Persisting randomness in randomly growing discrete structures: graphs and search trees

Rudolf Grübel

Leibniz Universität Hannover, Germany

The successive discrete structures generated by a sequential algorithm from random input constitute a Markov chain that may exhibit long term dependence on its first few input values. Using examples from random graph theory and search algorithms we show how such persistence of randomness can be detected and quantified with techniques from discrete potential theory. We also show that this approach can be used to obtain strong limit theorems in cases where previously only distributional convergence was known.

Keywords: Boundary theory, Markov chains, random graphs, search trees.

1 Introduction

Given a sequence $t_1, t_2, \ldots$ of input values, a sequential algorithm produces an output sequence $x_1, x_2, \ldots$, where the next output x_{n+1} depends on the current state x_n and the next input t_{n+1} only. For cases where x_n is a discrete structure, such as a permutation or a graph, and where the input values are realizations of independent random variables with the same distribution, the output sequence is a Markov chain $X = (X_n)_{n \in \mathbb{N}}$ that is adapted to a combinatorial family $\mathbb{F}$ in the sense that X_n takes its values in the subset $\mathbb{F}_n \subset \mathbb{F}$ of objects with base parameter n. Markov chains of this type often exhibit *persisting randomness*: Informally, this means that the influence of early values does not disappear as time goes by; formally, it means that the tail σ-field $\mathcal{T}(X)$ associated with X is not trivial. Further, such chains eventually leave every fixed finite subset of $\mathbb{F}$ with probability 1, which leads to the related problem of finding a state space completion that captures the information contained in $\mathcal{T}(X)$.

The classical example is the Pólya urn: Initially, at time $n = 1$, the urn contains one red and one blue ball. At time $n = 2, 3, \ldots$, a ball is selected uniformly at random and put back, together with another ball of the same colour. Here $\mathbb{F}_n$ is the set of the pairs $(i, j) \in \mathbb{N}_0 \times \mathbb{N}_0$ with $i + j + 1 = n$, where i and j are the number of balls of the two colours added up to time $n \in \mathbb{N}$. A suitable augmentation of the state space $\mathbb{F} = \mathbb{N}_0 \times \mathbb{N}_0$ is obtained by regarding a sequence $((i_n, j_n))_{n \in \mathbb{N}}$ with $i_n + j_n \to \infty$ as convergent if and only if $(i_n + 1)/(i_n + j_n + 2)$ (the proportion of red balls) tends to a value $\alpha \in [0, 1]$ as $n \to \infty$, which leads to a state space completion $\bar{\mathbb{F}}$ that may be represented by $\mathbb{F} \cup [0, 1]$. With respect to this convergence we have $X_n \to X_\infty$ almost surely, where X_∞ generates the (non-trivial) tail σ-field $\mathcal{T}(X)$.

Discrete potential theory provides a general method for the construction of such state space boundaries. This was initiated in a fundamental paper by Doob (1959) and has been applied to Pólya urns by Blackwell

and Kendall (1964). A recent textbook treatment is given in Woess (2009); see also the survey by Sawyer (1997). Many authors have used boundary theory for the analysis of random walks on discrete structures; see Kaĭmanovich and Vershik (1983) for a very influential review, and the more recent monograph by Woess (2000).

In the present paper we regard the discrete structures themselves as states of a stochastic process. Some standard search algorithms have recently been investigated from this point of view in Evans et al. (2012), and the results have been used in Grübel (2014) to prove strong limit theorems for functionals of the output sequence, such as the path length or Wiener index of search trees, that have attracted the attention of many researchers. We address the phenomenon of randomness persistence with these tools, specifically in connection with some popular models for random graphs, and for search algorithms. Further, we show that the method can be used to obtain almost sure convergence for the structures themselves or for functionals of the structures in cases where previously only convergence in distribution was known.

In the next section we provide some background on discrete potential theory and the Doob-Martin compactification. In order to keep this short we restrict ourselves to combinatorial Markov chains, where the state space is graded by the time parameter. Section 3 gives some elementary examples from random graph theory; the present author is not aware of any previous use of the Doob-Martin compactification in connection with (general) graph limits. In Section 4 we consider search trees, where we can build on the work of Evans et al. (2012) and Grübel (2014). In a final section we collect some comments on related work and provide further pointers to the literature.

We hope that such results contribute to the theoretical understanding of the growth models and algorithms. From an entirely practical point of view persisting randomness should be of interest as a strong dependence on the first few input values may be an entirely unwelcome aspect of an algorithm that the practitioner may have to address, for example by an additional randomization step.

2 An ultrashort summary of Markov chain boundary theory

A Markov chain is a sequence $X = (X_n)_{n \in \mathbb{N}}$ of random variables that take their values in some countable set $\mathbb{F}$, the state space, such that the Markov property holds,

$$P(X_{n+1} = x_{n+1}|X_n = x_n, \ldots, X_1 = x_1) \ = \ P(X_{n+1} = x_{n+1}|X_n = x_n) \tag{1}$$

for all $n \in \mathbb{N}$, $x_1, \ldots, x_{n+1} \in \mathbb{F}$. In the cases we are interested in there will be a canonical state $e \in \mathbb{F}$ with $X_1 = e$, and the transitions are homogeneous in time, which means that for some function $p : \mathbb{F} \times \mathbb{F} \to [0, 1]$,

$$P(X_{n+1} = y|X_n = x) = p(x, y) \quad \text{for all } x, y \in \mathbb{F}, \ n \in \mathbb{N}.$$

These are the transition probabilities; together with the starting point e they determine the distribution of the stochastic process X. We also assume that

$$P(X_n = x \text{ for some } n \in \mathbb{N}) \ > \ 0 \ \text{ for all } x \in \mathbb{F}. \tag{2}$$

In words: Every state has a chance to be visited—the chain is weakly irreducible.

Boundary theory provides an approach to the asymptotics of chains that 'leave the state space' in the sense that $\lim_{n \to \infty} P(X_n \in S) = 0$ for every finite set $S \subset \mathbb{F}$. It gives the 'right' extension (completion, compactification) $\bar{\mathbb{F}}$ of the state space, in the sense that

$$X_n \to X_\infty \ \text{ as } n \to \infty \ \text{ with probability } 1 \tag{3}$$

for some random variable X_∞ with values in the boundary $\partial\mathbb{F}$ of $\mathbb{F}$ in $\bar{\mathbb{F}}$, and that

$$\sigma(X_\infty) =_{\text{a.s.}} \mathcal{T}(X) := \bigcap_{n=1}^\infty \sigma\big(\{X_m \,:\, m \geq n\}\big). \tag{4}$$

In words: The limit generates the tail σ-field of the process, up to null sets. For property (4) we assume that X has the space-time property, by which we mean that each state can be visited at one particular point in time only. The combinatorial Markov chains in Section 1 are such space-time processes.

This feat is achieved by the Doob-Martin compactification, where we regard a sequence $(y_n)_{n\in\mathbb{N}} \subset \mathbb{F}$ as convergent if the conditional probabilities $P(X_1 = x_1, \ldots, X_m = x_m | X_n = y_n)$ converge as $n \to \infty$ for all fixed $m \in \mathbb{N}$, $x_1, \ldots, x_m \in \mathbb{F}$. Due to the Markov property (1) the construction can be based on the Martin kernel K,

$$K(x, y) := \frac{P(X_n = y | X_m = x)}{P(X_n = y)}, \quad x, y \in \mathbb{F}, \, n > m,$$

where m and n are the time values associated with the states x and y respectively; here weak irreducibility (2) is important. Indeed, manipulations of elementary conditional probabilities lead to

$$P(X_1 = x_1, \ldots, X_m = x_m | X_n = y_n) = K(x_m, y_n)\, P(X_1 = x_1, \ldots, X_m = x_m),$$

which connects the convergence condition on the conditional probabilities to the convergence of the values of the Martin kernel. We mention in passing that this approach to Doob-Martin convergence is equivalent to the usual approach via potential kernels; see also (Evans et al., 2012, Section 3) and (Evans et al., 2014, Section 2).

From a general point of view, any family $\mathcal{F}$ of functions $f : \mathbb{F} \to \mathbb{R}$ that separates the points of $\mathbb{F}$ leads to an embedding of $\mathbb{F}$ into the space $\mathbb{R}^{\mathcal{F}}$ of functions from $\mathcal{F}$ to $\mathbb{R}$ via

$$x \mapsto \big(f \mapsto f(x)\big).$$

On $\mathbb{R}^{\mathcal{F}}$ we use the topology of pointwise convergence. If all $f \in \mathcal{F}$ are continuous (which they automatically are if we endow $\mathbb{F}$ with the discrete topology) and bounded, then the embedding is continuous and its range is a product of bounded intervals, hence compact by Tychonov's theorem. This is a variant of the Stone-Čech compactification, see (Kelley, 1955, p.152f). In this construction, all $f \in \mathcal{F}$ have a unique continuous extension to the whole of the compactified space. Alternatively, for a countable family $\mathcal{F}$, a suitable metric can be defined on $\mathbb{F}$ using these functions, such that the associated completion has these properties; see (Woess, 2009, p.187).

In our present setup, the Doob-Martin compactification arises by taking $\mathcal{F}$ to be the set of functions $y \mapsto K(x, y)$, $x \in \mathbb{F}$. We use the same symbol for the extended functions and denote boundary elements by lower case Greek letters. With this construction, (3) and (4) are satisfied. In addition, we have the following remarkable properties: First, all non-negative harmonic functions $h : \mathbb{F} \to \mathbb{R}$ can be written as mixtures of the functions $K(\cdot, \alpha)$, $\alpha \in \partial\mathbb{F}$. To be precise we recall that $h : \mathbb{F} \to \mathbb{R}$ is harmonic if $h(x) = \sum_{y\in\mathbb{F}} p(x, y) h(y)$ for all $x \in \mathbb{F}$. Then for each such h with $h \geq 0$ and $h(e) = 1$ there is a probability measure μ_h on (the Borel subsets of) the boundary $\partial\mathbb{F}$ such that

$$h(x) = \int K(x, \alpha)\, \mu_h(d\alpha) \quad \text{for all } x \in \mathbb{F}. \tag{5}$$

The distribution μ_1 of the limit X_∞ represents the (trivial) harmonic function $h \equiv 1$. Secondly, conditioned on a limit value $X_\infty = \alpha$, the process is again a Markov chain, with transition probabilities p^h given by

$$p^h(x,y) \;=\; \frac{1}{K(x,\alpha)}\, p(x,y)K(y,\alpha). \tag{6}$$

This is an instance of Doob's h-transform, with $K(\cdot,\alpha)$ the corresponding harmonic function h. Of course, the interpretation of these transforms by a conditioning on the final value is a natural consequence of the initial idea of conditioning on the values y_n at time n and then letting n tend to ∞.

As it is central to our theme of persisting randomness we briefly explain why (and how) X_∞ generates the tail σ-field, up to null sets, that is, why property (4) holds.

The limit is obviously $\mathcal{T}(X)$-measurable, which means that $\sigma(X_\infty) \subset \mathcal{T}(X)$. For the other direction we need, for each tail event A, a Borel subset B of $\partial\mathbb{F}$ such that

$$P\big(A \triangle X_\infty^{-1}(B)\big) \;=\; 0, \tag{7}$$

where $X_\infty^{-1}(B) := \{\omega \in \Omega \,:\, X_\infty(\omega) \in B\}$. Let $A \in \mathcal{T}(X)$ and let 1_A be the associated indicator function; we may assume that $\kappa := P(A) > 0$. As the state space is graded in the sense that it can be written as the disjoint union of the 'slices' $\mathbb{F}_n$ of states that are possible at time n, we can define $h : \mathbb{F} \to [0,\infty)$ by setting $h(x) = \kappa^{-1}P(A|X_n = x)$ for $x \in \mathbb{F}_n$, $n \in \mathbb{N}$. With (1) it follows that h is harmonic, and it turns out that the measure μ_h representing h as in (5) has a density Φ with respect to the distribution μ_1 of X_∞. The set required in (7) can now be given as $B = \Phi^{-1}(\{\kappa^{-1}\})$.

In particular, if the distribution of X_∞ is concentrated on a single value of the boundary then $\mathcal{T}(X)$ is P-trivial, so randomness 'disappears in the limit'.

3 Graph limits

Our basis in this section is the recent monograph by Lovász (2012), which also gives references to the original research articles. Let $\mathbb{G}[n]$ be the set of simple graphs $G = (V, E)$ with vertex set $V = [n] := \{1, \ldots, n\}$. The set $\mathbb{G}[1]$ has only one element, the graph $e = G_1$ with the single node 1 and no edges. A number of popular models for randomly growing graphs fits into the framework of combinatorial Markov chains, with state space $\mathbb{F} = \mathbb{G} := \bigcup_{n=1}^\infty \mathbb{G}[n]$ and start at G_1. We work out the boundary for two of them, the uniform attachment process, and the Erdős-Rényi graphs, where we consider two variants of the latter. We note that the state space compactifications are abstract constructions so that the only uniqueness that we may expect is up to homeomorphisms; usually there are many possibilities for a concrete description.

On its own the question of how to define limits of finite graphs, interpreted as the search for a completion or compactification of the countable set $\mathbb{G}$, does not involve any probability and, of course, it can have quite different answers depending on the specific circumstances. For example, we might distinguish between sparse and dense graphs, referring to the rate of growth of the number $e(G_n)$ of edges $E(G_n)$ in relation to the number $v(G_n)$ of vertices $V(G_n)$ of G_n in a sequence $(G_n)_{n\in\mathbb{N}} \subset \mathbb{G}$. For the dense case the notion of subgraph sampling has turned out to be important (there are several equivalent definitions): For two graphs $G, H \in \mathbb{G}$ let $t(H, G)$ be the number of possibilities to embed H into G or, more formally, with $\Gamma(H, G)$ the set of injective functions $\phi : V(H) \to V(G)$, let

$$T(H, G) := \big\{\phi \in \Gamma(H, G) \,:\, \{\phi(i), \phi(j)\} \in E(G) \Leftrightarrow \{i, j\} \in E(H)\big\}, \tag{8}$$

$$t(H,G) := \#T(H,G), \quad \rho(H,G) := \frac{(v(G) - v(H))!}{v(G)!} t(H,G). \tag{9}$$

We then say that a sequence $(G_n)_{n \in \mathbb{N}}$ converges if for all $H \in \mathbb{G}$ the relative number $\rho(H, G_n)$ of these possibilities converges as a sequence of real numbers. The value $\rho(H, G_n)$ can be interpreted as the probability that, choosing $m = v(H)$ elements of $V(G_n)$ randomly and without replacement, the subgraph of G_n induced on these nodes is isomorphic to H. The convergence may be rephrased in a somewhat abstract manner: We define an embedding of $\mathbb{G}$ into the set $[0,1]^{\mathbb{G}}$ of functions on $\mathbb{G}$ with values in the unit interval by

$$G \mapsto \big(H \mapsto \rho(H,G)\big), \quad G \in \mathbb{G}, \tag{10}$$

and then consider the closure of the range of the embedding as a compactification of $\mathbb{G}$. Note that the function space is compact with respect to pointwise convergence by Tychonov's theorem. Viewed this way, the similarity to the Doob-Martin compactification becomes apparent, where we use the embedding

$$x \mapsto \big(y \mapsto K(x,y)\big), \quad x \in \mathbb{F}, \tag{11}$$

based on the Martin kernel instead.

Returning to the Markov chain models of randomly growing graphs, we first consider the *uniform attachment* model; see (Lovász, 2012, Example 11.39). In order to describe its dynamics suppose that we are in state $G_n \in \mathbb{G}[n]$ at time n. We then construct $G_{n+1} \in \mathbb{G}[n+1]$ by adding those edges $\{i, j\} \subset [n+1]$ not (yet) in G_n with probability $1/(n+1)$, independently of each other. Let $X = (X_n)_{n \in \mathbb{N}}$ be the corresponding Markov chain, which has state space $\mathbb{G}$ and starts at G_1.

Let $\Delta := \{\{i,j\} \in \mathbb{N} \times \mathbb{N} : 1 \le i < j\}$. For each $\{i,j\} \in \Delta$ we define the edge indicator $1_{\{i,j\}} : \mathbb{G} \to \{0,1\}$ to have the value 1 for a graph G if $\{i,j\}$ is an element of the edge set $E(G)$ of G, and 0 otherwise. Recall that $X_n = (V(X_n), E(X_n))$ is a random graph with $V(X_n) = [n]$; let M_n be the corresponding (random) adjacency matrix. Expressed in graph theoretical terms, part (a) of the following result shows that the Doob-Martin convergence associated with the uniform attachment graphs is the same as pointwise convergence of the adjacency matrices M_n. The limit M may be regarded as the adjacency matrix of the limit graph $X_\infty = (V(X_\infty), E(X_\infty))$, with vertex set $V(X_\infty) = \mathbb{N}$.

Theorem 1 (a) *The Doob-Martin boundary of the uniform attachment process X consists of the set $\{0,1\}^\Delta$, where convergence of a sequence $(G_n)_{n \in \mathbb{N}}$ of graphs $G_n \in \mathbb{G}[n]$ to a limit $M \in \{0,1\}^\Delta$ means that the edge indicators $1_{\{i,j\}}(G_n)$ converge to $M(i,j)$ as $n \to \infty$, for each $\{i,j\} \in \Delta$.*

(b) *With probability 1, X_∞ is equal to $M \equiv 1$.*

Proof: We compute the Martin kernel. Let $G_m \in \mathbb{G}[m]$, $G_n \in \mathbb{G}[n]$, $m < n$, be such that $E(G_m) \subset E(G_n)$. For any $\{i,j\} \in E(G_n) \setminus E(G_m)$ and any $l \in \{m+1, \dots, n\}$ let $q_{i,j,l}$ be the probability that this edge appears in the X-sequence from time l onwards. Clearly, if $i < j \le m$, and with the understanding that an empty product has the value 1,

$$P\big(\{i,j\} \in E(X_n) \big| X_m = G_m\big) = \sum_{l=m+1}^{n} q_{i,j,l}$$

$$= \sum_{l=m+1}^{n} \left(\prod_{k=m+1}^{l-1} \left(1 - \frac{1}{k}\right) \right) \frac{1}{l}$$

$$= \sum_{l=m+1}^{n} \frac{m}{l(l-1)} \quad = \quad 1 - \frac{m}{n}.$$

If $i < j$ and $j > m$,

$$P\big(\{i,j\} \in E(X_n)\big|X_m = G_m\big) \;=\; \sum_{l=j}^{n} q_{i,j,l} \;=\; 1 - \frac{j-1}{n}.$$

Let $E_j(G) := \{1 \le i < j : \{i,j\} \in E(G)\}$ and $e_j(G) := \#E_j(G)$. In order to go from G_m to G_n the edges in $E_j(G_n) \setminus E_j(G_m)$ with $j = 2, \ldots, m$ and those in $E_j(G_n)$ with $j = m+1, \ldots, n$ have to enter the graph at some time $l \in \{m+1, \ldots, n\}$, and the edges $\{i,j\}, 1 \le i < j$, not in $E_j(G_n), j = 2, \ldots, n$, must remain unchosen. Hence, by independence,

$$P(X_n = G_n | X_m = G_m) \;=\; \prod_{j=2}^{m} \Big(1 - \frac{m}{n}\Big)^{e_j(G_n) - e_j(G_m)} \Big(\frac{m}{n}\Big)^{j-1-e_j(G_n)}$$

$$\cdot \prod_{j=m+1}^{n} \Big(1 - \frac{j-1}{n}\Big)^{e_j(G_n)} \Big(\frac{j-1}{n}\Big)^{j-1-e_j(G_n)}.$$

Using this with $m = 1$ we get

$$P(X_n = G_n) \;=\; \prod_{j=2}^{n} \Big(1 - \frac{j-1}{n}\Big)^{e_j(G_n)} \Big(\frac{j-1}{n}\Big)^{j-1-e_j(G_n)}.$$

Taking ratios we arrive at

$$K(G_m, G_n) \;=\; K_1(G_m, G_n) \cdot K_2(G_m, G_n) \cdot K_3(G_m, G_n)$$

with

$$K_1(G_m, G_n) := \prod_{j=2}^{m} \Big(1 - \frac{m}{n}\Big)^{-e_j(G_m)}, \quad K_2(G_m, G_n) := \prod_{j=2}^{m} \Big(1 - \frac{m+1-j}{n+1-j}\Big)^{e_j(G_n)},$$

and

$$K_3(G_m, G_n) := \prod_{j=2}^{m} \Big(\frac{j-1}{m}\Big)^{e_j(G_n)+1-j}.$$

Because of $e_j(G) \le j - 1$ for all $G \in \mathbb{G}$, the first two of these factors will converge as $n \to \infty$ with m fixed, and the respective limits will always be 1.

Now let $F_{i,m}, 1 \le i < m$, be the graph with node set $[m]$ and a single edge $\{i,m\}$. If this edge appears in G_n, then

$$K_3(F_{i,m}, G_n) \;=\; \prod_{j=2}^{m} \Big(\frac{j-1}{m}\Big)^{e_j(G_n)+1-j} \;\ge\; 1$$

in view of $e_j(G_n) \leq j - 1$. As $K(F_{i,m}, G_n) = 0$ if $\{i, m\} \notin E(G_n)$ this means that Doob-Martin convergence implies that the limits

$$M(i, m) := \lim_{n \to \infty} 1_{\{i,m\}}(G_n) \in \{0, 1\} \tag{12}$$

exist for all $\{i, m\} \in \Delta$. On the other hand, from (12) we obtain the existence of the limits

$$\lim_{n \to \infty} e_j(G_n) \in \{0, \ldots, j - 1\} \tag{13}$$

for all $j \in \mathbb{N}$, $j \geq 2$, which in turn implies the convergence of $K_3(G_m, G_n)$. Taken together this characterizes Doob-Martin convergence as stated in part (a). In (12) and (13) convergence means that the sequence elements do not change from some index n onwards.

For the proof of part (b) let $\tau_{ij} := \inf\{n \in \mathbb{N} : \{i, j\} \in E(X_n)\}$ be the entry time of the edge $\{i, j\}$; we need to show that $P(\tau_{ij} < \infty) = 1$ for all $\{i, j\} \in \Delta$. This, however, is an easy consequence of the construction of X as we have, for $n \geq j$,

$$P(\tau_{ij} > n) = \prod_{l=j}^{n} P(\tau_{ij} > l \,|\, \tau_{ij} \geq l) = \prod_{l=j}^{n} \left(1 - \frac{1}{l+1}\right) = \frac{j}{n+1}.$$

$\square$

As a consequence of part (b) of the theorem, the tail σ-field of the uniform attachment process is trivial. Further, using (6), the chain conditioned on some limit value $M \in \{0, 1\}^\Delta$ can easily be described as follows: We proceed as before, but only edges $\{i, j\} \in \Delta$ with $M(i, j) = 1$ are allowed to enter.

Remark (a) An embedding interpretation as in (10) and (11) of the topology in Theorem 1 results if we identify a graph with the values of the edge indicators,

$$G \mapsto \big(\{i, j\} \mapsto 1_{\{i,j\}}(G)\big), \quad G \in \mathbb{G}. \tag{14}$$

Here the boundary is the full function space $\{0, 1\}^\Delta$, which is usually not the case.

(b) The graph sequence generated by the uniform attachment model converges in the sampling topology too, and in fact to what is arguably a more interesting limit; see (Lovász, 2012, Proposition 11.40). However, this 'more global' topology does not capture the tail information. To be specific, consider a random variable ξ with values in $\mathbb{N}$ and define a random element α of the boundary by $\alpha(i, j) = 0$ if $j = \xi$ and $\alpha(i, j) = 1$ otherwise, and let Y be the corresponding h-transform. In Y the random node ξ remains isolated forever. Then $\mathcal{T}(Y) =_{\text{a.s.}} \sigma(\xi)$, which means that some randomness persists. The 'local' topology in Theorem 1 detects this, whereas from the global point of view Y and the original chain X are asymptotically indistinguishable.

(c) Whereas (14) is an embedding in the strict sense of being one-to-one, (10) is not: If G and G' are of the same isomorphism type, then the functions $\rho(\cdot, G)$ and $\rho(\cdot, G')$ coincide; see also Theorem 5.29 and its proof in Lovász (2012). $\triangleleft$

The second model that we consider is perhaps the most famous of all random discrete structures: To obtain the Erdős-Rényi-Gilbert graph (Lovász, 2012, p.8) or binomial random graph (Janson et al., 2000, p.2) Y_n with node set $[n]$ and parameter $\theta \in [0, 1]$ we include each of the $\binom{n}{2}$ possible edges with probability θ, independently of each other. In contrast to the uniform attachment model discussed above the variables Y_n are now defined for each n separately, with

$$P(Y_n = G_n) \;=\; \theta^{e(G_n)}(1 - \theta)^{\binom{n}{2} - e(G_n)} \quad \text{for all } G_n \in \mathbb{G}[n], \tag{15}$$

but there is a well-known and canonical method to combine these into a Markov chain $Y = (Y_n)_{n \in \mathbb{N}}$: In order to move from Y_n to Y_{n+1} we add the node $n + 1$ and then, independently of each other, each of the edges $\{i, n + 1\}$, $1 \le i \le n$, with probability θ.

A moment's thought reveals that, in this model, none of the randomness will ever go away, which is the other extreme as compared with tail triviality. Indeed, it is possible to reconstruct the complete sequence $G_1, \ldots, G_n$ from its last element G_n, which implies that $\mathcal{T}(Y) = \sigma(Y)$. Roughly, for such chains with perfect memory 'the sequence is the limit'; see also (Evans et al., 2012, Section 9). In order to formalize this, let $\Psi_m^n(G_n)$ be the graph in $\mathbb{G}[m]$ that $G_n \in \mathbb{G}[n]$ induces on $[m]$, that is, we delete the nodes $m + 1, \ldots, n$ and the incident edges. This defines a family $\{\Psi_m^n : 1 \le m \le n < \infty\}$ of functions $\Psi_m^n : \mathbb{G}[n] \to \mathbb{G}[m]$ that is consistent in the sense that $\Psi_l^m \circ \Psi_m^n = \Psi_l^n$ whenever $l \le m \le n$. Let $\mathbb{G}_{\text{proj}} \subset \mathbb{G}[1] \times \mathbb{G}[2] \times \mathbb{G}[3] \times \cdots$ be the set of all sequences $(G_n)_{n \in \mathbb{N}}$ with the properties that $G_n \in \mathbb{G}[n]$ for all $n \in \mathbb{N}$, and $\Psi_m^n(G_n) = G_m$ for all $m, n \in \mathbb{N}$ with $m < n$. This set is known as the projective (or inverse) limit associated with the sequence $(\mathbb{G}[n])_{n \in \mathbb{N}}$ and the family $\{\Psi_m^n : 1 \le m \le n < \infty\}$. In the set of sequences, we regard a sequence (of sequences) as convergent if the respective elements at any particular position $l \in \mathbb{N}$ 'freeze', i.e. converge in the discrete topology on $\mathbb{G}[l]$. With the discrete topology the individual components are compact in view of $\#\mathbb{G}[l] < \infty$, so that their (infinite) product is compact. The set $\mathbb{G}_{\text{proj}}$ is closed therein, hence compact too.

There is a slightly different point of view that connects this abstract procedure to the material in the next section and that is also useful for the description of probability measures on the projective limit: The transition graph of a perfect memory chain on $\mathbb{G}$ is a rooted and locally finite tree, with root G_1 and directed edges (G_n, G_{n+1}), $G_n = \Psi_n^{n+1}(G_{n+1})$. The projective limit then coincides with the boundary of the ends compactification of the transition tree. For a node G of this tree with k vertices let

$$A_G := \big\{ (G_n)_{n \in \mathbb{N}} \in \mathbb{G}_{\text{proj}} : G_k = G \big\}$$

be the set paths through G. It is easy to see that a probability measure μ on the ends compactification is completely specified by the values $\mu(A_G)$, $G \in \mathbb{G}$.

Theorem 2 (a) *The boundary of the Doob-Martin compactification of $\mathbb{G}$ with respect to Y is given by the projective limit $\mathbb{G}_{\text{proj}}$.*

(b) *The distribution μ of Y_∞ is given by $\mu(A_G) = P(Y_m = G)$ if $G \in \mathbb{G}[m]$.*

Proof: Again, we look at the Martin kernel: If G_m is on the unique path from G_1 to G_n then $P(Y_n = G_n, Y_m = G_m) = P(Y_n = G_n)$ so that $K(G_m, G_n) = 1/P(Y_m = G_m)$, and $K(G_m, G_n) = 0$ otherwise. From this part (a) of the theorem follows easily. For (b) we note that a boundary point $(G_n)_{n \in \mathbb{N}}$ is in A_G if and only if $G_m = G$. □

In this situation conditioning on a limit value leads to a deterministic motion along the sequence of graphs that represents the limit. Part (b) and (15) imply that the limit distribution is diffuse. In particular, the tail σ-field is not trivial, as we have already noted before.

The perfect memory property is a consequence of the labelling of the nodes in the order of their appearance and the fact that in the step from n to $n+1$ only edges incident to $n+1$ are added. Uniform attachment graphs do not have this property; for example, if $\{1,2\} \in E(X_3)$ then it is not clear whether this edge has been added at time 2 or 3.

We now show that in the perfect memory case random relabelling may lead to a more interesting topology. We recall that the group $\mathbb{S}[n]$ of permutations of $[n]$ acts on $\mathbb{G}[n]$, meaning that each $\pi_n \in \mathbb{S}[n]$ defines a function on and to $\mathbb{G}[n]$ by mapping $G = ([n], E(G))$ to

$$\pi_n(G) := \big([n], \big\{\{\pi_n(i), \pi_n(j)\} : \{i,j\} \in E(G)\big\}\big). \tag{16}$$

Now let Π_n, $n \in \mathbb{N}$, be a sequence of independent random variables, with Π_n uniformly distributed on $\mathbb{S}[n]$. We define $X = (X_n)_{n \in \mathbb{N}}$ inductively by $X_1 \equiv G_1$, $X_{n+1} = \Pi_{n+1}(\tilde{X}_{n+1})$, where $\tilde{X}_{n+1}$ is constructed from X_n as in the chain Y above: $V(\tilde{X}_{n+1}) = [n+1]$ and $E(\tilde{X}_{n+1})$ is obtained from $E(X_n)$ by adding each of the edges $\{i, n+1\}$, $i \in [n]$, independently with probability p. In view of the fact that the transition from X_n to X_{n+1} only involves X_n and quantities that are independent of X_n the process $X = (X_n)_{n \in \mathbb{N}}$ is again a Markov chain and it continues to be adapted to $\mathbb{G}$. Also, the distribution (15) is invariant under $\mathbb{S}[n]$ as the action (16) does not change the number of edges. This implies that X_n and the variable Y_n from the perfect memory version without random relabelling have the same distribution (but they will in general not be equal). Below, we will refer to the sequence X as the Erdős-Rényi chain with parameter θ.

The following result shows that the Doob-Martin compactification for this model leads to the sampling topology mentioned at the beginning of this section; see (8) and (9).

Theorem 3 *Let X be the Erdős-Rényi chain with parameter θ, $0 < \theta < 1$.*

(a) A sequence $(G_n)_{n \in \mathbb{N}}$ with $G_n \in \mathbb{G}[n]$ for all $n \in \mathbb{N}$ converges in the Doob-Martin compactification associated with X if and only if the sequences $\big(\rho(H, G_n)\big)_{n \in \mathbb{N}}$ converge for every fixed $H \in \mathbb{G}$.

(b) With probability 1, X_∞ is equal to the function

$$H \mapsto \theta^{e(H)} (1-\theta)^{\binom{v(H)}{2} - e(H)}, \quad H \in \mathbb{G}.$$

Proof: (a) Let $G_m \in \mathbb{G}[m]$, $G_n \in \mathbb{G}[n]$ with $n > m$ be given. The random relabellings in steps $m+1, \ldots, n$ move the nodes $1, \ldots, m$ of G_m to different positions in $[n]$. This defines a random injective function $\Phi : [m] \to [n]$, where all $n!/(n-m)!$ possible values ϕ of Φ are equally likely. For $\phi \notin T(G_m, G_n)$, with T as in (8), we clearly have $P(X_n = G_n | X_m = G_m, \Phi = \phi) = 0$, whereas for $\phi \in T(G_m, G_n)$,

$$P(X_n = G_n | X_m = G_m, \Phi = \phi) = \theta^{e(G_n) - e(G_m)} (1-\theta)^{\binom{n}{2} - e(G_n) - \binom{m}{2} + e(G_m)}. \tag{17}$$

Note that the right hand side does not depend on ϕ. Using (15) a decomposition with respect to the value of Φ now gives

$$K(G_m, G_n) = \frac{\sum_{\phi \in T(G_m, G_n)} P(X_n = G_n | X_m = G_m, \Phi = \phi) \, P(\Phi = \phi | X_m = G_m)}{P(X_n = G_n)}$$

$$= \ \frac{(n-m)! \ t(G_m, G_n)}{n! \ \theta^{e(G_m)}(1-\theta)^{\binom{m}{2}-e(G_m)}} \ = \ \frac{\rho(G_m, G_n)}{P(X_m = G_m)}.$$

Clearly, this converges for fixed G_m as $n \to \infty$ if and only if $\rho(G_m, G_n)$ does.

(b) This follows from the sampling interpretation of ρ and the independence of the edge indicators. □

Again, part (b) implies that the tail σ-field of X is trivial. Further, the Erdős-Rényi chains with different parameter values are easily seen to be h-transforms of each other.

Theorem 3 identifies the Doob-Martin boundary of the Erdős-Rényi chain as a subset of the set of all functions $G_\infty : \mathbb{G} \to [0,1]$. For a concise description of this subset, by 'graphons', we refer the reader to Lovász (2012).

4 Search trees

The nodes of the complete binary tree $\mathbb{V} = \{0,1\}^\star$ are finite 0-1 sequences (or words) $u = (u_1, \ldots, u_k)$; we write $u0 = (u_1, \ldots, u_k, 0)$ and $u1 = (u_1, \ldots, u_k, 1)$ for the left and right child of u respectively and, if $k = |u| > 0$, $\bar{u} = (u_1, \ldots, u_{k-1})$ for its direct ancestor. The concatenation of $u = (u_1, \ldots, u_k)$ and $v = (v_1, \ldots, v_l)$ is given by $u + v = (u_1, \ldots, u_k, v_1, \ldots, v_l)$. We further write $\partial \mathbb{V} = \{0,1\}^\infty$ for the ends compactification of the tree $\mathbb{V}$.

By a binary tree we mean a subset $x \subset \mathbb{V}$ that is prefix-stable or, equivalently, contains the ancestor of each of its non-root elements. If $u \notin x$, $\bar{u} \in x$, then we call u external, and we write ∂x for the set of external nodes of x. The (fringe) subtree of x rooted at u is given by $x(u) := \{v \in \mathbb{V} : u + v \in x\}$. Let $\mathbb{B}_n$ be the set of binary trees with n nodes; $\mathbb{B} := \bigcup_{n=1}^\infty \mathbb{B}_n$. Prefix stability implies that any $x \in \mathbb{B}$ can be regarded as a contiguous subset of $\mathbb{V}$ and hence be described by its *boundary function*

$$B_x : \partial \mathbb{V} \to \mathbb{N}, \quad (u_i)_{i \in \mathbb{N}} \mapsto \min\{k \in \mathbb{N} : (u_1, \ldots, u_k) \notin x\}.$$

(This seems to be the most natural term, but in view of all the other occurrences of boundaries in the present paper, 'frontier' may be a sensible alternative.)

Given a sequence $(t_n)_{n \in \mathbb{N}}$ of pairwise distinct real numbers the binary search tree (BST) algorithm generates a sequence $(x_n)_{n \in \mathbb{N}}$ of labelled binary trees as follows: The first value is stored at the root node; given x_n the next value t_{n+1} is stored at the first empty node found when travelling through x_n, moving from u to $u0$ if the new value is smaller that the label of an occupied node and to $u1$ otherwise. This is one of the standard algorithm for searching and also arises in the context of sorting; see Knuth (1973), Mahmoud (1992) and Drmota (2009). Suppose that the t_i's are realizations of independent random variables η_i, $i \in \mathbb{N}$, with the same continuous distribution. Then the random binary trees X_n obtained for $\eta_1, \ldots, \eta_n$, $n \in \mathbb{N}$, can be collected into a Markov chain $X = (X_n)_{n \in \mathbb{N}}$ with a simple transition structure: X_1 is the tree that consists of the root node $\emptyset$ only, and in the transition from X_n to X_{n+1} one of the $n+1$ external nodes of X_n is chosen uniformly at random and incorporated into the tree. The BST chain has $\mathbb{B}$ as its state space, and $P(X_n \in \mathbb{B}_n) = 1$ for all $n \in \mathbb{N}$.

The Doob-Martin compactification of $\mathbb{B}$ with respect to X and the distribution of the limit X_∞ were obtained by Evans et al. (2012) and can be described as follows: $\partial \mathbb{B}$ is the set of probability measures μ on $\partial \mathbb{V}$. Convergence to $\mu \in \partial \mathbb{B}$ of a sequence $(x_n)_{n \in \mathbb{N}} \subset \mathbb{B}$ means that $a_n := \#x_n \to \infty$ and that the relative number $\#x_n(u)/a_n$ of nodes in the subtree rooted at $u \in \mathbb{V}$ converges to $\mu(A_u)$ for all $u \in \mathbb{V}$,

where A_u consists of all infinite 0-1 sequences with prefix u. As with the transition tree in the previous section, the values $\mu(A_u)$, $u \in \mathbb{V}$, determine μ.

We have $P(X_\infty(A_u) > 0) = 1$ for all $u \in \mathbb{V}$. The distribution of X_∞ is a probability measure on $\partial \mathbb{B}$, hence on the set of probability measures on $\partial \mathbb{V}$, where the latter is endowed with the σ-field generated by the projections $\mu \mapsto \mu(A)$, A a Borel subset of $\partial \mathbb{V}$. This distribution has the (characterizing) property that the random variables

$$\xi_u := \frac{X_\infty(A_{u0})}{X_\infty(A_u)}, \quad u \in \mathbb{V}, \tag{18}$$

are independent and uniformly distributed on the unit interval. This in turn implies that $X_\infty(A_u)$ can be written as the product of independent, identically distributed random variables, a fact that we will use repeatedly below.

In the present section we apply this to the asymptotics of the random functions B_{X_n}, $n \in \mathbb{N}$. The idea of describing randomly growing sets by their boundary appears in connection with models now known under the acronym 'IDLA' (internal diffusion limited aggregation). This subject area was initiated by Diaconis and Fulton (1991), an early important contribution is Lawler et al. (1992). Both papers deal with integer lattices, but the basic model has since then been applied to various other infinite discrete background sets, for example to the 'comb' by Huss and Sava (2012). BST chains may be seen as an IDLA variant on the background set $\mathbb{V}$, where the exploration process is a reinforced random walk in the sense that the probabilities of moving from u to $u0$ and $u1$ respectively depend on the number of previous particles that have travelled along the respective edge. The BST boundary functions have earlier been investigated under the name of 'silhouette' in Grübel (2005, 2009), where they were regarded as functions on the unit interval via

$$\partial \mathbb{V} \ni u = (u_k)_{k \in \mathbb{N}} \ \mapsto \ \beta(u) := \frac{1}{2} \ + \ \sum_{k=1}^{\infty} \frac{2u_k - 1}{2^{k+1}} \in [0, 1] \tag{19}$$

(binary rationals do not matter as X_∞ has no atoms). Figure 1 shows the boundary functions of X_n for various n, with pseudorandom data, where (19) has been used to display B_{X_n} as a function on $[0, 1]$.

We begin with two real-valued functionals of the boundary functions. First we consider the growth of the trees along a fixed path through the infinite binary tree.

Theorem 4 *Let $u \in \partial \mathbb{V}$ be fixed. Then the tail σ-field of the sequence $(B_{X_n}(u))_{n \in \mathbb{N}}$ is P-trivial.*

A proof can easily be obtained on using the well-known connection to records: The BST dynamics imply that $(B_{X_n}(u))_{n \in \mathbb{N}}$ is identical in distribution to the sequence $(S_n)_{n \in \mathbb{N}}$, $S_n = \sum_{k=1}^{n} \zeta_k$, of partial sums of independent random variables ζ_k with $P(\zeta_k = 1) = 1 - P(\zeta_k = 0) = 1/k$, $k \in \mathbb{N}$, which also appears when counting records in random samples. It follows that $((n, S_n))_{n \in \mathbb{N}}$ is a Markov chain with state space $\mathbb{F} = \{(n, i) : n \in \mathbb{N}, \ i \in [n]\}$ and transition probabilities

$$p\big((n, k), (n+1, k+1)\big) \ = \ 1 - p\big((n, k), (n+1, k)\big) \ = \ \frac{1}{n+1}.$$

The structural similarity to the Pólya urn mentioned in Section 1 should be apparent. In the records chain, a sequence $((n, k_n))_{n \in \mathbb{N}}$ of states converges in the sense introduced in Section 2 if and only if

$$\lim_{n \to \infty} \frac{k_n}{\log n} = \alpha \in [0, \infty],$$

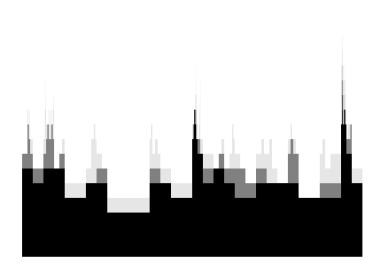**Fig. 1:** The subgraph of B_{X_n}, for $n = 50$ (black), $n = 100$ (gray) and $n = 200$ (light gray).

which leads to $\bar{\mathbb{F}} = \mathbb{F} \cup [0, \infty]$. This can be proved by 'path-counting', the asymptotics of unsigned Stirling numbers of the first kind, and an interesting monotonicity argument; see Gnedin and Pitman (2005) and the references given there. In the compactification, S_n tends to the constant value 1, which implies triviality of the tail σ-field as explained at the end of Section 2.

Let

$$H(0) = 0, \quad H(n) = \sum_{k=1}^{n} 1/k \ \text{ for } n \in \mathbb{N},$$

be the harmonic numbers. From the representation of $B_{X_n}(u)$ as a sum of independent Bernoulli variables we obtain the expected value $\mathbb{E}B_{X_n}(u) = H(n)$ and, using $H(n) \sim \log n$, the distributional convergence

$$\frac{1}{\sqrt{\log n}}\big(B_{X_n}(u) - \log n\big) \ \to_{\text{distr}} \ Z \ \text{ as } n \to \infty, \tag{20}$$

where Z has a standard normal distribution. However, by Theorem 4, there is no transformation of the random variables $B_{X_n}(u)$, $u \in \partial\mathbb{V}$ fixed, that leads to strong convergence with a non-degenerate limit.

For the second functional we integrate the boundary functions with respect to the measure λ on $\partial\mathbb{V}$ given by $\lambda(A_u) = 2^{-|u|}$. This is the unique normalized Haar measure if we regard the set of infinite 0-1 sequences as a compact group under the pointwise addition modulo 2. Recall from (18) that $\xi_u = X_\infty(A_{u0})/X_\infty(A_u)$, and let

$$C(t) \ = \ 1 + \frac{1}{2}\big(\log(t) + \log(1 - t)\big), \quad 0 < t < 1.$$

Lemma 5 *The random variables* $L_{\infty,k} := \sum_{|u|<k} 2^{-|u|}C(\xi_u)$, $k \in \mathbb{N}$, *converge almost surely and in* L^2 *as* $k \to \infty$.

Proof: A straightforward calculation shows that $\mathbb{E}C(\xi_u) = 0$ and $\mathrm{var}(C(\xi_u)) = 1 - (\pi^2/12)$. In particular, using independence of the ξ_u's,

$$\mathrm{var}\Big(\sum_{|u|=k} 2^{-|u|} C(\xi_u) \Big) = 2^{-k}\Big(1 - \frac{\pi^2}{12} \Big),$$

which implies that $\big(\sum_{|u|\leq k} 2^{-|u|} C(\xi_u), \sigma(\{\xi_u \,:\, |u| \leq k\})\big)_{k\in\mathbb{N}}$ is an L^2-bounded martingale, so that the corresponding limit theorem can be used. □

We write $L_\infty = \sum_{v\in\mathbb{V}} 2^{-|v|} C(\xi_v)$ for the limit of $L_{\infty,k}$ as $k \to \infty$. In this series we do not have absolute convergence: For $k \in \mathbb{N}$ fixed the mean of the random variable $\sum_{|u|=k} 2^{-|u|} |C(\xi_u)|$ is a positive value that does not depend on k.

Theorem 6 *Let* $L_n := \int B_{X_n}\, d\lambda$. *Then* $L_n - H(n) \to L_\infty$ *with probability 1 as* $n \to \infty$.

Proof: Let $\mathcal{F}_n$ be the σ-field generated by the first n variables $X_1, \ldots, X_n$ of the BST chain. We first show that

$$Z_n := \mathbb{E}[L_\infty | \mathcal{F}_n] = \sum_{u\in X_n} 2^{-|u|} - H(n). \tag{21}$$

By (18), the family $\{\xi_u \,:\, u \in \mathbb{V}\}$ is a function of the limit X_∞ of the BST sequence. Hence, using the Markov property of the latter, $\mathbb{E}[C(\xi_u)|\mathcal{F}_n] = \mathbb{E}[C(\xi_u)|X_n]$. From Proposition 2 and Lemma 4 in Grübel (2014) it is known that the distribution of ξ_u given $\#X_n(u0) = i$ and $\#X_n(u1) = j$ is the beta distribution with parameters $i+1$ and $j+1$, and that for a random variable ζ with this distribution we have $\mathbb{E}\log(\zeta) = H(i) - H(i+j+1)$. In particular, $\mathbb{E}[C(\xi_u)|X_n] = 0$ on $\{u \notin X_n\}$. For $u \in X_n$, using $\#X_n(u) = \#X_n(u0) + \#X_n(u1) + 1$ we obtain

$$\mathbb{E}[C(\xi_u)|X_n] = 1 + \frac{1}{2}\big(H(\#X_n(u0)) + H(\#X_n(u1))\big) - H(\#X_n(u)),$$

which may be written as

$$2^{-|u|}\mathbb{E}[C(\xi_u)|\mathcal{F}_n] = 2^{-|u|} - \big(\psi_n(u) - \psi_n(u0) - \psi_n(u1)\big)$$

with $\psi_n(u) := 2^{-|u|} H\big(\#X_n(u)\big)$. A summation by parts, see (Grübel, 2014, Lemma 5), now gives

$$\sum_{u\in X_n} \big(\psi_n(u) - \psi_n(u0) - \psi_n(u1)\big) = \psi_n(\emptyset) - \sum_{u\in\partial X_n} \psi_n(u).$$

Obviously, $\psi_n(u) = 0$ for $u \in \partial X_n$, which completes the proof of (21).

The integral defining L_n may be rewritten as follows,

$$\int B_{X_n}\, d\lambda = \sum_{k=1}^\infty k\, 2^{-k} \#\{u \in \partial X_n : |u| = k\} = \sum_{u\in\partial X_n} |u|\, 2^{-|u|} = \sum_{u\in X_n} 2^{-|u|}, \tag{22}$$

where the last equality can easily be proved by induction. Combining this with (21) and the convergence of L^2-bounded martingales we obtain the assertion. □

As a sum of independent and non-degenerate random variables the limit L_∞ is not almost surely constant; in particular, the tail σ-field of the L-sequence is not P-trivial.

A strong limit theorem for $L_n - H(n)$ has already been obtained in (Grübel, 2009) by proving directly that $(L_n - H(n), \mathcal{F}_n)_{n \in \mathbb{N}}$ is an L^2-bounded martingale. Our approach here differs insofar as it replaces the search for a suitable martingale by projecting the limit on the natural filtration $(\mathcal{F}_n)_{n \in \mathbb{N}}$ of the Markov chain, and it also provides a representation of the limit in terms of X_∞. The representation in turn leads to an interpretation of the limit as a distance from X_∞ to the Haar measure λ: Let $\mathcal{G}_k$ be the σ-field on $\partial \mathbb{V}$ generated by the sets A_u with $|u| = k$. The Kullback-Leibler divergence $\mathrm{KL}(\mu_1, \mu_2)$ of two measures μ_1 and μ_2 on a measure space $(\Omega, \mathcal{F})$, with $\mathcal{F}$ generated by a partition $A_1, \ldots, A_l$ of Ω, is given by

$$\mathrm{KL}(\mu_1, \mu_2) := \sum_{j=1}^{l} \mu_1(A_j) \log\left(\frac{\mu_2(A_j)}{\mu_1(A_j)}\right).$$

We write $\mu|_\mathcal{G}$ for the restriction of the measure μ to a sub-σ-field $\mathcal{G}$ of its domain $\mathcal{F}$.

Theorem 7 *With probability 1, $L_\infty = \lim_{k \to \infty} \mathrm{KL}(X_\infty|_{\mathcal{G}_k}, \lambda|_{\mathcal{G}_k})$.*

Proof: If we restrict the sum in the definition of L_∞ to the nodes of depth less than k then we obtain with $\phi(u) := 2^{-|u|} \log(X_\infty(A_u))$, using (18) and a summation by parts as in the proof of Theorem 6,

$$
\begin{aligned}
\sum_{|u|<k} 2^{-|u|} C(\xi_u) &= \sum_{|u|<k} 2^{-|u|}\left(1 + 2^{-1}\left(\log \frac{X_\infty(A_{u0})}{X_\infty(A_u)} + \log \frac{X_\infty(A_{u1})}{X_\infty(A_u)}\right)\right) \\
&= \sum_{|u|<k} 2^{-|u|} + \sum_{|u|<k}\left(\phi(u0) + \phi(u1) - \phi(u)\right) \\
&= k + \sum_{|u|=k} \phi(u) - \phi(\emptyset) \\
&= \sum_{|u|=k} k 2^{-|u|} + \sum_{|u|=k} 2^{-k} \log(X_\infty(A_u)) \quad = \mathrm{KL}(X_\infty|_{\mathcal{G}_k}, \lambda|_{\mathcal{G}_k}).
\end{aligned}
$$

$\square$

It is tempting to think of the limit as the Kullback-Leibler divergence of X_∞ and λ. Note, however, that the density $Z_k : \mathbb{V} \to [0, \infty)$ of $X_\infty|_{\mathcal{G}_k}$ with respect to $\lambda|_{\mathcal{G}_k}$ is given by

$$Z_k(u) = \frac{X_\infty(A_u)}{2^{-|u|}} = \prod_{j=1}^{k}(2\tilde{\xi}_j),$$

where $\tilde{\xi}_j$ is equal to either ξ_v or $1 - \xi_v$ with the ξ-variables as in (18), depending on the value of the jth entry of the sequence u, and with v the corresponding length $j - 1$ prefix of u. From this representation as a product of independent, identically distributed and non-degenerate random variables with mean 1 it follows that

$$P\left(\liminf_{k \to \infty} Z_k(u) = 0\right) = P\left(\limsup_{k \to \infty} Z_k(u) = \infty\right) = 1.$$

In particular, X_∞ and λ are mutually singular with probability 1. Clearly, some cancellation occurs in the sum defining L_∞, due to the fact that X_∞ is a random measure. xs
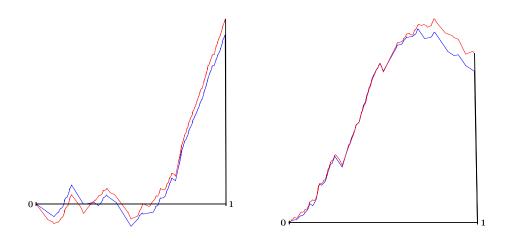

Fig. 2: Two values of Y_n, with $n = 500$ (blue) and $n = 1000$ (red).

We return to the boundary functions. From Theorem 4 it is clear that we cannot expect these to converge pointwise; see also Figure 1. Further, the asymptotic normality in (20) shows that, at a specific point, the functions increase roughly as $\log n$ but that there are fluctuations of the order $\sqrt{\log n}$. Hence, apart from shifting, some smoothing is needed, as has already been noticed in (Grübel, 2009). We first adapt the smoothing procedure introduced in (Grübel, 2009) to $\partial\mathbb{V}$ instead of $[0, 1]$ as domain of the random functions. For this we define a total order on $\partial\mathbb{V}$ by setting $u \prec v$ for $u = (u_k)_{k\in\mathbb{N}}$, $v = (v_k)_{k\in\mathbb{N}}$, with $u \ne v$, if and only if $u_l = 0$ and $v_l = 1$ in the first position l where the sequences differ. With the IDLA connection in mind we are now led to normalizing and smoothing B_{X_n} to Y_n given by

$$Y_n(u) := \int_{v \prec u} \left(B_{X_n}(v) - H(n) \right) \lambda(dv)$$

(recall that the nth harmonic number is the expectation of $B_{X_n}(v)$ for each $v \in \partial\mathbb{V}$). It is easy to deduce from (Grübel, 2009, Theorem 8) that Y_n converges in distribution to a process with continuous paths. However, as the Y_n's are all defined on the same probability space, it makes sense to ask whether these variables themselves converge.

Figure 2 shows the values of $Y_n(\omega)$ for two ω's, with $n = 500$ and $n = 1000$ respectively, where instead of two such ω's in the left and the right part of the figure two separate streams of numbers were used that the present author regards as plausible substitutes for truly random numbers: The two sequences were generated from alternating blocks of ten digits in the decimal expansion of $\pi - 3$, so that the left stream begins with $t_1 = 0.1415926535$, $t_2 = 0.2643383279$, ..., whereas the right stream has $t_1 = 0.8979323846$, $t_2 = 0.5028841971$ and so on. As in Figure 1, $\partial\mathbb{V}$ is mapped to $[0, 1]$ by the function β defined in (19) in order to be able to draw the functions.

The figure supports the conjecture that the random functions Y_n themselves converge, and that the limit is not a fixed function. Incidentally, it also demonstrates the influence of the first few values on the output

of the BST algorithm: The long-term proportion $\#X_n((0))/n$ of nodes in the left subtree is equal to the value of the first input variable, for example, and for the above π-data the t_1-values are quite different.

The theorem below confirms this conjecture. The theorem also provides a representation of the limit process in terms of the Doob-Martin limit X_∞ of the BST sequence and, in fact, its proof is closely connected to this representation.

We need to specify what convergence of the random functions Y_n means. For this, we define a metric d on $\partial\mathbb{V}$ by $d(u,v) = 2^{-l+1}$ for $u,v \in \partial\mathbb{V}$, $u \neq v$, where l is the first coordinate in which the two sequences differ as in the definition of the total order on $\partial\mathbb{V}$; also, $l-1 = |u \wedge v|$ where $u \wedge v$ denotes the longest common prefix (last common ancestor) of u and v. This turns $\partial\mathbb{V}$ into a compact metric space; we write $C(\partial\mathbb{V})$ for the set of continuous functions on $(\partial\mathbb{V}, d)$. Endowed with the supremum norm, $\|f\|_\infty = \sup_{u \in \mathbb{V}} |f(u)|$, $C(\partial\mathbb{V})$ becomes a Banach space.

Further, for $u = (u_k)_{k \in \mathbb{N}} \in \partial\mathbb{V}$ let $K(u) := \{k \in \mathbb{N} : u_k = 1\}$, and $u(k) := (u_1, \ldots, u_{k-1}, 0)$ for $k \in K(u)$. Generalizing the notation introduced above in connection with the second functional of the boundary functions we write

$$L_\infty(u) := \sum_{v \geq u} 2^{-|v|} C(\xi_v), \quad u \in \mathbb{V}, \tag{23}$$

where '$\geq$' now refers to prefix order. Clearly, $L_\infty(u)$ has the same distribution as $2^{-|u|} L_\infty(\emptyset)$, and we know from Lemma 5 that $L_\infty(\emptyset) = L_\infty$ has zero mean and finite variance.

We require two auxiliary results.

Lemma 8 (a) *For each $u \in \partial\mathbb{V}$, the random variables*

$$Y'_{\infty,m}(u) := \sum_{k \in K(u), k \leq m} L_\infty(u(k)), \quad m \in \mathbb{N},$$

converge almost surely and in L^2 as $m \to \infty$.

(b) *For each $u \in \partial\mathbb{V}$, the random variables*

$$Y''_{\infty,m}(u) := \sum_{k \in K(u), k \leq m} 2^{-k} \log\big(2^k X_\infty(A_{u(k)})\big), \quad m \in \mathbb{N},$$

converge almost surely and in L^2 as $m \to \infty$.

Proof: (a) For all $m, n \in \mathbb{N}$ we have

$$\big\|Y'_{\infty,m}(u) - Y'_{\infty,n}(u)\big\|_2 = \mathbb{E}\bigg\| \sum_{\substack{k \in K(u), \\ m \wedge n < k \leq m \vee n}} L_\infty\big(u(k)\big)\bigg\|_2 \leq 2^{-(m \wedge n - 1)/2} \|L_\infty\|_2.$$

This shows that that $(Y'_{\infty,m}(u))_{m \in \mathbb{N}}$ is a Cauchy sequence in L^2 and that, with Y'_∞ the limit,

$$\mathbb{E}\big(Y'_{\infty,m} - Y'_\infty\big)^2 = O(2^{-m}).$$

In particular, using Chebyshev's inequality, we get

$$\sum_{m=1}^\infty P(|Y'_{\infty,m} - Y'_\infty| > \epsilon) < \infty \quad \text{for all } \epsilon > 0,$$

which implies almost sure convergence.

(b) It follows from (18) that

$$\log X_\infty(A_u) = \sum_{j=1}^{|u|} \log \tilde{\xi}_j,$$

with $\tilde{\xi}_1, \ldots, \tilde{\xi}_{|u|}$ independent and uniformly distributed on the unit interval; see also the discussion following Theorem 7. This leads to

$$\left\| \log\big(2^{-|u|} X_\infty(A_u)\big) \right\|_2 = O(|u|).$$

Using this bound on the L^2-norm of the individual summands we can now proceed as in the proof of part (a). $\qquad \square$

In view of Lemma 8 it makes sense to define two random functions $Y'_\infty = (Y'_\infty(u))_{u \in \partial \mathbb{V}}$ and $Y''_\infty = (Y''_\infty(u))_{u \in \partial \mathbb{V}}$ on $\partial \mathbb{V}$ by

$$Y'_\infty(u) = \sum_{k \in K(u)} L_\infty(u(k)), \quad Y''_\infty(u) = \sum_{k \in K(u)} 2^{-k} \log\big(2^k X_\infty(A_{u(k)})\big). \tag{24}$$

The random functions Y'_∞ and Y''_∞ may also be regarded as stochastic processes with time parameter $u \in \partial \mathbb{V}$.

Lemma 9 (a) *With probability 1, the processes Y'_∞ and Y''_∞ have continuous paths.*

(b) *Both processes are integrable in the sense that*

$$\mathbb{E}\|Y'_\infty\|_\infty < \infty \ \text{ and } \ \mathbb{E}\|Y''_\infty\|_\infty < \infty.$$

Proof: We consider Y'_∞ first. For the proof of continuity we adapt the well-known chaining argument, see e.g. (Kallenberg, 1997, p.35), to the present situation. Let

$$\rho_k := \max\big\{|L_\infty(v)| : v \in \mathbb{V}, \ |v| = k\big\}, \ \ k \in \mathbb{N}.$$

For nodes v on a fixed level k the variables $L_\infty(v)$ are independent. Using $\rho_k^2 \le \sum_{|v|=k} L_\infty(v)^2$ we obtain

$$\mathbb{E}\rho_k^2 \le \sum_{|v|=k} \mathbb{E}L_\infty(v)^2 = \sum_{|v|=k} \text{var}(L_\infty(v)) = 2^{-k}\text{var}(L_\infty),$$

from which it follows that

$$\mathbb{E}\Big(\sum_{k=1}^\infty 2^{k/2} \rho_k^2\Big) < \infty. \tag{25}$$

This implies that on a set A of probability 1 we have

$$\rho_k(\omega) \le C(\omega)\, 2^{-k/4} \ \text{ for all } k \in \mathbb{N}, \ \omega \in A,$$

with some $C(\omega) < \infty$ that does not depend on k. Suppose now that $u, v \in \mathbb{V}$ are such that $d(u, v) \le \epsilon$. Then the first $l = l(\epsilon) = \lceil -\log_2(\epsilon) \rceil$ entries of u and v coincide, so that their connecting path does not go below height l. With Lemma 8 and the triangle inequality we therefore get

$$\left| Y'_\infty(u)(\omega) - Y'_\infty(v)(\omega) \right| \;\le\; 2 \sum_{k=l+1}^{\infty} \rho_k(\omega) \;\le\; 13\, C(\omega)\, \epsilon^{1/4}$$

whenever $\omega \in A$. This implies that almost all paths of Y'_∞ are continuous.

For the proof of integrability we first note that $\|Y'_\infty\|_\infty \le \sum_{k=1}^{\infty} \rho_k$. Using (25) and

$$\left(\sum_{k=1}^{\infty} |a_k|\, |b_k| \right)^2 \;\le\; \left(\sum_{k=1}^{\infty} a_k^2 \right) \left(\sum_{k=1}^{\infty} b_k^2 \right)$$

with $a_k = 2^{-k/4}$ we see that the upper bound has finite mean.

As in the proof of the previous lemma, the arguments used for Y'_∞ can be transferred to the other process Y''_∞: We now put

$$\sigma_k := \max\left\{ 2^{-k} \left| \log(2^k X_\infty(A_v)) \right| \,:\, v \in \mathbb{V},\, |v| = k \right\}, \quad k \in \mathbb{N},$$

and again, we will show that these decrease rapidly enough as $k \to \infty$. However, we no longer have independence of the individual random variables in the maximum, so we need a different argument. As in (Grübel, 2014) in connection with the maximum of the probabilities $X_\infty(A_u)$, $|u| = k$, we use the connection to the branching random walks discussed by Biggins (1977). This rests upon the observation that the variables

$$Y_u := -\log X_\infty(A_u), \quad |u| = k,$$

are the positions of the members of the kth generation in a branching random walk with offspring distribution δ_2, meaning that each particle has exactly two descendants, and with

$$Z := \delta_{-\log \xi} + \delta_{-\log(1-\xi)}, \quad \mathcal{L}(\xi) = \mathrm{unif}(0, 1),$$

for the point process of the positions of the children relative to their parent. Let

$$m(\theta) := \mathbb{E}\left(\int e^{\theta t}\, Z(dt) \right) \;=\; \frac{2}{1-\theta}, \quad \theta < 1, \tag{26}$$

and let $Z_+^{(k)}(t)$ be the number of particles in generation k that are located to the right of t. The random measure $Z^{(k)}$ is the kth convolution power of Z, which leads to

$$\mathbb{E}\left(\int e^{\theta t} Z^{(k)}(dt) \right) \;=\; m(\theta)^k,$$

so that

$$P\left(\max\{Y_u : |u| = k\} \ge t \right) \;\le\; \mathbb{E}Z_+^{(k)}(t)$$
$$\le\; e^{-\theta t}\, \mathbb{E}\left(\int_{[t,\infty)} e^{\theta s} Z^{(k)}(ds) \right)$$

$$\leq\ e^{-\theta t}\,m(\theta)^k$$

for all $t \in \mathbb{R}$, $0 \leq \theta < 1$ and $k \in \mathbb{N}$. Using (26) we get, with $\theta = 1/2$,

$$\begin{aligned}
\mathbb{E}\big(\max\{Y_u : |u| = k\}\big) &= \int_0^\infty P\big(\max\{Y_u : |u| = k\} \geq t\big)\,dt \\
&\leq \Big(\frac{5}{4}\Big)^k + \int_{(5/4)^k}^\infty e^{-t/2}\,4^k\,dt \\
&= O\Big(\Big(\frac{4}{3}\Big)^k\Big),
\end{aligned}$$

so that, for some constant $C < \infty$,

$$\mathbb{E}\bigg(\sum_{k=1}^\infty \Big(\frac{5}{4}\Big)^k \sigma_k\bigg) \ \leq\ C \sum_{k=1}^\infty \Big(\frac{5}{4}\Big)^k 2^{-k}\,\mathbb{E}\big(\max\{Y_u : |u| = k\}\big)\ <\ \infty.$$

Using this instead of (25) we can now proceed as in the first part of the proof. $\qquad\square$

There is obviously room to spare in the above chaining inequalities; tightening these leads to path properties beyond continuity.

In the proof of our final result we will use infinite-dimensional martingales; see (Neveu, 1975, Chapter V-2). For this, we require separability of the Banach space $\big(C(\partial\mathbb{V}), \|\cdot\|_\infty\big)$: The sets A_u, $u \in \mathbb{V}$, are closed and open in $\partial\mathbb{V}$, so their indicator functions are continuous. Moreover, the intersection of two such sets is again of this form, and the indicator functions separate the points of $\partial\mathbb{V}$. The required separability now follows on using the Stone-Weierstraß theorem.

Theorem 10 *With probability 1, Y_n converges in $\big(C(\partial\mathbb{V}), \|\cdot\|_\infty\big)$ to $Y_\infty := Y_\infty' + Y_\infty''$ as $n \to \infty$.*

Proof: For all $u \in \mathbb{V}$, $n \in \mathbb{N}$,

$$\begin{aligned}
\int_{A_u} B_{X_n}\,d\lambda &= \sum_{v \in \partial X_n, v \geq u} |v|\,2^{-|v|} \\
&= \sum_{w \in \partial X_n(u)} (|u| + |w|)\,2^{-|u|-|w|} \\
&= 2^{-|u|}\bigg(|u| \sum_{w \in \partial X_n(u)} 2^{-|w|} + \sum_{w \in \partial X_n(u)} |w|\,2^{-|w|}\bigg) \\
&= 2^{-|u|}\bigg(|u| + \sum_{w \in X_n(u)} 2^{-|w|}\bigg) \\
&= |u|\,2^{-|u|} + \sum_{v \in X_n, v \geq u} 2^{-|v|}.
\end{aligned}$$

Further, with $L_\infty(u)$ as in (23),

$$\mathbb{E}[L_\infty(u)|\mathcal{F}_n]\ =\ \mathbb{E}\Big[\sum_{v \in \mathbb{V}, v \geq u} 2^{-|v|} C(\xi_v)\Big|\mathcal{F}_n\Big]$$

$$= \sum_{v \in X_n, v \geq u} \left(2^{-|v|} - \big(\psi_n(u) - \psi_n(u0) - \psi_n(u1)\big)\right)$$

$$= \sum_{v \in X_n(u), v \geq u} 2^{-|v|} - 2^{-|u|} H\big(\#X_n(u)\big),$$

where we have used the same notation and the same arguments as in the proof of (21). Combining these we get, for all $u \in \mathbb{V}$,

$$\int_{A_u} \big(B_{X_n} - H(n)\big) \, d\lambda = \mathbb{E}[L_\infty(u)|\mathcal{F}_n] + 2^{-|u|}\big(H(\#X_n(u)) - H(n) + |u|\big). \tag{27}$$

Now let $u = (u_k)_{k \in \mathbb{N}} \in \partial\mathbb{V}$ be such that $\#K(u) < \infty$. The integration range appearing in the definition of $Y_n(u)$ can be decomposed as follows,

$$I_u := \{v \in \partial\mathbb{V} : v \prec u\} = \sum_{k \in K(u)} A_{u(k)}.$$

With (27) we thus obtain, setting $Y'_n := \mathbb{E}\big[Y'_\infty|\mathcal{F}_n\big]$,

$$Y_n(u) = \sum_{k \in K(u)} \int_{A_u} \big(B_{X_n} - H(n)\big) d\lambda$$

$$= Y'_n(u) + \sum_{k \in K(u)} 2^{-k}\big(H(\#X_n(u(k))) - H(n) + k\big).$$

From Lemma 9 we know that Y'_∞ is a $C(\partial\mathbb{V})$-valued integrable random variable. Hence $(Y'_n, \mathcal{F}_n)_{n \in \mathbb{N}}$ is a martingale with values in the separable Banach space $\big(C(\partial\mathbb{V}), \|\cdot\|_\infty\big)$, and by (Neveu, 1975, Proposition V.2.6) Y'_n converges almost surely in this space to Y'_∞ as $n \to \infty$.

It remains to prove that, as $n \to \infty$,

$$\sum_{k \in K(u)} 2^{-k}\big(H(\#X_n(u)) - H(n) + k\big) \to Y''_\infty(u), \tag{28}$$

with probability 1 and, with both sides regarded as functions of u, in $\big(C(\partial\mathbb{V}), \|\cdot\|_\infty\big)$. For this, we first show that

$$\mathbb{E}\big[\log\big(X_\infty(A_u)\big|\mathcal{F}_n\big] = H\big(\#X_n(u)\big) - H(n) \quad \text{for all } u \in \mathbb{V}, n \in \mathbb{N}. \tag{29}$$

Clearly, for $u = \emptyset$, both sides of (29) are equal to 0. If $u = (u_1, \ldots, u_k, u_{k+1})$ with $u_{k+1} = 0$ then, from (18),

$$\log X_\infty(A_u) = \log X_\infty(A_{\bar{u}}) + \log \xi_{\bar{u}}.$$

Further, we know from the proof of Theorem 6 that

$$\mathbb{E}[\log \xi_{\bar{u}}|\mathcal{F}_n] = H\big(\#X_n(\bar{u}0)\big) - H\big(\#X_n(\bar{u})\big).$$

Hence, if (29) holds for $\bar{u}$, then so it does for u. The same arguments work in the case $u_{k+1} = 1$, with $1 - \xi_{\bar{u}}$ and $\#X_n(\bar{u}1)$ instead of $\xi_{\bar{u}}$ and $\#X_n(\bar{u}0)$ respectively. This completes the induction proof for (29).

As in the first part of the proof we now get

$$\mathbb{E}\big[Y''_\infty(u)\big|\mathcal{F}_n\big] \;=\; \sum_{k \in K(u)} 2^{-k}\Big(H\big(\#X_n(u(k))\big) - H(n) + k\Big).$$

From this (28) follows on using the infinite-dimensional martingale convergence theorem again. □

5   Comments and complements

We collect some references to related work and also put the above results into a larger perspective.

(a) The approach of the present paper is not limited to graphs and search trees but may be used quite generally in the context of combinatorial Markov chains. For an elementary introduction to such processes and their boundaries, with many examples and algorithms, see Grübel (2013) (written in simple German).

(b) In concrete cases, the results provided by a general method such as the Doob-Martin approach can often be obtained more directly, using the additional structures then present. For example, in Grübel (2014) a proof of the basic BST result from Evans et al. (2012) is given that is based on the BST algorithm; this direct approach also leads to a representation of X_∞ in terms of the input sequence. Obviously, the same applies to the theory of graph limits, but the exposition of a common structure provided by a general theory may lead to a deeper understanding of such individual cases.

(c) As seen above, the boundary theory approach may lead to strong limit theorems for discrete structures and their functionals, occasionally improving on previous results. In Grübel (2014) such an amplification from convergence in distribution to convergence of the random variables is carried out for the Wiener index of search trees, where distributional convergence had earlier been obtained by Neininger (2002) with the contraction method. In both cases it is instructive to compare the proofs, which are quite different and seem to be less involved for the stronger result (once the Doob-Martin compactification has been worked out). The records chain provides an example where we have distributional convergence with a non-degenerate limit, but where a strong limit is necessarily degenerate, i.e. constant.

(d) At a qualitative level functionals of discrete random structures may have a non-trivial tail σ-field, which we interpret as persisting randomness, or they may not, even if the structures themselves show such a persistence; see Theorem 4. A similar phenomenon has been observed in connection with the subtree size profile of binary search trees by Dennert and Grübel (2010).

References

J. D. Biggins. Chernoff's theorem in the branching random walk. *J. Appl. Probability*, 14(3):630–636, 1977.

D. Blackwell and D. Kendall. The Martin boundary of Pólya's urn scheme, and an application to stochastic population growth. *J. Appl. Probability*, 1:284–296, 1964.

F. Dennert and R. Grübel. On the subtree size profile of binary search trees. *Combin. Probab. Comput.*, 19(4):561–578, 2010.

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 (1993), 1991. Commutative algebra and algebraic geometry, II (Italian) (Turin, 1990).

J. L. Doob. Discrete potential theory and boundaries. *J. Math. Mech.*, 8:433–458; erratum 993, 1959.

M. Drmota. *Random trees. An interplay between combinatorics and probability*. Springer, Wien, 2009.

S. N. Evans, R. Grübel, and A. Wakolbinger. Trickle-down processes and their boundaries. *Electron. J. Probab.*, 17:no. 1, 58, 2012.

S. N. Evans, R. Grübel, and A. Wakolbinger. Doob–Martin boundary of Rémy's tree growth chain. *ArXiv e-prints*, Nov. 2014.

A. Gnedin and J. Pitman. Exchangeable Gibbs partitions and Stirling triangles. *Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI)*, 325(Teor. Predst. Din. Sist. Komb. i Algoritm. Metody. 12):83–102, 244–245, 2005.

R. Grübel. A hooray for Poisson approximation. In *2005 International Conference on Analysis of Algorithms*, Discrete Math. Theor. Comput. Sci. Proc., AD, pages 181–191 (electronic). Assoc. Discrete Math. Theor. Comput. Sci., Nancy, 2005.

R. Grübel. On the silhouette of binary search trees. *Ann. Appl. Probab.*, 19(5):1781–1802, 2009.

R. Grübel. Kombinatorische Markov-Ketten. *Math. Semesterber.*, 60(2):185–215, 2013.

R. Grübel. Search trees: Metric aspects and strong limit theorems. *Ann. Appl. Probab.*, 24:1269–1297, 2014.

W. Huss and E. Sava. Internal aggregation models on comb lattices. *Electron. J. Probab.*, 17:no. 30, 21, 2012.

S. Janson, T. Łuczak, and A. Rucinski. *Random Graphs*. Wiley, New York, 2000.

V. A. Kaĭmanovich and A. M. Vershik. Random walks on discrete groups: boundary and entropy. *Ann. Probab.*, 11(3):457–490, 1983.

O. Kallenberg. *Foundations of modern probability*. Springer, New York, 1997.

J. L. Kelley. *General topology*. D. Van Nostrand Company, Inc., Toronto-New York-London, 1955.

D. E. Knuth. *The art of computer programming. Volume 3, Sorting and searching*. Addison-Wesley Publishing Co., Reading, Mass.-London-Don Mills, Ont., 1973.

G. F. Lawler, M. Bramson, and D. Griffeath. Internal diffusion limited aggregation. *Ann. Probab.*, 20(4): 2117–2140, 1992.

L. Lovász. *Large networks and graph limits*, volume 60 of *American Mathematical Society Colloquium Publications*. American Mathematical Society, Providence, RI, 2012.

H. M. Mahmoud. *Evolution of random search trees*. John Wiley & Sons Inc., New York, 1992.

R. Neininger. The Wiener index of random trees. *Combin. Probab. Comput.*, 11(6):587–597, 2002.

J. Neveu. *Discrete-parameter martingales*. North-Holland, Amsterdam, revised edition, 1975.

S. A. Sawyer. Martin boundaries and random walks. In *Harmonic functions on trees and buildings (New York, 1995)*, volume 206 of *Contemp. Math.*, pages 17–44. Amer. Math. Soc., Providence, RI, 1997.

W. Woess. *Random walks on infinite graphs and groups*. Cambridge University Press, Cambridge, 2000.

W. Woess. *Denumerable Markov chains. Generating functions, boundary theory, random walks on trees*. European Mathematical Society (EMS), Zürich, 2009.