

RANDOM LAMINATIONS AND MULTITYPE BRANCHING PROCESSES

NICOLAS CURIEN

Dma-Ens, 45 rue d'Ulm 75005 Paris, France

email: nicolas.curien@ens.fr

YUVAL PERES

Microsoft Research, Redmond, Washington, USA

email: peres@microsoft.com

Submitted March 27, 2011, accepted in final form July 5, 2011

AMS 2000 Subject classification: 60C05, 60D05

Keywords: Random trees, simple random walk, random snake

Abstract

We consider multitype branching processes arising in the study of random laminations of the disk. We classify these processes according to their subcritical or supercritical behavior and provide Kolmogorov-type estimates in the critical case corresponding to the random recursive lamination process of [1]. The proofs use the infinite dimensional Perron-Frobenius theory and quasi-stationary distributions.

1 Introduction

In this note we are interested in multitype branching processes that arise in the study of random recursive laminations. In order to introduce and motivate our results, let us briefly recall the basic construction of [1]. Consider a sequence $U_1, V_1, U_2, V_2, \dots$ of independent random variables, which are uniformly distributed over the unit circle \mathbb{S}_1 . We then construct inductively a sequence L_1, L_2, \dots of random closed subsets of the closed unit disk $\overline{\mathbb{D}}$. To start with, L_1 is set to be the (Euclidean) chord $[U_1 V_1]$ with endpoints U_1 and V_1 . Then at step $n + 1$, we consider two cases. Either the chord $[U_{n+1} V_{n+1}]$ intersects L_n , and we put $L_{n+1} = L_n$. Or the chord $[U_{n+1} V_{n+1}]$ does not intersect L_n , and we put $L_{n+1} = L_n \cup [U_{n+1} V_{n+1}]$. Thus, for every integer $n \geq 1$, L_n is a disjoint union of random chords. See Fig. 1.

A fragment of L_n is a connected component of $\overline{\mathbb{D}} \setminus L_n$. These fragments have a natural genealogy that we now describe. The first fragment, $\overline{\mathbb{D}}$, is represented by \emptyset . Then the first chord $[U_1 V_1]$ splits $\overline{\mathbb{D}}$ into two fragments, which are viewed as the offspring of \emptyset . We then order these fragments in a random way: With probability $1/2$, the first child of \emptyset , which is represented by 0 , corresponds to the largest fragment and the second child, which is represented by 1 , corresponds to the other fragment. With probability $1/2$ we do the contrary. We then iterate this device (see Fig. 2) so that each fragment appearing during the splitting process is labeled by an element of the infinite binary

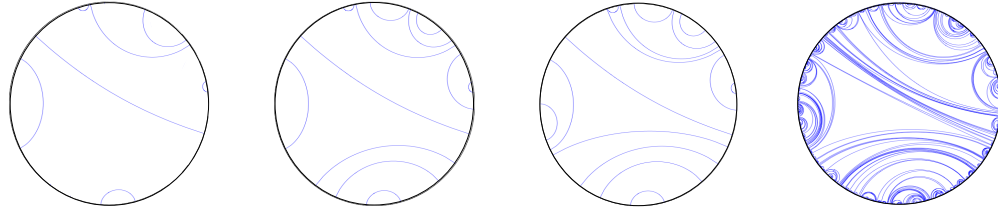


Figure 1: An illustration of the process creating the sequence $(L_n)_{n \geq 1}$. We use hyperbolic chords rather than Euclidean chords for aesthetic reasons.

tree

$$\mathbb{T}_2 = \bigcup_{n \geq 0} \{0, 1\}^n, \quad \text{where } \{0, 1\}^0 = \{\emptyset\}.$$

If F is a fragment, we call *end* of F , any connected component of $F \cap \mathbb{S}_1$. For convenience, the full disk $\overline{\mathbb{D}}$ is viewed as a fragment with 0 end. Consequently, we can associate to any $u \in \mathbb{T}_2$ a label $\ell(u)$ that corresponds to the number of ends of the corresponding fragment in the above process. Lemma 5.5 of [1] then entails that this random labeling of \mathbb{T}_2 is described by the following branching mechanism: For any $u \in \mathbb{T}_2$ labeled $m \geq 0$, choose $m_1 \in \{0, 1, \dots, m\}$ uniformly at random and assign the values $1 + m_1$ and $1 + m - m_1$ to the two children of u . This is the multitype branching process we will be interested in. See Fig. 2.

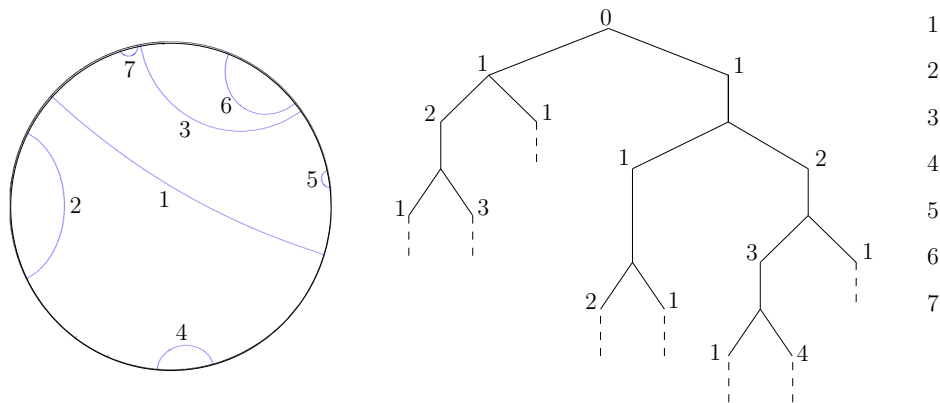


Figure 2: On the left-hand side, the first 7 chords of the splitting process. On the right-hand side, the associated branching process corresponding to the number of ends of the fragments at their creations. Notice that we split the fragments according to the order of appearance of the chords, thus the binary tree on the right-hand side seems stretched.

We can also define a random labeling by using the above branching mechanism but starting with a value $a \geq 0$ at the root \emptyset of \mathbb{T}_2 , the probability distribution of this process will be denoted P_a

and its relative expectation E_a . A ray is an infinite geodesic path $\mathbf{u} = (u_1, u_2, \dots) \in \{0, 1\}^{\mathbb{N}}$ starting from the root \emptyset in \mathbb{T}_2 . For any ray $\mathbf{u} = (u_1, \dots, u_n, \dots)$ or any word of finite length $u = (u_1, \dots, u_n)$, we denote by $[\mathbf{u}]_i$ or $[u]_i$ the word (u_1, \dots, u_i) for $1 \leq i \leq n$, and $[u]_0 = \emptyset$.

Theorem ([1, Lemma 5.5]). *Almost surely, there exists no ray \mathbf{u} along which all the labels starting from 4 are bigger than or equal to 4,*

$$\mathbf{P}_4(\exists \mathbf{u} \in \{0, 1\}^{\mathbb{N}} : \ell([\mathbf{u}]_i) \geq 4, \forall i \geq 0) = 0.$$

The starting label 4 does not play any special role and can be replaced by any value bigger than 4. This theorem was proved and used in [1] to study certain properties of the random closed subset $L_\infty = \overline{\cup L_n}$, and in particular to prove that it is almost surely a maximal lamination (roughly speaking that the complement of L_∞ is made of disjoint triangles), see [1, Proposition 5.4]. One of the purposes of this note is to provide quantitative estimates related to this theorem. Specifically let

$$G_n = \{u \in \{0, 1\}^n : \ell([u]_i) \geq 4, \forall i \in \{0, 1, \dots, n\}\}$$

be the set of paths in \mathbb{T}_2 joining the root to the level n along which the labels are bigger than or equal to 4.

Theorem 1.1. *The expected number of paths starting from the root and reaching level n along which the labels starting from 4 are bigger than or equal to 4 satisfies*

$$\mathbf{E}_4 [\#G_n] \xrightarrow{n \rightarrow \infty} \frac{4}{e^2 - 1}. \tag{1}$$

Furthermore, there exist two constants $0 < c_1 < c_2 < \infty$ such that the probability that $G_n \neq \emptyset$ satisfies

$$\frac{c_1}{n} \leq \mathbf{P}_4(G_n \neq \emptyset) \leq \frac{c_2}{n}. \tag{2}$$

Remark 1.2. *These estimates are reminiscent of the critical case for Galton-Watson processes with finite variance $\sigma^2 < \infty$. Indeed if H_n denotes the number of vertices at height n in such a process then $\mathbf{E}[H_n] = 1$ and Kolmogorov’s estimate [2] implies that $\mathbf{P}(H_n \neq 0) \sim \frac{2}{\sigma^2 n}$.*

The proof of Theorem 1.1 relies on identifying the quasi-stationary distribution of the labels along a fixed ray conditioned to stay bigger than or equal to 4. This is done in Section 2. In Section 3, we also study analogues of this branching random walk on the k -ary tree, for $k \geq 3$, coming from a natural generalization of the process $(L_n)_{n \geq 0}$ where we replace chords by triangles, squares... see Fig. 3.

We prove in these cases that there is no critical value playing the role of 4 in the binary case.

Acknowledgments. The first author thanks Microsoft Research and the University of Washington, where most of this work was done, for their hospitality. We are also grateful to Jean-François Le Gall for precious comments and suggestions on a first version of this note.

2 The critical case

2.1 A martingale

Fix an arbitrary ray \mathbf{u}_0 in \mathbb{T}_2 , for example $\mathbf{u}_0 = (0, 0, 0, 0, 0, \dots)$ and define $X_n = \ell([\mathbf{u}_0]_n)$ for $n \geq 0$, so that X_n is the value at the n -th vertex on the fixed ray \mathbf{u}_0 of the \mathbb{T}_2 -indexed walk ℓ starting from

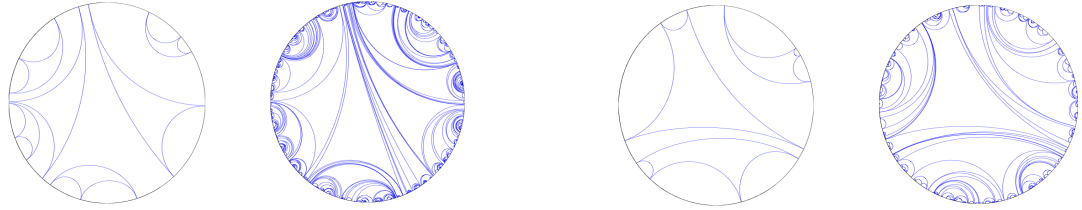


Figure 3: Extension of the process $(L_n)_{n \geq 1}$ where we throw down triangles or squares instead of chords.

$x_0 \geq 4$ at the root. Then $(X_n)_{n \geq 0}$ is a homogeneous Markov chain with transition probabilities given by

$$P_2(x, y) = \frac{1}{x + 1} \mathbf{1}_{1 \leq y \leq x+1}.$$

We first recall some results derived in [1]. If \mathcal{F}_n is the canonical filtration of $(X_n)_{n \geq 0}$ then a straightforward calculation leads to $\mathbf{E}_{x_0}[X_{n+1} | \mathcal{F}_n] = 1 + X_n/2$, hence the process $M_n = 2^n(X_n - 2)$ is a martingale starting from $x_0 - 2$. For $i \geq 1$, we let T_i be the stopping time $T_i = \inf\{n \geq 0 : X_n = i\}$, and $T = T_1 \wedge T_2 \wedge T_3$. By the stopping theorem applied to the martingale $(M_n)_{n \geq 0}$, we obtain for every $n \geq 0$,

$$x_0 - 2 = \mathbf{E}_{x_0}[M_{n \wedge T}] = \mathbf{E}_{x_0}[-2^{T_1} \mathbf{1}_{\{T_1 = T \leq n\}}] + 0 + \mathbf{E}_{x_0}[2^{T_3} \mathbf{1}_{\{T_3 = T \leq n\}}] + \mathbf{E}_{x_0}[2^n(X_n - 2)\mathbf{1}_{\{T > n\}}].$$

One can easily check from the transition kernel of the Markov chain $(X_n)_{n \geq 0}$ that for every $i \geq 1$, $\mathbf{P}_{x_0}[T_1 = T = i] = \mathbf{P}_{x_0}[T_2 = T = i] = \mathbf{P}_{x_0}[T_3 = T = i]$. Hence, the equality in the last display becomes

$$x_0 - 2 = \mathbf{E}_{x_0}[2^n(X_n - 2)\mathbf{1}_{\{T > n\}}],$$

or equivalently

$$x_0 - 2 = 2^n \mathbf{P}_{x_0}[T > n] \mathbf{E}_{x_0}[X_n - 2 | T > n]. \tag{3}$$

Our strategy here is to compute the stationary distribution of X_n conditionally on the non-extinction event $\{T > n\}$, in order to prove the convergence of $\mathbf{E}_4[X_n | T > n]$ and finally to get asymptotics for $\mathbf{P}_4[T > n]$. Before any calculation, we make a couple of simple remarks. Obviously $\mathbf{E}_{x_0}[X_n - 2 | T > n] \geq 2$, and thus we get $2^n \mathbf{P}_{x_0}(T > n) \leq \frac{x_0 - 2}{2}$. Since there are exactly 2^n paths joining the root \emptyset of \mathbb{T}_2 to the level n , we deduce that the number $\#G_n$ of paths joining \emptyset to the level n along which the labels are bigger than or equal to 4 satisfies

$$\mathbf{E}_{x_0}[\#G_n] \leq \frac{x_0 - 2}{2}. \tag{4}$$

Notice that a simple argument shows that if $4 \leq x_0 \leq x_1$ then the chain X_n starting from x_0 and the chain X'_n starting from x_1 can be coupled in such a way that $X_n \leq X'_n$ for all $n \geq 0$.

2.2 The quasi-stationary distribution

We consider the substochastic matrix of the Markov chain X_n killed when it reaches 1, 2 or 3: This is the matrix $(\tilde{P}_2(x, y))_{x, y \geq 4}$ given by

$$\tilde{P}_2(x, y) = \frac{1}{x+1} \mathbf{1}_{y \leq x+1}.$$

We will show that \tilde{P}_2 is a 2-positive matrix, in the sense of [3, Lemma 1]. For that purpose we seek left and right non-negative eigenvectors of \tilde{P}_2 for the eigenvalue 1/2. In other words we look for two sequences $(g(x))_{x \geq 4}$ and $(f(x))_{x \geq 4}$ of non-negative real numbers such that $f(4) = g(4) = 1$ (normalization) and for every $x \geq 4$

$$g(x) = 2 \sum_{y \geq 4} g(y) \tilde{P}_2(y, x) = 2 \sum_{y=(x-1) \vee 4}^{\infty} \frac{g(y)}{y+1}, \tag{5}$$

$$f(x) = 2 \sum_{y \geq 4} \tilde{P}_2(x, y) f(y) = \frac{2}{x+1} \sum_{y=4}^{x+1} f(y). \tag{6}$$

We start with the left eigenvector g . From (5), we get $g(5) = g(4) = 1$, and $g(i) - g(i+1) = \frac{2}{i} g(i-1)$ for $i \geq 5$. Letting

$$G(z) = \sum_{i \geq 4} \frac{z^{i+1}}{i+1} g(i), \quad 0 \leq z < 1,$$

the last observations lead to the following differential equation for G

$$2G(z) = z^{-1}(z-1)G'(z) + z^3,$$

with the condition $G(z) = z^5/5 + o(z^5)$. A simple computation yields $G(z) = 3/4 \exp(2z)(z-1)^2 + (z^3/2 + 3z^2/4 - 3/4)$. After normalization, the generating function $G_{1/2}(z) = \sum_{i \geq 4} g_{1/2}(i)z^i$ of the unique probability distribution $g_{1/2}$ which is a left eigenvector for the eigenvalue 1/2 is given by

$$G_{1/2}(z) = \frac{z}{2} (\exp(2z)(z-1) + z + 1),$$

that is

$$g_{1/2}(i) = \frac{2^{i-3}(i-3)}{(i-1)!} \mathbf{1}_{i \geq 4}.$$

This left eigenvector is called the *quasi-stationary distribution* of X_n conditioned on non-extinction. For the right eigenvector f , a similar approach using generating functions is possible, but it is also easy to check by induction that

$$f(i) = \frac{i-2}{2} \mathbf{1}_{i \geq 4},$$

satisfies (6). Hence the condition (iii) of Lemma 1 in [3] is fulfilled (in particular $\sum f(x)g(x) < \infty$) and the substochastic matrix \tilde{P}_2 is 2-positive. For every $x \geq 4$, set $q_n(x) = \mathbf{P}_4(X_n = x \mid T > n) = \mathbf{P}_4(T > n)^{-1} \tilde{P}_2^n \nu(x)$ where ν stands for the “vector” $(\nu_i)_{i \geq 4}$ with $\nu_4 = 1$ and $\nu_i = 0$ if $i \geq 5$. Theorem 3.1 of [3] then implies that

$$q_n(x) \xrightarrow{n \rightarrow \infty} g_{1/2}(x). \tag{7}$$

Unfortunately this convergence does not immediately imply that $\mathbf{E}_4[X_n \mid T > n] \rightarrow \mathbf{E}[X]$ where X is distributed according to $g_{1/2}$. But this will follow from the next proposition.

Proposition 2.1. For every $n \geq 0$ the sequence $\left(\frac{q_n(x)}{g_{1/2}(x)} \right)_{x \geq 4}$ is decreasing.

Proof. By induction on $n \geq 0$. For $n = 0$ the statement is true. Suppose it holds for $n \geq 0$. By the definition of q_{n+1} , for $x \geq 4$ we have

$$\begin{aligned} q_{n+1}(x) &= \mathbf{P}_4(X_{n+1} = x \mid T > n + 1) \\ &= \frac{1}{\mathbf{P}_4(T > n + 1)} \sum_{z \geq 4} \mathbf{P}_4(X_n = z, X_{n+1} = x, T > n) \\ &= \frac{\mathbf{P}_4(T > n)}{\mathbf{P}_4(T > n + 1)} \sum_{z \geq (x-1) \vee 4} \frac{q_n(z)}{z + 1} \end{aligned} \quad (8)$$

We need to verify that, for every $x \geq 4$, we have $q_{n+1}(x)g_{1/2}(x+1) \geq q_{n+1}(x+1)g_{1/2}(x)$ or equivalently, using (8) and (5) with $g = g_{1/2}$, that

$$\left(\sum_{z \geq x \vee 4} \frac{g_{1/2}(z)}{z + 1} \right) \left(\sum_{z \geq (x-1) \vee 4} \frac{q_n(z)}{z + 1} \right) \geq \left(\sum_{z \geq (x-1) \vee 4} \frac{g_{1/2}(z)}{z + 1} \right) \left(\sum_{z \geq x \vee 4} \frac{q_n(z)}{z + 1} \right)$$

For $x = 4$ this inequality holds. Otherwise, if $x > 4$, we have to prove that

$$q_n(x-1) \sum_{z \geq x \vee 4} \frac{g_{1/2}(z)}{z + 1} \geq g_{1/2}(x-1) \sum_{z \geq x \vee 4} \frac{q_n(z)}{z + 1}. \quad (9)$$

Set $A_x = \frac{q_n(x-1)}{g_{1/2}(x-1)}$ to simplify notation. The induction hypothesis guarantees that $q_n(z) \leq A_x g_{1/2}(z)$ for every $z \geq x$, and therefore

$$\sum_{z \geq x \vee 4} \frac{q_n(z)}{z + 1} \leq A_x \sum_{z \geq x \vee 4} \frac{g_{1/2}(z)}{z + 1}.$$

This gives the bound (9) and completes the proof of the proposition. \square

By Proposition 2.1 we have for every $x \geq 1$, $\frac{q_n(x)}{g_{1/2}(x)} \leq \frac{q_n(4)}{g_{1/2}(4)} \leq C$, where $C = \sup_{n \geq 0} \frac{q_n(4)}{g_{1/2}(4)} < \infty$ by (7). This allows us to apply dominated convergence to get

$$\mathbf{E}_4[X_n \mid T > n] = \sum_{x \geq 4} x q_n(x) \xrightarrow{n \rightarrow \infty} \sum_{x \geq 4} x g_{1/2}(x) = G'_{1/2}(1) = \frac{e^2 + 3}{2}.$$

Using (3) we then conclude that

$$2^n \mathbf{P}_4[T > n] \xrightarrow{n \rightarrow \infty} \frac{4}{e^2 - 1}. \quad (10)$$

2.3 Proof of Theorem 1.1

We first introduce some notation. We denote the tree \mathbb{T}_2 truncated at level n by $\mathbb{T}_2^{(n)}$. For every $u = (u_1, \dots, u_n) \in \{0, 1\}^n$, and every $j \in \{0, 1, \dots, n\}$, recall that $[u]_j = (u_1, \dots, u_j)$, and if $j \geq 1$,

also set $[u]_j^* = (u_1, \dots, u_{j-1}, 1 - u_j)$. We say that $j \in \{0, 1, \dots, n - 1\}$ is a *left turn* (resp. *right turn*) of u if $u_{j+1} = 0$ (resp. $u_{j+1} = 1$). A *down step* of u is a time $j \in \{0, 1, \dots, n - 1\}$ such that

$$\ell([u]_j) - \ell([u]_{j+1}) \geq 2.$$

Note that if j is a down step of u then $\ell([u]_{j+1}^*) = 2 + \ell([u]_j) - \ell([u]_{j+1}) \geq 4$. The set of all $j \in \{0, 1, \dots, n - 1\}$ that are left turns, resp. right turns, resp. down steps, of u is denoted by $L(u)$, resp. $R(u)$, resp. $D(u)$. We endow \mathbb{T}_2 with the lexicographical order \preceq , and say that a path $u \in \{0, 1\}^n$ is on the left (resp. right) of $v \in \{0, 1\}^n$ if $u \preceq v$ (resp. $v \preceq u$). A vertex of $\{0, 1\}^n$ will be identified with the path it defines in $\mathbb{T}_2^{(n)}$. If $u, v \in \mathbb{T}_2$ we let $u \wedge v$ be the last common ancestor of u and v .

Proof of Theorem 1.1. LOWER BOUND. We use a second moment method. Recall that

$$G_n = \{u \in \{0, 1\}^n : \ell([u]_i) \geq 4, \forall i \in \{0, 1, \dots, n\}\}$$

is the set of all paths in $\mathbb{T}_2^{(n)}$ from the root to the level n along which the labels are bigger than or equal to 4. A path in G_n is called "good". Using (10), we can compute the expected number of good paths and get

$$\mathbf{E}_4[\#G_n] = 2^n \mathbf{P}_4[T > n] \xrightarrow{n \rightarrow \infty} \frac{4}{e^2 - 1},$$

as $n \rightarrow \infty$, which proves the convergence (1) in the theorem. For $u \in G_n$ and $j \in \{0, 1, \dots, n\}$, we let $\text{Right}(u, j)$ be the set of all good paths to the right of u that diverge from u at level j ,

$$\text{Right}(u, j) = \{v \in G_n : u \preceq v \text{ and } u \wedge v = [u]_j\}.$$

In particular, if j is a right turn for u , that is $u_{j+1} = 1$, then $\text{Right}(u, j) = \emptyset$. Furthermore $\text{Right}(u, n) = \{u\}$. Let us fix a path $u \in \{0, 1\}^n$, and condition on $u \in G_n$ and on the labels along u . Let $j \in \{0, 1, 2, \dots, n\}$. Note that the first vertex of a path in $\text{Right}(u, j)$ that is not an ancestor of u is $[u]_{j+1}^*$ and its label is $2 + \ell([u]_j) - \ell([u]_{j+1})$, so if we want $\text{Right}(u, j)$ to be non-empty, the time j must be a down step of u . If j is a left turn and a down step for u , the subtree $\{w \in \mathbb{T}_2^{(n)} : w \wedge [u]_j^* = [u]_j^*\}$ on the right of $[u]_j$ is a copy of $\mathbb{T}_2^{(n-j-1)}$, whose labeling starts at $\ell([u]_{j+1}^*)$. Hence thanks to (4) we get

$$\mathbf{E}_4[\#\text{Right}(u, j) \mid u \in G_n, (\ell([u]_i))_{0 \leq i \leq n}] \leq \frac{\ell([u]_{j+1}^*) - 2}{2} = \frac{\ell([u]_j) - \ell([u]_{j+1})}{2}.$$

Since the labels along the ancestral line of u cannot increase by more than one at each step, if $u \in G_n$ we have $\sum_{i=0}^{n-1} |\ell([u]_{i+1}) - \ell([u]_i)| \mathbf{1}_{i \in D(u)} \leq n$. Combining these inequalities, we obtain

$$\mathbf{E}_4 \left[\sum_{j=0}^n \#\text{Right}(u, j) \mid u \in G_n, (\ell([u]_i))_{0 \leq i \leq n} \right] \leq \frac{n}{2}.$$

We can now bound $\mathbf{E}_4[\#G_n^2]$ from above:

$$\begin{aligned}
\mathbf{E}_4[\#G_n^2] &\leq 2\mathbf{E}_4\left[\sum_{u\in\{0,1\}^n}\sum_{u\leq v}\mathbf{1}_{u\in G_n}\mathbf{1}_{v\in G_n}\right] \\
&= 2\sum_{u\in\{0,1\}^n}\mathbf{P}_4(u\in G_n)\mathbf{E}_4\left[\sum_{u\leq v}\mathbf{1}_{v\in G_n}\mid u\in G_n\right] \\
&= 2\sum_{u\in\{0,1\}^n}\mathbf{P}_4(u\in G_n)\mathbf{E}_4\left[\sum_{j=0}^n\#\text{Right}(u,j)\mid u\in G_n\right] \\
&\leq n,
\end{aligned} \tag{11}$$

where we used $\mathbf{E}_4[\#G_n] \leq 1$ from (4). The lower bound of Theorem 1.1 directly follows from the second moment method : Using (1) and (11) we get the existence of $c_1 > 0$ such that

$$\mathbf{P}_4(\#G_n > 0) \geq \frac{\mathbf{E}_4[\#G_n]^2}{\mathbf{E}_4[(\#G_n)^2]} \geq \frac{c_1}{n}. \tag{12}$$

UPPER BOUND. We will first provide estimates on the number of down steps of a fixed path $u \in \{0, 1\}^n$. Recall that $L(u)$, $R(u)$ and $D(u)$ respectively denote the left turns, right turns, and down steps times of u .

Lemma 2.2. *There exists a constant $c_3 > 0$ such that, for every $n \geq 0$ and every $u_0 \in \{0, 1\}^n$*

$$\mathbf{P}_4(u_0 \in G_n, \#D(u_0) \leq c_3n) \leq c_3^{-1}2^{-n} \exp(-c_3n).$$

Proof. We use the notation of Section 2.1. For any set $A \subset \{0, 1, \dots, n-1\}$ and $m \in \{0, 1, \dots, n-\#A\}$, with the notation $N_n^A = \#\{j \in \{0, 1, \dots, n-1\} \setminus A : X_j = 5\}$ we have from [1, formula (27)]

$$\mathbf{P}_4\left[X_{j+1} \geq (X_j - 1) \vee 4, \forall j \in \{0, 1, \dots, n-1\} \setminus A, N_n^A = m\right] \leq \left(\frac{1}{2}\right)^m \left(\frac{3}{7}\right)^{n-m-\#A},$$

We will first obtain crude estimates for N_n^A . Note that $N_n^A \leq N_n^\emptyset$ and that $\sup_{i \geq 1} P_2(i, 5) = \frac{1}{5}$, so that for any $B \subset \{0, 1, \dots, n\}$ we have

$$\mathbf{P}_4(X_i = 5, \forall i \in B) \leq 5^{-\#B}.$$

By summing this bound over all choices of B with $\#B \geq m$ we get $\mathbf{P}_4(N_n^\emptyset \geq m) \leq 2^n 5^{-m}$ for every $m \in \{0, 1, \dots, n\}$. Let $\kappa_1 \in (0, 1/2)$ and $\kappa_2 \in (0, 1)$ such that $\kappa_1 + \kappa_2 < 1$. We have

$$\begin{aligned}
&\mathbf{P}_4(u_0 \in G_n, \#D(u_0) \leq \kappa_1n) \\
&\leq \mathbf{P}_4(u_0 \in G_n, \#D(u_0) \leq \kappa_1n, N_n^\emptyset \leq \kappa_2n) + \mathbf{P}_4(N_n^\emptyset \geq \kappa_2n) \\
&\leq \sum_{\substack{A \subset \{0, 1, \dots, n-1\} \\ \#A \leq \kappa_1n}} \mathbf{P}_4\left[X_{j+1} \geq (X_j - 1) \vee 4, \forall j \in \{0, 1, \dots, n-1\} \setminus A; N_n^A \leq \kappa_2n\right] + \mathbf{P}_4(N_n^\emptyset \geq \kappa_2n) \\
&\leq ([\kappa_2n] + 1) \sum_{\substack{A \subset \{0, 1, \dots, n-1\} \\ \#A \leq \kappa_1n}} \left(\frac{7}{6}\right)^{[\kappa_2n]} \left(\frac{3}{7}\right)^{n-[\kappa_1n]} + \mathbf{P}_4(N_n^\emptyset \geq \kappa_2n)
\end{aligned}$$

Bounding $(\lfloor \kappa_2 n \rfloor + 1)$ by n and the size of a subset A in the sum by $\kappa_1 n$, using the above remark we get that the last display is less than

$$\leq n^2 \binom{n}{\lfloor \kappa_1 n \rfloor} \left(\frac{7}{6}\right)^{\lfloor \kappa_2 n \rfloor} \left(\frac{3}{7}\right)^{n - \lfloor \kappa_1 n \rfloor} + 2^n 5^{-\lfloor \kappa_2 n \rfloor}. \tag{13}$$

Notice that for every $C > 1$ we can choose $\kappa_1 > 0$ small enough so that $n \binom{n}{\lfloor \kappa_1 n \rfloor} \leq C^n$ for n large enough. Furthermore

$$\left(\frac{7}{6}\right)^{\lfloor \kappa_2 n \rfloor} \left(\frac{3}{7}\right)^{n - \lfloor \kappa_1 n \rfloor} = 2^{-n} 2^{\lfloor \kappa_1 n \rfloor} \left(\frac{6}{7}\right)^{n - \lfloor \kappa_1 n \rfloor - \lfloor \kappa_2 n \rfloor},$$

and by choosing κ_1 even smaller if necessary we can ensure that the right hand side of (13) is bounded by $c_3^{-1} 2^{-n} \exp(-c_3 n)$ for some $c_3 > 0$. \square

We use the last lemma to deduce that

$$n \mathbf{P}_4(\exists u \in G_n, \#D(u) \leq c_3 n) \leq \frac{n}{c_3} \exp(-c_3 n) \xrightarrow{n \rightarrow \infty} 0. \tag{14}$$

We now argue on the event $E_L = \{\exists u \in G_n, \#(D(u) \cap L(u)) \geq c_3 n/2\}$. On this event there exists a path $u \in G_n$ with at least $c_3 n/2$ down steps which are also left turns. Conditionally on this event we consider the left-most path P of G_n satisfying these properties, that is

$$P = \min_{\preceq} \{u \in G_n, \#(D(u) \cap L(u)) \geq c_3 n/2\}.$$

A moment's thought shows that conditionally on P and on the values of the labels along the ancestral line of P , the subtrees of $\mathbb{T}_2^{(n)}$ hanging on the right-hand side of P , that are the offspring of the points $[P]_{j+1}^*$ for $j \in L(P)$, are independent and distributed as labeled trees started at $\ell([P]_{j+1}^*)$.

Hence conditionally on P and on the labels $(\ell([P]_i), 0 \leq i \leq n)$, for any $j \in L(P) \cap D(P)$ the expected number of paths belonging to the set $\text{Right}(P, j)$ (defined in the proof of the lower bound) is

$$\begin{aligned} \mathbf{E}_4 \left[\# \text{Right}(P, j) \mid P, (\ell([P]_i))_{0 \leq i \leq n} \right] &= 2^{n-j-1} \mathbf{P}_{\ell([P]_{j+1}^*)}(T > n - j - 1) \\ &\geq 2^{n-j-1} \mathbf{P}_4(T > n - j - 1) \\ &\geq \kappa_3 > 0, \end{aligned} \tag{15}$$

where κ_3 is a positive constant independent of n whose existence follows from (10). Thus we have

$$\begin{aligned} \mathbf{E}_4[\#G_n \mid E_L] &= \mathbf{E}_4 \left[\mathbf{E} \left[\sum_{j=0}^n \# \text{Right}(P, j) \mid P, (\ell([P]_i))_{0 \leq i \leq n} \right] \mid E_L \right] \\ &\geq \kappa_3 \mathbf{E}_4[\#(D(P) \cap L(P)) \mid E_L]. \\ &\geq \frac{c_3 \kappa_3}{2} n. \end{aligned} \tag{16}$$

Since $\mathbf{P}_4(E_L) \leq \mathbf{E}_4[\#G_n] / \mathbf{E}_4[\#G_n \mid E_L]$ we can use (1) to obtain $\mathbf{P}_4(E_L) \leq \kappa_4/n$ for some constant $\kappa_4 > 0$. By a symmetry argument, the same bound holds for the event $E_R = \{\exists u \in G_n, \#(D(u) \cap R(u)) \geq c_3 n/2\}$. Since $\{G_n \neq \emptyset\}$ is the union of the events E_R, E_L and $\{\exists u \in G_n, \#D(u) \leq c_3 n\}$, we easily deduce the upper bound of the theorem from the previous considerations and (14). \square

3 Extensions

Fix $k \geq 2$. We can extend the recursive construction presented in the introduction by throwing down polygons instead of chords: This will yield an analogue of the multitype branching process on the full k -ary tree. Formally if x_1, \dots, x_k are k (distinct) points of \mathbb{S}_1 we denote by $\text{Pol}(x_1, \dots, x_k)$ the convex closure of $\{x_1, \dots, x_k\}$ in $\overline{\mathbb{D}}$. Let $(U_{i,j} : 1 \leq j \leq k, i \geq 1)$ be independent random variables that are uniformly distributed over \mathbb{S}_1 . We construct inductively a sequence L_1^k, L_2^k, \dots of random closed subsets of the closed unit disk $\overline{\mathbb{D}}$. To start with, L_1^k is $\text{Pol}(U_{1,1}, \dots, U_{1,k})$. Then at step $n + 1$, we consider two cases. Either the polygon $P_{n+1} := \text{Pol}(U_{n+1,1}, \dots, U_{n+1,k})$ intersects L_n^k , and we put $L_{n+1}^k = L_n^k$. Or the polygon P_{n+1} does not intersect L_n^k , and we put $L_{n+1}^k = L_n^k \cup P_k$. Thus, for every integer $n \geq 1$, L_n^k is a disjoint union of random k -gons. In a way very similar to what we did in the introduction we can identify the genealogy of the fragments appearing during this process with the complete k -ary tree

$$\mathbb{T}_k = \bigcup_{i \geq 0} \{0, 1, \dots, k - 1\}^i, \quad \text{where } \{0, 1, \dots, k - 1\}^0 = \{\emptyset\}.$$

Then the number of ends of the fragments created during this process gives a labeling ℓ^k of \mathbb{T}_k whose distribution can be described inductively by the following branching mechanism (this is an easy extension of [1, Lemma 5.5]): For $u \in \mathbb{T}_k$ with label $m \geq 0$ we choose a decomposition $m = m_1 + m_2 + \dots + m_k$ with $m_1, m_2, \dots, m_k \in \{0, 1, \dots, m\}$, uniformly at random among all $\binom{m+k-1}{k-1}$ possible choices, and we assign the labels $m_1 + 1, m_2 + 1, \dots, m_k + 1$ to the children of \emptyset . Again the distribution of the labeling ℓ^k of \mathbb{T}_k obtained if we use the above branching mechanism but started from $a \geq 0$ at the root will be denoted by \mathbf{P}_a and its expectation by \mathbf{E}_a . We use the same notation as in the binary case and are interested in a similar question: For which $a \geq 0$ does there exist with positive probability a ray \mathbf{u} such that $\ell^k([\mathbf{u}]_i) \geq a$ for every $i \geq 0$? Specifically, the value a is called *subcritical* for the process $(\ell^k(u), u \in \mathbb{T}_k)$ when there exists a constant $c > 0$ such that

$$\mathbf{P}_a(\exists u \in \{0, 1, \dots, k - 1\}^n : \ell^k([u]_i) \geq a, \forall i \in \{0, 1, \dots, n\}) \leq \exp(-cn).$$

It is called *supercritical* when there exists a constant $c > 0$ such that we have both

$$\begin{cases} \mathbf{P}_a(\exists \mathbf{u} \in \{0, 1, \dots, k - 1\}^{\mathbb{N}} : \ell^k([\mathbf{u}]_i) \geq a, \forall i \in \{0, 1, \dots\}) \geq c, \\ \mathbf{E}_a[\#\{u \in \{0, 1, \dots, k - 1\}^n : \ell^k([u]_i) \geq a, \forall i \in \{0, 1, \dots, n\}\}] \geq \exp(cn). \end{cases}$$

Note that a deterministic argument shows that if $k \geq 2$ and $a = 2$, there always exists a ray with labels greater than or equal to 2, also when $k = 2$ and $a = 3$ there exists a ray with labels greater than 3. The case $k = 2$ and $a = 4$ has been treated in our main theorem. We have the following classification of all remaining cases:

Theorem 3.1. *We have the following properties for the process ℓ^k ,*

- for $k = 2$ and $a \geq 5$ the process is subcritical,
- for $k = 3$ the process is subcritical for $a \geq 4$, and supercritical for $a = 3$,
- for $k \geq 4$ and $a \geq 3$ the process is subcritical.

Proof. SUPERCRITICAL CASE $k = 3$ AND $a = 3$. We will prove that for $k = 3$ and $a = 3$, the process is supercritical. Similarly as in Section 2.1 we consider the tree-indexed process ℓ^3 on a fixed ray of \mathbb{T}_3 , say $\{0, 0, 0, \dots\}$. Then the process Y_n given by the n -th value of ℓ^3 started from 3 along this ray is a homogeneous Markov chain with transition matrix given by

$$P_3(x, y) = \frac{2(x + 2 - y)}{(x + 1)(x + 2)} \mathbf{1}_{1 \leq y \leq x+1}.$$

We introduce the stopping times $T_i = \inf\{n \geq 0, Y_n = i\}$ for $i = 1, 2$ and set $T = T_1 \wedge T_2$. We consider a modification of the process Y_n that we denote \bar{Y}_n , which has the same transition probabilities as Y_n on $\{1, 2, 3, 4\}$, but the transition between 4 and 5 for Y_n is replaced by a transition from 4 to 4 for \bar{Y}_n . Thus we have $\bar{Y}_n \leq 4$ and an easy coupling argument shows that we can construct Y_n and \bar{Y}_n simultaneously in such a way that $\bar{Y}_n \leq Y_n$ for all $n \geq 0$. Hence we have the following stochastic inequality

$$\bar{T} \leq T,$$

with an obvious notation for \bar{T} . To evaluate \bar{T} we consider the subprocess $\bar{Y}_{n \wedge \bar{T}}$ which is again a Markov chain whose transition matrix restricted to $\{3, 4\}$ is

$$\begin{pmatrix} 1/5 & 1/10 \\ 1/5 & 1/5 \end{pmatrix}.$$

The largest eigenvalue λ_{\max} of this matrix is greater than 0.34, which implies that

$$\mathbf{P}_3(T > n) \geq \mathbf{P}_3(\bar{T} > n) \geq \kappa_5(0.34)^n,$$

for some constant $\kappa_5 > 0$ independent of n . It follows that the expected number of paths starting at the root \emptyset of \mathbb{T}_3 that have labels greater than or equal to 3 up to level n , which is $3^n \mathbf{P}_3(T > n)$, eventually becomes strictly greater than 1: There exists $n_0 \geq 1$ such that $\mathbf{P}_3(T > n) > 3^{-n}$ for $n \geq n_0$. A simple coupling argument shows that the process ℓ^3 started from $a \geq 3$ stochastically dominates the process ℓ^3 started from 3. Consequently, if we restrict our attention to the levels that are multiples of n_0 and declare that v is a descendant of u if along the geodesic between u and v the labels of ℓ^3 are larger than 3, then this restriction stochastically dominates a supercritical Galton-Watson process. Hence the value 3 is supercritical for ℓ^3 .

SUBCRITICAL CASE $k = 3$ AND $a = 4$. As in the binary case we let

$$\tilde{P}_3(x, y) = \frac{2(x + 2 - y)}{(x + 1)(x + 2)} \mathbf{1}_{4 \leq y \leq x+1},$$

be the substochastic matrix of the process Y_n started at 4 and killed when it hits 1, 2 or 3. We will construct a positive vector $(h(x))_{x \geq 4}$ such that $\sum_x h(x) < \infty$ and

$$h \cdot \tilde{P}_3 \leq \lambda h, \tag{17}$$

for some positive $\lambda < 1/3$, where we use the notation $h \cdot \tilde{P}_3(y) = \sum_x h(x) \tilde{P}_3(x, y)$. This will imply that

$$\mathbf{P}_4(T > n) \leq \frac{\sum_x h(x)}{h(4)} \lambda^n,$$

where T is the first hitting time of $\{1, 2, 3\}$ by the process Y_n started at 4. The subcriticality of the case $k = 3$ and $a = 4$ follows from the preceding bound since there are 3^n paths up to

level n and $\lambda < 1/3$. To show the existence of a positive vector h satisfying (17) we begin by studying the largest eigenvalue of a finite approximation of the infinite matrix \tilde{P}_3 . To be precise let $\tilde{P}_3^{(30)} = (\tilde{P}_3(i, j))_{4 \leq i, j \leq 30}$. A numerical computation with Maple[©] gives

$$\lambda_{\max} := \max \left\{ \text{Eigenvalues}(\tilde{P}_3^{(30)}) \right\} \simeq 0.248376642883065 < 1/3.$$

The vector $(h(x))_{x \geq 4}$ is then constructed as follows. Let $(h(x))_{4 \leq x \leq 30}$ be an eigenvector associated with the largest eigenvalue λ_{\max} of $\tilde{P}_3^{(30)}$, such that $\min_{4 \leq x \leq 30} h(x) = h(30) = 1$. Note that the vector h can be chosen to have positive coordinates by the Perron-Frobenius theorem and it is easy to verify that $x \rightarrow h(x)$ is decreasing. For $x \geq 31$ we then let

$$h(x) = 13^{x-30} \left(\frac{30!}{x!} \right)^2.$$

We now verify that this vector satisfies (17) with λ slightly greater than λ_{\max} . Suppose first that $y \in \{4, \dots, 30\}$. In this case $\sum_{4 \leq x \leq 30} h(x) \tilde{P}_3(x, y)$ equals $\lambda_{\max} h(y)$ by definition, whereas the contribution of $\sum_{x \geq 31} h(x) \tilde{P}_3(x, y)$ is less than $\sum_{x \geq 31} h(x) < 0.014$, thus

$$h \cdot \tilde{P}_3(y) \leq 0.263h(y). \quad (18)$$

Now, if $y \geq 31$ we have

$$\begin{aligned} & \sum_{x \geq y-1} h(x) \tilde{P}_3(x, y) \\ & \leq 13^{y-30} \left(13^{-1} \tilde{P}_3(y-1, y) \left(\frac{30!}{(y-1)!} \right)^2 + \tilde{P}_3(y, y) \left(\frac{30!}{y!} \right)^2 + \sum_{x \geq y+1} 13^{x-y} \left(\frac{30!}{x!} \right)^2 \right) \\ & \leq 13^{y-30} \left(\frac{30!}{y!} \right)^2 \left(\frac{2}{13} \frac{y^2}{y(y+1)} + \frac{4}{(y+1)(y+2)} + \sum_{i \geq 1} 13^i \left(\frac{y!}{(y+i)!} \right)^2 \right) \\ & \leq 0.3 \cdot h(y) \end{aligned}$$

which proves (17).

OTHER CRITICAL CASES. The other critical cases are treated in the same way. We only provide the reader with the numerical values of the maximal eigenvalues of the truncated substochastic matrices that are very good approximations of the maximal eigenvalues of the infinite matrices,

$$\max\{\text{eigenvalues}(P_2(i, j))_{5 \leq i, j \leq 30}\} \simeq 0.433040861268365 < 1/2,$$

$$\max\{\text{eigenvalues}(P_4(i, j))_{3 \leq i, j \leq 30}\} \simeq 0.231280689028977 < 1/4.$$

□

References

- [1] N. Curien and J.-F. Le Gall. Random recursive triangulations of the disk via fragmentation theory. *Annals of Probab. (to appear)*.
- [2] A. N. Kolmogorov. Zur lösung einer biologischen aufgabe [german: On the solution of a problem in biology]. *Izv. NII Matem. Mekh. Tomskogo Univ.*, 2:7–12, 1938.
- [3] E. Seneta and D. Vere-Jones. On quasi-stationary distributions in discrete-time Markov chains with a denumerable infinity of states. *J. Appl. Probability*, 3:403–434, 1966. [MR0207047](#)