our first result (Theorem 1) shows that, for a large family of permuton samples, the asymptotic behavior of the BST height is the same as the one found by Devroye for uniform permutations, namely that $h(\mathcal{T}\langle\sigma_\mu^n\rangle)$ is asymptotically equivalent to $c^* \log n$. Our second result (Theorem 13) studies another type of limit for the sequence of BSTs, using the formalism of *subtree size convergence* recently introduced by Grübel in [13]. In this setting and under some mild assumption, we prove convergence of the BST associated with permuton samples, where the limit object depends on the permuton only through its "derivative" at the left edge $\{0\} \times [0, 1]$ of the unit square $[0, 1]^2$.

In the remaining part of the introduction, we present the model of permuton samples and introduce some notation. Our main results are then stated and proved in Sections 2 and 3, and extra results are discussed in Section 4.

1.2 Our model: binary search trees of permuton samples

There is a natural way to map a (generic) finite set of points $\mathcal{P} \subset \mathbb{R}^2$ to a permutation $\sigma\langle\mathcal{P}\rangle$ and a binary search tree $\mathcal{T}\langle\mathcal{P}\rangle$, which we describe now. Let $\mathcal{P} = \{(x_1, y_1), \ldots, (x_n, y_n)\}$ be a set of points in $\mathbb{R}^2$ with distinct x- and distinct y-coordinates, and let $\{(x_{(1)}, y_{(1)}), \ldots, (x_{(n)}, y_{(n)})\}$ be its reordering such that $x_{(1)} < \ldots < x_{(n)}$. Then there exists a unique permutation $\sigma = \sigma\langle\mathcal{P}\rangle$ of $\{1, \ldots, n\}$ such that $(y_{(1)}, \ldots, y_{(n)})$ and $(\sigma_1, \ldots, \sigma_n)$ are in the same relative order. We let $\mathcal{T}\langle\mathcal{P}\rangle := \mathcal{T}\langle y_{(1)}, \ldots, y_{(n)}\rangle$ and note that the trees $\mathcal{T}\langle\mathcal{P}\rangle$ and $\mathcal{T}\langle\sigma\langle\mathcal{P}\rangle\rangle$ have the same shape since their underlying data have the same relative order. These constructions are illustrated in Figure 2.
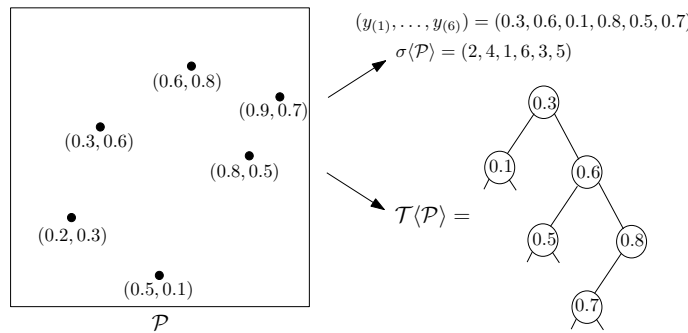
Figure 2 A set of points in $\mathbb{R}^2$ and its associated permutation and binary search tree.

Now consider a probability measure μ on $\mathbb{R}^2$ and take a set $\mathcal{P}_\mu^n$ of n i.i.d. points in $\mathbb{R}^2$ with distribution μ. In order to make sure that the associated permutation and BST are well-defined, we need the coordinates of the points to be all distinct. To this extent, we assume for the rest of this work that the projections of μ on both axes have no atom. Moreover, since the permutation and the shape of the tree only depend on the relative positions of the points, without loss of generality we can re-scale μ so that its support is in $[0,1]^2$ and both its marginals are uniform (see [5, Remark 1.2] for details). Such measures are called *permutons*, and are natural limit objects for large permutations (see e.g. [2, 14]). The associated model of random permutations $\sigma\langle\mathcal{P}_\mu^n\rangle$ will then simply be denoted by σ_μ^n. This is a broad generalization of the uniform measure on permutations of size n, which corresponds to $\mu = \mathrm{Leb}_{[0,1]^2}$. Such models have been considered in the literature under various perspectives, see e.g. [5, 9, 11, 12, 15].

In the current paper, we are interested in the binary search tree $\mathcal{T}\langle\sigma_\mu^n\rangle$ of this random permutation model. Since we will be interested only in the shape of this tree (height in Section 2, subtree size convergence in Section 3), we may and will equivalently consider the tree $\mathcal{T}\langle\mathcal{P}_\mu^n\rangle$ instead of $\mathcal{T}\langle\sigma_\mu^n\rangle$. Furthermore, for convenience, we shall work with a Poisson point process $\mathcal{P}_\mu^N$ with intensity $n\mu$, instead of the point process $\mathcal{P}_\mu^n$. This new process has random size $N \sim \mathrm{POISSON}(n)$, and conditionally given N it contains i.i.d. points distributed under μ. This enables useful independence properties, which make the proofs of our results easier. In the full paper [8], we explain in great detail how to "de-Poissonize" our results.

1.3 Some probabilistic notation

Throughout this paper, "with high probability" (w.h.p.) means "with probability tending to 1, as n tends to ∞". We also use the notation $X_n = o_\mathbb{P}(Y_n)$ to say that X_n/Y_n converges to 0 in probability, and we write $X \preceq Y$ (resp. $X \succeq Y$) to denote that X is stochastically smaller (resp. larger) than Y.

2   First main result: universal behavior of the BST height

2.1   Statement of the result and proof strategy

We denote by $h(\mathcal{T})$ the height of a tree $\mathcal{T}$, i.e. the maximal distance from a leaf to the root. As mentioned in Section 1.1, Devroye [10] proved that for uniformly random permutations σ^n of size n, the quantity $h\left(\mathcal{T}\langle\sigma^n\rangle\right)/\log n$ converges in probability and in L^p (for all $p \geq 1$) to a constant c^*, defined as the unique solution to $c\log(2e/c) = 1$ with $c \geq 2$. We provide a sufficient condition on a permuton μ, under which the same result holds for $h\left(\mathcal{T}\langle\mathcal{P}_\mu^N\rangle\right)$. In the following, a permuton μ is said to satisfy assumption (A1) if μ has a bounded density ρ on the unit square $[0,1]^2$, which is continuous and positive on a neighborhood of $\{0\} \times [0,1]$.

▶ **Theorem 1** (Universality of BST height for permuton samples). *Let μ be a permuton satisfying assumption* (A1), *and let $\mathcal{P}_\mu^N$ be a Poisson point process with intensity $n\mu$. Then, as $n \to \infty$, the following convergence holds in probability and in L^p for all $p \geq 1$:*

$$\frac{h\left(\mathcal{T}\langle\mathcal{P}_\mu^N\rangle\right)}{c^*\log n} \longrightarrow 1\,.$$

Let us briefly overview the proof strategy of Theorem 1. We shall decompose the BST drawn from a permuton sample as a top tree, to which hanging trees are attached. To this end, consider $\beta \in (0,1)$ and set $\mathcal{P}(\beta) := \mathcal{P}_\mu^N \cap ([0,\beta] \times [0,1])$. Then set $K_\beta := |\mathcal{P}(\beta)|$ and let $y_{(1)} < \cdots < y_{(K_\beta)}$ be the ordered y-coordinates of the points in $\mathcal{P}(\beta)$. For each $0 \leq k \leq K_\beta$, define $I_k = (y_{(k)}, y_{(k+1)})$ with the convention $y_{(0)} = 0$ and $y_{(K_\beta+1)} = 1$. Finally, for each k, define $\mathcal{P}_k(\beta) := \mathcal{P}_\mu^N \cap \left((\beta,1] \times I_k\right)$. We call $\mathcal{T}\langle\mathcal{P}(\beta)\rangle$ and $\left(\mathcal{T}\langle\mathcal{P}_k(\beta)\rangle\right)_{0 \leq k \leq K_\beta}$ respectively the *top tree* and the *hanging trees* of $\mathcal{T}\langle\mathcal{P}_\mu^N\rangle$. One can see that the top and hanging trees are indeed subtrees of $\mathcal{T}\langle\mathcal{P}_\mu^N\rangle$. Furthermore, the entire tree can be reconstructed by grafting the hanging trees to some nodes of the top tree. In particular, this yields the following lemma:

▶ **Lemma 2.** *For any $\beta \in (0,1)$:*

$$h\left(\mathcal{T}\langle\mathcal{P}(\beta)\rangle\right) \leq h\left(\mathcal{T}\langle\mathcal{P}_\mu^N\rangle\right) \leq h\left(\mathcal{T}\langle\mathcal{P}(\beta)\rangle\right) + 1 + \max_{0 \leq k \leq K_\beta}\left\{h\left(\mathcal{T}\langle\mathcal{P}_k(\beta)\rangle\right)\right\}.$$
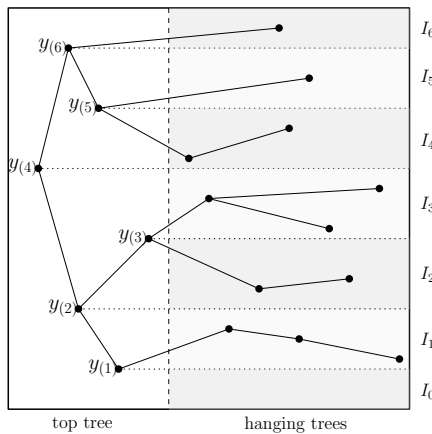
Figure 3 A sample of points and its associated BST, decomposed as top and hanging trees. The BST has been rotated of 90 degrees to the left, so that it can be drawn directly on the set of points.

See Figure 3 for an illustration. Thus, controlling the height of $\mathcal{T}\langle\mathcal{P}_\mu^N\rangle$ amounts to controlling the heights of its top and hanging trees. This is done via different approaches: in Sections 2.2 and 2.3 we prove that the top tree has height $(c^* + o_{\mathbb{P}}(1))\log n$ for well chosen β, and in Sections 2.4 and 2.5 we prove that the hanging trees all have height $o_{\mathbb{P}}(\log n)$. Finally, we combine these estimates in Section 2.6 to conclude the proof of Theorem 1.

2.2 Height modification by adding/removing points

We rely on comparison arguments to prove our results: the basic idea is to (locally) compare the density of our permuton to a constant density, for which we can apply Devroye's result. However, while Poisson point processes possess nice monotonicity properties with respect to their intensities, BSTs are much trickier to handle. Indeed, one can see that adding a single point to a point set may halve the height of the associated BST. In this section, we develop adequate tools for such comparison arguments.

We start with a simple lemma about genealogies in a BST, easily derived by construction.

▶ **Lemma 3.** *Let $y = (y_1, \ldots, y_n)$ be a list of distinct numbers and $\mathcal{T} = \mathcal{T}\langle y\rangle$ be the associated BST. If $i < j$ are two indices then the following are equivalent:*

- *y_i is an ancestor of y_j in $\mathcal{T}$ (the converse cannot hold);*
- *there is no $k < i$ such that y_k is between y_i and y_j, i.e. such that $(y_i - y_k)(y_j - y_k) < 0$.*

A *chain* in a tree $\mathcal{T}$ is a subset C of its nodes such that for every pair (v, w) in C, either v is an ancestor of w, or the converse. We note that the height of $\mathcal{T}$ is the maximal size of a chain, minus 1. By extension, if $\mathcal{P} = \{(x_1, y_1), \ldots, (x_n, y_n)\}$ is a generic point set, we say that $C \subseteq \mathcal{P}$ is a chain of $\mathcal{T}\langle\mathcal{P}\rangle$ if the corresponding nodes form a chain. Using Lemma 3, the following result is proved immediately.

▶ **Lemma 4.** *Let $\mathcal{P}_- \subseteq \mathcal{P}_+$ be two point sets with distinct x- and distinct y-coordinates. Then, for any chain C of $\mathcal{T}\langle\mathcal{P}_+\rangle$, the set $C \cap \mathcal{P}_-$ is a chain of $\mathcal{T}\langle\mathcal{P}_-\rangle$. Consequently, if $\mathcal{C}$ is a chain of maximal size in $\mathcal{T}\langle\mathcal{P}_+\rangle$, we have*

$$h\left(\mathcal{T}\langle\mathcal{P}_-\rangle\right) \geq h\left(\mathcal{T}\langle\mathcal{P}_+\rangle\right) - \left|\mathcal{C} \cap (\mathcal{P}_+ \setminus \mathcal{P}_-)\right|.$$

Combining the above lemma with standard thinning properties of Poisson point processes, we get the following useful proposition.

▶ **Proposition 5.** *Let $\rho_- \leq \rho_+$ be two intensity functions defined on the same support $S \subseteq \mathbb{R}^2$, and $\mathcal{P}_-, \mathcal{P}_+$ be two Poisson point processes with intensities ρ_- and ρ_+. Then, we have*

$$h\left(\mathcal{T}\langle\mathcal{P}_-\rangle\right) \succeq \text{BINOMIAL}\left(1 + h\left(\mathcal{T}\langle\mathcal{P}_+\rangle\right), \inf_{(x,y)\in S} \frac{\rho_-(x,y)}{\rho_+(x,y)}\right) - 1.$$

Proof. Write $r := \inf_{(x,y)\in S} \frac{\rho_-(x,y)}{\rho_+(x,y)}$ where, by convention, $\frac{\rho_-(x,y)}{\rho_+(x,y)} = 1$ if $\rho_+(x,y) = 0$. We couple $\mathcal{P}_+$ and $\mathcal{P}_-$ according to the classical thinning process, meaning that $\mathcal{P}_-$ is constructed by keeping each point (x, y) of $\mathcal{P}_+$ independently with probability $\rho_-(x,y)/\rho_+(x,y) \geq r$.

Let $\mathcal{C}$ be a chain of maximal size in $\mathcal{P}_+$, and set $K := |\mathcal{C} \cap \mathcal{P}_-|$. By Lemma 4:

$$h\left(\mathcal{T}\langle\mathcal{P}_-\rangle\right) \geq h\left(\mathcal{T}\langle\mathcal{P}_+\rangle\right) - \left|\mathcal{C} \cap (\mathcal{P}_+ \setminus \mathcal{P}_-)\right| = |\mathcal{C}| - 1 - \left|\mathcal{C} \cap (\mathcal{P}_+ \setminus \mathcal{P}_-)\right| = K - 1.$$

Conditionally given $\mathcal{P}_+$ we have $K \succeq \text{BINOMIAL}\left(|\mathcal{C}|, r\right)$, and this concludes the proof. ◀

2.3   Controlling the height of the top tree

We can now use our tools to compare the BST of Poisson point processes with the BST of uniformly random permutations.

▶ **Proposition 6.** *Let $R = [x_1, x_2] \times [y_1, y_2]$ be a rectangle with non-empty interior and $\rho : R \to (0, \infty)$ be a continuous, positive intensity function. For each integer n, let $\mathcal{P}_\rho^N$ be a Poisson point process with intensity $n\rho$. Let $0 < m \leq M < \infty$ be such that $m \leq \rho \leq M$, and write $\eta := \frac{M-m}{m}$. Then for any $\varepsilon > 0$ we have:*

$$\lim_{n \to \infty} \mathbb{P}\left[\left|\frac{h\left(\mathcal{T}\langle \mathcal{P}_\rho^N \rangle\right)}{c^* \log n} - 1\right| > \eta + \varepsilon\right] = 0. \tag{1}$$

Moreover, for any $p > 0$, the sequence $\left(\frac{h\left(\mathcal{T}\langle \mathcal{P}_\rho^N \rangle\right)}{\log n}\right)^p$ is uniformly integrable.

Proof. Write $\zeta := (x_2 - x_1)(y_2 - y_1) > 0$ for the area of R. Also note that $M/m = 1 + \eta$ and $m/M \geq 1 - \eta$. Using Proposition 5 with $\rho_- = n\rho$ and $\rho_+ = nM$ on R, we obtain

$$h\left(\mathcal{T}\langle \mathcal{P}_\rho^N \rangle\right) \succeq \text{BINOMIAL}\left(1 + h\left(\mathcal{T}\langle \mathcal{P}_+ \rangle\right), \frac{m}{M}\right) - 1,$$

where $\mathcal{T}\langle \mathcal{P}_+ \rangle$ is the BST of a uniform permutation of random size $\text{POISSON}\left(n\zeta M\right)$. According to [10, Theorem 5.1], $h\left(\mathcal{T}\langle \mathcal{P}_+ \rangle\right)$ then behaves as $c^* \log(|\mathcal{P}_+|)$ as $n \to \infty$ in probability, which is itself close to $c^* \log n$. Since $\text{BINOMIAL}\left(a \log n, \, m/M\right)$ is concentrated around $(am/M) \log n$, we deduce:

$$h\left(\mathcal{T}\langle \mathcal{P}_\rho^N \rangle\right) \geq \frac{m}{M}\left(c^* \log n - o_\mathbb{P}(\log n)\right) \geq \left(1 - \eta - o_\mathbb{P}(1)\right)c^* \log n.$$

Similarly, using Proposition 5 with $\rho_- = nm$ and $\rho_+ = n\rho$ we obtain

$$h\left(\mathcal{T}\langle \mathcal{P}_- \rangle\right) \succeq \text{BINOMIAL}\left(1 + h\left(\mathcal{T}\langle \mathcal{P}_\rho^N \rangle\right), \frac{m}{M}\right) - 1, \tag{2}$$

where $\mathcal{T}\langle \mathcal{P}_- \rangle$ is the BST of a uniform permutation of random size $\text{POISSON}\left(n\zeta m\right)$. We proceed as before to conclude the proof of Equation (1).

For the uniform integrability claim, it suffices to establish boundedness of $\mathbb{E}\left[\frac{h\left(\mathcal{T}\langle \mathcal{P}_\rho^N \rangle\right)^p}{\log(n)^p}\right]$ in n, for all $p > 0$. Conditionally given $h\left(\mathcal{T}\langle \mathcal{P}_\rho^N \rangle\right)$, write $S_n + 1$ for a random variable with distribution $\text{BINOMIAL}\left(1 + h\left(\mathcal{T}\langle \mathcal{P}_\rho^N \rangle\right), \frac{m}{M}\right)$. Then, using Hoeffding's inequality:

$$\mathbb{P}\left[S_n < \frac{m}{2M}\left(1 + h\left(\mathcal{T}\langle \mathcal{P}_\rho^N \rangle\right)\right) - 1 \,\Big|\, h\left(\mathcal{T}\langle \mathcal{P}_\rho^N \rangle\right)\right] \leq e^{-\frac{m^2}{2M^2}\left(1 + h\left(\mathcal{T}\langle \mathcal{P}_\rho^N \rangle\right)\right)}$$

and therefore, by discriminating according to this event for any $n \geq e$:

$$\mathbb{E}\left[\frac{h\left(\mathcal{T}\langle \mathcal{P}_\rho^N \rangle\right)^p}{\log(n)^p}\right] \leq \mathbb{E}\left[h\left(\mathcal{T}\langle \mathcal{P}_\rho^N \rangle\right)^p e^{-\frac{m^2}{2M^2}\left(1 + h\left(\mathcal{T}\langle \mathcal{P}_\rho^N \rangle\right)\right)}\right] + \mathbb{E}\left[\frac{\left((2M/m) \cdot (S_n + 1) - 1\right)^p}{\log(n)^p}\right].$$

Since the function $x \mapsto x^p e^{-\frac{m^2}{2M^2}(1+x)}$ is bounded over $\mathbb{R}_+$, the first term is bounded in n. For the second term, we use $(a + b)^p \leq 2^{p-1}(a^p + b^p)$ along with (2) to deduce:

$$\mathbb{E}\left[\frac{\left((2M/m) \cdot (S_n + 1)\right)^p}{\log(n)^p}\right] \leq 2^{p-1}\left(\frac{2M}{m}\right)^p\left(\mathbb{E}\left[\frac{h\left(\mathcal{T}\langle \mathcal{P}_- \rangle\right)^p}{\log(n)^p}\right] + \frac{1}{\log(n)^p}\right)$$

which is bounded in n by [10, Lemma 3.1] and Poisson estimates (indeed, recall that $\mathcal{T}\langle \mathcal{P}_- \rangle$ is the BST of a uniform permutation of random size $\text{POISSON}\left(n\zeta m\right)$). This concludes the proof. ◀

The weakness of the previous proposition is that η, which depends on the rectangle under consideration, might be large. In the next statement we show that, for continuous positive densities ρ, it is possible to choose rectangles for which the corresponding η is small.

▶ **Corollary 7.** *Let D be a compact domain in the plane and $\rho : D \to (0, \infty)$ be a continuous, positive intensity function. Then for any $\varepsilon > 0$, there exists $\beta > 0$ such that for any rectangle $R = [x_1, x_1 + \beta] \times [y_1, y_2]$ with non-empty interior contained in D:*

$$\lim_{n \to \infty} \mathbb{P}\left[\left|\frac{h\left(\mathcal{T}\langle \mathcal{P}_\rho^N \cap R\rangle\right)}{c^* \log n} - 1\right| > \varepsilon\right] = 0\,.$$

In particular, taking $x_1 = y_1 = 0$ and $y_2 = 1$, the tree $\mathcal{T}\langle \mathcal{P}_\rho^N \cap R\rangle$ is the top tree $\mathcal{T}\langle \mathcal{P}(\beta)\rangle$ defined in Section 2.1. This top tree therefore has height $(c^* + \varepsilon) \log n$, for small enough β and under assumption (A1).

Proof. Let $\varepsilon > 0$ and assume that $\varepsilon < \min_D \rho$. By uniform continuity of ρ, we can find $\beta > 0$ such that for any $(x, y), (x', y') \in D$, the inequality $|x - x'| + |y - y'| \leq \beta$ implies $|\rho(x, y) - \rho(x', y')| \leq \varepsilon$. Then consider $R = [x_1, x_1 + \beta] \times [y_1, y_2]$ contained in D. Define

$$f : y \in [y_1, y_2] \mapsto \int_{y_1}^y \rho(x_1, t)dt \quad \text{and} \quad g : y \in [y_1, y_2] \mapsto y_1 + (y_2 - y_1)f(y)/f(y_2)\,.$$

The function g is a $\mathcal{C}^1$ increasing map from $[y_1, y_2]$ onto itself. Let $\widetilde{\mathcal{P}}$ denote the set of points obtained after applying the transformation $(x, y) \mapsto (x, g(y))$ to $\mathcal{P}_\rho^N \cap R$. This transformation does not change the relative orders of points, therefore $\mathcal{T}\langle \widetilde{\mathcal{P}}\rangle$ and $\mathcal{T}\langle \mathcal{P}_\rho^N \cap R\rangle$ have the same shape. Additionally, $\widetilde{\mathcal{P}}$ follows the law of a Poisson point process with intensity

$$n \frac{\rho(x, g^{-1}(y))}{g'(g^{-1}(y))} = n \frac{f(y_2)}{y_2 - y_1} \frac{\rho(x, g^{-1}(y))}{\rho(x_1, g^{-1}(y))}$$

on R. Thus we can apply Proposition 6 with $\eta = \frac{2\varepsilon}{\min_D \rho - \varepsilon}$ to obtain:

$$\lim_{n \to \infty} \mathbb{P}\left[\left|\frac{h\left(\mathcal{T}\langle \tilde{\mathcal{P}}\rangle\right)}{c^* \log n} - 1\right| > \eta + \varepsilon\right] = 0\,.$$

Since this holds for any small enough $\varepsilon > 0$, and $\eta \to 0$ as $\varepsilon \to 0$, the result follows. ◀

2.4 Extreme deviation bounds, via monotone subsequences

It remains to argue that the hanging trees simultaneously all have height $o_{\mathbb{P}}(\log n)$. A "typical" horizontal band in Figure 3 contains $\mathcal{O}(1)$ points, but their maximum is actually $\mathcal{O}(\log n)$ (Proposition 11). The hanging trees are themselves BSTs of point processes, and therefore they individually have height $\mathcal{O}(\log \log n) \ll \log n$. To have this bound for all $\mathcal{O}(n)$ hanging trees simultaneously, we need adequate deviation estimates for the BST height of point processes. Such estimates are provided by Devroye for uniform BSTs [10], but the monotonicity properties of BSTs are not good enough to use direct comparison arguments. We solve this by relating the BST height of a point set to its longest monotone subsequences, for which we have good monotonicity properties and deviation bounds.

Let σ be a permutation of $\{1, \ldots, n\}$. An increasing subsequence of σ is a sequence of indices $i_1 < \cdots < i_k$ such that $\sigma(i_1) < \cdots < \sigma(i_k)$. The maximum length of an increasing subsequence of σ is then denoted by $\mathrm{LIS}(\sigma)$. We define similarly $\mathrm{LDS}(\sigma)$, the maximum length of a decreasing subsequence of σ.

▶ **Lemma 8.** *For any permutation σ, we have $h\left(\mathcal{T}\langle\sigma\rangle\right) \leq \mathrm{LIS}\left(\sigma\right) + \mathrm{LDS}\left(\sigma\right)$.*

Proof. Let $i_1 < \cdots < i_k$ be a sequence of integers such that $\sigma(i_1), \ldots, \sigma(i_k)$ label nodes on a chain C of $\mathcal{T}\langle\sigma\rangle$. Define $\mathcal{I}_\mathcal{R}$ (resp. $\mathcal{I}_\mathcal{L}$) as the family of i_j's such that the node following $\sigma(i_j)$ in C lies in its right subtree (resp. left subtree). By construction, $\mathcal{I}_\mathcal{R} \cup \{i_k\}$ and $\mathcal{I}_\mathcal{L} \cup \{i_k\}$ form respectively an increasing and a decreasing subsequence of σ. The lemma follows. ◄

Combining this lemma with [6, Proposition 3.2], we get that if ρ is an integrable function then $h\left(\mathcal{T}\langle\mathcal{P}_\rho^N\rangle\right) = o_\mathbb{P}(n)$. We will need a more quantitative version of this, valid only for bounded functions ρ. We start with the following lemma, proved by a straightforward application of the first moment method.

▶ **Lemma 9.** *For each integer n, let σ^n be a uniform permutation of $\{1, \ldots, n\}$. Then:*

$$\mathbb{P}\left[\mathrm{LIS}\left(\sigma^n\right) \geq \frac{n}{\log n}\right] \leq \exp\left(-n + o(n)\right).$$

Using the previous two lemmas and standard techniques, we obtain a useful corollary:

▶ **Corollary 10.** *For any $M > 0$ and $\varepsilon > 0$, there exists $n_0 = n_0(M, \varepsilon)$ such that the following holds. For any $0 < \zeta \leq 1$, any function $\rho : [0,1]^2 \to [0,\infty)$ bounded by M and supported on some rectangle $[a,b] \times [c,d]$ with $(b-a)(d-c) \leq \zeta$, and for any integer $n > n_0/\zeta$:*

$$\mathbb{P}\left[h\left(\mathcal{T}\langle\mathcal{P}_\rho^N\rangle\right) > 2\varepsilon\zeta n\right] \leq 4\exp\left(-\tfrac{\varepsilon}{2}\zeta n \log(\zeta n)\right).$$

We refer to [8] for the full proof. The key argument is that LIS and LDS are, unlike the height of BSTs, monotone in their arguments: if $\mathcal{P}_- \subseteq \mathcal{P}_+$ are generic point sets, then $\mathrm{LIS}\left(\sigma\langle\mathcal{P}_-\rangle\right) \leq \mathrm{LIS}\left(\sigma\langle\mathcal{P}_+\rangle\right)$, and likewise for LDS. We can thus compare $\mathcal{P}_\rho^N$ to a *homogeneous* Poisson point process with higher intensity, and use extreme deviation bounds for the monotone subsequences of the latter.

2.5   Controlling the height of the hanging trees

Throughout the rest of this section, we use the notation of Section 2.1. For each integer $0 \leq k \leq K_\beta$, we let $\zeta_k := |I_k|$ be the vertical length of the band $(\beta, 1] \times I_k$.

▶ **Proposition 11.** *Let μ be a permuton. Assume that there exists $\beta > 0$ such that $\mu_{/[0,\beta] \times [0,1]}$ has a continuous and positive density $\rho : [0,\beta] \times [0,1] \to (0,\infty)$. Then the following holds.*
1. *There exists $\alpha > 0$ such that $\max_k \zeta_k \leq \alpha \frac{\log n}{n}$ w.h.p. as $n \to \infty$.*
2. *All powers of $\frac{1}{\log n} \max_k \left|\mathcal{P}_k(\beta)\right|$ are uniformly integrable.*

Sketch of proof. The first item can be derived using standard results on "maximal spacings". Indeed, by a thinning procedure, $\max_k \zeta_k$ is bounded above by the largest gap among $\mathrm{POISSON}\left(n\beta m\right)$ i.i.d. uniform variables in $[0,1]$, where m is a lower bound for ρ. This is known to concentrate around $\log(n)/(n\beta m)$ [16], which proves the first item. To prove the second item, we can use that conditionally given $\mathcal{P}(\beta)$, the number $\left|\mathcal{P}_k(\beta)\right|$ has distribution $\mathrm{POISSON}\left(n(1-\beta)\zeta_k\right)$. Then, conclude with item (1) and Poisson estimates. ◄

▶ **Proposition 12.** *Let μ be a permuton satisfying* (A1)*. Then for any $\beta \in (0,1)$ we have the following convergence in probability as n goes to infinity:*

$$\frac{1}{\log n} \max_{0 \leq k \leq K_\beta} \left\{h\left(\mathcal{T}\langle\mathcal{P}_k(\beta)\rangle\right)\right\} \longrightarrow 0.$$

Proof. From Proposition 11, item (1), there exists $\alpha > 0$ such that w.h.p. $\max_k \zeta_k < \alpha \frac{\log n}{n}$. Work under this event, and conditionally given $\mathcal{P}(\beta)$. Then for each k, $\mathcal{P}_k(\beta)$ is distributed like a Poisson point process with intensity $n\rho_{/[\beta,1] \times [y_{(k)}, y_{(k+1)}]}$. For each k, Corollary 10 applies with ρ restricted to $[\beta,1] \times [y_{(k)}, y_{(k+1)}]$ and $\zeta = \alpha \frac{\log n}{n}$. With a union bound, we deduce:

$$\mathbb{P}\left[\max_{0 \leq k \leq |\mathcal{P}(\beta)|} h\left(\mathcal{T}\langle \mathcal{P}_k(\beta)\rangle\right) > \delta \log n\right] \leq 4(|\mathcal{P}(\beta)| + 1) \exp\left(-\frac{\delta}{4\alpha} \alpha \log n \, \log(\alpha \log n)\right)$$

for $\zeta n = \alpha \log n$ large enough. But w.h.p. we have $\max_k \zeta_k < \alpha \frac{\log n}{n}$ and $|\mathcal{P}(\beta)| < n$, so the unconditioned probability tends to 0 as $n \to \infty$. This proves the proposition. ◄

2.6 Concluding the proof of the height theorem

Proof of Theorem 1. Fix $\varepsilon > 0$. Let D be a compact neighborhood of $\{0\} \times [0,1]$ on which ρ is continuous and positive, and let $\beta = \beta(\varepsilon) > 0$ be given by Corollary 7 applied to ρ on D. Therefore, if $\mathcal{T}\langle \mathcal{P}(\beta)\rangle := \mathcal{T}\langle \mathcal{P}_\mu^N \cap ([0,\beta] \times [0,1])\rangle$ denotes the top tree of $\mathcal{T}\langle \mathcal{P}_\mu^N \rangle$:

$$\lim_{n \to \infty} \mathbb{P}\left[\left|\frac{h\left(\mathcal{T}\langle \mathcal{P}(\beta)\rangle\right)}{c^* \log n} - 1\right| > \varepsilon\right] = 0.$$

Furthermore, by Proposition 12, the quantity

$$\frac{1}{\log n} \max_{0 \leq k \leq K_\beta} \left\{h\left(\mathcal{T}\langle \mathcal{P}_k(\beta)\rangle\right)\right\}$$

converges in probability to 0. Combined with Lemma 2, this implies that $\frac{1}{\log n} h\left(\mathcal{T}\langle \mathcal{P}_\mu^N \rangle\right)$ converges in probability to c^*.

Together with Proposition 6 and Proposition 11, Lemma 2 also implies uniform integrability of all powers of $\frac{1}{\log n} h\left(\mathcal{T}\langle \mathcal{P}_\mu^N \rangle\right)$. Therefore the convergence holds in L^p for all $p \geq 1$, concluding the proof of Theorem 1. ◄

3 Second main result: subtree size convergence of the BSTs

3.1 Some definition, and statement of the result

Next, we state a limit theorem for $\mathcal{T}\langle \mathcal{P}_\mu^N \rangle$, in the sense of the *subtree size convergence* recently introduced by Grübel [13]. We start by recalling this notion of convergence.

Identify nodes in a binary tree with finite words in the alphabet $\{0, 1\}$ as follows: the empty word $\varnothing$ corresponds to the root, and for a node v encoded by w, the words $w0$ and $w1$ encode respectively the left and right children of v. Let $\mathbb{V} = \{0, 1\}^*$ be the set of all finite words on $\{0, 1\}$, representing all nodes of the complete infinite binary tree. A labeled tree is then identified with a function from a subset of $\mathbb{V}$ to $\mathbb{R}$, where the domain of the function is the set of nodes in the tree, and a node is mapped to its label. In particular, $\mathcal{T}(v)$ denotes the label of the node v in $\mathcal{T}$. We also write $v \in \mathcal{T}$ to indicate that the node v is in $\mathcal{T}$. Given a finite tree $\mathcal{T}$ and a node $v \in \mathbb{V}$, define

$$t(\mathcal{T}, v) := \frac{1}{|\mathcal{T}|} \left|\left\{u \in \mathcal{T} : v \preceq u\right\}\right|,$$

where $v \preceq u$ means that v is a prefix of u. In words, $t(\mathcal{T}, v)$ is the proportion of nodes in $\mathcal{T}$ which are descendants of v.

Further write Ψ for the set of functions $\psi : \mathbb{V} \to [0,1]$ such that $\psi(\varnothing) = 1$ and for any $v \in \mathbb{V}$, we have $\psi(v) = \psi(v0) + \psi(v1)$. Then a sequence of binary trees $(\mathcal{T}^n)_{n\in\mathbb{N}}$ is said to converge to a function $\psi \in \Psi$ if and only if $t(\mathcal{T}^n, v) \to \psi(v)$ for all $v \in \mathbb{V}$. If that is the case, we write $\mathcal{T}_n \xrightarrow{\text{ssc}} \psi$ and refer to this as *subtree size convergence*.

We now define two important objects before stating our result. For any complete BST $\mathcal{T} : \mathbb{V} \to (0,1)$, we define $\mathcal{T}_{\text{left}} : \mathbb{V} \to \mathbb{R}$ as follows. First, for any $v \in \{0\}^*$, let $\mathcal{T}_{\text{left}}(v) := 0$. Then if $v = v'10^k$ for some $k \geq 0$, let $\mathcal{T}_{\text{left}}(v) := \mathcal{T}(v')$. Informally, $\mathcal{T}_{\text{left}}(v)$ is the right-most ancestor of v to its left. Define similarly $\mathcal{T}_{\text{right}}$ such that $\mathcal{T}_{\text{right}}(v) := 1$ for any $v \in \{1\}^*$ and $\mathcal{T}_{\text{right}}(v) := \mathcal{T}(v')$ whenever $v = v'01^k$ for some $k \geq 0$. We note that this definition implies that $\mathcal{T}_{\text{left}}(v) < \mathcal{T}(v) < \mathcal{T}_{\text{right}}(v)$ for any $v \in \mathbb{V}$.

Given a probability measure m on $[0,1]$ without atoms, write $\psi_m \in \Psi$ for the following random object. First, let $Y = (Y_1, Y_2, \ldots)$ be an i.i.d. variables distributed according to m and write $\mathcal{T}^m := \mathcal{T}\langle Y \rangle$ for the corresponding (infinite) BST. Then, let $\psi_m := \mathcal{T}^m_{\text{right}} - \mathcal{T}^m_{\text{left}}$. This is well-defined, since $\mathcal{T}^m$ is a.s. complete [10, Theorem 6.1]. It is immediate to check that indeed $\psi_m \in \Psi$ (almost surely). See Figure 4 for an example.
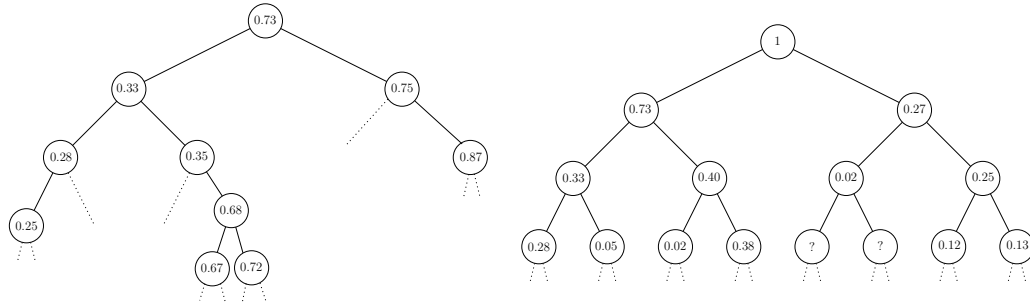
Figure 4 Example of realizations of $\mathcal{T}^m$ and ψ_m. Note that we do not have enough data to compute two of the values of ψ_m on nodes in the third level.

We can now state our second main result. A permuton is said to satisfy assumption (A2) if there exists a probability measure μ_0 on $[0,1]$, without atoms, such that

$$\frac{1}{x}\mu\big([0,x] \times \cdot\big) \xrightarrow[x\to 0]{} \mu_0 \tag{3}$$

for the weak topology. Assumption (A2) is weaker than (A1): in particular, (A2) holds whenever μ admits a continuous density on a neighborhood of $\{0\} \times [0,1]$.

▶ **Theorem 13** (Subtree size convergence of BSTs for permuton samples). *Let μ be a permuton satisfying* (A2). *The following convergence in distribution holds for the subtree size topology:*

$$\mathcal{T}\langle \mathcal{P}^N_\mu \rangle \xrightarrow{\text{ssc}} \psi_{\mu_0}.$$

Note that the limit depends on μ only through μ_0. The assumption that μ_0 does not have atoms is important. A first difficulty when μ_0 has some atom is that the BST $\mathcal{T}\langle Y_1, Y_2, \ldots \rangle$ where $Y_1, Y_2, \ldots,$ are i.i.d. variables with distribution μ_0 is ill-defined, since some of the Y_i's are equal. We can also see that, in this case, the limit of $\mathcal{T}\langle \mathcal{P}^n_\mu \rangle$ may not depend only on μ_0. Indeed, consider the permutons μ^1 and μ^2 supported by the sets $y \equiv \frac{1}{2} + x \mod 1$ and $y \equiv \frac{1}{2} - x \mod 1$. They both satisfy (3) with $\mu_0 = \delta_{1/2}$, but it is easy to see that their BSTs have different limits in the sense of subtree size convergence.

3.2 Preliminaries to the proof

We start with a variant of the Glivenko–Cantelli theorem for triangular arrays.

▶ **Proposition 14.** *Let μ be a probability measure with a finite fourth moment, and distribution function $F(x) := \mu((-\infty, x])$. For each $n \geq 1$, let $(X_{i,n})_{1 \leq i \leq n}$ be i.i.d. random variables with common distribution μ and let $F_n(x) := \frac{1}{n} \left| \{ i \leq n : X_{i,n} \leq x \} \right|$ be their empirical distribution function. Then F_n converges a.s. uniformly to F.*

Proof. A classical fourth moment computation, together with Borel–Cantelli lemma – see e.g. [4, Theorem 6.1] – shows that, for any fixed x, $F_n(x)$ converges a.s. to $F(x)$. The rest of the proof is similar to that of the classical Glivenko–Cantelli theorem which considers a single sequence X_i of i.i.d. random variables instead of a triangular array, but does not require a fourth moment condition; see e.g. [4, Theorem 20.6]. ◀

Under assumption (A2), we can prove convergence in distribution of the leftmost points in $\mathcal{P}_\mu^N$. The proof of the following proposition is rather technical, and can be found in [8].

▶ **Proposition 15.** *Let μ be a permuton satisfying* (A2), *and let $\mathcal{P}_\mu^N = \left\{ \left(X_i^N, Y_i^N \right), 1 \leq i \leq N \right\}$ be a Poisson point process with intensity $n\mu$. Let $\left(\left(X_{(i)}^N, Y_{(i)}^N \right) \right)_{1 \leq i \leq N}$ be its reordering such that $X_{(1)}^N < \cdots < X_{(N)}^N$. Then, for any fixed $K \geq 1$, we have the following convergence in distribution:*

$$\left(Y_{(1)}^N, \ldots, Y_{(K)}^N \right) \xrightarrow[n \to \infty]{} (Y_k)_{1 \leq k \leq K}$$

where $(Y_k)_{1 \leq k \leq K}$ is a sequence of i.i.d. random variables distributed according to μ_0.

Finally, with the notation of Section 3.1, we can use the functions $\mathcal{T}_{\text{left}}$ and $\mathcal{T}_{\text{right}}$ to describe the descendants of nodes in $\mathcal{T}$. The proof is straightforward.

▶ **Lemma 16.** *Let $y_1, \ldots, y_n$ be distinct numbers and let $\mathcal{T} := \mathcal{T}\langle y_1, \ldots, y_n \rangle$ be the corresponding BST. Let u be a node in $\mathcal{T}$ and let k be such that $\mathcal{T}(u) = y_k$. Then:*

$$t(\mathcal{T}, u) = \frac{1}{|\mathcal{T}|} \left| \left\{ y_k, \ldots, y_n \right\} \cap \left(\mathcal{T}_{\text{left}}(u), \mathcal{T}_{\text{right}}(u) \right) \right|.$$

3.3 Proof of subtree size convergence

Proof of Theorem 13. Write $\mathcal{T}^N := \mathcal{T}\langle \mathcal{P}_\mu^N \rangle$. Since the subtree size topology is by definition the pointwise convergence of the function $(t(., u))_{u \in \mathbb{V}}$, we need to prove the convergence of finite-dimensional distributions. Namely we need to prove that, for any $d \geq 1$ and $u_1, \ldots, u_d \in \mathbb{V}$, we have the following convergence in distribution as $n \to \infty$:

$$\left(t(\mathcal{T}^N, u_i) \right)_{i \leq d} \longrightarrow \left(\psi_\mu(u_i) \right)_{i \leq d}. \tag{4}$$

Recall the notation of Proposition 15. Using Skorohod's representation theorem [3, Section 6], we might assume that the convergence

$$\left(Y_{(1)}^N, \ldots, Y_{(K)}^N \right) \xrightarrow[n \to \infty]{} (Y_k)_{1 \leq k \leq K} \tag{5}$$

holds almost surely. Since μ_0 has no atoms, the numbers $(Y_k)_{k \geq 1}$ are a.s. distinct. Moreover the tree $\mathcal{T}\langle Y_1, Y_2, \ldots \rangle$ has a.s. shape $\mathbb{V}$. Consequently, a.s. there exists a (random) threshold K such that all nodes $u_i, i \leq d$ belong to $\mathcal{T}\langle Y_1, \ldots, Y_K \rangle$. Using (5), there exists a (random)

threshold n_0 such that for all $n \geq n_0$, the relative order of $(Y_{(1)}^N, \ldots, Y_{(K)}^N)$ is the same as that of $(Y_1, \ldots, Y_K)$. Hence the trees $\mathcal{T}_K^N := \mathcal{T}\langle Y_{(1)}^N, \ldots, Y_{(K)}^N \rangle$ and $\mathcal{T}_K^\infty := \mathcal{T}\langle Y_1, \ldots, Y_K \rangle$ have the same shape T_K. Moreover for any v in T_K, the labels $\mathcal{T}_K^N(v)$ and $\mathcal{T}_K^\infty(v)$ equal $Y_{(i)}^N$ and Y_i respectively, for the *same* index i. Therefore $\mathcal{T}_K^N(v) \to \mathcal{T}_K^\infty(v)$ as $n \to \infty$, a.s. in the probability space created by the application of Skorohod's representation theorem.

Now, using Lemma 16 and the fact that each u_i is filled in $\mathcal{T}^N$ before step $K = \mathcal{O}_\mathbb{P}(1)$:

$$t\left(\mathcal{T}^N, u_i\right) = \frac{1}{N} \left| \left\{ Y_1^N, \ldots, Y_N^N \right\} \cap \left(\mathcal{T}_{\text{left}}^N(u_i), \mathcal{T}_{\text{right}}^N(u_i) \right) \right| + o_\mathbb{P}(1).$$

Consider the empirical distribution function $F_N(y) := \frac{1}{N} \left| \{Y_1^N, \ldots, Y_N^N\} \cap (-\infty, y) \right|$. Then:

$$t\left(\mathcal{T}^N, u_i\right) = F_N\left(\mathcal{T}_{\text{right}}^N(u_i)\right) - F_N\left(\mathcal{T}_{\text{left}}^N(u_i)\right) + o_\mathbb{P}(1).$$

The random variable $N \sim \textsc{Poisson}(n)$ is well-concentrated around n, and conditionally given N, the points $Y_1^N, \ldots, Y_N^N$ are i.i.d. random variables in $[0, 1]$. Since μ is a permuton, their common (conditional) distribution is the uniform distribution. From Proposition 14, we infer that F_n converges a.s. uniformly on $[0, 1]$ to the identity function (the earlier use of Skorohod's representation theorem implies that the $(Y_i^n)_{1 \leq i \leq n}$ are coupled in a nontrivial way for different values of n, but Proposition 14 applies nevertheless).

Moreover, the above discussion implies that $\mathcal{T}_{\text{right}}^N(u_i)$ and $\mathcal{T}_{\text{left}}^N(u_i)$ converge a.s. to $\mathcal{T}_{\text{right}}^\infty(u_i)$ and $\mathcal{T}_{\text{left}}^\infty(u_i)$ respectively. Therefore, a.s. in the probability space created by the application of Skorohod's representation theorem, for all $i \leq d$ we have:

$$t\left(\mathcal{T}^N, u_i\right) = \mathcal{T}_{\text{right}}^\infty(u_i) - \mathcal{T}_{\text{left}}^\infty(u_i) + o_\mathbb{P}(1) = \psi_{\mu_0}(u_i) + o_\mathbb{P}(1).$$

Since a.s. (joint) convergence implies (joint) convergence in distribution, (4) is proved.   ◀

4   Extra results

In this last section, we briefly discuss some additional results and open questions. More details can be found in the full paper [8].

De-Poissonization. As mentioned in the introduction, it is possible to state Theorems 1 and 13 for $\mathcal{P}_\mu^n$ (a set of n i.i.d. points under μ) instead of $\mathcal{P}_\mu^N$ (a Poisson point process with intensity $n\mu$). In other words, it is possible to "de-Poissonize" the random size N into a deterministic size n. This is rather technical, and hinges on the comparison method of Proposition 5 along with standard estimates on the Poisson law.

Optimality of hypotheses in Theorem 1. Assumption (A1) is in some sense optimal for the upper bound on the BST height. Indeed, in [8] we exhibit two permutons which do not quite satisfy (A1), and whose BSTs are much deeper.

Universality of the lower bound for the BST height. On the other hand, we could not construct a permuton μ such that $h\left(\mathcal{T}\langle\mathcal{P}_\mu^n\rangle\right)$ is asymptotically smaller than $c^* \log(n)$. This leads us to conjecture that the BSTs of permuton samples always satisfy this lower bound. In [8], we prove a partial result in this direction.

─── **References** ───

1   Louigi Addario-Berry and Benoît Corsini. The height of Mallows trees. *Ann. Probab.*, 49(5):2220–2271, 2021. doi:10.1214/20-AOP1503.

2   Frédérique Bassino, Mathilde Bouvel, Valentin Féray, Lucas Gerin, Mickaël Maazoun, and Adeline Pierrot. Universal limits of substitution-closed permutation classes. *J. Eur. Math. Soc. (JEMS)*, 22(11):3565–3639, 2020. doi:10.4171/JEMS/993.

3 Patrick Billingsley. *Convergence of probability measures.* Wiley Ser. Probab. Stat. Chichester: Wiley, 2nd ed. edition, 1999.

4 Patrick Billingsley. *Probability and measure. Anniversary edition.* Hoboken, NJ: John Wiley & Sons, 2012.

5 Jacopo Borga, Sayan Das, Sumit Mukherjee, and Peter Winkler. Large deviation principle for random permutations. *International Mathematics Research Notices*, rnad096, 2023. `doi:10.1093/imrn/rnad096`.

6 Jacopo Borga, Ewain Gwynne, and Xin Sun. Permutons, meanders, and sle-decorated liouville quantum gravity, 2022. Preprint arXiv:2207.02319. `doi:10.48550/arXiv.2207.02319`.

7 Benoît Corsini. The height of record-biased trees. *Random Structures & Algorithms*, 62(3):623–644, 2023. `doi:10.1002/rsa.21110`.

8 Benoît Corsini, Victor Dubach, and Valentin Féray. Binary search trees of permuton samples, 2024. Preprint arXiv:2403.03151. `doi:10.48550/arXiv.2403.03151`.

9 Jean-Dominique Deuschel and Ofer Zeitouni. Limiting curves for i.i.d. records. *Ann. Probab.*, 23(2):852–878, 1995.

10 Luc Devroye. A note on the height of binary search trees. *Journal of the ACM (JACM)*, 33(3):489–498, 1986.

11 Victor Dubach. Increasing subsequences of linear size in random permutations and the Robinson–Schensted tableaux of permutons, 2023. Preprint arXiv:2307.05768. `doi:10.48550/arXiv.2307.05768`.

12 Victor Dubach. Locally uniform random permutations with large increasing subsequences. *Combinatorial Theory*, 3(3), 2023. `doi:10.5070/C63362784`.

13 Rudolf Grübel. A note on limits of sequences of binary trees. *Discrete Mathematics & Theoretical Computer Science*, 25(Analysis of Algorithms), 2023. `doi:10.46298/dmtcs.10968`.

14 Carlos Hoppen, Yoshiharu Kohayakawa, Carlos Gustavo Moreira, Balázs Ráth, and Rudini Menezes Sampaio. Limits of permutation sequences. *J. Comb. Theory, Ser. B*, 103(1):93–113, 2013. `doi:10.1016/j.jctb.2012.09.003`.

15 Jonas Sjöstrand. Monotone subsequences in locally uniform random permutations. *Ann. Probab.*, 51(4):1502–1547, 2023. `doi:10.1214/23-AOP1624`.

16 Eric Slud. Entropy and maximal spacings for random partitions. *Z. Wahrscheinlichkeitstheor. Verw. Geb.*, 41:341–352, 1978. `doi:10.1007/BF00533604`.

The Recurrence/Transience of Random Walks on a Bounded Grid in an Increasing Dimension

Shuma Kumamoto ✉
Graduate School of Mathematical Science, Kyushu University, Fukuoka, Japan

Shuji Kijima ✉ 🄓
Faculty of Data Science, Shiga University, Hikone, Japan

Tomoyuki Shirai ✉ 🄓
Institute of Mathematics for Industry, Kyushu University, Fukuoka, Japan

―――― **Abstract** ――――

It is celebrated that a simple random walk on $\mathbb{Z}$ and $\mathbb{Z}^2$ returns to the initial vertex v infinitely many times during infinitely many transitions, which is said *recurrent*, while it returns to v only finite times on $\mathbb{Z}^d$ for $d \geq 3$, which is said *transient*. It is also known that a simple random walk on a growing region on $\mathbb{Z}^d$ can be recurrent depending on growing speed for any *fixed d*. This paper shows that a simple random walk on $\{0, 1, \ldots, N\}^n$ with an *increasing n* and a fixed N can be recurrent depending on the increasing speed of n. Precisely, we are concerned with a specific model of a *random walk on a growing graph* (*RWoGG*) and show a phase transition between the recurrence and transience of the random walk regarding the growth speed of the graph. For the proof, we develop a *pausing coupling* argument introducing the notion of *weakly less homesick as graph growing* (*weakly LHaGG*).

2012 ACM Subject Classification Theory of computation → Random walks and Markov chains

Keywords and phrases Random walk, dynamic graph, recurrence, transience, coupling

Digital Object Identifier 10.4230/LIPIcs.AofA.2024.22

1 Introduction

The recurrence or transience is a classical and fundamental topic of random walks on *infinite* graphs, see e.g., [16]: let $X_0, X_1, X_2, \ldots$ be a random walk (or a Markov chain) on an infinite state space V, e.g., $V = \mathbb{Z}$, with $X_0 = v$ for $v \in V$. The random walk is said to be *recurrent* at the initial state v if

$$\sum_{t=1}^{\infty} \Pr[X_t = v] = \infty \tag{1}$$

holds, otherwise it is said to be *transient*. Intuitively, (1) means that the random walk is "expected" to return to the initial state infinitely many times. An interesting fact is that a simple random walk on $\mathbb{Z}$ or $\mathbb{Z}^2$ is recurrent, while a simple random walk on $\mathbb{Z}^d$ is transient for $d \geq 3$, cf. [16].

Analysis of random walks on *dynamic graphs* has been developed in several contexts. Random walks in random environment is a popular subject in probability theory, where self-interacting random walks including reinforced random walks and excited random walks have been intensively investigated as a relatively tractable non-Markovian process, see e.g., [9, 5, 15, 29, 30, 21]. The recurrence or transience of a random walk in a random environment

is a major topic there, particularly random walks on growing subgraphs of $\mathbb{Z}^d$ or infinitely growing trees are the major targets [11, 12, 18, 1]. In distributed computing, analysis of algorithms including random walks on dynamic graphs attracts increasing attention because networks are often dynamic [7, 22, 2, 28]. Searching or covering networks, related to hitting or cover times of random walks, are major topics there [8, 3, 14, 4, 24, 6, 20].

Existing works.   As we stated above, a simple random walk on the infinite integer grid $\mathbb{Z}^d$ is recurrent for $d = 1$ and 2, while it is transient for $d \geq 3$. Dembo et al. [12] investigated a random walk on an infinitely *growing* region of $\mathbb{Z}^d$ and showed a phase transition, that is roughly speaking a random walk is recurrent if and only if $\sum_{t=1}^{\infty} \pi_t(\mathbf{0}) = \infty$ holds under certain conditions, where π_t denotes the stationary distribution of the transition matrix at time t. Huang [18] extended the argument of [12] and gave a similar or essentially the same phase transition for more general graphs. The proofs are based on the edge conductance and the central limit theorem on the assumption that every vertex of the dynamic graph has a degree at most *constant* to time (or the size of the graph). Those arguments are sophisticated and enhanced using the argument of evolving set and the heat kernel by recent works [10, 13].

Kumamoto et al. [23] were concerned with a specific model called *random walk on growing graph* (*RWoGG*), which is parametrized by $\mathfrak{d} \colon \mathbb{Z}_{\geq 0} \to \mathbb{Z}_{\geq 0}$ representing the growing (inverse) speed of the graph. Then, they investigated a simple random walk on $\{0, 1\}^n$ *with an increasing* n, and showed that the random walk is recurrent if $\sum_{n=1}^{\infty} \mathfrak{d}(n)/2^n = \infty$, otherwise transient. Notice that the degree of every vertex of the $\{0, 1\}^n$ skeleton graph infinitely grows as $n \to \infty$. They introduced the notion of *less-homesickness as graph growing* (*LHaGG*) and gave a proof by a coupling argument, which is easier than the arguments based on the conductance or heat kernel, for this specific object. However, the proof technique is not simply applicable to a simple random walk on $\{0, \ldots, N\}^n$ with an increasing n (and a fixed N), and it remained as future work.

Result.   This paper is concerned with the RWoGG model (see Sec. 2.1), and shows a phase transition by the growing speed regarding a random walk being recurrent/transient for a lazy simple random walk on $\{0, \ldots, N\}^n$ with an increasing n and a fixed N. For this purpose, we introduce the notion of *weakly* less-homesick as graph growing (*weakly* LHaGG; see Sec. 3) and show sufficient conditions for a weakly LHaGG RWoGG to be recurrent (Thm. 2) or transient (Thm. 4). The notion of weakly LHaGG is quite intuitive and natural, but we have to develop a new technique of *pausing coupling* to prove that a lazy simple RWoGG is weakly LHaGG. Then, we give the threshold $\sum_{k=1}^{\infty} \mathfrak{d}(k)/(2N)^k = \infty$ of the phase transition (Thm. 6).

Other related works.   It is another celebrated fact that a simple random walk on an infinite k-ary tree is transient for $k \geq 2$ [26, 27]. Amir et al. [1] introduced a random walk in a changing environment model and investigated the recurrence and transience of random walks in the model. They gave a conjecture about the conditions of the recurrence and transience regarding the limit of a graph sequence and proved it for trees. Huang's work [18] implicitly implies a phase transition between the recurrence and transience of the random walk on a growing k-ary tree regarding the growing speed of the graph, based on the conductance arguments. Kumamoto et al. [23] explicitly showed the phase transition for a growing k-ary tree under the RWoGG model, where they employ a coupling argument.

There is a lot of work on the recurrence or transience of random walks on growing trees in the context of self-interacting random walks including reinforced random walks and excited random walks, e.g., [19, 17]. They are non-Markovian processes, and in a bit different line from [12, 1, 18, 23] and this paper.

Related to the cover time, which is another major topic on random walks, Cooper and Frieze [8] investigated the covering rate of a random walk on the "web-graph" model, where the graph grows at a constant speed. Kijima et al. [20] introduced the RWoGG model, where the growing (inverse) speed of a graph is parameterized by $\mathfrak{d}\colon \mathbb{Z}_{\geq 0} \to \mathbb{Z}_{\geq 0}$, and they investigated its covering rate.

Organization. As a preliminary, we describe the model of random walk on growing graph (RWoGG) in Section 2. Section 3 introduces the notion of weakly LHaGG, and presents some general theorems for sufficient conditions that a weakly LHaGG RWoGG is recurrent/transient. Section 4 shows a phase transition between the recurrence and transience of a lazy simple random walk on $\{0, \ldots, N\}^n$ with an increasing n.

2 Preliminaries

2.1 Model

A growing graph is a sequence of (static) graphs $\mathcal{G} = \mathcal{G}_0, \mathcal{G}_1, \mathcal{G}_2, \ldots$ where $\mathcal{G}_t = (\mathcal{V}_t, \mathcal{E}_t)$ for $t = 0, 1, 2, \ldots$ denotes a graph[1] with a finite vertex set $\mathcal{V}_t$ and an edge set $\mathcal{E}_t \subseteq \binom{\mathcal{V}_t}{2}$. For simplicity, this paper assumes[2] $\mathcal{V}_t \subseteq \mathcal{V}_{t+1}$ and $\mathcal{E}_t \subseteq \mathcal{E}_{t+1}$. In this paper, we assume $|\mathcal{V}_\infty| = \infty$, otherwise the subject is trivial; that is always recurrent. A random walk on a growing graph is a Markovian series $X_t \in \mathcal{V}_t$ $(t = 0, 1, 2, \ldots)$.

In particular, this paper is concerned with a specific model, described as follows, cf. [20]. A *random walk on a growing graph* (*RWoGG*), in this paper, is formally characterized by a 3-tuple of functions $\mathcal{D} = (\mathfrak{d}, G, P)$. The function $\mathfrak{d}\colon \mathbb{Z}_{>0} \to \mathbb{Z}_{\geq 0}$ denotes the duration. For convenience, let $T_n^{\mathfrak{d}} := \sum_{i=1}^n \mathfrak{d}(i)$ for $n = 1, 2, \ldots$ and $T_0^{\mathfrak{d}} = 0$. We call the time interval[3] $[T_{n-1}^{\mathfrak{d}}, T_n^{\mathfrak{d}})$ *phase n* for $n = 1, 2, \ldots$; thus $T_{n-1}^{\mathfrak{d}} = \sum_{i=1}^{n-1} \mathfrak{d}(i)$ is the beginning of the n-th phase, but we also say that $T_{n-1}^{\mathfrak{d}}$ is the end of the $(n-1)$-st phase, for convenience. The function $G\colon \mathbb{Z}_{>0} \to \mathfrak{G}$ represents the graph $G(n) = (V(n), E(n))$ for the phase n, where $\mathfrak{G}$ denotes the set of all (static) graphs, i.e., our growing graph $\mathcal{G}$ satisfies $\mathcal{G}_t = G(n)$ for $t \in [T_{n-1}^{\mathfrak{d}}, T_n^{\mathfrak{d}})$. Similarly, the function $P\colon \mathbb{Z}_{>0} \to \mathfrak{M}$ is a function representing the "transition probability" of a random walk on graph $G(n)$ where $\mathfrak{M}$ denotes the set of all transition matrices.

In summary, a RWoGG X_t $(t = 0, 1, 2, \ldots)$ characterized by $\mathcal{D} = (\mathfrak{d}, G, P)$ is temporally a time-homogeneous finite Markov chain according to $P(n)$ with the state space $V(n)$ during the time interval $[T_{n-1}^{\mathfrak{d}}, T_n^{\mathfrak{d}})$. Suppose $X_0 = o$ for $o \in V(1)$. We define the return probability at o by

$$R_{\mathfrak{d}}(t) = \Pr[X_t = o] \ (= \Pr[X_t = o \mid X_0 = o]) \tag{2}$$

at each time $t = 0, 1, 2, \ldots$. We say o is recurrent by RWoGG $\mathcal{D} = (\mathfrak{d}, G, P)$ if

$$\sum_{t=1}^{\infty} R_{\mathfrak{d}}(t) = \infty \tag{3}$$

holds, otherwise, i.e., $\sum_{t=1}^{\infty} R_{\mathfrak{d}}(t)$ is upper bounded, o is *transient* by $\mathcal{D}$.

---

[1] Every static graph is simple and undirected in this paper, for simplicity of the arguments.
[2] Thus, the current position does not disappear in the next step.
[3] Let $[T_{n-1}^{\mathfrak{d}}, T_n^{\mathfrak{d}}) = \{T_{n-1}^{\mathfrak{d}}, T_{n-1}^{\mathfrak{d}} + 1, \ldots, T_n^{\mathfrak{d}} - 1\}$, for convenience. Notice that $|[T_{n-1}^{\mathfrak{d}}, T_n^{\mathfrak{d}})| = \mathfrak{d}(n)$.

2.2 Mixing time

Next, we briefly introduce some terminology for random walks on static graphs, or time-homogeneous Markov chains, according to [25]. Suppose that $X_0, X_1, X_2, \ldots$ is a random walk on a static graph $G = (V, E)$ characterized by a time-homogeneous transition matrix $P = (P(u, v)) \in \mathbb{R}_{\geq 0}^{V \times V}$ where $P(u, v) = \Pr[X_{t+1} = v \mid X_t = u]$. A transition matrix P is *irreducible* if $\forall u, v \in V$, $\exists t > 0$, $P^t(u, v) > 0$. A transition matrix P is *aperiodic* if $\gcd\{t > 0 \,;\, P^t(v, v) > 0\} = 1$ for any $v \in V$. A Markov chain is *ergodic* if it is irreducible and aperiodic. A probability distribution π over V is a *stationary distribution* if it satisfies $\pi P = \pi$. It is well known that an ergodic P has a unique stationary distribution [25].

Let

$$d(t) := \max_{x \in V} \| P^t(x, \cdot) - \pi \|_{TV}. \tag{4}$$

Then, the *mixing time* of P is given by

$$t_{mix}(\epsilon) := \min \{t \,;\, d(t) \leq \epsilon\} \tag{5}$$

for $\epsilon \in (0, 1)$. We will use the following fact in the proof of Lemma 3 appearing later.

▶ **Lemma 1.** *Suppose P is ergodic. Let π_v denote the probability of $v \in V$ in the stationary distribution of P. If $t \geq t_{mix}(\frac{\pi_v}{4})$ then $d(t) \leq \frac{\pi_v}{2}$ holds.*

Proof. Let

$$\overline{d}(t') := \max_{x, y \in V} \left\| P^{t'}(x, \cdot) - P^{t'}(y, \cdot) \right\|_{TV}.$$

It is known that $d(t) \leq \overline{d}(t) \leq 2d(t)$ holds (cf. Lemma 4.10 in [25]). For convenience, let $t = t_{mix}(\frac{\pi_v}{4}) + s$ for $s \geq 0$. Then,

$$
\begin{aligned}
d(t) &\leq \overline{d}(t) && \text{(by Lem. 4.10 in [25])} \\
&\leq \overline{d}\left(t_{mix}(\tfrac{\pi_v}{4})\right) \overline{d}(s) && \text{(by the submultiplicativity of } \overline{d} \text{ (cf. Lem. 4.11 in [25]))} \\
&\leq \overline{d}\left(t_{mix}(\tfrac{\pi_v}{4})\right) && \text{(since } \overline{d}(s) \leq 1) \\
&\leq 2d\left(t_{mix}(\tfrac{\pi_v}{4})\right) && \text{(since } \overline{d}(t) \leq 2d(t) \text{ (cf. Lem. 4.10 in [25]))} \\
&\leq 2\frac{\pi_v}{4} && \text{(by (5))}
\end{aligned}
$$

and we obtain the claim. ◀

3 Recurrence and Transience

This section presents sufficient conditions that RWoGG $\mathcal{D} = (\mathfrak{d}, G, P)$ is recurrent/transient. Let $t_{mix}^{G(k)}(\epsilon)$ denote the mixing time of $P(k)$ and let $\pi_o^{G(k)}$ denote the probability of a vertex $o \in V$ in the stationary distribution of $P(k)$ in the following. In this paper, we are mainly concerned with RWoGG $\mathcal{D} = (\mathfrak{d}, G, P)$ satisfying

$$\sum_{k=1}^{\infty} \tau^*(k) p(k) < \infty \tag{6}$$

where $p(k) := \pi_o^{G(k)}$ and $\tau^*(k) := t_{mix}^{G(k)}\left(\frac{p(k)}{4}\right)$. Roughly speaking, the condition (6) means that the mixing times of $\mathcal{D}$ are not very large.

3.1 Recurrence

This section gives a sufficient condition that o is recurrent by $\mathcal{D}$.

▶ **Theorem 2.** *Suppose* $\mathcal{D} = (\mathfrak{d}, G, P)$ *satisfies* (6). *If* $\mathfrak{d}$ *satisfies*

$$\sum_{k=1}^{\infty} \mathfrak{d}(k)p(k) = \infty \tag{7}$$

then the initial vertex o *is recurrent by* $\mathcal{D}$ *where* $p(k) = \pi_o^{G(k)}$.

To prove Theorem 2, we prove the following lemma.

▶ **Lemma 3.** *Any RWoGG* $\mathcal{D} = (\mathfrak{d}, G, P)$ *satisfies*

$$\sum_{t=1}^{T_n^{\mathfrak{d}}} R_{\mathfrak{d}}(t) \geq \frac{1}{2} \sum_{k=1}^{n} \mathfrak{d}(k)p(k) - \frac{1}{2} \sum_{k=1}^{n} \tau^*(k)p(k) \tag{8}$$

for any $n \geq 1$*, where recall* $\tau^*(k) := t_{mix}^{G(k)}\left(\frac{p(k)}{4}\right)$ *and* $p(k) := \pi_o^{G(k)}$.

Proof. We prove the claim by an induction with respect to n. For $n = 1$, we prove

$$\sum_{t=1}^{\mathfrak{d}(1)} R_{\mathfrak{d}}(t) \geq \frac{1}{2}\left(\mathfrak{d}(1) - \tau^*(1)\right)p(1) \tag{9}$$

holds, where recall $T_1^{\mathfrak{d}} = \mathfrak{d}(1)$ by definition. We consider two cases whether $\mathfrak{d}(1) \leq \tau^*(1)$ or not. If $\mathfrak{d}(1) \leq \tau^*(1)$ then the right hand side of (8) ≤ 0. Clearly the left hand side of (8) ≥ 0, and we obtain (8). Suppose $\mathfrak{d}(1) > \tau^*(1)$. Notice that

$$\left| R_{\mathfrak{d}}(t) - \pi_o^{G(1)} \right| \leq 2\frac{p(1)}{4} = \frac{1}{2}p(1)$$

holds for $t \geq \tau^*(1)$ by Lemma 1. It implies

$$R_{\mathfrak{d}}(t) \geq \pi_o^{G(1)} - \frac{1}{2}p(1) = \frac{1}{2}p(1)$$

for $t \geq \tau^*(1)$, where recall $p(1) = \pi_o^{G(1)}$ by definition. Then,

$$\sum_{t=1}^{\mathfrak{d}(1)} R_{\mathfrak{d}}(t) \geq \sum_{t=\tau^*(1)}^{\mathfrak{d}(1)} R_{\mathfrak{d}}(t) \geq \sum_{t=\tau^*(1)}^{\mathfrak{d}(1)} \frac{1}{2}p(1) = \frac{1}{2}(\mathfrak{d}(1) - \tau^*(1))p(1)$$

holds. We obtain (9).

Inductively assuming (8) holds for n, we prove it for $n+1$. Noting that $T_{n+1}^{\mathfrak{d}} = T_n^{\mathfrak{d}} + \mathfrak{d}(n+1)$,

$$\sum_{t=1}^{T_{n+1}^{\mathfrak{d}}} R_{\mathfrak{d}}(t) = \sum_{t=1}^{T_n^{\mathfrak{d}}} R_{\mathfrak{d}}(t) + \sum_{t=1}^{\mathfrak{d}(n+1)} R_{\mathfrak{d}}(T_n^{\mathfrak{d}} + t)$$

$$\geq \frac{1}{2} \sum_{k=1}^{n} \mathfrak{d}(k)p(k) - \frac{1}{2} \sum_{k=1}^{n} \tau^*(k)p(k) + \sum_{t=1}^{\mathfrak{d}(n+1)} R_{\mathfrak{d}}(T_n^{\mathfrak{d}} + t) \tag{10}$$

holds since $\sum_{t=1}^{T_{n+1}^{\mathfrak{d}}} R_{\mathfrak{d}}(t) \geq \frac{1}{2}\sum_{k=1}^{n}\mathfrak{d}(k)p(k) - \frac{1}{2}\sum_{k=1}^{n}\tau^*(k)p(k)$ holds by the inductive assumption. Concerning the third term of (10), we can prove

$$\sum_{t=1}^{\mathfrak{d}(n+1)} R_{\mathfrak{d}}(T_n^{\mathfrak{d}} + t) \geq \frac{1}{2}(\mathfrak{d}(n+1) - \tau^*(n+1))p(n+1)$$

in a similar way as (9). Therefore,

$$(10) \geq \frac{1}{2}\sum_{k=1}^{n}\mathfrak{d}(k)p(k) - \frac{1}{2}\sum_{k=1}^{n}\tau^*(k)p(k) + \frac{1}{2}\left(\mathfrak{d}(n+1) - \tau^*(n+1)\right)p(n+1)$$

$$= \frac{1}{2}\sum_{k=1}^{n+1}\mathfrak{d}(k)p(k) - \frac{1}{2}\sum_{k=1}^{n+1}\tau^*(k)p(k)$$

holds. We obtain the claim.   ◄

Now, we prove Theorem 2.

Proof of Theorem 2. Recall the assumption (6),

$$\sum_{t=1}^{T_n^{\mathfrak{d}}} R_{\mathfrak{d}}(t) \geq \frac{1}{2}\sum_{k=1}^{n}\mathfrak{d}(k)p(k) - \frac{1}{2}\sum_{k=1}^{n}\tau^*(k)p(k) \qquad \text{(by Lemma 3)}$$

$$\geq \frac{1}{2}\sum_{k=1}^{n}\mathfrak{d}(k)p(k) - \frac{1}{2}\sum_{k=1}^{\infty}\tau^*(k)p(k) \qquad (\tau^*(k)p(k) \geq 0 \text{ for any } k)$$

$$= \frac{1}{2}\sum_{k=1}^{n}\mathfrak{d}(k)p(k) - C_1 \qquad \text{(by (6))} \qquad (11)$$

hold, where C_1 is a positive constant. Thus, the hypothesis $\sum_{k=1}^{\infty}\mathfrak{d}(k)p(k) = \infty$ implies $\sum_{t=1}^{\infty} R_{\mathfrak{d}}(t) = \infty$, which is what we want.   ◄

Even if (6) does not hold, Lemma 3 implies that the weaker condition

$$\lim_{n\to\infty}\sum_{k=1}^{n}(\mathfrak{d}(k) - \tau^*(k))p(k) = \infty$$

guarantees the recurrence.

3.2   Weakly less homesick as graph growing

Before giving a sufficient condition for transience, we introduce the notion of weakly less-homesickness as graph growing, which is a relationship between RWoGGs and plays an important role in our analysis. Let $\mathcal{D}_f = (f, G, P)$ and $\mathcal{D}_{f'} = (f', G', P')$ be RWoGGs, and let $R_f(t)$ and $R_{f'}(t)$ respectively denote their return probabilities at time $t = 1, 2, \ldots$. We say $\mathcal{D}_{f'} = (f', G', P')$ is *weakly less-homesick* than $\mathcal{D}_f = (f, G, P)$ at time t if

$$\sum_{k=1}^{t} R_f(k) \geq \sum_{k=1}^{t} R_{f'}(k) \qquad\qquad (12)$$

holds.

In particular, this paper is mainly concerned with the weakly less-homesick relation between $\mathcal{D}_f = (f, G, P)$ and $\mathcal{D}_g = (g, G, P)$ with the same P (and G). We say $P \colon \mathbb{Z}_{>0} \to \mathcal{M}$ is *weakly less-homesick as graph growing (weakly LHaGG)*[4] if $\mathcal{D}_g = (g, G, P)$ is weakly less-homesick than $\mathcal{D}_f = (f, G, P)$ whenever

$$\sum_{k=1}^{n} f(k) \geq \sum_{k=1}^{n} g(k) \tag{13}$$

holds for any $n \in \mathbb{Z}_{>0}$, where we remark that G and P are common in $\mathcal{D}_f$ and $\mathcal{D}_g$. The condition (13) implies the graph in $\mathcal{D}_g$ grows faster than $\mathcal{D}_f$, intuitively.

3.3 Transience

Next, we give a sufficient condition that o is transient by $\mathcal{D}$.

▶ **Theorem 4.** *Suppose RWoGG $\mathcal{D} = (\mathfrak{d}, G, P)$ is weakly LHaGG and satisfies* (6). *If $\mathfrak{d}$ satisfies*

$$\sum_{k=2}^{\infty} \mathfrak{d}(k) p(k-1) < \infty \tag{14}$$

then the initial vertex o is transient by $\mathcal{D}$ where $p(k) = \pi_o^{G(k)}$.

To prove Theorem 4, we prove the following lemma.

▶ **Lemma 5.** *Suppose RWoGG $\mathcal{D} = (\mathfrak{d}, G, P)$ is weakly LHaGG. Let*

$$g(k) := \max\left(\mathfrak{d}(k), \tau^*(k)\right) \tag{15}$$

for $k \geq 1$. Then, the sum of return probabilities of the RWoGG $\mathcal{D}_g = (g, G, P)$ satisfies

$$\sum_{t=1}^{T_{n+1}^g} R_g(t) \leq g(1) + \frac{3}{2} \sum_{k=2}^{n+1} g(k) p(k-1) \tag{16}$$

for $n \geq 1$.

Proof. We prove the claim for each fixed $n = 1, 2, \ldots$. Let

$$f(k) := \begin{cases} g(k) & (k \leq n-1) \\ \infty & (k = n). \end{cases} \tag{17}$$

Let Z_t $(t = 0, 1, 2, \ldots)$ denote a RWoGG $\mathcal{D}_f = (f, G, P)$ with $Z_0 = o$. Let $R_f(t)$ denote the return probability of Z_t, i.e., $R_f(t) = \Pr[Z_t = o] = \Pr[Z_t = o \mid Z_0 = o]$. Clearly, $T_n^f \geq T_n^g$ holds for any $n \geq 1$, hence the weakly LHaGG assumption implies

$$\sum_{t=1}^{T} R_g(t) \leq \sum_{t=1}^{T} R_f(t) \tag{18}$$

for any T.

---

[4] Strictly speaking, weakly LHaGG should be a property of the sequence of transition matrices $P(1), P(2), P(3), \ldots$. For convenience of the notation, we say $\mathcal{D} = (f, G, P)$ is weakly LHaGG, in this paper.

Suppose $Z_{T^g_{n-1}} = v$. Then, Z_t for $t \in (T^g_n, T^g_{n+1}]$ is nothing but a time-homogeneous random walk according to $P(n)$ with the "initial state" $Z_{T^g_{n-1}} = v$. For convenience, let $t = T^g_{n-1} + t'$, then

$$t' \geq g(n) \geq \tau^*(n) \tag{19}$$

by (15) and (17). This implies

$$\sum_{t=1}^{T^g_{n+1}} R_g(t) = \sum_{k=1}^{n+1} \sum_{t=T^g_{k-1}+1}^{T^g_k} R_f(t)$$

$$= \sum_{t=1}^{T^g_1} R_f(t) + \sum_{k=2}^{n+1} \sum_{t=T^g_{k-1}+1}^{T^g_k} R_f(t)$$

$$\leq g(1) + \sum_{k=1}^{n} \sum_{t=T^g_k+1}^{T^g_{k+1}} R_f(t)$$

$$= g(1) + \sum_{k=1}^{n} \sum_{t=T^g_k+1}^{T^g_{k+1}} \sum_{v \in V(k-1)} \Pr\left[Z_t = o \mid Z_{T^g_{k-1}} = v\right] \Pr\left[Z_{T^g_{k-1}} = v \mid Z_0 = o\right]$$

$$= g(1) + \sum_{k=1}^{n} \sum_{t'=1}^{g(k+1)} \sum_{v \in V(k-1)} \Pr\left[Z_{T^g_k+t'} = o \mid Z_{T^g_{k-1}} = v\right] \Pr\left[Z_{T^g_{k-1}} = v \mid Z_0 = o\right]$$

$$\leq g(1) + \sum_{k=1}^{n} \sum_{t'=1}^{g(k+1)} \max_{v \in V(k-1)} \Pr\left[Z_{T^g_k+t'} = o \mid Z_{T_{k-1}} = v\right]$$

$$\leq g(1) + \sum_{k=1}^{n} \sum_{t'=1}^{g(k+1)} \left(p(k) + \frac{p(k)}{2}\right) \qquad \text{(by (19))}$$

$$= g(1) + \sum_{k=1}^{n} \sum_{t'=1}^{g(k+1)} \frac{3}{2} p(k)$$

$$= g(1) + \frac{3}{2} \sum_{k=1}^{n} g(k+1) p(k)$$

$$= g(1) + \frac{3}{2} \sum_{k=2}^{n+1} g(k) p(k-1) \tag{20}$$

holds. The claim is clear by (18) and (20).   ◀

Now, we are ready to prove Theorem 4.

Proof of Theorem 4. Let

$$g(k) := \max\left(\mathfrak{d}(k), \tau^*(k)\right)$$

for $k \geq 1$. Notice that $g(k) \leq \mathfrak{d}(k) + \tau^*(k)$ holds. We calculate $\sum_{t=1}^{\infty} R_{\mathfrak{d}}(t)$ using Lemma 5:

$$\sum_{t=1}^{\infty} R_{\mathfrak{d}}(t) = \lim_{n\to\infty} \sum_{t=1}^{T^g_{n+1}} R_{\mathfrak{d}}(t) \leq \lim_{n\to\infty} \sum_{t=1}^{T^g_{n+1}} R_g(t) \qquad \text{(since weakly LHaGG)}$$

$$\leq \lim_{n\to\infty} \left\{ g(1) + \frac{3}{2} \sum_{k=2}^{n+1} g(k)p(k-1) \right\} \qquad \text{(by Lemma 5)}$$

$$= g(1) + \frac{3}{2} \sum_{k=2}^{\infty} g(k)p(k-1) \tag{21}$$

$$\leq \mathfrak{d}(1) + \tau^*(1) + \frac{3}{2} \sum_{k=2}^{\infty} \mathfrak{d}(k)p(k-1) + \frac{3}{2} \sum_{k=2}^{\infty} \tau^*(k)p(k-1)$$

$$\leq \frac{3}{2} \sum_{k=2}^{\infty} \mathfrak{d}(k)p(k-1) + C \qquad \text{(by (6))}$$

holds with some constant C. Now it is easy to see that (14) implies $\sum_{t=1}^{\infty} R_{\mathfrak{d}}(t) < \infty$, meaning that $\mathcal{D} = (\mathfrak{d}, G, P)$ is transient. ◀

4 Random Walk on Growing Dimension Boxes

This section is concerned with a random walk on growing dimension boxes $\mathcal{D} = (\mathfrak{d}, G, P)$. Let $G(n) = (V(n), E(n))$ be a graph given by

$$V(n) := \{0, \ldots, N\}^{n_0+n-1}$$
$$E(n) := \{(x, y) \, ; \, x, y \in V(n), \|x - y\|_1 = 1\}$$

where n and N are (fixed) positive integers. Let $o \in V(n)$ denote the origin vertex. Let $P^{G(n)}$ for $n \geq 1$ denote the transition probability of a lazy simple random walk on the static graph $G(n)$, which is given by

$$P_{x,y}^{G(n)} = \begin{cases} \frac{1}{2} & \text{(if } x = y) \\ \frac{1}{4(n_0+n-1)} & \text{(if } \|x - y\|_1 = 1, \ x_k \neq y_k \text{ and } x_k \notin \{0, N\}) \\ \frac{1}{2(n_0+n-1)} & \text{(if } \|x - y\|_1 = 1, \ x_k \neq y_k \text{ and } x_k \in \{0, N\}) \\ 0 & \text{(otherwise)} \end{cases}$$

for $x, y \in V(n)$. Then, we are concerned with a RWoGG X_t ($t = 0, 1, 2, \ldots$) according to $\mathcal{D} = (\mathfrak{d}, G, P)$. If the graph grows at time t, we assume $X_t = (x_1, \ldots, x_{n_0+n-1}) = (x_1, \ldots, x_{n_0+n-1}, 0)$.

▶ **Theorem 6.** *If* $\mathcal{D} = (\mathfrak{d}, G, P)$ *satisfies*

$$\sum_{k=1}^{\infty} \frac{\mathfrak{d}(k)}{(2N)^k} = \infty$$

then o is recurrent, otherwise o is transient.

We will prove Theorem 6 based on Theorems 2 and 4. As a preliminary step, we remark two facts. One is about the stationary distribution of $P^{G(n)}$, and it is not difficult to observe that

$$p(n) = \frac{1}{(2N)^{n_0+n-1}} \tag{22}$$

holds. The other is about the mixing time of $P^{G(n)}$, and we can prove

$$\tau^*(n) \leq 8N^2 \log_2(2N)(n_0 + n - 1)^3 \tag{23}$$

by a standard coupling technique. Therefore, random walk on growing boxes satisfies (6). Then, it is not difficult to see that Theorem 6 follows from the following Lemma 7.

▶ **Lemma 7.** *Random walk on growing dimension boxes is weakly LHaGG.*

Before the proof of Lemma 7, we prove Theorem 6.

Proof of Theorem 6. Notice that $\mathcal{D}$ satisfies (23) and it is weakly LHaGG by Lemma 7, meaning that $\mathcal{D}$ satisfies the hypotheses of Theorems 2 and Theorem 4. By (22),

$$p(n) = \frac{1}{(2N)^{n_0-1}} \frac{1}{(2N)^n}$$

where remark that $\frac{1}{(2N)^{n_0-1}}$ is a constant since N and n_0 are constants. If $\sum_{k=1}^{\infty} \frac{\partial(k)}{(2N)^k} = \infty$ then

$$\sum_{k=1}^{\infty} \partial(k)p(k) = \frac{1}{(2N)^{n_0-1}} \sum_{k=1}^{\infty} \frac{\partial(k)}{(2N)^k} = \infty$$

holds, which implies o is recurrent by Theorem 2. Similarly, if $\sum_{k=1}^{\infty} \frac{\partial(k)}{(2N)^k} \leq C$ holds for some constant then

$$\sum_{k=2}^{\infty} \partial(k)p(k-1) = \frac{1}{(2N)^{n_0-2}} \sum_{k=2}^{\infty} \frac{\partial(k)}{(2N)^k} \leq \frac{1}{(2N)^{n_0-2}} C$$

holds, which implies o is transient by Theorem 4.   ◀

4.1   Proof of Lemma 7

We prove Lemma 7 by an artificial coupling. Due to the page limitation, we here explain a proof sketch.

Let $\boldsymbol{X} = X_0, X_1, \ldots$ be a RWoGG according to $\mathcal{D}_f = (f, G, P)$, and let $R_f(t)$ ($t = 0, 1, 2, \ldots$) denote its return probability. Similarly, let $\boldsymbol{Y} = Y_0, Y_1, \ldots$ be a RWoGG according to $\mathcal{D}_g = (g, G, P)$, and let $R_g(t)$ ($t = 0, 1, 2, \ldots$) denote its return probability. Note that $X_0 = Y_0 = o$. Suppose that

$$\sum_{k=1}^{n} f(k) \geq \sum_{k=1}^{n} g(k) \tag{24}$$

holds for any $n \geq 1$. Then, we couple $\boldsymbol{X}$ and $\boldsymbol{Y}$ time *asynchronously*, so that $X_t \leq Y_t$ holds for any $t = 0, 1, 2, \ldots$, which is established in three steps by the following Lemmas 8–10.

▶ **Lemma 8.** *Suppose $\boldsymbol{X}$ and $\boldsymbol{Y}$ satisfy*

$$X_t = o, \ Y_{t'} = o,$$

for $t \leq t'$. Then, there is a coupling of $\boldsymbol{X}$ and $\boldsymbol{Y}$ such that

$$\min\{r \ ; \ r \geq 0, \ X_{t+r} \neq o\} = \min\{r \ ; \ r \geq 0, \ Y_{t'+r} \neq o\}, \tag{25}$$

i.e., X and Y stay at the origin vertex o for exactly the same r steps, where we define $\min \emptyset = \infty$ for convenience.

Proof. Notice that each of X and Y remains at the origin vertex o with probability $\frac{1}{2}$, and leaves the origin vertex o with probability $\frac{1}{2}$ independent of dimensions. Then, we can construct a coupling of X and Y. ◄

▶ **Lemma 9.** *Suppose that $|X_t| = |Y_{t'}| = 1$, where $t \leq t'$. Then, there is a coupling of $\boldsymbol{X}$ and $\boldsymbol{Y}$ such that*

$$\min\left\{r \;;\; r > 0,\; X_{t+r} = o\right\} \leq \min\left\{r \;;\; r > 0,\; Y_{t'+r} = o\right\}, \tag{26}$$

i.e., X returns to the origin vertex o in a fewer steps than Y.

Sketch of proof. Without loss of generality, we may assume[5] that $X_t = (X_t^1, X_t^2, \ldots, X_t^{n_t}) = (1, 0, \ldots, 0)$ and $Y_{t'} = (Y_{t'}^1, Y_{t'}^2, \ldots, Y_{t'}^{m_{t'}}) = (1, 0, \ldots, 0)$, where we remark n_t and m_t respectively denote the dimensions of X_t and Y_t.

Let $I(t) \in \{1, \ldots, n_t\}$ denote the index selected in the transition from X_{t-1} to X_t, and let $J(t) \in \{1, \ldots, m_t\}$ denote the index selected in the transition from Y_{t-1} to Y_t. For example, when $X_{t-1} = (0, 0, 0)$, $I(t) = 1$ and $X_t^1 = X_{t-1}^1 + 1$ then $X_t = (1, 0, 0)$. Then, we couple $\{I(t+r)\}_{r=1,2,\ldots}$ and $\{J(t+r)\}_{r=1,2,\ldots}$. For $\theta_{t+r} \in \{1, 2, \ldots, n_{t+r}\}$, let

$$\Psi_k(\theta_{t+r}) := \left\{\{\omega_{t+r}\}_{r\in\mathbb{N}} \,\middle|\, \omega_{k'} > n_{t+r} \text{ for } k' < k \text{ and } \omega_k = \theta_{t+r}\right\}$$

and

$$\widetilde{\Psi}(t+r) := \left\{\{\omega_{t+r}\}_{r\in\mathbb{N}} \,\middle|\, \omega_{k'} > n_{t+r} \text{ for } k' \geq 1\right\}$$

for $r \geq 1$. Let $\boldsymbol{W} = \{W_s\}_{s\in\mathbb{N}}$ satisfy $W_s := J(t'+s)$ for $s \geq 1$. Suppose $I(t+r) = \theta_{t+r}$. Let $s_1 = k$ such that $\boldsymbol{W} \in \Psi_k(\theta_{t+1})$, and let $S(1) := s_1$. Recursively, let $s_r = k'$ such that $\boldsymbol{W} \in \Psi_{k'}(\theta_{t+r})$ for $r \geq 2$, and let $S(r) := S(r-1) + s_r$. Let $\Psi(\theta_{t+r}) := \bigcup_{k=1}^{\infty} \Psi_k(\theta_{t+r})$.

Firstly, we claim that

$$\Pr[I(t+r) = \theta_{t+r}] = \Pr\left[\boldsymbol{W} \in \Psi(\theta_{t+r})\right] + \frac{1}{n_{t+r}} \Pr\left[\boldsymbol{W} \in \widetilde{\Psi}(t+r)\right]. \tag{27}$$

Clearly,

$$\Pr[I(t+r) = \theta_{t+r}] = \frac{1}{n_{t+r}}$$

holds for the left-hand-side of (27). Notice that

$$t + r \leq t' + S(r) \tag{28}$$

holds. Then, we can prove

$$\Pr\left[\boldsymbol{W} \in \Psi(\theta_{t+r})\right] + \frac{1}{n_{t+r}} \Pr\left[\boldsymbol{W} \in \widetilde{\Psi}(t+r)\right] = \frac{1}{n_{t+r}} \tag{29}$$

holds, which implies (27).

Next, we prove for any r and $i \leq n_{t+r}$ that

$$X_{t+r}^i \leq Y_{t'+s}^i \tag{30}$$

for $S(r) \leq s < S(r+1)$. We consider two cases whether $i \leq n_{t+1}$ or not.

---

[5] Suppose that $|Y_t| = |Y_t'| = 1$. Let $Y_t^i = Y_t'^j = 1$. There is the coupling of $\boldsymbol{Y}$ and $\boldsymbol{Y}'$ such that $Y_t = o$ if and only if $Y_t' = o$.

(i) Consider the case $i \leq n_{t+1}$. Recall that $X_t^i = Y_{t'}^i$ for $i \leq n_t$. We inductively prove that

$$X_{t+r}^i = Y_{t'+s}^i \tag{31}$$

for $S(r) \leq s < S(r+1)$ with respect to r. Notice that

$$Y_{t'+S(r)}^i = Y_{t'+s}^i \tag{32}$$

for $S(r) \leq s < S(r+1)$ since $I(t'+s) > n_{t+r+1}$ for $S(r) < s < S(r+1)$. Suppose i is chosen l times for $X_t, \ldots, X_{t+r}$ in the r steps, i.e.,

$$|\{r' \mid I(t+r') = i, \ 1 \leq r' \leq r\}| = l.$$

Notice that i is chosen l times for $Y_{t'} \ldots, Y_{t'+S(r')}$ in the $S(r')$ steps. Then,

$$\Pr\left[X_{t+r'}^i - X_{t+r'-1}^i = z \mid I(t+r') = i\right]$$
$$= \Pr\left[Y_{t'+S(r')}^i - Y_{t'+S(r')-1}^i = z \mid J(t'+S(r')) = i\right]$$

holds for $z \in \{-1, 0, 1\}$. The inductive assumption $X_{t+r'-1}^i = Y_{t'+S(r')-1}^i$ implies

$$X_{t+r'}^i = Y_{t'+S(r')}^i. \tag{33}$$

We obtain (31).

(ii) Consider the case $i > n_{t+1}$. Suppose i is chosen l times for $X_t, \ldots, X_{t+r}$ in the r steps, i.e.,

$$|\{r' \mid I(t+r') = i, \ 1 \leq r' \leq r\}| = l,$$

and let r'' denote the minimum satisfying $I(t+r'') = i$ for $1 \leq r'' \leq r$. Clearly, $X_{t+r''-1}^i \leq Y_{t'+S(r'')-1}^i$. If $X_{t+r''-1}^i = Y_{t'+S(r'')-1}^i$, we can prove (30) for any s satisfying $S(r) \leq s < S(r+1)$ in a similar way as the case (i).

Thus, we consider the case $X_{t+r''-1}^i < Y_{t+S(r'')-1}^i$. Then, we can couple the transitions $X_{t+r'-1}^i \mapsto X_{t+r'}^i$ and $Y_{t+S(r')-1}^i \mapsto Y_{t+S(r')}^i$ such that

$$\Pr\left[X_{t+r'}^i - X_{t+r'-1}^i = 0 \mid I(t+r') = i\right]$$
$$= \Pr\left[\left|Y_{t'+S(r')}^i - Y_{t'+S(r')-1}^i\right| = 1 \mid J(t'+S(r')) = i\right] = \frac{1}{2}$$
$$\Pr\left[\left|X_{t+r'}^i - X_{t+r'-1}^i\right| = 1 \mid I(t+r') = i\right]$$
$$= \Pr\left[Y_{t'+S(r')}^i - Y_{t'+S(r')-1}^i = 0 \mid J(t+S(r')) = i\right] = \frac{1}{2}$$

hold. Recall that $X_{t+r''-1}^i < Y_{t+S(r'')-1}^i$ implies $X_{t+r''-1}^i + 1 \leq Y_{t+S(r'')-1}^i$. Thus, the coupling implies $X_{t+r''}^i \leq Y_{t'+S(r'')}^i$. For other $r''' \in \{r' \ ; \ I(t+r') = i, \ 1 \leq r' \leq r\}$, we can inductively prove (30) in a similar way.

Therefore, if Y returns to the origin vertex o at time $t' + S(r)$ then X returns to the origin vertex o before the time $t+r$ by (30). Clearly $t+r \leq t' + S(r)$ by (28). We obtain the claim. ◀

▶ **Lemma 10.** *Let*

$$\tau_o := \min\left\{r \ ; \ r > 0, \ X_{t+r} = o\right\},$$
$$\tau_o' := \min\left\{r \ ; \ r \geq 0, \ Y_{t'+S(\tau_o)+r} = o\right\}.$$

If $Y_{t'+S(\tau_o)} \neq o$ then there is a coupling of $\mathbf{X}$ and $\mathbf{Y}$ such that $X_{t+\tau_o} = Y_{t'+S(\tau_o)+\tau_o'} = o$, i.e., X stops its time until Y returns to the origin vertex o.

Proof. Since

$$\Pr\left[\tau_o' < \infty \text{ or } \tau_o' = \infty \mid Y_{t'+S(\tau_o)} \neq o\right] = 1$$

holds for any $Y_{t'+S(\tau_o)} \in \{0, 1, \dots, N\}^{m_{t'+S(\tau_o)}}$. ◄

We prove Lemma 7 using Lemmas 8–10.

Proof of Lemma 7. Let

$$\tau_o^x(n) := \min\{t \; ; \; t > \tau_o^x(n-1) \; , \; X_t = o\},$$
$$\tau_o^y(n) := \min\{t \; ; \; t > \tau_o^y(n-1) \; , \; Y_t = o\}$$

for $n \geq 1$. For convenience, let $\tau_o^x(0) := 0$ and $\tau_o^y(0) := 0$. To begin with, we prove that there is a coupling of X and Y such that

$$\tau_o^x(n) \leq \tau_o^y(n) \tag{34}$$

for any $n \geq 0$. For $n = 0$, (34) is obvious. Inductively assuming that (34) holds for n, we prove it for $n+1$. If $\tau_o^y(n+1) = \infty$ then we have

$$\tau_o^x(n+1) \leq \tau_o^y(n+1)$$

clearly, and we obtain (34) in this case. Then, we consider the case of $\tau_o^y(n+1) < \infty$. By Lemma 10, X can stop at the vertex o at time $\tau_o^x(n)$ by the time $\tau_o^y(n)$. Therefore, we can consider the coupling of $X_{\tau_o^x(n)}$ and $Y_{\tau_o^y(n)}$. By Lemma 8, there exists $t_n \geq 1$ such that

$$X_{\tau_o^x(n)+s} = o \quad \text{and} \quad Y_{\tau_o^y(n)+s} = o \tag{35}$$

for any s satisfying $0 \leq s < t_n$, and

$$X_{\tau_o^x(k)+t_n} \neq o \quad \text{and} \quad Y_{\tau_o^y(n)+t_n} \neq o \tag{36}$$

hold. If $t_n > 1$ then (35) implies

$$\tau_o^x(n+1) = \tau_o^x(n) + 1 \quad \text{and} \quad \tau_o^y(n+1) = \tau_o^y(n) + 1.$$

This means that we have

$$\tau_o^x(n+1) \leq \tau_o^y(n+1)$$

by the inductive assumption (34), and we obtain the equation (34) in the case $t_n > 1$. Then, we consider the case $t_n = 1$. Notice that (35) and (36) imply

$$|X_{\tau_o^x(k)+t_n}| = 1, \; |Y_{\tau_o^y(k)+t_n}| = 1,$$

and hence Lemma 9 implies that there is a coupling of X and Y such that

$$\tau_o^x(n+1) \leq \tau_o^y(n+1) \tag{37}$$

holds. Therefore, we obtain (34) in the case $t_n = 1$. Thus, we obtain (34) for any $n+1$. It is not difficult to see from (34) that the random walk on growing dimension boxes is weakly LHaGG. ◄

References

1   Gideon Amir, Itai Benjamini, Ori Gurel-Gurevich, and Gady Kozma. Random walk in changing environment. *Stochastic Processes and their Applications*, 130(12):7463–7482, 2020. `doi:10.1016/j.spa.2020.08.003`.

2   John Augustine, Gopal Pandurangan, and Peter Robinson. Distributed algorithmic foundations of dynamic networks. *SIGACT News*, 47(1):69–98, March 2016. `doi:10.1145/2902945.2902959`.

3   Chen Avin, Michal Koucký, and Zvi Lotker. How to explore a fast-changing world (cover time of a simple random walk on evolving graphs). In Luca Aceto, Ivan Damgård, Leslie Ann Goldberg, Magnús M. Halldórsson, Anna Ingólfsdóttir, and Igor Walukiewicz, editors, *Automata, Languages and Programming*, ICALP 2008, pages 121–132, 2008.

4   Chen Avin, Michal Koucký, and Zvi Lotker. Cover time and mixing time of random walks on dynamic graphs. *Random Structures & Algorithms*, 52(4):576–596, 2018. `doi:10.1002/rsa.20752`.

5   Itai Benjamini and David Wilson. Excited Random Walk. *Electronic Communications in Probability*, 8:86–92, 2003. `doi:10.1214/ECP.v8-1072`.

6   Leran Cai, Thomas Sauerwald, and Luca Zanetti. Random walks on randomly evolving graphs. In Andrea Werneck Richa and Christian Scheideler, editors, *Structural Information and Communication Complexity*, SIROCCO 2020, pages 111–128, 2020.

7   Colin Cooper. Random walks, interacting particles, dynamic networks: Randomness can be helpful. In Adrian Kosowski and Masafumi Yamashita, editors, *Structural Information and Communication Complexity*, SIROCCO 2011, pages 1–14, 2011.

8   Colin Cooper and Alan Frieze. Crawling on simple models of web graphs. *Internet Mathematics*, 1(1):57–90, 2003.

9   Burgess Davis. Reinforced random walk. *Probability Theory and Related Fields*, 84(2):203–229, 1990. `doi:10.1007/BF01197845`.

10   Amir Dembo, Ruojun Huang, Ben Morris, and Yuval Peres. Transience in growing subgraphs via evolving sets. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 53(3):1164–1180, 2017. `doi:10.1214/16-AIHP751`.

11   Amir Dembo, Ruojun Huang, and Vladas Sidoravicius. Monotone interaction of walk and graph: recurrence versus transience. *Electronic Communications in Probability*, 19:1–12, 2014. `doi:10.1214/ECP.v19-3607`.

12   Amir Dembo, Ruojun Huang, and Vladas Sidoravicius. Walking within growing domains: recurrence versus transience. *Electronic Journal of Probability*, 19:1–20, 2014. `doi:10.1214/EJP.v19-3272`.

13   Amir Dembo, Ruojun Huang, and Tianyi Zheng. Random walks among time increasing conductances: heat kernel estimates. *Probability Theory and Related Fields*, 175(1):397–445, October 2019. `doi:10.1007/s00440-018-0894-1`.

14   Oksana Denysyuk and Luís Rodrigues. Random walks on evolving graphs with recurring topologies. In Fabian Kuhn, editor, *Distributed Computing*, DISC 2014, pages 333–345, 2014.

15   Dmitry Dolgopyat, Gerhard Keller, and Carlangelo Liverani. Random walk in Markovian environment. *The Annals of Probability*, 36(5):1676–1710, 2008. `doi:10.1214/07-AOP369`.

16   Richard Durrett. *Probability: theory and examples*. Cambridge University Press, Fifth edition, 2019.

17   Daniel Figueiredo, Giulio Iacobelli, Roberto Oliveira, Bruce Reed, and Rodrigo Ribeiro. On a random walk that grows its own tree. *Electronic Journal of Probability*, 26:1–40, 2021. `doi:10.1214/20-EJP574`.

18   Ruojun Huang. On random walk on growing graphs. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 55(2):1149–1162, 2019. `doi:10.1214/18-AIHP913`.

19   Giulio Iacobelli, Daniel R. Figueiredo, and Giovanni Neglia. Transient and slim versus recurrent and fat: Random walks and the trees they grow. *Journal of Applied Probability*, 56(3):769–786, 2019. `doi:10.1017/jpr.2019.43`.

20   Shuji Kijima, Nobutaka Shimizu, and Takeharu Shiraga. How many vertices does a random walk miss in a network with moderately increasing the number of vertices? In *Proceedings of the Thirty-Second Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA 2021, pages 106–122, 2021. `doi:10.1137/1.9781611976465.8`.

21   Elena Kosygina and Martin PW Zerner. Excited random walks: results, methods, open problems. *Bulletin of the Institute of Mathematics Academia Sinica (New Series)*, 8(1), 2013.

22   Fabian Kuhn and Rotem Oshman. Dynamic networks: models and algorithms. *SIGACT News*, 42(1):82–96, March 2011. `doi:10.1145/1959045.1959064`.

23   Shuma Kumamoto, Shuji Kijima, and Tomoyuki Shirai. An analysis of the recurrence/transience of random walks on growing trees and hypercubes, 2024. `arXiv:2405.09102`.

24   Ioannis Lamprou, Russell Martin, and Paul Spirakis. Cover time in edge-uniform stochastically-evolving graphs. *Algorithms*, 11(10), 2018. `doi:10.3390/a11100149`.

25   David A. Levin and Yuval Peres. *Markov chains and mixing times*. American Mathematical Society, Second edition, 2017.

26   Russell Lyons. Random walks and percolation on trees. *The Annals of Probability*, 18(3):931–958, 1990.

27   Russell Lyons and Yuval Peres. *Probability on Trees and Networks*. Cambridge University Press, USA, 2017.

28   Othon Michail and Paul G. Spirakis. Elements of the theory of dynamic networks. *Communications of the ACM*, 61(2):72, January 2018. `doi:10.1145/3156693`.

29   Laurent Saloff-Coste and Jessica Zúñiga. Merging for time inhomogeneous finite Markov chains, Part I: Singular values and stability. *Electronic Journal of Probability*, 14:1456–1494, 2009. `doi:10.1214/EJP.v14-656`.

30   Laurent Saloff-Coste and Jessica Zúñiga. Merging for time inhomogeneous finite Markov chains, Part II: Nash and log-Sobolev inequalities. *The Annals of Probability*, 39(3):1161–1203, 2011.

The Alternating Normal Form of Braids and Its Minimal Automaton

Vincent Jugé ✉ 🏠 🆔
LIGM, CNRS & Univ Gustave Eiffel, Marne-la-Vallée, France
IRIF, CNRS & Université Paris Cité, France

June Roupin
LIGM, CNRS & Univ Gustave Eiffel, Marne-la-Vallée, France

─── **Abstract** ───

The alternating normal form of braids is a well-known normal form on standard braid monoids. This normal form is regular: the language it identifies with is regular. We give a characterisation of the minimal automaton of this language and compute its size, both in terms of number of states and of transitions, depending on the number of generators of the monoid.

1 Introduction

The group of braids with n strands, commonly denoted by B_n, is the group of isotopy classes of geometric braids with n strands. In [2], E. Artin proved that this group enjoyed the following finite presentation:

$$B_n = \left\langle \sigma_1, \ldots, \sigma_{n-1} \;\middle|\; \begin{array}{l} \sigma_i \sigma_j = \sigma_j \sigma_i \text{ when } j \geqslant i+2 \\ \sigma_i \sigma_{i+1} \sigma_i = \sigma_{i+1} \sigma_i \sigma_{i+1} \end{array} \right\rangle.$$

The relations $\sigma_i \sigma_j = \sigma_j \sigma_i$ are called *commutation* relations; relations $\sigma_i \sigma_{i+1} \sigma_i = \sigma_{i+1} \sigma_i \sigma_{i+1}$ are called *braid* relations. They come with the *simplification* relations $\sigma_i \sigma_i^{-1} = \varepsilon$, where ε denotes the neutral element of the group. These three kinds of relations are illustrated in Figure 1. Braid elements $\sigma_1, \ldots, \sigma_{n-1}$ are called *Artin generators*.
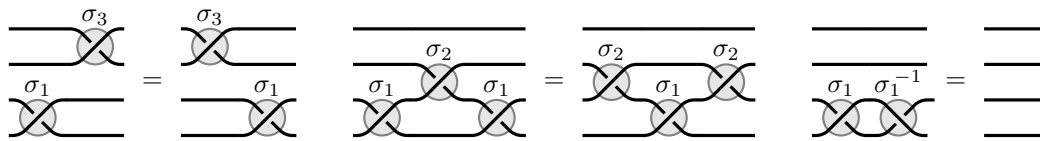
Figure 1 Commutation relation $\sigma_1 \sigma_3 = \sigma_3 \sigma_1$, braid relation $\sigma_1 \sigma_2 \sigma_1 = \sigma_2 \sigma_1 \sigma_2$ and simplification relation $\sigma_1 \sigma_1^{-1} = \varepsilon$. The latter relation is valid only in the group B_n.

Braid groups enjoy numerous algebraic, combinatorial and geometric properties, many of which are connected with the study of the (standard) braid monoid B_n^+: this is the monoid positively generated by the generators σ_i, i.e., the least subset of B_n containing generators $\sigma_1, \ldots, \sigma_{n-1}$ (but not their inverses) and stable by product:

$$B_n^+ = \left\langle \sigma_1, \ldots, \sigma_{n-1} \;\middle|\; \begin{array}{l} \sigma_i \sigma_j = \sigma_j \sigma_i \text{ when } j \geqslant i+2 \\ \sigma_i \sigma_{i+1} \sigma_i = \sigma_{i+1} \sigma_i \sigma_{i+1} \end{array} \right\rangle^+.$$

Here are some properties of the monoid B_n^+ [1, 4, 6, 8, 12, 14]:

a) The braid monoid B_n^+ is simplifiable: whenever $\alpha\beta\gamma = \alpha\beta'\gamma$, we have $\beta = \beta'$.

b) The left-divisibility ordering, defined by $\alpha \leqslant_{\mathsf{L}} \beta$ whenever there exists a braid $\gamma \in B_n^+$ (also denoted by $\alpha^{-1}\beta$) such that $\alpha\gamma = \beta$, is a lattice: any two elements α and β have a greatest common divisor $\alpha \wedge \beta$ and a least common multiple $\alpha \vee \beta$.

c) Similarly, the right-divisibility ordering, defined by $\beta \geqslant_{\mathsf{R}} \alpha$ whenever there exists a braid $\gamma \in B_n^+$ (also denoted by $\beta\alpha^{-1}$) such that $\beta = \gamma\alpha$, is a lattice.

**d)** The braid $\Delta_n = \sigma_1(\sigma_2\sigma_1)(\sigma_3\sigma_2\sigma_1)\cdots(\sigma_{n-1}\sigma_{n-2}\cdots\sigma_2\sigma_1)$, called the *Garside element* of the monoid B_n^+, is the least common multiple of the family of Artin generators for both the left- and right-divisibility orderings. Both its left and its right divisors coincide with the set of positive braids in which any two strands cross each other at most once; such braids are called *simple* braids. Furthermore, Δ_n obeys the relations $\Delta_n\sigma_i = \sigma_{n-i}\Delta_n$, which makes the inner automorphism $\phi_n\colon \beta \mapsto \Delta_n^{-1}\beta\Delta_n$ an involution of B_n^+.

e) The function $\mathsf{s}_n\colon \beta \mapsto \beta \wedge \Delta_n$, which selects the largest simple left divisor of a braid β, obeys the identity $\mathsf{s}_n(\alpha\beta) = \mathsf{s}_n(\alpha\mathsf{s}_n(\beta))$.

Properties **a)** to **c)** give rise to recursive decompositions, some of which we will focus on below: Property **a)** allows factoring a braid into factors on which we will be able to work independently, and Properties **b)** and **c)** will allow, under some conditions, to select the *largest* divisor of a braid that belongs to a given set. For instance, the following result is a consequence of Properties **a)** and **b)**:

f) The submonoid B_{n-1}^+, generated by $\sigma_1, \ldots, \sigma_{n-2}$ and called a *parabolic* submonoid of B_n^+, is a sub-lattice of B_n^+. Thus, each braid $\beta \in B_n^+$ has a largest left divisor in B_{n-1}^+, denoted by $\mathsf{h}_n(\beta)$ and called the *n-head* of β: its left divisors are the left divisors of β that belong to B_{n-1}^+. The corresponding right divisor $\mathsf{h}_n(\beta)^{-1}\beta$, denoted by $\mathsf{b}_n(\beta)$, is called the *n-body* of β.

The effect of the functions s_4, h_4 and b_4 and of the automorphism ϕ_4 on the braid $\beta = \sigma_2\sigma_3\sigma_3\sigma_1$ is illustrated in Figure 2: s_4 selects the largest simple left divisor of β, which is not necessarily a prefix of the representation of β we started from; h_4 selects the largest left divisor of β that can be written without using the generator σ_3, and b_4 selects the corresponding right divisor; finally, ϕ_4 replaces each generator σ_i of β by σ_{4-i}.
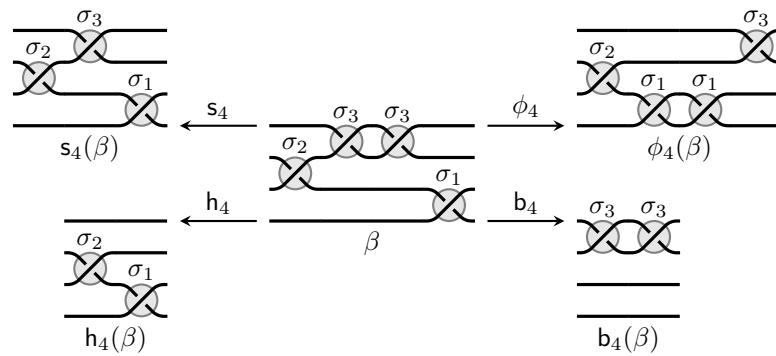
Figure 2 Applying s_4, h_4, b_4 and ϕ_4 to the braid $\beta = \sigma_2\sigma_3\sigma_3\sigma_1$.

The above presentation identifies each braid $\beta \in B_n^+$ with an equivalence class of words over the alphabet $\mathbb{A}_n = \{\sigma_1, \ldots, \sigma_{n-1}\}$. A *normal form* is then a language containing exactly one word $\mathsf{NF}(\beta)$ in each equivalence class β, which will be a preferred representative of β.

For a normal form to be useful, the following tasks should be as easy as possible [8]:

- deciding whether a word w belongs to the normal form;
- transforming a word w representing a braid β into the word $\mathsf{NF}(\beta)$;
- computing, given two words $w = \mathsf{NF}(\beta)$ and $w' = \mathsf{NF}(\beta')$, the word $\mathsf{NF}(\beta\beta')$.

In this article, we focus on the first question, for which a possible answer is: "the normal form should be a regular set, and its minimal automaton should be small".

In that context, among a plethora of other normal forms, let us mention three similar normal forms on braids: the Garside normal form [7], the lexicographically minimal normal form [13] and the alternating normal form [3, 5]: the Garside normal form is the most well-known normal form on braid monoids, and all three are regular.

▶ **Definition 1.** *The Garside normal form of a braid $\beta \in B_n^+$ is inductively defined as the following factorisation of β into simple braids: we set $\mathsf{Gar}_n(\beta) = \beta$ when β is simple, and $\mathsf{Gar}_n(\beta) = \mathsf{s}_n(\beta)\mathsf{Gar}_n(\mathsf{s}_n(\beta)^{-1}\beta)$ otherwise. If necessary, each simple divisor can then be written as a product of generators σ_i.*

▶ **Definition 2.** *The lexicographically minimal normal form of a braid $\beta \in B_n^+$ is denoted by $\mathsf{LexMin}_n(\beta)$. It is the word representing the braid $\beta \in B_n^+$ that is minimal for the lexicographic ordering induced by the ordering $\sigma_1 < \sigma_2 < \cdots < \sigma_{n-1}$ on Artin generators.*

Alternatively, the word $\mathsf{LexMin}_n(\beta)$ may be inductively defined by $\mathsf{LexMin}_2(\sigma_1^k) = \sigma_1^k$ or, if $n \geqslant 3$, by $\mathsf{LexMin}_n(\beta) = \mathsf{LexMin}_{n-1}(\beta)$ when $\beta \in B_{n-1}^+$, and

$$\mathsf{LexMin}_n(\beta) = \mathsf{LexMin}_{n-1}(\mathsf{h}_n(\beta))\sigma_{n-1}\mathsf{LexMin}_n((\mathsf{h}_n(\beta)\sigma_{n-1})^{-1}\beta)$$

otherwise.

▶ **Definition 3.** *The alternating normal form of a braid $\beta \in B_n^+$ is denoted by $\mathsf{Alt}_n(\beta)$. It is inductively defined by $\mathsf{Alt}_2(\sigma_1^k) = \sigma_1^k$ or, if $n \geqslant 3$, by $\mathsf{Alt}_n(\beta) = \mathsf{Alt}_{n-1}(\beta)$ when $\beta \in B_{n-1}^+$, and $\mathsf{Alt}_n(\beta) = \mathsf{Alt}_{n-1}(\mathsf{h}_n(\beta))\phi_n(\mathsf{Alt}_n(\phi_n(\mathsf{b}_n(\beta))))$ otherwise*[1].

This normal form is tightly connected to the rotating normal form [11], a similar normal form defined on the *dual* braid monoid B_n^{+*}, which is the sub-monoid of B_n positively generated by braids of the form $\sigma_u\sigma_{u-1}\cdots\sigma_{v+1}\sigma_v\sigma_{v+1}^{-1}\sigma_{v+2}^{-1}\cdots\sigma_u^{-1}$. Below, we study the alternating normal form, by constructing explicitly its minimal automaton and counting its states and transitions.

Figure 3 presents the three representatives of the braid $\beta = \sigma_3\sigma_2\sigma_2\sigma_3\sigma_1\sigma_1\sigma_3$ that belong to the Garside, lexicographically minimal and alternating normal forms, illustrating that these three words may all differ from each other.

Property **e)** provides us with a co-deterministic automaton that, once given as input a word $w = w_0 w_1 \cdots w_{k-1}$ representing a braid β, computes at each step the braid $\mathsf{s}_n(w_{\geqslant i})$, where $w_{\geqslant i} = w_i w_{i+1} \cdots w_{k-1}$: indeed, it suffices to observe that $\mathsf{s}_n(w_{\geqslant i}) = \mathsf{s}_n(w_i \mathsf{s}_n(w_{\geqslant i+1}))$, and to precompute s_n on braids of the form $\sigma_i\gamma$, where σ_i is an Artin generator and γ is simple. This automaton itself helps proving that the three above normal forms are regular: we check that w is in

1) Garside normal form by verifying that it starts with a prefix $w_{<i} = w_0 w_1 \cdots w_{i-1}$ representing $\mathsf{s}_n(w)$, and then that $w_{\geqslant i}$ is in Garside normal form;

---

[1] One may often find a *mirrored* version of this normal form, in which, instead of extracting the largest left divisor of β in B_{n-1}^+, one extracts the largest right divisor. The languages induced by both versions are mirrors of each other.
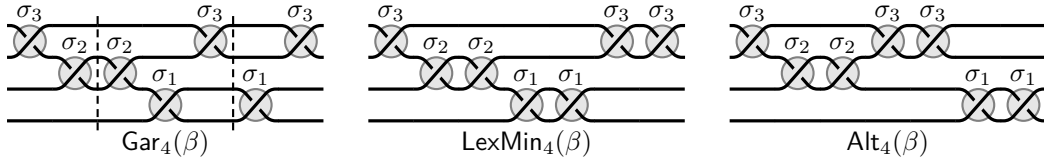
Figure 3 Normal forms of the braid $\beta = \sigma_3\sigma_2\sigma_2\sigma_3\sigma_1\sigma_1\sigma_3$. Vertical dashed bars separate the simple braids $\sigma_3\sigma_2$, $\sigma_2\sigma_3\sigma_1$ and $\sigma_1\sigma_3$ into which β was factored to give its Garside normal form.

2) lexicographically minimal normal form by verifying that each letter w_i is the least Artin generator that left-divides $\mathsf{s}_n(w_{\geqslant i})$;

3) alternating normal form by finding the smallest index i such that $w_i = \sigma_{n-1}$ (if any), verifying that σ_{n-1} is the only Artin generator left-dividing $\mathsf{s}_n(w_{\geqslant i})$, and verifying that $w_{<i}$ and $\phi_n(w_{\geqslant i})$ are in alternating normal form.

Similar arguments would prove that the rotating normal form is also regular.

Although such observations yield automata recognising the three above normal forms, these automata are non-deterministic, and determinising them might result in unreasonably large deterministic automata. Explicit minimal automata for the lexicographically minimal normal form were constructed in [9]; there, it is proved that the minimal automaton recognising $\mathsf{LexMin}_n(B_n^+)$ has $2F_{2n+1} - n(n+1)/2 - 2 \approx 2\Phi^{2n+1}/\sqrt{5}$ states, where F_k is the kth Fibonacci number and $\Phi = (1 + \sqrt{5})/2$ is the Golden Ratio. Non-necessarily minimal automata for the rotating normal form were constructed in [11]. The minimal automaton of the Garside normal form, viewed as a language over the alphabet $\mathsf{s}_n(B_n^+)$ of simple braids, has 2^{n-1} states and ob_n transitions, where ob_n is the nth *ordered Bell number*. Finally, the minimal automata for the Garside normal form (viewed as a language over $\mathbb{A}_n$, choosing a canonical representative of each simple braid), for the alternating normal form and for the rotating normal form have net yet been investigated.

In this article, we prove the following results.

▶ **Theorems 19 & 20.** *The minimal automaton of the language* $\mathsf{Alt}_n(B_n^+)$ *is an explicit automaton* $\mathscr{A}_n$ *with* s_n *states and* t_n *transitions, where* $s_1 = 1$, $t_1 = 0$, *and*

$$s_n = \frac{25 \times 2^{2n-3} - 9n^2 + 3n + 7}{27} \ \text{ and } \ t_n = \frac{(225n - 290)2^{2n-5} - 9n^3 - 9n^2 + 93n - 77}{81}$$

whenever $n \geqslant 2$.

In particular, $\mathscr{A}_n$ has asymptotically $25 \times 2^{2n-3}/27$ states and an average of $3n/4 - 29/30$ transitions per state; this largely exceeds the growth rate of the monoid B_n^+, which is only $3.233636\ldots$ [10]. Like the minimal automaton of $\mathsf{LexMin}_n(B_n^+)$, the size of $\mathscr{A}_n$ is exponential in n; the exponent is larger, being $2^2 = 4$ here instead of $\Phi^2 \approx 2.618$.

2 Characterising words in alternating normal form

In this section, we briefly present key results paving the way for Theorem 10, which is an "automata-flavoured" characterisation of words $w \in \mathbb{A}_n^*$ in alternating normal form. These results are based on the *ad hoc* notions of *chain*, or *chain containment* and of *rigid chain containment*, the latter two being braid invariants. Their full proofs are omitted in this paper.

▶ **Definition 4.** *Let $u \geqslant v$ be two integers. The braid word $\sigma_{u \to v} = \sigma_u \sigma_{u-1} \cdots \sigma_v$, which is the only factorisation of the braid it represents, is called a (u, v)-chain. Then, we say that a word $w \in \mathbb{A}_n^*$ contains a (u, v)-chain if $\sigma_{u \to v}$ is a subword of w, i.e., if w admits a factorisation of the form $w = w^{(u)} \sigma_u w^{(u-1)} \sigma_{u-1} \cdots w^{(v)} \sigma_v w^{(v-1)}$. If, furthermore, no factor $w^{(i)}$ contains any occurrence of the letters σ_i or σ_{i+1}, we say that w rigidly contains a (u, v)-chain.*

Although the braid word $\sigma_u \sigma_{u+1} \cdots \sigma_v$ is not a chain when $u < v$, it will also be denoted by $\sigma_{u \to v}$; considering such words may be useful since ϕ_n exchanges $\sigma_{u \to v}$ and $\sigma_{(n-u) \to (n-v)}$.

▶ **Lemma 5.** *Let u and v be two integers such that $u \geqslant v$, and let w and w' be two words representing the same braid $\beta \in B_n^+$. If w contains a (u, v)-chain, so does w'. Similarly, if w rigidly contains a (u, v)-chain, so does w'; in that case, we have $\beta \geqslant_\mathsf{R} \sigma_{u \to v}$.*

Proof idea. It suffices to treat the case where w and w' are related by a single commutation or braid relation. Then, if w rigidly contains a (u, v)-chain, an induction on i shows that $\beta \geqslant_\mathsf{R} \sigma_{i \to v}$ whenever $u \geqslant i \geqslant v$. ◀

This generalises the property that containing a letter σ_u, i.e., a (u, u)-chain, is a braid invariant.

▶ **Lemma 6.** *Let $v \leqslant n - 1$ be an integer. A braid $\beta \in B_n^+$ contains an $(n - 1, v)$-chain if and only if its n-body contains a letter σ_v.*

Proof idea. Let $\mathsf{h}_n(\beta)$ and $\mathsf{b}_n(\beta)$ be represented by two braid words $w \in \mathbb{A}_{n-1}^*$ and $w' \in \mathbb{A}_n^*$. If their concatenation ww' contains an $(n - 1, v)$-chain, the leftmost letter of that chain must already belong to w', and so must its rightmost letter σ_v. Conversely, if $\mathsf{b}_n(\beta)$ contains a letter σ_v, so does w', and each occurrence of a letter $\sigma_i \neq \sigma_{n-1}$ in w' must be preceded by an occurrence of the letter σ_{i+1}, thereby proving that w' contains an $(n - 1, v)$-chain. ◀

To obtain the desired characterisation, we introduce the notions of *left* and *right sets* of a braid.

▶ **Definition 7.** *The* left set *of a braid β is defined as the set $\mathsf{L}(\beta) = \{i \colon \sigma_i \leqslant_\mathsf{L} \beta\}$, and the* right set *of β is defined as the set $\mathsf{R}(\beta) = \{i \colon \beta \geqslant_\mathsf{R} \sigma_i\}$.*

▶ **Lemma 8.** *Let $v \leqslant n - 1$ be an integer and $\beta \in B_n^+$ be a braid such that $\mathsf{L}(\beta) = \{n - 1\}$. Either β is a chain or there exists an integer $v \leqslant n - 1$ such that $\sigma_{(n-1) \to v} \sigma_v$ is a prefix of each word representing β.*

Proof idea. Assuming that β is not a chain, let $\sigma_{(n-1) \to v} \sigma_u$ be a left divisor of β in which v is chosen minimal. If $u \geqslant v + 1$, then $u - 1 \in \mathsf{L}(\beta)$, which is impossible; Lemma 6 proves that $u \geqslant v - 1$, and the minimality of v forbids the case $u = v - 1$. ◀

The interest of these notions arises from the following result, which relates each braid $\beta \in B_n^+$ with braids $\beta \sigma_{n \to v} \in B_{n+1}^+$:

▶ **Proposition 9.** *Let $v \leqslant n$ be an integer. A braid $\beta \in B_n^+$ contains an $(n - 1, v - 1)$-chain if and only if $\mathsf{L}(\beta) = \mathsf{L}(\beta \sigma_{n \to v})$.*

Proof idea. If β contains no $(n - 1, v - 1)$-chain, $\mathsf{b}_n(\beta)$ contains no letter σ_{v-1}: it is the commutative product of two braids γ and γ', with generators in $\{\sigma_1, \ldots, \sigma_{v-2}\}$ and $\{\sigma_v, \ldots, \sigma_{n-1}\}$, respectively. But then, $\beta \sigma_{n \to v} = \mathsf{h}_n(\beta) \sigma_{n \to v} \gamma \phi_\uparrow(\gamma')$, where the morphism $\phi_\uparrow$ maps each generator σ_i such that $i \geqslant v$ to the generator σ_{i+1}. Since $\mathsf{h}_n(\beta)$ belongs to B_{n-1}^+, it commutes with σ_n, which ends up left-dividing $\beta \sigma_{n \to v}$, but not β. ◀

Conversely, if β contains an $(n-1, v-1)$-chain, every word w representing $\beta\sigma_{n\to v}$ both contains an $(n-1, v-1)$-chain and rigidly contains an (n, v)-chain. However, we can prove that each occurrence of a letter σ_i of the former chain lies to the left of the occurrence of the letter σ_{i+1} of the latter chain. Thus, $w_{\geqslant 1}$ rigidly contains an (n, v)-chain, and Lemma 5 proves that the braid β' represented by $w_{\geqslant 1}$ is right-divided by $\sigma_{n\to v}$. This means that $\beta = w_0(\beta'\sigma_{n\to v}^{-1})$ is left-divided by w_0, this reasoning being valid for each letter $w_0 \in \mathsf{L}(\beta\sigma_{n\to v})$. ◀

From these results, we can derive the following characterisation of the alternating normal form $\mathsf{Alt}_n(B_n^+)$.

▶ **Theorem 10.** *A word $w \in \mathbb{A}_n^*$ belongs to $\mathsf{Alt}_n(B_n^+)$ if and only if $n = 2$ or $n \geqslant 3$ and w has a (necessarily unique) factorisation $w = w^{(0)}\phi_n(w^{(1)})\phi_n^2(w^{(2)})\cdots\phi_n^k(w^{(k)})$ such that:*
1. *each of the words $w^{(0)}, w^{(1)}, \ldots, w^{(k)}$ belongs to $\mathsf{Alt}_{n-1}(B_{n-1}^+)$;*
2. *each of the words $\phi_{n-1}(w^{(1)}), \phi_{n-1}(w^{(2)}), \ldots, \phi_{n-1}(w^{(k)})$ belongs to $\mathsf{Alt}_{n-1}(B_{n-1}^+)$ and starts with the letter σ_{n-2};*
3. *for all $i \geqslant 1$ and $v \geqslant 1$, if $\sigma_{(n-1)\to v}$ is a prefix of $\phi_n(w^{(i+1)})\phi_n^2(w^{(i+2)})\cdots\phi_n^{k-i}(w^{(k)})$, the word $w^{(i)}$ contains an $(n-2, v-1)$-chain.*

Proof idea. Given a braid $\beta \in B_n^+$ such that $w = \mathsf{Alt}_n(\beta)$, the factors $w^{(i)}$ are the n-heads of the braids $\beta^{(0)}, \beta^{(1)}, \ldots$ given by $\beta^{(0)} = \beta$ and $\beta^{(i+1)} = \phi_n(\mathsf{b}_n(\beta^{(i)}))$. Statement **1** is true by induction on n. Then, for all $i \geqslant 1$, we have $\mathsf{L}(\mathsf{h}_n(\beta^{(i)})) \subseteq \mathsf{L}(\beta^{(i)}) = \{1\}$; thus, $w^{(i)}$ starts with the letter σ_1, and ϕ_n exchanges the alternating normal forms of the braids $\mathsf{h}_n(\beta^{(i)})$ and $\phi_n(\mathsf{h}_n(\beta^{(i)}))$. In other words, $\phi_{n-1}(w^{(i)})$ starts with the letter σ_{n-2} and coincides with the word $\mathsf{Alt}_{n-1}(\phi_{n-1}(\mathsf{h}_n(\beta^{(i)})))$, which proves statement **2**. Finally, when $i \geqslant 1$, an induction on k proves that $w^{(i)}\phi_n(w^{(i+1)})\cdots\phi_n^{i+k}(w^{(i+k)}) = \mathsf{Alt}_n(\beta^{(i)})$, and since $\mathsf{L}(\beta^{(i)}) = \{1\}$, it must coincide with its (non-empty) subset $\mathsf{L}(w^{(i)})$, thereby making statement **3** a consequence of Lemma 6 and Proposition 9.

Conversely, given a factorisation $w^{(0)}\phi_n(w^{(1)})\cdots\phi_n^k(w^{(k)})$ of a word $w \in \mathbb{A}_n^*$ that makes statements **1** to **3** valid, we prove by induction on n and k that each word $w^{(i)}$ is the alternating normal form of the n-head of the braid $\beta^{(i)}$, where $\beta^{(0)}$ is the word represented by w and $\beta^{(i+1)} = \phi_n(\mathsf{b}_n(\beta^{(i)}))$. The induction hypothesis proves that both words $w' = w^{(0)}$ and $w'' = w^{(1)}\phi_n(w^{(2)})\cdots\phi_n^{k-1}(w^{(k)})$ are the alternating normal forms of braids $\beta' \in B_{n-1}^+$ and $\beta'' \in B_n^+$, and it remains to prove that $\mathsf{L}(\beta'') \subseteq \{1\}$, which we do by induction on k.

This is vacuously true when $k = 0$ and, when $k \geqslant 1$, the induction hypothesis ensures that $w^{(1)}$ and $w^{(2)}\phi_n(w^{(3)})\cdots\phi_n^{k-2}(w^{(k)})$ are the alternating forms of $\mathsf{h}_n(\beta'')$ and $\phi_n(\mathsf{b}_n(\beta''))$. Since $\mathsf{L}(\mathsf{b}_n(\beta'')) \subseteq \{n-1\}$, Lemma 8 proves that $\mathsf{s}_n(\mathsf{b}_n(\beta''))$ coincides with a chain $\sigma_{n-1\to v}$ that is a prefix of each word representing $\mathsf{b}_n(\beta'')$, including $\phi_n(w^{(2)})\phi_n^2(w^{(3)})\cdots\phi_n^{k-1}(w^{(k)})$. Thus, statement **3** proves that $w^{(1)}$, or, equivalently, $\mathsf{h}_n(\beta'')$, contains an $(n-2, v-1)$-chain. Consequently, since $\mathsf{L}(\beta'') = \mathsf{L}(\mathsf{h}_n(\beta'')\mathsf{b}_n(\beta'')) = \mathsf{L}(\mathsf{h}_n(\beta'')\mathsf{s}_n(\mathsf{b}_n(\beta''))) = \mathsf{L}(\mathsf{h}_n(\beta'')\sigma_{n-1\to v})$, Proposition 9 proves that $\mathsf{L}(\beta'') = \mathsf{L}(\mathsf{h}_n(\beta''))$. But $\phi_{n-1}(w^{(1)})$ is the alternating normal form of a braid that must coincide with $\phi_{n-1}(\mathsf{h}_n(\beta''))$, and its first letter is σ_{n-2}, which means that the $(n-1)$-head of $\phi_{n-1}(\mathsf{h}_n(\beta''))$ is empty, i.e., that $\mathsf{L}(\phi_{n-1}(\mathsf{h}_n(\beta'')) \subseteq \{n-2\}$. It follows, as desired, that $\mathsf{L}(\mathsf{h}_n(\beta'')) \subseteq \{1\}$. ◀

3 Minimal automata

In this section, we explicitly build the minimal automaton of the language $\mathscr{L}_n = \mathsf{Alt}_n(B_n^+)$. In order to do so, a crucial step lies in building the minimal automaton of the language $\mathsf{Alt}_n(\mathsf{b}_n(B_n^+))$. Noting that $\mathsf{b}_n(B_n^+)$ coincides with the set of braids $\beta \in B_n^+$
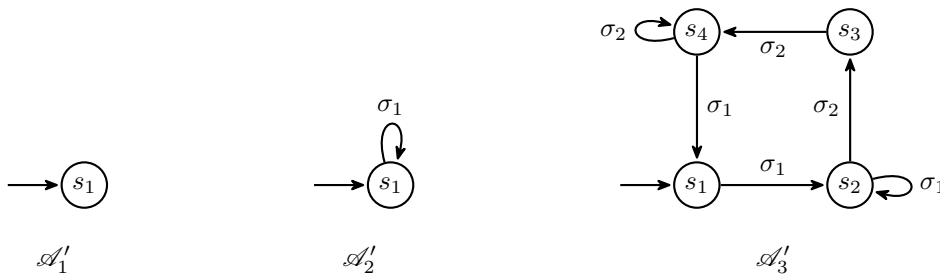
■ **Figure 4** Minimal automata $\mathscr{A}_1'$ to $\mathscr{A}_3'$.

such that $\mathsf{L}(\beta) \subseteq \{\sigma_{n-1}\}$, it turns out that $\phi_n(\mathsf{Alt}_n(\beta)) = \mathsf{Alt}_n(\phi_n(\beta))$ whenever $\beta \in \mathsf{b}_n(B_n^+)$. Thus, we also look at the language $\mathscr{L}_n' = \mathsf{Alt}_n(\phi_n(\mathsf{b}_n(B_n^+)))$, which is connected to $\mathscr{L}_n$ by the relation $\mathscr{L}_n = \mathscr{L}_{n-1}\phi_n(\mathscr{L}_n')$. In particular, $\mathscr{L}_n'$ is the language of alternating normal forms of braids $\beta \in B_n^+$ such that $\mathsf{L}(\beta) \subseteq \{\sigma_1\}$; viewing B_n^+ as a subset of B_{n+1}^+ proves that $\mathscr{L}_n' \subseteq \mathscr{L}_{n+1}'$, which is the reason why we chose to study $\mathscr{L}_n'$ and not its conjugate $\phi_n(\mathscr{L}_n')$.

Both languages $\mathscr{L}_n$ and $\mathscr{L}_n'$ are prefix-closed: when they contain a word w, they also contain all its prefixes. Thus, we identify the minimal automaton of $\mathscr{L}_n$ with a tuple $\mathscr{A}_n = (V_n, \mathbb{A}_n, \delta_n, \iota_n)$, in which V_n denotes the set of states, all of which are accepting; $\mathbb{A}_n$ is the alphabet; $\delta_n \colon V_n \times \mathbb{A}_n \mapsto V_n$ denotes the transition function; and $\iota_n \in V_n$ denotes the initial state of the automaton. Similarly, we identify the minimal automaton of $\mathscr{L}_n'$ with a tuple $\mathscr{A}_n' = (V_n', \mathbb{A}_n, \delta_n', \iota_n')$, and we shall first focus on constructing the automata $\mathscr{A}_n'$. However, in order to define this latter automaton, we will first define an auxiliary automaton $\mathscr{A}_n'' = (V_n'', \mathbb{A}_n, \delta_n'', \iota_n'')$ that will recognise $\mathscr{L}_n'$ but not be minimal; the automaton $\mathscr{A}_n'$ will be built by minimising $\mathscr{A}_n''$.

The languages $\mathscr{L}_1'$ and $\mathscr{L}_2'$ consist of the empty word and of all words on the alphabet $\mathbb{A}_2 = \{\sigma_1\}$, respectively. Then, when $n \geqslant 3$, we focus on building the automata $\mathscr{A}_n$ and $\mathscr{A}_n'$ based on the automata $\mathscr{A}_{n-1}$ and $\mathscr{A}_{n-1}'$. Of course, we may use Theorem 10 directly to compute any automaton $\mathscr{A}_n$; for instance, $\mathscr{L}_3'$ consists of words of the form $(\sigma_1\sigma_1^*\sigma_2^2\sigma_2^*\sigma_1)^*$ and their prefixes, from which we deduce the automaton $\mathscr{A}_3'$ given in Figure 4.

Below, we proceed in four steps: first, we present a few preliminary results; second, we construct an automaton $\mathscr{A}_n''$ that recognises the language $\mathscr{L}_n'$; third, we construct the minimal automaton $\mathscr{A}_n'$ of $\mathscr{L}_n'$; fourth, we construct the minimal automaton $\mathscr{A}_n$ of $\mathscr{L}_n$ itself.

3.1 Preliminary results

When $n \geqslant 3$, and due to Theorem 10, a word $w \in \mathbb{A}_n^*$ belongs to $\mathscr{L}_n'$ if and only if w has a factorisation $w = w^{(0)}\phi_n(w^{(1)})\cdots\phi_n^k(w^{(k)})$ such that:

2.' each word $w^{(i)}$ belongs to $\mathscr{L}_{n-1}'$;

3.' for all $i \geqslant 0$ and $v \geqslant 1$, if $\sigma_{(n-1)\to v}$ is a prefix of $\phi_n(w^{(i+1)})\phi_n^2(w^{(i+2)})\cdots\phi_n^{i+k}(w^{(k)})$, the word $w^{(i)}$ contains a $(n-2, v-1)$-chain.

These are variants of criteria **2** and **3**. Criterion **3'** also requires, while reading a word $w \in \mathscr{L}_{n-1}'$, recalling the least integer v (if any) for which w contains an $(n-2, v)$-chain. This integer is denoted by $\mathsf{ch}_{n-1}(w)$; when w contains no $(n-2, v)$-chain at all, i.e., when $w \in \mathbb{A}_{n-2}^*$, we set $\mathsf{ch}_{n-1}(w) = n-1$. In particular,

i) if $k = 0$, we simply have $\mathsf{ch}_{n-1}(w) = n-1$;

ii) if $k = 1$, the integer $\mathsf{ch}_{n-1}(w)$ may vary between 2 and $n-2$;

iii) if $k \geqslant 2$, the word w contains the $(n-2, 1)$-chain $\sigma_{(n-2)\to 1}$, and $\mathsf{ch}_{n-1}(w) = 1$.

The following results allow determining $\mathsf{ch}_{n-1}(w)$ in case **ii)**.

▶ **Lemma 11.** *Let w be a word belonging to $\mathscr{L}_n$, and let σ_u be its rightmost letter, or $\sigma_u = \sigma_1$ if w is empty. For all $v \leqslant n-1$, the word $w\sigma_{u \to v}$ also belongs to $\mathscr{L}_n$; furthermore, if w belongs to $\mathscr{L}_n'$, so does $w\sigma_{u \to v}$. Finally, if $n \geqslant 2$ and w is a non-empty word in $\mathscr{L}_n'$, there exist at least two integers x and y such that $w\sigma_x$ and $w\sigma_y$ belong to $\mathscr{L}_{n+1}'$.*

Proof. We prove both statements of Lemma 11 separately. First, let ℓ be the length of w, and let $w' = w\sigma_{u \to v}$. The first statement being immediate when $\ell = 0$, we assume that $\ell \geqslant 1$. Since the braid word $\sigma_u \sigma_{u \to v}$ is the only representative of its braid, we have $\mathsf{s}_n(\sigma_u \sigma_{u \to v}) = \sigma_u$. A backward induction on i proves then that $\mathsf{s}_n(w_{\geqslant i}) = \mathsf{s}_n(w'_{\geqslant i})$ for all $i \leqslant \ell - 1$. It follows from the remark **3)** of page 4 that $w\sigma_{u \to v} \in \mathscr{L}_n$, and that $\mathsf{L}(w\sigma_{u \to v}) = \mathsf{L}(w)$, thereby proving that $w\sigma_{u \to v} \in \mathscr{L}_n'$ if $w \in \mathscr{L}_n'$.

   We prove the last statement by induction on n. If $n = 2$, the word w is of the form $w = \sigma_1^\ell$, and $w\sigma_1$ and $w\sigma_2$ belong to $\mathscr{L}_3'$. If $n \geqslant 3$, let $w = w^{(0)} \phi_n(w^{(1)}) \cdots \phi_n^k(w^{(k)})$ be the factorisation of w given in Theorem 10. If $k = 0$, the word w belongs to $\mathscr{L}_{n-1}'$, and the induction hypothesis also proves that there exist two integers x and y for which $w\sigma_x$ and $w\sigma_y$ belong to $\mathscr{L}_n'$. Otherwise, $k \geqslant 1$, and w both ends with some letter σ_u and contains the letter σ_{n-1}, which proves that both $w\sigma_u$ and $w\sigma_n$ belong to $\mathscr{L}_{n+1}'$. ◀

▶ **Lemma 12.** *Let w be a word belonging to both $\mathscr{L}_{n-1}'$ and $\mathscr{L}_{n-2}'\phi_n(\mathscr{L}_{n-2}')$, and let $w^{(0)}$ be its longest prefix belonging to $\mathbb{A}_{n-2}^*$:*

- *if w has a factorisation of the form $w = w^{(0)}\sigma_{(n-2) \to v}$, then $\mathsf{ch}_{n-1}(w) = v$;*
- *otherwise, let v be the least integer such that $w\sigma_v \in \mathscr{L}_{n-1}'$: we have $\mathsf{ch}_{n-1}(w) = v + 1$.*

Proof. The first part of Lemma 12 being immediate, we focus on the second part. In that case, let $w^{(1)}$ be the suffix of w such that $w = w^{(0)}w^{(1)}$, and let $\beta \in B_{n-1}^+$ be the braid represented by w. By construction, $w^{(1)}$ belongs to $\phi_{n-1}(\mathbb{A}_{n-2}^*)$, and $w^{(1)} = \mathsf{Alt}_{n-1}(\mathsf{b}_{n-1}(\beta))$. Furthermore, by Lemma 6, $\mathsf{ch}_{n-1}(w)$ is simply the least letter of $w^{(1)}$, say, y.

   In particular, when $z \leqslant y - 2$, the word $w\sigma_z$ contains no $(n-2, z)$-chain, and Lemma 6 prevents it from belonging to $\mathscr{L}_{n-1}'$. Conversely, the word $\phi_{n-1}(w^{(1)}\sigma_{y-1}) = \phi_{n-1}(w^{(1)})\sigma_{n-y}$ satisfies both criteria **2'** and **3'** that mark it as a member of $\mathscr{L}_{n+1-y}'$, and thus of $\mathscr{L}_{n-1}'$. Moreover, $w^{(1)}$ is not a chain, so that the maximal chains that are prefixes of $w^{(1)}$ and of $w^{(1)}\sigma_{y-1}$ coincide with each other. Consequently, $w\sigma_{y-1}$ itself belongs to $\mathscr{L}_{n-1}'$. ◀

   As a consequence of Lemma 12, for each word w in $\mathscr{L}_{n-1}'$, the integer $\mathsf{ch}_{n-1}(w)$ depends only on whether $\sigma_{n-2}\sigma_1$ is a subword of w and, if not, on the residual of w (i.e., on the set $\{x \in \mathbb{A}_{n-1}^* : wx \in \mathscr{L}_{n-1}'\}$). For each automaton $\mathscr{A} = (V, \mathbb{A}_{n-1}, \delta, \iota)$ recognising the language $\mathscr{L}_{n-1}'$, the residual of a word $w \in \mathscr{L}_{n-1}'$ depends only on the state $s = \delta(\iota, w)$ of $\mathscr{A}$ to which w is mapped. Thus, below, for each state s of $\mathscr{A}_{n-1}'$, we simply note $\mathsf{ch}_{n-1}(s)$ the common value of the integers $\mathsf{ch}_{n-1}(w)$ when $\delta'(\iota_{n-1}', w) = s$ and $\sigma_{n-2}\sigma_1$ is not a subword of w; if no such word w exists, we set $\mathsf{ch}_{n-1}(s) = 1$.

3.2 Construction and correctness of the automaton $\mathscr{A}_n''$

Here, we give a construction of an automaton $\mathscr{A}_n''$ that recognises the language $\mathscr{L}_n'$. This automaton *looks like* its minimal equivalent $\mathscr{A}_n'$ represented in Figure 5 when $n = 4$. The semantics of its states is given in the beginning of the proof of Proposition 13.

▶ **Proposition 13.** *Given an integer $n \geqslant 4$, let $\mathscr{A}_{n-1}' = (V_{n-1}', \mathbb{A}_{n-1}, \delta_{n-1}', \iota_{n-1}')$ be the minimal automaton recognising $\mathscr{L}_{n-1}'$. The language $\mathscr{L}_n'$ is recognised by the (deterministic, non-minimal) automaton $\mathscr{A}_n'' = (V_n'', \mathbb{A}_n, \delta_n'', \iota_n'')$ defined as follows. The state set of $\mathscr{A}_n''$ is given by $V_n'' = ((V_{n-1}' \times \{\top, \bot\}) \cup P_n'') \times \{\mathsf{Id}_n, \phi_n\}$, where we set $P_n'' = \{p_i^j : 2 \leqslant j \leqslant i \leqslant n-1\}$; its initial state is $\iota_n'' = (\iota_{n-1}', \bot, \mathsf{Id}_n)$; and its transition function δ_n'' is given by:*

a. $\delta_n''((s, \mathsf{f}, \mathsf{Id}_n), \sigma_i) = (\delta_{n-1}'(s, \sigma_i), \mathsf{f}, \mathsf{Id}_n)$ *when* $\mathsf{ch}_{n-1}(s) \neq 2$ *or* $i \neq 1$;

b. $\delta_n''((s, \mathsf{f}, \mathsf{Id}_n), \sigma_1) = (\delta_{n-1}'(s, \sigma_1), \top, \mathsf{Id}_n)$ *when* $\mathsf{ch}_{n-1}(s) = 2$;

c. $\delta_n''((p_i^j, \mathsf{Id}_n), \sigma_{n+1-i}) = (p_{i-1}^j, \mathsf{Id}_n)$ *when* $p_{i-1}^j \in P_n$;

d. $\delta_n''((p_i^j, \mathsf{Id}_n), \sigma_{n-i}) = (\delta_{n-1}'(\imath_{n-1}', \sigma_{1 \to (n-i)} \sigma_{n-i}), \bot, \mathsf{Id}_n)$ *when* $p_i^j \in P_n$;

e. $\delta_n''((s, \bot, \mathsf{Id}_n), \sigma_{n-1}) = (p_{n-1}^j, \phi_n)$ *when* $3 \leqslant j \leqslant n-1$ *and* $\mathsf{ch}_{n-1}(s) = j-1$;

f. $\delta_n''((s, \top, \mathsf{Id}_n), \sigma_{n-1}) = (p_{n-1}^2, \phi_n)$;

g. $\delta_n''((s, \mathsf{f}, \phi_n), \sigma_i) = (s', \mathsf{f}', \phi_n^{k+1})$ *when* $\delta_n''((s, \mathsf{f}, \mathsf{Id}_n), \sigma_{n-i}) = (s', \mathsf{f}', \phi_n^k)$;

h. $\delta_n''((s, \mathsf{f}, \phi_n^k), \sigma_i)$ *is not defined in all other cases.*

Proof. Let $w = w^{(0)} \phi_n(w^{(1)}) \cdots \phi_n^k(w^{(k)})$ be the factorisation a word $w \in \mathscr{L}_n'$ given in Theorem 10. Here is the intended semantics of the state $\delta_n''(\imath_n'', w)$ to which w is sent. The state $\delta_n''(\imath_n'', w)$ shall be of the form $(s, \mathsf{f}, \phi_n^k)$ or (p_i^j, ϕ_n^k), thereby indicating whether the factorisation of w contains an odd or an even number of factors. In general, i.e., when $k = 0$ or when $\phi_n(w^{(k)})$ is *not* a chain, we shall set $s = (\delta_{n-1}'(\imath_{n-1}', w^{(k)}), \mathsf{f}, \phi_n^k)$, where f is a boolean flag set to $\mathsf{f} = \top$ (i.e., $\mathsf{f} = \mathbf{true}$) if $w^{(k)}$ contains an $(n-2, 1)$-chain, and $\mathsf{f} = \bot$ (i.e., $\mathsf{f} = \mathbf{false}$) otherwise. However, if $k \geqslant 1$ and $\phi_n(w^{(k)})$ is an $(n-1, v)$ chain, let $x = \mathsf{ch}_{n-1}(w^{(k-1)})$: the state $\delta_n''(\imath_n'', w)$ shall be the pair (p_v^{x+1}, ϕ_n^k), thereby indicating that $w^{(k)}$ is an $(n-1, v)$-chain that can be extended to form an $(n-1, x+1)$-chain, but not more.

We can now prove by induction on ℓ that, for each word $w \in \mathscr{L}_n'$ of length ℓ, the state $\delta_n''(\imath_n'', w)$ has the intended semantics, thereby demonstrating that $\mathscr{A}_n''$ recognises the language $\mathscr{L}_n'$. The base case $\ell = 0$ follows from our choice for $\imath_n''$; now, assuming that $\ell \geqslant 1$ and that $\ell - 1$ satisfies the inductive property, we wish to prove that ℓ also satisfies this property, by considering the cases **a** and **h** separately.

Case **a** is the general case obtained when facing a state $\delta_n''(\imath_n'', w)$ for which k is even: reading a letter σ_i will just let the word $w^{(k)}$ grow inside the language $\mathscr{L}_{n-1}'$, without changing the flag f or parity of k. By contrast, case **b** happens when f changes from $\bot$ to $\top$: indeed, the word $w^{(k)}$ already contained an $(n-2, 2)$-chain, and adding a letter σ_1 yields an $(n-2, 1)$-chain. Thus, the transition labelled σ_1 targets the state $(\delta_{n-1}'(s, \sigma_1), \top, \mathsf{Id}_n)$ instead of the state $(\delta_{n-1}'(s, \sigma_1), \bot, \mathsf{Id}_n)$ that might otherwise have been expected.

Then, cases **c** and **d** focus on transitions leaving a state (p_i^j, Id_n): either we read the letter σ_{n+1-i} (in case **c**) and we just transformed the $(n-2, i)$-chain $\phi_n(w^{(k)})$ into an $(n-2, i-1)$-chain $\phi_n(w^{(k)}) \sigma_{i-1}$, or we read the letter σ_{n-i} (in case **c**), in which case $w^{(k)} \sigma_{n-i}$ stops being a chain, and we should just remember the same amount of information as if we were reading the word $w^{(k)} \sigma_{n-i}$ instead of $w^{(0)} \phi_n(w^{(1)}) w^{(2)} \cdots w^{(k)} \sigma_{n-i}$.

Cases **e** and **f** focus on transitions labelled σ_{n-1}: they target the state $(p_{n-1}^{x+1} s, \phi_n)$, where x is the least integer such that $w^{(k)}$ contained an $(n-2, x)$-chain; that integer x is given by Lemma 12 in case **e**, where $w^{(k)}$ contains no $(n-2, 1)$-chain; it is just 1 in case **f**.

Finally, case **g** replicates all cases **a** to **f**, but when k is odd instead of even. And case **h** consists in observing that cases **a** to **g** already cover all possible transitions of the automaton $\mathscr{A}_n''$. ◀

3.3 Construction, correctness and minimality of the automaton $\mathscr{A}_n'$

When $n \geqslant 4$, the automaton $\mathscr{A}_n''$ is not minimal. A first reason is that some states have the same residuals, and should thus be merged; we recall that the residual of a state $s \in V_n''$ is the language $\mathscr{L}_n''(s)$ of those words $w \in \mathbb{A}_n^*$ for which $\delta_n''(s, w)$ exists and is accepting. A second reason is that, when $n \geqslant 5$, some states of $\mathscr{A}_n''$ are not even accessible. Thus, we shall transform $\mathscr{A}_n''$ into the minimal automaton $\mathscr{A}_n'$ of $\mathscr{L}_n'$.

▶ **Definition 14.** *Given an integer $n \geqslant 4$, let $\mathscr{A}''_n = (V''_n, \mathbb{A}_n, \delta''_n, \imath'_n)$ be the automaton built in Proposition 13. The automaton $\mathscr{A}'_n = (V'_n, \mathbb{A}_n, \delta'_n, \imath'_n)$ is defined from $\mathscr{A}''_n$ by*

- *merging the states $(\imath'_{n-1}, \bot, \phi^k_n)$ and $(p^{n-1}_{n-1}, \phi^k_n)$ for each $\phi^k_n \in \{\mathsf{Id}_n, \phi_n\}$; this amounts to deleting the state $(p^{n-1}_{n-1}, \phi^k_n)$ and redirecting toward $(\imath'_{n-1}, \bot, \phi^k_n)$ those transitions of $\mathscr{A}''_n$ that were targeted toward $(p^{n-1}_{n-1}, \phi^k_n)$;*
- *deleting states of the form $((p^j_i, \mathsf{Id}_{n-1}), \bot, \phi^k_n)$, where $p^j_i \in P'_{n-1}$ and $\phi^k_n \in \{\mathsf{Id}_n, \phi_n\}$.*

In particular, the actual state set of $\mathscr{A}'_n$ is $V'_n = ((\overline{V}'_{n-1} \times \{\bot\}) \cup (V'_{n-1} \times \{\top\}) \cup P'_n) \times \{\mathsf{Id}_n, \phi_n\}$, where $P'_n = P''_n \setminus \{p^{n-1}_{n-1}\}$ and $\overline{V}'_{n-1} = V'_{n-1} \setminus (P'_{n-1} \times \{\mathsf{Id}_{n-1}\})$.

The reason why we shall merge the states $\imath''_n = (\imath'_{n-1}, \bot, \mathsf{Id}_n)$ and $(p^{n-1}_{n-1}, \mathsf{Id}_n)$ is that both states have only one outgoing transition, labelled by σ_1, and whose target state is $\delta''_n(\imath''_n, \sigma_1)$. Similarly, we shall merge the states $(\imath'_{n-1}, \bot, \phi_n)$ and (p^{n-1}_{n-1}, ϕ_n). Below, we will consider $(p^{n-1}_{n-1}, \phi^k_n)$ as an alias for $(\imath'_{n-1}, \bot, \phi^k_n)$.

Then, we shall also delete states of the form $((p^j_i, \mathsf{Id}_{n-1}), \bot, \phi^k_n)$ because they are not accessible. Indeed, by construction of δ''_{n-1}, every path in $\mathscr{A}''_{n-1}$ (or in $\mathscr{A}'_{n-1}$) and ending in such a state $(p^j_i, \mathsf{Id}_{n-1})$ must have previously visited a state of the form (s, ϕ_{n-1}). Thus, every word w for which $\delta'_{n-1}(\imath'_{n-1}, w) = (p^j_i, \mathsf{Id}_{n-1})$ already contains an $(n-2, 1)$-chain, thereby proving that $\delta''_n(\imath'_n, w) = ((p^j_i, \mathsf{Id}_{n-1}), \top, \mathsf{Id}_n)$. It follows, as announced, that $((p^j_i, \mathsf{Id}_{n-1}), \bot, \mathsf{Id}_n)$ is inaccessible, and we prove similarly that $((p^j_i, \mathsf{Id}_{n-1}), \bot, \phi_n)$ is also inaccessible.

An example of this construction, when $n = 4$, is given in Figure 5, where we start from the 4-state automaton $\mathscr{A}'_3$, whose states are denoted by s_1 to s_4, and obtain the 20-state automaton $\mathscr{A}'_4$. We wish we had represented the automaton $\mathscr{A}'_5$, thereby showing why replacing V'_{n-1} by $\overline{V}'_{n-1}$ is important, but $\mathscr{A}'_5$ contains 86 states, which is difficult to read. For the sake of readability, each state $(s_i, \mathsf{f}, \phi^k_n)$ is denoted by s^{f}_i, with a bar when $\phi^k_n = \phi_n$, i.e., when k is odd. We added the dangling states (p^j_i, ϕ^k_n), which are also denoted by p^j_i, with a bar when $\phi^k_n = \phi_n$.

While the above paragraphs prove that the automaton $\mathscr{A}'_n$ recognises the same language as $\mathscr{A}''_n$, i.e., the language $\mathscr{L}'_n$, we shall now prove that its states are accessible and have pairwise distinct residuals. Once again, we proceed by induction and, since this result is clear when $n \leqslant 3$, we assume that $n \geqslant 4$ and that $\mathscr{A}'_{n-1}$ is already known to be minimal.

In Lemma 15, we prove not only that each state of $\mathscr{A}'_n$ is accessible, but also most states can be reached via a word that does not contain any $(n-1, 1)$-chain. This will be crucial toward proving, in Lemma 16, that for any two distinct states s and s', there exists a word that can be read from s but not from s', or from s' but not from s.

▶ **Lemma 15.** *The automaton $\mathscr{A}'_n$ is strongly connected. Furthermore, for each state s in $\overline{V}'_n$, i.e., each state s distinct from the states (p^j_i, Id_n) for which $p^j_i \in P'_n$, there exists a word w that does not contain any $(n-1, 1)$-chain and such that $\delta'_n(\imath'_n, w) = s$.*

Proof. The result being visibly correct when $n \leqslant 3$, let us assume that $n \geqslant 4$. Given a state $s \in V'_{n-1}$, the induction hypothesis ensures that there exists a word $w \in \mathscr{L}'_{n-1}$ without $(n-1, 1)$-chain for which $\delta'_{n-1}(\imath'_{n-1}, w) = s$. Since $\mathscr{A}'_{n-1}$ is strongly connected, it also tells us that there exists a non-empty word $w' \in \mathscr{L}'_{n-1}$ for which $\delta'_{n-1}(\imath'_{n-1}, \sigma_1 w') = \imath'_{n-1}$; it follows that $\delta'_n(\imath'_n, w'') = (\imath'_{n-1}, \top, \mathsf{Id}_n)$, where $w'' = \sigma_{1 \to (n-2)} \sigma_{(n-2) \to 1} w'$.

Finally, for each state $p^j_i \in P'_n$, the word $w^j_i = \sigma_{1 \to (n-2)} \sigma_{(n-2) \to (j-1)} \sigma_{(n-1) \to i}$ obeys the relation $\delta'_n(\imath'_n, w^j_i) = (p^j_i, \phi_n)$; identifying the states $(\imath'_{n-1}, \bot, \phi_n)$ and p^{n-1}_{n-1} also makes this construction valid when $p^j_i = p^{n-1}_{n-1}$. We complete the proof by observing that:

- $(s, \bot, \mathsf{Id}_n) = \delta'_n(\imath'_n, w)$;
- $(s, \top, \mathsf{Id}_n) = \delta'_n(\imath'_n, w'' w)$;
- $(p^j_i, \phi_n) = \delta'_n(\imath'_n, w^j_i)$, including when $p^j_i = p^{n-1}_{n-1}$, i.e., when $(p^j_i, \phi_n) = (\imath'_{n-1}, \bot, \phi_n)$;

Figure 5 Automaton $\mathscr{A}'_4$, which contains four copies of $\mathscr{A}'_3$ (circled in gray, two of which miss an edge) and a few dangling states p_i^j. Erasing dashed edges gives us an automaton that recognises the language $\mathscr{L}'_4 \cap (\mathscr{L}'_3\phi_4(\mathscr{L}'_3))$, in which p_2^2 and p_3^2 are no longer accessible, and whose states have been split into four classes, coloured in brown, blue, green and red: a state s lies in the blue (resp., green, red) class when $\mathsf{ch}_4(w) = 2$ (resp., 3, 4) for all the words $w \in \mathscr{L}'_3\phi_4(\mathscr{L}'_3)$ such that $\delta'_4(\imath'_4, w) = s$. Once a dashed edge has been taken, we just have $\mathsf{ch}_4(w) = 1$. Similarly, we have $\mathsf{ch}_3(w) = 1$ (resp., 2, 3) for all the words $w \in \mathscr{L}'_3$ such that $\delta'_4(\imath'_4, w) = s$ when $s \in \{s_1^\top, s_2^\top, s_3^\top, s_4^\top\}$ (resp., $\{s_3^\bot, s_4^\bot\}$, $\{s_1^\bot, s_2^\bot\}$).

- $(s, \bot, \phi_n) = \delta'_n((\imath'_{n-1}, \bot, \phi_n), \phi_n(w))$;
- $(s, \top, \phi_n) = \delta'_n((\imath'_{n-1}, \bot, \phi_n), \phi_n(w''w))$;
- $(p_i^j, \mathsf{Id}_n) = \delta'_n((\imath'_{n-1}, \bot, \phi_n), \phi_n(w_i^j))$, including when $(p_i^j, \mathsf{Id}_n) = (\imath'_{n-1}, \bot, \mathsf{Id}_n) = \imath'_n$. ◀

▶ **Lemma 16.** *The states of $\mathscr{A}'_n$ have pairwise distinct residuals.*

Proof. Let s and s' be distinct states of $\mathscr{A}'_n$, and let w and w' be non-empty words in $\mathscr{L}'_n$ such that $\delta'_n(\imath'_n, w) = s$ and $\delta'_n(\imath'_n, w') = s'$. In addition, let $w^{(0)}\phi_n(w^{(1)})\cdots\phi_n^k(w^{(k)})$ and $w'^{(0)}\phi_n(w'^{(1)})\cdots\phi_n^\ell(w'^{(\ell)})$ be their factorisations given by Theorem 10; we assume that w and w' were chosen so that k and ℓ are minimal.

We first prove that the states of the form $(t, \mathsf{f}, \mathsf{Id}_n)$ or (p_i^j, Id_n) have pairwise distinct residuals. We call such states "Id_n-states", as opposed to "ϕ_n-states":

- If s is of the form (p_i^i, Id_n), including if $i = n - 1$, the only letter in $\mathscr{L}'_n(s)$ is σ_{n-i}. On the contrary, $\mathscr{L}'_n(s')$ contains at least two letters, both if s' is of the form (p_i^j, Id_n) with $i \neq j$, since two such letters are σ_{n-i} or σ_{n+1-i}, or if s' is of the form $(t', \mathsf{f}', \mathsf{Id}_n)$, because of Lemma 11 (which we can use because we forced w and w' to be non-empty). Thus, each state (p_i^i, Id_n) has a residual distinct from all other Id_n-states.

- If s is of the form (p_i^j, Id_n), it is the only state such that $\delta'_n(s, \sigma_{(n+1-j)\to(n-i)}) = (p_i^i, \mathsf{Id}_n)$. Thus, the state $\delta'_n(s', \sigma_{(n+1-j)\to(n-i)})$ either fails to exist or does not coincide with (p_i^i, Id_n): in both cases, s and s' have distinct residuals.
- If both s and s' are of the form $(t, \mathsf{f}, \mathsf{Id}_n)$ and $(t', \mathsf{f}', \mathsf{Id}_n)$, either $t \neq t'$, in which case the induction hypothesis proves that $\mathscr{L}'_n(s) \cap \mathscr{L}'_{n-1} = \mathscr{L}'_{n-1}(t) \neq \mathscr{L}'_{n-1}(t') = \mathscr{L}'_n(s') \cap \mathscr{L}'_{n-1}$, or $\mathsf{f} \neq \mathsf{f}'$, in which case the chain $\sigma_{(n-1)\to 2}$ belongs to exactly one of the residuals $\mathscr{L}'_n(s)$ and $\mathscr{L}'_n(s')$.

Similarly, all ϕ_n-states have distinct residuals.

Finally, assume that s is an Id_n-state and that s' is a ϕ_n-state. Let σ_u be the last letter of w, with $u \leqslant n-2$, and let $z = \mathsf{ch}_{n-1}(s)$. We have $\delta'_n(s, \sigma_{u\to(n-1)}) = (p_{n-1}^{z+1}, \phi_n)$. On the contrary, either the state $s_1 = \delta'_n(s', \sigma_u)$ is a ϕ_n-state, in which case the state $s_2 = \delta'_n(s', \sigma_{u\to(n-2)})$ is also a ϕ_n-state and $\delta'_n(s', \sigma_{u\to(n-1)}) = \delta'_n(s_2, \sigma_{n-1})$ differs from (p_{n-1}^{z+1}, ϕ_n), or s_1 is an Id_n-state of the form $(p_{n-1}^i, \mathsf{Id}_n)$, in which case $\delta'_n(s', \sigma_{u\to(n-1)}) = \delta'_n(s_1, \sigma_{(u+1)\to(n-1)})$ is also an Id_n-state. ◀

As a consequence of Lemmas 15 and 16, we obtain the following result.

▶ **Proposition 17.** *The automaton $\mathscr{A}'_n$ is the minimal automaton of the language $\mathscr{L}'_n$.*

3.4 Construction, correctness and minimality of the automaton $\mathscr{A}_n$

We finally construct the automata $\mathscr{A}_n$ as follows.

▶ **Definition 18.** *First, $\mathscr{A}_1 = \mathscr{A}'_1$ is the automaton with one unique (necessarily initial) state and no transition, and $\mathscr{A}_2 = \mathscr{A}'_2$ is the automaton with one unique state and one loop labelled by σ_1. Then, when $n \geqslant 3$, the state set and initial state of $\mathscr{A}_n$ are defined by $V_n = V_{n-1} \cup V'_n$ and $\imath_n = \imath_{n-1}$, and the transition function δ_n is defined as follows:*
- *$\delta_n(s, \sigma_u) = \delta_{n-1}(s, \sigma_u)$ when $s \in V_{n-1}$ and $u \leqslant n-2$;*
- *$\delta_n(s, \sigma_{n-1}) = \delta'_n(\imath'_n, \sigma_1)$ when $s \in V_{n-1}$;*
- *$\delta_n(s, \sigma_u) = \delta'_n(s, \sigma_{n-u})$ when $s \in V'_n$*

The relation $\mathscr{L}_n = \mathscr{L}_{n-1}\phi_n(\mathscr{L}'_n)$ proves that this automaton recognises the set $\mathscr{L}_n$. Then, we prove by induction that each state s of $\mathscr{A}_n$ is accessible: when $s \in V_{n-1}$, this is just the induction hypothesis, and when $s \in V'_n$, Lemma 15 proves that s is accessible from $\delta'_n(\imath'_n, \sigma_1)$, which is itself accessible *via* the one-letter word σ_{n-1}.

Our last task consists in proving that any two states s and s' of $\mathscr{A}_n$ have pairwise distinct residuals. When s and s' belong to V_{n-1}, this is the induction hypothesis, and when s and s' belong to V'_n, this is the irreducibility of $\mathscr{A}'_n$. Finally, when $s \in V_{n-1}$ and $s' \in V'_n$, let w be a word such that $\delta_n(\imath_n, w) = s$, and let σ_u be its last letter, or $\sigma_u = \sigma_1$ in case w is empty. Then, let $w' = \sigma_{u\to 1}$. Lemma 11 proves that ww' belongs to $\mathscr{L}_{n-1}$, so that $\delta_n(s, w'\sigma_{n-1}) = \delta'_n(\imath'_n, \sigma_1)$; this is a state of $\mathscr{A}'_n$ that differs from all states (p_i^j, Id_n). By contrast, if they ever exist, the state $\delta_n(s', w') = \delta'_n(s', \phi_n(w'))$ is a ϕ_n-state, so that $\delta_n(s', w'\sigma_{n-1}) = \delta'_n(\delta'_n(s', \phi_n(w')), \sigma_1)$ is a state $(p_{n-1}^i, \mathsf{Id}_n)$ – which may be the state $(p_{n-1}^{n-1}, \mathsf{Id}_n) = \imath'_n$. From the above discussion results the following theorem.

▶ **Theorem 19.** *The automaton $\mathscr{A}_n$ is the minimal automaton of the language $\mathsf{Alt}_n(B_n^+)$.*

4 Size of the minimal automata

This final section is devoted to evaluate the size of the minimal automaton $\mathscr{A}_n$ of the language $\mathsf{Alt}_n(B_n^+)$, both in terms of states and of transitions.

▶ **Theorem 20.** *The automaton $\mathscr{A}_n$ has s_n states and t_n transitions, where $s_1 = 1$, $t_1 = 0$, and*

$$s_n = \frac{25 \times 2^{2n-3} - 9n^2 + 3n + 7}{27} \ \text{and} \ t_n = \frac{(225n - 290)2^{2n-5} - 9n^3 - 9n^2 + 93n - 77}{81}$$

whenever $n \geqslant 2$.

Proof. Below, let s'_n and t'_n denote the number of states and of transitions of the automaton $\mathscr{A}'_n$. First, we have $s'_1 = s'_2 = 1$, $s'_3 = 4$ and $t'_1 = 0$, $t'_2 = 1$, $t'_3 = 6$. Then, when $n \geqslant 3$, note that $|P'_n| = |P'_{n-1}| + (n-2)$. It follows that

$$s'_n = |V'_n| = 2((|V'_{n-1}| - |P'_{n-1}|) + |V'_{n-1}| + |P'_n|) = 4s'_{n-1} + 2(n-2),$$

and an immediate induction proves that $s'_n = (25 \times 2^{2n-5} - 6n + 4)/9$ for all $n \geqslant 3$.

Furthermore, each state (p_i^j, ϕ_n^k) is of out-degree 2 when $i > j$, and 1 when $i = j$; the latter cases occurs $n - 3$ times, and thus, $2|P'_n| - (n-3)$ transitions leave a state (p_i^j, Id_n). In addition, those Id_n-states of $\mathscr{A}'_n$ from which one can read a letter σ_{n-1} are the states $(s, \mathsf{f}, \mathsf{Id}_n)$ for which $\mathsf{f} = \top$ or s is a ϕ_{n-1}-state of $\mathscr{A}'_{n-1}$; there are s'_{n-1} states in the first family, and $s'_{n-1}/2$ states in the second family but not in the first one. Consequently, in total, there are

$$(t'_{n-1} - (2|P'_{n-1}| - (n-4))) + t'_{n-1} + (2|P'_n| - (n-3)) + 3s'_{n-1}/2$$

transitions leaving Id_n-states, and $t'_n = 4t'_{n-1} + 3s'_{n-1} + 2(2n-5)$. Hence, another induction proves that $t'_n = ((25n - 35)2^{2n-7} - 2n + 4)/3$ for all $n \geqslant 3$.

Finally, $s_1 = s_2 = 1$, $t_1 = 0$ and $t_2 = 1$, whereas $s_n = s_{n-1} + s'_n$ and $t_n = t_{n-1} + s_{n-1} + t'_n$ for all $n \geqslant 3$. Thus, an easy induction proves once again that $s_n = (25 \times 2^{2n-3} - 9n^2 + 3n + 7)/27$ and $t_n = ((225n - 290)2^{2n-5} - 9n^3 - 9n^2 + 93n - 77)/81$ whenever $n \geqslant 2$. ◀

5 Open problems and perspectives

The above study of the minimal automaton leaves wide open a few questions, which we intend to explore in follow-up work.

Linear-time recognition algorithm

Precomputing in time $\mathcal{O}(s_n + t_n) = \mathcal{O}(n2^n)$ the automaton $\mathscr{A}_n$ gives us an algorithm that will then detect in time $\mathcal{O}(\ell)$ whether an ℓ-letter word $w \in \mathbb{A}_n^*$ is in alternating normal form. However, when ℓ is small, this precomputation may seem prohibitively costly. Instead, our recursive description of the automata $\mathscr{A}_n$ and $\mathscr{A}'_n$ also provides us with a simple algorithm that will run in time $\mathcal{O}(n\ell)$. Indeed, we can simulate the execution of a path in $\mathscr{A}_n$ as follows:

1. a pointer indicates which is the largest letter σ_{k-1} we have read so far, which means that we are currently reading a word in the automaton $\mathscr{A}'_k$;
2. each state of $\mathscr{A}'_k$ can be represented by a list of the form $s = (p_u^v, \phi_i^?, \mathsf{f}_{i+1}, \phi_{i+1}^?, \ldots, \mathsf{f}_k, \phi_k^?)$, where $p_u^v \in P'_i$, each flag f_j is a boolean $\bot$ or $\top$, and each morphism $\phi_j^?$ is either Id_j or ϕ_j;
3. our recursive description of the transition function δ_n makes it easy to compute in time $\mathcal{O}(n)$ the list that represents the state $\delta_n(s, \sigma_a)$ for all letters $\sigma_a \in \mathbb{A}_n$, provided that this state is well-defined.

Automaticity

Another natural question concerns the automaticity of the alternating normal form, which can be summarised as follows. For each generator $\sigma_i \in \mathbb{A}_n$, we wish to recognise those pairs of words (w, w') representing braids β and β' such that $\beta = \sigma_i \beta'$ (this is *left* automaticity) or $\beta = \beta' \sigma_i$ (this is *right* automaticity). In practice, the words w and w' having distinct lengths, the pair (w, w') shall be represented as a word on the alphabet $(\mathbb{A}_n \cup \{\bullet\})^2$, where $\bullet$ is a padding symbol; *synchronous* automaticity requires that the only padding symbol should be the rightmost symbol of w', and *asynchronous* automaticity allows placing padding symbols wherever we want. It is shown in [5, Proposition 6.10] that the alternating normal form is not asynchronously left-automatic. It might still be right-automatic; this should be the subject of a subsequent article.

Rotating normal form

A close cousin to the alternating normal form is the rotating normal form, already mentioned in the introduction. This normal form is not defined on the standard braid monoid B_n^+ itself, but on the dual braid monoid B_n^{+*} positively generated by the generators $\sigma_{i,j} = (\sigma_{(i+1)\to j})^{-1}\sigma_{i\to j}$. This monoid enjoys properties similar to the properties **a)** to **e)** of page 2. However, there, the Garside element is a braid δ_n for which the inner automorphism $\varphi_n \colon \beta \mapsto \delta_n^{-1}\beta\delta_n$ is not an involution of B_n^{+*} when $n \geqslant 3$, but is of order n.

Local criteria similar to Theorem 10 were found in [11], which help characterising words in rotating normal form and proving that this normal form is regular. Nevertheless, the resulting automaton is not yet guaranteed to be minimal, being analogous to our automaton $\mathscr{A}_n''$ rather than to its minimal variant $\mathscr{A}_n'$. Thus, we intend to replicate our study of the alternating normal form to the rotating normal form, possibly studying its right automaticity as well.

Random generation

In [13], V. Gebhardt and J. González-Meneses focus on the problem of generating uniformly at random a braid $\beta \in B_n^+$ of length $\ell \geqslant 0$. By identifying each braid β with the word $\mathsf{MinLex}_n(\beta)$, they reduce this problem to that of generating a word of length ℓ in a regular language, whose minimal automaton they computed. A crucial step is then to compute the number of paths of length ℓ, in the automaton, that may leave a given state s. Doing so efficiently requires:

- identifying the set of *minimal forbidden patterns*, i.e., the minimal words (for the prefix ordering) that do not label a path leaving the state s;
- applying inclusion-exclusion formulas based on that set;
- using structural properties of that set to perform only polynomially many (and not exponentially many) calls to inclusion-exclusion formulas.

As a result, they obtain an algorithm that generates β in time $\mathcal{O}(n^4 \log(n)\ell^3 \log(\ell))$. It might be possible to adapt this approach to efficiently count paths leaving a state of $\mathscr{A}_n$, thereby obtaining another sampling algorithm, with a similar complexity.

References

1   Sergei Adian. Fragments of the word Δ in the braid group. *Matematicheskie Zametki*, 36(1):25–34, 1984. `doi:10.1007/BF01139549`.

2   Emil Artin. Theorie der Zöpfe. In *Abhandlungen aus dem mathematischen Seminar der Universität Hamburg*, volume 4, pages 47–72. Springer, 1925. `doi:10.1007/BF02950718`.

3   Serge Burckel. The wellordering on positive braids. *Journal of Pure and Applied Algebra*, 120(1):1–17, 1997. `doi:10.1016/S0022-4049(96)00072-2`.

4   Ruth Charney. Artin groups of finite type are biautomatic. *Mathematische Annalen*, 292(1):671–683, 1992. `doi:10.1007/BF01444642`.

5   Patrick Dehornoy. Alternating normal forms for braids and locally Garside monoids. *Journal of pure and applied algebra*, 212(11):2413–2439, 2008. `doi:10.1016/j.jpaa.2008.03.027`.

6   Patrick Dehornoy, François Digne, Eddy Godelle, Daan Krammer, and Jean Michel. *Foundations of Garside theory*, volume 22. Citeseer, 2015. `doi:10.4171/139`.

7   Pierre Deligne. Les immeubles des groupes de tresses généralisés. *Inventiones mathematicae*, 17:273–302, 1972. `doi:10.1007/BF01406236`.

8   David Epstein. *Word processing in groups*. CRC Press, 1992. `doi:10.1201/9781439865699`.

9   Ramón Flores and Juan González-Meneses. On lexicographic representatives in braid monoids. *Journal of Algebraic Combinatorics*, 52(4):561–597, 2020. `doi:10.1007/s10801-019-00913-7`.

10 Ramón Flores and Juan González-Meneses. On the growth of Artin–Tits monoids and the partial theta function. *Journal of Combinatorial Theory, Series A*, 190:105623, 2022. `doi:10.1016/j.jcta.2022.105623`.

11 Jean Fromentin. The rotating normal form of braids is regular. *Journal of Algebra*, 501:545–570, 2018. `doi:10.1016/j.jalgebra.2018.01.001`.

12 Frank Garside. The braid group and other groups. *The Quarterly Journal of Mathematics*, 20(1):235–254, 1969. `doi:10.1093/qmath/20.1.235`.

13 Volker Gebhardt and Juan González-Meneses. Generating random braids. *Journal of Combinatorial Theory, Series A*, 120(1):111–128, 2013. `doi:10.1016/j.jcta.2012.07.003`.

14 Jean Michel. A note on words in braid monoids. *Journal of Algebra*, 215(1):366–377, 1999. `doi:10.1006/jabr.1998.7723`.

Analysis of Regular Sequences: Summatory Functions and Divide-And-Conquer Recurrences

Clemens Heuberger[1] ✉ ⌂ iD
Department of Mathematics, University of Klagenfurt, Austria

Daniel Krenn ✉ ⌂ iD
Fachbereich Mathematik, Paris Lodron University of Salzburg, Austria

Tobias Lechner ✉
Department of Mathematics, University of Klagenfurt, Austria

──── **Abstract** ────

In the asymptotic analysis of regular sequences as defined by Allouche and Shallit, it is usually advisable to study their summatory function because the original sequence has a too fluctuating behaviour. It might be that the process of taking the summatory function has to be repeated if the sequence is fluctuating too much. In this paper we show that for all regular sequences except for some degenerate cases, repeating this process finitely many times leads to a "nice" asymptotic expansion containing periodic fluctuations whose Fourier coefficients can be computed using the results on the asymptotics of the summatory function of regular sequences by the first two authors of this paper.

In a recent paper, Hwang, Janson, and Tsai perform a thorough investigation of divide-and-conquer recurrences. These can be seen as 2-regular sequences. By considering them as the summatory function of their forward difference, the results on the asymptotics of the summatory function of regular sequences become applicable. We thoroughly investigate the case of a polynomial toll function.

1 Introduction

1.1 Overview

The aim of [7] is the study of the asymptotic behaviour of the summatory functions of regular sequences [1] – in simplest terms, a sequence x is called q-regular for some integer $q \geq 2$ if there are square matrices $A_0, \ldots, A_{q-1}$, a row vector u and a column vector w such that for all integers $n \geq 0$,

$$x(n) = u A_{n_0} \ldots A_{n_{\ell-1}} w \tag{1}$$

---

[1] Corresponding author

where $(n_{\ell-1}, \ldots, n_0)$ is the $q$ary expansion of n; an alternative definition will be given in Definition 1. Regular sequences have been introduced by Allouche and Shallit [1]; a plethora of examples have also been given in the same publication. We highlight two prototypical examples at this point: the binary sum of digits function and the worst case number of comparisons in merge sort.

The main result of [7] is that the summary function $N \mapsto \sum_{0 \le n < N} x(n)$ of a q-regular sequence x has an asymptotic expansion

$$\sum_{0 \le n < N} x(n) = \sum_{\substack{\lambda \in \sigma(C) \\ |\lambda| > R}} N^{\log_q \lambda} \sum_{0 \le k < m_C(\lambda)} \frac{(\log N)^k}{k!} \, \Phi_{\lambda k}(\log_q N) + O\big(N^{\log_q R}(\log N)^\kappa\big) \tag{2}$$

as $N \to \infty$, where the $\Phi_{\lambda k}$ are suitable 1-periodic continuous functions and $\sigma(C)$, m_C, R, κ are a set, a function, and two quantities, respectively, depending on the regular sequence and which will be explained in Theorem 3 below. An algorithm is given to compute the Fourier coefficients of the periodic functions. The main question is whether there are $\lambda \in \sigma(C)$ with $|\lambda| > R$. In this case, we say that we established a *good* asymptotic expansion for the summary function of the regular sequence. Otherwise, (2) reduces to an error term. Note that discussing the question of whether the periodic fluctuations vanish is beyond the scope of this paper.

Studying the summary function was motivated by the fact that in several well-known examples of regular sequences, the sequences themselves are fluctuating too much so that it is impossible to establish a good asymptotic expansion for the regular sequence itself. For instance, for the binary sum of digits function s_2, we have $s_2(2^k - 1) = k$ and $s_2(2^k) = 1$ for all integers $k \ge 0$, so the most precise asymptotic expansion for $s_2(n)$ is $s_2(n) = O(\log n)$ for $n \to \infty$. However, the summary function might admit a good asymptotic expansion: For the summary function of the binary sum of digits function, we have $\sum_{0 \le n < N} s_2(n) = \frac{1}{2} N \log_2 N + N\Phi(\log_2 N)$ for some 1-periodic continuous function Φ as $N \to \infty$; see Delange [3]. In this particular example, there is no error term; in general, an error term is to be expected.

However, *a priori*, it is not clear whether the summary function of a regular sequence will be smooth enough so that a good asymptotic expansion can be established. In fact, it is known [1, Theorems 2.6 and 2.5] that the forward difference $n \mapsto x(n+1) - x(n)$ of a regular sequence x is again regular. This means that the summary function of the forward difference of the binary sum of digits function equals the binary sum of digits function and no good asymptotic expansion can be obtained. Thus, as we are able to go forth (summary function) and back (forward difference), the question arises whether for every regular sequence, there is a non-negative integer k such that its k-fold summary function admits a good asymptotic expansion. In this paper, we prove that this is the case for all regular sequences except for some degenerate cases (Theorem 5).

For other regular sequences, the sequence itself might admit a good asymptotic expansion. One example are sequences associated with divide-and-conquer schemes [11, 12], for example, the worst case analysis of the number of comparisons in the merge sort algorithm. These are closely related to the so-called "master theorems"; see the discussion in [12]. These sequences are easily seen [12, Equation (2.1)] to be regular sequences (as long as the toll function is regular). While Hwang, Janson, and Tsai [12] provide a direct proof for the asymptotic behaviour and give plenty of examples, the question is whether these results can also be obtained by using the results in [7]. In the present paper, we see such a sequence as the summary function of its forward difference, and we show that for polynomial toll functions, we get a good asymptotic expansion in the vast majority of cases. The result is formulated in

Theorems 7 and 8. In contrast to [12], we are not constrained to cases where the toll function is asymptotically smaller than the sequence and Fourier coefficients can be computed using the results of [7].

The remaining paper is structured as follows. In Section 1.2, we recall the definition and the relevant results on regular sequences. This is followed in Section 1.3 by the statement of our new result on the k-fold summatory function. In Section 1.4, we present the state of the art for divide-and-conquer sequences and state our version the result in Section 1.5. An explicit example is discussed in Section 1.6. Sections 2 and 3 are devoted to the proofs of our theorems.

1.2 Regular Sequences: Definition and State of the Art

We recall the definition[2] of a *regular sequence*; see Allouche and Shallit [1, 2] for characterisations, properties, and an abundance of examples.

▶ **Definition 1.** *Let $q \geq 2$ be an integer. A sequence $x \in \mathbb{C}^{\mathbb{N}_0}$ is said to be q-regular[3] if there are a non-negative integer D, a family $A = (A_r)_{0 \leq r < q}$ of $D \times D$ matrices over $\mathbb{C}$, a vector $u \in \mathbb{C}^{1 \times D}$ and a vector-valued sequence $v \in (\mathbb{C}^{D \times 1})^{\mathbb{N}_0}$ such that for all $n \in \mathbb{N}_0$, we have*

$$x(n) = uv(n),$$

and such that for all $0 \leq r < q$ and all $n \in \mathbb{N}_0$, we have

$$v(qn + r) = A_r v(n). \tag{3}$$

We call $(u, A, v(0))$ a linear representation *of x and v the* right vector-valued sequence *associated with this linear representation.*

Note that (1) easily follows from (3) by induction; the other direction is contained in [1, Lemma 4.1].

In [7] asymptotic properties were studied. To formulate an abbreviated version of its main result, we first need to recall the notion of the joint spectral radius of a set of square matrices as bounds on matrix products are relevant in view of the representation (1). We fix a vector norm $\| \cdot \|$ on $\mathbb{C}^D$ and consider its induced matrix norm.

▶ **Definition 2.** *Let D be a positive integer and $\mathcal{G}$ be a finite set of $D \times D$ matrices over $\mathbb{C}$.*
1. *The joint spectral radius of $\mathcal{G}$ is defined as*

$$\rho(\mathcal{G}) := \lim_{k \to \infty} \sup\{\|G_1 \dots G_k\|^{1/k} \,|\, G_1, \dots, G_k \in \mathcal{G}\}.$$

2. *We say that $\mathcal{G}$ has the* simple growth property *if*

$$\|G_1 \dots G_k\| = O(\rho(\mathcal{G})^k)$$

holds for all $G_1, \dots, G_k \in \mathcal{G}$ and $k \to \infty$.

For a family $G = (G_i)_{i \in I}$ of $D \times D$ matrices, we set $\rho(G) := \rho(\{G_i \,|\, i \in I\})$ and we say that G has the simple growth property if $\{G_i \,|\, i \in I\}$ has the simple growth property.

---

[2] Strictly speaking, this is an algorithmic characterisation of a regular sequence which is equivalent to the definition given by Allouche and Shallit [1], who first introduced this concept: they define a sequence x to be q-regular if the kernel

$$\left\{ x \circ (n \mapsto q^j n + r) \,\middle|\, j, r \in \mathbb{N}_0 \text{ with } 0 \leq r < q^j \right\}$$

is contained in a finitely generated module.
[3] In the standard literature, the basis is frequently denoted by k instead of our q here.

We note that the joint spectral radius and the simple growth property are independent of the chosen norm; cf. [9, Remark 4.2].

For a square matrix M, let $\sigma(M)$ denote the set of eigenvalues of M and by $m_M(\lambda)$ the size of the largest Jordan block of M associated with some $\lambda \in \mathbb{C}$. In particular, we have $m_M(\lambda) = 0$ if $\lambda \notin \sigma(M)$. Finally, we let $\{z\} := z - \lfloor z \rfloor$ denote the fractional part of a real number z. We use Iverson's convention: For a statement S, we set $[\![S]\!] = 1$ if S is true and 0 otherwise; see also Graham, Knuth and Patashnik [6, p. 24].

▶ **Theorem 3** ([7, Theorem A], [4, 5]). *Let x be a q-regular sequence with linear representation (u, A, w), and set*

$$B_r := \sum_{0 \leq s < r} A_s, \qquad\qquad C := \sum_{0 \leq s < q} A_s$$

for $0 \leq r < q$.

We choose $R > 0$ as follows: If A has the simple growth property, then we set $R = \rho(A)$. Otherwise, we choose $R > \rho(A)$ such that there is no eigenvalue $\lambda \in \sigma(C)$ with $\rho(A) < |\lambda| \leq R$.

Then we have

$$\sum_{0 \leq n < N} x(n) = \sum_{\substack{\lambda \in \sigma(C) \\ |\lambda| > R}} N^{\log_q \lambda} \sum_{0 \leq k < m_C(\lambda)} \frac{(\log N)^k}{k!} \, \Phi_{\lambda k}(\{\log_q N\})$$

$$+ O\big(N^{\log_q R}(\log N)^{\max\{m_C(\lambda)\,:\,|\lambda|=R\}}\big) \quad (4)$$

as $N \to \infty$, where $\Phi_{\lambda k}$ are suitable 1-periodic functions. If there are no eigenvalues $\lambda \in \sigma(C)$ with $|\lambda| \leq R$, the O-term can be omitted.

For $|\lambda| > R$ and $0 \leq k < m_C(\lambda)$, the function $\Phi_{\lambda k}$ is Hölder continuous with any exponent smaller than $\log_q(|\lambda|/R)$.

Note that [7] also contains results on how to compute the Fourier coefficients of the periodic fluctuations $\Phi_{\lambda k}$.

1.3   Summatory Functions of Regular Sequences

As announced in Section 1.1, within this paper, we show that for all regular sequences except for some degenerate cases, there is a non-negative integer k such that the k-fold summatory function admits a good asymptotic expansion. In order to formulate our result, we first fix a notation for summatory functions.

▶ **Definition 4.** *For a sequence $x \colon \mathbb{N}_0 \to \mathbb{C}^D$ (for some positive integer D), define the sequence $\Sigma x \colon \mathbb{N}_0 \to \mathbb{C}^D$ by*

$$(\Sigma x)(N) = \sum_{0 \leq n < N} x(n).$$

We use the convention that Σ binds more strongly than evaluation, i.e., we write $\Sigma x(N)$ instead of $(\Sigma x)(N)$.

We are now able to formulate our result.

▶ **Theorem 5.** *Let x be a q-regular sequence with linear representation (u, A, w) and set $C := \sum_{0 \leq r < q} A_r$. Assume that C has a non-zero eigenvalue.*

Then there is a non-negative integer k such that $\Sigma^k x$ admits a good asymptotic expansion.

This theorem is proved in Section 2.

1.4 Divide-and-Conquer Sequences: Definition and State of the Art

Hwang, Janson, and Tsai [12] study sequences x with

$$x(n) = \alpha x\left(\left\lfloor \frac{n}{2} \right\rfloor\right) + \beta x\left(\left\lceil \frac{n}{2} \right\rceil\right) + g(n) \tag{5}$$

for $n \geq 2$, where α and β are two given positive constants, g is a given function, called the toll function, and $x(1)$ is given.

The simplest version of their result is summarised in the following theorem; more general (weaker assumptions on g) versions are also available.

▶ **Theorem 6** ([12, Corollary 2.14]). *Let x be a sequence satisfying* (5). *Assume that there is an $\varepsilon > 0$ such that $g(n) = O(n^{\log_2(\alpha+\beta)-\varepsilon})$. Then*

$$x(n) = n^{\log_2(\alpha+\beta)}\Phi(\{\log_2 n\}) + O(n^{\log_2(\alpha+\beta)-\varepsilon})$$

for $n \to \infty$ where Φ is a continuous, 1-periodic function.

1.5 Divide-and-Conquer Sequences: Polynomial Toll Function

For divide-and-conquer sequences g, a sequence satisfying the recurrence (5) can be seen as a regular sequence: It is not hard to see that we have

$$
\begin{aligned}
x(2n) &= (\alpha + \beta)x(n) + g(2n), \\
x(2n+1) &= \alpha x(n) + \beta x(n+1) + g(2n+1)
\end{aligned}
\tag{6}
$$

for $n \geq 1$. Thus x is a 2-recursive sequence in the sense of [9] and therefore 2-regular by [9, Corollary D]. Alternatively, a linear representation for x can also be constructed directly from (6): the associated right vector-valued sequence consists of $n \mapsto x(n)$, $n \mapsto x(n+1)$ and the right vector-valued sequence associated to a linear representation of g; the matrices of the linear representation can then easily be reconstructed from (6) and the linear representation of g. The fact that (6) holds only for $n \geq 1$ (instead of $n \geq 0$) can be fixed; see [1, Proof of Lemma 4.1] or [9, Theorem B].

As announced in Section 1.1, our goal is to see what can be said about the asymptotics of $x(n)$ for $n \to \infty$ using Theorem 3. While the method works for arbitrary regular toll functions; see Remark 14 (although good asymptotic expansions cannot be guaranteed in all cases), we formulate our main result for polynomial toll functions; first versions are contained in the master's thesis [14] of the third author.

▶ **Theorem 7.** *Let $g(n) = \sum_{i=0}^{k} c_i n^i$ be a polynomial of degree $k \geq 1$, x be a sequence satisfying* (5). *Then the asymptotic behaviour of $x(n)$ for $n \to \infty$ can be described as follows, where Φ and Ψ are 1-periodic continuous functions.*

- **Case 1a.** *If $\alpha + \beta > 2^k$ and $2^k > \max\{\alpha, \beta\}$, then*

$$x(n) = n^{\log_2(\alpha+\beta)}\Phi(\{\log_2 n\}) + n^k \Psi(\{\log_2 n\}) + O(n^{\log_2 \max\{\alpha,\beta\}}).$$

- **Case 1b.** *If $\alpha + \beta > 2^k$ and $\max\{\alpha, \beta\} \geq 2^k$, then*

$$x(n) = n^{\log_2(\alpha+\beta)}\Phi(\{\log_2 n\}) + O(n^{\log_2 \max\{\alpha,\beta\}}(\log n)^{[\![\max\{\alpha,\beta\}=2^k]\!]}).$$

- **Case 2.** *If $\alpha + \beta = 2^k$, then*

$$x(n) = n^k(\log n)\Phi(\{\log_2 n\}) + n^k \Psi(\{\log_2 n\}) + O(n^{\log_2 \max\{\alpha,\beta\}+[\![\alpha=\beta]\!]\varepsilon})$$

for any $\varepsilon > 0$.

- **Case 3.** *If $2^k > \alpha + \beta > 2^{k-1}$, then*

$$x(n) = n^k \Phi(\{\log_2 n\}) + n^{\log_2(\alpha+\beta)} \Psi(\{\log_2 n\})$$
$$+ O(n^{\log_2 \max\{\alpha,\beta,2^{k-1}\} + [\![\max\{\alpha,\beta\}=2^{k-1}]\!]\varepsilon} (\log n)^{[\![\max\{\alpha,\beta\}<2^{k-1}]\!]})$$

for any $\varepsilon > 0$.
- **Case 4.** *If $2^{k-1} \geq \alpha + \beta$, then*

$$x(n) = n^k \Phi(\{\log_2 n\}) + O(n^{k-1}(\log n)^E),$$

where

$$E := 1 + [\![\alpha + \beta = 2^{k-1}]\!]([\![k \geq 2 \ and \ c_{k-1} \neq 0]\!] + [\![k = 1 \ and \ d_0 + d_1 \neq 0]\!])$$

with

$$d_0 := (1 - \beta)x(1) - g(1) + g(0), \qquad\qquad d_1 := g(1) - (1 - \beta)x(1).$$

This theorem is proved in Section 3.

The case of a constant toll function is somewhat simpler.

▶ **Theorem 8.** *Let $g(n) = c_0$ be a constant toll function and let x be a sequence satisfying* (5). *Let d_0 and d_1 be defined as in Theorem 7. Then the asymptotic behaviour of $x(n)$ for $n \to \infty$ can be described as follows, where Φ is a 1-periodic continuous function.*
- **Case 1.** *If $d_0 = d_1 = 0$, then*

$$x(n) = n^{\log_2(\alpha+\beta)} \Phi(\{\log_2 n\}).$$

- **Case 2a.** *If $d_0 \neq 0$ or $d_1 \neq 0$, and $\alpha + \beta > 1$, then*

$$x(n) = n^{\log_2(\alpha+\beta)} \Phi(\{\log_2 n\})$$
$$+ O(n^{\log_2 \max\{\alpha,\beta,1\} + [\![\max\{\alpha,\beta\}=1]\!]\varepsilon} (\log n)^{[\![\max\{\alpha,\beta\}<1]\!]})$$

for any $\varepsilon > 0$.
- **Case 2b.** *If $d_0 \neq 0$ or $d_1 \neq 0$, and $\alpha + \beta \leq 1$, then*

$$x(n) = O((\log n)^{[\![\alpha+\beta=1 \ and \ d_0+d_1 \neq 0]\!]}).$$

This theorem is also proved in Section 3.

1.6 Example

We conclude this introductory section with one example to illustrate the results.

▶ **Example 9.** Consider the divide-and-conquer algorithm for finding the minimum and the maximum of a list of n elements. The number $x(n)$ of comparisons needed satisfies (5) for $n \geq 3$ with $a = b = 1$ and $g(n) = 2$ for $n \geq 3$ and with $x(1) = 0$ and $x(2) = 1$; cf. [11, Example 3.2].

By Theorem 8 (and [9, Theorem B] to deal with the fact that the divide-and-conquer recurrence is only valid for $n \geq 3$ instead of $n \geq 2$; see Section 4 for details), we get

$$x(n) = n\Phi(\{\log_2 n\}) + O(n^\varepsilon)$$

for some 1-periodic continuous function Φ and any $\varepsilon > 0$. The Fourier coefficients of Φ can be computed; cf. Figure 1.
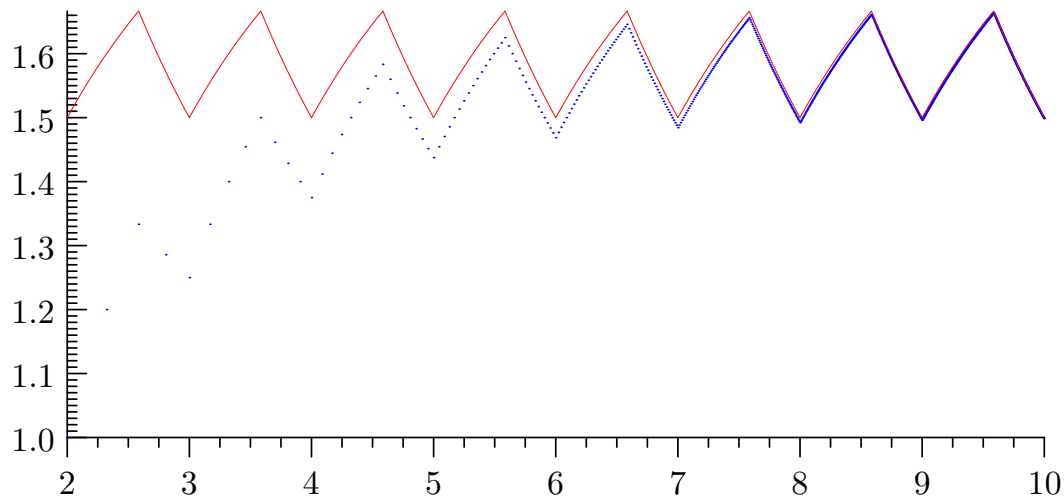
■ **Figure 1** Comparison of the 1-periodic function Φ in the asymptotic expansion determined using Theorem 8 with the empirical values of the sequence.

2　Summatory Functions: Proof of Theorem 5

Before proving Theorem 5, we collect two lemmata on the linear representations of summatory functions and k-fold summatory functions.

The following lemma is implicitly shown in [7, Lemma 12.2], however, it is crucial for our purposes, so we provide a precise formulation and will prove it for self-containedness.

▶ **Lemma 10.** *Let x be a q-regular sequence with linear representation (u, A, w) and associated right vector-valued sequence v. Set*

$$B_r := \sum_{0 \le s < r} A_s \text{ for } 0 \le r < q \text{ and } C := \sum_{0 \le s < q} A_s. \tag{7}$$

Then we have

$$\Sigma v(qN + r) = C \, \Sigma v(N) + B_r v(N) \tag{8}$$

for all $N \ge 0$ and $0 \le r < q$.

Additionally, Σx is regular with linear representation $(\widetilde{u}, \widetilde{A}, \widetilde{w})$ with

$$\widetilde{u} := (u, 0),$$

$$\widetilde{A}_r := \begin{pmatrix} C & B_r \\ 0 & A_r \end{pmatrix} \text{ for } 0 \le r < q,$$

$$\widetilde{w} := \begin{pmatrix} 0 \\ w \end{pmatrix};$$

the associated right vector-valued sequence is $\begin{pmatrix} \Sigma v \\ v \end{pmatrix}$.

Proof. By definition, we have $\Sigma v(0) = 0$. Let $N \ge 0$ and $0 \le r < q$. Then

$$\Sigma v(qN + r) = \sum_{0 \le n < qN} v(n) + \sum_{qN \le n < qN + r} v(n).$$

Replacing n by $qm + s$ for $m \in \mathbb{Z}$ and $0 \le s < q$ in the first sum and replacing n by $qN + s$ for $0 \le s < r$ in the second sum yields

$$\Sigma v(qN + r) = \sum_{0 \le m < N} \sum_{0 \le s < q} v(qm + s) + \sum_{0 \le s < r} v(qN + s).$$

Using the linear representation yields

$$\Sigma v(qN + r) = \sum_{0 \le m < N} \sum_{0 \le s < q} A_s v(m) + \sum_{0 \le s < r} A_s v(N) = C \, \Sigma v(N) + B_r v(N).$$

In other words, we have shown (8).

As we have $x = uv$, we also have

$$\Sigma x = u \, \Sigma v = u \, \Sigma v + 0 = \widetilde{u} \begin{pmatrix} \Sigma v \\ v \end{pmatrix}.$$

We conclude that Σx has the given linear representation and associated right vector-valued sequence. ◀

▶ **Remark 11.** In [7, Lemma 12.2], a very similar result appears in Equation (12.1) there. The difference between Lemma 10 and that equation is an additional summand $(I - A_0)[\![qN + r > 0]\!]$.

The reason for the additional summand is that in general, $f(0) = A_0 f(0)$ (with the notations there) does not hold, cf. also [10] for a discussion of this condition.

Iterating the results in Lemma 10 leads to the following lemma.

▶ **Lemma 12.** *Let x be a q-regular sequence with linear representation (u, A, w), $k \ge 1$, and use the notations from (7). Then $\Sigma^k x$ is q-regular with linear representation $(\widetilde{u}, \widetilde{A}, \widetilde{w})$ where $\widetilde{A}_r$ is a block upper triangular matrix with diagonal blocks $q^{k-1}C$, $q^{k-2}C$, ..., qC, C, A_r for $0 \le r < q$ and $\widetilde{u}$ and $\widetilde{w}$ are vectors; the associated right vector-valued sequence is $\left(\Sigma^k v, \Sigma^{k-1} v, \dots, \Sigma v, v\right)^\top$.*

Proof. We claim that for $m \ge 1$, there are matrices $M_{m,0}$, ..., $M_{m,m-1}$ such that

$$\Sigma^m v(qN + r) = q^{m-1} C \, \Sigma^m v(N) + \sum_{0 \le j < m} M_{m,j} \Sigma^j v(N) \tag{9}$$

holds for all $N \ge 0$ and $0 \le r < q$.

We show (9) by induction on m. For $m = 1$, this is (8). To show (9) for m replaced by $m + 1$, we use (8) for v replaced by the regular sequence with associated right vector-valued sequence $\left(\begin{smallmatrix} \Sigma v \\ v \end{smallmatrix}\right)$ studied in Lemma 10 and the linear representation given there. We obtain

$$\Sigma^m \begin{pmatrix} \Sigma v \\ v \end{pmatrix}(qN + r) = q^{m-1} \begin{pmatrix} qC & \sum_{0 \le r < q} B_r \\ 0 & C \end{pmatrix} \Sigma^m \begin{pmatrix} \Sigma v \\ v \end{pmatrix}(N) + \sum_{0 \le j < m} \widetilde{M}_{m,j} \Sigma^j \begin{pmatrix} \Sigma v \\ v \end{pmatrix}(N)$$

for suitable matrices $\widetilde{M}_{m,j}$ for $0 \le j < m$. Considering the first block row of this equation and collecting terms by powers of Σ leads to (9) with m replaced by $m + 1$.

Using (9) for $1 \le m \le k$ yields the linear representation as described in the lemma. ◀

Proof of Theorem 5. Let ρ be the joint spectral radius of A and r the spectral radius (largest absolute value of an eigenvalue) of C.

For some fixed k which will be chosen appropriately later, Lemma 12 yields a linear representation of $\Sigma^k x$ with the properties given there. Let $\widetilde{C} := \sum_{0 \le r < q} \widetilde{A}_r$.

The k-fold summatory function $\Sigma^k x$ admits a good asymptotic expansion if the spectral radius of $\widetilde{C}$ is larger than the joint spectral radius of $\widetilde{A}$. So we compute both.

By Lemma 12, $\widetilde{C}$ is a block upper triangular matrix with diagonal blocks $q^k C, \ldots, qC,$ C. So the spectral radius of $\widetilde{C}$ is $q^k r$.

The joint spectral radius of a family of block upper triangular matrices is the maximum of the joint spectral radii of the diagonal blocks; see [13, Proposition 1.5]. This implies that the joint spectral radius of $\widetilde{A}$ is $\max\{q^{k-1}r, q^{k-2}r, \ldots, qr, r, \rho\} = \max\{q^{k-1}r, \rho\}$.

It is clear that $q^k r > q^{k-1} r$ so the only condition which needs to be satisfied is $q^k r > \rho$. Such a k exists because $r > 0$ (as C has a non-zero eigenvalue). ◀

3 Divide-and-Conquer Recurrences: Proof of Theorems 7 and 8

For the proof of Theorems 7 and 8, we will first consider a general regular toll function g and summarise our findings in the general case in Remark 14. Afterwards, we will specialise to a polynomial toll function.

As announced in Section 1.1, we write x as the summatory function of the forward difference of x, i.e.,

$$x(N) = \sum_{0 \le n < N} (x(n+1) - x(n)) + x(0)$$

for $N \ge 0$. Note that strictly speaking, $x(0)$ is not defined in (5). However, we may assume that $g(1)$ and $g(0)$ are somehow defined: they are not used in (5), but we can extend the definition of g if $g(0)$ and $g(1)$ should be undefined.

In order to use Theorem 3, we need a linear representation of the forward difference of x. A first step is the following lemma.

▶ **Lemma 13.** *Let x be a sequence satisfying* (5) *for some toll function g and set $x(0) := 0$ and $h(n) := x(n+1) - x(n)$ for $n \ge 0$. Then*

$$\begin{aligned}
h(2n) &= \beta h(n) + g(2n+1) - g(2n) + d_0 \delta_0(n), \\
h(2n+1) &= \alpha h(n) + g(2n+2) - g(2n+1) + d_1 \delta_0(n)
\end{aligned} \tag{10}$$

for $n \ge 0$ with d_0, d_1 as in Theorem 7 and $\delta_0(n) := [\![n = 0]\!]$ for $n \ge 0$.

Proof. We can rewrite (6) as

$$\begin{aligned}
x(2n) &= (\alpha + \beta)x(n) + g(2n) - g(0)\delta_0(n), \\
x(2n+1) &= \alpha x(n) + \beta x(n+1) + g(2n+1) + ((1-\beta)x(1) - g(1))\delta_0(n)
\end{aligned} \tag{11}$$

for $n \ge 0$.

Then (10) follows from

$$\begin{aligned}
h(2n) &= x(2n+1) - x(2n), \\
h(2n+1) &= x(2n+2) - x(2n+1)
\end{aligned}$$

and inserting (11) into these equations. ◀

▶ **Remark 14.** Note that h occurs only as $h(n)$ on the right-hand side of of (10). Therefore, a right vector-valued sequence associated with h can be constructed by using $h(n)$ in its first component, then whatever is needed to express $g(2n + 2) - g(2n + 1)$ and $g(2n + 1) - g(2n)$, followed by $\delta_0(n)$. The matrices A_0 and A_1 of the linear representation will thus be in block triangular form; the upper left diagonal elements of A_0 and A_1 being β and α, respectively.

Assume that the contributions of g to the linear representation are small in comparison to α and β. Then the joint spectral radius of the linear representation of h will be $\max\{\alpha, \beta\}$, whereas the matrix C as in Theorem 3 will have a dominant eigenvalue $\alpha + \beta$, which is larger than the joint spectral radius. Thus we will have a good asymptotic expansion in this case.

We now turn to the case of a polynomial toll function.

▶ **Lemma 15.** *Let x and h be as in Lemma 13 with a polynomial toll function $g(n) = \sum_{i=0}^{k} c_i n^i$ for some $k \geq 0$ and some constants $c_0, \ldots, c_k$ with $c_k \neq 0$.*
Set

$$b_{0j} := \sum_{i=j+1}^{k} \binom{i}{j} 2^j c_i, \qquad\qquad b_{1j} := \sum_{i=j+1}^{k} \binom{i}{j}(2^i - 2^j) c_i,$$

$$a_{0ij} := [\![j = i]\!] 2^i, \qquad\qquad a_{1ij} := \binom{i}{j} 2^j$$

for $0 \leq i < k$ and $0 \leq j < k$ and

$$b_r := (b_{r(k-1)}, \ldots, b_{r0}), \qquad\qquad \widetilde{A}_r := (a_{rij})_{\substack{i=k-1,\ldots,0, \\ j=k-1,\ldots,0}},$$

$$\mu_r := \begin{cases} \beta & \text{if } r = 0, \\ \alpha & \text{if } r = 1, \end{cases} \qquad\qquad A_r := \begin{pmatrix} \mu_r & b_r & d_r \\ 0 & \widetilde{A}_r & 0 \\ 0 & 0 & [\![r = 0]\!] \end{pmatrix}$$

for $r \in \{0, 1\}$ and

$$u := (1, 0, \ldots, 0) \in \mathbb{C}^{1 \times (k+2)}, \qquad w := \begin{cases} (x(1), 0, \ldots, 0, 1, 1)^\top \in \mathbb{C}^{(k+2) \times 1} & \text{if } k \geq 1, \\ (x(1), 1)^\top \in \mathbb{C}^{2 \times 1} & \text{if } k = 0. \end{cases}$$

Then (u, A, w) is a linear representation for h.
If $d_0 = d_1 = 0$, then $(\widetilde{u}, \widetilde{A}, \widetilde{w})$ is also a linear representation for h where $\widetilde{u}$, $\widetilde{A}$, and $\widetilde{w}$ arise from from u, A, and w by removing the last column, the last row and column, and the last row, respectively.

We remark that A_r is an upper triangular matrix for $r \in \{0, 1\}$ because $\binom{i}{j} = 0$ for $j > i$ (and indices in $\widetilde{A}_0$ and $\widetilde{A}_1$ are decreasing).

Proof. As a right vector-valued sequence v, we choose

$$n \mapsto (h(n), n^{k-1}, \ldots, 1, \delta_0(n))^\top.$$

We immediately check that $v(0) = w$ and that $uv(n) = h(n)$ holds for all $n \geq 0$.

Using the binomial theorem repeatedly, we get

$$g(2n+1) - g(2n) = \sum_{i=0}^{k} c_i \sum_{j=0}^{i-1} \binom{i}{j} 2^j n^j \qquad\qquad = \sum_{j=0}^{k-1} b_{0j} n^j,$$

$$g(2n+2) - g(2n+1) = \sum_{i=0}^{k} c_i \sum_{j=0}^{i-1} \binom{i}{j} 2^j n^j (2^{i-j} - 1) = \sum_{j=0}^{k-1} b_{1j} n^j,$$

$$(2n)^i = 2^i n^i \qquad\qquad = \sum_{j=0}^{k-1} a_{0ij} n^j \qquad \text{for } 0 \le i < k,$$

$$(2n+1)^i = \sum_{j=0}^{i} \binom{i}{j} 2^j n^j \qquad\qquad = \sum_{j=0}^{k-1} a_{1ij} n^j \qquad \text{for } 0 \le i < k.$$

We verify that (10) translates into

$$v(2n + r) = A_r v(n)$$

for $r \in \{0, 1\}$ and $n \ge 0$.

If $d_0 = d_1 = 0$, then A_0 and A_1 are block diagonal matrices. The lower right block is not taken into account when multiplying by u, so the lower right block can be omitted.

Thus the result follows.   ◀

▶ **Lemma 16.** *Let x and h be as in Lemma 13 and g, (u, A, w) be as in Lemma 15. Assume that $k \ge 1$. Set $C = A_0 + A_1$.*

Then $\rho(A) = \max\{\alpha, \beta, 2^{k-1}\}$ and $\sigma(C) = \{\alpha + \beta, 2^k, 2^{k-1}, \dots, 2, 1\}$. If $\max\{\alpha, \beta\} \ne 2^{k-1}$, then A has the simple growth property.

Proof. By Lemma 15, A_0 is an upper triangular matrix with diagonal elements β, 2^{k-1}, ..., 1, 1 and A_1 is an upper triangular matrix with diagonal elements α, 2^{k-1}, ..., 1, 0.

The joint spectral radius of a set of upper triangular matrices is the maximum of the diagonal elements of the matrices; see [13, Proposition 1.5]. This implies that $\rho(A) = \max\{\alpha, \beta, 2^{k-1}\}$. By [9, Lemma 4.5], A has the simple growth property if the joint spectral radius of A occurs only once as a maximum of corresponding diagonal elements of A_0 and A_1. This is the case if $\max\{\alpha, \beta\} \ne 2^{k-1}$.

We also conclude that C is an upper triangular matrix with diagonal elements $\alpha + \beta$, 2^k, 2^{k-1}, ..., 2, 1. Thus the assertion for $\sigma(C)$ follows.   ◀

Proof of Theorem 7. To prove Theorem 7 using Theorem 3, we need to determine the eigenvalues of C which are greater or equal than the joint spectral radius of A_0 and A_1 (with the notations of Lemma 16) and the size of the largest Jordan block associated with any such eigenvalue.

As $\rho(A) = \max\{\alpha, \beta, 2^{k-1}\} \ge 2^{k-1} \ge 1$, it is clear that the only relevant eigenvalues of C are contained in the set $\{\alpha + \beta, 2^k, 2^{k-1}\}$.

The main case distinction of Theorem 7 concerns the order of $\alpha + \beta$, 2^k, and 2^{k-1}.

▪ **Case 1:** $\alpha + \beta > 2^k$. This implies that $\max\{\alpha, \beta\} > 2^{k-1}$ and therefore $\rho(A) = \max\{\alpha, \beta\}$ and A has the simple growth property. The eigenvalue $\alpha + \beta$ of C is larger than the joint spectral radius and is a simple eigenvalue. The eigenvalue 2^k is also a simple eigenvalue. If it is larger than the joint spectral radius, we are in Case 1a and have two asymptotic terms larger than the error term. If $2^k \le \max\{\alpha, \beta\}$, we are in Case 1b and have one asymptotic term larger than the error term; there is a logarithmic factor in the error term if and only if 2^k equals the joint spectral radius.

- **Case 2:** $\alpha + \beta = 2^k$. In this case, $2^k = \alpha + \beta$ has algebraic multiplicity 2 as an eigenvalue of C. We note that $C - 2^k I$ has the shape

$$C - 2^k I = \begin{pmatrix} 2^k - 2^k & \binom{k}{k-1} 2^k c_k & b' \\ 0 & 2^k - 2^k & b'' \\ 0 & 0 & C' \end{pmatrix} = \begin{pmatrix} 0 & k 2^k c_k & b' \\ 0 & 0 & b'' \\ 0 & 0 & C' \end{pmatrix}$$

for some vectors b', b'' and some upper triangular matrix C' with diagonal elements $2^{k-1} - 2^k$, $2^{k-2} - 2^k$, ..., $2^0 - 2^k$. As $c_k \neq 0$ by assumption, the kernel of $C - 2^k I$ has dimension 1. We conclude that the size $m_C(2^k)$ of the largest Jordan block of C associated with the eigenvalue 2^k equals 2. So we have a logarithmic factor in the asymptotic main term.

We also note that $\max\{\alpha, \beta\} \geq 2^{k-1}$ holds in this case with equality if and only if $\alpha = \beta = 2^{k-1}$. So the joint spectral radius $\rho(A)$ equals $\max\{\alpha, \beta\}$ and A has the simple growth property unless $\alpha = \beta = 2^{k-1}$.

- **Case 3:** $2^k > \alpha + \beta > 2^{k-1}$. In this case, we have $\max\{\alpha, \beta\} > 2^{k-2}$ and C has two simple dominant eigenvalues 2^k and $\alpha + \beta$. We do not have additional information about the joint spectral radius. If $\max\{\alpha, \beta\} \neq 2^{k-1}$, then A has the simple growth property. If $\max\{\alpha, \beta\} < 2^{k-1}$, then the joint spectral radius of A is 2^{k-1}. As C has 2^{k-1} as an eigenvalue as well, there is a logarithmic factor in the error term in exactly this situation.

- **Case 4:** $2^{k-1} \geq \alpha + \beta$. In this case, we have $\max\{\alpha, \beta\} < 2^{k-1}$, so $\rho(A) = 2^{k-1}$ and A has the simple growth property. There is only one eigenvalue of C larger than this joint spectral radius, namely 2^k. We have to determine $m_C(2^{k-1})$ in order to find out the exponent of $\log n$ in the error term. The algebraic multiplicity of 2^{k-1} as an eigenvalue of C equals $1 + [\![\alpha + \beta = 2^{k-1}]\!]$. So if $\alpha + \beta < 2^{k-1}$, we have $m_C(2^{k-1}) = 1$ and a factor $\log n$ in the error term.

We now consider the case $\alpha + \beta = 2^{k-1}$ and $k \geq 2$. We note that $C - 2^{k-1} I$ has the shape

$$C - 2^{k-1} I = \begin{pmatrix} 2^{k-1} - 2^{k-1} & \binom{k}{k-1} 2^k c_k & \binom{k-1}{k-2} 2^{k-1} c_{k-1} + \binom{k}{k-2} 2^k c_k & b' \\ 0 & 2^k - 2^{k-1} & \binom{k-1}{k-2} 2^{k-2} & b'' \\ 0 & 0 & 2^{k-1} - 2^{k-1} & b''' \\ 0 & 0 & 0 & C' \end{pmatrix}$$

$$= \begin{pmatrix} 0 & k 2^k c_k & (k-1) 2^{k-1} c_{k-1} + k(k-1) 2^{k-1} c_k & b' \\ 0 & 2^{k-1} & (k-1) 2^{k-2} & b'' \\ 0 & 0 & 0 & b''' \\ 0 & 0 & 0 & C' \end{pmatrix}$$

for suitable vectors b', b'', b''' and a regular upper triangular matrix C'. Subtracting $2k c_k$ times the second row from the first row does not change the kernel, so we get

$$\ker(C - 2^{k-1} I) = \ker \begin{pmatrix} 0 & 0 & (k-1) 2^{k-1} c_{k-1} & b' - 2k c_k b'' \\ 0 & 2^{k-1} & (k-1) 2^{k-2} & b'' \\ 0 & 0 & 0 & b''' \\ 0 & 0 & 0 & C' \end{pmatrix}.$$

We conclude that $\dim \ker(C - 2^{k-1} I) = 1 + [\![c_{k-1} = 0]\!]$ and therefore $m_C(2^{k-1}) = 1 + [\![c_{k-1} \neq 0]\!]$.

Finally we turn to the case that $\alpha + \beta = 2^{k-1}$ and $k = 1$. Then we have

$$C - I = \begin{pmatrix} 0 & 2c_1 & d_0 + d_1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

If $d_0 = d_1 = 0$, by the last statement of Lemma 15, the last row and column of $C - I$ are omitted and 1 has algebraic multiplicity 1 as an eigenvalue of C. We conclude that $\dim \ker C - I = 1 + [\![d_0 + d_1 = 0]\!] - [\![d_0 = 0]\!][\![d_1 = 0]\!]$ and therefore $m_C(2^{k-1}) = 1 + [\![d_0 + d_1 \neq 0]\!]$.

So, to summarise, $m_C(2^{k-1}) = 1 + E$ where E is defined in Theorem 7. ◄

Proof of Theorem 8. If $d_0 = d_1 = 0$, then Lemma 15 yields $A_0 = (\beta)$, $A_1 = (\alpha)$, and $C = (\alpha + \beta)$, so the joint spectral radius of A equals $\max\{\alpha, \beta\}$ which is strictly less than the unique eigenvalue $\alpha + \beta$ of C. As there is no eigenvalue of C less than the joint spectral radius of A, there is no error term. The result follows in this case.

From now on, we assume that $d_0 \neq 0$ or $d_1 \neq 0$. Lemma 15 yields

$$A_0 = \begin{pmatrix} \beta & d_0 \\ 0 & 1 \end{pmatrix}, \qquad A_1 = \begin{pmatrix} \alpha & d_1 \\ 0 & 0 \end{pmatrix}, \qquad C = \begin{pmatrix} \alpha + \beta & d_0 + d_1 \\ 0 & 1 \end{pmatrix}.$$

We see that the joint spectral radius of A is $\max\{\alpha, \beta, 1\}$ and that C has eigenvalues $\alpha + \beta$ and 1. It is now easy to deduce the assertions of the theorem. ◄

4 Details on Example 9

We proceed as outlined in Remark 14. Setting $x(0) = 0$ as usual, we have

$$x(n) = x(\lfloor n/2 \rfloor) + x(\lceil n/2 \rceil) + 2 - [\![n = 2]\!] - 2[\![n = 1]\!] - 2[\![n = 0]\!]$$

for $n \geq 0$. Equivalently, we have

$$x(2n) = 2x(n) + 2 - [\![n = 1]\!] - 2[\![n = 0]\!],$$
$$x(2n + 1) = x(n) + x(n + 1) + 2 - 2[\![n = 0]\!]$$

for $n \geq 0$. Setting $h(n) := x(n + 1) - x(n)$ leads to

$$h(2n) = h(n) + [\![n = 1]\!],$$
$$h(2n + 1) = h(n) + [\![n = 0]\!]$$

for $n \geq 0$. This defines a 2-regular sequence with a linear representation (u, A, w) with associated right vector-valued sequence v defined by $v(n) = (h(n), \delta_1(n), \delta_0(n))$ with

$$A_0 = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad\qquad A_1 = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix},$$
$$u = (1, 0, 0), \qquad\qquad w = (0, 0, 1)^\top.$$

Here, δ_1 is defined by $\delta_1(n) := [\![n = 1]\!]$ for $n \geq 0$.

We can now use SageMath[4] to compute Fourier coefficients and to produce Figure 1.

---

[4] The code for this example is available at `https://gitlab.com/cheuberg/fluctuation-find-min-max` (mirrored at `https://arxiv.org/src/2403.06589/anc`); it uses the code accompanying [7] which is available at at `https://gitlab.com/dakrenn/regular-sequence-fluctuations`.

References

1   Jean-Paul Allouche and Jeffrey Shallit. The ring of k-regular sequences. *Theoret. Comput. Sci.*, 98(2):163–197, 1992. `doi:10.1016/0304-3975(92)90001-V`.

2   Jean-Paul Allouche and Jeffrey Shallit. *Automatic sequences: Theory, applications, generalizations.* Cambridge University Press, Cambridge, 2003. `doi:10.1017/CBO9780511546563`.

3   Hubert Delange. Sur la fonction sommatoire de la fonction "somme des chiffres". *Enseign. Math. (2)*, 21:31–47, 1975. `doi:10.5169/seals-47328`.

4   Philippe Dumas. Joint spectral radius, dilation equations, and asymptotic behavior of radix-rational sequences. *Linear Algebra Appl.*, 438(5):2107–2126, 2013. `doi:10.1016/j.laa.2012.10.013`.

5   Philippe Dumas. Asymptotic expansions for linear homogeneous divide-and-conquer recurrences: Algebraic and analytic approaches collated. *Theoret. Comput. Sci.*, 548:25–53, 2014. `doi:10.1016/j.tcs.2014.06.036`.

6   Ronald L. Graham, Donald E. Knuth, and Oren Patashnik. *Concrete mathematics. A foundation for computer science.* Addison-Wesley, second edition, 1994.

7   Clemens Heuberger and Daniel Krenn. Asymptotic analysis of regular sequences. *Algorithmica*, 82(3):429–508, 2020. `doi:10.1007/s00453-019-00631-3`.

8   Clemens Heuberger, Daniel Krenn, and Tobias Lechner. Fluctuation Find Min and Max. Software, version 1.0., Austrian Science Fund (FWF) [10.55776/DOC78], swhId: `swh:1:dir:e3d6789813ee435280117108c7bfd47809aeecc9` (visited on 2024-07-10). URL: `https://gitlab.com/cheuberg/fluctuation-find-min-max`.

9   Clemens Heuberger, Daniel Krenn, and Gabriel F. Lipnik. Asymptotic analysis of q-recursive sequences. *Algorithmica*, 84(9):2480–2532, 2022. `doi:10.1007/s00453-022-00950-y`.

10   Clemens Heuberger, Daniel Krenn, and Gabriel F. Lipnik. A note on the relation between recognisable series and regular sequences, and their minimal linear representations. *J. Symbolic Comput.*, 2024. `doi:10.1016/j.jsc.2023.102295`.

11   Hsien-Kuei Hwang, Svante Janson, and Tsung-Hsi Tsai. Exact and asymptotic solutions of a divide-and-conquer recurrence dividing at half: Theory and applications. *ACM Trans. Algorithms*, 13(4):Art. 47, 43 pp., October 2017. `doi:10.1145/3127585`.

12   Hsien-Kuei Hwang, Svante Janson, and Tsung-Hsi Tsai. Identities and periodic oscillations of divide-and-conquer recurrences splitting at half. *Adv. in Appl. Math.*, 155:Paper No. 102653, 53, 2024. `doi:10.1016/j.aam.2023.102653`.

13   Raphaël Jungers. *The joint spectral radius. Theory and applications*, volume 385 of *Lecture Notes in Control and Information Sciences*. Springer-Verlag, Berlin, 2009. `doi:10.1007/978-3-540-95980-9`.

14   Tobias Lechner. Application of theory for regular sequences on divide and conquer algorithms. Master's thesis, University of Klagenfurt, 2024. `urn:nbn:at:at-ubk:1-52067`. URL: `https://resolver.obvsg.at/urn:nbn:at:at-ubk:1-52067`.

Patricia's Bad Distributions

Louigi Addario-Berry ✉ 🄳
Department of Mathematics and Statistics, McGill University, Montréal, Canada

Pat Morin ✉ 🄳
School of Computer Science, Carleton University, Ottawa, Canada

Ralph Neininger[1] ✉ 🄳
Institute of Mathematics, Goethe University Frankfurt, Germany

──── **Abstract** ────

The height of a random PATRICIA tree built from independent, identically distributed infinite binary strings with arbitrary diffuse probability distribution μ on $\{0,1\}^{\mathbb{N}}$ is studied. We show that the expected height grows asymptotically sublinearly in the number of leaves for any such μ, but can be made to exceed any specific sublinear growth rate by choosing μ appropriately.

1 Introduction and results

The PATRICIA tree is a space efficient data structure for strings which improves on the trie. For the purpose of this note it is sufficient to introduce these tree structures for binary strings: Label the nodes of the complete infinite rooted binary tree by the elements of $\cup_{k=0}^{\infty}\{0,1\}^k$, starting at the root with $\emptyset$ and left and right child of a node labelled $v \in \{0,1\}^k$ with $v0$ and $v1$, respectively. Here, for $v \in \{0,1\}^k$ with $v = (v_1, \ldots, v_k)$ we abbreviate v as $v = v_1 \ldots v_k$ and denote $vi := v_1 v_2 \ldots v_k i$ for $i = 0, 1$.

The coming definitions are depicted in Figure 1. For distinct infinite binary strings $x_1, \ldots, x_n \in \{0,1\}^{\mathbb{N}}$ a finite tree called a *trie* (or *radix search tree*) to represent the strings $x_1, \ldots, x_n$ is constructed by first associating with each x_i the infinite path in $\cup_{k=0}^{\infty}\{0,1\}^k$ consisting of the nodes whose labels are the prefixes of x_i. The node labelled with the shortest such prefix that is not a prefix of any x_j with $j \in \{1, \ldots, n\} \setminus \{i\}$ becomes a leaf in the trie representing string x_i for $i = 1, \ldots, n$. The resulting tree, which is a finite binary tree with n leaves, is the trie representing $x_1, \ldots, x_n$. Next, starting from the trie, all vertices with out-degree 1 (i.e. with exactly one child) are deleted and the resulting gaps are closed by merging the two nodes which formed a deleted edge. This results in the PATRICIA tree, which was introduced independently by Morrison [21] and Gwehenberger [12] and first systematically analysed by Knuth [17]. The PATRICIA tree contains all the information needed to retrieve the strings and to perform operations such as sorting, searching and selecting; for broad expositions, see [18, 20, 25].

---

[1] Corresponding author

PATRICIA trees have been analysed assuming various probabilistic models for the input strings; where usually the infinite strings are assumed to be independent and identically distributed over $\{0,1\}^{\mathbb{N}}$. Note that atoms of such a distribution result in identical strings with positive probability, and in this case the construction of the trie does not lead to a finite tree. Hence, the law of the strings is usually assumed to be diffuse (non-atomic). Special cases of such diffuse probability distributions have been considered in the analysis of algorithms on strings such as the Bernoulli models, Markov model, dynamical sources or the density model; see [23, 24, 6, 4, 14, 11, 19, 1, 15, 13] and the references given in these papers.
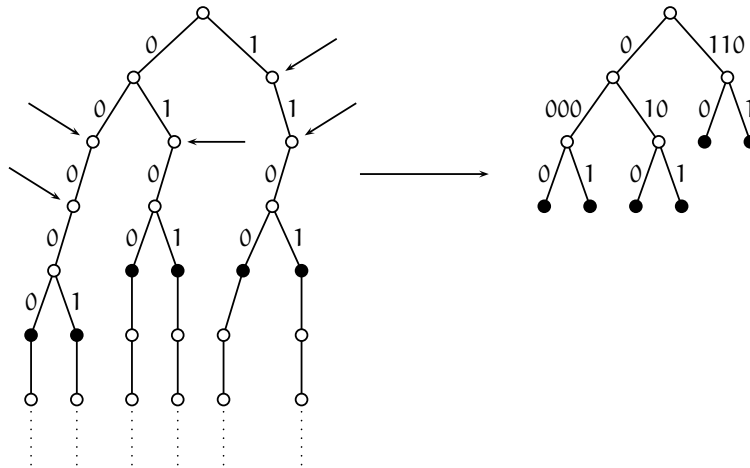
■ **Figure 1** On the left the trie for the strings $00000\ldots$, $00001\ldots$, $0100\ldots$, $0101\ldots$, $1100\ldots$, and $1101\ldots$ is shown. Its leaves are the full black vertices, the indicated children of the full black vertices do not belong to the trie. Vertices with out-degree 1 within the trie are indicated by arrows. On the right the resulting PATRICIA tree by deleting corresponding edges is shown.

In the present note we focus on the height of a PATRICIA tree, which is the maximal (graph) distance of any leaf from the root. The asymptotic behavior of the height of tries and PATRICIA trees under the Bernoulli models is covered by Pittel [23, 24] and Devroye [5, 7]. For example, for the height H_n^{syB} of the PATRICIA tree constructed from n independent strings under the symmetric Bernoulli model, i.e. all bits being independent and Bernoulli$(\frac{1}{2})$ distributed, Pittel [23] obtained as $n \to \infty$ that

$$\frac{H_n^{\mathrm{syB}}}{\log_2 n} \to 1 \text{ almost surely.}$$

This shows an asymptotic 50% improvement of the PATRICIA tree over the trie, for which the limit constant for the same probabilistic model is 2 instead of 1. For general diffuse laws concentration of the height of PATRICIA trees is studied (assuming only independence of the infinite strings not necessarily identical distribution) by Devroye [8] based on results from [2]; see also [16] for concentration of the height of PATRICIA trees in the Bernoulli model.

While such studies aim to show that the height behaves well with respect to applications from algorithms, Evans and Wakolbinger [9, 10] studied these random tree structures as tree-valued transient Markov chains from the perspective of Doob–Martin boundary theory. They asked (private communication) how high PATRICIA trees can grow for arbitrary diffuse probability distributions of the strings (see [10, Section 5] for specific examples). The subject

of the present note is to answer this question by Theorems 1 and 2: The expected height grows always sublinearly, but can be made to exceed any fixed sublinear growth rate by the choice of an appropriate diffuse law.

For a diffuse probability distribution μ on $\{0,1\}^{\mathbb{N}}$ and $(\Xi^{(j)})_{j \in \mathbb{N}}$ a sequence of independent and identically distributed random strings with law μ we denote by H_n^{μ} the height of the PATRICIA tree constructed from $\Xi^{(1)}, \ldots, \Xi^{(n)}$.

▶ **Theorem 1.** *For all diffuse probability distributions μ on $\{0,1\}^{\mathbb{N}}$ we have, as $n \to \infty$, that*

$$\frac{\mathbb{E}[H_n^{\mu}]}{n} \to 0, \quad and \quad \frac{H_n^{\mu}}{n} \to 0 \ almost \ surely.$$

▶ **Theorem 2.** *For any sequence $\alpha = (\alpha_n)_{n \in \mathbb{N}}$ of positive numbers with $\alpha_n \to \infty$ as $n \to \infty$ there exists a diffuse probability distribution $\nu = \nu^{(\alpha)}$ on $\{0,1\}^{\mathbb{N}}$ such that for all n sufficiently large*

$$\frac{\mathbb{E}[H_n^{\nu}]}{n/\alpha_n} \to \infty, \quad and \quad \frac{H_n^{\nu}}{n/\alpha_n} \to \infty \ almost \ surely.$$

We call a law ν on $\{0,1\}^{\mathbb{N}}$ causing large expected heights $\mathbb{E}[H_n^{\nu}]$ *bad* since such laws are undesirable from the point of view of the efficiency of algorithms based on PATRICIA trees. The remaining part of the present note contains proofs of these two theorems.

▶ Remark 1. For the density model, which is a subclass of the diffuse distributions on $\{0,1\}^{\mathbb{N}}$, the asymptotics of Theorem 1 were obtained by Devroye [6, page 419]. There, also bad distributions with asymptotic properties as in our Theorem 2 are constructed for sequences $\alpha_n = n^{\varepsilon}$ with $0 < \varepsilon < 1$.

2   Proofs

2.1   Proof of Theorem 1

We start with a technical observation:

▶ **Lemma 3.** *Suppose μ is a diffuse probability distribution on $\{0,1\}^{\mathbb{N}}$, and let $\Xi = (\xi_i)_{i \in \mathbb{N}}$ be random with law μ. Then for all ε there exists $k = k(\varepsilon) \in \mathbb{N}$ such that for any string $v = v_1 \ldots v_k \in \{0,1\}^k$, $\mathbb{P}(\xi_1 \ldots \xi_k = v_1 \ldots v_k) < \varepsilon$.*

Proof. Suppose for a contradiction that there exists $\varepsilon > 0$ such that for all $k \in \mathbb{N}$ there is a string $v_1 \ldots v_k \in \{0,1\}^k$ such that $\mathbb{P}(\xi_1 \ldots \xi_k = v_1 \ldots v_k) \geq \varepsilon$. Then by a compactness argument shown below there exists an infinite string $v = (v_i)_{i \in \mathbb{N}} \in \{0,1\}^{\mathbb{N}}$ such that for all $k \in \mathbb{N}$, $\mathbb{P}(\xi_1 \ldots \xi_k = v_1 \ldots v_k) \geq \varepsilon$. The events $\{\xi_1 \ldots \xi_k = v_1 \ldots v_k\}$ are decreasing in k, so this implies that

$$\mathbb{P}(\Xi = v) = \lim_{k \to \infty} \mathbb{P}(\xi_1 \ldots \xi_k = v_1 \ldots v_k) \geq \varepsilon \,,$$

which contradicts the assumption that μ is diffuse.

It remains to show the existence of the infinite string $v = (v_i)_{i \in \mathbb{N}} \in \{0,1\}^{\mathbb{N}}$ such that for all $k \in \mathbb{N}$, $\mathbb{P}(\xi_1 \ldots \xi_k = v_1 \ldots v_k) \geq \varepsilon$. Consider $\{0,1\}$ as a topological space with the discrete topology (all subsets being open) and $\{0,1\}^{\mathbb{N}}$ as the product space with the product topology. As a product of compact spaces $\{0,1\}^{\mathbb{N}}$ is compact. It is also a Hausdorff space. The projections $\Pi_k : \{0,1\}^{\mathbb{N}} \to \{0,1\}^k$ given by

$$(v_i)_{i \in \mathbb{N}} \overset{\Pi_k}{\longmapsto} v_1 \ldots v_k$$

are continuous for all $k \in \mathbb{N}$. Hence, the set

$$
\begin{aligned}
V_k &:= \{(v_i)_{i \in \mathbb{N}} \in \{0,1\}^{\mathbb{N}} \mid \mathbb{P}(\xi_1 \ldots \xi_k = v_1 \ldots v_k) \geq \varepsilon\} \\
&= \bigcup_{\substack{v_1 \ldots v_k \in \{0,1\}^k \\ \mathbb{P}(\xi_1 \ldots \xi_k = v_1 \ldots v_k) \geq \varepsilon}} \Pi_k^{-1}(\{v_1 \ldots v_k\})
\end{aligned}
$$

is closed and thus compact in $\{0,1\}^{\mathbb{N}}$. This implies that $(V_k)_{k \in \mathbb{N}}$ is a nested sequence of non-empty, compact sets. Now, Cantor's intersection theorem implies

$$
\bigcap_{k=1}^{\infty} V_k \neq \emptyset.
$$

Any element v of $\bigcap_{k=1}^{\infty} V_k$ has the desired property.   ◀

Proof of Theorem 1. Fix a diffuse probability distribution μ on $\{0,1\}^{\mathbb{N}}$. Let $\Xi^{(j)} = (\xi_i^{(j)})_{i \in \mathbb{N}}$ for $j \in \mathbb{N}$ be independent, identically distributed with law μ and denote by T_n the PATRICIA tree built from $\Xi^{(1)}, \ldots, \Xi^{(n)}$.

We first show that $H_n^{\mu}/n \to 0$ almost surely. Fix any $\varepsilon \in (0, 1/4)$. Let $k = k(\varepsilon)$ be as in Lemma 3, so that for any string $v = v_1 \ldots v_k \in \{0,1\}^k$, if $\Xi = (\xi_i)_{i \in \mathbb{N}}$ has law μ then $\mathbb{P}(\xi_1 \ldots \xi_k = v_1 \ldots v_k) < \varepsilon$. To prove $H_n^{\mu}/n \to 0$ almost surely we first show that

$$
\mathbb{P}(\exists n_0 \, \forall n \geq n_0 : H_n^{\mu} \leq k + 2\varepsilon n) = 1. \tag{1}
$$

Note that if the event

$$
E_{n,k} := \bigcup_{v_1 \ldots v_k \in \{0,1\}^k} \{|\{1 \leq j \leq n : \xi_1^{(j)} \ldots \xi_k^{(j)} = v_1 \ldots v_k\}| \geq 2\varepsilon n\}
$$

does not occur then the subtrees of T_n rooted at nodes $v \in \{0,1\}^k$ all have at most $2\varepsilon n$ leaves and so height less than $2\varepsilon n$; thus if $E_{n,k}$ does not occur then $H_n^{\mu} \leq k + 2\varepsilon n$. It follows that

$$
\begin{aligned}
&\mathbb{P}(\exists n_0 \, \forall n \geq n_0 : H_n^{\mu} \leq k + 2\varepsilon n) \\
&\geq \mathbb{P}(E_{n,k} \text{ occurs for at most finitely many values } n) \\
&= \mathbb{P}\left(\left(\limsup_{n \to \infty} E_{n,k}\right)^c\right),
\end{aligned}
$$

so to prove (1) it suffices to show that the probability of $\limsup_{n \to \infty} E_{n,k}$ is 0. For this, simply note that

$$
\begin{aligned}
\mathbb{P}(E_{n,k}) &\leq \sum_{v_1 \ldots v_k \in \{0,1\}^k} \mathbb{P}(|\{1 \leq j \leq n : \xi_1^{(j)} \ldots \xi_k^{(j)} = v_1 \ldots v_k\}| \geq 2\varepsilon n) \\
&\leq 2^k \mathbb{P}(Y_n \geq 2\varepsilon n),
\end{aligned}
$$

where Y_n has the Binomial distribution $\mathrm{Bin}(n, \varepsilon)$; the second inequality holds since the events that $\xi_1^{(j)} \ldots \xi_k^{(j)} = v_1 \ldots v_k$ are independent for distinct $1 \leq j \leq n$, and each has probability at most ε. A Chernoff bound then gives

$$
\mathbb{P}(E_{n,k}) \leq 2^k e^{-\varepsilon n/2}.
$$

Since this is summable, it follows by the first Borel–Cantelli lemma that

$$
\mathbb{P}\left(\limsup_{n \to \infty} E_{n,k}\right) = 0,
$$

hence we obtain (1). Now, note that for any $m_0 \in \mathbb{N}$,

$$\left\{ \frac{H_n^\mu}{n} \to 0 \right\} = \bigcap_{m=m_0}^{\infty} \bigcup_{n_0=1}^{\infty} \bigcap_{n=n_0}^{\infty} \left\{ \frac{H_n^\mu}{n} \leq \frac{3}{m} \right\}.$$

Thus, for $\varepsilon = \frac{1}{m}$ with fixed $m \geq m_0$ we can choose n sufficiently large so that $k(\varepsilon)/n \leq \varepsilon$ and obtain

$$\{H_n^\mu \leq k(\varepsilon) + 2\varepsilon n\} \subset \left\{ \frac{H_n^\mu}{n} \leq \frac{3}{m} \right\}$$

and see that (1) implies $H_n^\mu/n \to 0$ almost surely.

Finally, note that by construction of the PATRICIA tree we deterministically have $H_n^\mu \leq n-1$, thus $H_n^\mu/n \leq 1$. Hence, we obtain from $H_n^\mu/n \to 0$ almost surely and dominated convergence that $\mathbb{E}[H_n^\mu]/n \to 0$. ◀

2.2 Proof of Theorem 2

As building blocks for our bad distributions we first define a set of auxiliary probability distributions $(\mu_N, N \in \mathbb{N})$, on $\{0,1\}^{\mathbb{N}}$ as follows. For fixed $N \in \mathbb{N}$ we choose T uniformly at random from $\{1, \ldots, N^2\}$. Independently of T, let $(B_i)_{i \in \mathbb{N}}$ be independent Bernoulli($\frac{1}{2}$)-distributed random variables. Then define a sequence $(\vartheta_i)_{i \in \mathbb{N}}$ by

$$\vartheta_i = \begin{cases} 0, & \text{if } i < T, \\ 1, & \text{if } i = T, \\ B_{i-T}, & \text{if } i > T. \end{cases} \tag{2}$$

Now, μ_N is defined as the law of the string $\Theta = (\vartheta_i)_{i \in \mathbb{N}}$. Note that by definition μ_N is diffuse for all $N \in \mathbb{N}$. We use the notation

$$\langle \Theta \rangle := \min\{i \in \mathbb{N} \mid \vartheta_i = 1\}$$

for the index of the first entry of Θ equal to 1.

▶ **Lemma 4.** *For any $n \in \{1, \ldots, N\}$ we have $\mathbb{E}[H_n^{\mu_N}] \geq n - 2$.*

Proof. Let $1 \leq n \leq N \in \mathbb{N}$ and $\Theta^{(1)}, \ldots, \Theta^{(n)}$ be i.i.d. with law μ_N. We consider the set $A := \{\langle \Theta^{(1)} \rangle, \ldots, \langle \Theta^{(n)} \rangle\} \subset \{1, \ldots, N^2\}$. By construction of the PATRICIA tree we have

$$H_n^{\mu_N} \geq |A| - 1, \tag{3}$$

where $|A|$ denotes the cardinality of A, i.e., the number of distinct elements within the set $\{\langle \Theta^{(1)} \rangle, \ldots, \langle \Theta^{(n)} \rangle\}$. For all $1 \leq i < j \leq n$ we have $\mathbb{P}(\langle \Theta^{(i)} \rangle = \langle \Theta^{(j)} \rangle) = 1/N^2$. Hence, we obtain

$$\mathbb{E}[|A|] \geq n - \mathbb{E}\left[\sum_{1 \leq i < j \leq n} \mathbf{1}_{\{\langle \Theta^{(i)} \rangle = \langle \Theta^{(j)} \rangle\}} \right] \geq n - \frac{n^2}{2N^2} \geq n - 1, \tag{4}$$

since $n \leq N$. Now, (3) and (4) imply the assertion. ◀

Proof of Theorem 2. Without loss of generality we may assume that $\alpha_n = o(n)$. There exists an $n_0 \in \mathbb{N}$ such that $\alpha_n \geq 8$ for all $n \geq n_0$. We define $\beta_n := \lfloor \log_2 \alpha_n \rfloor - 2$ and a sequence $(A(n))_{n \in \mathbb{N}}$ as a generalized inverse of $(\beta_n)_{n \in \mathbb{N}}$ by

$$A(n) := \max\{m \in \mathbb{N} \mid \beta_m \leq n\}, \quad n \in \mathbb{N}. \tag{5}$$

First, a probability distribution $\mu^{(\alpha)}$ on $\{0,1\}^{\mathbb{N}}$ is obtained in two stages. Let G be a random variable with geometric distribution with parameter $\frac{1}{2}$, i.e., with $\mathbb{P}(G = k) = (\frac{1}{2})^k$ for $k \in \mathbb{N}$. Then define a sequence $(\lambda_i)_{i \in \mathbb{N}}$ by

$$\lambda_i = \begin{cases} 0, & \text{if } i < G, \\ 1, & \text{if } i = G, \\ \vartheta_{i-G}, & \text{if } i > G, \end{cases} \tag{6}$$

where $\Theta = (\vartheta_i)_{i \in \mathbb{N}}$, conditional on $\{G = k\}$, has law $\mu_{A(k)}$ defined in (2) with $A(\cdot)$ defined in (5). We then define $\mu = \mu^{(\alpha)}$ as the law of $\Lambda = (\lambda_i)_{i \in \mathbb{N}}$. Since the $\mu_{A(k)}$ are diffuse, we obtain that μ is diffuse.

Now, let $\Lambda^{(j)} = (\lambda_i^{(j)})_{i \in \mathbb{N}}$ for $j \in \mathbb{N}$ be independent with law μ. For $n \geq n_0$, by construction,

$$X_n := \left| \left\{ 1 \leq j \leq n \; : \; \left(\lambda_1^{(j)}, \ldots, \lambda_{\beta_n}^{(j)} \right) = (0, \ldots, 0, 1) \right\} \right|$$

is $\text{Bin}(n, 2^{-\beta_n})$-distributed. To get rid of the floors in the definition of β_n denote by X'_n a $\text{Bin}(n, 4/\alpha_n)$-distributed random variable. Note that $2^{-\beta_n} \geq 4/\alpha_n$. By Okamoto's inequality, see [22] or [3, Exercise 2.12] we have

$$\mathbb{P}\left(X_n < \frac{2n}{\alpha_n} \right) \leq \mathbb{P}\left(X'_n - \frac{4n}{\alpha_n} < -\frac{2n}{\alpha_n} \right)$$

$$\leq \exp\left(-\frac{n(2/\alpha_n)^2}{2(4/\alpha_n)(1 - 4/\alpha_n)} \right) \leq \exp\left(-\frac{n}{2\alpha_n} \right).$$

Hence, with high probability at least $\lceil 2n/\alpha_n \rceil$ of the n strings start with the prefix $(0, \ldots, 0, 1)$ of length β_n and thus have suffixes $(\lambda_{\beta_n+1}^{(j)}, \lambda_{\beta_n+2}^{(j)}, \ldots)$ drawn independently from $\mu_{A(\beta_n)}$ for the respective j. For all $n \geq n_0$ we have $\lceil 2n/\alpha_n \rceil \leq n \leq A(\beta_n)$. Hence, by Lemma 4, $\lceil 2n/\alpha_n \rceil$ such strings cause an expected height of at least $2n/\alpha_n - 2$. Together we obtain for all sufficiently large n, note also $\alpha_n = o(n)$, that

$$\mathbb{E}[H_n^{\mu}] \geq \mathbb{P}\left(X_n \geq \frac{2n}{\alpha_n} \right) \mathbb{E}\left[H_n^{\mu} \Big| X_n \geq \frac{2n}{\alpha_n} \right]$$

$$\geq \left(1 - \exp\left(-\frac{n}{2\alpha_n} \right) \right) \left(\frac{2n}{\alpha_n} - 2 \right)$$

$$\geq \frac{n}{\alpha_n}.$$

Since the sequence $(\log \alpha_n)$ tends to infinity the present proof implies the existence of a diffuse probability distribution $\nu = \nu^{(\alpha)}$ on $\{0,1\}^{\mathbb{N}}$ such that $\mathbb{E}[H_n^{\nu}] \geq n/\log \alpha_n$ for all sufficiently large $n \in \mathbb{N}$, hence

$$\frac{\mathbb{E}[H_n^{\nu}]}{n/\alpha_n} \to \infty.$$

To prove the second statement of Theorem 2 we use the following bound from Devroye [8, page 21]: for any diffuse probability distribution μ on $\{0,1\}^{\mathbb{N}}$ and any $t > 0$,

$$\mathbb{P}\left(H_n^{\mu} \leq \mathbb{E}[H_n^{\mu}] - t \right) \leq \exp\left(-\frac{t^2}{2\mathbb{E}[H_n^{\mu} + 1]} \right) \leq \exp\left(-\frac{t^2}{2n} \right). \tag{7}$$

We now consider the probabilities

$$\mathbb{P}\left(H_n^{\nu} \leq \frac{n}{\log^2 \alpha_n} \right) = \mathbb{P}\left(H_n^{\nu} \leq \mathbb{E}[H_n^{\nu}] - \left(\mathbb{E}[H_n^{\nu}] - \frac{n}{\log^2 \alpha_n} \right) \right) \tag{8}$$

and note that for all sufficiently large n we have

$$\mathbb{E}[H_n^\nu] - \frac{n}{\log^2 \alpha_n} \geq \frac{n}{\log \alpha_n} - \frac{n}{\log^2 \alpha_n} = n \frac{\log(\alpha_n) - 1}{\log^2 \alpha_n} \geq \frac{n}{\log^2 n}. \tag{9}$$

Combining (7)–(9) we obtain

$$\mathbb{P}\left(H_n^\nu \leq \frac{n}{\log^2 \alpha_n}\right) \leq \exp\left(-\frac{n}{2\log^4 n}\right)$$

for all sufficiently large n. Since these upper bounds are summable it follows from the first Borel–Cantelli Lemma that $\liminf_{n\to\infty} H_n^\nu/(n/\log^2 \alpha_n) \geq 1$ almost surely, hence

$$\frac{H_n^\nu}{n/\alpha_n} \to \infty \text{ almost surely.}$$

Thus, ν has the properties claimed in Theorem 2. ◀

---- **References** ----

1   Ali Akhavi, Fréderic Paccaut, and Brigitte Vallée. Building Sources of Zero Entropy: Rescaling and Inserting Delays. In Mark Daniel Ward, editor, *33rd International Conference on Probabilistic, Combinatorial and Asymptotic Methods for the Analysis of Algorithms (AofA 2022)*, volume 225 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 1:1–1:28, Dagstuhl, Germany, 2022. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. `doi:10.4230/LIPIcs.AofA.2022.1`.

2   Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. A sharp concentration inequality with applications. *Random Structures Algorithms*, 16(3):277–292, 2000. `doi:10.1002/(SICI) 1098-2418(200005)16:3<277::AID-RSA4>3.0.CO;2-1`.

3   Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration inequalities*. Oxford University Press, Oxford, 2013. A nonasymptotic theory of independence, With a foreword by Michel Ledoux. `doi:10.1093/acprof:oso/9780199535255.001.0001`.

4   Julien Clément, Philippe Flajolet, and Brigitte Vallée. Dynamical sources in information theory: a general analysis of trie structures. *Algorithmica*, 29(1-2):307–369, 2001. Average-case analysis of algorithms (Princeton, NJ, 1998). `doi:10.1007/BF02679623`.

5   Luc Devroye. A note on the probabilistic analysis of patricia trees. *Random Structures & Algorithms*, 3(2):203–214, 1992. `doi:10.1002/rsa.3240030209`.

6   Luc Devroye. A study of trie-like structures under the density model. *Ann. Appl. Probab.*, 2(2):402–434, 1992. URL: `http://links.jstor.org/sici?sici=1050-5164(199205) 2:2<402:ASOTSU>2.0.CO;2-F&origin=MSN`.

7   Luc Devroye. Universal limit laws for depths in random trees. *SIAM Journal on Computing*, 28(2):409–432, 1998. `doi:10.1137/S0097539795283954`.

8   Luc Devroye. Universal asymptotics for random tries and PATRICIA trees. *Algorithmica*, 42(1):11–29, 2005. `doi:10.1007/s00453-004-1137-7`.

9   Steven N. Evans and Anton Wakolbinger. Radix sort trees in the large. *Electronic Communications in Probability*, 22(none):1–13, 2017. `doi:10.1214/17-ECP77`.

10 Steven N. Evans and Anton Wakolbinger. PATRICIA bridges. In *Genealogies of interacting particle systems*, volume 38 of *Lect. Notes Ser. Inst. Math. Sci. Natl. Univ. Singap.*, pages 231–265. World Sci. Publ., Hackensack, NJ, [2020] ©2020.

11 Michael Fuchs, Hsien-Kuei Hwang, and Vytas Zacharovas. An analytic approach to the asymptotic variance of trie statistics and related structures. *ArXiv*, abs/1303.4244, 2013. URL: `https://api.semanticscholar.org/CorpusID:14429040`.

12 Gernot Gwehenberger. Anwendung einer binären Verweiskettenmethode beim Aufbau von Listen. *Elektronische Rechenanlagen*, 10:223–226, 1968.

13   Jasper Ischebeck. Central limit theorems for fringe trees in patricia tries. *Preprint arXiv:2305.14900 [math.PR]*, 2023. `doi:10.48550/arXiv.2305.14900`.

14   Svante Janson. Renewal theory in the analysis of tries and strings. *Theor. Comput. Sci.*, 416:33–54, January 2012. `doi:10.1016/j.tcs.2011.10.007`.

15   Svante Janson. Central limit theorems for additive functionals and fringe trees in tries. *Electronic Journal of Probability*, 27(none):1–63, 2022. `doi:10.1214/22-EJP776`.

16   Charles Knessl and Wojciech Szpankowski. Limit laws for the height in PATRICIA tries. *J. Algorithms*, 44(1):63–97, 2002. Analysis of algorithms. `doi:10.1016/S0196-6774(02)00212-2`.

17   Donald E. Knuth. *The art of computer programming. Volume 3*. Addison-Wesley Series in Computer Science and Information Processing. Addison-Wesley Publishing Co., Reading, Mass.-London-Don Mills, Ont., 1973. Sorting and searching.

18   Donald E. Knuth. *The art of computer programming. Vol. 3*. Addison-Wesley, Reading, MA, 1998. Sorting and searching, Second edition [of MR0445948].

19   Kevin Leckey, Ralph Neininger, and Wojciech Szpankowski. *Towards More Realistic Probabilistic Models for Data Structures: The External Path Length in Tries under the Markov Model*, pages 877–886. Society for Industrial and Applied Mathematics, 2013. `doi:10.1137/1.9781611973105.63`.

20   Hosam M. Mahmoud. *Evolution of random search trees*. Wiley-Interscience Series in Discrete Mathematics and Optimization. John Wiley & Sons, Inc., New York, 1992. A Wiley-Interscience Publication.

21   Donald R. Morrison. Patricia—practical algorithm to retrieve information coded in alphanumeric. *Journal of the ACM*, 15:514–534, 1968.

22   Masashi Okamoto. Some inequalities relating to the partial sum of binomial probabilities. *Ann. Inst. Statist. Math.*, 10:29–35, 1958. `doi:10.1007/BF02883985`.

23   B. Pittel. Asymptotical growth of a class of random trees. *Ann. Probab.*, 13(2):414–427, 1985. `doi:10.1214/aop/1176993000`.

24   Boris Pittel. Paths in a random digital tree: limiting distributions. *Adv. in Appl. Probab.*, 18(1):139–155, 1986. `doi:10.2307/1427240`.

25   Wojciech Szpankowski. *Average case analysis of algorithms on sequences*. Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley-Interscience, New York, 2001. With a foreword by Philippe Flajolet. `doi:10.1002/9781118032770`.

Limit Laws for Critical Dispersion on Complete Graphs

Umberto De Ambroggio ✉
Department of Mathematics, LMU Munich, Germany

Tamás Makai ✉
Department of Mathematics, LMU Munich, Germany

Konstantinos Panagiotou ✉ ⌂
Department of Mathematics, LMU Munich, Germany

Annika Steibel ✉
Department of Mathematics, LMU Munich, Germany

───── **Abstract** ─────

We consider a synchronous process of particles moving on the vertices of a graph G, introduced by Cooper, McDowell, Radzik, Rivera and Shiraga (2018). Initially, M particles are placed on a vertex of G. In subsequent time steps, all particles that are located on a vertex inhabited by at least two particles jump independently to a neighbour chosen uniformly at random. The process ends at the first step when no vertex is inhabited by more than one particle; we call this (random) time step the *dispersion time*.

In this work we study the case where G is the complete graph on n vertices and the number of particles is $M = n/2 + \alpha n^{1/2} + o(n^{1/2})$, $\alpha \in \mathbb{R}$. This choice of M corresponds to the critical window of the process, with respect to the dispersion time. We show that the dispersion time, if rescaled by $n^{-1/2}$, converges in p-th mean, as $n \to \infty$ and for any $p \in \mathbb{R}$, to a continuous and almost surely positive random variable T_α. We find that T_α is the absorption time of a standard logistic branching process, thoroughly investigated by Lambert (2005), and we determine its expectation. In particular, in the middle of the critical window we show that $\mathbb{E}[T_0] = \pi^{3/2}/\sqrt{7}$, and furthermore we formulate explicit asymptotics when $|\alpha|$ gets large that quantify the transition into and out of the critical window. We also study the random variable counting the *total number of jumps* that are performed by the particles until the dispersion time is reached and prove that, if rescaled by $n \ln n$, it converges to $2/7$ in probability.

1 Introduction

The *dispersion process* introduced by Cooper, McDowell, Radzik, Rivera and Shiraga [2] consists of particles moving on the vertices of a given graph G. A particle is said to be *happy* if there are no other particles occupying the same vertex and *unhappy* otherwise. Initially, $M \geq 2$ (unhappy) particles are placed on some vertex of G. Subsequently, at discrete time steps, all unhappy particles move *simultaneously* and *independently* to a neighbouring
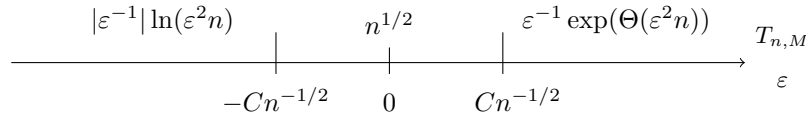
35th International Conference on Probabilistic, Combinatorial and Asymptotic Methods for the Analysis of Algorithms (AofA 2024).
Editors: Cécile Mailler and Sebastian Wild; Article No. 26; pp. 26:1–26:12
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

$$|\varepsilon^{-1}|\ln(\varepsilon^2 n) \qquad n^{1/2} \qquad \varepsilon^{-1}\exp(\Theta(\varepsilon^2 n)) \qquad T_{n,M}$$

$$-Cn^{-1/2} \qquad 0 \qquad Cn^{-1/2} \qquad \varepsilon$$

Figure 1 The typical order of $T_{n,M}$ when $M = (1+\varepsilon)n/2$ and $|\varepsilon| = o(1)$. Note that $|\varepsilon^{-1}|\ln(\varepsilon^2 n)$ and $\varepsilon^{-1}\exp(\varepsilon^2 n)$ are in $\Theta(\sqrt{n})$ when $|\varepsilon| = \Theta(n^{-1/2})$, and so the transition into and out of the critical window is smooth.

vertex selected uniformly at random, while the happy particles remain in place. The process terminates at the first time step at which all particles are happy; we call this (random) time step the *dispersion time*.

It is clear that if the number of particles is small – compared to the number of vertices in the graph – then the dispersion time should be small as well. Intuitively, increasing the number of particles makes it more and more difficult for the particles to disperse quickly. This transition from 'fast' to 'slow' dispersion is quite well-understood and sharp when the underlying graph is the complete graph on n vertices with loops, in which case we write $T_{n,M}$ for the dispersion time started with M particles at an arbitrary vertex. The typical order of $T_{n,M}$ changes rather abruptly around $M = n/2$. Indeed, if we write $M = M(n) = (1+\varepsilon)n/2 \in \mathbb{N}$ for some sequence $\varepsilon = \varepsilon(n) \in [-1, 1]$, then in [2] it was established that $T_{n,M}$ is typically

- at most logarithmic in n when $\limsup_{n\to\infty} \varepsilon < 0$ and
- at least exponential in n when $\liminf_{n\to\infty} \varepsilon > 0$.

The details of this apparent and abrupt transition from logarithmic to exponential time are obviously of great interest and were investigated further in [3], where the authors studied the typical order and the tails of $T_{n,M}$ when $\varepsilon = o(1)$, that is, when $M = n/2 + o(n)$. In this setting they showed that for any constant $C > 0$, if $\varepsilon \le -Cn^{-1/2}$, then the process typically finishes in $\Theta(|\varepsilon|^{-1}\ln(\varepsilon^2 n))$ steps, while if $\varepsilon \ge Cn^{-1/2}$, then a much larger number $\varepsilon^{-1}\exp(\Theta(\varepsilon^2 n))$ of steps is required. Moreover, within the *critical window* corresponding to the range $|\varepsilon| = O(n^{-1/2})$, they showed that the process typically runs for $\Theta(n^{1/2})$ steps, making the transition into and out of the critical window smooth, see also Figure 1.

In this paper we will perform a fine analysis of the dispersion process within the critical window, that is, when $M = n/2 + O(\sqrt{n})$. Our first main result establishes that the dispersion time, scaled by $n^{-1/2}$, converges in distribution to some continuous and almost surely positive random variable. For a sequence of real-valued random variables $(Z_n)_{n\in\mathbb{N}}$ and a random variable Z we write $Z_n \xrightarrow{d} Z$ to denote that the sequence $(Z_n)_{n\in\mathbb{N}}$ converges to Z in distribution.

▶ **Theorem 1.** *Let $\alpha \in \mathbb{R}$ and $M = M(n) = n/2 + \alpha\sqrt{n} + o(\sqrt{n}) \in \mathbb{N}$. Then there is a continuous and almost surely positive random variable T_α such that, as $n \to \infty$,*

$$n^{-1/2}T_{n,M} \xrightarrow{d} T_\alpha \ .$$

Within the proof of Theorem 1 we derive an explicit description of the distribution of T_α. In order to specify it at this point we need to step back a bit and introduce some notation and present some facts about the process. Let us write U_t for the (random) number of unhappy particles at the end of step t, so that $U_0 = M$, and let us fix some $\delta > 0$. As we will argue in Section 3, U_t drops rather quickly to $\Theta(n^{1/2})$ particles. In particular, with probability at least $1 - \delta$, after $t^* \sim \frac{4}{7}\delta n^{1/2}$ steps we have that $U_{t^*} \sim n^{1/2}/\delta$; here and everywhere else "$\sim$" will stand for "$= (1 + o(1))$" and asymptotic statements are, unless stated explicitly otherwise, with respect to $n \to \infty$ and uniform in all other parameters. After t^* the process
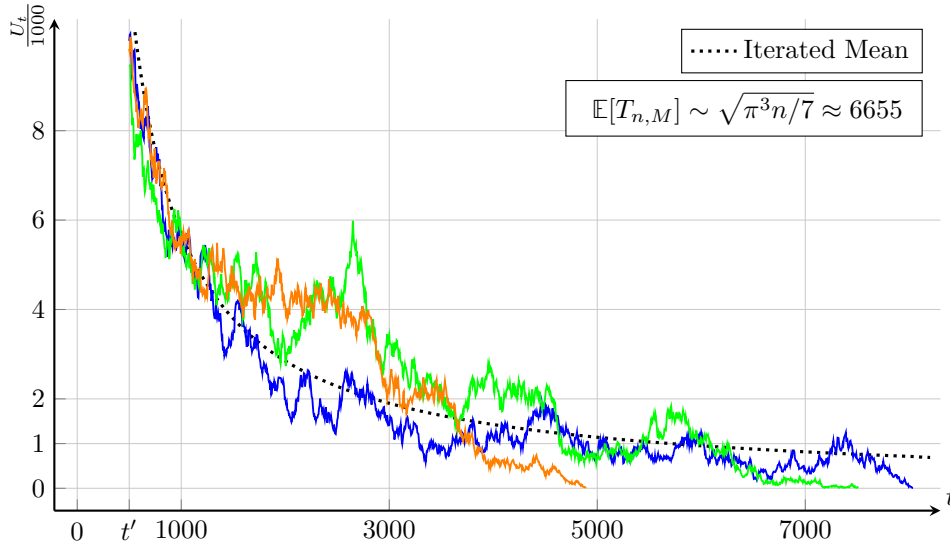
Figure 2 Three sample runs of the dispersion process with $n = 10^7$ and $M = n/2$, where we depict the number of unhappy particles U_t, divided by 1000, at each step t. The trajectory is revealed only after $t' = 500$, where $U_{t'} \approx 10^4 \approx 3\sqrt{n}$ in all cases. The dotted line represents the iterated mean of U_t. For the asymptotics of $\mathbb{E}[T_{n,M}]$ see (4).

$(U_t)_{t \geq t^*}$ of unhappy particles starts fluctuating significantly, see Figure 2 for outcomes of a simulation study when $M = n/2$. In order to get a grip on it, we scale time and space by a factor of $n^{1/2}$ and establish that $(n^{-1/2} U_{t^* + \lfloor s\sqrt{n} \rfloor})_{s \geq 0}$ converges weakly to a diffusion process. Here weak convergence denotes, as usual, convergence in $D([0, T], \mathbb{R})$ for all $T < \infty$, where $D([0, T], \mathbb{R})$ represents the space of all right-continuous functions from $[0, T]$ to $\mathbb{R}$ with left-limits.

▶ **Lemma 2.** *Let $\alpha \in \mathbb{R}$ and $M = M(n) = n/2 + \alpha\sqrt{n} + o(\sqrt{n}) \in \mathbb{N}$. Let $\delta > 0$ and let*

$$T_{n,M,\delta} := \inf\{t > 0 : U_t \leq n^{1/2}/\delta\}$$

be the first step at which there are at most $n^{1/2}/\delta$ unhappy particles. Then, as $n \to \infty$, weakly

$$\left(n^{-1/2} U_{T_{n,M,\delta} + \lfloor sn^{1/2} \rfloor}\right)_{s \geq 0} \to X,$$

where X is a logistic branching process starting from $X_0 = \delta^{-1}$. In particular, if we denote by B a standard Brownian motion, then X uniquely satisfies the SDE

$$dX_s = \left(2\alpha X_s - \frac{7}{4}X_s^2\right)ds + \sqrt{X_s}dB_s, \quad s > 0, \quad and \quad X_0 = \delta^{-1}. \tag{1}$$

For more background on SDEs in general and the specific equation encountered here we refer to Section 2. Let us mention only that stochastic processes satisfying (1) are well-studied and are also called in the literature *logistic Feller diffusions* or *Feller diffusions with logistic growth*. Generally, such processes satisfy an SDE of the form

$$dX_s = (aX_s - cX_s^2)ds + \sqrt{\gamma X_s}dB_s, \ s > 0, \ \text{with initial condition } X_0 = x \geq 0, \tag{2}$$

where $a \in \mathbb{R}$ and $c, \gamma > 0$. They appear in the context of population dynamics and stochastically extend the deterministic logistic growth model that describes the evolution of a population under the influences of natural birth, mortality and inter-individual competition. A prime source on the topic is Lambert [7], who provides a thorough and detailed discussion of the properties of solutions to (2).

With Lemma 2 at hand we show in Section 3, see Lemma 13 there, that the first step at which the unhappy particles vanish, divided by $n^{1/2}$, converges in distribution to the *absorption time* of X, that is, the first time when X hits zero. Letting $\delta \to 0$ then yields the claimed statement. In particular, T_α in Theorem 1 is the absorption time of the limiting solution of (1) when the initial condition $X_0 \to \infty$; this limiting process, called *standard logistic branching process*, is well-defined and well-studied, see for example [7] and Section 2.2.

The explicit descriptions of X and T_α pave the way to obtain further bits of information. To achieve this we will exploit the following bounds, stating that $n^{-1/2}T_{n,M}$ has exponential tails, that are an immediate consequence of the main theorems in [3].

▶ **Theorem 3.** *Let $\alpha \in \mathbb{R}$ and $M = n/2 + \alpha\sqrt{n} + o(\sqrt{n}) \in \mathbb{N}$. Then there is a constant $c_\alpha > 0$ such that for all sufficiently large n*

$$\mathbb{P}\big(T_{n,M} \leq n^{1/2}/Ac_\alpha\big) \leq e^{-A} \quad and \quad \mathbb{P}\big(T_{n,M} > Ac_\alpha n^{1/2}\big) \leq e^{-A}, \qquad A \geq 1.$$

Together with our Theorem 1 this implies that for any $p \in \mathbb{R}$ we even obtain convergence in $\mathcal{L}^p$, in particular

$$n^{-p/2}\mathbb{E}\big[T_{n,M}^p\big] \sim \mathbb{E}\big[T_\alpha^p\big], \quad p \in \mathbb{R}. \tag{3}$$

We also obtain, without including the proof here, for $M = n/2 + \alpha\sqrt{n} + o(\sqrt{n})$ the series representation

$$\mathbb{E}[T_\alpha] = \lim_{n\to\infty} \mathbb{E}\big[n^{-1/2}T_{n,M}\big] = \frac{\pi^{3/2}}{\sqrt{7}} + \frac{1}{\sqrt{7}}\sum_{m\geq 1}\frac{\Gamma(\frac{m+1}{2})}{m!}\left(\frac{8\alpha}{\sqrt{7}}\right)^m t_m, \quad \alpha \in \mathbb{R},$$

where $\Gamma(\cdot)$ is the Gamma function and

$$t_m := \sum_{k\geq 0}\frac{2}{(\frac{m+1}{2}+2k)(\frac{m+3}{2}+2k)} = H_{(m-1)/4} - H_{(m-3)/4}, \quad m \in \mathbb{N}_0,$$

and $H_x = \sum_{k\geq 1}\left(\frac{1}{k} - \frac{1}{k+x}\right)$ denotes the "x-th harmonic number". Let us highlight the specific case $\alpha = 0$: when we are essentially *at* the critical point, then we obtain the beautiful formula

$$\mathbb{E}[T_0] = \lim_{n\to\infty}\mathbb{E}\big[n^{-1/2}T_{n,n/2+o(\sqrt{n})}\big] = \frac{\pi^{3/2}}{\sqrt{7}}, \tag{4}$$

which is in the interval 2.104 ± 0.001, see also Figure 2. Our methods also allow us to study the behavior of the transition in and out of the critical window, that is, $\mathbb{E}[T_\alpha]$ when $\alpha \to -\infty$ or $\alpha \to \infty$. Indeed we are able to show the following asymptotics

$$\mathbb{E}[T_\alpha] \overset{\alpha\to-\infty}{\sim} \frac{\ln|\alpha|}{|\alpha|} \qquad \text{and} \qquad \mathbb{E}[T_\alpha] \overset{\alpha\to\infty}{\sim} \frac{\sqrt{7\pi}}{8}\frac{e^{16\alpha^2/7}}{\alpha^2}.$$

So, when α gets big, then $\mathbb{E}[T_\alpha]$ behaves (up to polynomial corrections) quadratic exponential in α; already for $\alpha = 3$ we obtain the enormous value $\mathbb{E}[T_3] \approx 5.894 \cdot 10^7$. On the other hand, for negative α we get a moderate polynomial behavior with logarithmic corrections. Note

that the large $|\alpha|$ asymptotics presented here are in perfect accordance with the transition in and out of the critical window, see also Figure 1 and the discussion at the beginning of the introduction.

Our second main result addresses the total number of jumps $\sum_{t\geq 0} U_t$ performed by the particles. In contrast to the dispersion time, the total number of jumps, scaled by $n \ln n$, converges to a fixed quantity.

▶ **Theorem 4.** *Let $\alpha \in \mathbb{R}$ and $M = M(n) = n/2 + \alpha\sqrt{n} + o(\sqrt{n}) \in \mathbb{N}$. Then*

$$\frac{1}{n \ln n} \sum_{t \geq 0} U_t \xrightarrow{d} \frac{2}{7} \ .$$

In particular, each of the $M \sim n/2$ particles performs on average typically $\sim \frac{4}{7} \ln n$ jumps before everybody settles, and this is independent of α. Indeed, our aforementioned analysis of the early steps in Section 3, that is, the first $o(n^{1/2})$ steps, shows that there are already $\sim \frac{2}{7} n \ln n$ jumps in those steps of the process. With Lemma 2 and Theorem 1 in mind, it is not surprising that the remaining $\Theta(n^{1/2})$ steps only contribute an additional of $O(n)$ number jumps, as $n^{-1/2} U_t$ is typically bounded for $t = \Theta(n^{1/2})$.

Theorem 1 and Lemma 2 actually suggest that a much stronger statement should be true. We know that $(n^{-1/2} U_{T_{n,M,\delta} + \lfloor sn^{1/2} \rfloor})_{s \geq 0}$ converges weakly to a logistic branching process X, and so the total number of jumps should be close to $n^{1/2} A$, where $A := \int_0^\infty X_s ds$, plus the additional $\frac{2}{7} n \ln n$ jumps from the first $T_{n,M,\delta}$ steps. Thus the variations in the total number of jumps should be linear in n; that is, there should be a (non-trivial) random variable S such that

$$n^{-1} \left(\sum_{t \geq 0} U_t - \frac{2}{7} n \ln n \right) \xrightarrow{d} S.$$

We leave it as an open problem to prove this conjecture.

Variations on the Theme

Our work opens up opportunities for studying a variety of models that are related to the dispersion process or extensions of it. In a general setting, *happiness* can be defined as a property of individual vertices and particles. More specifically, each vertex may have a *capacity*, which, if exceeded, deems all particles on that vertex as unhappy. On the other side, each particle p may have a *stress level*, which dictates an upper bound on the particles that share a vertex with p so that p is still happy. We leave it as an open problem to study the precise behavior in a general setting, where for example the empirical distributions of the capacities and the stress levels fulfill appropriate convergence properties.

In a different line of research it would be challenging to provide detailed studies of dispersion processes on graphs different than the complete graph. We believe, for example, that our results also hold if the underlying graph is a sufficiently dense Erdős-Rényi random graph $G_{n,p}$, which is obtained by retaining independently each edge of the complete graph on n vertices with probability p. In particular, if, say, $p = \omega(n^{-1/2})$, guaranteeing that the minimum degree is much larger than $\sqrt{n}$, then similar results as in Theorem 1 should hold, as the process finishes after $O(\sqrt{n})$ rounds if the graph is complete. However, it might be the case that even on much sparser graphs the behavior does not change (since, for example, in most steps just an $O(\sqrt{n})$ number of particles move). We consider it as an important and eminent challenge to study the effect of the edge probability p on the distribution of the dispersion time.

Related work

The dispersion process was also studied by Frieze and Pegden [6], who, apart from the dispersion time, also considered the *dispersion distance* on the infinite line. They showed that the dispersion distance is $\Theta(n)$ when there are n particles in the system, improving upon previous results in [2]. A similar setup was considered by Shang [10], who studied the dispersion distance on the infinite line in a non-uniform dispersion process.

Processes where particles move on the vertices of a graph have been widely studied over the past decades; we refer the reader to [2] for references. Concerning processes whose scope is to *disperse* particles on a discrete structure, arguably the best known such model is Internal Diffusion Limited Aggregation (IDLA), see [1, 4, 8]. In this model, particles sequentially start (one at a time) from a specific vertex designated as the origin. Each particle moves randomly until it finds an unoccupied vertex; then it occupies it forever, meaning that it does not move at subsequent process steps.

Another related and well-studied class of models are Activated Random Walks (ARWs) that evolve on the d-dimensional lattice, see [9] for an extensive review. Roughly speaking, we place particles on $\mathbb{Z}^d$, and some of them are initially active while others are asleep. The rules of the process are then as follows. Whenever a particle is alone on a vertex, it falls asleep with a certain rate. On the other hand, active particles jump according to independent random choices, and whenever they encounter a particle that is asleep, they wake it up.

Outline

In Section 2 we present the main tool used during our proof, namely diffusion approximation, and then in Section 3 we include a brief derivation of some of the aforementioned results, primarily Lemma 2 and Theorem 1. The full paper containing all proofs is available at `arXiv:2403.05372`.

2   Probabilistic Preliminaries

2.1   Diffusion Approximation

A main tool that we will use in the proof of Theorem 1 is the concept of *diffusion approximation*, which allows us to approximate a Markov chain $(Y^{(n)})_{n \in \mathbb{N}}$ with values in $\mathbb{R}$, by a continuous-time stochastic process. More specifically, we examine convergence properties of $(Y^{(n)})_{n \in \mathbb{N}}$ to a process satisfying a stochastic differential equation (SDE)

$$dX_s = b(X_s)ds + \sigma(X_s)dB_s, \quad s > 0, \tag{5}$$

where $b, \sigma : \mathbb{R} \to \mathbb{R}$ are suitable functions and B is a 1-dimensional standard Brownian motion. In this section we provide an overview of the necessary results from stochastic calculus. Additionally, we collect some properties of the limit process that will emerge within the proof of Theorem 1. In what follows we denote discrete time by $t \in \mathbb{N}_0$ (so, for example, $Y^{(n)} = (Y_t^{(n)})_{t \in \mathbb{N}_0}$), whereas $s \geq 0$ represents continuous time.

Let us consider (5). A *(weak) solution to* (5) *with initial value* $X_0 = x \in \mathbb{R}$ is a triple $(X, B, \mathcal{P})$, where $\mathcal{P} = (\Omega, \mathcal{F}, (\mathcal{F}_s)_{s \geq 0}, \mathbb{P})$ is a filtered probability space with the filtration satisfying the usual conditions, i.e. $(\mathcal{F}_s)_{s \geq 0}$ is right-continuous and complete. Further, $X = (X_s)_{s \geq 0}$ and $B = (B_s)_{s \geq 0}$ are continuous stochastic processes that are adapted to $(\mathcal{F}_s)_{s \geq 0}$ such that

- B is a standard 1-dimensional Brownian motion with respect to $(\mathcal{F}_s)_{s \geq 0}$, i.e. B is a standard Brownian motion and $B_s - B_r$ is independent of $\mathcal{F}_r$ for any $0 \leq r < s$;

■ X_s satisfies (5) and the initial condition, i.e.

$$X_s = x + \int_0^s b(X_r)dr + \int_0^s \sigma(X_r)dB_r, \quad s \geq 0.$$

Moreover, we say that there is *(weak) uniqueness* if whenever $(X, B, \mathcal{P})$ and $(\tilde{X}, \tilde{B}, \tilde{\mathcal{P}})$ solve (5) weakly and satisfy $X_0 = \tilde{X}_0$, then X and $\tilde{X}$ have the same law.

In order to get the diffusion approximation to work, we construct a sequence of right-continuous and continuous-time stochastic processes from the given sequence $(Y^{(n)})_{n\in\mathbb{N}}$ of discrete time Markov chains by using constant interpolation between the time points. Then, under appropriate conditions specified in the subsequent theorem, $(Y^{(n)})_{n\in\mathbb{N}}$ converges weakly to the solution of an SDE. With the necessary concepts at hand we are now ready to present our main tool, and we refer for example to [5, Ch. 8] for an extensive treatment.

▶ **Theorem 5** (Diffusion Approximation). *Let $b, \sigma : \mathbb{R} \to \mathbb{R}$ be continuous functions and assume that for any $x \in \mathbb{R}$ the SDE (5) possesses a unique solution such that $X_0 = x$. Furthermore, let $h : \mathbb{N} \to \mathbb{R}_+$ be a sequence with $\lim_{n\to\infty} h(n) = 0$ and for all $n \in \mathbb{N}$ let $Y^{(n)} = (Y_t^{(n)})_{t\in\mathbb{N}_0}$ be a discrete-time Markov chain with values in $S^{(n)} \subseteq \mathbb{R}$. Define, for all $t \in \mathbb{N}_0$, $x \in S^{(n)}$*

$$b^{(n)}(x) := \frac{\mathbb{E}\big[Y_{t+1}^{(n)} - x \mid Y_t^{(n)} = x\big]}{h(n)}, \quad a^{(n)}(x) := \frac{\mathbb{E}\big[(Y_{t+1}^{(n)} - x)^2 \mid Y_t^{(n)} = x\big]}{h(n)},$$

and $\gamma_p^{(n)}(x) := \mathbb{E}\big[|Y_{t+1}^{(n)} - x|^p \mid Y_t^{(n)} = x\big]/h(n)$ for $p \geq 2$. Let $a := \sigma^2$ and assume that for all $R < \infty$

$$\lim_{n\to\infty} \sup_{x\in S^{(n)}, |x|\leq R} |b^{(n)}(x) - b(x)| = 0, \quad \lim_{n\to\infty} \sup_{x\in S^{(n)}, |x|\leq R} |a^{(n)}(x) - a(x)| = 0,$$

and

$$\lim_{n\to\infty} \sup_{x\in S^{(n)}, |x|\leq R} \gamma_p^{(n)}(x) = 0 \quad \text{for some } p \geq 2.$$

Finally, assume that $Y_0^{(n)} \to x$ as $n \to \infty$. Then $(Y_{\lfloor s/h(n)\rfloor}^{(n)})_{s\geq 0}$ converges weakly to a strong Markov process X that satisfies the SDE (5) with $X_0 = x$.

2.2 The (Standard) Logistic Branching Process

We already discussed in the introduction that the processes that will be relevant here are the so-called logistic branching processes, given by the solution of

$$dX_s = (aX_s - cX_s^2)ds + \sqrt{\gamma X_s}dB_s, \ s > 0,$$

with $X_0 = x \geq 0$, $a \in \mathbb{R}$ and $c, \gamma > 0$, see also (2) (and (1) for the particular case that will appear here). In the remainder of this section we collect some key properties that will be handy. The first one is about the existence and uniqueness of solutions, see [7].

▶ **Lemma 6.** *For all initial states $x \geq 0$ and for all $a \in \mathbb{R}$ and $c, \gamma > 0$, there exists a unique solution $(X_{a,c,\gamma,x}, B_{a,c,\gamma,x}, \mathcal{P}_{a,c,\gamma,x})$ to (2). Moreover, $X_{a,c,\gamma,x}$ is non-negative.*

In what follows it will be convenient to consider a specific choice of the filtered probability space $\mathcal{P}_{a,c,\gamma,x} = (\Omega', \mathcal{F}', (\mathcal{F}_s')_{s\geq 0}, \mathbb{P}')$ (where all components depend on the parameters a, c, γ, x) from the previous lemma that we construct as follows. Let Ω be the space of all continuous maps $[0, \infty) \to \mathbb{R}$ and let X be the coordinate process given by $X_s(\xi) = \xi(s)$ for

all $s \geq 0$ and $\xi \in \Omega$. Additionally, consider the σ-algebra $\mathcal{F} = \sigma\{X_s \mid s \geq 0\}$ and equip the measurable space $(\Omega, \mathcal{F})$ with the filtration $(\mathcal{F}_s)_{s \geq 0}$ given by $\mathcal{F}_s = \sigma\{X_r \mid 0 \leq r \leq s\}$ for all $s \geq 0$, which we may complete and right-continuously extend in order to fulfil the usual conditions. Via the map $\Omega' \ni \xi' \mapsto X_{a,c,\gamma,x}(\xi') \in \Omega$ it is possible to switch from $\mathcal{P}_{a,c,\gamma,x}$ to the canonical probability space $(\Omega, \mathcal{F}, (\mathcal{F}_s)_{s \geq 0}, \mathbb{P}_{a,c,\gamma,x})$, where $\mathbb{P}_{a,c,\gamma,x}$ is the probability measure given by $\mathbb{P}_{a,c,\gamma,x}(A) = \mathbb{P}'((X_{a,c,\gamma,x})^{-1}(A))$ for all $A \in \mathcal{F}$. By this particular choice we obtain that the coordinate process X on $(\Omega, \mathcal{F}, (\mathcal{F}_s)_{s \geq 0}, \mathbb{P}_{a,c,\gamma,x})$ has the same law as $X_{a,c,\gamma,x}$ under $\mathbb{P}'$, i.e. under $\mathbb{P}_{a,c,\gamma,x}$ the process X satisfies (2). The following corollary is now an immediate consequence of Lemma 6, and a similar construction was also performed in [7].

▶ **Corollary 7.** *For all initial states $x \geq 0$ and for all $a \in \mathbb{R}$ and $c, \gamma > 0$, there is a unique solution $(X, B_{a,c,\gamma,x}, \mathcal{P}_{a,c,\gamma,x})$ to (2), where X is the coordinate process and thus independent of a, c, γ, x. Moreover, X is non-negative $\mathbb{P}_{a,c,\gamma,x}$-almost surely, where $\mathbb{P}_{a,c,\gamma,x}$ denotes the probability measure of $\mathcal{P}_{a,c,\gamma,x}$.*

For the rest of this paper we will adopt the above procedure and consider solutions to (2) only with respect to the canonical probability space $(\Omega, \mathcal{F}, (\mathcal{F}_s)_{s \geq 0}, \mathbb{P}_{a,c,\gamma,x})$. Our main object of interest will be the time at which the logistic Feller diffusion X hits zero, which under $\mathbb{P}_{a,c,\gamma,x}$ is given by the stopping time

$$T(\xi) = \inf\{s \geq 0 : \xi(s) = 0\}, \quad \xi \in \Omega.$$

The author of [7] establishes that T is finite $\mathbb{P}_{a,c,\gamma,x}$-almost surely. Moreover, $X_s = 0$ for all $s \geq T$ under $\mathbb{P}_{a,c,\gamma,x}$, i.e. upon hitting zero the process becomes constant, which is why we also refer to T as *absorption time*.

Within our context, it will be necessary to consider solutions to (2) with initial value $x \to \infty$. The required results are covered by the following statement, whose proof can be found in [7] and for which we define the function $\theta : [0, \infty) \to \mathbb{R}$ by

$$\theta(\lambda) := \int_0^\lambda \exp\left(\frac{\gamma}{4c}v^2 - \frac{a}{c}v\right) dv, \quad \lambda \geq 0. \tag{6}$$

▶ **Lemma 8.** *For all $x \geq 0, a \in \mathbb{R}$ and $c, \gamma > 0$, the expectation of T under $\mathbb{P}_{a,c,\gamma,x}$ is finite and*

$$\mathbb{E}_{a,c,\gamma,x}[T] = \frac{1}{c} \int_0^\infty \frac{\theta(\lambda)}{\lambda \theta'(\lambda)}(1 - \exp(-x\lambda))\, d\lambda.$$

In addition, the measures $(\mathbb{P}_{a,c,\gamma,x})_{x \geq 0}$ converge weakly, as $x \to \infty$, to the law $\mathbb{P}_{a,c,\gamma,\infty}$ of the so-called standard logistic branching process. *Under $\mathbb{P}_{a,c,\gamma,\infty}$, the hitting time T is a continuous random variable which is finite almost surely and has finite expectation given by*

$$\mathbb{E}_{a,c,\gamma,\infty}[T] = \sup_{x \geq 0} \mathbb{E}_{a,c,\gamma,x}[T] = \frac{1}{c} \int_0^\infty \frac{\theta(\lambda)}{\lambda \theta'(\lambda)}\, d\lambda.$$

3 Proof Strategy & Some Details

In the following lemma we investigate the early phase of the process. In particular we are interested in the number of steps and the number of jumps until the number of unhappy particles drops to $\Theta(n^{1/2})$.

▶ **Lemma 9.** *Let $\epsilon, \delta > 0, \alpha \in \mathbb{R}$ and $M = n/2 + \alpha\sqrt{n} + o(\sqrt{n}) \in \mathbb{N}$. Let*

$$T_{n,M,\delta} := \inf\left\{t > 0 : U_t \le n^{1/2}/\delta\right\}.$$

Then, for all sufficiently small $\delta > 0$ and all sufficiently large n, with probability at least $1 - \delta$,

$$\left|T_{n,M,\delta} - \frac{4}{7}\delta n^{1/2}\right| \le \epsilon\delta n^{1/2} \quad and \quad \left|\sum_{0 \le t \le T_{n,M,\delta}} U_t - \frac{2}{7}n\ln n\right| \le \epsilon n \ln n.$$

In particular, (roughly) $\frac{4}{7}\delta n^{1/2}$ steps are required to drop below $n^{1/2}/\delta$ unhappy particles, and at this step the accumulated number of unhappy particles, which corresponds to the total number of jumps, is (roughly) $\frac{2}{7}n\ln n$. The lemma is established by considering the number of unhappy particles for a relatively short number of steps, where the change of the process can be precisely controlled by means of martingale concentration, exploiting the subgaussian nature of the increments. We omit the details due to space limitations.

We focus on the late phase, which uses the diffusion approximation toolbox. We write

$$U_{t+1} - U_t = X_{t+1} - Y_{t+1},$$

where X_{t+1} stands for the number of particles that were happy at step t but become unhappy in step $t + 1$ (because some particle which was unhappy at time t moved onto their vertex) and Y_{t+1} is the number of unhappy particles at time t that become happy at step $t + 1$ (because at time $t + 1$ they are alone on the vertex that they occupy). Moreover, define

$$X_{t+1,h} := \mathbb{1}[h \in \mathcal{U}_{t+1}] \quad and \quad Y_{t+1,u} := \mathbb{1}[u \in \mathcal{H}_{t+1}]$$

where $\mathcal{H}_{t+1}/\mathcal{U}_{t+1}$ is the set of happy/unhappy particles at time $t + 1$ and, so that we can write

$$X_{t+1} = \sum_{h \in \mathcal{H}_t} X_{t+1,h} \quad and \quad Y_{t+1} = \sum_{u \in \mathcal{U}_t} Y_{t+1,u}.$$

It is clear that, given U_t, we can compute $\mathbb{E}[X_{t+1,h}], \mathbb{E}[Y_{t+1,u}]$ and $\mathbb{E}[X_{t+1,h}Y_{t+1,u}]$ for any $h \in \mathcal{H}_t$ and $u \in \mathcal{U}_t$; the details are omitted. With this at hand, we then establish asymptotics of the drift and variation for the number of unhappy particles, which we describe in the following two lemmas.

▶ **Lemma 10.** *Let $\varepsilon = \varepsilon(n) = o(1)$, $u : \mathbb{N} \to \mathbb{N}$ and $M = M(n) := (1 + \varepsilon)n/2 \in \mathbb{N}$. Then, uniformly,*

$$\mathbb{E}\big[U_{t+1} - U_t \mid U_t = u\big] = \varepsilon u - \frac{u^2}{n}\left(\frac{7}{4} + \frac{3\varepsilon}{4}\right) + O\left(\frac{u}{n} + \frac{u^3}{n^2}\right).$$

▶ **Lemma 11.** *Let $\varepsilon = \varepsilon(n) = o(1)$ and $u : \mathbb{N} \to \mathbb{N}$ be such that $u = o(n^{2/3})$ and $M = M(n) := (1 + \varepsilon)n/2 \in \mathbb{N}$. Then, uniformly,*

$$\mathbb{E}\big[(U_{t+1} - U_t)^2 \mid U_t = u\big] = u + o(\varepsilon u^2 + u).$$

To continue we introduce the (continuous) time-shifted process

$$U'_s := U_{\lfloor s \rfloor + T_{n,M,\delta}}, \quad s \ge 0.$$

By applying Theorem 5 we will show that $(n^{-1/2}U'_{s\sqrt{n}})_{s \ge 0}$ converges weakly to a diffusion. Note that the following lemma is just a reformulation of Lemma 2 in the Introduction, as (7) corresponds to the SDE (1).

▶ **Lemma 12.** *Let $\delta > 0$. As $n \to \infty$, the process $(n^{-1/2}U'_{s\sqrt{n}})_{s\geq 0}$ converges weakly to a process X that satisfies*

$$dX_s = \left(2\alpha X_s - \frac{7}{4}X_s^2\right)ds + \sqrt{X_s}dB_s, \quad s > 0, \quad and \quad X_0 = \delta^{-1}. \tag{7}$$

Proof. We will apply Theorem 5 with $h = h(n) = n^{-1/2}$ and $Y_t^{(n)} := n^{-1/2}U'_t$ for $t \in \mathbb{N}_0$. First, note that it is necessary to extend the SDE (7) in a way that it has a unique solution not only for all initial values $x \geq 0$, but for all $x \in \mathbb{R}$. To this end, write $a^+ = \max\{a, 0\}$ for $a \in \mathbb{R}$ and consider the SDE

$$dX_s = \left(2\alpha X_s^+ - \frac{7}{4}(X_s^+)^2\right)ds + \sqrt{X_s^+}dB_s, \quad s > 0, \text{ with } X_0 = x \in \mathbb{R}. \tag{8}$$

Note that if the initial value x is negative, then $X = x$ uniquely satisfies this SDE. For $x \geq 0$, recall that Corollary 7 guarantees the existence of a unique solution $(X, B_{2\alpha,7/4,1,x}, \mathcal{P}_{2\alpha,7/4,1,x})$ to (7) with $X_0 = x$ and such that $X \geq 0$ almost surely. Hence, if $x \geq 0$, (8) coincides with (7) with initial value $X_0 = x$ and we conclude that (8) possesses a unique solution for all $x \in \mathbb{R}$.

Next, we employ Lemmas 10 and 11 with $\varepsilon(n) = 2\alpha n^{-1/2} + o(n^{-1/2})$, as $M = n/2 + \alpha n^{1/2} + o(n^{1/2})$. For this purpose, let $R < \infty$ and consider $x \in S^{(n)} \subseteq \{0, n^{-1/2}, 2n^{-1/2}, ..., n^{1/2}\}$ with $|x| \leq R$. Then, Lemma 10 with $u = xn^{1/2}$ implies that

$$b^{(n)}(x) = \frac{\mathbb{E}\left[n^{-1/2}U'_{t+1} - n^{-1/2}U'_t \mid n^{-1/2}U'_t = x\right]}{n^{-1/2}} = 2\alpha x - \frac{7}{4}x^2 + o\left(R + R^3\right).$$

Further, as $xn^{1/2} = o(n^{2/3})$ due to $|x| \leq R$, it follows from Lemma 11 with $u = xn^{1/2}$ that

$$a^{(n)}(x) = \frac{\mathbb{E}\left[(n^{-1/2}U'_{t+1} - n^{-1/2}U'_t)^2 \mid n^{-1/2}U'_t = x\right]}{n^{-1/2}} = x + o\left(R^2\right). \tag{9}$$

We therefore obtain that for any $R < \infty$

$$\lim_{n\to\infty} \sup_{x \in S^{(n)}, |x| \leq R} \left|b^{(n)}(x) - \left(2\alpha x - \frac{7}{4}x^2\right)\right| = 0 \text{ and } \lim_{n\to\infty} \sup_{x \in S^{(n)}, |x| \leq R} |a^{(n)}(x) - x| = 0. \tag{10}$$

Moreover, we show

$$\lim_{n\to\infty} \sup_{x \in S^{(n)}, |x| \leq R} |\gamma_3^{(n)}(x)| = 0, \tag{11}$$

and

$$U'_0 = U_{T_{n,M,\delta}} \sim n^{1/2}/\delta \text{ with probability } 1 - o(1). \tag{12}$$

The last two facts, whose proof is omitted here, together with (10) and the existence of a unique solution to (8) guarantee that we can apply Theorem 5 to conclude that $(n^{-1/2}U'_{s\sqrt{n}})_{s\geq 0}$ converges weakly to a process X that satisfies (7) with $X_0 = 1/\delta$, and the proof is finished. ◀

Recall from Corollary 7 that $(X, B_{2\alpha,7/4,1,x}, \mathcal{P}_{2\alpha,7/4,1,x})$ represents a solution of (7) with initial value $x \geq 0$ and that the corresponding hitting time of zero is given by

$$T = \inf\{s \geq 0 : X_s = 0\} \tag{13}$$

under the probability measure $\mathbb{P}_{2\alpha,7/4,1,x}$. The next statement asserts that $n^{-1/2}T'_{n,M,\delta}$, where $T'_{n,M,\delta} := T_{n,M} - T_{n,M,\delta}$, converges in distribution to T under $\mathbb{P}_{2\alpha,7/4,1,1/\delta}$.

▶ **Lemma 13.** *Let $\delta > 0$. Then, as $n \to \infty$, $n^{-1/2}T'_{n,M,\delta} \overset{d}{\to} T$, with T given by (13) under the probability measure $\mathbb{P}_{2\alpha,7/4,1,1/\delta}$.*

Proof. Let $s \geq 0$ and recall from Section 2.1 that, under $\mathbb{P}_{2\alpha,7/4,1,1/\delta}$, $X_s = 0$ is equivalent to $T \leq s$. Similarly, $U_{\lfloor s \rfloor} = 0$ if and only if $T_{n,M} \leq s$, from which we obtain that $T'_{n,M,\delta} = T_{n,M} - T_{n,M,\delta} \leq s$ if and only if $U'_s = U_{\lfloor s \rfloor + T_{n,M,\delta}} = 0$. Hence,

$$\mathbb{P}_{2\alpha,7/4,1,1/\delta}(X_s = 0) = \mathbb{P}_{2\alpha,7/4,1,1/\delta}(T \leq s) \quad \text{and} \quad \mathbb{P}(U'_s = 0) = \mathbb{P}(T'_{n,M,\delta} \leq s).$$

As Lemma 12 entails $\lim_{n\to\infty} \mathbb{P}\big(n^{-1/2}U'_{s\sqrt{n}} = 0\big) = \mathbb{P}_{2\alpha,7/4,1,1/\delta}(X_s = 0)$, we conclude

$$\lim_{n\to\infty} \mathbb{P}(n^{-1/2}T'_{n,M,\delta} \leq s) = \mathbb{P}_{2\alpha,7/4,1,1/\delta}(T \leq s). \qquad \blacktriangleleft$$

With the convergence in distribution shown in the previous lemma at hand, we are now in the position to prove Theorem 1.

Proof of Theorem 1. Recall from Lemma 9 that $T_{n,M,\delta} \leq \delta n^{1/2}$ with probability at least $1 - \delta$ for δ sufficiently small and n large enough. As $T_{n,M,\delta}$ is non-negative, this implies that

$$n^{-1/2}T'_{n,M,\delta} \leq n^{-1/2}T_{n,M} \leq \delta + n^{-1/2}T'_{n,M,\delta} \text{ with probability at least } 1 - \delta.$$

Applying Lemma 13, we therefore obtain that for all $s \geq 0$

$$\lim_{n\to\infty} \mathbb{P}(n^{-1/2}T_{n,M} \leq s) \leq \lim_{n\to\infty} \mathbb{P}(n^{-1/2}T'_{n,M,\delta} \leq s) = \mathbb{P}_{2\alpha,7/4,1,1/\delta}(T \leq s)$$

and

$$\lim_{n\to\infty} \mathbb{P}\big(n^{-1/2}T_{n,M} \geq s\big) \leq \lim_{n\to\infty} \mathbb{P}\big(n^{-1/2}T_{n,M} \geq s, T_{n,M,\delta} \leq \delta n^{1/2}\big) + \delta$$
$$\leq \mathbb{P}_{2\alpha,7/4,1,1/\delta}\big(T \geq s - \delta\big) + \delta.$$

Note that $\mathbb{P}_{2\alpha,7/4,1,x_1}(T \geq \tau) \geq \mathbb{P}_{2\alpha,7/4,1,x_2}(T \geq \tau)$ for all $x_1 > x_2 \geq 0$ and $\tau \geq 0$, since X is almost surely continuous and needs a positive and finite amount of time to drop from x_1 to x_2. So, since according to Lemma 8 we have $\lim_{\delta\to 0} \mathbb{P}_{2\alpha,7/4,1,1/\delta} = \mathbb{P}_{2\alpha,7/4,1,\infty}$ and T is continuous, it therefore follows that

$$\lim_{\delta\to 0} \mathbb{P}_{2\alpha,7/4,1,1/\delta}\big(T \geq s - \delta\big) \leq \lim_{\delta\to 0} \mathbb{P}_{2\alpha,7/4,1,\infty}\big(T \geq s - \delta\big) = \mathbb{P}_{2\alpha,7/4,1,\infty}\big(T \geq s\big),$$

which yields

$$\lim_{n\to\infty} \mathbb{P}(n^{-1/2}T_{n,M} \leq s) = \mathbb{P}_{2\alpha,7/4,1,\infty}(T \leq s), \quad s \geq 0.$$

Thus, $n^{-1/2}T_{n,M} \overset{d}{\to} T_\alpha$, where T_α satisfies

$$\mathbb{P}(T_\alpha \leq s) = \mathbb{P}_{2\alpha,7/4,1,\infty}(T \leq s), \quad s \geq 0.$$

Moreover, $\mathbb{P}_{2\alpha,7/4,1,\infty}(T > 0) = 1$ implies that T_α is positive almost surely, and this completes the proof of Theorem 1. $\blacktriangleleft$

References

1 K. Bringmann, F. Kuhn, K. Panagiotou, U. Peter, and H. Thomas. Internal DLA: Efficient simulation of a physical growth model. In *Automata, Languages, and Programming*, pages 247–258, Berlin, Heidelberg, 2014. Springer Berlin Heidelberg.

2 C. Cooper, A. McDowell, T. Radzik, N. Rivera, and T. Shiraga. Dispersion processes. *Random Structures Algorithms*, 53(4):561–585, 2018. `doi:10.1002/rsa.20822`.

3 U. De Ambroggio, T. Makai, and K. Panagiotou. Dispersion on the complete graph, 2023. An extended abstract appeared in the *Proceedings of EUROCOMB '23*. `arXiv:2306.02474`.

4 P. Diaconis and W. Fulton. A growth model, a game, an algebra, Lagrange inversion, and characteristic classes. *Rend. Sem. Mat. Univ. Politec. Torino*, 49(1):95–119, 1991.

5 R. Durrett. *Stochastic calculus – a practical introduction*. Probability and Stochastics Series. CRC Press, Boca Raton, FL, 1996.

6 A. Frieze and W. Pegden. A note on dispersing particles on a line. *Random Structures Algorithms*, 53(4):586–591, 2018. `doi:10.1002/rsa.20821`.

7 A. Lambert. The branching process with logistic growth. *Ann. Appl. Probab.*, 15(2):1506–1535, 2005. `doi:10.1214/105051605000000098`.

8 G. F. Lawler, M. Bramson, and D. Griffeath. Internal diffusion limited aggregation. *Ann. Probab.*, 20(4):2117–2140, 1992.

9 L. T. Rolla. Activated Random Walks on $\mathbb{Z}^d$. *Probability Surveys*, 17:478–544, 2020. `doi:10.1214/19-PS339`.

10 Y. Shang. Longest distance of a non-uniform dispersion process on the infinite line. *Inform. Process. Lett.*, 164:106008, 5, 2020. `doi:10.1016/j.ipl.2020.106008`.

Asymptotic Enumeration of Rooted Binary Unlabeled Galled Trees with a Fixed Number of Galls

Lily Agranat-Tamir[1] ✉
Department of Biology, Stanford University, CA, USA

Michael Fuchs ✉
Department of Mathematical Sciences, National Chengchi University, Taipei, Taiwan

Bernhard Gittenberger ✉
Department of Discrete Mathematics and Geometry, Technische Universität Wien, Austria

Noah A. Rosenberg ✉
Department of Biology, Stanford University, CA, USA

Abstract

Galled trees appear in problems concerning admixture, horizontal gene transfer, hybridization, and recombination. Building on a recursive enumerative construction, we study the asymptotic behavior of the number of rooted binary unlabeled (normal) galled trees as the number of leaves n increases, maintaining a fixed number of galls g. We find that the exponential growth with n of the number of rooted binary unlabeled normal galled trees with g galls has the same value irrespective of the value of $g \geq 0$. The subexponential growth, however, depends on g; it follows $c_g n^{2g-3/2}$, where c_g is a constant dependent on g. Although for each g, the exponential growth is approximately 2.4833^n, summing across *all* g, the exponential growth is instead approximated by the much larger 4.8230^n.

1 Introduction

Rooted binary trees are a staple of mathematical phylogenetic analysis, as they are used to represent diverse biological processes taking place in time – including the evolution of species, the evolution of genes among those species, and the divergence of populations [9, 21, 24]. The root represents a common ancestor, and the leaves represent subsequent biological entities, often in the present day. Viewed as objects evolving in time, by extension of "vertical" inheritance that occurs in genetic transmission from parents to offspring, biological divergences are viewed as taking place vertically on the tree. Mathematical phylogenetic analyses of trees have produced rich contributions to algorithmic and combinatorial studies.

Certain evolutionary events, however, involve *merging* rather than *divergence* of biological lineages. Such events include the recombination that occurs during gamete formation, population admixture, horizontal gene transfer, and hybridization. To describe processes that include these events, we must look beyond trees to *phylogenetic networks* [14, 17, 18].

---

[1] corresponding author
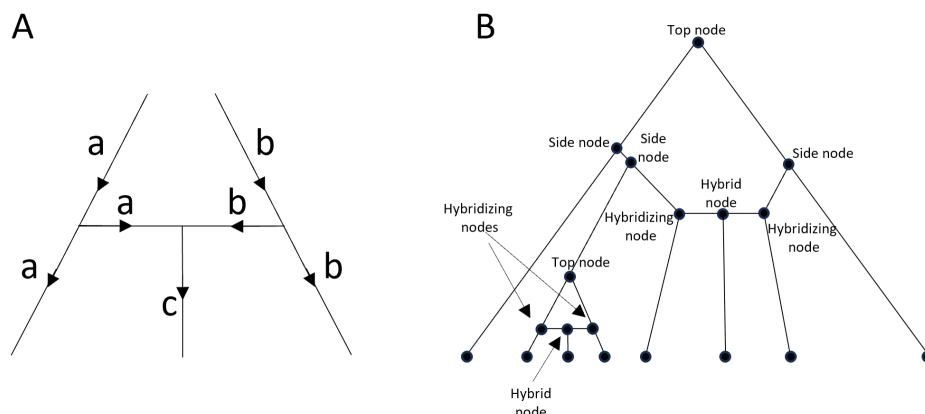
Figure 1 Features of a gall in a galled tree. (A) A gall as a representation of a biological merging event. Biological lineages a and b each bifurcate, with one branch of each bifurcation merging to form lineage c. (B) Nomenclature for the various nodes in a gall.

Among the phylogenetic networks, galled trees are some of the simplest. As their name suggests, they are tree-like, but they can contain certain internal nodes with in-degree 2 and out-degree 1, representing permitted classes of mergings. Galled trees are named for the growths, termed *galls*, which appear in plants but which do not disrupt their tree-like structure. They were first introduced in the study of recombination [15, 16, 23].

Mathematically, a galled tree allows each vertex or edge in a graph to be contained in at most one cycle. An additional requirement is needed for galled trees to be meaningful for biological processes such as hybridization. In a hybridization event, two biological lineages, a and b, each bifurcate; a merging event occurs between two branches, one from each bifurcation, producing a third lineage, c (Figure 1A). The structure of the event requires that when viewed graphically, a gall – a cycle in the graph – contains at least four nodes. These include a *top node*, two *hybridizing nodes*, and one *hybrid node*. Additional *side nodes* are permitted, and we regard the hybridizing nodes as special side nodes (Figure 1B). The requirement that galls have at least these four nodes (i.e. the top node must not be a hybridizing node) is equivalent to a requirement that galled trees be *normal*.

Many enumerative problems on galled trees have been investigated [3, 4, 5, 22]; this study concerns rooted binary unlabeled normal galled (non-plane) trees. Given number of galls g, as the number of leaves $n \to \infty$, what is the growth of the size of this class? The case of $g = 0$ is the enumeration of rooted binary unlabeled trees, and we previously studied $g = 1$ [1]. Building on a recurrence for rooted binary unlabeled normal galled trees with n leaves and g galls, we obtain a generating function for $g = 2$. We find the asymptotic behavior of the number of trees with n leaves and $g = 2$ galls, and we obtain asymptotics for each $g > 2$. In our main result, Theorem 10, we report that the number of galled trees with n leaves and g galls has the form $\beta_g n^{2g - \frac{3}{2}} \rho^{-n}$, where ρ is the radius of convergence of the generating function for the $g = 0$ case, and β_g is a constant that depends solely on g.

2 Definitions

We define our concepts formally. We assume that all networks and trees are binary; we usually drop the term *binary*. A *rooted phylogenetic network* is a directed acyclic graph in which four properties hold. (i) There exists a unique node with in-degree 0 and out-degree 2. This node
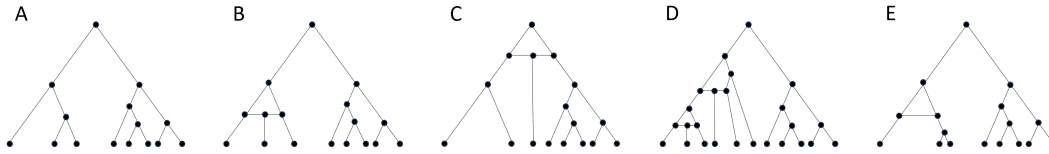
Figure 2 Rooted binary unlabeled galled trees. (A) A tree with no galls. (B) A galled tree with one gall. (C) A galled tree with a root gall. (D) A galled tree with two galls. (E) A galled tree that is not a normal galled tree and that is not included in the class of galled trees that we enumerate.

is the *root node*. (ii) *Leaf nodes* possess in-degree 1 and out-degree 0. (iii) Non-leaf, non-root nodes possess in-degree 2 and out-degree 1 or in-degree 1 and out-degree 2. (iv) Edges are directed away from the root. Nodes with in-degree 2 and out-degree 1 are *reticulation nodes* (or *hybrid nodes*). Nodes with in-degree 1 and out-degree 2 are *tree nodes*.

A *rooted galled tree* is a rooted phylogenetic network with three additional properties. (v) Each reticulation node a_r has a unique ancestor node r so that exactly two non-overlapping paths of edges connect r to a_r. Ignoring the direction of the edges, the two paths from r to a_r produce a cycle C_r. The cycle is termed a *gall*. (vi) Consider galls C_r and C_s, associated with reticulation nodes a_r and a_s, $a_r \neq a_s$. The sets of nodes in the galls C_r and C_s are disjoint. (vii) Ancestor node r and reticulation node a_r are separated by two or more edges. Condition (vii) encodes the requirement that we consider only *normal* galled trees (Figure 2).

We generally drop the terms *rooted* and *normal*, and refer only to *galled trees*, and where a distinction is necessary, *labeled* and *unlabeled* galled trees. Although a galled tree is not technically a tree due to the presence of cycles, we continue to refer to galled trees as trees. We similarly refer to the galled trees rooted at internal nodes of a galled tree as *subtrees*. Our view of galls as representations of biological merging events leads us to depict hybridizing nodes and their associated hybrid node on a horizontal line, representing the simultaneity of these nodes when a galled tree is taken to represent a structure evolving in time [2, 20].

A basic result describes the maximal number of galls possible in a galled tree with n leaves. A gall contains three or more descendant subtrees: one from the reticulation node, two from the hybridizing nodes, and one for each additional side node. Hence, the smallest galled tree possesses $n = 3$ leaves. Adding a gall to a galled tree involves replacing one subtree with at least three subtrees, so that each gall adds at least two leaves. For a tree with g galls, the number of leaves satisfies $n \geq 2g + 1$, or $g \leq \lfloor \frac{n-1}{2} \rfloor$ [20].

We will need to consider *compositions*, ordered lists of positive integers that sum to a specified value. We denote by $C(a, b)$ the compositions of a natural number a into b parts. $C(a, b)$ is the set of ordered lists of positive integers of length b, $(i_1, i_2, \ldots, i_b)$, with sum equal to a. We denote by $C_p(a, b)$ the subset of $C(a, b)$ containing the *palindromic* compositions of a, that is, the compositions $(i_1, i_2, \ldots, i_b)$ for which $i_j = i_{b-j+1}$ for each j from 1 to b.

3 Previous work

We review a number of results. The rooted binary unlabeled galled trees generalize the rooted binary unlabeled trees without galls. Letting U_n denote the number of rooted binary unlabeled trees with no galls and letting $\mathcal{U}(t)$ denote the generating function $\sum_{n \geq 0} U_n t^n$,

$$\mathcal{U}(t) = t + \frac{1}{2}\mathcal{U}^2(t) + \frac{1}{2}\mathcal{U}(t^2). \tag{1}$$

Denoting the radius of convergence by ρ, as $t \to \rho^-$, we have $\mathcal{U}(t) \sim 1 - \gamma\sqrt{1 - t/\rho}$, where $\gamma \approx 1.1300$ and $\rho \approx 0.4027$ [8, p. 55] [10, pp. 476-477]. The asymptotic approximation for the number of rooted binary unlabeled trees (with no galls) is,

$$U_n = [t^n]\mathcal{U}(t) \sim \frac{\gamma}{2\Gamma(\frac{1}{2})}n^{-\frac{3}{2}}\rho^{-n}. \tag{2}$$

In our previous work on rooted binary unlabeled normal galled trees [1] (henceforth "unlabeled galled trees"), we obtained a recursion enumerating the A_n unlabeled galled trees with n leaves and another recursion enumerating the $E_{n,g}$ unlabeled galled trees with a specified number of galls g. We specifically considered the case of $g = 1$. We also studied the asymptotics of A_n and $E_{n,1}$ through their generating functions. The generating function for unlabeled galled trees, considering all possible numbers of galls, was found to be [1, eq. 36]

$$\mathcal{A}(t) = t + \frac{1}{2}\mathcal{A}^2(t) + \frac{1}{2}\mathcal{A}(t^2) + 1 - \frac{1}{1 - \mathcal{A}(t)} + \frac{\mathcal{A}(t)}{2[1 - \mathcal{A}(t)]^2} + \frac{\mathcal{A}(t)}{2[1 - \mathcal{A}(t^2)]}. \tag{3}$$

The three leftmost terms, identical to the generating function $\mathcal{U}(t)$ (eq. (1)), arise from the galled trees in which two subtrees descend immediately from the root. The other terms arise from galled trees with a gall that contains the root, a *root gall*.

Using the *asymptotics of implicit tree-like classes* theorem [10, pp. 467-468], we obtained the asymptotics of the number of galled trees with n leaves, A_n [1, eq. 42]: $A_n = [t^n]\mathcal{A}(t) \sim \left[\delta/\left(2\Gamma(\frac{1}{2})\right)\right]n^{-\frac{3}{2}}\alpha^{-n}$, where $\delta \approx 0.2793$ and $\alpha \approx 0.2073$. $\mathcal{A}(t)$ has convergence radius about half that of $\mathcal{U}(t)$, so that galled trees are much more numerous than the trees without galls.

We also derived the generating function $\mathcal{E}_1(t)$ and asymptotic growth of the number of unlabeled galled trees with exactly one gall. We state these results as propositions.

▶ **Proposition 1** ([1], eq. 48). *The generating function $\mathcal{E}_1(t)$ for the number of unlabeled galled trees with 1 gall satisfies*

$$\mathcal{E}_1(t) = \frac{1}{1 - \mathcal{U}(t)} - \frac{1}{[1 - \mathcal{U}(t)]^2} + \frac{\mathcal{U}(t)}{2[1 - \mathcal{U}(t)]^3} + \frac{\mathcal{U}(t)}{2[1 - \mathcal{U}(t)][1 - \mathcal{U}(t^2)]}. \tag{4}$$

▶ **Proposition 2** ([1], eq. 50). *The asymptotic growth of the number $E_{n,1}$ of unlabeled galled trees with n leaves and 1 gall satisfies*

$$E_{n,1} \sim \frac{1}{2\gamma^3\Gamma(\frac{3}{2})}n^{\frac{1}{2}}\rho^{-n} = \frac{1}{\gamma^3\sqrt{\pi}}n^{\frac{1}{2}}\rho^{-n}. \tag{5}$$

Proposition 2 follows from the fact that as $t \to \rho^-$, $\mathcal{E}_1(t) \sim 1/[2\gamma^3(1 - t/\rho)^{\frac{3}{2}}]$. $\mathcal{E}_1(t)$ in eq. (4) depends on $\mathcal{U}(t)$. Eq. (5) clarifies that the exponential growth of the number of unlabeled galled trees with one gall is the same as that of the number of unlabeled galled trees with no galls; only the subexponential growth differs. We will generalize this result.

4 Recursion

4.1 Recursion for g galls, $E_{n,g}$

In [1, eq. 27], we obtained a recursion for $E_{n,g}$, the number of unlabeled galled trees with n leaves and exactly g galls; Table 3 reported the numerical values $E_{n,g}$ up to $n = 18$. The base cases are $E_{1,0} = 1$ and $E_{1,g} = 0$ for $g \geq 1$. We also write $E_{m,\ell} = 0$ when m is not a positive integer, ℓ is not a positive integer, or both.

▶ **Proposition 3.** *For (n,g) with $n \geq 2$ and $0 \leq g \leq \lfloor \frac{n-1}{2} \rfloor$, the number of unlabeled galled trees with n leaves and g galls is*

$$E_{n,g} = \frac{1}{2}\Bigg[\Bigg(\sum_{\mathbf{c}\in C(n,2)} \sum_{\mathbf{d}\in C(g+2,2)} \prod_{i=1}^{2} E_{c_i,d_i-1}\Bigg) + E_{\frac{n}{2},\frac{g}{2}} \tag{6}$$

$$+ \Bigg(\sum_{k=3}^{n}(k-2)\sum_{\mathbf{c}\in C(n,k)} \sum_{\mathbf{d}\in C(g-1+k,k)} \prod_{i=1}^{k} E_{c_i,d_i-1}\Bigg) \tag{7}$$

$$+ \Bigg(\sum_{a=1}^{\lfloor\frac{n-1}{2}\rfloor} \sum_{\mathbf{c}\in C_p(n,2a+1)} \sum_{\mathbf{d}\in C_p\big(g-1+(2a+1),2a+1\big)} \prod_{i=1}^{a+1} E_{c_i,d_i-1}\Bigg)\Bigg]. \tag{8}$$

The approach is to use a recursion at the root node. We sum over all products of possible counts of subtrees, each with fewer than n leaves. Pairs of galled trees that are reflections of one another over the root – or the axis connecting the top node to the reticulation node of the root gall – are the same unlabeled galled tree, explaining the leading $\frac{1}{2}$. We add back terms for galled trees that are symmetric over the root, which are not double-counted.

Line (6) in Proposition 3 enumerates galled trees with n leaves and g galls that do not have a root gall. The first term traverses combinations of numbers of leaves in the two subtrees summing to n by traversing compositions $\mathbf{c}$ of n into 2 parts $\big(\mathbf{c} \in C(n,2)\big)$. It also traverses combinations of placements of the g galls in the two subtrees. Because subtrees can possess 0 galls, these combinations are identified from compositions of $g+2$ into 2 parts, subtracting 1 gall in each part $\big(\mathbf{d} \in C(g+2,2)\big)$. The second term adds back the galled trees with identical subtrees; this term is nonzero only if both n and g are even.

Line (7) counts galled trees with n leaves and g galls that do have a root gall. It traverses the possible number k of subtrees descending from side nodes, hybridizing nodes, and the hybrid node of the root gall (3 to n, the number of leaves). It then traverses the $k-2$ possible nodes in the root gall where the hybrid node can be placed: all k nodes except immediate descendants of the root. We then traverse the possible combinations of the n leaves and $g-1$ remaining (non-root) galls into the k subtrees, again allowing subtrees with no galls.

Line (8) adds back half the galled trees with n leaves and g galls that have a root gall and that are symmetric over the reticulation node. Here, a is the possible number of subtrees of the root gall on each side of the reticulation node, so that the root gall has $2a+1$ subtrees in total. The composition of the n leaves into $2a+1$ subtrees and the composition of the $g-1$ galls into those subtrees are both palindromic. Given these compositions, a tree is specified by its subtrees of one side of the reticulation node and the subtree of the reticulation node.

4.2 Recursion for two galls, $E_{n,2}$

For $g=2$, for $n \geq 2$, the recursion for $E_{n,g}$ becomes

$$E_{n,2} = \frac{1}{2}\Bigg[\Bigg(\sum_{c=1}^{n-1}\sum_{d=0}^{2} E_{c,d}E_{n-c,2-d}\Bigg) + E_{\frac{n}{2},1}$$

$$+ \sum_{k=3}^{n}(k-2)\sum_{\mathbf{c}\in C(n,k)} \sum_{\mathbf{d}\in C(k+1,k)} \prod_{i=1}^{k} E_{c_i,d_i-1}$$

$$+ \sum_{a=1}^{\lfloor\frac{n-1}{2}\rfloor} \sum_{\mathbf{c}\in C_p(n,2a+1)} \sum_{\mathbf{d}\in C_p(2a+2,2a+1)} \prod_{i=1}^{a+1} E_{c_i,d_i-1}\Bigg]$$

$$
= \frac{1}{2} \Bigg[\bigg(2 \sum_{m=1}^{n-1} U_m E_{n-m,2} + \sum_{m=1}^{n-1} E_{m,1} E_{n-m,1} \bigg) + E_{\frac{n}{2},1}
$$

$$
+ \sum_{k=3}^{n} (k-2) \sum_{m=k-1}^{n-1} \sum_{\mathbf{c} \in C(m,k-1)} \bigg(\prod_{i=1}^{k-1} U_{c_i} \bigg) k E_{n-m,1}
$$

$$
+ \sum_{a=1}^{\lfloor \frac{n-1}{2} \rfloor} \sum_{\mathbf{c} \in C_p(n,2a+1)} \bigg(\prod_{i=1}^{a} U_{c_i} \bigg) E_{c_{a+1},1} \Bigg]. \tag{9}
$$

Recall here that $E_{m,1} = 0$ if $m \notin \mathbb{N}$. In the first line, m gives the number of leaves in the "left" subtree of the root and $n - m$ is the number in the "right" subtree (the left–right distinction is solely for convenience, as we consider non-plane trees, in which the particular embedding of a tree in the plane is disregarded). In the second line, k is the number of subtrees of the root gall, m is the number of leaves across the $k - 1$ subtrees of the root gall that *do not* contain a gall, and $n - m$ is the number of leaves in the subtree with the second gall.

5 Analysis of $E_{n,2}$

5.1 Generating function

Using the recursion in eq. (9), we now find the generating function of $E_{n,2}$, which we define by $\mathcal{E}_2(t) = \sum_{n \geq 0} E_{n,2} t^n$. Eq. (9) holds for all $n \geq 0$ because $E_{n,2} = 0$ for $n \leq 4$ and $E_{n,1} = 0$ for $n \leq 2$. We can add terms involving U_0, $E_{0,1}$, and $E_{0,2}$, all of which equal zero. Then

$$
\mathcal{E}_2(t) = \sum_{n \geq 0} E_{n,2} t^n = \frac{1}{2} \Bigg[\underbrace{\sum_{n \geq 0} \bigg(\Big(2 \sum_{m=0}^{n} U_m E_{n-m,2} \Big) + \Big(\sum_{m=0}^{n} E_{m,1} E_{n-m,1} \Big) + E_{\frac{n}{2},1} \bigg) t^n}_{\mathcal{E}_{2_{\mathrm{i}}}(t)}
$$

$$
+ \underbrace{\sum_{n \geq 0} \bigg(\sum_{k=3}^{n} (k-2) k \sum_{m=k-1}^{n-1} \sum_{\mathbf{c} \in C(m,k-1)} \bigg(\prod_{i=1}^{k-1} U_{c_i} \bigg) E_{n-m,1} \bigg) t^n}_{\mathcal{E}_{2_{\mathrm{ii}}}(t)}
$$

$$
+ \underbrace{\sum_{n \geq 0} \bigg(\sum_{a=1}^{\lfloor \frac{n-1}{2} \rfloor} \sum_{\mathbf{c} \in C_p(n,2a+1)} \bigg(\prod_{i=1}^{a} U_{c_i} \bigg) E_{c_{a+1},1} \bigg) t^n}_{\mathcal{E}_{2_{\mathrm{iii}}}(t)} \Bigg]. \tag{10}
$$

We now simplify the three terms of $\mathcal{E}_2(t)$:

$$
\mathcal{E}_{2_{\mathrm{i}}}(t) = 2 \sum_{m \geq 0} \sum_{n \geq m} (U_m t^m)(E_{n-m,2} t^{n-m}) + \sum_{m \geq 0} \sum_{n \geq m} (E_{m,1} t^m)(E_{n-m,1} t^{n-m}) + \sum_{n \geq 0} E_{\frac{n}{2},1} t^n
$$

$$
= 2 \sum_{m \geq 0} (U_m t^m) \sum_{\ell \geq 0} (E_{\ell,2} t^\ell) + \sum_{m \geq 0} (E_{m,1} t^m) \sum_{\ell \geq 0} (E_{\ell,1} t^\ell) + \sum_{n \geq 0} E_{n,1} t^{2n}
$$

$$
= 2 \mathcal{U}(t) \mathcal{E}_2(t) + \mathcal{E}_1^2(t) + \mathcal{E}_1(t^2). \tag{11}
$$

For $\mathcal{E}_{2_{\mathrm{ii}}}(t)$, we obtain

$$
\begin{aligned}
\mathcal{E}_{2_{\mathrm{ii}}}(t) &= \sum_{k\geq 3}(k-2)k \sum_{m\geq k-1}\sum_{\mathbf{c}\in C(m,k-1)}\prod_{i=1}^{k-1}U_{c_i}t^{c_i}\sum_{n\geq m}E_{n-m,1}t^{n-m} \\
&= \sum_{k\geq 3}(k-2)k\sum_{i_1\geq 0}\sum_{i_2\geq 0}\cdots\sum_{i_{k-1}\geq 0}U_{i_1}U_{i_2}\cdots U_{i_{k-1}}t^{i_1+i_2+\ldots+i_{k-1}}\sum_{\ell\geq 0}E_{\ell,1}t^\ell \\
&= \sum_{k\geq 3}(k-2)k\,\mathcal{U}^{k-1}(t)\,\mathcal{E}_1(t)=\mathcal{E}_1(t)\Big[\sum_{k\geq 2}(k^2-1)\mathcal{U}^k(t)\Big] \\
&= \mathcal{E}_1(t)\Big[\Big(\sum_{k\geq 0}k^2\mathcal{U}^k(t)\Big)-\mathcal{U}(t)-\Big(\sum_{k\geq 0}\mathcal{U}^k(t)\Big)+1+\mathcal{U}(t)\Big] \\
&= \mathcal{E}_1(t)\Big[\frac{\mathcal{U}(t)+\mathcal{U}^2(t)}{[1-\mathcal{U}(t)]^3}-\frac{1}{1-\mathcal{U}(t)}+1\Big]. \tag{12}
\end{aligned}
$$

Finally, $\mathcal{E}_{2_{\mathrm{iii}}}(t)$ becomes

$$
\begin{aligned}
\mathcal{E}_{2_{\mathrm{iii}}}(t) &= \sum_{a\geq 1}\sum_{m\geq a}\sum_{\mathbf{c}\in C(m,a)}\prod_{i=1}^{a}U_{c_i}t^{2c_i}\sum_{n\geq 2m}E_{n-2m,1}t^{n-2m} \\
&= \sum_{a\geq 1}\sum_{i_1\geq 0}\sum_{i_1\geq 0}\cdots\sum_{i_a\geq 0}U_{i_1}U_{i_2}\cdots U_{i_a}t^{2i_1+2i_2+\ldots+2i_a}\sum_{\ell\geq 0}E_{\ell,1}t^\ell \\
&= \sum_{a\geq 1}\mathcal{U}^a(t^2)\,\mathcal{E}_1(t)=\frac{\mathcal{E}_1(t)}{1-\mathcal{U}(t^2)}-\mathcal{E}_1(t). \tag{13}
\end{aligned}
$$

Summing the three parts, we obtain the following proposition.

▶ **Proposition 4.** *The generating function $\mathcal{E}_2(t)$ for the number of unlabeled galled trees with 2 galls satisfies*

$$
\mathcal{E}_2(t)=\frac{\mathcal{E}_1(t)}{2[1-\mathcal{U}(t)]}\left[\mathcal{E}_1(t)+\frac{\mathcal{U}(t)+\mathcal{U}^2(t)}{[1-\mathcal{U}(t)]^3}-\frac{1}{1-\mathcal{U}(t)}+\frac{1}{1-\mathcal{U}(t^2)}\right]+\frac{\mathcal{E}_1(t^2)}{2[1-\mathcal{U}(t)]}. \tag{14}
$$

5.2 Asymptotic analysis

To analyze the asymptotics of $\mathcal{E}_2(t)$ as $t\to\rho^-$, we take the highest-order terms in Proposition 4, that is, the terms with the highest power of $1-t/\rho$ in the denominator. We recall $\mathcal{U}(t)\sim 1-\gamma\sqrt{1-t/\rho}$. From Proposition 1, $\mathcal{E}_1(t)\sim 1/[2\gamma^3(1-t/\rho)^{\frac{3}{2}}]$. We have:

$$
\mathcal{E}_2(t)\sim\frac{\mathcal{E}_1^2(t)}{2[1-\mathcal{U}(t)]}+\frac{2\mathcal{E}_1(t)}{2[1-\mathcal{U}(t)]^4}=\frac{5}{8\gamma^7(1-t/\rho)^{7/2}}. \tag{15}
$$

To obtain a result for the coefficients $E_{n,2}$, we use the transfer formula (Corollary VI.1, page 392 and Theorem VI.4, page 393 in [10]) – according to which, if $f(t)$ is Δ-analytic with a singularity at b, and $f(t)\sim(1-\frac{t}{b})^{-a}$ as $\frac{t}{b}\to 1$ with t in Δ, and $a\notin\{0,-1,-2,\ldots\}$, then $[t^n]f(t)\sim n^{a-1}b^{-n}/\Gamma(a)$. Here, ρ fulfills the role of b and $\frac{7}{2}$ that of a.

▶ **Proposition 5.** *The asymptotic growth of the number $E_{n,2}$ of unlabeled galled trees with n leaves and 2 galls satisfies*

$$
E_{n,2}\sim\frac{5}{8\gamma^7\Gamma(\frac{7}{2})}n^{\frac{5}{2}}\rho^{-n}=\frac{1}{3\gamma^7\sqrt{\pi}}n^{\frac{5}{2}}\rho^{-n}. \tag{16}
$$

We note the appearance of ρ^{-n} and $n^{5/2}$ to obtain the following corollary.

▶ **Corollary 6.** *The exponential growth of $\mathcal{E}_2(t)$ is the same as that of $\mathcal{U}(t)$ and $\mathcal{E}_1(t)$; however, its subexponential growth is greater.*

6   Analysis of $E_{n,g}$

6.1   Generating function

We denote the generating function of the number of galled trees with exactly g galls by $\mathcal{E}_g(t) = \sum_{n \geq 0} E_{n,g} t^n$. Similarly to the case of $g = 2$, we use the recursion we had calculated for $E_{n,g}$ in Proposition 3 to derive the generating function. From Proposition 3, we can decompose the generating function by

$$
\mathcal{E}_g(t) = \frac{1}{2}\Bigg[\underbrace{\sum_{n \geq 0} \Bigg(\Bigg(\sum_{\mathbf{c} \in C(n,2)} \sum_{\mathbf{d} \in C(g+2,2)} \prod_{i=1}^{2} E_{c_i,d_i-1} \Bigg) + E_{\frac{n}{2},\frac{g}{2}} \Bigg) t^n}_{\mathcal{E}_{g_{\mathrm{i}}}(t)}
$$

$$
+ \underbrace{\sum_{n \geq 0} \Bigg(\sum_{k=3}^{n} (k-2) \sum_{\mathbf{c} \in C(n,k)} \sum_{\mathbf{d} \in C(g-1+k,k)} \prod_{i=1}^{k} E_{c_i,d_i-1} \Bigg) t^n}_{\mathcal{E}_{g_{\mathrm{ii}}}(t)}
$$

$$
+ \underbrace{\sum_{n \geq 0} \Bigg(\sum_{a=1}^{\lfloor \frac{n-1}{2} \rfloor} \sum_{\mathbf{c} \in C_p(n,2a+1)} \sum_{\mathbf{d} \in C_p(g-1+2a+1,2a+1)} \prod_{i=1}^{a+1} E_{c_i,d_i-1} \Bigg) t^n}_{\mathcal{E}_{g_{\mathrm{iii}}}(t)} \Bigg]. \tag{17}
$$

where $E_{n,g} = 0$ for pairs with $n = 0$ or $n = 1$ and $g \geq 1$. The terms in the decomposition are

$$
\mathcal{E}_{g_{\mathrm{i}}}(t) = 2 \sum_{m \geq 0} \sum_{n \geq m} (U_m t^m)(E_{n-m,g} t^{n-m}) + \sum_{j=1}^{g-1} \sum_{m \geq 0} \sum_{n \geq m} (E_{m,j} t^m)(E_{n-m,g-j} t^{n-m})
$$

$$
+ \sum_{n \geq 0} E_{\frac{n}{2},\frac{g}{2}} t^n
$$

$$
\mathcal{E}_{g_{\mathrm{ii}}}(t) = \sum_{\ell=1}^{g-1} \sum_{k \geq 3} (k-2) \binom{k}{\ell} \sum_{m \geq k-\ell} \sum_{\mathbf{c} \in C(m,k-\ell)} \prod_{i=1}^{k-\ell} U_{c_i} t^{c_i}
$$

$$
\times \sum_{n \geq m} \sum_{\tilde{\mathbf{c}} \in C(n-m,\ell)} \sum_{\mathbf{d} \in C(g-1,\ell)} \prod_{j=1}^{\ell} E_{\tilde{c}_j,d_j} t^{\tilde{c}_j} \tag{18}
$$

$$
\mathcal{E}_{g_{\mathrm{iii}}}(t) = \sum_{\ell=0}^{\lfloor \frac{g-1}{2} \rfloor} \sum_{a \geq 1} \binom{a}{\ell} \sum_{m_1 \geq a-\ell} \sum_{\mathbf{c} \in C(m_1,a-\ell)} \prod_{i=1}^{a-\ell} U_{c_i} t^{2c_i}
$$

$$
\times \sum_{m \geq m_1+\ell} \sum_{\tilde{\mathbf{c}} \in C(m-m_1,\ell)} \sum_{b=\ell}^{\lfloor \frac{g-1}{2} \rfloor} \sum_{\mathbf{d} \in C(b,\ell)} \prod_{j=1}^{\ell} E_{\tilde{c}_j,d_j} t^{2c_j} \sum_{n \geq 2m} E_{n-2m,g-1-2b} t^{n-2m}, \tag{19}
$$

where it is convenient to denote U_n by $E_{n,0}$ for terms with $g - 1 - 2b = 0$ in $\mathcal{E}_{g_{\mathrm{iii}}}(t)$.

In $\mathcal{E}_{g_{\mathrm{i}}}(t)$, j is the number of galls in the left subtree of the root, supposing both subtrees possess at least one gall. In $\mathcal{E}_{g_{\mathrm{ii}}}(t)$, ℓ is the number of subtrees of the root gall that possess at least one gall; k is the number of subtrees of the root gall, so that $\binom{k}{\ell}$ counts ways to select which ℓ subtrees possess galls; and m is the number of leaves in the $k - \ell$ remaining subtrees.

Similarly, in $\mathcal{E}_{g_{\mathrm{iii}}}(t)$, for symmetric root galls, ℓ is the number of subtrees of the left side of the root gall that contain galls; a is the number of subtrees of the left side of the root gall; m_1 is the sample size in the $a - \ell$ subtrees that do not possess galls; $m - m_1$ is the sample size in the ℓ subtrees that do possess galls; and b is the number of galls in those ℓ subtrees.

We now solve each part of the decomposition:

$$\mathcal{E}_{g_{\mathrm{i}}}(t) = 2 \sum_{m \geq 0} (U_m t^m) \sum_{\ell \geq 0} (E_{\ell,g} t^\ell) + \sum_{j=1}^{g-1} \sum_{m \geq 0} (E_{m,j} t^m) \sum_{\ell \geq 0} (E_{\ell,g-j} t^\ell) + \sum_{n \geq 0} E_{n,\frac{g}{2}} t^{2n}$$

$$= 2\mathcal{U}(t)\, \mathcal{E}_g(t) + \left(\sum_{j=1}^{g-1} \mathcal{E}_j(t)\, \mathcal{E}_{g-j}(t) \right) + \mathcal{E}_{\frac{g}{2}}(t^2). \tag{20}$$

where $\mathcal{E}_\ell(t) = 0$ for $\ell \notin \mathbb{N}$. The second part produces

$$\mathcal{E}_{g_{\mathrm{ii}}}(t) = \sum_{\ell=1}^{g-1} \sum_{k \geq \max(\ell,3)} (k-2)\binom{k}{\ell} \sum_{i_1 \geq 0} \sum_{i_2 \geq 0} \cdots \sum_{i_{k-\ell} \geq 0} U_{i_1} U_{i_2} \cdots U_{i_{k-\ell}} t^{i_1+i_2+\ldots+i_{k-\ell}}$$

$$\times \sum_{\mathbf{d} \in C(g-1,\ell)} \sum_{j_1 \geq 0} \sum_{j_2 \geq 0} \cdots \sum_{j_\ell \geq 0} E_{j_1,d_1} E_{j_2,d_2} \cdots E_{j_\ell,d_\ell} t^{j_1+j_2+\ldots+j_\ell}$$

$$= \sum_{\ell=1}^{g-1} \left(\sum_{k \geq \max(\ell,3)} (k-2)\binom{k}{\ell} \mathcal{U}^{k-\ell}(t) \right) \sum_{\mathbf{d} \in C(g-1,\ell)} \prod_{j=1}^{\ell} \mathcal{E}_{d_j}(t)$$

$$= \sum_{\ell=1}^{g-1} \left(\frac{3\mathcal{U}(t) - 2 + \ell}{[1 - \mathcal{U}(t)]^{\ell+2}} + [\![\ell = 1]\!] \right) \sum_{\mathbf{d} \in C(g-1,\ell)} \prod_{j=1}^{\ell} \mathcal{E}_{d_j}(t). \tag{21}$$

Here, $[\![\cdot]\!]$ denotes the Iverson bracket. Finally, for the third part,

$$\mathcal{E}_{g_{\mathrm{iii}}}(t) = \sum_{\ell=0}^{\lfloor \frac{g-1}{2} \rfloor} \sum_{a \geq 1} \binom{a}{\ell} \sum_{i_1 \geq 0} \sum_{i_2 \geq 0} \cdots \sum_{i_{a-\ell} \geq 0} U_{i_1} U_{i_2} \cdots U_{i_{a-\ell}} t^{2i_1+2i_2+\ldots+2i_{a-\ell}}$$

$$\times \sum_{b=\ell}^{\lfloor \frac{g-1}{2} \rfloor} \sum_{\mathbf{d} \in C(b,\ell)} \sum_{j_1 \geq 0} \sum_{j_2 \geq 0} \cdots \sum_{j_\ell \geq 0} E_{j_1,d_1} E_{j_2,d_2} \cdots E_{j_\ell,d_\ell} t^{2j_1+2j_2+\ldots2j_\ell}$$

$$\times \sum_{j \geq 0} E_{j,g-1-2b} t^j$$

$$= \sum_{\ell=0}^{\lfloor \frac{g-1}{2} \rfloor} \left(\sum_{a \geq 1} \binom{a}{\ell} \mathcal{U}^{a-\ell}(t^2) \right) \sum_{b=\ell}^{\lfloor \frac{g-1}{2} \rfloor} \sum_{\mathbf{d} \in C(b,\ell)} \left(\prod_{j=1}^{\ell} \mathcal{E}_{d_j}(t^2) \right) \mathcal{E}_{g-1-2b}(t)$$

$$= \sum_{\ell=0}^{\lfloor \frac{g-1}{2} \rfloor} \left(\frac{1}{[1 - \mathcal{U}(t^2)]^{\ell+1}} - [\![\ell = 0]\!] \right) \sum_{b=\ell}^{\lfloor \frac{g-1}{2} \rfloor} \sum_{\mathbf{d} \in C(b,\ell)} \left(\prod_{j=1}^{\ell} \mathcal{E}_{d_j}(t^2) \right) \mathcal{E}_{g-1-2b}(t). \tag{22}$$

6.2 Asymptotic analysis

$\mathcal{E}_g(t)$ is the sum $\frac{1}{2}[\mathcal{E}_{g_{\mathrm{i}}}(t) + \mathcal{E}_{g_{\mathrm{ii}}}(t) + \mathcal{E}_{g_{\mathrm{iii}}}(t)]$ (eq. (17)). We denote $\mathcal{E}'_{g_{\mathrm{i}}}(t) = \left(\sum_{j=1}^{g-1} \mathcal{E}_j(t)\, \mathcal{E}_{g-j}(t) \right) + \mathcal{E}_{\frac{g}{2}}(t^2)$ and have $\mathcal{E}_g(t) = \frac{1}{2[1-\mathcal{U}(t)]}[\mathcal{E}'_{g_{\mathrm{i}}}(t) + \mathcal{E}_{g_{\mathrm{ii}}}(t) + \mathcal{E}_{g_{\mathrm{iii}}}(t)]$. From eqs. (20)-(22), $\mathcal{E}_g(t)$ is a rational function in $\mathcal{U}(t)$ and $\mathcal{E}_\ell(t)$ for $1 \leq \ell \leq g-1$, as well as in $\mathcal{U}(t^2)$ and $\mathcal{E}_\ell(t^2)$ for $1 \leq \ell \leq g-1$.

▶ **Proposition 7.** *The generating function $\mathcal{E}_g(t)$ for the number of unlabeled galled trees with g galls satisfies as $t \to \rho^-$*

$$\mathcal{E}_g(t) \sim \frac{\delta_g}{\gamma^{4g-1}(1 - t/\rho)^{2g-1/2}}, \tag{23}$$

where δ_g is a constant dependent on g satisfying $\delta_1 = \frac{1}{2}$, and for $g \geq 2$,

$$\delta_g = \frac{1}{2} \sum_{\ell=1}^{g-1} \left[\delta_\ell \delta_{g-\ell} + (\ell+1) \sum_{\mathbf{d} \in C(g-1,\ell)} \prod_{j=1}^{\ell} \delta_{d_j} \right]. \tag{24}$$

Proof. We proceed by induction. The claim holds for $g = 1$ (Proposition 1) and $g = 2$ (eq. (15)), with $\delta_2 = \frac{1}{2}[\frac{1}{2}\frac{1}{2} + 2\frac{1}{2}] = \frac{5}{8}$. We assume inductively that for $\ell = 1, 2, \ldots, g-1$, $\mathcal{E}_\ell(t) \sim \delta_\ell/[\gamma^{4\ell-1}(1 - t/\rho)^{2\ell-1/2}]$, with constants δ_ℓ as in eq. (24). By the inductive hypothesis, the convergence radius of $\mathcal{E}_\ell(t)$ for each ℓ, $1 \leq \ell \leq g-1$, is ρ. Because $t^2 < t$ for $t < \rho$, $\mathcal{U}(t^2)$ and $\mathcal{E}_\ell(t^2)$ can be treated as constants when finding the asymptotic behavior of $\mathcal{E}_g(t)$. As a result, using the inductive hypothesis, all terms in $\mathcal{E}_g(t)$ take the form $c/[\gamma^m(1 - t/\rho)^{m/2}]$, and we must find the terms with the maximal power of $1/\sqrt{1 - t/\rho}$.

We examine $\mathcal{E}'_{g_i}(t)$, $\mathcal{E}_{g_{iii}}(t)$, and then $\mathcal{E}_{g_{ii}}(t)$. By the inductive hypothesis,

$$\mathcal{E}'_{g_i}(t) \sim \sum_{j=1}^{g-1} \left[\frac{\delta_j}{\gamma^{4j-1}(1 - t/\rho)^{2j-1/2}} \cdot \frac{\delta_{g-j}}{\gamma^{4(g-j)-1}(1 - t/\rho)^{2(g-j)-1/2}} \right]$$
$$\sim \sum_{j=1}^{g-1} \frac{\delta_j \delta_{g-j}}{\gamma^{4g-2}(1 - t/\rho)^{2g-1}} \tag{25}$$

$$\mathcal{E}_{g_{iii}}(t) \sim \sum_{\ell=0}^{\lfloor \frac{g-1}{2} \rfloor} \sum_{b=\ell}^{\lfloor \frac{g-1}{2} \rfloor} \left(\frac{1}{[1 - \mathcal{U}(\rho^2)]^{\ell+1}} \sum_{\mathbf{d} \in C(b,\ell)} \prod_{j=1}^{\ell} \mathcal{E}_{d_j}(\rho^2) \right) \frac{\delta_{g-1-2b}}{\gamma^{4g-8b-5}(1 - t/\rho)^{2g-4b-5/2}}. \tag{26}$$

Because the largest power of $1/(1 - t/\rho)$ in $\mathcal{E}_{g_{iii}}(t)$ is less than $2g - 1$, its largest power in $\mathcal{E}'_{g_i}(t)$, $\mathcal{E}_{g_{iii}}(t)$ does not affect the asymptotics of $\mathcal{E}_g(t)$.

For $\mathcal{E}_{g_{ii}}(t)$, for any $\ell = 1, 2, \ldots, g-1$, two quantities determine the power of $1/\sqrt{1 - t/\rho}$: both $\sum_{\mathbf{d} \in C(g-1,\ell)} \prod_{j=1}^{\ell} \mathcal{E}_{d_j}(t)$ and $[3\mathcal{U}(t) - 2 + \ell]/[1 - \mathcal{U}(t)]^{\ell+2} + [\![\ell = 1]\!]$. First, according to the inductive hypothesis, for each ℓ, $1 \leq \ell \leq g-1$, noting $\sum_{j=1}^{\ell} d_j = g-1$,

$$\sum_{\mathbf{d} \in C(g-1,\ell)} \prod_{j=1}^{\ell} \mathcal{E}_{d_j}(t) \sim \sum_{\mathbf{d} \in C(g-1,\ell)} \prod_{j=1}^{\ell} \frac{\delta_{d_j}}{\gamma^{4d_j-1}(1 - t/\rho)^{2d_j-1/2}}$$
$$\sim \sum_{\mathbf{d} \in C(g-1,\ell)} \frac{\prod_{j=1}^{\ell} \delta_{d_j}}{\gamma^{4g-4-\ell}(1 - t/\rho)^{2g-2-\ell/2}}. \tag{27}$$

Second, for ℓ, $1 \leq \ell \leq g-1$, from $\mathcal{U}(t) \sim 1 - \gamma\sqrt{1 - t/\rho}$,

$$\left(\frac{3\mathcal{U}(t) - 2 + \ell}{[1 - \mathcal{U}(t)]^{\ell+2}} + [\![\ell = 1]\!] \right) \sim \frac{\ell+1}{\gamma^{\ell+2}(1 - t/\rho)^{(\ell+2)/2}}. \tag{28}$$

Combining eqs. (27) and (28), we obtain

$$
\mathcal{E}_{g_{\mathrm{ii}}}(t) \sim \sum_{\ell=1}^{g-1} \sum_{\mathbf{d}\in C(g-1,\ell)} \frac{\prod_{j=1}^{\ell}\delta_{d_j}}{\gamma^{4g-4-\ell}(1-t/\rho)^{2g-2-\ell/2}} \cdot \frac{\ell+1}{\gamma^{\ell+2}(1-t/\rho)^{(\ell+2)/2}}
$$

$$
\sim \sum_{\ell=1}^{g-1} \frac{(\ell+1)\sum_{\mathbf{d}\in C(g-1,\ell)}\prod_{j=1}^{\ell}\delta_{d_j}}{\gamma^{4g-2}(1-t/\rho)^{2g-1}}. \tag{29}
$$

The proof is concluded by noting

$$
\mathcal{E}_{g}(t) \sim \left[\sum_{j=1}^{g-1} \frac{\delta_j \delta_{g-j}}{\gamma^{4g-2}(1-t/\rho)^{2g-1}} + \sum_{\ell=1}^{g-1} \frac{(\ell+1)\sum_{\mathbf{d}\in C(g-1,\ell)}\prod_{j=1}^{\ell}\delta_{d_j}}{\gamma^{4g-2}(1-t/\rho)^{2g-1}} \right] \frac{1}{2\gamma(1-t/\rho)^{1/2}}
$$

$$
\sim \frac{\sum_{\ell=1}^{g-1}\left[\delta_\ell \delta_{g-\ell} + (\ell+1)\sum_{\mathbf{d}\in C(g-1,\ell)}\prod_{j=1}^{\ell}\delta_{d_j}\right]}{2\gamma^{4g-1}(1-t/\rho)^{2g-1/2}}
$$

$$
\sim \frac{\delta_g}{\gamma^{4g-1}(1-t/\rho)^{2g-1/2}}. \tag{30}
$$

◀

▶ **Theorem 8.** *The asymptotic growth of the number $E_{n,g}$ of unlabeled galled trees with n leaves and a fixed number of galls $g \geq 1$ satisfies*

$$
E_{n,g} \sim \frac{\delta_g}{\gamma^{4g-1}\Gamma(2g-\frac{1}{2})} n^{2g-\frac{3}{2}} \rho^{-n} \sim \frac{2^{2g-1}\delta_g}{\gamma^{4g-1}(4g-3)!!\sqrt{\pi}} n^{2g-\frac{3}{2}}\rho^{-n}. \tag{31}
$$

Proof. The first step follows from the transfer formula. For the second step of eq. (31), we recall $\Gamma(n+\frac{1}{2}) = [(2n-1)!!/2^n]\sqrt{\pi}$ with and $2g - \frac{1}{2} = (2g-1) + \frac{1}{2}$. ◀

The δ_g have a relationship with the Catalan numbers, $C_m = \binom{2m}{m}/(m+1)$.

▶ **Proposition 9.** *The numbers $\{\delta_g\}_{g\geq 1}$ satisfy $2^{2g-1}\delta_g = C_{2g-1}$.*

Proof. We prove the result by showing that the generating function $\mathcal{D}(t) = \sum_{g\geq 1} 2^{2g-1}\delta_g t^{2g-1}$ is the odd part of the generating function of the Catalan numbers, $\mathcal{C}_O(t) = \sum_{g\geq 1} C_{2g-1} t^{2g-1}$.

$\mathcal{C}_O(t)$ satisfies $\mathcal{C}_O(t) = \frac{1}{2}\sum_{n\geq 0}[C_n t^n - C_n(-t)^n] = \sum_{n\geq 1} C_{2n-1} t^{2n-1}$, where $\mathcal{C}(t) = (1-\sqrt{1-4t})/(2t)$ is the generating function of the Catalan numbers. Hence, $\mathcal{C}_O(t) = \left[1 - \frac{1}{2}(\sqrt{1-4t} + \sqrt{1+4t})\right]/(2t)$. From the recursion for δ_g (Proposition 7),

$$
\mathcal{D}(t) = t + \sum_{g\geq 2}\left(\sum_{\ell=1}^{g-1} 2^{2g-2}\delta_\ell \delta_{g-\ell}\right) t^{2g-1} + \sum_{g\geq 2}\left[\sum_{\ell=1}^{g-1}(\ell+1)2^{2g-2}\sum_{\mathbf{d}\in C(g-1,\ell)}\prod_{j=1}^{\ell}\delta_{d_j}\right]t^{2g-1}
$$

$$
= t + \left[\sum_{\ell\geq 1} 2^{2\ell-1}\delta_\ell t^{2\ell-1}\sum_{g\geq \ell+1} 2^{2(g-\ell)-1}\delta_{g-\ell} t^{2(g-\ell)-1}\right]t
$$

$$
+ \left[\sum_{\ell\geq 1}(\ell+1)(2t)^\ell \sum_{g\geq \ell+1}\sum_{\mathbf{d}\in C(g-1,\ell)}\prod_{j=1}^{\ell} 2^{2d_j-1}\delta_{d_j} t^{2d_j-1}\right]t
$$

$$
= t + t\mathcal{D}^2(t) + t\sum_{\ell\geq 1}(\ell+1)[2t\mathcal{D}(t)]^\ell
$$

$$
= t + t\mathcal{D}^2(t) + \frac{2t^2\mathcal{D}(t)}{[1-2t\mathcal{D}(t)]^2} + \frac{2t^2\mathcal{D}(t)}{1-2t\mathcal{D}(t)}. \tag{32}
$$

Solving for $\mathcal{D}(t)$, we obtain four solutions, only one of which has the correct limit of 0 as $t \to 0$; this root is equal to $\mathcal{C}_O(t)$. ◀

▶ **Theorem 10.** *The number of unlabeled galled trees with n leaves and any fixed number of galls $g \geq 0$ has asymptotic approximation*

$$E_{n,g} \sim \frac{2^{2g-1}}{(2g)! \, \gamma^{4g-1} \sqrt{\pi}} n^{2g-\frac{3}{2}} \rho^{-n}. \tag{33}$$

■ **Table 1** The subexponential portion $c_g n^{2g-\frac{3}{2}}$ of the growth $c_g n^{2g-\frac{3}{2}} \rho^{-n}$ with the number of leaves n of $E_{n,g}$, the number of galled trees with exactly g galls. Quantities are computed according to eq. (2) for $g = 0$ and Theorems 8 and 10 for $g \geq 1$.

Number of galls g	Exact constant c_g	Approximate value of c_g	$n^{2g-\frac{3}{2}}$
0	$\frac{\gamma}{2\sqrt{\pi}}$	0.3188	$n^{-\frac{3}{2}}$
1	$\frac{1}{\gamma^3\sqrt{\pi}}$	0.3910	$n^{\frac{1}{2}}$
2	$\frac{5}{15\gamma^7\sqrt{\pi}} = \frac{8}{24\gamma^7\sqrt{\pi}} = \frac{1}{3\gamma^7\sqrt{\pi}}$	0.0799	$n^{\frac{5}{2}}$
3	$\frac{42}{945\gamma^{11}\sqrt{\pi}} = \frac{32}{720\gamma^{11}\sqrt{\pi}} = \frac{2}{45\gamma^{11}\sqrt{\pi}}$	0.0065	$n^{\frac{9}{2}}$
4	$\frac{429}{135135\gamma^{15}\sqrt{\pi}} = \frac{128}{40320\gamma^{15}\sqrt{\pi}} = \frac{1}{315\gamma^{15}\sqrt{\pi}}$	2.8638×10^{-4}	$n^{\frac{13}{2}}$
5	$\frac{4862}{34459425\gamma^{19}\sqrt{\pi}} = \frac{512}{3628800\gamma^{19}\sqrt{\pi}} = \frac{2}{14175\gamma^{19}\sqrt{\pi}}$	7.8062×10^{-6}	$n^{\frac{17}{2}}$

Proof. The Catalan numbers satisfy $C_n = 2^n(2n-1)!!/(n+1)!$, so that

$$\frac{2^{2g-1}\delta_g}{(4g-3)!!} = \frac{C_{2g-1}}{(4g-3)!!} = \frac{2^{2g-1}[2(2g-1)-1]!!}{(4g-3)!! \, (2g-1+1)!} = \frac{2^{2g-1}}{(2g)!}.$$

The case of $g = 0$ is included, as $E_{n,0} \sim [2^{-1}/(\gamma^{-1}\sqrt{\pi})]n^{-\frac{3}{2}}\rho^{-n} = [\gamma/2\sqrt{\pi}]n^{-\frac{3}{2}}\rho^{-n} \sim U_n$. ◀

Table 1 depicts the subexponential growth of $E_{n,g}$ for each g from 1 to 5. For $g = 1$ and $g = 2$, the theorem recovers the values obtained in Propositions 2 and 5.

▶ **Corollary 11.** *The exponential growth of the number $E_{n,g}$ of unlabeled trees with n leaves and a fixed number of galls $g \geq 1$ is the same as that of U_n, the number of unlabeled trees with no galls; however, the subexponential growth is greater by a factor of $4n^2/[\gamma^4(2g+1)(2g+2)]$.*

7 Discussion

We have studied the number of rooted binary unlabeled galled trees with a fixed number of galls, analyzing the exponential growth of this quantity as the number of leaves increases. We have found that the exponential growth, with the increase in the number of leaves n, of the number of galled trees with a fixed number of galls is independent of the number of galls g (Corollary 11). This independence includes the case of $g = 0$ galls, the classic case of rooted binary unlabeled trees. It also implies that the number of galled trees whose number of galls is in some finite set G also has this same exponential growth.

The exponential growth with n of the number of galled trees with fixed g or with g in a finite set of values contrasts with the much greater increase in A_n, the number of galled trees with no restriction on the number of galls. This much larger growth for A_n is explained

by the increase in the subexponential component with increasing g of the number of galled trees with n leaves and g galls, and the fact that with no maximum number of galls, as n increases, the number of terms in $A_n = \sum_{g \geq 0}^{\lfloor (n-1)/2 \rfloor} E_{n,g}$ grows without bound.

Our analysis produced a recursion for the Catalan numbers with odd indices: $C_{2n-1} = \sum_{m=1}^{n-1} C_{2m-1} C_{2(n-m)-1} + \sum_{m=1}^{n-1}(m+1)2^m \sum_{\mathbf{d} \in C(n-1,m)} C_{2d_j-1}$. The first part comes from terms of $C_n = \sum_{m=0}^{n-1} C_m C_{(n-1)-m}$ with odd m and $(n-1)-m$; the second substitutes a sum involving Catalan numbers with odd index for terms with even m and $(n-1)-m$.

The difference across values of g in the growth of the number of trees with exactly $g \geq 0$ galls lies in the subexponential component, $c_g n^{2g-\frac{3}{2}}$. Related problems involving labeled phylogenetic networks show this same pattern, in which incrementing a constant associated with network complexity does change the subexponential growth but not the exponential growth. In particular, this pattern is seen with increasingly many reticulation nodes in various network classes [6, 7, 11, 12, 13, 19]; the subexponential growth often includes a factor of n^2, as in our case.

Note additionally that beginning from $g = 1$, the constant c_g in the asymptotic approximation for $E_{n,g}$ decreases with g (eq. (31), Table 1). This property also holds for the labeled normal networks of Fuchs et al. [11, 12, 13].

The study here deals with the asymptotic enumeration of galled trees when the number of *galls* is fixed. Using the bivariate function $\mathcal{A}(t, u) = \sum_{n \geq 0} \sum_{g \geq 0} E_{n,g} t^n u^g$, Section 5.6 of our previous study of galled trees showed that for a fixed number of *leaves*, the number of galls follows an asymptotic normal distribution [1, eq. 56]. The marginal analysis fixing the number of galls contributes a perspective on the bivariate distribution different from that of the previous analysis.

We comment that we could potentially have derived our generating functions by the symbolic method [10]. Our approach instead began with constructive enumeration of possible cases, continuing the analysis based on a recursion derived in our previous study of galled trees [1] in order to find the generating functions. The symbolic method, which we defer to a subsequent article, potentially leads to simpler derivations that enable quick comparisons of relationships among enumerations for different types of galled trees.

By analyzing the asymptotics of $E_{n,g}$ for arbitrary g, this work solves unsolved problems from [1], who only analyzed $E_{n,1}$ and $A_n = \sum_{g \geq 0}^{\lfloor (n-1)/2 \rfloor} E_{n,g}$. The analysis has potential to assist in other scenarios with unlabeled phylogenetic networks indexed by a fixed quantity.

References

1 L. Agranat-Tamir, S. Mathur, and N. A. Rosenberg. Enumeration of rooted binary unlabeled galled trees. *Bulletin of Mathematical Biology*, 86:45, 2024. `doi:10.1007/s11538-024-01270-8`.

2 F. Bienvenu, A. Lambert, and M. Steel. Combinatorial and stochastic properties of ranked tree-child networks. *Random Structures and Algorithms*, 60:653–689, 2022. `doi:10.1002/rsa.21048`.

3 M. Bouvel, P. Gambette, and M. Mansouri. Counting phylogenetic networks of level 1 and 2. *Journal of Mathematical Biology*, 81:1357–1395, 2020. `doi:10.1007/s00285-020-01543-5`.

4 G. Cardona and L. Zhang. Counting and enumerating tree-child networks and their subclasses. *Journal of Computer and System Sciences*, 114:84–104, 2020. `doi:10.1016/j.jcss.2020.06.001`.

5 K.-Y. Chang, W.-K. Hon, and S. V. Thankachan. Compact encoding for galled-trees and its applications. In *2018 Data Compression Conference*, pages 297–306, Snowbird, UT, 2018. `doi:10.1109/DCC.2018.00038`.

6   Y. S. Chang and M. Fuchs. Counting Phylogenetic Networks with Few Reticulation Vertices: Galled and Reticulation-Visible Networks. *Bull. Math. Biol.*, 86(7):76, 2024. `doi:10.1007/s11538-024-01309-w`.

7   Y.-S. Chang, M. Fuchs, and G.-R. Yu. Galled tree-child networks. In *35th International Conference on Probabilistic, Combinatorial and Asymptotic Methods for the Analysis of Algorithms*, volume 302, Article 2 of *Leibniz International Proceedings in Informatics, LIPIcs*. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Wadern, 2024.

8   L. Comtet. *Advanced Combinatorics*. Reidel, Boston, 1974. `doi:10.1007/978-94-010-2196-8`.

9   J. Felsenstein. *Inferring Phylogenies*. Sinauer Associates Inc., Sunderland MA, 2004.

10   P. Flajolet and R. Sedgewick. *Analytic Combinatorics*. Cambridge University Press, Cambridge, 2009. `doi:10.1017/CBO9780511801655`.

11   M. Fuchs, B. Gittenberger, and M. Mansouri. Counting phylogenetic networks with few reticulation vertices: tree-child and normal networks. *Australasian Journal of Combinatorics*, 73:385–423, 2019.

12   M. Fuchs, B. Gittenberger, and M. Mansouri. Counting phylogenetic networks with few reticulation vertices: exact enumeration and corrections. *Australasian Journal of Combinatorics*, 81:257–282, 2021.

13   M. Fuchs, E.-Y. Huang, and G.-R. Yu. Counting phylogenetic networks with few reticulation vertices: a second approach. *Discrete Applied Mathematics*, 320:140–149, 2022. `doi:10.1016/j.dam.2022.03.026`.

14   D. Gusfield. *ReCombinatorics*. MIT Press, Cambridge MA, 2014.

15   D. Gusfield, S. Eddhu, and C. Langley. Efficient reconstruction of phylogenetic networks with constrained recombination. In *Proceedings of the IEEE Computer Society Conference on Bioinformatics*, pages 363–374, 2003. `doi:10.1109/CSB.2003.1227337`.

16   D. Gusfield, S. Eddhu, and C. H. Langley. The fine structure of galls in phylogenetic networks. *INFORMS Journal on Computing*, 16:459–469, 2004. `doi:10.1287/ijoc.1040.0099`.

17   D. H. Huson, R. Rupp, and C. Scornavacca. *Phylogenetic Networks: Concepts, Algorithms and Applications*. Cambridge University Press, Cambridge, 2010. `doi:10.1017/CBO9780511974076`.

18   S. Kong, J. C. Pons, L. Kubatko, and K. Wicke. Classes of explicit phylogenetic networks and their biological and mathematical significance. *Journal of Mathematical Biology*, 84:47, 2022. `doi:10.1007/s00285-022-01746-y`.

19   M. Mansouri. Counting general phylogenetic networks. *Australasian Journal of Combinatorics*, 83:40–86, 2022.

20   S. Mathur and N. A. Rosenberg. All galls are divided into three or more parts: recursive enumeration of labeled histories for galled trees. *Algorithms in Molecular Biology*, 18:1, 2023. `doi:10.1186/s13015-023-00224-4`.

21   C. Semple and M. Steel. *Phylogentics*. Oxford University Press, Oxford, 2003.

22   C. Semple and M. Steel. Unicyclic networks: compatibility and enumeration. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 3:84–91, 2006. `doi:10.1109/TCBB.2006.14`.

23   Y. S. Song. A concise necessary and sufficient condition for the existence of a galled-tree. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 3:186–191, 2006. `doi:10.1109/TCBB.2006.15`.

24   T. Warnow. *Computational Phylogenetics*. Cambridge University Press, Cambridge, 2018. `doi:10.1017/9781316882313`.

Sharpened Localization of the Trailing Point of the Pareto Record Frontier

James Allen Fill[1] ✉ 🏠 🆔
Department of Applied Mathematics and Statistics, The Johns Hopkins University,
Baltimore, MD, USA

Daniel Q. Naiman ✉ 🏠
Department of Applied Mathematics and Statistics, The Johns Hopkins University,
Baltimore, MD, USA

Ao Sun ✉
Department of Applied Mathematics and Statistics, The Johns Hopkins University,
Baltimore, MD, USA

─── **Abstract** ───

For $d \geq 2$ and i.i.d. d-dimensional observations $X^{(1)}, X^{(2)}, \ldots$ with independent Exponential(1) coordinates, we revisit the study by Fill and Naiman (*Electron. J. Probab.*, 25:Paper No. 92, 24 pp., 2020) of the boundary (relative to the closed positive orthant), or "frontier", F_n of the closed Pareto record-setting (RS) region $\mathrm{RS}_n := \{0 \leq x \in \mathbb{R}^d : x \not\prec X^{(i)} \text{ for all } 1 \leq i \leq n\}$ at time n, where $0 \leq x$ means that $0 \leq x_j$ for $1 \leq j \leq d$ and $x \prec y$ means that $x_j < y_j$ for $1 \leq j \leq d$. With $x_+ := \sum_{j=1}^d x_j = \|x\|_1$, let

$$F_n^- := \min\{x_+ : x \in F_n\} \quad \text{and} \quad F_n^+ := \max\{x_+ : x \in F_n\}.$$

Almost surely, there are for each n unique vectors $\lambda_n \in F_n$ and $\tau_n \in F_n$ such that $F_n^+ = (\lambda_n)_+$ and $F_n^- = (\tau_n)_+$; we refer to λ_n and τ_n as the *leading* and *trailing* points, respectively, of the frontier. Fill and Naiman provided rather sharp information about the typical and almost sure behavior of F^+, but somewhat crude information about F^-, namely, that for any $\varepsilon > 0$ and $c_n \to \infty$ we have

$$\mathbb{P}(F_n^- - \ln n \in (-(2+\varepsilon)\ln\ln\ln n, c_n)) \to 1$$

(describing typical behavior) and almost surely

$$\limsup \frac{F_n^- - \ln n}{\ln\ln n} \leq 0 \text{ and } \liminf \frac{F_n^- - \ln n}{\ln\ln\ln n} \in [-2, -1].$$

In this extended abstract we use the theory of *generators* (minima of F_n) together with the first- and second-moment methods to improve considerably the trailing-point location results to

$$F_n^- - (\ln n - \ln\ln\ln n) \xrightarrow{\mathrm{P}} -\ln(d-1)$$

(describing typical behavior) and, for $d \geq 3$, almost surely

$$\limsup[F_n^- - (\ln n - \ln\ln\ln n)] \leq -\ln(d-2) + \ln 2$$
$$\text{and } \liminf[F_n^- - (\ln n - \ln\ln\ln n)] \geq -\ln d - \ln 2.$$

───────────

[1] Corresponding author

1   Introduction, background, and main results

Notation.   Throughout this extended abstract we abbreviate the kth iterate of natural logarithm ln by L_k and L_1 by L, and we write $x_+ := \sum_{j=1}^{d} x_j$ and $x_\times := \prod_{j=1}^{d} x_j$ for the sum and product, respectively, of coordinates of the d-dimensional vector $x = (x_1, \dots, x_d)$. When $0 \le x$ the sum x_+ equals the ℓ^1-norm $\|x\|_1$, but we use the notation x_+ more generally. We denote coordinate-wise maximum and minimum of vectors by $\vee$ and $\wedge$, respectively.

Unless otherwise noted, all results of this extended abstract hold for any dimension $d \ge 2$.

The study of univariate records is well established ([1] is a standard reference), but that of multivariate records remains under vigorous development. Fill and Naiman [6] studied the stochastic process (F_n), where F_n is the boundary, or "frontier", for *Pareto records* (consult Definitions 1.1–1.2) in general dimension d when the observed sequence of points $X^{(1)}, X^{(2)}, \dots$ are assumed (as they are throughout this extended abstract, except where otherwise noted) to be i.i.d. (independent and identically distributed) copies of a d-dimensional random vector X with independent Exponential(1) coordinates X_j. Their main goal was to sharpen (in various senses) the assertion in Bai et al. [2] "that nearly all maxima occur in a thin strip sandwiched between [the] two parallel hyper-planes"

$$x_+ = L\,n - L_3\,n - L[4(d-1)] \quad \text{and} \quad x_+ = L\,n + 4(d-1)\,L_2\,n.$$

They did this largely by studying (separately) the maximum and minimum sums of coordinates for points lying in F_n. The results for the maximum sum were rather sharp; less so for the minimum sum. The main aim of this extended abstract is to use the theory of generators (minima of F_n) and the first- and second-moment methods to improve considerably their results about the minimum sum.

1.1   Pareto records and the record-setting region

For the reader's convenience, and with the permission of the authors and the copyright holder, this short subsection is excerpted largely verbatim from [6, Section 1.1].

We begin with some definitions. For a positive integer n, let $[n] := \{1, \dots, n\}$. Thus $[d]^{[n]}$ denotes the set of all functions from $[n]$ into $[d]$, or simply the set of all n-tuples with each entry in $\{1, \dots, d\}$. For d-dimensional vectors $x = (x_1, \dots, x_d)$ and $y = (y_1, \dots, y_d)$, write $x \prec y$ (respectively, $x \le y$) to mean that $x_j < y_j$ (resp., $x_j \le y_j$) for $j \in [d]$. (We caution that, with this convention, $\le$ is weaker than $\preceq$, the latter meaning "$\prec$ or $=$"; indeed, $(0,0) \le (0,1)$ but we have neither $(0,0) \prec (0,1)$ nor $(0,0) = (0,1)$. This distinction will be important for some of our later discussion of generators.) The notation $x \succ y$ means $y \prec x$, and $x \ge y$ means $y \le x$; the notation $x < y$ means $x \le y$ but $x \ne y$, and $y > x$ means $x < y$.

▶ **Definition 1.1.**
(a) *We say that $X^{(k)}$ is a* (Pareto) record *(or that it* sets *a record at time k) if $X^{(k)} \not\prec X^{(i)}$ for all $1 \le i < k$.*
(b) *If $1 \le k \le n$, we say that $X^{(k)}$ is a* current record *(or* remaining record*, or* maximum*) at time n if $X^{(k)} \not\prec X^{(i)}$ for all $1 \le i \le n$.*

For $n \geq 1$ (or $n \geq 0$, with the obvious conventions) let ρ_n ($\equiv \rho_{d,n}$) denote the number of remaining records at time n (when the dimension is d).

▶ **Definition 1.2.**

(a) *The* record-setting region *at time n is the (random) closed set of points*

$$\mathrm{RS}_n := \{x \in \mathbb{R}^d : 0 \leq x \not\prec X^{(i)} \text{ for all } 1 \leq i \leq n\}.$$

(b) *We call the (topological) boundary of RS_n (relative to the closed positive orthant determined by the origin) its* frontier *and denote it by F_n.*
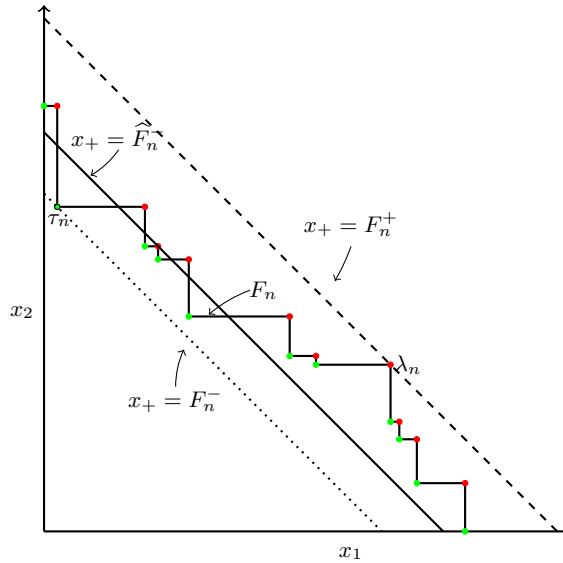
Figure 1 Record frontier F_n based on n observations (for some $n \geq 10$) resulting in 10 current records (shown in red), with the three hyperplanes $x_+ = F_n^+$, $x_+ = F_n^-$, and $x_+ = \widehat{F}_n^-$, the leading point λ_n and the trailing point τ_n. Concerning the three hyperplanes, see Definition 1.4 and (1.2). Generators (see Definition 4.1) are shown in green.

▶ **Remark 1.3.** The terminology in Definition 1.2(a) is natural since the next observation $X^{(n+1)}$ sets a record if and only if it falls in the record-setting region. Note that

$$\mathrm{RS}_n = \{x \in \mathbb{R}^d : 0 \leq x \not\prec X^{(i)} \text{ for all } 1 \leq i \leq n$$
$$\text{such that } X^{(i)} \text{ is a current record at time } n\},$$

and that the current records at time n all belong to RS_n but lie on its frontier. Observe also that F_n is a closed subset of RS_n.

This extended abstract primarily concerns the stochastic process (F_n), and specifically the process F^- as defined (along with the process F^+) next (see Figure 1).

▶ **Definition 1.4.** *Recalling that F_n denotes the frontier of RS_n, let*

$$F_n^- := \min\{x_+ : x \in F_n\} \quad and \quad F_n^+ := \max\{x_+ : x \in F_n\}. \tag{1.1}$$

Almost surely, there are for each n unique vectors $\lambda_n \in F_n$ and $\tau_n \in F_n$ such that $F_n^+ = \lambda_n$ and $F_n^- = \tau_n$; we call λ_n and τ_n the leading *and* trailing *points, respectively, of the frontier.*

Since the sets RS_n decrease (weakly) with n, we have the following trivial consequence.

▶ **Lemma 1.5.** *The process F^- has nondecreasing sample paths.*

1.2   The record-setting frontier; our two main theorems

Fill and Naiman first showed, in a precise sense [6, Theorem 1.4], that the difference between the sum of coordinates (call it Y_n) of a "generic" current record at time n and $\mathrm{L}\, n$ converges in distribution to standard Gumbel. They next translated results from classical extreme value theory due to Kiefer [7] to the setting of multivariate records to produce rather sharp typical-behavior and almost-sure results about the process F^+. For completeness, we repeat their main result [6, Theorem 1.8] for F^+ here, except that we have rather effortlessly extended part (b) of that theorem using Kiefer's "first proof" as described in [6, proof of Theorem 1.8(b)]. We remark that the difference between the top-boundary threshold at about $\mathrm{L}\, n + d\, \mathrm{L}_2\, n$ and bottom-boundary threshold at about $\mathrm{L}\, n + (d-1)\, \mathrm{L}_2\, n$ is a noteworthy feature of F_n^+ discussed further in [6, Section 1.3].

▶ **Theorem 1.6** (Kiefer [7]). *Consider the process F^+ defined at* (1.1).
(a) <u>Typical behavior of F^+</u>: $F_n^+ - [\mathrm{L}\, n + (d-1)\, \mathrm{L}_2\, n - \mathrm{L}((d-1)!)] \xrightarrow{\mathcal{L}} G$.
(b) <u>Top boundaries for F^+</u>: *For any sequence $b_n \to \infty$ that is ultimately monotone increasing,*

$$\mathbb{P}(F_n^+ \geq b_n \text{ i.o.}) = 1 \text{ or } 0 \text{ according as } \sum e^{-b_n} b_n^{d-1} \text{ diverges or converges.}$$

In particular, for any $k \geq 2$ we have

$$\mathbb{P}\left(F_n^+ \geq \mathrm{L}\, n + d\, \mathrm{L}_2\, n + \sum_{i=3}^{k} \mathrm{L}_i\, n + c\, \mathrm{L}_{k+1}\, n \text{ i.o.}\right) = \begin{cases} 1 & \text{if } c \leq 1; \\ 0 & \text{if } c > 1. \end{cases}$$

(c) <u>Bottom boundaries for F^+</u>:

$$\mathbb{P}(F_n^+ \leq \mathrm{L}\, n + (d-1)\, \mathrm{L}_2\, n - \mathrm{L}_3\, n - \mathrm{L}((d-1)!) + c \text{ i.o.}) = \begin{cases} 1 & \text{if } c \geq 0; \\ 0 & \text{if } c < 0. \end{cases}$$

From Theorem 1.6 it follows in particular that

$$\frac{F_n^+ - \mathrm{L}\, n}{\mathrm{L}_2\, n} \xrightarrow{\mathrm{P}} d - 1$$

and

$$\liminf \frac{F_n^+ - \mathrm{L}\, n}{\mathrm{L}_2\, n} = d - 1 < d = \limsup \frac{F_n^+ - \mathrm{L}\, n}{\mathrm{L}_2\, n} \text{ a.s.}$$

The results derived in [6] for F^- are much less sharp than for F^+. For the reader's convenience, we repeat those results here. Although parts (a) and (c1) were stated with coefficient -3 [rather than $-(2+c)$] for the $\mathrm{L}_3\, n$ term, the improvement we have noted here is pointed out in [6, Remark 3.3].

▶ **Theorem 1.7** ([6], Theorem 1.12). *Consider the process F^- defined at* (1.1).
(a) <u>Typical behavior of F^-</u>:

$$\mathbb{P}(F_n^- \leq \mathrm{L}\, n - (2 + c)\, \mathrm{L}_3\, n) \to 0 \text{ if } c > 0, \text{ and}$$

$$\mathbb{P}(F_n^- \geq \mathrm{L}\, n + c_n) \to 0 \text{ if } c_n \to \infty.$$

(b) <u>Top outer boundaries for F^-</u>: $\mathbb{P}(F_n^- \geq \mathrm{L}\, n + c\, \mathrm{L}_2\, n \text{ i.o.}) = 0$ if $c > 0$.

(c1) <u>Bottom outer boundaries for F^-</u>: $\mathbb{P}(F_n^- \leq \mathrm{L}\, n - (2 + c)\, \mathrm{L}_3\, n \text{ i.o.}) = 0$ if $c > 0$.

(c2) <u>A bottom inner boundary for F^-</u>: $\mathbb{P}(F_n^- \leq \mathrm{L}\, n - \mathrm{L}_3\, n \text{ i.o.}) = 1$. ⌟

Recall that for real-valued random variables Z_n and real numbers a_n, the condition $Z_n = O_{\mathrm{p}}(a_n)$ means that Z_n/a_n is bounded in probability.

The first of two main results of this extended abstract, Theorem 1.8, sharpens Theorem 1.7 considerably. In light of (i) the constant-order variability for a "generic" current record at time n described in the opening paragraph of this subsection and (ii) Theorem 1.6(a), we find it quite surprising that, properly centered but not scaled, F_n^- has a limit in probability.

▶ **Theorem 1.8.** *Consider the process F^- defined at (1.1).*

(a) <u>Typical behavior of F^-</u>:

$$F_n^- = \mathrm{L}\, n - \mathrm{L}_3\, n - \mathrm{L}(d - 1) + O_{\mathrm{p}}\left(\frac{\mathrm{L}_3\, n}{\mathrm{L}_2\, n}\right).$$

(b) <u>Top outer boundaries for F^-</u>: *If $d \geq 3$, then*

$$\mathbb{P}(F_n^- \geq \mathrm{L}\, n - \mathrm{L}_3\, n - \mathrm{L}(d - 2) + \mathrm{L}\, 2 + c \text{ i.o.}) = 0 \text{ if } c > 0.$$

(c) <u>Bottom outer boundaries for F^-</u>:

$$\mathbb{P}(F_n^- \leq \mathrm{L}\, n - \mathrm{L}_3\, n - \mathrm{L}\, d - \mathrm{L}\, 2 - c \text{ i.o.}) = 0 \text{ if } c > 0.$$

Theorem 1.8 gives rise immediately to the following succinct corollary.

▶ **Corollary 1.9.** *Consider the process F^- defined at (1.1).*

(a) <u>Typical behavior of F^-</u>:

$$F_n^- - (\mathrm{L}\, n - \mathrm{L}_3\, n) \xrightarrow{\mathrm{P}} - \mathrm{L}(d - 1)$$

and thus

$$\frac{F_n^- - \mathrm{L}\, n}{\mathrm{L}_3\, n} \xrightarrow{\mathrm{P}} -1$$

and, yet more crudely,

$$\frac{F_n^- - \mathrm{L}\, n}{\mathrm{L}_2\, n} \xrightarrow{\mathrm{P}} 0.$$

(b) <u>Almost sure behavior for F^-</u>:

$$\lim \frac{F_n^- - \mathrm{L}\, n}{\mathrm{L}_2\, n} = 0 \text{ a.s.}$$

Further, for fixed $d \geq 3$ we have the refinement

$$F_n^- = \mathrm{L}\, n - \mathrm{L}_3\, n + O(1) \text{ a.s.}$$

▶ Remark 1.10. We do not know how to improve Theorem 1.7(b) when $d = 2$.

Suppose now that instead of F_n^- we consider the somewhat larger quantity

$$\widehat{F}_n^- := (\text{minimum coordinate-sum of any current record at time } n). \tag{1.2}$$

(See Figure 1.) Our second main theorem concerns the process $\widehat{F}^-$; in summary, the same results hold for $\widehat{F}^-$ as for F^- in Theorem 1.8, with a sharper remainder term for $\widehat{F}^-$ in part (a).

▶ **Theorem 1.11.** *Consider the process $\widehat{F}^-$ defined at* (1.2).

(a) <u>Typical behavior of $\widehat{F}^-$</u>:

$$\widehat{F}_n^- = \mathrm{L}\, n - \mathrm{L}_3\, n - \mathrm{L}(d-1) + O_{\mathrm{p}}\!\left(\frac{1}{\mathrm{L}_2\, n}\right).$$

(b) <u>Top outer boundaries for $\widehat{F}^-$</u>: *If $d \geq 3$, then*

$$\mathbb{P}(\widehat{F}_n^- \geq \mathrm{L}\, n - \mathrm{L}_3\, n - \mathrm{L}(d-2) + \mathrm{L}\, 2 + c \text{ i.o.}) = 0 \text{ if } c > 0.$$

(c) <u>Bottom outer boundaries for $\widehat{F}^-$</u>:

$$\mathbb{P}(\widehat{F}_n^- \leq \mathrm{L}\, n - \mathrm{L}_3\, n - \mathrm{L}\, d - \mathrm{L}\, 2 - c \text{ i.o.}) = 0 \text{ if } c > 0.$$

As a corollary, the process $\widehat{F}^-$ satisfies the same assertions as for F^- in Corollary 1.9.

▶ **Remark 1.12.** Combining Theorems 1.8 and 1.11, we find that there is little difference between the two processes, in the sense that

$$\widehat{F}_n^- - F_n^- \xrightarrow{\ \mathrm{P}\ } 0,$$

because in fact $0 \leq \widehat{F}_n^- - F_n^- = O_{\mathrm{p}}\!\left(\frac{\mathrm{L}_3\, n}{\mathrm{L}_2\, n}\right)$.

▶ **Remark 1.13.** Extending Theorem 1.11, we conjecture that

$$(\mathrm{L}_2\, n)\left(\widehat{F}_n^- - [\mathrm{L}\, n - \mathrm{L}_3\, n - \mathrm{L}(d-1)]\right) \tag{1.3}$$

has a nondegenerate limiting distribution. This is discussed further in Remark 3.3.

1.3 Outline of extended abstract

The proof of Theorem 1.8 relies on Theorem 1.11, so we tackle the latter first. In Sections 2–3 we apply the first moment method and the second moment method, respectively, to the number of remaining records with suitably small coordinate-sum; this leads to the proof of Theorem 1.11 in Appendix B. In Sections 4–5 of this extended abstract we review and extend the theory of generators developed in [5]. In Section 6 we apply the first moment method to the number of generators with suitably small coordinate sum; this, together with the upper bounds on $\widehat{F}^-$ in Theorem 1.11, leads to the proof of Theorem 1.8 in Section 7.

▶ **Remark 1.14.** Because $F_n^- \leq \widehat{F}_n^-$, Theorem 1.8(b) follows immediately from Theorem 1.11(b), as does Theorem 1.11(c) from Theorem 1.8(c).

More notation.   Throughout the extended abstract, the boundaries we consider will without exception have the form

$$b_n := \mathrm{L}\, n - \mathrm{L}_3\, n - \mathrm{L}\, c_n \text{ with } c_n > 0 \text{ and } c_n = \Theta(1). \tag{1.4}$$

Also, we will often use the notation

$$\beta_n := n e^{-b_n} \tag{1.5}$$

The dimension $d \geq 2$ will always remain fixed as $n \to \infty$.

2   Stochastic lower bound on $\widehat{F}_n^-$ via the first moment method

In this section we show how to obtain a suitable stochastic lower bound on $\widehat{F}_n^-$. See Proposition 2.3 for the result. The idea, for a suitably chosen sequence (b_n) is to apply the first moment method (computation of sufficiently small mean, together with application of Markov's inequality) to the count $\rho_n(b_n)$, where

$$\rho_n(b) := \#\{\text{remaining records } r \text{ at epoch } n \text{ with } r_+ \le b\}. \tag{2.1}$$

Asymptotic determination of the mean is obtained by suitably modifying the asymptotic determination of the mean of $\rho_n = \rho_n(\infty)$ in [2, Section 2].

2.1   Upper (and lower) asymptotic bound(s) on mean

In the next lemma we determine detailed asymptotics for the mean of $\rho_n(b_n)$ when (b_n) is a boundary of interest in establishing Theorems 1.8 and 1.11. The proof is rather elementary, but we defer it to Appendix A. We define

$$J_j(x) := \int_x^\infty (\mathrm{L}\, z)^j e^{-z}\, \mathrm{d}z. \tag{2.2}$$

and note that $J_j(x) \sim (\mathrm{L}\, x)^j e^{-x}$ as $x \to \infty$.

▶ **Lemma 2.1.** *With the notation and assumptions of (1.4)–(1.5) and (2.2), as $n \to \infty$ we have*

$$\mathbb{E}\, \rho_n(b_n) = [1 + O(n^{-1}(\mathrm{L}_2\, n)^2)] \frac{1}{(d-1)!} \sum_{j=0}^{d-1} (-1)^j \binom{d-1}{j} (\mathrm{L}\, n)^{d-1-j} J_j(\beta_n), \tag{2.3}$$

or, equivalently,

$$\mathbb{E}\, \rho_n(b_n) = \frac{1}{(d-1)!} \sum_{j=0}^{d-1} (-1)^j \binom{d-1}{j} (\mathrm{L}\, n)^{d-1-j} J_j(\beta_n) + O(n^{-1}(\mathrm{L}\, n)^{d-1-c_n}(\mathrm{L}_2\, n)^2). \tag{2.4}$$

▶ Remark 2.2. We need only lead-order asymptotics for the mean in this section, but (as seen in the proof of Lemma 3.1 found in the full-length paper) we require much more detailed asymptotics for it in the next section – asymptotics with an additive $o(1)$ remainder term, as we have in (2.4).

2.2   Stochastic lower bound on $\widehat{F}_n^-$

We are now in position to apply Markov's inequality to bound the probability of the event $\{\widehat{F}_n^- \le b_n\} = \{\rho_n(b_n) \ge 1\}$.

▶ **Proposition 2.3** (Stochastic lower bound on $\widehat{F}_n^-$). *With the notation and assumptions of (1.4), as $n \to \infty$ we have*

$$\mathbb{P}(\widehat{F}_n^- \le b_n) \le \mathbb{E}\, \rho_n(b_n) = (1 + o(1)) \frac{1}{(d-1)!} (\mathrm{L}\, n)^{d-1-c_n}.$$

3   Stochastic upper bound on $\widehat{F}_n^-$ via second moment method

In this section we show how to obtain a suitable stochastic upper bound on $\widehat{F}_n^-$ (and thus also on F_n^-). See Proposition 3.2 for the result. The idea, for a suitably chosen sequence (b_n), is to apply the second moment method (computation of sufficiently large mean and sufficiently small variance, together with application of Chebyshev's inequality) to the count $\rho_n(b_n)$ [recall the definition (2.1)], which almost surely equals

$$\rho_n^\circ(b_n) := \#\{\text{remaining records } r \text{ at epoch } n \text{ with } r_+ < b_n\}. \tag{3.1}$$

For the mean, we will use Lemma 2.1. The bound on the variance of $\rho_n(b_n)$ is obtained by suitably modifying the already quite technical asymptotic determination of the variance of $\rho_n = \rho_n(\infty)$ in [2, Section 2]; the determination here is quite a bit more technical still.

3.1   Upper bound on variance

We next show that the standard deviation of $\rho_n(b_n)$ is of smaller order of magnitude than the mean – and by enough so that our proof (found in the full-length paper) of Theorem 1.8(b) (for $\widehat{F}^-$, which implies the result for F^-) using the first Borel–Cantelli lemma will succeed. The rather long and rather computationally technical proof of the following result is left for the full-length paper, where the reverse inequality (not needed in this extended abstract) is also established.

▶ **Lemma 3.1.** *With the notation and assumptions of* (1.4)*, as $n \to \infty$ we have*

$$\operatorname{Var} \rho_n(b_n) \le (1 + o(1)) \, \mathbb{E}\, \rho_n(b_n). \tag{3.2}$$

3.2   Stochastic upper bound on $\widehat{F}_n^-$

We are now in position to utilize Chebyshev's inequality to provide a bound on $\mathbb{P}(\widehat{F}_n^- \ge b_n) = \mathbb{P}(\rho_n^\circ(b_n) = 0) = \mathbb{P}(\rho_n(b_n) = 0)$.

▶ **Proposition 3.2** (Stochastic upper bound on $\widehat{F}_n^-$)**.** *With the notation and assumptions of* (1.4)*, as $n \to \infty$ we have*

$$\mathbb{P}(F_n^- \ge b_n) \le \mathbb{P}(\widehat{F}_n^- \ge b_n) \le (1 + o(1))\,(d-1)!(\mathrm{L}\,n)^{-(d-1-c_n)} = O((\mathrm{L}\,n)^{-(d-1-c_n)}).$$

Proof. The first asserted inequality follows because $F_n^- \le \widehat{F}_n^-$. Moreover, using Chebyshev's inequality, Lemma 3.1, and Lemma 2.1, we find

$$\begin{aligned}
\mathbb{P}(\widehat{F}_n^- \ge b_n) &\le \mathbb{P}(\rho_n(b_n) = 0) = \mathbb{P}(\rho_n(b_n) - \mathbb{E}\,\rho_n(b_n) \le -\,\mathbb{E}\,\rho_n(b_n)) \\
&\le \frac{\operatorname{Var} \rho_n(b_n)}{[\mathbb{E}\,\rho_n(b_n)]^2} \le (1 + o(1)) \, [\mathbb{E}\,\rho_n(b_n)]^{-1} \\
&= (1 + o(1))\,(d-1)!(\mathrm{L}\,n)^{-(d-1-c_n)} \\
&= O((\mathrm{L}\,n)^{-(d-1-c_n)}),
\end{aligned}$$

as desired.   ◀

▶ **Remark 3.3.** Lemma 3.1 and the reverse inequality established in the full-length paper suggest that the law of $\rho_n(b_n)$ might be well approximated by a Poisson distribution with the same mean, but, after attempts using the Stein–Chen method (see, e.g., [4]) or the method of moments, we have been unable to prove such an approximation even in the case that $\mathbb{E}\,\rho_n(b_n)$ has a limit $\lambda \in (0, \infty)$. For fixed $a \in \mathbb{R}$, let $R_n(a)$ denote $\rho_n(b_n)$ when

$$b_n = \mathrm{L}\,n - \mathrm{L}_3\,n - \mathrm{L}(d-1) + \frac{a}{\mathrm{L}_2\,n}, \tag{3.3}$$

i.e., when $c_n = (d-1)e^{-a/\operatorname{L}_2 n}$ in (1.4). Even if a Poisson approximation should fail, we certainly conjecture that $R_n(a)$ converges in distribution to a nondegenerate $R(a)$ as $n \to \infty$ with $\mathbb{P}(R(a) = 0)$ continuous and strictly decreasing in a. In that case, it follows that (1.3) has limiting distribution function $a \mapsto \mathbb{P}(R(a) \geq 1)$.

In particular, if $R(a)$ is Poisson distributed for every a, then (1.3) converges in distribution to $-G^*$, where G^* has a Gumbel distribution with location $-\frac{\operatorname{L}[(d-1)!]}{d-1}$ and scale $\frac{1}{d-1}$.

4 Characterization of generators

The unpublished manuscript [5] by Fill and Naiman developed the concept of generators of multivariate records mainly in connection with an importance-sampling algorithm for generating (simulating) records. We shall find the same concept crucial for our improvement Theorem 1.8(c) to Theorem 1.7(c2), the latter of which was established using a quite different idea, namely, a certain geometric lemma [6, Lemma 3.1]. Accordingly, in this section and the next we review and extend the theory of generators developed in [5]. In this section we provide a characterization of the set of generators that is useful in counting them.

▶ **Definition 4.1.** *Suppose $x \in [0, \infty)^d$.*

(a) *The* closed positive orthant generated *(or* determined*) by x is the set*

$$O_x^+ := \{y \in [0, \infty)^d : y \geq x\}.$$

(b) *The minimum points of the frontier F_n are called* generators. *We denote the set of generators at time n by G_n.*

▶ Remark 4.2.

(a) The record-setting region RS_n equals the union $\cup_{g \in G_n} O_g^+$ of closed positive orthants. The elements of G_n are called generators because RS_n is the up-set in $[0, \infty)^d$ generated by G_n with respect to the partial order $\leq$.

(b) The almost surely unique generator with minimum coordinate-sum is the trailing point τ_n, just as the remaining record with maximum coordinate-sum is the leading point λ_n.

There are 11 generators in Figure 1, including the trailing point τ_n at the intersection of F_n and the dotted hyperplane (line) marked with $x_+ = F_n^-$. In terminology we shall establish shortly, 9 of these are interior (i.e., 2-dimensional) generators and 2 of them are 1-dimensional generators.

We now proceed to characterize the set of generators.

Denote the $\rho \equiv \rho_n$ current records at a given time n by $r^{(1)}, \ldots, r^{(\rho)}$ (listed here in arbitrary, but fixed, order). The record-setting region $S \equiv \mathrm{RS}_n$ is then the closed set

$$S = \cap_{i=1}^{\rho} \left[\cup_{k=1}^d O^+\left(r_k^{(i)} e^{(k)}\right) \right] = \cup_{k_1=1}^d \cdots \cup_{k_\rho=1}^d \cap_{i=1}^{\rho} O^+\left(r_{k_i}^{(i)} e^{(k_i)}\right)$$
$$= \cup_{k_1=1}^d \cdots \cup_{k_\rho=1}^d O^+\left(\vee_{i=1}^{\rho} r_{k_i}^{(i)} e^{(k_i)}\right) = \cup_{k \in [d]^{[\rho]}} O^+\left(R_1^{(\Pi_1(k))}, \ldots, R_d^{(\Pi_d(k))}\right),$$

where $e^{(k)}$ denotes the kth standard basis vector and for $j \in [d]$ and $k \in [d]^{[\rho]}$ we have defined the ordered partition $\Pi(k) = (\Pi_1(k), \ldots, \Pi_d(k))$ of $[\rho]$ by

$$\Pi_j(k) := k^{-1}(\{j\}) = \{i \in [\rho] : k_i = j\},$$

and for $j \in [d]$ and $P \subseteq [\rho]$ we have defined

$$R_j^{(P)} := \vee_{i \in P} r_j^{(i)}.$$

Therefore we have the neat representation

$$S = \bigcup O^+ \left(R_1^{(\Pi_1)}, \ldots, R_d^{(\Pi_d)} \right), \tag{4.1}$$

where the union here is taken over all ordered partitions $\Pi = (\Pi_1, \ldots, \Pi_d)$ of $[\rho]$ into d sets; each Π_j is allowed to be empty, in which case $R_j^{(\Pi_j)} := 0$. This shows immediately that every element of $G \equiv G_n$ has in each coordinate either 0 or the value of some record in that coordinate.

To simplify our characterization of generators, we begin by considering only "interior" generators. For any point $x \in O_0^+$, let $\nu(x)$ denote the set of non-zero coordinates of x, and observe that x lies in the interior of O_0^+ if and only if $\nu(x) = [d]$. We call such a point x an *interior* point.

Observe that a point x of the form $\left(R_1^{(\Pi_1)}, \ldots, R_d^{(\Pi_d)} \right)$ appearing in (4.1) is interior if and only if all the cells Π of the partition are nonempty. Next, note that $x \in (0, \infty)^d$ is of such a form if and only if there exist d distinct indices $i_1, \ldots, i_d$ such that $x_j = r_j^{(i_j)}$ for $j \in [d]$.

We are now in position to state and (in Appendix C) prove a characterization of the set I of interior generators. (Note that $I \subset G \subset S$.)

▶ **Theorem 4.3.** *A point* $g \in [0, \infty)^d$ *belongs to* I *if and only if*
  (i) $g \in S$, *and*
  (ii) *there exist* d *distinct indices* $i_1, \ldots, i_d$ *such that*

$$g_j = r_j^{(i_j)} = \min \left\{ r_j^{(i_\ell)} : \ell \in [d] \right\} \text{ for every } j \in [d]. \tag{4.2}$$

▶ Remark 4.4. Theorem 4.3 gives an injection from the set of interior generators into the set of ordered d-tuples of remaining records.

Now that we have characterized the interior generators, it is straightforward to characterize G in terms of projections of the current records to lower-dimensional coordinate subspaces, but some care must be taken to ensure that the almost sure property of having no coordinate ties remains true after projection. To begin a careful description, given a subset $T = \{j_1, \ldots, j_t\}$ of $[d]$ with $|T| = t \in [d]$ and $1 \leq j_1 < \cdots < j_t \leq d$, define the *projection mapping* $\pi_T : \mathbb{R}^d \to \mathbb{R}^t$ by

$$\pi_T(x_1, \ldots, x_d) := (x_{j_1}, \ldots, x_{j_t}),$$

and define the *injection mapping* $\iota_T : \mathbb{R}^t \to \mathbb{R}^d$ by

$$\iota_T(x_1, \ldots, x_t) := \vee_{k=1}^t x_{j_k} e^{(j_k)}.$$

Recall that $\nu(x)$ denotes the set of nonzero coordinates of a point $x \in [0, \infty)^d$. Define the set of *T-generators* to be the set

$$G_T := G \cap \{x : \nu(x) = T\}$$

and observe that G is the disjoint union

$$G = \cup_{T \subseteq [d]} G_T.$$

This observation, together with a characterization of each G_T, thus provides a characterization of G. A characterization of each G_T is obtained by combining the following theorem with Theorem 4.3.

To set up the statement of the theorem, consider the image

$$R_T := \pi_T(R) = \left\{ \pi_T(r^{(i)}) : i \in [\rho] \right\} \subset \mathbb{R}^{|T|}$$

under π_T of the set $R := \left\{ r^{(i)} : i \in [\rho] \right\}$ of current records, and note that R_T inherits the property of "no ties in any coordinate" from R. Let I_T denote the set of interior generators of R_T, and let $G'_T := \iota_T(I_T)$ denote the injection of I_T into $\mathbb{R}^d$.

▶ **Theorem 4.5.** *For every $T \subseteq [d]$ we have $G_T = G'_T$.*

In light of Theorem 4.5 (which is proved in Appendix C), we call the number of nonzero coordinates of a generator its *dimension*. Figure 2 shows the generators of various dimensions for an example with $d = 3$.
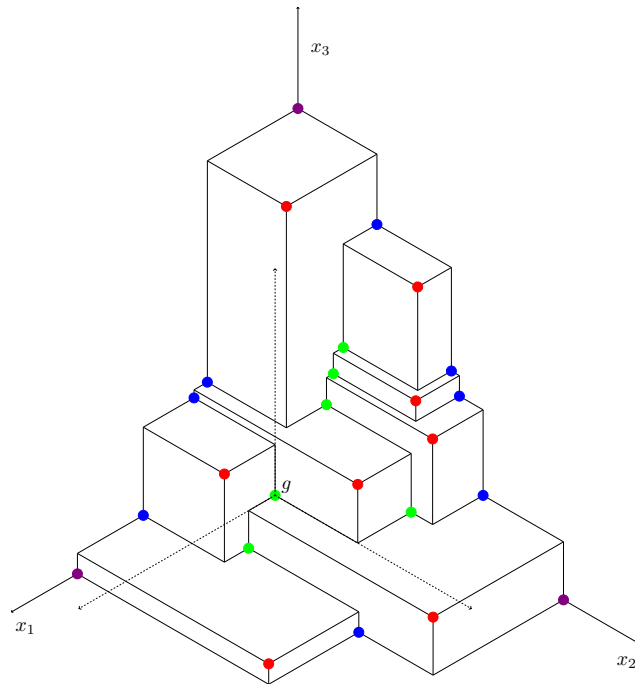
Figure 2 Example of a record frontier in dimension $d = 3$ with $\rho = 8$ remaining records shown in red and the resulting $\gamma = 17$ generators: three one-dimensional generators shown in violet, eight two-dimensional generators shown in blue, and six three-dimensional (interior) generators shown in green. The lower boundary of one of the orthants O_g^+ is shown using dashed lines.

▶ **Example 4.6.** Suppose $d = 4$ and the current records are $(2, 8, 3, 7)$ and $(5, 1, 4, 6)$. Then $|G| = 8$, because $|G_T| = 1$ for precisely eight nonempty subsets T of $[4]$ and $|G_T| = 0$ otherwise. The eight subsets T for which $|G_T| = 1$ are

$$G_{\{1\}} = \{(5, 0, 0, 0)\}; \ G_{\{2\}} = \{(0, 8, 0, 0)\}; \ G_{\{3\}} = \{(0, 0, 4, 0)\};$$
$$G_{\{4\}} = \{(0, 0, 0, 7)\}; \ G_{\{1,2\}} = \{(2, 1, 0, 0)\}; \ G_{\{1,4\}} = \{(2, 0, 0, 6)\};$$
$$G_{\{2,3\}} = \{(0, 1, 3, 0)\}; \ G_{\{3,4\}} = \{(0, 0, 3, 6)\}.$$

Thus there are four one-dimensional generators, four two-dimensional generators, and no generators with dimension exceeding two.

5   The expected number of generators

The proof of Theorem 1.8(c) requires a tight upper bound on the expected number of generators at time n with suitably small coordinate-sum. In this section we warm up with a result of independent interest, giving an asymptotic approximation for the expected total number of generators at time n. We remark in passing that such an approximation proves useful in the analysis of the importance-sampling record-generating scheme described in [5, Sections 2–4].

5.1   Exact expressions

Let $\gamma_{d,n}$ (respectively, $\iota_{d,n}$) denote the number of generators (resp., interior generators) after a given number n of d-dimensional observations. Our first result relates the expectations of these two quantities.

▶ **Lemma 5.1.** *For integers $d \geq 0$ and $n \geq 0$, we have*

$$\mathbb{E}\,\gamma_{d,n} = \sum_{k=0}^{d} \binom{d}{k} \mathbb{E}\,\iota_{k,n}, \tag{5.1}$$

Proof. This is immediate from Theorem 4.5 and the discussion preceding that theorem.   ◀

In Lemma 5.1, note that $\iota_{0,n} = \delta_{0,n}$: There is a single 0-dimensional generator (namely, the origin in $\mathbb{R}^d$) if $n = 0$ and no 0-dimensional generators otherwise. Also note that $\iota_{d,n} = 0$ if $n < d$.

The next result (proved in Appendix C) gives an exact expression for $\mathbb{E}\,\iota_{d,n}$ for $n \geq d \geq 1$. We write $n^{\underline{k}}$ for the falling factorial power

$$n(n-1)\cdots(n-k+1) = k!\binom{n}{k}.$$

▶ **Lemma 5.2.** *For integers $n \geq d \geq 1$, we have*

$$\mathbb{E}\,\iota_{d,n} = n^{\underline{d}} \int_{(0,1]^d} x_\times^{d-1}(1 - x_\times)^{n-d}\,\mathrm{d}x. \tag{5.2}$$

▶ Remark 5.3.

(a) The exact expression (5.2) in Lemma 5.2 may be compared to a similar expression for $\mathbb{E}\,\rho_{d,n}$ derived in [2, Section 2]: For $d \geq 1$ and $n \geq 1$ we have

$$\mathbb{E}\,\rho_{d,n} = n \int_{(0,1]^d} (1 - x_\times)^{n-1}\,\mathrm{d}x. \tag{5.3}$$

In fact, by expanding the factor $x_\times^{d-1}$ appearing in the integrand in (5.2) as

$$[1 - (1 - x_\times)]^{d-1} = \sum_{j=0}^{d-1} (-1)^j \binom{d-1}{j} (1 - x_\times)^j,$$

one sees that the expected counts of interior generators and expected counts of remaining records are related by

$$\mathbb{E}\,\iota_{d,n} = n^{\underline{d}} \sum_{j=0}^{d-1} (-1)^j \frac{\binom{d-1}{j}}{n - d + j + 1} \mathbb{E}\,\rho_{d,n-d+j+1}$$

for $n \geq d \geq 1$. But we do not know of any use for this connection.

(b) An alternative expression to (5.3) is

$$\mathbb{E}\,\rho_{d,n} = \sum_{j=1}^{n} (-1)^{j-1} \binom{n}{j} j^{-(d-1)} =: \widehat{H}_n^{(d-1)},$$

a so-called *Roman harmonic number* studied by [8], [9], [10].

5.2   Asymptotics

From here we follow the same outline as for the expected number of remaining records in Bai et al. [2] to obtain an asymptotic expansion for $\mathbb{E}\,\iota_{d,n}$ (see our Theorem 5.7, the main result of Section 5). Accordingly, we begin by considering a Poissonized analogue of $\mathbb{E}\,\iota_{d,n}$, whose proof is rather simple and is included in the full-length paper.

▶ **Lemma 5.4.** *For integers $d \geq 1$ and $n \geq 0$, define*

$$\hat{\iota}_{d,n} := n^d \int_{[0,1)^d} x_\times^{d-1} \exp(-n x_\times)\,\mathrm{d}x.$$

Then, for fixed d, as $n \to \infty$ we have

$$\hat{\iota}_{d,n} = (\mathrm{L}\,n)^{d-1} \sum_{j=0}^{d-1} \frac{(-1)^j \Gamma^{(j)}(d)}{j!(d-1-j)!} (\mathrm{L}\,n)^{-j} + O((n\,\mathrm{L}\,n)^{d-1} e^{-n}).$$

We next bound the difference between $\hat{\iota}_{d,n}$ and

$$\tilde{\iota}_{d,n} := n^d \int_{[0,1)^d} x_\times^{d-1} (1 - x_\times)^n\,\mathrm{d}x. \tag{5.4}$$

▶ **Lemma 5.5.** *For fixed $d \geq 1$, as $n \to \infty$ we have*

$$0 \leq \hat{\iota}_{d,n} - \tilde{\iota}_{d,n} = O(n^{-1}(\mathrm{L}\,n)^{d-1}).$$

Proof. We utilize the elementary inequality

$$e^{-nt}(1 - nt^2) \leq (1-t)^n \leq e^{-nt}$$

for $n \geq 1$ and $0 \leq t \leq 1$ (see [3, Lemma 5]). This yields

$$0 \leq \hat{\iota}_{d,n} - \tilde{\iota}_{d,n} \leq n^{d+1} \int_{[0,1)^d} x_\times^{d+1} \exp(-n x_\times)\,\mathrm{d}x.$$

Proceeding just as in the proof of Lemma 5.4, we find that the last expression here is $O(n^{-1}(\mathrm{L}\,n)^{d-1})$. ◀

▶ **Theorem 5.6.** *For fixed $d \geq 1$, as $n \to \infty$ the expected number of interior generators at time n in dimension d satisfies*

$$\mathbb{E}\,\iota_{d,n} = (\mathrm{L}\,n)^{d-1} \sum_{j=0}^{d-1} \frac{(-1)^j \Gamma^{(j)}(d)}{j!(d-1-j)!} (\mathrm{L}\,n)^{-j} + O(n^{-1}(\mathrm{L}\,n)^{d-1}).$$

Proof. Comparing (5.2) and (5.4) and then invoking Lemma 5.5, we see that

$$
\mathbb{E}\,\iota_{d,n} = \frac{n^{\underline{d}}}{(n-d)^d}\tilde{\iota}_{d,n-d} = [1 + O(n^{-1})]\,\tilde{\iota}_{d,n-d}
$$
$$
= [1 + O(n^{-1})]\,[\hat{\iota}_{d,n-d} + O(n^{-1}(\mathrm{L}\,n)^{d-1})]
$$
$$
= [1 + O(n^{-1})]\,\hat{\iota}_{d,n-d} + O(n^{-1}(\mathrm{L}\,n)^{d-1}).
$$

But, according to Lemma 5.4,

$$
\hat{\iota}_{d,n-d} = [\mathrm{L}(n-d)]^{d-1}\sum_{j=0}^{d-1}\frac{(-1)^j\Gamma^{(j)}(d)}{j!(d-1-j)!}[\mathrm{L}(n-d)]^{-j} + O((n\,\mathrm{L}\,n)^{d-1}e^{-n})
$$
$$
= (\mathrm{L}\,n)^{d-1}\sum_{j=0}^{d-1}\frac{(-1)^j\Gamma^{(j)}(d)}{j!(d-1-j)!}(\mathrm{L}\,n)^{-j} + O(n^{-1}(\mathrm{L}\,n)^{d-2}).
$$

Thus

$$
\mathbb{E}\,\iota_{d,n} = [1 + O(n^{-1})]\,(\mathrm{L}\,n)^{d-1}\sum_{j=0}^{d-1}\frac{(-1)^j\Gamma^{(j)}(d)}{j!(d-1-j)!}(\mathrm{L}\,n)^{-j} + O(n^{-1}(\mathrm{L}\,n)^{d-1})
$$
$$
= (\mathrm{L}\,n)^{d-1}\sum_{j=0}^{d-1}\frac{(-1)^j\Gamma^{(j)}(d)}{j!(d-1-j)!}(\mathrm{L}\,n)^{-j} + O(n^{-1}(\mathrm{L}\,n)^{d-1}),
$$

as claimed. ◀

Combining (5.1) and (5.2), we can obtain an exact expression for $\mathbb{E}\,\gamma_{d,n}$. Similarly, combining (5.1) and Theorem 5.6 we obtain the following asymptotic expansion in powers of logarithm for $\mathbb{E}\,\gamma_{d,n}$ after a little rearrangement.

▶ **Theorem 5.7.** *For fixed $d \geq 1$, as $n \to \infty$ the expected number of generators at time n in dimension d satisfies*

$$
\mathbb{E}\,\gamma_{d,n} = (\mathrm{L}\,n)^{d-1}\sum_{j=0}^{d-1}a_{d,j}(\mathrm{L}\,n)^{-j} + O(n^{-1}(\mathrm{L}\,n)^{d-1}),
$$

where

$$
a_{d,j} := \sum_{k=0}^{j}\binom{d}{d-j+k}\frac{(-1)^k\Gamma^{(k)}(d-j+k)}{k!(d-1-j)!}.
$$

▶ Remark 5.8. Concerning Theorem 5.7:
(a) In particular, $a_{d,0} = 1$, so $\mathbb{E}\,\gamma_{d,n}$ has lead-order asymptotics

$$
\mathbb{E}\,\gamma_{d,n} = (\mathrm{L}\,n)^{d-1} + O((\mathrm{L}\,n)^{d-2});
$$

this is $(d-1)!$ times as large as the lead-order asymptotics for the expected number of remaining records, namely,

$$
\mathbb{E}\,\rho_{d,n} = \frac{(\mathrm{L}\,n)^{d-1}}{(d-1)!} + O((\mathrm{L}\,n)^{d-2}).
$$

(b) For $d = 2$ and $n \geq 0$, we have

$$\mathbb{E}\,\gamma_{2,n} = H_n + 1 = \mathbb{E}\,\rho_{2,n} + 1,$$

where $H_n := \sum_{k=1}^{n} k^{-1}$ is the nth harmonic number; and in fact it is easy to see that $\gamma_{2,n} = \rho_{2,n} + 1$. For $d = 3$ and $n \geq 0$, we have

$$\mathbb{E}\,\gamma_{3,n} = H_n^2 + H_n^{(2)} + 1 = 2\,\mathbb{E}\,\rho_{3,n} + 1,$$

where $H_n^{(2)} := \sum_{k=1}^{n} k^{-2}$ is the nth second-order harmonic number; and in fact $\gamma_{3,n} = 2\rho_n + 1$, as established in [5, Corollary 6.6]. There is not such a simple relationship between the exact values of $\rho_{d,n}$ and $\gamma_{d,n}$ for $d \geq 4$; confer [5, Remark 6.7].

(c) We hope to extend the work of Section 5.2 by finding at least lead-order asymptotics for the variance, and also a normal approximation or other limit theorem, for the number $\gamma_{d,n}$ of generators after n observations.

6 Stochastic lower bound on F_n^- via the first moment method

In this section we show how to obtain a suitable stochastic lower bound on F_n^-. See Proposition 6.2 for the result. The idea, for a suitably chosen sequence (b_n) is to apply the first moment method (computation of sufficiently small mean, together with application of Markov's inequality) to the count

$$\gamma_n(b) := (\text{number of generators at epoch } n \text{ with coordinate-sum} \leq b).$$

The bound on the mean of $\gamma_n(b_n)$ is obtained by suitably modifying the proof of Theorem 5.7 [compare also the similar treatment of $\rho_n(b_n)$ in Section 3 and the full-length paper].

▶ **Lemma 6.1.** *With the notation and assumptions of* (1.4)–(1.5) *and* (2.2), *as $n \to \infty$ we have*

$$\mathbb{E}\,\gamma_n(b_n) \leq (1 + o(1)) \frac{(\mathrm{L}\,n)^{d-1}}{(d-1)!} (c_n\,\mathrm{L}_2\,n)^{d-1} (\mathrm{L}\,n)^{-c_n}. \tag{6.1}$$

Proof. We will be very brief here. Following very closely along the lines of Section 5, one finds that

$$
\begin{aligned}
\mathbb{E}\,\gamma_n(b_n) &\sim \frac{1}{(d-1)!} \int_{ne^{-b_n}}^{\mathrm{L}\,n} (\mathrm{L}\,n - \mathrm{L}\,z)^{d-1} z^{d-1} e^{-z} \, \mathrm{d}z \\
&\leq \frac{(\mathrm{L}\,n)^{d-1}}{(d-1)!} \int_{ne^{-b_n}}^{\infty} z^{d-1} e^{-z} \, \mathrm{d}z \\
&\sim \frac{(\mathrm{L}\,n)^{d-1}}{(d-1)!} (ne^{-b_n})^{d-1} \exp(-ne^{-b_n}) \\
&= \frac{(\mathrm{L}\,n)^{d-1}}{(d-1)!} (c_n\,\mathrm{L}_2\,n)^{d-1} (\mathrm{L}\,n)^{-c_n}. \qquad \blacktriangleleft
\end{aligned}
$$

We are now in position to utilize Markov's inequality.

▶ **Proposition 6.2** (Stochastic lower bound on F_n^-)**.** *Fix $d \geq 2$. If $1 \leq c_n = O(1)$ and*

$$b \equiv b_n := \mathrm{L}\,n - \mathrm{L}_3\,n - \mathrm{L}\,c_n,$$

then

$$\mathbb{P}(F_n^- \leq b_n) \leq \mathbb{E}\,\gamma_n(b_n) \leq (1 + o(1)) \frac{(\mathrm{L}\,n)^{d-1}}{(d-1)!} (c_n\,\mathrm{L}_2\,n)^{d-1} (\mathrm{L}\,n)^{-c_n}.$$

7   Proof of Theorem 1.8

In this section we prove Theorem 1.8.

Proof of Theorem 1.8.

(a) This follows readily from Propositions 6.2 and 3.2 (or one can invoke Theorem 1.11 instead of Proposition 3.2).

(b) As noted in Remark 1.14, this is immediate from Theorem 1.11(b), already established in Appendix B.

(c) This follows in the same fashion as our given proof of Theorem 1.11(c), now using Proposition 6.2 in place of Proposition 2.3. We leave the routine details to the reader.   ◀

References

1 Barry C. Arnold, N. Balakrishnan, and H. N. Nagaraja. *Records.* Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons, Inc., New York, 1998. A Wiley-Interscience Publication. `doi:10.1002/9781118150412`.

2 Zhi-Dong Bai, Luc Devroye, Hsien-Kuei Hwang, and Tsung-Hsi Tsai. Maxima in hypercubes. *Random Structures Algorithms*, 27(3):290–309, 2005. `doi:10.1002/rsa.20053`.

3 Zhi-Dong Bai, Hsien-Kuei Hwang, Wen-Qi Liang, and Tsung-Hsi Tsai. Limit theorems for the number of maxima in random samples from planar regions. *Electron. J. Probab.*, 6:no. 3, 41, 2001. `doi:10.1214/EJP.v6-76`.

4 A.D. Barbour, Lars Holst, and Svante Janson. *Poisson Approximation.* Oxford studies in Probability. Oxford University Press, New York, 1992.

5 James Allen Fill and Daniel Q. Naiman. Generating Pareto records, 2019. arXiv:1901.05621. `arXiv:1901.05621`.

6 James Allen Fill and Daniel Q. Naiman. The Pareto record frontier. *Electron. J. Probab.*, 25:Paper No. 92, 24, 2020. `doi:10.1214/20-ejp492`.

7 J. Kiefer. Iterated logarithm analogues for sample quantiles when $p_n \downarrow 0$. *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971), Vol. I: Theory of statistics*, pages 227–244, 1972.

8 Daniel E. Loeb and Gian-Carlo Rota. Formal power series of logarithmic type. *Adv. Math.*, 75(1):1–118, 1989. `doi:10.1016/0001-8708(89)90079-0`.

9 Steven Roman. The logarithmic binomial formula. *Amer. Math. Monthly*, 99(7):641–648, 1992. `doi:10.2307/2324994`.

10 J. Sesma. The Roman harmonic numbers revisited. *J. Number Theory*, 180:544–565, 2017. `doi:10.1016/j.jnt.2017.05.009`.

A   Proof of Lemma 2.1

This appendix is devoted to the (elementary) proof of Lemma 2.1.

Proof of Lemma 2.1. We will prove (2.3) by separately considering (a) upper and (b) lower bounds. Before beginning, we note that the mean in question has the exact expression

$$\mathbb{E}\,\rho_n(b_n) = n \int_{x \geq 0:\, x_+ \leq b_n} e^{-x_+}(1-e^{-x_+})^{n-1}\, \mathrm{d}x = \frac{n}{(d-1)!} \int_0^{b_n} y^{d-1} e^{-y}(1-e^{-y})^{n-1}\, \mathrm{d}y. \quad \text{(A.1)}$$

A key technical tool we will use is the pair of elementary inequalities

$$e^{-nt}(1 - nt^2) \leq (1-t)^n \leq e^{-nt} \qquad\qquad\qquad\qquad \text{(A.2)}$$

for $n \geq 1$ and $0 \leq t \leq 1$ (see [3, Lemma 5]). Also, note from the definition (2.2) of the function J_j that

$$J_j(x) \sim (\mathrm{L}\, x)^j e^{-x} \text{ as } x \to \infty \tag{A.3}$$

and that for $1 \leq x < y$ we have

$$0 < J_j(x) - J_j(y) \leq (\mathrm{L}\, y)^j (e^{-x} - e^{-y}) = (\mathrm{L}\, y)^j e^{-y}(e^{y-x} - 1). \tag{A.4}$$

(a) Utilizing the upper bound in (A.2) immediately we derive

$$
\begin{aligned}
&\mathbb{E}\, \rho_{n+1}(b_{n+1}) \\
&= \frac{n+1}{(d-1)!} \int_0^{b_{n+1}} y^{d-1} e^{-y}(1 - e^{-y})^n \, \mathrm{d}y \\
&\leq \frac{n+1}{(d-1)!} \int_0^{b_{n+1}} y^{d-1} \exp\left(-n e^{-y} - y\right) \mathrm{d}y \\
&= \frac{1 + n^{-1}}{(d-1)!} \int_{ne^{-b_{n+1}}}^n (\mathrm{L}\, n - \mathrm{L}\, z)^{d-1} e^{-z} \, \mathrm{d}z \\
&\leq \frac{1 + n^{-1}}{(d-1)!} \int_{ne^{-b_{n+1}}}^{n+1} [\mathrm{L}(n+1) - \mathrm{L}\, z]^{d-1} e^{-z} \, \mathrm{d}z \\
&= \frac{1 + n^{-1}}{(d-1)!} \sum_{j=0}^{d-1} (-1)^j \binom{d-1}{j} [\mathrm{L}(n+1)]^{d-1-j} \int_{ne^{-b_{n+1}}}^{n+1} (\mathrm{L}\, z)^j e^{-z} \, \mathrm{d}z \\
&= \frac{1 + n^{-1}}{(d-1)!} \sum_{j=0}^{d-1} (-1)^j \binom{d-1}{j} [\mathrm{L}(n+1)]^{d-1-j} [J_j(ne^{-b_{n+1}}) - J_j(n+1)].
\end{aligned}
\tag{A.5}
$$

That is,

$$\mathbb{E}\, \rho_n(b_n) \leq \frac{1 + (n-1)^{-1}}{(d-1)!} \sum_{j=0}^{d-1} (-1)^j \binom{d-1}{j} (\mathrm{L}\, n)^{d-1-j} [J_j((n-1)e^{-b_n}) - J_j(n)]. \tag{A.6}$$

By the note following (2.2), $J_j(n) \sim (\mathrm{L}\, n)^j e^{-n}$. Moreover, by (A.4) we have

$$
\begin{aligned}
&0 < J_j((n-1)e^{-b_n}) - J_j(\beta_n) \\
&\leq (\mathrm{L}\, \beta_n)^j e^{-\beta_n} [\exp(e^{-b_n}) - 1] \\
&\sim (\mathrm{L}\, \beta_n)^j e^{-\beta_n} e^{-b_n} = (\mathrm{L}\, \beta_n)^j e^{-\beta_n} n^{-1} c_n \mathrm{L}_2\, n = O\!\left((\mathrm{L}\, \beta_n)^j e^{-\beta_n} n^{-1} \mathrm{L}_2\, n\right).
\end{aligned}
$$

Thus

$$
\begin{aligned}
\mathbb{E}\, \rho_n(b_n) &\leq \frac{1}{(d-1)!} \sum_{j=0}^{d-1} (-1)^j \binom{d-1}{j} (\mathrm{L}\, n)^{d-1-j} J_j(\beta_n) + O(n^{-1}(\mathrm{L}\, n)^{d-1-c_n} \mathrm{L}_2\, n) \\
&= [1 + O(n^{-1} \mathrm{L}_2\, n)] \frac{1}{(d-1)!} \sum_{j=0}^{d-1} (-1)^j \binom{d-1}{j} (\mathrm{L}\, n)^{d-1-j} J_j(\beta_n),
\end{aligned}
$$

bettering the claim in the upper-bound direction for the mean at (2.3).

(b) Utilizing the lower bound in (A.2), we find from (A.1) that

$$
\begin{aligned}
\mathbb{E}\,\rho_n(b_n) &= \frac{n}{(d-1)!} \int_0^{b_n} y^{d-1} e^{-y} (1 - e^{-y})^{n-1} \, \mathrm{d}y \\
&\geq \frac{n}{(d-1)!} \int_0^{b_n} y^{d-1} e^{-y} (1 - e^{-y})^n \, \mathrm{d}y \\
&\geq \frac{n}{(d-1)!} \int_0^{b_n} y^{d-1} \exp\left(-ne^{-y} - y\right) \left(1 - ne^{-2y}\right) \, \mathrm{d}y.
\end{aligned}
\tag{A.7}
$$

We derive that the added term in (A.7) satisfies

$$
\begin{aligned}
&\frac{n}{(d-1)!} \int_0^{b_n} y^{d-1} \exp\left(-ne^{-y} - y\right) \mathrm{d}y \\
&= \frac{1}{(d-1)!} \int_{\beta_n}^n (\mathrm{L}\,n - \mathrm{L}\,z)^{d-1} e^{-z} \, \mathrm{d}z \\
&= \frac{1}{(d-1)!} \sum_{j=0}^{d-1} (-1)^j \binom{d-1}{j} (\mathrm{L}\,n)^{d-1-j} [J_j(\beta_n) - J_j(n)].
\end{aligned}
$$

But $J_j(n) \sim e^{-n}(\mathrm{L}\,n)^j$, whence the added term in (A.7) is lower-bounded by

$$
\frac{1}{(d-1)!} \sum_{j=0}^{d-1} (-1)^j \binom{d-1}{j} (\mathrm{L}\,n)^{d-1-j} J_j(\beta_n) - O(e^{-n}(\mathrm{L}\,n)^{d-1}).
$$

So it remains to show that the subtracted term in (A.7) can be absorbed into the remainder term in (2.4), which we will do in similar (but easier) fashion to upper-bounding $\mathbb{E}\,\rho_n(b_n)$. Indeed, the subtracted term satisfies

$$
0 < \frac{n^2}{(d-1)!} \int_0^{b_n} y^{d-1} \exp\left(-ne^{-y} - 3y\right) \mathrm{d}y = \frac{n^{-1}}{(d-1)!} \int_{\beta_n}^n z^2 (\mathrm{L}\,n - \mathrm{L}\,z)^{d-1} e^{-z} \, \mathrm{d}z
\tag{A.8}
$$

$$
\begin{aligned}
&\leq \frac{n^{-1}(\mathrm{L}\,n)^{d-1}}{(d-1)!} \int_{\beta_n}^\infty z^2 e^{-z} \, \mathrm{d}z \\
&\sim \frac{n^{-1}(\mathrm{L}\,n)^{d-1}}{(d-1)!} \beta_n^2 e^{-\beta_n} \\
&= \frac{n^{-1}(\mathrm{L}\,n)^{d-1}}{(d-1)!} c_n^2 (\mathrm{L}_2\,n)^2 (\mathrm{L}\,n)^{-c_n} = O(n^{-1}(\mathrm{L}_2\,n)^2 (\mathrm{L}\,n)^{d-1-c_n}) \\
&= O\left(n^{-1}(\mathrm{L}_2\,n)^2 \frac{1}{(d-1)!} \sum_{j=0}^{d-1} (-1)^j \binom{d-1}{j} (\mathrm{L}\,n)^{d-1-j} J_j(\beta_n)\right),
\end{aligned}
$$

as desired.   ◀

B   Proof of Theorem 1.11

In this appendix we prove Theorem 1.11.

Proof of Theorem 1.11.

(a) This follows readily from Propositions 2.3 and 3.2. Here are some details. For $a \in \mathbb{R}$, let

$$
b_n(a) := \mathrm{L}\,n - \mathrm{L}_3\,n - \mathrm{L}(d-1) + \frac{a}{\mathrm{L}_2\,n},
\tag{B.1}
$$

as at (3.3); this is an instance of (1.4) with $c_n = (d-1)e^{-a/\operatorname{L_2}n}$. By Proposition 2.3,

$$\mathbb{P}(\widehat{F}_n^- \le b_n(a)) \le (1 + o(1))\frac{1}{(d-1)!}(\operatorname{L}n)^{d-1-c_n} \to \frac{1}{(d-1)!}e^{(d-1)a};$$

the last expression here tends to 0 as $a \to -\infty$. Similarly, by Proposition 3.2,

$$\mathbb{P}(\widehat{F}_n^- \ge b_n(a)) \le (1 + o(1))(d-1)!(\operatorname{L}n)^{-(d-1-c_n)} \to (d-1)!e^{-(d-1)a},$$

and the last expression here tends to 0 as $a \to \infty$. It follows that the sequence of distributions of (1.3) is tight, i.e., that Theorem 1.11(a) holds.

(b) Like F^- (Lemma 1.5), the process $\widehat{F}^-$ has nondecreasing sample paths. From this it follows that if (b_n) is (ultimately) monotone nondecreasing and (n_j) is any strictly increasing sequence of positive integers, then

$$\{\widehat{F}_n^- \ge b_n \text{ i.o.}(n)\} \subseteq \{\widehat{F}_{n_{j+1}}^- \ge b_{n_j} \text{ i.o.}(j)\}.$$

To complete the proof of part (b), we choose $b_n \equiv \operatorname{L}n - \operatorname{L_3}n - \operatorname{L}(d-2) + \operatorname{L}2 + c$ with $c > 0$ and $n_j \equiv 2^j$, bound $\mathbb{P}(\widehat{F}_{n_{j+1}}^- \ge b_{n_j})$ using Proposition 3.2, and apply the first Borel–Cantelli lemma.

Here are the details. If n is even, then

$$\begin{aligned} b_{n/2} &= \operatorname{L}(n/2) - \operatorname{L_3}(n/2) - \operatorname{L}(d-2) + \operatorname{L}2 + c = \operatorname{L}n - \operatorname{L_3}(n/2) - \operatorname{L}(d-2) + c \\ &\ge \operatorname{L}n - \operatorname{L_3}n - \operatorname{L}(d-2) + c, \end{aligned}$$

the last expression being the one in Proposition 3.2 with $c_n \equiv e^{-c}(d-2)$. Thus, by that proposition,

$$\begin{aligned} \mathbb{P}(\widehat{F}_{n_{j+1}}^- \ge b_{n_j}) &= \mathbb{P}(\widehat{F}_{n_{j+1}}^- \ge b_{n_{j+1}/2}) \\ &\le \mathbb{P}(\widehat{F}_{n_{j+1}}^- \ge \operatorname{L}n_{j+1} - \operatorname{L_3}n_{j+1} - \operatorname{L}(d-2) + c) \\ &= O((\operatorname{L}n_{j+1})^{-[d-1-e^{-c}(d-2)]}) = O((j+1)^{-[1+(1-e^{-c})(d-2)]}), \end{aligned}$$

which is summable.

(c) To prove part (c) [which, as noted in Remark 1.14, will also follow immediately once we prove Theorem 1.8(c)], we begin with an argument similar to that for part (b). If (b_n) is (ultimately) monotone nondecreasing and (n_j) is any strictly increasing sequence of positive integers, then

$$\{F_n^- \le b_n \text{ i.o.}(n)\} \subseteq \{F_{n_j}^- \le b_{n_{j+1}} \text{ i.o.}(j)\}.$$

To complete the proof of part (c), we choose $b_n \equiv \operatorname{L}n - \operatorname{L_3}n - \operatorname{L}d - \operatorname{L}2 - c$ with $c > 0$ and $n_j \equiv 2^j$, bound $\mathbb{P}(F_{n_j}^- \le b_{n_{j+1}})$ using Proposition 2.3, and apply the first Borel–Cantelli lemma.

Here are the details. First note that

$$b_{2n} = \operatorname{L}(2n) - \operatorname{L_3}(2n) - \operatorname{L}d - \operatorname{L}2 - c \le \operatorname{L}n - \operatorname{L_3}n - \operatorname{L}d - c,$$

the bounding expression being the one in Proposition 2.3 with $c_n \equiv e^c d$. Thus, by that proposition,

$$\begin{aligned} \mathbb{P}(F_{n_j}^- \le b_{n_{j+1}}) &\le \mathbb{P}(F_{n_j}^- \le \operatorname{L}n_j - \operatorname{L_3}n_j - \operatorname{L}d - c) \\ &= O((\operatorname{L}n_j)^{-[e^c d - (d-1)]}) = O(j^{-[1+(e^c-1)d]}), \end{aligned}$$

which is summable. ◀

C   Proofs of Theorems 4.3 and 4.5 and Lemma 5.2

Proof of Theorem 4.3. First suppose $g \in I$. Then (i) is automatic from the definition of I. Moreover, we know from our earlier discussion of generators that (ii) holds for $g = \left(R_1^{(\Pi_1)}, \ldots, R_d^{(\Pi_d)} \right)$ with the possible exception of the second equality in (4.2). But if that equality does not hold, let $j, \ell \in [d]$ with $j \neq \ell$ satisfy

$$r_j^{(i_\ell)} < R_j^{(\Pi_j)}.$$

We then move i_ℓ from the cell Π_ℓ to the cell Π_j in order to form a new partition, call it Π'. Then

$$g > \left(R_1^{(\Pi'_1)}, \ldots, R_d^{(\Pi'_d)} \right) \in S,$$

so g is not a generator.

Next we prove the converse. If g has these two properties, then $g \in (0, \infty)^d$ belongs to S, so all that is left to show is that g is a minimum (with respect to $\leq$) of S. Suppose that $x < g$; we will complete the proof by showing that $x \notin S$.

Let j_0 satisfy $x_{j_0} < g_{j_0}$. Then

$$x_{j_0} < g_{j_0} = r_{j_0}^{(i_{j_0})} \tag{C.1}$$

using (4.2) for the equality. Additionally, for $j \neq j_0$ we have

$$x_j \leq g_j < r_j^{(i_{j_0})}, \tag{C.2}$$

where the second inequality holds by (4.2) because

$$g_j = r_j^{(i_j)} = \min \left\{ r_j^{(i_\ell)} : \ell \in [d] \right\},$$

which almost surely is strictly smaller than $r_j^{(i_{j_0})}$ because $i_j \neq i_{j_0}$. Combining (C.1) and (C.2), we see that $x \prec r^{(i_{j_0})}$, and so $x \notin S$.   ◀

Proof of Theorem 4.5. Let $t = |T|$. There is no loss of generality (and there is some ease in notation) in supposing that $T = [t]$, and thus $x \in G_T$ if and only if $x \in G$ and $x_{t+1} = \cdots = x_d = 0$. Let $x = (x_1, \ldots, x_t, 0, \ldots, 0)$ satisfy $\nu(x) = t$. We will show that $x \in G_T$ – equivalently, that $x \in G$ – if and only if $\pi_T(x) \in I_T$ – equivalently, that $x \in \iota_T(I_T) = G'_T$.

Indeed, for x to be a generator, there are two requirements: (i) $x \in S$, and (ii) x is a minimum of S. The requirement (i) is that for each i there should exist $j \in [d]$ such that $x_j \geq r_j^{(i)}$. However, since we assume that $r^{(i)} \succ 0$, such j must belong to $[t]$. We have thus argued that x is in $S = \mathrm{RS}(R)$ (the record-setting region determined by the points in R) if and only if $\pi_T(x) \in \mathrm{RS}(R_T)$.

The requirement (ii) is that $y < x$ must imply $y \notin S$. But note that $y < x$ if and only if y is of the form $y = (y_1, \ldots, y_t, 0, \ldots, 0)$ with $\pi_T(y) < \pi_T(x)$. Thus requirement (ii) can be rephrased thus: If $y = (y_1, \ldots, y_t, 0, \ldots, 0)$ with $\pi_T(y) < \pi_T(x)$, then $y \notin \mathrm{RS}$ – equivalently, by what we argued in connection with requirement (i), that $\pi_T(y) \notin \mathrm{RS}(R_T)$.

So we have argued that x is a generator if and only if $\pi_T(x) \in I_T$, i.e., if and only if $x \in G'_T$. This is as desired.   ◀

Proof of Lemma 5.2. To facilitate the statement and proof of Lemma 5.2, and in order to follow more closely the analogous treatment of remaining records in [2, Section 2], we may and do switch from Exponential(1) observation coordinates to observations uniformly distributed in $[0,1)^d$. Referring to Theorem 4.3(ii), let us say that the d-tuple $(X^{(i_1)}, \ldots, X^{(i_d)})$ of observations (where the indices i_j are distinct elements of $[n]$) *generates* an epoch-n interior generator g if those d observations are all remaining records at epoch n and

$$g_j = X_j^{(i_j)} = \min\{X_j^{(i_\ell)} : \ell \in [d]\} \text{ for every } j \in d.$$

Note that every interior generator is generated by precisely one such generating d-tuple. Thus $\mathbb{E}\,\iota_{d,n}$ equals $n^{\underline{d}}$ times the probability that $(X^{(1)}, \ldots, X^{(d)})$ generates an interior generator. Condition on the value $\mathbf{y} := (x^{(1)}, \ldots, x^{(d)})$ of this d^2-tuple. According to Theorem 4.3, in order for $\mathbf{y}$ to generate an interior generator, two conditions are required. One is that

$$x_j^{(\ell)} \geq x_j^{(j)} \text{ for every } \ell, j \in d \text{ with } \ell \neq j. \tag{C.3}$$

Let $x := \left(x_1^{(1)}, \ldots, x_d^{(d)}\right)$. The other condition is that the remaining $n - d$ observations each need to fall outside O_x^+, guaranteeing the condition $x \in S$ required by Theorem 4.3(i).

Therefore,

$$\mathbb{E}\,\iota_{d,n} = n^{\underline{d}} \int_{\mathbf{y}:(\text{C.3}) \text{ holds}} \left[1 - \prod(1 - x_j^{(j)})\right]^{n-d} \,d\mathbf{y},$$

a d^2-dimensional integral which reduces effortlessly to a d-dimensional integral:

$$\mathbb{E}\,\iota_{d,n} = n^{\underline{d}} \int_{[0,1)^d} \left[\prod(1 - x_j)\right]^{d-1} \left[1 - \prod(1 - x_j)\right]^{n-d} \,dx$$

$$= n^{\underline{d}} \int_{(0,1]^d} x_\times^{d-1}(1 - x_\times)^{n-d} \,dx,$$

as desired. ◀

Statistics of Parking Functions and Labeled Forests

Stephan Wagner ✉ 📷

Institute of Discrete Mathematics, TU Graz, Austria
Department of Mathematics, Uppsala University, Sweden

Mei Yin ✉ 📷

Department of Mathematics, University of Denver, CO, USA

—— Abstract ——

In this paper we obtain some new results on the enumeration of parking functions and labeled forests. We introduce new statistics both for parking functions and for labeled forests that are connected to each other by means of a bijection. We determine the joint distribution of two statistics on parking functions and their counterparts on labeled forests. Our results on labeled forests also serve to explain the mysterious equidistribution between two seemingly unrelated statistics in parking functions recently identified by Stanley and Yin and give an explicit bijection between the two statistics.

1 Introduction

In this paper we obtain some new results on the enumeration of parking functions and labeled forests. We introduce new statistics both for parking functions and for labeled forests and connect them by means of a bijection. We determine the joint distribution of two statistics on parking functions and their counterparts on labeled forests. Our enumerative results for parking functions and for labeled forests inform each other. In particular, our results on labeled forests serve to explain the mysterious equidistribution between two seemingly unrelated statistics in parking functions recently identified by Stanley and Yin [16] and give an explicit bijection between the two statistics.

We begin with the necessary definitions. In the parking function scenario due to Konheim and Weiss [9], there are n parking spaces on a one-way street, labeled $1, 2, \ldots, n$ in consecutive order. A line of $m \leq n$ cars enters the street, one by one. The ith car drives to its preferred spot π_i and parks there if possible; if the spot is already occupied then the car parks in the first available spot after that. The list of preferences $\pi = (\pi_1, \ldots, \pi_m)$ is called a *parking function* if all cars successfully park. (The parking function is called "classical" when $m = n$.) We denote the set of parking functions by $\mathrm{PF}(m, n)$, where m is the number of cars and n is the number of parking spots. Using the pigeonhole principle, we see that a parking function

$\pi \in \mathrm{PF}(m, n)$ must have at most one value $= n$, at most two values $\geq n - 1$, and for each k at most k values $\geq n - k + 1$, and any such function is a parking function. Equivalently, π is a parking function if and only if

$$\#\{k : \pi_k \leq i\} \geq m - n + i, \ \ \text{for } i = n - m + 1, \ldots, n. \tag{1}$$

We make two immediate observations from (1). The first observation is that parking functions are invariant under the action of the symmetric group $\mathfrak{S}_m$ permuting the m cars, that is, permuting the list of preferences π. The second observation is that when some π_i takes values in the set $\{1, 2, \ldots, n - m + 1\}$, changing π_i to any other value in the set $\{1, 2, \ldots, n - m + 1\}$ has no effect on π being a parking function.

One of the most fundamental results on parking functions is that the number of parking functions is $|\mathrm{PF}(m, n)| = (n - m + 1)(n + 1)^{m-1}$. A famous combinatorial proof in the classical case was given by Pollak (unpublished but recounted in [5] and [12]). See also Pitman and Stanley [15] for a generalization of Pollak's circle argument. The combinatorial argument boils down to the following easily verified statement: Let G denote the group of all m-tuples $(a_1, \ldots, a_m) \in [n + 1]^m$ with componentwise addition modulo $n + 1$. Let H be the subgroup of G generated by $(1, 1, \ldots, 1)$. Then every coset of H contains exactly $n - m + 1$ parking functions. Interpreted probabilistically, the combinatorial operation involves assigning m cars on a circle with $n + 1$ spots and recording those car assignments where spot $n + 1$ is left empty after circular rotation. Since there are $n - m + 1$ missing spaces for the assignment of any preference sequence, any preference sequence π has $n - m + 1$ rotations that are valid parking functions. Our parking function proofs will be based on refinements of Pollak's proof technique, where we investigate the individual parking statistics for each car the moment it is parked on the circle.

This new line of approach first appeared in a paper by Stanley and Yin [16] and is extended in this paper, where we introduce the new statistics *leading elements* and *size of level set* on parking functions $\pi \in \mathrm{PF}(m, n)$ and examine their joint distributions. The statistic leading elements was introduced in [16] earlier for classical parking functions $\pi \in \mathrm{PF}(n, n)$ and counts the total number of cars whose desired spot is the same as that of the first car. It was shown in [16, Theorem 4.2] via a generating function approach that for classical parking functions the leading elements statistic is equidistributed with the widely-studied 1's statistic that counts the total number of cars whose desired spot is spot 1. This feature of parking functions is quite mysterious as these two parking function statistics seem unrelated and are not of the same nature. While the leading elements statistic is invariant under circular rotation, it does not satisfy permutation symmetry as permuting the entries might change the first element. On the other hand, though the 1's statistic is invariant under permuting all the entries, it does not exhibit circular rotation invariance. Indeed, only one out of $n + 1$ rotations of an assignment of n cars on a circle with $n + 1$ spots gives a valid parking function. It is thus intriguing what is hidden behind the pair of statistics (leading elements, 1's). By casting parking functions in the context of labeled forests, this question will be answered in our paper. We explain the gist of our argument below.

Let $[n] = \{1, 2, \ldots, n\}$ and $[n]_0 = \{0, 1, \ldots, n\}$. Let $\mathcal{T}(n)$ denote the set of all *rooted trees* T on the vertex set $[n]_0$ with root 0. More generally, let $\mathcal{F}(m, n)$ denote the set of all *rooted forests* F with $n + 1$ vertices and m edges (equivalently, $n - m + 1$ distinct tree components) such that a specified set of $n - m + 1$ vertices are the roots of the different trees. We label the roots of F by $\{01, 02, \ldots, 0(n - m + 1)\}$ and the non-root vertices by $\{1, 2, \ldots, m\}$. The fact that the cardinalities of classical parking functions and of rooted trees are the same, i.e.,

$$|\mathrm{PF}(n, n)| = |\mathcal{T}(n)|,$$

and more generally

$$|\operatorname{PF}(m,n)| = |\mathcal{F}(m,n)|,$$

has motivated much work in the study of connections between the two combinatorial structures. One bijective construction between parking functions and labeled forests goes back to Foata and Riordan [5]. Their construction is for the special case $m = n$ and is referred to as a breadth first search (with a queue) on rooted trees. See Yan [17, Section 1.2.3] and also Chassaing and Marckert [1]. We will show that under the bijective correspondence induced by breadth first search, the seemingly unrelated leading elements statistic and the 1's statistic for classical parking functions both become degree statistics. One of them (the root degree) is classical, while the other (degree of the parent of a fixed vertex) appears to be new.

Generally, statistics based on degrees and other aspects of labeled trees have been studied extensively. A well-known generalization of Cayley's tree theorem includes the degrees of all vertices as additional statistics [7, 14]. Another interesting example is the enumeration of labeled trees with respect to their indegrees: there are two versions of defining indegrees, both leading to the same enumeration formula. In the global orientation (see e.g. [14]), all edges are oriented towards the root; in the local orientation (see [13]), they are oriented towards the higher label. For many more interesting statistics of labeled trees, see for instance [3] (descents), [11] (inversions, which were also connected to parking functions in [6]), or [10] (runs).

This paper is organized as follows. In Section 2 we extend the statistic leading elements (denoted $\operatorname{lel}(\pi)$) for classical parking functions that was studied by Stanley and Yin [16] to general parking functions $\pi \in \operatorname{PF}(m,n)$. We also introduce a new statistic, size of level set (denoted $\operatorname{slev}(\pi)$), for parking functions $\pi \in \operatorname{PF}(m,n)$ that extends the notion of the 1's statistic (denoted $\operatorname{ones}(\pi)$) for classical parking functions and study the joint distribution of the level set statistic and the leading elements statistic. We establish the generating function for the pair of statistics $(\operatorname{slev}(\pi), \operatorname{lel}(\pi))$ using variations of Pollak's argument.

In Section 3, we apply the aforementioned bijection between parking functions and labeled forests that is based on breadth first search. By means of this bijection, we find that the pair of statistics $(\operatorname{slev}(\pi), \operatorname{lel}(\pi))$ on the set of parking functions $\operatorname{PF}(m,n)$ is equidistributed with the pair of statistics $(\deg(0), \deg(p))$ on the set of rooted forests $\mathcal{F}(m,n)$, where $\deg(0)$ is the root degree of a rooted forest (the total number of children of all roots $01, 02, \ldots, 0(n-m+1)$), and $\deg(p)$ is the number of children of the parent p of the vertex labeled 1 (by degree, we generally mean more precisely the number of children of a vertex in a rooted tree, which is 1 less than the degree in the graph-theoretical sense for non-root vertices).

The pair of statistics $(\deg(0), \deg(p))$ is further considered in Section 4. In particular, we directly prove a formula for the number of rooted forests in $\mathcal{F}(m,n)$ for which $\deg(0)$ and $\deg(p)$ take on given values by means of a combinatorial argument. In the special case $m = n$ (i.e., for labeled trees), the two statistics $\deg(0)$ and $\deg(p)$ also have the same distribution. We provide an explicit bijection for this fact.

In Section 5, we finally examine the asymptotic properties of the statistics investigated in our paper using standard probabilistic tools.

2 Statistics on parking functions

In this section we investigate the joint distribution of the pair of statistics $(\operatorname{slev}(\pi), \operatorname{lel}(\pi))$ on parking functions $\pi \in \operatorname{PF}(m,n)$. The precise definitions of the individual statistics read as follows:

▬ Leading elements: total number of cars whose desired spot is the same as that of the first car, denoted $\mathrm{lel}(\pi)$. This statistic was recently introduced (in the special case $m = n$) by Stanley and Yin [16].

▬ Size of level set: total number of cars whose desired spot is in the range $\{1, 2, \ldots, n-m+1\}$, denoted $\mathrm{slev}(\pi)$. When $m = n$, the level set statistic reduces to the 1's statistic for classical parking functions, which counts the total number of cars whose desired spot is spot 1, often denoted $\mathrm{ones}(\pi)$. The level set statistic $\mathrm{slev}(\pi)$ has not been considered before, but the 1's statistic $\mathrm{ones}(\pi)$ has been widely studied.

Our results for $\mathrm{PF}(m, n)$ are extensions of corresponding results for classical parking functions $\mathrm{PF}(n, n)$ in [16]. As mentioned in the introduction, we will expand upon Pollak's ingenious circle argument [5] for the street parking model to derive our results.

The following lemma was proven before using other means, see for example Kenyon and Yin [8, Corollary 3.4]. Our direct combinatorial argument below will shed light on the structure of parking functions and will be useful in the proof of Theorem 4. As implied by the necessary and sufficient condition (1), changing $\pi_1 = 1$ to whichever value below $n - m + 1$ will still keep π a parking function. The number of parking functions $\pi \in \mathrm{PF}(m, n)$ with $\pi_1 \in \{1, 2, \ldots, n - m + 1\}$ is thus $n - m + 1$ times the number of parking functions $\pi \in \mathrm{PF}(m, n)$ with $\pi_1 = 1$.

▶ **Lemma 1.** *We have*

$$\#\{\pi \in \mathrm{PF}(m, n) : \pi_1 = 1\} = (n - m + 2)(n + 1)^{m-2},$$

which implies that

$$\#\{\pi \in \mathrm{PF}(m, n) : \pi_1 \in \{1, 2, \ldots, n - m + 1\}\} = (n - m + 1)(n - m + 2)(n + 1)^{m-2}.$$

Proof. The statement is trivially true for $m = 1$. For $m \geq 2$, we assign cars $2, \ldots, m$ independently on a circle of length $n + 1$. Taking circular rotation into consideration, the car assignments give rise to $(n - m + 2)(n + 1)^{m-2}$ valid parking functions. Note that car 1 will always be able to park if its desired spot is spot 1. Our conclusion readily follows. ◀

The following lemma allows us to split a parking function in $\mathrm{PF}(m, n)$ into an arbitrary map whose range is precisely the set $\{1, 2, \ldots, n - m + 1\}$ (that is relevant for the statistic slev) and a parking function on a smaller domain. It will be very useful in proving our results on the distribution of statistics on $\mathrm{PF}(m, n)$.

▶ **Lemma 2.** *Consider a function $\pi : [m] \to [n]$. Fix the elements of π that are equal to one of $1, 2, \ldots, n-m+1$, and suppose that there are $s \geq 0$ such elements (this is precisely $\mathrm{slev}(\pi)$). Let the other elements be $\pi_{j_1}, \pi_{j_2}, \ldots, \pi_{j_{m-s}}$, and define a new function $\tilde{\pi} : [m - s] \to [m - 1]$ by $\tilde{\pi}_i = \pi_{j_i} - (n - m + 1)$. Then π is a parking function in $\mathrm{PF}(m, n)$ if and only if $\tilde{\pi}$ is a parking function in $\mathrm{PF}(m - s, m - 1)$.*

▶ Remark 3. For $s = 0$, there is no valid parking function in view of (1). This is consistent with the fact that $\mathrm{PF}(m, m - 1)$ is (trivially) empty.

Proof. We make use of the characterization (1) of parking functions. For any $i > m - n$, we have

$$\#\{k : \pi_k \leq i\} = s + \#\{k : \tilde{\pi}_k \leq i - (n - m + 1)\}.$$

So the condition in (1) is equivalent to

$$\#\{k : \tilde{\pi}_k \leq i - (n - m + 1)\} \geq m - n + i - s$$

for $i = m - n + 1, m - n + 2, \ldots, n$. Substituting $h = i - (n - m + 1)$, this becomes

$$\#\{k : \tilde{\pi}_k \leq h\} \geq h + 1 - s \tag{2}$$

for $h = 0, 1, \ldots, m-1$. This is precisely the condition for a parking function in $\mathrm{PF}(m-s, m-1)$, except for one detail: the conditions start at $h = 0$ rather than $h = s$. However, for $h < s$, (2) is trivially satisfied. This completes the proof. ◀

Lemma 2 means that every parking function in $\mathrm{PF}(m, n)$ can be uniquely decomposed into an arbitrary function $\pi_a : A \to [n - m + 1]$ on a set $A \subseteq [m]$ of cardinality s and a function that is equivalent to a parking function π_p in $\mathrm{PF}(m - s, m - 1)$.

As a consequence of the decomposition in Lemma 2, we will now be able to prove results on the distributions of the statistics $\mathrm{slev}(\pi)$ and $\mathrm{lel}(\pi)$. We start with the joint distribution of $\mathrm{slev}(\pi)$ and $\mathrm{lel}(\pi)$, which is determined by the following theorem.

▶ **Theorem 4.** *Let $s, t \geq 1$. We have*

$$\#\{\pi \in \mathrm{PF}(m, n) \colon \mathrm{slev}(\pi) = s \text{ and } \mathrm{lel}(\pi) = t\}$$
$$= \binom{m - 2}{s - 1, t - 1, m - s - t} (n - m + 1)^s (m - 1)^{m-s-t+1}$$
$$+ \binom{m - 1}{t - 1, s - t, m - s} s(n - m + 1)(n - m)^{s-t} m^{m-s-1}.$$

We will revisit this formula later in the context of rooted forests. The generating function for the joint distribution of $\mathrm{slev}(\pi)$ and $\mathrm{lel}(\pi)$ is obtained in a straightforward fashion by summing over all s and t.

▶ **Corollary 5.**

$$\sum_{\pi \in \mathrm{PF}(m,n)} x^{\mathrm{slev}(\pi)} y^{\mathrm{lel}(\pi)} = (n - m + 1)xy \left[(m - 1)((n - m + 1)x + y + m - 1)^{m-2} \right.$$
$$\left. + (xy + (n - m)x + 1)(xy + (n - m)x + m)^{m-2} \right]. \tag{3}$$

Proof of Theorem 4. Let us count parking functions $\pi \in \mathrm{PF}(m, n)$ for which $\mathrm{slev}(\pi) = s$. Lemma 2 shows that we can decompose π into an arbitrary function π_a from A to $[n - m + 1]$, where $|A| = s$ and $A \subseteq [m]$, and a (function equivalent to a) parking function π_p in $\mathrm{PF}(m - s, m - 1)$. For $\mathrm{lel}(\pi)$, we distinguish two cases:

▪ The spot of the first car does not lie in $\{1, 2, \ldots, n - m + 1\}$. In this case, $1 \notin A$, and the value of $\mathrm{lel}(\pi)$ is determined by the function π_p. Recall that by Pollak's argument, for every possible map from $[m - s]$ to $[m]$, there are s possible rotations that will turn it into a parking function in $\mathrm{PF}(m - s, m - 1)$. Thus in a randomly chosen parking function, each car (other than the first) takes the same spot as car 1 with the same probability $\frac{1}{m}$, and all the cars are independent. So $\mathrm{lel}(\pi)$ follows a binomial distribution in this case, and there are $s\binom{m-s-1}{t-1}(m - 1)^{m-s-t}$ possibilities for π_p such that $\mathrm{lel}(\pi) = t$. Since there are $\binom{m-1}{s}$ choices for the set A (the domain of π_a) and $(n - m + 1)^s$ choices for the function π_a itself, we obtain a total contribution of

$$\binom{m - 1}{s} \cdot (n - m + 1)^s \cdot s \binom{m - s - 1}{t - 1} (m - 1)^{m-s-t}$$
$$= \binom{m - 2}{s - 1, t - 1, m - s - t} (n - m + 1)^s (m - 1)^{m-s-t+1}.$$

The corresponding generating function is

$$\sum_{s=1}^{m} \binom{m-1}{s}(n-m+1)^s ys(m-1+y)^{m-s-1}x^s$$
$$= (n-m+1)xy(m-1)((n-m+1)x+y+m-1)^{m-2}.$$

- The spot of the first car lies in $\{1, 2, \ldots, n-m+1\}$. Then $1 \in A$, and $\mathrm{lel}(\pi)$ is completely determined by the function π_a. Each element of $A \setminus \{1\}$ has the same probability $\frac{1}{n-m+1}$ of being mapped to the same element as car 1 by π_a, and all these elements are independent. So given s, $\mathrm{lel}(\pi)$ follows a binomial distribution in this case as well, and given $\pi_a(1)$, there are $\binom{s-1}{t-1}(n-m)^{s-t}$ possibilities for the map π_a such that $\mathrm{lel}(\pi) = t$. There are now $\binom{m-1}{s-1}$ choices for the set A, $n-m+1$ choices for the spot of the first car, and sm^{m-s-1} possible choices for π_p, so this case yields a contribution of

$$\binom{s-1}{t-1}(n-m)^{s-t} \cdot \binom{m-1}{s-1}(n-m+1) \cdot sm^{m-s-1}$$
$$= \binom{m-1}{t-1, s-t, m-s}s(n-m+1)(n-m)^{s-t}m^{m-s-1}.$$

The generating function associated with this case is

$$\sum_{s=1}^{m} \binom{m-1}{s-1}y(n-m+1)(n-m+y)^{s-1}sm^{m-s-1}x^s$$
$$= (n-m+1)xy(xy+(n-m)x+m)^{m-2}(xy+(n-m)x+1).$$

Combining the two contributions, we obtain Theorem 4 and Corollary 5. ◀

Specializing the generating function by setting $x = 1$ or $y = 1$, we immediately obtain the distributions of $\mathrm{lel}(\pi)$ and $\mathrm{slev}(\pi)$. These are given in the following two corollaries.

▶ **Corollary 6.** *Taking $x = 1$ in (3), we have*

$$\sum_{\pi \in \mathrm{PF}(m,n)} y^{\mathrm{lel}(\pi)} = (n-m+1)y(y+n)^{m-1}.$$

▶ **Corollary 7.** *Taking $y = 1$ in (3), we have*

$$\sum_{\pi \in \mathrm{PF}(m,n)} x^{\mathrm{slev}(\pi)} = (n-m+1)x\left((n-m+1)x+m\right)^{m-1}.$$

Observe that the generating functions in Corollaries 6 and 7 are identical for $m = n$ (up to renaming the variable). In this case, $\mathrm{slev}(\pi)$ becomes the statistic $\mathrm{ones}(\pi)$ (number of 1's in the parking function).

3   Breadth first search

In this section we explore the implications of the breadth first search (BFS) algorithm connecting parking functions $\mathrm{PF}(m,n)$ and rooted forests $\mathcal{F}(m,n)$. This allows us to transfer results from parking functions to forests. Our construction will extend the corresponding construction between classical parking functions $\mathrm{PF}(n,n)$ and rooted trees $\mathcal{T}(n)$ by Foata and Riordan [5, Section 3]. That our construction is a bijection may be similarly argued as in [5],
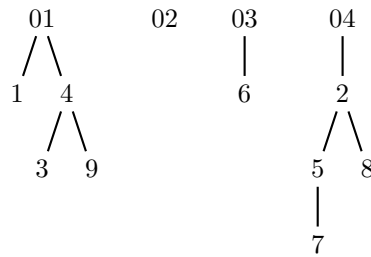
■ **Figure 1** Rooted spanning forest.

with minor adaptations. We will not go over all the technical details here, but will provide the explicit formulas for the generalized construction and illustrate the correspondence with a concrete example.

A forest $F \in \mathcal{F}(m, n)$ may be represented by an acyclic function f, where for a non-root vertex i, $f_i = j$ indicates that vertex j is the parent of vertex i in a tree component of the forest. Take $m = 9$ and $n = 12$. See Figure 1 representing an element of $F \in \mathcal{F}(9, 12)$, which corresponds to the acyclic function f given below:

$$
\begin{array}{ccccccccccc}
i & = & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\
f_i & = & 01 & 04 & 4 & 01 & 2 & 03 & 5 & 2 & 4
\end{array}
\tag{4}
$$

We read the vertices of the forest in breadth first search (BFS) order. That is, read root vertices in order first, then all vertices at level 1 (children of a root), then those at level 2 (distance 2 from a root), and so on, where vertices at a given level are naturally ordered in order of increasing predecessor, and, if they have the same predecessor, increasing order. See [17, Section 1.2.3] for a description of this graph searching algorithm in the language of computer science. Applying BFS to the forest F in Figure 1, we have

$$v_{01}, \ldots, v_{04}, v_5, \ldots, v_{13} = 01, 02, 03, 04, 1, 4, 6, 2, 3, 9, 5, 8, 7.$$

We let σ_f^{-1} be the vertex ordering once we remove the root vertices and σ_f be the *inverse order permutation* of σ_f^{-1}.

$$
\begin{array}{ccccccccccc}
i & = & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\
\sigma_f^{-1}(i) & = & 1 & 4 & 6 & 2 & 3 & 9 & 5 & 8 & 7 \\
\sigma_f(i) & = & 1 & 4 & 5 & 2 & 7 & 3 & 9 & 8 & 6
\end{array}
.
$$

We further let $t(f) = (r_1, \ldots, r_{12})$ with r_i recording the degree of v_i, starting with v_{01} and ending with v_{12} (ignoring the final vertex v_{13}), that is,

$$t(f) = (2, 0, 1, 1, 0, 2, 0, 2, 0, 0, 1, 0).$$

The sequence $t(f)$ is referred to as the *forest specification* of F.

Via the breadth first search, a generic forest $F \in \mathcal{F}(m, n)$ may thus be uniquely characterized by its associated specification $t(f)$ and order permutation σ_f. Furthermore, the pair $(t(f), \sigma_f)$ must satisfy certain *balance* and *compatibility* conditions. For exact definitions of these conditions, see [8, Section 2.2] and the references therein. Indeed, if we let $\mathcal{C}(m, n)$ be the set of all feasible pairs, then $\mathcal{C}(m, n)$ is in one-to-one correspondence with $\mathcal{F}(m, n)$. It turns out that $\mathcal{C}(m, n)$ is also in one-to-one correspondence with the set of parking functions $\mathrm{PF}(m, n)$, which we now describe.

For a parking function $\pi \in \mathrm{PF}(m,n)$, the associated *specification* is $s(\pi) = (r_1, \ldots, r_n)$, where $r_k = \#\{i : \pi_i = k\}$ records the number of cars whose parking preference is spot k. The *order permutation* $\tau_\pi \in \mathfrak{S}_m$, on the other hand, is defined by $\tau_\pi(i) = \#\{j : \pi_j < \pi_i, \text{ or } \pi_j = \pi_i \text{ and } j \leq i\}$, and so is the permutation that orders the list, without switching elements that are the same. In words, $\tau_\pi(i)$ is the position of the entry π_i in the non-decreasing rearrangement of π. For example, for $\pi = (3,1,3,1)$, $\tau_\pi(1) = 3$, $\tau_\pi(2) = 1$, $\tau_\pi(3) = 4$, and $\tau_\pi(4) = 2$. We can easily recover a parking function π by replacing i in τ_π with the ith smallest term in the sequence $1^{r_1} \ldots n^{r_n}$. As in the case of rooted forests, all feasible pairs $(s(\pi), \tau_\pi)$ for parking functions constitute the set $\mathcal{C}(m,n)$.

Combining the above perspectives, we see that the breadth first search algorithm bijectively connects parking functions and rooted forests, where $(t(f), \sigma_f) = (s(\pi), \tau_\pi)$. Continuing with our earlier example, for the forest $F \in \mathcal{F}(9,12)$ in Figure 1 with acyclic function representation f given by (4), we have

$$s(\pi) = (2, 0, 1, 1, 0, 2, 0, 2, 0, 0, 1, 0),$$

and

i	$=$	1	2	3	4	5	6	7	8	9
$\tau_\pi^{-1}(i)$	$=$	1	4	6	2	3	9	5	8	7
$\tau_\pi(i)$	$=$	1	4	5	2	7	3	9	8	6

We form the non-decreasing rearrangement sequence associated with $s(\pi)$:

$$1^2, 3^1, 4^1, 6^2, 8^2, 11^1 = 1, 1, 3, 4, 6, 6, 8, 8, 11.$$

Replacing i in τ_π with the ith smallest term in this sequence yields the corresponding parking function $\pi \in \mathrm{PF}(9,12)$ given below:

i	$=$	1	2	3	4	5	6	7	8	9
π_i	$=$	1	4	6	1	8	3	11	8	6

This bijective construction between parking functions and rooted forests has some interesting implications that are listed in the following theorem.

▶ **Theorem 8.** *The following statistics are equidistributed:*
- *The number of times π_i appears in a parking function $\pi \in \mathrm{PF}(m,n)$ equals the degree of the parent of vertex i in the corresponding forest $F \in \mathcal{F}(m,n)$.*
- *The number of times $1, 2, \ldots, n - m + 1$ appears in a parking function $\pi \in \mathrm{PF}(m,n)$ respectively equals the degree of the root vertex $01, 02, \ldots, 0(n-m+1)$ in the corresponding forest $F \in \mathcal{F}(m,n)$.*

Proof. This is due to our specific construction. From a forest F to a parking function π, we have

$$\pi_i = \begin{cases} j & \text{if } f_i = 0j \text{ for some } j = 1, 2, \ldots, n - m + 1, \\ (n - m + 1) + \sigma_f(f_i) & \text{otherwise.} \end{cases}$$

Conversely, from a parking function π to a forest F, we have

$$f_i = \begin{cases} 0j & \text{if } \pi_i = j \text{ for some } j = 1, 2, \ldots, n - m + 1, \\ \tau_\pi^{-1}(\pi_i - (n - m + 1)) & \text{otherwise.} \end{cases}$$

The second claim is clear. For the first claim, we note that $\pi_i = \pi_j$ corresponds to $f_i = f_j$, i.e., vertices i and j have the same parent.

In our illustrative example, the number of times each entry of π appears is given by the vector $\vec{w} = (2, 1, 2, 2, 2, 1, 1, 2, 2)$, whose entries coincide with the respective degree of the parent of vertex i for $i \in \{1, 2, \ldots, 9\}$. An immediate consequence of this fact is that

$$\sum_{i=1}^{9} w_i = \sum_{i=1}^{4} \deg^2(0i) + \sum_{i=5}^{13} \deg^2(i) = 15.$$

We also observe that the number of times $1, 2, 3, 4$ appears in π respectively agrees with the degrees of the roots $01, 02, 03, 04$ in the forest F, both yielding the vector $(2, 0, 1, 1)$. ◄

4 Statistics on trees and forests

In the bijection described in the previous section, the number of times π_1 occurs in the parking function (the statistic $\mathrm{lel}(\pi)$) corresponds to the degree $\deg(p)$ of the parent p of vertex 1 (see Theorem 8). Moreover, the total number of times $1, 2, \ldots, n - m + 1$ occur in the parking function (our statistic $\mathrm{slev}(\pi)$) corresponds to the total root degree $\deg(0)$, i.e., the sum of the degrees of all roots. Hence, the pair $(\deg(0), \deg(p))$ follows the same joint distribution as the pair $(\mathrm{slev}(\pi), \mathrm{lel}(\pi))$. The following generating function identity is therefore an automatic consequence of Corollary 5.

▶ **Theorem 9.**

$$\sum_{F \in \mathcal{F}(m,n)} x^{\deg(0)} y^{\deg(p)} = (n - m + 1)xy \left[(m - 1)((n - m + 1)x + y + m - 1)^{m-2} \right.$$
$$\left. + (xy + (n - m)x + 1)(xy + (n - m)x + m)^{m-2} \right]. \quad (5)$$

When $m = n$, the rooted spanning forest $F \in \mathcal{F}(m, n)$ reduces to a rooted tree $T \in \mathcal{T}(n)$. We recognize from Corollaries 6 and 7 that $\mathrm{ones}(\pi)$ and $\mathrm{lel}(\pi)$ are equidistributed. The breadth first search algorithm maps them to $\deg(0)$ and $\deg(p)$, respectively, so those are equidistributed as well. In words, the distribution of the number of children of the root 0 of a rooted labeled tree follows the same distribution as the number of children of the parent of vertex 1 (or indeed by symmetry the parent of any fixed vertex). The property of being parent of a specific vertex induces a bias towards higher degrees, which turns out to be equivalent to the bias induced by being the root (which necessarily has at least one child). The following procedure provides an explicit bijection for the equidistribution of $\deg(0)$ and $\deg(p)$.
1. Remove the edge connecting vertices 1 and p.
2. Connect vertices 0 and 1 by an edge.
3. Interchange vertices 0 and p.

See Figure 2 for the general procedure and Figure 3 for an example. This map has the extra benefit of being an involution. Moreover, the degrees of all vertices except 0 and p are preserved. Some nice features are hence introduced in the corresponding parking function bijection, where in our example

$$\pi = (8, 4, 5, 1, 2, 1, 1, 5, 6) \leftrightarrow \pi' = (5, 8, 2, 2, 1, 5, 5, 4, 9).$$

We see that $\mathrm{ones}(\pi)$ and $\mathrm{lel}(\pi)$ are switched, but the frequencies of the non-1 and non-leading elements are preserved up to permutation. In the example, the non-1 and non-leading elements in π are 5 (occurring twice), 2, 4, and 6. Those in π' are 2 (occurring twice), 4, 8, and 9.
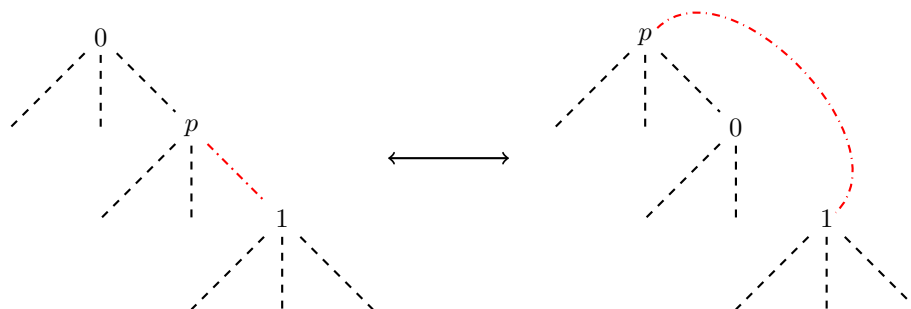
Figure 2 A bijective map between $\deg(0)$ and $\deg(p)$ (illustration).
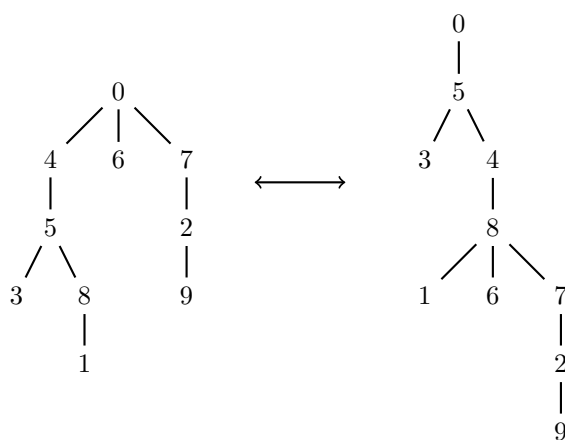
Figure 3 A bijective map between $\deg(0)$ and $\deg(p)$ (example).

The counting formula in our next proposition is equivalent to Theorem 4 in view of the bijection between parking functions and forests. In the following, we also illustrate it with a combinatorial proof of the statement in the setting of forests.

▶ **Proposition 10.** *Let $s, t \geq 1$. We have*

$$\#\{F \in \mathcal{F}(m,n)\colon \deg(0) = s \text{ and } \deg(p) = t\}$$
$$= \binom{m-2}{s-1, t-1, m-s-t}(n-m+1)^s(m-1)^{m-s-t+1}$$
$$+ \binom{m-1}{t-1, s-t, m-s}s(n-m+1)(n-m)^{s-t}m^{m-s-1}.$$

Proof. As a starting point, we recall the well-known result that the number of rooted forests with vertex set $[a]$ and b components whose root labels are given is ba^{a-b-1} (see [14]). Thus the number of such rooted forests *with one distinguished vertex* (possibly one of the roots) is ba^{a-b}, and by symmetry the number of such rooted forests where a vertex in the first component is distinguished must be a^{a-b}. Now in order to prove our formula, we distinguish two cases:
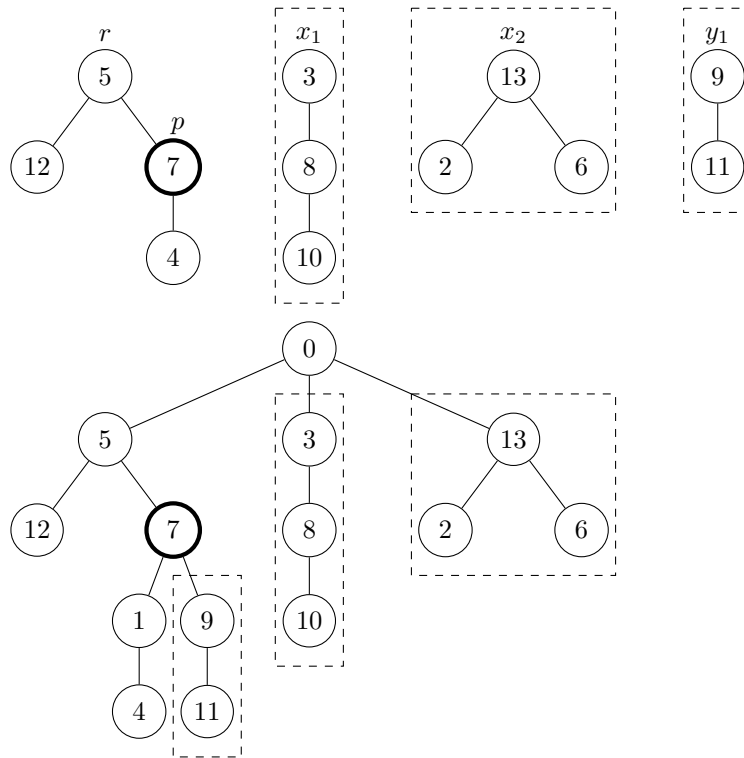
Case 1. The parent p of vertex 1 is not one of the roots.   A forest F with this property as well as $\deg(0) = s$ and $\deg(p) = t$ can be uniquely constructed as follows:

- Choose a label $r \in [m] \setminus \{1\}$ ($m - 1$ possibilities).
- Choose two disjoint sets of labels $\{x_1, x_2, \ldots, x_{s-1}\}$ and $\{y_1, y_2, \ldots, y_{t-1}\}$ from $[m] \setminus \{1, r\}$ (for which there are $\binom{m-2}{s-1, t-1, m-s-t}$ possibilities).
- Choose a rooted forest on $[m] \setminus \{1\}$ with root labels $r, x_1, x_2, \ldots, x_{s-1}, y_1, y_2, \ldots, y_{t-1}$ and a distinguished vertex p in the first component (there are $(m-1)^{m-s-t}$ possibilities, as explained above). Note that potentially $p = r$.
- Split the distinguished vertex p into two vertices, labeled p and 1 respectively, where p is the parent and 1 is the child, and all former children of p now become children of 1.
- Add roots $01, 02, \ldots, 0(n - m + 1)$, and connect each of the vertices $r, x_1, x_2, \ldots, x_{s-1}$ with one of these roots by an edge. There are $(n - m + 1)^s$ possibilities for this.
- Add edges between vertex p and $y_1, y_2, \ldots, y_{t-1}$. Note that p is indeed the parent of 1 in this construction, and that $\deg(0) = s$ as well as $\deg(p) = t$.

It is easy to reverse the procedure given a forest with $\deg(0) = s$ and $\deg(p) = t$ for which p is not one of the roots. So to summarize, we have

$$\binom{m - 2}{s - 1, t - 1, m - s - t}(n - m + 1)^s (m - 1)^{m-s-t+1}$$

possible forests in this case, which accounts for the first term in our formula. Case 1 is illustrated in the special case $n = m = 13$, $s = 3$ and $t = 2$ in Figure 4. The choice of root labels is $r = 5$, $\{x_1, x_2\} = \{3, 13\}$ and $\{y_1\} = \{9\}$.

Figure 4 Illustration of the procedure: the rooted forest (top) with the distinguished vertex p indicated by a thick node, and the final tree (bottom).

Case 2. The parent p of vertex 1 is one of the roots. Again, there is a unique way to construct all these forests:

- Select a set of labels $\{x_1, x_2, \ldots, x_{s-1}\}$ from $[m] \setminus \{1\}$ in $\binom{m-1}{s-1}$ ways.
- Construct a rooted forest with vertex set $[m]$ and root labels $1, x_1, x_2, \ldots, x_{s-1}$ (there are sm^{m-s-1} possibilities).
- Among the labels $x_1, x_2, \ldots, x_{s-1}$, choose the siblings of vertex 1; there are $\binom{s-1}{t-1}$ possible choices.
- Pick one of the $n - m + 1$ roots $01, 02, \ldots, 0(n - m + 1)$ as the parent p of vertex 1, and connect it and all the siblings chosen in the previous step to it by an edge.
- Lastly, pick one of the other $n - m$ roots as parent for each of the remaining vertices with label in the set $\{x_1, x_2, \ldots, x_{s-1}\}$. This step yields $(n - m)^{s-t}$ possibilities.

Putting these together gives us the second term and thus completes the proof.   ◄

5   Limit distributions

In this section we conclude the paper with some asymptotic observations on the statistics investigated in our paper. We consider the scenario where m is a linear function of n, and determine the limit distributions of the statistics lel and slev (for parking functions) and $\deg(0)$ as well as $\deg(p)$ (for rooted forests), respectively.

▶ **Proposition 11.** *Let s be a fixed positive integer, and take $m = cn$ for some $0 < c \leq 1$ as $n \to \infty$. Consider the parking preference $\pi \in \mathrm{PF}(m, n)$ chosen uniformly at random, and let $\mathrm{lel}_s(\pi)$ be the number of cars with the same preference as car s. Then $\mathrm{lel}_s(\pi) - 1 \xrightarrow{\mathrm{d}} \mathrm{Poisson}(c)$, i.e., for every fixed nonnegative integer j,*

$$\mathbb{P}\left(\mathrm{lel}_s(\pi) = 1 + j \mid \pi \in \mathrm{PF}(m, n)\right) \sim \frac{c^j e^{-c}}{j!}.$$

Proof. By permutation symmetry, we only need to prove this result for $s = 1$, where $\mathrm{lel}_s(\pi)$ is exactly $\mathrm{lel}(\pi)$. We divide both sides of the generating function for $\mathrm{lel}(\pi)$ in Corollary 6 through by $(n - m + 1)(n + 1)^{m-1}$. The right side becomes the probability generating function of $S(m, n) = 1 + \sum_{i=1}^{m-1} X_i$, where the X_i are independent Bernoulli random variables:

$$X_i = \begin{cases} 0, & \text{probability } n/(n + 1), \\ 1, & \text{probability } 1/(n + 1). \end{cases}$$

Hence we have a standard case of the law of rare events, leading to a Poisson limit distribution.
◄

▶ **Corollary 12.** *Let s be a fixed positive integer, and take $m = cn$ for some $0 < c \leq 1$ as $n \to \infty$. Consider the labeled forest $F \in \mathcal{F}(m, n)$ chosen uniformly at random, and let $\deg(p_s)$ be the degree of the parent p_s of vertex s. Then $\deg(p_s) - 1 \xrightarrow{\mathrm{d}} \mathrm{Poisson}(c)$.*

Proof. By permutation symmetry, we only need to prove this result for $s = 1$, where $\deg(p_s)$ is exactly $\deg(p)$. Since $\mathrm{lel}(\pi)$ and $\deg(p)$ are equidistributed (by Theorem 8), the statement follows from Proposition 11.   ◄

▶ **Proposition 13.** *Take $m = cn$ for some $0 < c < 1$ as $n \to \infty$. Consider the parking preference $\pi \in \mathrm{PF}(m, n)$ chosen uniformly at random. Then we have*

$$\frac{\mathrm{slev}(\pi) - c(1 - c)n}{\sqrt{c^2(1 - c)n}} \xrightarrow{\mathrm{d}} \mathcal{N}(0, 1).$$

Proof. We proceed as in the proof of Proposition 11 and divide both sides of the generating function for slev(π) in Corollary 7 through by $(n-m+1)(n+1)^{m-1}$. The right side becomes the probability generating function of $S(m,n) = 1 + \sum_{i=1}^{m-1} X_i$, where the X_i are independent Bernoulli random variables:

$$X_i = \begin{cases} 0, & \text{probability } m/(n+1), \\ 1, & \text{probability } (n-m+1)/(n+1). \end{cases}$$

The probabilities converge to c and $1-c$ respectively, and the standard central limit theorem applies. This means that slev(π) may be approximated by $\mathcal{N}(0,1)$ after standardization. ◄

▶ **Corollary 14.** *Take $m = cn$ for some $0 < c < 1$ as $n \to \infty$. Consider the labeled forest $F \in \mathcal{F}(m,n)$ chosen uniformly at random. Then we have*

$$\frac{\deg(0) - c(1-c)n}{\sqrt{c^2(1-c)n}} \xrightarrow{\text{d}} \mathcal{N}(0,1).$$

Proof. Since slev(π) and $\deg(0)$ are equidistributed (again by Theorem 8), this is the same proof as for Proposition 13. ◄

▶ **Remark 15.** The asymptotic analysis of the special case ones(π) for $\pi \in \mathrm{PF}(n,n)$ was conducted by Diaconis and Hicks [2]. The limit distribution of the root degree in labeled trees is classical [4, Example IX.6]. Both ones(π) -1 and $\deg(0) - 1$ can be approximated by Poisson(1).

References

1   Philippe Chassaing and Jean-François Marckert. Parking functions, empirical processes, and the width of rooted labeled trees. *Electron. J. Combin.*, 8(1):Research Paper 14, 19, 2001. `doi:10.37236/1558`.

2   Persi Diaconis and Angela Hicks. Probabilizing parking functions. *Adv. in Appl. Math.*, 89:125–155, 2017. `doi:10.1016/j.aam.2017.05.004`.

3   Ömer Eğecioğlu and Jeffrey B. Remmel. Bijections for Cayley trees, spanning trees, and their q-analogues. *J. Combin. Theory Ser. A*, 42(1):15–30, 1986. `doi:10.1016/0097-3165(86)90004-X`.

4   Philippe Flajolet and Robert Sedgewick. *Analytic Combinatorics.* Cambridge University Press, 2009.

5   Dominique Foata and John Riordan. Mappings of acyclic and parking functions. *Aequationes Math.*, 10:10–22, 1974. `doi:10.1007/BF01834776`.

6   António Guedes de Oliveira and Michel Las Vergnas. Parking functions and labeled trees. *Sém. Lothar. Combin.*, 65:Art. B65e, 10, 2010/12.

7   Frank Harary and Edgar M. Palmer. *Graphical enumeration.* Academic Press, New York-London, 1973.

8   Richard Kenyon and Mei Yin. Parking functions: from combinatorics to probability. *Methodol. Comput. Appl. Probab.*, 25(1):Paper No. 32, 30, 2023. `doi:10.1007/s11009-023-10022-5`.

9   Alan G. Konheim and Benjamin Weiss. An occupancy discipline and applications. *SIAM J. Appl. Math.*, 14:1266–1274, 1966.

10 Marie-Louise Lackner and Alois Panholzer. Runs in labelled trees and mappings. *Discrete Math.*, 343(9):111990, 16, 2020. `doi:10.1016/j.disc.2020.111990`.

11 C. L. Mallows and John Riordan. The inversion enumerator for labeled trees. *Bull. Amer. Math. Soc.*, 74:92–94, 1968. `doi:10.1090/S0002-9904-1968-11888-9`.

12 John Riordan. Ballots and trees. *J. Combinatorial Theory*, 6:408–411, 1969.

13   Heesung Shin and Jiang Zeng. A bijective enumeration of labeled trees with given indegree sequence. *J. Combin. Theory Ser. A*, 118(1):115–128, 2011. `doi:10.1016/j.jcta.2010.07.001`.

14   Richard P. Stanley. *Enumerative combinatorics. Vol. 2*, volume 62 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1999. `doi:10.1017/CBO9780511609589`.

15   Richard P. Stanley and Jim Pitman. A polytope related to empirical distributions, plane trees, parking functions, and the associahedron. *Discrete Comput. Geom.*, 27(4):603–634, 2002. `doi:10.1007/s00454-002-2776-6`.

16   Richard P. Stanley and Mei Yin. Some enumerative properties of parking functions, 2023. arXiv: 2306.08681.

17   Catherine H. Yan. Parking functions. In *Handbook of enumerative combinatorics*, Discrete Math. Appl. (Boca Raton), pages 835–893. CRC Press, Boca Raton, FL, 2015.

Depth-First Search Performance in Random Digraphs

Philippe Jacquet ✉ ⌂ 🆔
Inria Saclay Ile de France, 1 Rue Honoré d'Estienne d'Orves, 91120 Palaiseau, France

Svante Janson ✉ ⌂ 🆔
Department of Mathematics, Uppsala University, Sweden

──── **Abstract** ────

We present an analysis of the depth-first search algorithm in a random digraph model with independent outdegrees having an arbitrary distribution with finite variance. The results include asymptotics for the distribution of the stack index and depths of the search. The search yields a series of trees of finite size before and after the exploration of a giant tree. Our analysis mainly concerns the giant tree. Most results are first order. This analysis proposed by Donald Knuth in his next to appear volume of *The Art of Computer Programming* gives interesting insight in one of the most elegant and efficient algorithm for graph analysis due to Tarjan.

1 Introduction

This paper is a continuation of our earlier paper [6], which studies a special case using a simpler method. The motivation of both paper is a new section in Donald Knuth's *The Art of Computer Programming* [7], which is dedicated to Depth-First Search (DFS) in a digraph. The DFS is an important computing tool dedicated to the exploration of large unstructured dataset, mostly organised as large directed graphs, and the DFS is the fundation of the daily crawling process of the graph of the Web performed by Google [9]. Briefly, the DFS starts with an arbitrary vertex, and explores the arcs from that vertex one by one. When an arc is found leading to a vertex that has not been seen before, the DFS explores the arcs from it in the same way, in a recursive fashion, before returning to the next arc from its parent. This eventually yields a tree containing all descendants of the the first vertex (which is the root of the tree). If there still are some unseen vertices, the DFS starts again with one of them and finds a new tree, and so on until all vertices are found. We refer to [7] for details as well as for historical notes. (See also S1–S2 in Section 2.) Note that the digraphs in [7] and here are multi-digraphs, where loops and multiple arcs are allowed. (Although in our random model they are few and usually not important.) The DFS algorithm generates a spanning forest (the *depth-first forest*) in the digraph, with all arcs in the forest directed away from the roots. Our main purpose is to study the properties of the depth-first forest, starting with a random digraph G; in particular we study the distribution of the stack index and the depth of vertices in the depth-first forest.

The random digraph model that we consider has n vertices and a given outdegree distribution $\mathbf{P}$. The outdegrees (number of outgoing arcs) of the n vertices are independent random numbers with this distribution. The endpoint of each arc is uniformly selected at random among the n vertices, independently of all other arcs. (Therefore, an arc can loop

back to the starting vertex, and multiple arcs can occur.) We consider asymptotics as $n \to \infty$ for a fixed outdegree distribution. In [6], we studied the case when the outdegree distribution is geometric. In the present paper, we generalise this and let $\mathbf{P}$ be an arbitrary distribution; however, we assume throughout the paper that the outdegree distribution $\mathbf{P}$ has a finite second moment.

▶ Remark 1. Related results are proved by Enriquez, Faraud and Ménard [3] for DFS in an undirected Erdős–Rényi graph $G(n, \lambda/n)$; the case when $\lambda = 1 + \varepsilon$ for small $\varepsilon > 0$ is studied further by another argument by Diskin and Krivelevich [2]. DFS in the random digraph $D(n, p)$ has also been considered previously, for example in the proof of [8, Theorem 3]. Although this is for a different random graph model, DFS on $G(n, \lambda/n)$ is the same as DFS on the Erdős–Rényi digraph $D(n, \lambda/n)$, which is essentially the same as the digraph studied in the present paper with outdegree distribution $\mathrm{Po}(\lambda)$. Hence, the main result of [3], which shows convergence of the depth profile in the depth-first forest to a certain deterministic limit, is essentially the special case $\mathbf{P} = \mathrm{Po}(\lambda)$ of our result for the depths (Theorem 6). The proofs are quite different. See also Enriquez, Faraud, Ménard and Noiry [4], where related results are given for the (undirected) configuration model.

We analyze the process $d(t)$ of depths of the vertices, in the order they are found by the DFS. For the geometric outdegree distribution studied in [6], $d(t)$ is a Markov chain, which was used in our proofs. For general outdegree distributions, this is no longer true. We show in Section 2 that we can use the size $I(t)$ of the stack of arcs kept by the DFS as a substitute; this is a Markov chain, and we obtain limit results for the stack size with deviation in $O_{L^2}(n^{1/2})$. (See Section 1.1 below for notation.) In a second step (Section 3), this is used to derive limit results for the depths $d(t)$, but the results obtained are within deviation in $o_{\mathrm{p}}(n)$ which is close to the order of the result. We give also an alternative approach in Section 4 where the depths $d(t)$ are analysed directly by a different method; the results there are preliminary and less complete, but it seems that this method yields sharper results that the method in Section 3, with deviation within $O_{\mathrm{p}}(n^{\beta})$ with β arbitrary close to $4/5$; thus the worst case error is now negligible compared to the main order.

Many details and further results will be given in the forthcoming full paper.

1.1 Some notation

We denote the given outdegree distribution by $\mathbf{P}$. We let η denote random variables with this distribution. In particular, we denote the outdegree of vertex v by $\eta(v)$. Recall that our standing assumption is that these outdegrees are i.i.d. (independent and identically distributed) with $\eta(v) \sim \mathbf{P}$. We let v_t denote the t-th vertex found by the DFS, and simplify notation by letting $\eta_t := \eta(v_t)$ be its outdegree. It follows from the construction of the DFS that also the random variables η_t, $t = 1, \ldots, n$ are i.i.d. with distribution $\mathbf{P}$; this fundamental property will be used repeatedly without further mention.

We assume throughout that the second moment $\mathbb{E}\,\eta^2 < \infty$. This is essential for some results (e.g. results on asymptotic normality), but we conjecture that many results hold assuming only that the first moment $\mathbb{E}\,\eta < \infty$.

The mean outdegree, *i.e.*, the expectation $\mathbb{E}\,\eta$ of $\mathbf{P}$, is denoted by λ. In analogy with branching processes, we say that the random digraph is *subcritical* if $\lambda < 1$, *critical* if $\lambda = 1$, and *supercritical* if $\lambda > 1$.

Let $d(v)$ be the depth of vertex v, and let $d(t) := d(v_t)$.

As usual, w.h.p. means *with high probability*, *i.e.*, with probability $1 - o(1)$ as $n \to \infty$.

For random variables X_n and positive numbers a_n, we write $X_n = o_{\mathrm{p}}(a_n)$ if, as $n \to \infty$, $X_n/a_n \xrightarrow{\mathrm{P}} 0$, *i.e.*, if for every $\varepsilon > 0$, we have $\mathbb{P}(|X_n| > \varepsilon a_n) \to 0$. Furthermore, $X_n = O_{\mathrm{p}}(a_n)$ means that the family $\{X_n/a_n\}$ is bounded in probability, *i.e.*, if for every $\varepsilon > 0$ there exists C such that $\mathbb{P}(|X_n| > C a_n) < \varepsilon$ for all n. $X_n = o_{L^2}(a_n)$ means $\mathbb{E}\big[|X_n/a_n|^2\big] \to 0$, and $X_n = O_{L^2}(a_n)$ means $\mathbb{E}\big[|X_n/a_n|^2\big] = O(1)$.

Ge(p) denotes the geometric distribution with parameter $p \in (0,1]$; thus $\xi \sim \mathrm{Ge}(p)$ means that ξ is a random variable with

$$\mathbb{P}(\xi = k) = (1-p)^k p, \qquad k = 0, 1, 2, \ldots \tag{1.1}$$

We write $x^+ := \max(x, 0)$.

2 Stack index analysis

As said above, unlike the special case $\mathbf{P} = \mathrm{Ge}(1-p)$ studied in [6], in general the depth $d(t)$ does not follow a Markov chain. Therefore, we instead first consider the stack index $I(t)$ defined below, which does evolve as a Markov chain; many arguments for $I(t)$ below are similar to arguments for $d(t)$ in [6, Section 2].

The DFS can be regarded as keeping a stack of unexplored arcs, for which we have seen the start vertex but not the endpoint. The stack evolves as follows:

S1. If the stack is empty, pick a new vertex v that has not been seen before (if there is no such vertex, we have finished). Otherwise, pop the last arc from the stack, and reveal its endpoint v (which is uniformly random over all vertices). If v already is seen, repeat.

S2. (v is now a new vertex) Reveal the outdegree $\eta(v)$ of v and add to the stack $\eta(v)$ new arcs from v, with unspecified endpoints. GOTO S1.

Let $I(t)$ be the size of the stack when v_t is found (but before we add the arcs from v_t). After v_t is found, and the η_t arcs from v_t have been added to the stack, the stack size is thus $I(t) + \eta_t$. We next perform step S1 one or several times. As long as the stack is not empty, we find each time an already seen vertex with probability t/n, and in this case we repeat S1. Hence, conditioned on $I(t)$ and η_t, for $k \geqslant 1$, the probability that S1 is performed exactly k times is

$$\left(\frac{t}{n}\right)^{k-1}\left(1 - \frac{t}{n}\right), \tag{2.1}$$

provided $1 \leqslant k \leqslant I(t) + \eta_t$, and if none of these events occur, then S1 is repeated a last time and a new vertex is picked that will be the root of a new tree in the depth-first forest (unless we have finished the DFS). Note that (2.1) equals the probability $\mathbb{P}(\xi = k)$ with $\xi_t \sim \mathrm{Ge}(1 - t/n)$, see (1.1). Thus, we can write

$$I(t+1) = \big(I(t) + \eta_t - 1 - \xi_t\big)^+, \qquad 1 \leqslant t < n, \tag{2.2}$$

where $\xi_t \sim \mathrm{Ge}(1 - t/n)$ is a random variable independent of the history; more precisely, the random variables η_t $(1 \leqslant t \leqslant n)$ and ξ_t $(1 \leqslant t < n)$ are all independent. We start the stochastic recursion (2.2) with $I(1) = 0$.

We define

$$\zeta_t := \eta_t - \xi_t - 1. \tag{2.3}$$

Thus, (2.2) can be written $I(t+1) = \big(I(t) + \zeta_t\big)^+$. We see also that, for $1 \leqslant t < n$,

$$v_{t+1} \text{ is a root} \iff I(t) + \zeta_t < 0 \iff I(t+1) > I(t) + \zeta_t. \tag{2.4}$$

Note that the random variables ζ_t are independent, but have different distributions. We define also

$$\widetilde{I}(t) := \sum_{i=1}^{t-1} (\eta_i - 1 - \xi_i) = \sum_{i=1}^{t-1} \zeta_i \tag{2.5}$$

and

$$I_*(t) := \min_{1 \leqslant j \leqslant t} \widetilde{I}(j). \tag{2.6}$$

It follows from the recursion (2.2) and induction that

$$I(t) = \widetilde{I}(t) - I_*(t). \tag{2.7}$$

By (2.5) and (2.7),

$$\widetilde{I}(t+1) = \widetilde{I}(t) + \zeta_t = I_*(t) + I(t) + \zeta_t \tag{2.8}$$

and thus it follows from (2.4) and (2.6) that, for $1 \leqslant t < n$,

$$v_{t+1} \text{ is a root} \iff I(t) + \zeta_t < 0 \iff \widetilde{I}(t+1) < I_*(t) \iff I_*(t+1) < I_*(t). \tag{2.9}$$

Obviously, v_1 is also a root, with $I(1) = \widetilde{I}(1) = I_*(1) = 0$.

We have

$$\mathbb{E}\,\zeta_t = \mathbb{E}\,\eta_t - \mathbb{E}\,\xi_t - 1 = \lambda - \frac{t/n}{1 - t/n} - 1 = \lambda - \frac{1}{1 - t/n}. \tag{2.10}$$

Hence, uniformly in $t/n \leqslant \theta^*$ for any fixed $\theta^* < 1$,

$$\mathbb{E}\,\widetilde{I}(t) = \sum_{i=1}^{t-1} \mathbb{E}\,\zeta_i = (t-1)\lambda - \sum_{i=1}^{t-1} \frac{1}{1 - i/n} = n\widetilde{\iota}(t/n) + O(1), \tag{2.11}$$

where

$$\widetilde{\iota}(\theta) := \int_0^\theta \left(\lambda - \frac{1}{1 - x} \right) dx = \lambda\theta + \log(1 - \theta). \tag{2.12}$$

Let, as in [6],

$$\theta_0 := \left(1 - \lambda^{-1} \right)^+ = \begin{cases} 1 - \lambda^{-1}, & \lambda > 1, \\ 0, & \lambda \leqslant 1, \end{cases} \tag{2.13}$$

and let θ_1 is the largest root in $[0, 1)$ of $\widetilde{\iota}(\theta_1) = 0$; thus

$$\log(1 - \theta_1) = -\lambda\theta_1 \tag{2.14}$$

and θ_1 equals the survival probability of a Bienayme–Galton–Watson (BGW) process with $\mathrm{Po}(\lambda)$ offspring distribution. Define

$$\widetilde{\iota}^+(\theta) := [\widetilde{\iota}(\theta)]^+ = \begin{cases} \lambda\theta + \log(1 - \theta), & 0 \leqslant \theta \leqslant \theta_1, \\ 0, & \theta_1 \leqslant \theta \leqslant 1, \end{cases} \tag{2.15}$$

It is easy to see that if $\lambda \leqslant 1$, then $\widetilde{\iota}^+(\theta) = 0$ for all $\theta \in [0, 1]$, while if $\lambda > 1$, then $\widetilde{\iota}^+(\theta) > 0$ for $0 < \theta < \theta_1$ (where $0 < \theta_1 < 1$), with a maximum at $\theta_0 = 1 - \lambda^{-1}$.

We now can argue similarly to [6], using $I(t)$ instead of $d(t)$. The proof of [6, Theorem 2.4] applies with very minor differences, and yields:

▶ **Theorem 2.** *Suppose that the outdegree distribution has finite variance. Then*

$$\max_{1 \leqslant t \leqslant n} \left| I(t) - n\widetilde{\iota}^+(t/n) \right| = O_{L^2}(n^{1/2}). \tag{2.16}$$

Furthermore, for every $\varepsilon > 0$, we also have

$$\max_{1 \leqslant t \leqslant (1-\varepsilon)n} \left| \widetilde{I}(t) - n\widetilde{\iota}(t/n) \right| = O_{L^2}(n^{1/2}). \tag{2.17}$$

We can use Theorem 2 to show the following results, extending [6, Theorems 4.1 and 4.3] to general outdegree distributions. The proofs are similar to the ones in [6]; details will be given in the full paper.
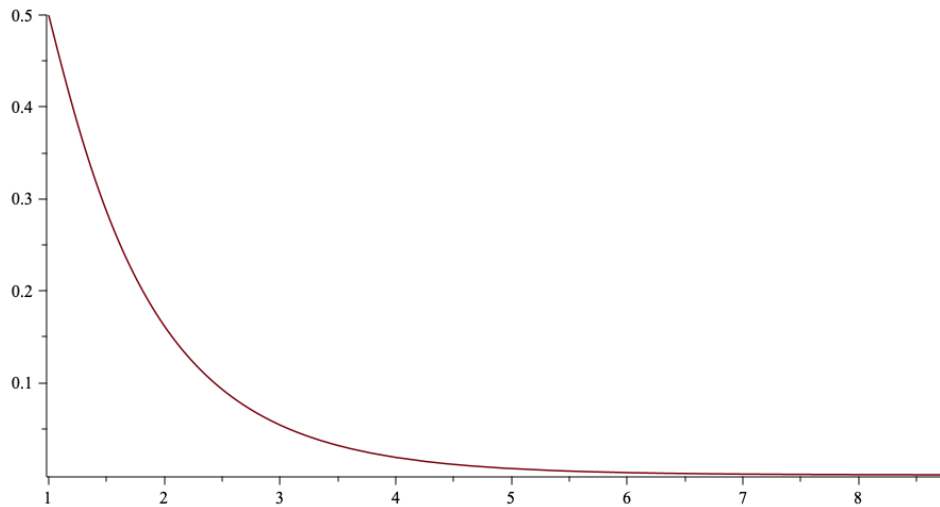
▶ **Theorem 3.** *Suppose that the outdegree distribution has finite variance. Let N be the number of trees in the depth-first forest. Then*

$$N = \psi n + O_{L^2}(n^{1/2}), \tag{2.18}$$

where

$$\psi := 1 - \theta_1 - \frac{\lambda}{2}(1 - \theta_1)^2. \tag{2.19}$$

Figure 1 shows the parameter ψ as a function of the average degree λ.

🟨 **Figure 1** ψ, as function of λ.

▶ **Theorem 4.** *Suppose that the outdegree distribution has finite variance. Let $\mathbf{T}_1$ be the largest tree in the depth-first forest.*
  (i) *If $\lambda \leqslant 1$, then $|\mathbf{T}_1| = o_{\mathrm{p}}(n)$.*
  (ii) *If $\lambda > 1$, then the largest tree in the depth-first forest has order $|\mathbf{T}_1| = \theta_1 n + O_{L^2}(n^{1/2})$. Furthermore, the second largest tree has order $|\mathbf{T}_2| = o_{\mathrm{p}}(n)$.*

More precise results will be given in the full paper.

3   Stack size and depth

We can recover the depth of the vertices from the stack size process, or more precisely, from the process $\widetilde{I}(t)$ defined above; this uses a method that has been used by several authors in the study of other random trees, see for example [5, (1.1)] and [1, Proposition 4].

When vertex v_t is found by the DFS, the stack consist of all future arcs from the ancestors of v_t in the depth-first forest. Hence, the stack size $I(t)$ is the total number of future arcs from the ancestors of v_t. As the DFS continues, it first explores the descendants of v_t. During this period, the ancestors of v_t are still ancestors of the current vertex, and their future arcs remain the same as when v_t was found. Hence, if v_s is a descendant of v_t, then $I(s) \geqslant I(t)$. Moreover, we then also have $I_*(s) = I_*(t)$ by (2.9), since no new root has been found, and thus, by (2.7), $\widetilde{I}(s) \geqslant \widetilde{I}(t)$. On the other hand, when the DFS has finished exploring decendants of v_t, then it has explored all arcs from v_t and the DFS next backtracks to the ancestors of v_t by exploring the future edges of the ancestors of v_t one by one. If another vertex v_u is found in this way, then $I(u) < I(t)$; furthermore, $I_*(u) = I_*(t)$ as above, and thus $\widetilde{I}(u) < \widetilde{I}(t)$. Finally, if no further new vertex is found in the same tree as v_t, then the DFS next finds a new root v_u; in this case, $I(u) = 0 \leqslant I(t)$, and, by (2.9), $I_*(u) < I_*(t)$. Hence, $\widetilde{I}(u) < \widetilde{I}(t)$ holds in this case too. This leads to the following characterisation. (Details will be given in the full paper.)

▶ **Lemma 5.**

(i) *For any $s, t \in [n]$, v_s is a descendant of v_t if and only if $s > t$ and*

$$\min_{t \leqslant i \leqslant s} \widetilde{I}(i) = \widetilde{I}(t). \tag{3.1}$$

(ii) *For any $t \geqslant 1$, the ancestors of v_t are the vertices v_s corresponding to the weak right-to-left minima of $(\widetilde{I}(s))_{1 \leqslant s \leqslant t}$, excluding $s = t$. Hence the depth $d(t)$ of v_t is the number of such right-to-left minima.*

Let $G_\eta(z) := \sum_k \mathbb{P}(\eta = k) z^k$ be the probability generating function of η. For $\theta \in [0, 1]$, let η_θ be a random variable obtained by thinning the outdegree η and keeping each arc with probability $1 - \theta$, independently; thus η_θ has probability generating function

$$G_{\eta_\theta}(z) = G_\eta\big(\theta + (1 - \theta)z\big) = G_\eta\big(1 - (1 - \theta)(1 - z)\big). \tag{3.2}$$

Note that the number of outgoing arcs from v_t to vertices that are not discovered earlier has the distribution of η_θ with $\theta = t/n$. Let $\rho(\theta)$ be the survivability probability that the Bienayme–Galton–Watson (BGW) process with offspring distribution η_θ has infinite size. It satisfies the identity:

$$1 - \rho(\theta) = G_{\eta_\theta}(1 - \rho(\theta)) = G_\eta\big(\theta + (1 - \theta)(1 - \rho(\theta))\big) = G_\eta\big(1 - (1 - \theta)\rho(\theta)\big). \tag{3.3}$$

We have $\mathbb{E}\,\eta_\theta = (1 - \theta)\lambda$. Hence, if $(1 - \theta)\lambda > 1$, then $\rho(\theta) > 0$; otherwise $\rho(\theta) = 0$.

Define

$$\widetilde{\ell}(\theta) := \int_0^\theta \rho(x)\,\mathrm{d}x, \qquad 0 \leqslant \theta \leqslant \theta_0, \tag{3.4}$$

and

$$\widetilde{\ell}^+(\theta) := \begin{cases} \widetilde{\ell}(\theta) = \int_0^\theta \rho(x)\,\mathrm{d}x, & 0 \leqslant \theta \leqslant \theta_0, \\ \widetilde{\ell}(\breve{\theta}), \text{ where } \breve{\theta} \in (0, \theta_0) \text{ and } \widetilde{\iota}(\breve{\theta}) = \widetilde{\iota}(\theta), & \theta_0 < \theta < \theta_1, \\ 0, & \theta_1 \leqslant \theta \leqslant 1. \end{cases} \tag{3.5}$$

If $\lambda \leqslant 1$, this simply means $\widetilde{\ell}(\theta) := 0$ for all $\theta \in [0, 1]$.

Using Lemma 5 and Theorem 2 we can obtain the following result; the detailed argument (given in the full paper) is unfortunately rather long.

▶ **Theorem 6.** *We have*

$$\max_{1 \leqslant t \leqslant n} \left| d(t) - n\widetilde{\ell}^+(t/n) \right| = o_{L^2}(n). \tag{3.6}$$

4 Depth analysis

In this section we sketch an alternative approach where we study the depth directly, without using the stack index. Details will be given in the full paper. In particular, for simplicity we approximate below generating functions and other quantities by their limits as $n \to \infty$, sometimes omitting careful estimates of the errors. We consider only the case $\lambda > 1$, since otherwise there is no giant tree and the depth d_t is small for all t.

Recall that the outdegree η has the p.g.f. $G_\eta(z) = \sum_k \mathbb{P}(\eta = k)z^k$, and that at time t, the number of outgoing arcs from v_t that lead to vertices not already seen has the distribution η_θ with p.g.f. $G_{\eta_\theta}(z) = \eta(\theta + (1-\theta)z)$, see (3.2).

4.1 Splitting the giant tree

Recall the definitions of θ_0 and θ_1 from (2.13)–(2.14). It is easily seen that these are, roughly, the proportions of discovered nodes when we find the top of the giant tree, and when we leave the giant tree, respectively. Let $\varepsilon \in (0,1)$ be a fixed number which will be specified later. We divide the interval $[0, n\theta_1]$ into three intervals: $\mathcal{U}_n = [0, \theta_0 n - n^{1-\varepsilon}]$ called the *climbing zone*, $\mathcal{D}_n = [\theta_0 n + n^{1-\varepsilon}, \theta_1 n]$, called the *descending zone*, and the band $\mathcal{B}_n = [\theta_0 n - n^{1-\varepsilon}, \theta_0 n + n^{1-\varepsilon}]$, called the *neutral zone*.

For $t \in \mathcal{U}_n$, we say that v_t is a *ladder vertex* if for all $t' \in \mathcal{U}_n$: $t' > t \Rightarrow d(t') \geqslant d(t)$. In other words, t is a right minimum of the function $d(t)$ in $\mathcal{U}$. For an integer i, we define $d_1^{-1}(i)$ as the last integer t in $\mathcal{U}_n$ such that $d(t) = i$ (provided such a t exists); note that then v_t is a ladder vertex. Also, we say that t is an *up-time* if v_t is a ladder vertex and t is the first time it is visited.

As a property of DFS, the ladder vertices are also revisited in the descending zone. Let v be a ladder vertex visited at time t in the climbing zone. When revisited the depth would be back at $d(t)$, see Figure 2.

4.2 Average interval between two up-times

Assume that $t \in \mathcal{U}_n$ is an up-time of a ladder vertex v_t with depth $d(t) = i$. Then (assuming that v_t is not the last ladder vertex) the next up-time t' finds the next ladder vertex $v_{t'}$. Clearly, $v_{t'}$ must be a child of v_t in the DFS, so its depth $d(t') = i + 1$, and it is the first child such that the DFS does not return to v_t before the end of $\mathcal{U}_n$. Given an up-time t, the future events in $\mathcal{U}_n$, including the value of t', are independent of the past history of the DFS. If we ignore the variation of θ in the interval $[t, t']$, then the DFS there can be regarded as a BGW tree $\mathcal{T}_\theta$ (rooted at v_t) with offspring distribution η_θ; since we assume that $\theta < \theta_0$, the BGW tree $\mathcal{T}_\theta$ is supercritical. We denote by $g_\theta(z)$ the generating function of the total number of vertices in $\mathcal{T}_\theta$ when the number is finite; it satisfies the fixed point equation:

$$g_\theta(z) = zG_{\eta_\theta}(g_\theta(z)). \tag{4.1}$$

thus $g_\theta(1) = 1 - \rho(\theta)$, the BGW tree extinction probability in (3.3).
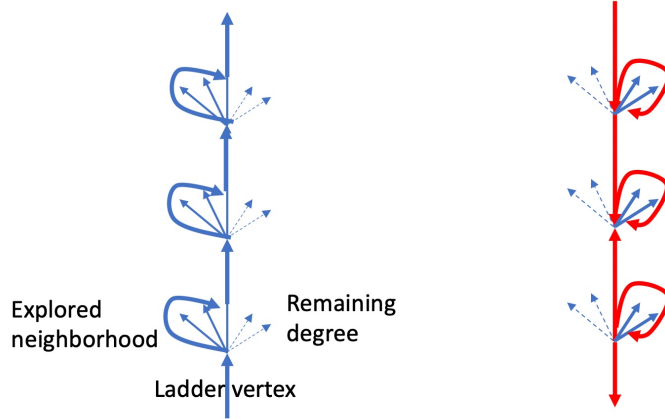
■ **Figure 2** The DFS in the giant tree: the climbing phase (left) and the descending phase (right) through ladder vertices.

We denote by $H(z, \theta)$ the probability generating function of the number of visited nodes in the subtrees of the root of $\mathcal{T}_\theta$ before hitting an infinite subtree, under the condition that the root has indeed an infinite subtree. Let H' denote the derivative with respect to the first variable z.

▶ **Theorem 7.** *We have the identity*

$$H(z, \theta) = 1 + \frac{z - 1}{1 - g_\theta(z)} \tag{4.2}$$

and thus the expected number of visited nodes in the BGW tree is

$$H'(1, \theta) = \frac{1}{1 - g_\theta(1)} = \frac{1}{\rho(\theta)} = O((\theta_0 - \theta)^{-1})). \tag{4.3}$$

The p.g.f. of the number of visited vertices between two consecutive up-times t and t' is thus asymptotically, with $\theta = t/n$

$$H\left(z, \theta \pm O\left(\frac{1}{n(\theta_0 - \theta)^2}\right)\right) \tag{4.4}$$

and the average number of new visited vertices is $H'(1, \theta) + O\left(\frac{1}{n(\theta_0 - \theta)^4}\right)$.

Proof. If k is the number of children of the root in the BGW tree $\mathcal{T}_\theta$, then the generating function of the number of subtrees before finding an infinite one is $\sum_{\ell < k} z^\ell (1 - \rho(\theta))^\ell \rho(\theta)$. The generating function of the number of visited vertices in the BGW tree before finding an infinite subtree is thus $z\rho(\theta) \sum_{\ell < k} (g_\theta(z))^\ell$. Summing over the offspring distribution we get the unconditional generating function

$$z\frac{\rho(\theta)}{1 - g_\theta(z)} \left(G_{\eta_\theta}(1) - G_{\eta_\theta}(g_\theta(z))\right) = \frac{\rho(\theta)}{1 - g_\theta(z)}(z - g_\theta(z))$$

$$= \rho(\theta) \left(\frac{z - 1}{1 - g_\theta(z)} + 1\right). \tag{4.5}$$

To get the p.g.f. $H(z, \theta)$ of the number of discovered nodes conditioned on $\mathcal{T}_\theta$ having at least one infinite subtree, one must divide by $\rho(\theta)$, which is the probability that the BGW tree is infinite. The error term comes from the fact that the fraction of discovered vertices varies

during the interval in the proportion of the size of the interval divided by n. If we look at the generating function $H(z, \theta)$ it turns out (see the appendix) that the interval multiplied by $(\theta_0 - \theta)^2$ is bounded in probability, thus a variation of θ of order $O\big((\theta_0 - \theta)^{-2}/n\big)$. ◀

Let Δ_i be the time interval between two consecutive up-times t and t' such that $d(t) = i$. To get the number $u(t, t')$ of up-times in an interval $(t, t']$ one must find the probability that $\Delta_i + \Delta_{i+1} + \cdots + \Delta_{i+k} \leqslant t' - t$. Since the up-times are renewal points in the interval $\mathcal{U}_n$, we have, given that t is an up-time,

$$
\begin{cases}
\mathbb{P}(\Delta_i + \Delta_{i+1} + \cdots + \Delta_{i+k} > t' - t) & = & \mathbb{P}(u(t, t') < k) \\
\mathbb{P}(\Delta_i + \Delta_{i+1} + \cdots + \Delta_{i+k} \leqslant t' - t) & = & \mathbb{P}(u(t, t') \geqslant k).
\end{cases}
\tag{4.6}
$$

Thus by Chebychev's inequality, we have the Chernoff type bounds

$$
\forall x > 0 : \begin{cases}
\mathbb{P}(u(t, t') < k) & \leqslant & H(e^x, t/n)^k e^{-(t'-t)x} \\
\mathbb{P}(u(t, t') \geqslant k) & \leqslant & H(e^{-x}, t'/n)^k e^{(t'-t)x}
\end{cases}
\tag{4.7}
$$

The estimate of these probabilities will come from an easy application of generating functions.

▶ **Lemma 8.** *When the quantity θ is smaller than but close to θ_0, we have $\rho(\theta) = 1 - g_\theta(1) = 2\frac{\lambda^3}{\lambda_2}(\theta_0 - \theta) + O((\theta_0 - \theta)^2)$ where $\lambda_2 = G_\eta''(1)$. Furthermore the generating function $g_\theta(z)$ has a radius of convergence which is $1 + \frac{\lambda^4}{2\lambda_2}(\theta_0 - \theta)^2 + O((\theta_0 - \theta)^3)$, and for z such that $|z - 1| \ll (\theta_0 - \theta)^2$ we have*

$$
g_\theta(z) = 1 - 2\frac{\lambda^3}{\lambda_2}(\theta_0 - \theta) + \frac{z-1}{\lambda(\theta_0 - \theta)} + O\Big(\frac{(z-1)^2}{(\theta_0 - \theta)^3} + (\theta_0 - \theta)^2\Big).
\tag{4.8}
$$

$$
H(z, \theta) = 1 + \frac{\lambda_2(z-1)}{2\lambda^3(\theta_0 - \theta)} + O\Big(\frac{(z-1)^2}{(\theta_0 - \theta)^3} + (\theta_0 - \theta)^2\Big)
\tag{4.9}
$$

$$
H'(1, \theta) = \frac{\lambda_2}{2\lambda^3(\theta_0 - \theta)} + O(1).
\tag{4.10}
$$

Proof. In Appendix A. ◀

▶ **Lemma 9.** *Let $\varepsilon > 0$ and $\frac{1-\varepsilon}{2} < \alpha < 1 - \epsilon$. For $a, A > 0$ there exist $B > 0$ and $C > 0$ such that for $t < t' \in \mathcal{U}_n$ with $an^{1-\varepsilon} \leqslant t' - t \leqslant An^{1-\varepsilon}$ we have*

$$
\mathbb{P}\left(u(t, t') \geqslant \frac{1}{H'(1, \theta)}(t' - t + Bn^\alpha)\right) < \exp(-Cn^{2\alpha - 1 - \varepsilon}),
\tag{4.11}
$$

$$
\mathbb{P}\left(u(t, t') < \frac{1}{H'(1, \theta')}(t' - t - Bn^\alpha)\right) < \exp(-Cn^{2\alpha - 1 - \varepsilon}).
\tag{4.12}
$$

Proof. Now t and t' can be within $n^{1-\varepsilon}$ from $n\theta_0$ which has an impact on the estimate of $\log H(e^x, \theta)$. For the super-critical BGW tree, we know from Lemma 8 and its proof that $H'(1, \theta) = O((\theta_0 - \theta)^{-1})$ and that, for $|x| \ll (\theta_0 - \theta)^2$,

$$
H(e^x, \theta) \leqslant \exp\left(xH'(1, \theta) + D\frac{x^2}{(\theta_0 - \theta)^3}\right)
\tag{4.13}
$$

for some $D > 0$.

For (4.11), let $t' - t = kH'(1, \theta) - Bn^\alpha$ for some $B > 0$ and $0 < \alpha < 1 - \varepsilon$. Thus $k = \frac{1}{H'(1, \theta)}(t' - t - Bn^\alpha)$ which makes k of the order $n^{1-\varepsilon}(\theta_0 - \theta) \gtrsim n^{1-2\varepsilon}$. We use the estimates (4.7). The quantity $H(e^x, \theta)^k e^{-(t'-t)x}$ in (4.7) is smaller than $\exp(\frac{Dkx^2}{(\theta_0 - \theta)^3} - Bxn^\alpha)$,

since the terms in $kxH'(1,\theta)$ cancel. The minimal value of this quantity is reached for $x = \frac{B}{2Dk} n^\alpha (\theta_0 - \theta)^3$ which is of order $\frac{B}{2D(t'-t)} n^\alpha (\theta_0 - \theta)^3 H'(1,\theta) \lesssim n^{\alpha-1+\varepsilon}(\theta_0 - \theta)^2 \ll (\theta_0 - \theta)^2$. This minimal value is $\exp(-\frac{B^2}{4kD}(\theta_0 - \theta)^3 H'(1,\theta)n^{2\alpha}) \leqslant \exp(-Cn^{2\alpha-1-\varepsilon})$ for some $C > 0$.

The proof of (4.12) is essentially the same.   ◀

At this point it is tempting to take the optimal value for α to be $(1-\varepsilon)/2$ with ε close to 0. But in fact this would ignore that the quantities $H'(1,\theta)$ and $H'(1,\theta')$ may differ by a factor bounded away from 1, which would introduce an error greater than the term Bn^α. This in fact reduce the possibilities for $1 - \varepsilon$ and α for a consistent deviation probability. The values $\varepsilon = 1/5$ and $\alpha = 3/5$ seem to be the limiting values that we can achieve as we show next.

▶ **Theorem 10.** *For all $\beta > \frac{4}{5}$ for all $t \in \mathcal{U}_n$ we have $d(t) = n \int_0^{t/n} \rho(\theta)d\theta + o_{\mathrm{p}}(n^\beta)$.*

Proof. Assume first that t is an up-time, and that there is an increasing sequence of up-times (not necessarily consecutive) $t_0, t_1, \ldots, t_\ell$ with $t_0 = 0$ and $t_\ell = t$ with the constraint that for all $i < \ell$: $t_{i+1} - t_i \leqslant An^{1-\varepsilon}$. Since $u(0,t) = \sum_i u(t_i, i+1)$, it turns out that the probability that there exists i such that $u(t_i, t_{i+1}) > \frac{1}{H'(1,t_i/n)}(t_{i+1} - t_i + Bn^\alpha)$ is smaller than $n \exp(-Cn^{2\alpha-1-\varepsilon})$ as shown in Lemma 9. Therefore with high probability $\sum_i u(t_i, t_{i+1}) \leqslant \sum_i \frac{t_{i+1}-t_i}{H'(1,t_i/n)} + O(n^{\alpha+\varepsilon})$, provided that $2\alpha - 1 - \varepsilon > 0$.

Similarly with high probability $\sum_i u(t_i, t_{i+1}) \geqslant \sum_i \frac{t_{i+1}-t_i}{H'(1,t_{i+1}/n)} - O(n^{\alpha+\varepsilon})$. Since both

$$\sum_i \frac{t_{i+1} - t_i}{H'(1, t_i/n)} = n \int_0^{t/n} \frac{d\theta}{H'(1,\theta)} + O(n^{1-\varepsilon}) \tag{4.14}$$

$$\sum_i \frac{t_{i+1} - t_i}{H'(1, t_{i+1}/n)} = n \int_0^{t/n} \frac{d\theta}{H'(1,\theta)} - O(n^{1-\varepsilon}) \tag{4.15}$$

it follows that with high probability $u(0,t) = n \int_0^{t/n} \frac{d\theta}{H'(1,\theta)} + O(n^{1-\varepsilon}) + O(n^{\alpha+\varepsilon})$, provided that $2\alpha - 1 - \varepsilon > 0$. The smallest error order is when $1 - \varepsilon = \alpha + \varepsilon$, thus when $1 - \varepsilon = \alpha + \varepsilon > 4/5$.

When t is not an up-time, the same result holds since the extra depth explored after the last up-time is $O_{\mathrm{p}}(n^{2\varepsilon})$, well within the error term $O(n^\beta)$.   ◀

4.3   Average interval between two down times

During the descent phase of the giant tree (*i.e.* the values of t in $\mathcal{D}_n$), the DFS revisits all ladder vertices met during the ascending phase. The process consists of exploring the remaining outgoing neighbours of the ladder vertices that have not been visited during the ascending phase.

We notice that given the ladder times, the remaining outdegrees of the ladder vertices form a sequence of independent random variable.

Let t the time at which a ladder vertex is again revisited. Let t' be the first time at which it was visited during the ascending period, thus $d(t') = d(t)$. Let $\theta = t/n \in (\theta_0, \theta_1)$ and $u(\theta) = t'/n$. We have the following theorem.

▶ **Theorem 11.** *The p.g.f. of the remaining degree is asymptotically*

$$\tilde{\eta}_\theta(z) = (1 - u(\theta))\frac{G_\eta(z) - 1 + \rho(u(\theta))}{z - 1 + (1 - u(\theta))\rho(u(\theta))}. \tag{4.16}$$

Proof. Let $\tilde{\eta}_{\theta,k}(z)$ be the generating function of the remaining degree of the vertex v_t assuming the latter has outdegree k. Then, asymptotically,

$$\tilde{\eta}_{\theta,k}(z) \quad = \quad \sum_{\ell < k} (1 - a(u(\theta))^\ell\, a(u(\theta)) z^{k-1-\ell} \tag{4.17}$$

$$= \quad a(u(\theta)) \frac{z^k - (1 - a(u(\theta)))^k}{z - 1 + a(u(\theta))} \tag{4.18}$$

where $a(\theta) = (1 - \theta)\rho(\theta)$ is the probability that a new vertex has not yet been visited and is the root of an infinite BGW tree.

Summing over all outdegrees weighted by their probabilities we get the unconditional generating function of the remaining outdegree: $a(u(\theta)) \frac{G_\eta(z) - G_\eta(1 - a(u(\theta)))}{z - 1 + a(u(\theta))}$ which is equal to $a(u(\theta)) \frac{G_\eta(z) - 1 + \rho(u(\theta))}{z - 1 + a(u(\theta))}$ since (3.3) yields $G_\eta(1 - a(u(\theta))) = G_{\eta_{u(\theta)}}\big(1 - \rho(u(\theta))\big) = 1 - \rho(u(\theta))$. The above expression has value $\rho(u(\theta))$ at $z = 1$ and therefore must be divided by this factor in order to get the p.g.f. $\tilde{\eta}_\theta(z)$. ◀

The "down-times" in $\mathcal{D}_n$ bear some symmetries with the up-times in $\mathcal{U}_n$, with time reversed.

▶ **Theorem 12.** *The p.g.f. of the number of visited vertices between two consecutive down-times is asymptotically*

$$\overline{H}_\theta(z) = (1 - u(\theta)) \frac{\frac{g_\theta(z)}{z} - 1 + \rho\left(u(\theta)\right)}{(1 - \theta)(g_\theta(z) - 1) + (1 - u(\theta))\rho\left(u(\theta)\right)} \tag{4.19}$$

and the average number of visited vertices is $\overline{H}'_\theta(1) = \frac{1}{\rho(u(\theta))} \frac{\lambda - \frac{1}{1 - u(\theta)}}{\frac{1}{1 - \theta} - \lambda}$.

Proof. We have $\overline{H}_\theta(z) = \tilde{\eta}_\theta(\theta + (1 - \theta)g_\theta(z))$, which simplifies using (3.2) and (4.1). ◀

Let $\bar{\Delta}_i$ be the number of new visited nodes between two consecutive down-times t and t' such that $d(t') = i$. To get the number $d(t, t')$ of down-times between t and t' we use the renewal property of the sequence of down-times. With a similar reasoning as in (4.6)–(4.7) we get

$$\begin{cases} \mathbb{P}(\bar{\Delta}_i + \bar{\Delta}_{i-1} + \cdots + \bar{\Delta}_{i-k} > t' - t) &= \mathbb{P}(d(t, t') < k) \\ \mathbb{P}(\bar{\Delta}_i + \bar{\Delta}_{i-1} + \cdots + \bar{\Delta}_{i-k} \leqslant t' - t) &= \mathbb{P}(d(t, t') \geqslant k). \end{cases} \tag{4.20}$$

and

$$\forall x > 0 \;: \begin{cases} \mathbb{P}(d(t, t') < k) &\leqslant (\bar{\Delta}_{t'/n}(e^x))^k e^{-(t'-t)x} \\ \mathbb{P}(d(t, t') \geqslant k) &\leqslant (\bar{\Delta}_{t/n}(e^{-x}))^k e^{(t'-t)x}. \end{cases} \tag{4.21}$$

We have a theorem whose proof is absolutely similar to the proof of Theorem 10:

▶ **Theorem 13.** *For all $\beta > \frac{4}{5}$ with high probability for all $t \in \mathcal{D}_n$ we have* $d(t) = n \int_{t/n}^{\theta_1} \frac{\mathrm{d}\theta}{\overline{H}'_\theta(1)} + O(n^\beta)$.

For $\theta \in [\theta_0, \theta_1]$, define $\widehat{\ell}(\theta) = \int_\theta^{\theta_1} \frac{\mathrm{d}\tau}{\overline{H}'_\theta(1)}$. Since $d(t) = d(t')$, where asymptotically $t/n = \theta \in [\theta_0, \theta_1]$ and $t'/n = u(\theta) \in [0, \theta_0]$, we obtain from Theorems 10 and 13, recalling (3.4),

$$\widehat{\ell}(\theta) = \int_0^{u(\theta)} \rho(x)\, \mathrm{d}x = \widetilde{\ell}(u(\theta)). \tag{4.22}$$

▶ **Theorem 14.** *We have the identity $u(\theta) = \breve{\theta}$ in (3.5).*

Proof. Theorem 12 yields

$$\widehat{\ell}'(\theta) = -\frac{1}{\overline{H}'_\theta(1)} = -\frac{\rho(u(\theta))}{\lambda - \frac{1}{1-u(\theta)}} \frac{1-(1-\theta)\lambda}{1-\theta}. \tag{4.23}$$

Furthermore $\widetilde{\ell}'(\theta) = \rho(\theta)$ by (3.4). Hence, (4.22) implies

$$\frac{\mathrm{d}u(\theta)}{\mathrm{d}\theta} = \frac{\widehat{\ell}'(\theta)}{\widetilde{\ell}'(u(\theta))} = \frac{\frac{1}{1-\theta}-\lambda}{\frac{1}{1-u(\theta)}-\lambda}. \tag{4.24}$$

Similarly, the identity $\widetilde{\imath}(\breve{\theta}) = \widetilde{\imath}(\theta)$ in (3.5) yields, recalling (2.12),

$$\frac{\mathrm{d}\breve{\theta}}{\mathrm{d}\theta} = \frac{\widetilde{\imath}'(\theta)}{\widetilde{\imath}'(\breve{\theta})} = \frac{\frac{1}{1-\theta}-\lambda}{\frac{1}{1-\breve{\theta}}-\lambda}. \tag{4.25}$$

Since $u(\theta_0) = \breve{\theta}_0 = \theta_0$, the differential equations (4.24) and (4.25) have the same solution. We call the function $\theta \mapsto \breve{\theta}$ the mirror function; note that it only depends of average outdegree without any further consideration on the details of the outdegree distribution. All computations done we have $\breve{\theta} = 1 + \frac{1}{\lambda}W_{-1}\left(-(1-\theta)e^{(\theta-1)\lambda}\right)$ using the branch $W_{-1}(\cdot)$ of the Lambert W function, see Figure 3 for $\lambda = 2$. ◀
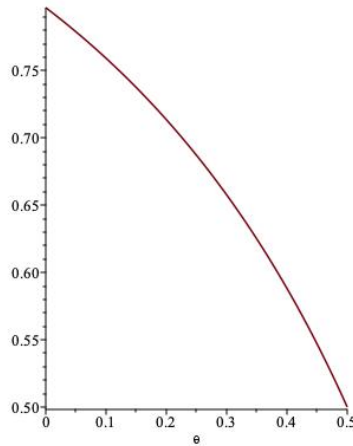
Figure 3 The mirror function for $\lambda = 2$.

▶ Remark 15. Theorems 6 and 13 show that $\widehat{\ell}(\theta) = \widetilde{\ell}^+(\theta)$ for $\theta \in [\theta_0, \theta_1]$. Furthermore, the identity $u(\theta) = \breve{\theta}$ follows also directly from Theorem 6 and (3.5) together with $d(t) = d(t')$.

5 An example

We take the example of a constant outdegree. In the case of outdegree 2, $G_\eta(z) = z^2$, $\lambda = 2$, and $\theta_0 = \frac{1}{2}$. Consider for simplicity only $\theta \in [0, \theta_0]$. Then, (3.3) has the explicit solution $\rho(\theta) = 1 - \left(\frac{\theta}{1-\theta}\right)^2$, and Theorems 2 and 6 show that if $t/n \to \theta$, then

$$\frac{1}{n}I(t) \xrightarrow{L^2} \widetilde{\imath}^+(\theta) = 2\theta + \log(1-\theta), \tag{5.1}$$

$$\frac{1}{n}d(t) \xrightarrow{L^2} \widetilde{\ell}^+(\theta) = -2\log(1-\theta) - \frac{\theta}{1-\theta}, \tag{5.2}$$

In the case of higher constant outdegree, say 8, ρ must be calculated from the implicit formula (3.3), for $\theta \leqslant \theta_0 = 1 - \frac{1}{8}$. Larger outdegrees can be treated similarly. Figure 4 displays the average depth and index for constant outdegree 2 and 8. We notice that for $\eta = 2$, the stack size is smaller than the depth. This might be a surprise because at each new discovered node the stack stores the whole set of outgoing arcs, while the depth increases by at most 1. The reason is that in this case $\eta = 2$, and thus, although the stack keeps all unexplored arcs for all ancestors, this is at most one arc for each ancestor and many ancestor have no arc left. For η larger this disappears, and the stack size becomes larger than the depth.

6 Conclusion

We have presented an analysis of the Depth-First Search algorithm by Tarjan in a model of random graphs recently introduced by Don Knuth. We have presented a version of the stack model which can be analyzed as a Markov chain and is much easier to analyse than the real depth. The latter requires new insights in a model of Bienaymé–Galton–Watson trees with a varying extinction probability, in particular when close to the sub-critical case.
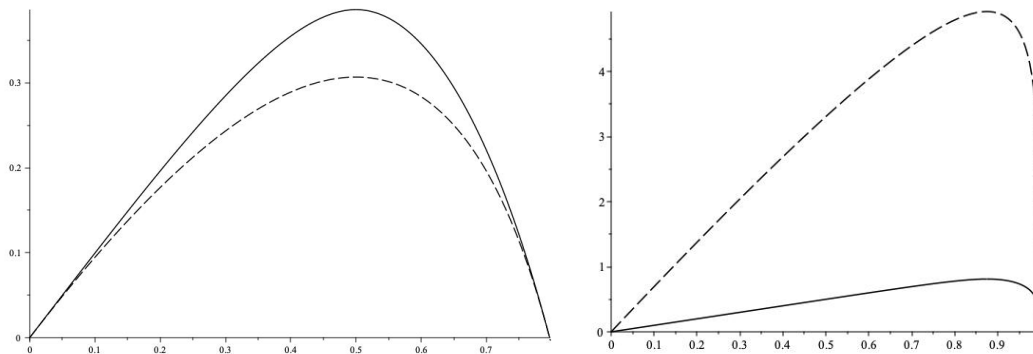
Figure 4 The limit of $\frac{1}{n}I(t)$ (dashed) and $\frac{1}{n}d(t)$ (solid) as functions of $\theta = t/n$ for $\theta \in [0, \theta_0]$, for the example $\eta = 2$ (left), $\eta = 8$ (right).

References

1 Jürgen Bennies and Götz Kersting. A random walk approach to galton–watson trees. *J. Theoret. Probab.*, 3:777–803, 2000.

2 Sahar Diskin and Michael Krivelevich. On the performance of the depth first search algorithm in supercritical random graphs. *Electron. J. Combin.*, 3, 2022.

3 Nathanaël Enriquez, Gabriel Faraud, and Laurent Ménard. Limiting shape of the depth first search tree in an erdős-rényi graph. *Random Structures Algorithms*, 2:501–516, 2020.

4 Nathanaël Enriquez, Gabriel Faraud, Laurent Ménard, and Nathan Noiry. Depth first exploration of a configuration model. *Electron. J. Probab.*, 53, 2022.

5 Jean-François Le Gall and Yves Le Jan. Branching processes in lévy processes: the exploration process. *Ann. Probab.*, 1:213–252, 1998.

6 Philippe Jacquet and Svante Janson. Depth-first search performance in a random digraph with geometric outdegree distribution. *La Matematica*, 2024.

7 Donald E. Knuth. *The Art of Computer Programming.* Preliminary draft, URL: `https://cs.stanford.edu/~knuth/fasc12a.ps.gz`, 2022.

8 Michael Krivelevich and Benny Sudakov. The phase transition in random graphs: a simple proof. *Random Structures Algorithms*, 2:131–138, 2013.

9 Jayant Madhavan, David Ko, Łucja Kot, Vignesh Ganapathy, Alex Rasmussen, and Alon Halevy. Google's deep web crawl. *Proceedings of the VLDB Endowment*, 2:1241–1252, 2208.

A   Appendix

Proof of Lemma 8. We want first to determine the first fixed point $g_\theta(1)$ of G_{η_θ}. We have the expansion

$$G_\eta(y) = 1 + \lambda(y-1) + \frac{\lambda_2}{2}(y-1)^2 + O((y-1)^3) \tag{A.1}$$

where λ_2 is $G_\eta''(1)$. Thus

$$G_{\eta_\theta}(y) = G_\eta(\theta + (1-\theta)y) = 1 + (1-\theta)\lambda(y-1) + (1-\theta)^2 \frac{\lambda_2}{2}(y-1)^2 + O((y-1)^3) \tag{A.2}$$

We look at the fixed point equation $G_{\eta_\theta}(y) - y = 0$ which is equivalent to

$$\lambda(\theta_0 - \theta)(y-1) + \frac{\lambda_2(1-\theta)^2}{2}(y-1)^2 + O((y-1)^3) = 0 \tag{A.3}$$

The natural root is 1 but the other root is $1 - 2\frac{\lambda(\theta_0-\theta)}{\lambda_2(1-\theta)^2} + O((\theta_0-\theta)^2)$ which yields the claimed result since $\frac{1}{1-\theta} = \lambda + O(\theta_0 - \theta)$.

To determine the function $g_\theta(z)$ one must solve the equation $zG_{\eta_\theta}(y) = y$ which writes:

$$1 - \frac{1}{z} + (1 + \lambda(\theta_0 - \theta) - \frac{1}{z})(y-1) + \frac{\lambda_2(1-\theta)^2}{2}(y-1)^2 + O((y-1)^3) = 0. \tag{A.4}$$

Retaining only the second order (the extra order will be in $O((\theta_0 - \theta)^3)$ we get an equation which is degenerate when $\left(1 + \lambda(\theta_0 - \theta) - \frac{1}{z}\right)^2 - 2\lambda_2(1-\theta)^2(1 - \frac{1}{z}) = 0$, *i.e.* when $1 - \frac{1}{z} = \frac{\lambda^2(\theta_0-\theta)^2}{2\lambda_2(1-\theta)^2} + O((\theta_0 - \theta)^3)$. In the above expression we can substitute $(1-\theta)$ with $\frac{1}{\lambda}$ to a $O((\theta_0 - \theta)^3)$ error term.

Assuming $|z - 1| \ll (\theta_0 - \theta)^2$ we have the expression of the root $y = g_\theta(z)$:

$$y = 1 - 2\frac{\lambda^3}{\lambda_2}(\theta_0 - \theta) + \frac{z-1}{\lambda(\theta_0-\theta)} + O\left(\frac{(z-1)^2}{(\theta_0-\theta)^3} + (\theta_0 - \theta)^2\right). \tag{A.5}$$

Now the expression of $H(z, \theta)$ satisfies

$$H(z,\theta) = 1 + \frac{z-1}{1 - g_\theta(z)} \tag{A.6}$$

$$= 1 + \frac{z-1}{2\frac{\lambda^3}{\lambda_2}(\theta_0-\theta) - \frac{z-1}{\lambda(\theta_0-\theta)} + O\left(\frac{(z-1)^2}{(\theta_0-\theta)^3} + (\theta_0-\theta)^2\right)} \tag{A.7}$$

$$= 1 + \frac{\lambda_2(z-1)}{2\lambda^3(\theta_0-\theta)}\left(\frac{1}{1 - \frac{\lambda_2(z-1)}{2\lambda^4(\theta_0-\theta)^2} + O\left(\frac{(z-1)^2}{(\theta_0-\theta)^4} + (\theta_0-\theta)\right)}\right) \tag{A.8}$$

$$= 1 + \frac{\lambda_2(z-1)}{2\lambda^3(\theta_0-\theta)} + \frac{\lambda_2^2(z-1)^2}{4\lambda^7(\theta_0-\theta)^3} + O\left(\frac{|z-1|^3}{(\theta_0-\theta)^5} + |z-1|\right) \tag{A.9}$$

$$= 1 + \frac{\lambda_2(z-1)}{2\lambda^3(\theta_0-\theta)} + O\left(\frac{(z-1)^2}{(\theta_0-\theta)^3} + |z-1|\right) \tag{A.10}$$

In particular,

$$H'(1,\theta) = \frac{\lambda_2}{2\lambda^3(\theta_0-\theta)} + O(1). \tag{A.11}$$

Since the terms added after the unity are small compared to 1 we can equivalently state that

$$H(z, \theta) = \exp\left(\frac{\lambda_2(z-1)}{2\lambda^3(\theta_0 - \theta)} + O\left(\frac{(z-1)^2}{(\theta_0 - \theta)^3} + |z - 1|\right)\right). \tag{A.12}$$

We also have by (4.2), still assuming $|z - 1| \ll (\theta_0 - \theta)^2$,

$$\begin{aligned}
H(z, \theta) - 1 - (z-1)H'(1, \theta) &= \frac{z-1}{1 - g_\theta(z)} - \frac{z-1}{1 - g_\theta(1)} \\
&= (z-1)\frac{g_\theta(z) - g_\theta(1)}{(1 - g_\theta(z))(1 - g_\theta(1))}
\end{aligned} \tag{A.13}$$

Furthermore, by taking the derivative of the fixed point equation $zG_{\eta_\theta}(g_\theta(z)) = g_\theta(z)$, we find, using (A.2) and (A.5),

$$\begin{aligned}
g'_\theta(z) &= \frac{G_{\eta_\theta}(g_\theta(z))}{1 - zG'_{\eta_\theta}(g_\theta(z))} = \frac{g_\theta(z)/z}{-\lambda(\theta_0 - \theta) - (1 - \theta)^2\lambda_2(g_\theta(z) - 1) + o(|\theta_0 - \theta|^2)} \\
&= \frac{1}{\lambda(\theta_0 - \theta) + o(|\theta_0 - \theta|^2)} = O\left((\theta_0 - \theta)^{-1}\right).
\end{aligned} \tag{A.14}$$

Thus (A.13) and the mean-value theorem together with (A.5) yield

$$H(z, \theta) - 1 - (z-1)H'(1, \theta) = O\left(|z - 1|^2/(\theta_o - \theta)^3\right). \tag{A.15}$$

We can rewrite this as, for $|x| \ll (\theta_0 - \theta)^2$, using the estimate (A.11),

$$H(e^x, \theta) = \exp\left(xH'(1, \theta) + O\left(\frac{x^2}{(\theta_0 - \theta)^3}\right)\right). \tag{A.16}$$

◀