

Random sequential nearest-neighbor coloring on trees

Anne-Laure Basdevant*, Arvind Singh†

June 23, 2026

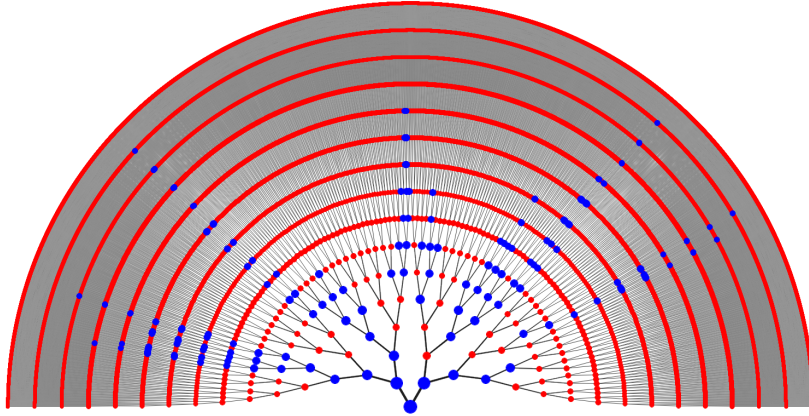


Figure 1: Simulation of the coloring process on a binary tree of height 15. All the leaves are seeds initially colored in red while the root is the seed of the blue cluster.

Abstract

We study a nearest-neighbor coloring process in which vertices are revealed in random order and inherit the color of the closest vertex revealed before them. This model is a discrete analogue of coloring processes previously studied by Preater [10] and Aldous [1] in Euclidean spaces. We focus here on regular trees and analyze the associated genealogy of color inheritance. In contrast with the Euclidean case, the genealogical graph on an infinite regular tree is not connected: it has infinitely many infinite one-ended components, each with a distinct asymptotic direction, while every vertex has only finitely many descendants. We also describe how this structure is modified in the presence of finitely many initial seeds. Finally, we study local limits of the coloring on finite regular trees as their height tends to infinity, for two natural seed configurations: two fixed seeds, and one blue seed at the root with red seeds at the leaves.

1 Introduction

We study a random coloring procedure on graphs, in which vertices are colored sequentially at random, each vertex taking, at the time it is picked, the color of one of its closest already colored vertices. This model can be interpreted as a growth process driven by local competition between different species (*i.e.*, colors). Alternatively, it may also be interpreted as a variant of the classical voter model where new settlers take the political opinion of one of their already settled neighbors.

Sequential nearest-neighbor rules of this kind have appeared in two closely related forms. In the uncolored version, each arriving point is linked to its closest predecessor, giving the on-line nearest-neighbor graph, or random nearest-neighbor tree, whose geometry has been studied in several Euclidean settings [8, 9, 7, 4]. In the colored version, introduced by Preater and Aldous and further studied in [10, 1, 3], the focus is instead on the random partition generated by color inheritance. More precisely, in

*LPSM, Sorbonne Université, France, anne.laure.basdevant@normalesup.org

†CNRS and LMO, Université Paris-Saclay, France arvind.singh@universite-paris-saclay.fr

the latter Euclidean model, the authors consider the unit cube $[-1, 1]^d$. Initially, the origin is colored in blue and the boundary of the cube in red. Subsequently, independent, uniformly sampled random points fall in $[-1, 1]^d$ and, upon arrival, each point takes the color of the nearest point that has appeared so far. This procedure ultimately creates a random coloring of the unit cube¹. One of the main results in [10, 3] is that there exists a.s. an open region containing the origin that is fully blue, *i.e.*, all red vertices remain at a positive distance from the initial blue vertex. On the other hand, the converse question of whether there may exist blue vertices arbitrarily close to the red boundary with positive probability is more delicate and still remains open.

In contrast with these Euclidean works, the present paper deals with a discrete setting in which the underlying graph is a regular tree. This change of geometry leads to rather different phenomena and makes the genealogical structure of the process a natural object of study. Before specializing to trees, let us first describe the model on an arbitrary finite connected graph $G = (V, E)$. Given a set of initially colored vertices called *seeds* and a uniformly random permutation of the non-seed vertices (which encodes the order in which the vertices are revealed), the coloring procedure is defined as follows. When a vertex v is revealed, it takes the color of one of the already colored vertices that minimize the graph distance to v ; if several such vertices exist, one chooses uniformly among them. At the end of the procedure, all vertices are colored.

The same model can equivalently be formulated in terms of arrival times, in a way that naturally extends to infinite graphs: one assigns to each vertex v a time $\tau_v \geq 0$, interpreted as its arrival time, and reveals the vertices in increasing order of these times. On a finite graph, if the variables $(\tau_v)_{v \in V}$ are i.i.d. with a continuous distribution, then their increasing order induces a uniform random permutation of the vertices and we recover the previous construction of the coloring. On an infinite graph, there is in general no way to reveal the vertices one after another in increasing order of their arrival times, since the collection (τ_v) need not admit a smallest element. Nevertheless, the variables (τ_v) still encode the relative order of arrival of the vertices and make it possible to reconstruct the genealogy by a local backward exploration. Vertices s with arrival time $\tau_s = 0$ are declared to be seeds, have no ancestor, and are assigned a fixed color. If v has positive arrival time, we write $d(\cdot, \cdot)$ for the graph distance and set

$$r(v) := \min\{d(v, w) : w \in V, \tau_w < \tau_v\}. \quad (1)$$

The parent of v is defined as the vertex at distance $r(v)$ with an arrival time strictly smaller than τ_v (chosen uniformly at random if there is more than one). Iterating this parent relation yields an ancestral line for each vertex and, in turn, a genealogical forest \mathcal{G} on the set of vertices V . Each connected component of \mathcal{G} is a tree with all its vertices of the same color.

In the present paper, we focus on the case where the graph G is a regular tree. Our main purpose is to study the genealogy generated by the coloring procedure. In the infinite-volume setting, we show that this genealogy is a forest with infinitely many connected components and investigate its geometric structure. We then return to finite regular trees and derive from the infinite-volume picture several consequences for the corresponding finite-volume colorings and their local limits.

1.1 Main results

We now introduce some notation and state the main results of the paper.

1.1.1 Structure of the genealogical graph on a regular infinite tree.

We consider the infinite regular tree \mathcal{T} in which every vertex has $K + 1$ neighbors, where $K \geq 2$ is fixed. We fix a vertex $o \in \mathcal{T}$, which we call the root. For $u, v \in \mathcal{T}$, we write $d(u, v)$ for the graph distance between u and v , and $|v| := d(o, v)$ for the height of the vertex v . For $v \in \mathcal{T}$ and $r \in \mathbb{N}$, we denote by $B(v, r)$ (resp. $S(v, r)$) the ball (resp. sphere) of radius r and center v .

¹This construction produces only a countable set of colored sample points but can be subsequently extended to a coloring of the whole cube by assigning to each location the color of the points in its infinitesimal neighborhood (with an additional “gray” interface corresponding to accumulation points from both colors).

We denote by $\partial\mathcal{T}$ the boundary at infinity of \mathcal{T} , that is, the set of infinite rays² starting from the root. Finally, for $v \in \mathcal{T}$, we denote by \mathcal{T}_v the descendant subtree rooted at v , namely the induced subtree consisting of all vertices $w \in \mathcal{T}$ such that the geodesic path from o to w passes through v .

We first consider the genealogical graph \mathcal{G} associated with the coloring process on \mathcal{T} without seeds. More precisely, we fix an i.i.d. sequence of arrival times $(\tau_v, v \in \mathcal{T})$ sampled according to a continuous distribution³ on $(0, \infty)$. Therefore, all vertices have a positive arrival time (there are no initial seeds) and, for each vertex v , we can define its parent $p(v)$ by the exploration procedure explained in (1). This defines an oriented graph \mathcal{G} on \mathcal{T} , called the *genealogical graph*, whose oriented edges are of the form $(v, p(v))$. We call its connected components the *genealogical components*. By construction, every connected component of \mathcal{G} is necessarily an oriented tree⁴.

We stress that \mathcal{G} has the same vertex set as \mathcal{T} but is not compatible with the graph structure of \mathcal{T} since $p(v)$ is not, in general, a neighbor of v . In particular, the genealogical components are trees but are not subtrees of \mathcal{T} .

For each vertex $v \in \mathcal{T}$, we denote by

$$\Gamma(v) := (v, p(v), p^{(2)}(v), \dots) \quad (2)$$

its ancestral path, where $p^{(k)}(v)$ is the k -th iterate of the parent map. We observe that two vertices u and v belong to the same component if and only if their ancestral paths eventually coalesce, i.e. there exist integers $m, n \geq 0$ such that

$$p^{(m+k)}(u) = p^{(n+k)}(v) \quad \text{for all } k \geq 0.$$

For $v \in \mathcal{T}$, we also define the set of all its descendants:

$$D(v) := \{w \in \mathcal{T} : v \in \Gamma(w)\}.$$

The following theorem describes the structure of \mathcal{G} .

Theorem 1 (The genealogical graph without seeds). *Consider the coloring process on \mathcal{T} without seeds and let \mathcal{G} denote the associated genealogical graph. Then the following assertions hold almost surely:*

1. *The genealogical graph \mathcal{G} has infinitely many connected components. Furthermore, denoting by N_n the number of connected components of \mathcal{G} intersecting the sphere $S(o, n)$, there exists a constant $c > 0$ (depending only on K) such that*

$$\liminf_{n \rightarrow \infty} \frac{N_n}{\#S(o, n)} \geq c \quad \text{a.s.}$$

2. *Every connected component of \mathcal{G} is an infinite one-ended oriented tree, i.e. for every $v \in \mathcal{T}$, its ancestral path $\Gamma(v)$ is infinite while its set of descendants $D(v)$ is finite and satisfies*

$$\mathbb{E}[\#D(v) \mid \tau_v] < \infty \quad \text{a.s.}$$

3. *Every connected component \mathcal{C} of \mathcal{G} admits a limit direction, i.e. there exists an infinite ray $\xi = (w_i)_{i \geq 1} \in \partial\mathcal{T}$ such that, for every vertex v in the component and every $i \geq 1$, the vertices of $\Gamma(v)$ eventually belong to the subtree \mathcal{T}_{w_i} . We denote this limit by*

$$\text{dir } \mathcal{C} \in \partial\mathcal{T}.$$

4. *For every vertex $v \in \mathcal{T}$, there exists, almost surely, a connected component \mathcal{C} such that $\text{dir } \mathcal{C} \in \partial\mathcal{T}_v$.*
5. *If $\mathcal{C}, \mathcal{C}'$ are two distinct connected components of \mathcal{G} , then*

$$\text{dir } \mathcal{C} \neq \text{dir } \mathcal{C}' \quad \text{a.s.}$$

²We call a ray a geodesic starting from the root, i.e., a non-backtracking path $o = v_0, v_1, v_2, \dots$ where v_{i+1} is a neighbor of v_i and $v_{i+1} \neq v_{i-1}$.

³The choice of the distribution does not matter provided it has no atoms. In practice, it will be convenient to choose either a uniform or exponential distribution.

⁴By convention, edges in the genealogical graph are always oriented from a vertex to its parent.

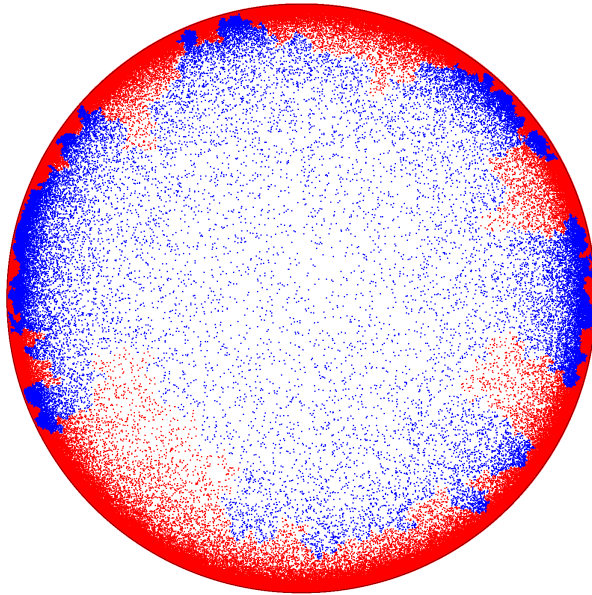


Figure 2: Coloring on the Poincaré hyperbolic disk starting from a single blue point at the origin and a red ideal boundary. Simulation for 10^7 points.

Remark 2. 1. *The fact that the genealogical graph \mathcal{G} has infinitely many connected components is related to the exponential growth of the underlying tree \mathcal{T} . If one considers instead the lattice $G = \mathbf{Z}^d$ for $d \geq 2$ as the base graph, then the associated genealogical graph is a.s. connected. This result can be proved using arguments similar to those used by Aldous to study the coalescence time for the "coupled EA process" (Section 3 of [1]). The arguments of [1] are for the continuous Poissonian model in the plane but can be extended, with some adjustments, to the lattice case for any dimension $d \geq 2$.*

2. *Here, we are working with a base graph \mathcal{T} that is a tree, but the acyclic nature of the graph does not seem fundamental. We believe that the description of \mathcal{G} given in Theorem 1 should remain true for more general vertex-transitive graphs with exponential growth. We also expect that Theorem 1 should admit a continuous analogue for the Poissonian coloring process in hyperbolic space. This is the subject of work in progress (see Figure 2 for a simulation of the coloring process on the Poincaré hyperbolic disk).*

Theorem 3 (The genealogical graph with a finite number of seeds). *Let \mathcal{G} denote the genealogical graph of the coloring process on \mathcal{T} without seeds. Write the decomposition of \mathcal{G} into its connected components as:*

$$\mathcal{G} = \bigsqcup_{i \in I} \mathcal{C}_i.$$

Fix $M \geq 1$ and vertices $\tilde{s}_1, \dots, \tilde{s}_M \in \mathcal{T}$ chosen as seeds. Let $\tilde{\mathcal{G}}$ denote the genealogical graph constructed⁵ with the same i.i.d. sequence of arrival times $(\tau_v, v \in \mathcal{T})$ except that now $\tau_{\tilde{s}_j} = 0$ for all $1 \leq j \leq M$. Then the decomposition of $\tilde{\mathcal{G}}$ into connected components satisfies

$$\tilde{\mathcal{G}} = \bigsqcup_{i \in I} \tilde{\mathcal{C}}_i \sqcup \bigsqcup_{j=1}^M \tilde{\mathcal{U}}_j,$$

where

⁵In this coupled construction of \mathcal{G} and $\tilde{\mathcal{G}}$, we also require ties (when there are multiple possible ancestors) to be broken in the same way, see Section 2.1 for more details.

1. For each $i \in I$, $\tilde{\mathcal{C}}_i$ is an infinite oriented subtree of \mathcal{C}_i . In particular, we have $\text{dir } \tilde{\mathcal{C}}_i = \text{dir } \mathcal{C}_i$.
2. For each $j \in \{1, \dots, M\}$, $\tilde{\mathcal{U}}_j$ is an oriented tree rooted at \tilde{s}_j . Moreover,
 - (a) For $v \in \tilde{\mathcal{U}}_j$, $v \neq \tilde{s}_j$, the set $\tilde{D}(v)$ of descendants of v in $\tilde{\mathcal{U}}_j$ is finite with $\mathbb{E}[\#\tilde{D}(v) \mid \tau_v] < \infty$.
 - (b) The component $\tilde{\mathcal{U}}_j$ is infinite if and only if \tilde{s}_j has infinite degree. Furthermore, we have

$$\mathbb{P}(\tilde{s}_j \text{ has infinite degree in } \tilde{\mathcal{U}}_j) = \begin{cases} 1 & \text{if } \{v \in \mathcal{T}, d(v, \tilde{s}_j) = \min_{i \leq M} d(v, \tilde{s}_i)\} \text{ is infinite} \\ 0 & \text{otherwise} \end{cases}.$$

Remark 4. We note that if there are $M \leq K + 1$ seeds on a $(K + 1)$ -regular tree, then necessarily all the seeds have infinite degree in the genealogical graph.

1.1.2 Coloring on finite regular trees and local limits

In the second part of this paper, we describe the coloring process on large finite regular trees and study its limit as the height of the tree tends to infinity. For $\ell \geq 0$, let

$$\mathcal{T}^{(\ell)} := \mathcal{T} \cap B(o, \ell)$$

denote the finite regular tree of height ℓ .

Coloring with seeds at the root and at the leaves. First, we consider the coloring process on $\mathcal{T}^{(\ell)}$, where seeds are placed at the root and at every leaf of $\mathcal{T}^{(\ell)}$ (that is, every vertex at height ℓ). This setting mimics, in some sense, the model studied in the continuous setting on $[-1, 1]^d$, with a blue point at the origin and the boundary of the cube colored red. Thus, here the seed at the root propagates the blue color to its descendants, whereas all seeds at the leaves propagate the same red color to their descendants (see Figure 1 for an illustration of this process). We denote by $B^{(\ell)}$ (resp. $R^{(\ell)}$) the final sets of blue (resp. red) vertices.

Note that for $\ell < \ell'$, there is a natural coupling for the coloring processes on $\mathcal{T}^{(\ell)}$ and $\mathcal{T}^{(\ell')}$ obtained by considering the same ordering of the vertices (*i.e.*, the same arrival times) and simply ignoring vertices v with height $|v| > \ell$ when coloring $\mathcal{T}^{(\ell)}$. For the particular seed configuration considered here, this coupling is monotone in the sense that

$$B^{(\ell)} \subset B^{(\ell')} \quad \text{a.s. for all } \ell < \ell'. \quad (3)$$

We are particularly interested in the following quantities:

$$M_\ell := \max\{|v| : v \in B^{(\ell)}\}, \quad m_\ell := \min\{|v| : v \in R^{(\ell)}\}, \quad (4)$$

that is, the maximal height reached by the blue region and the minimal height reached by the red region. By (3), the sequences $(m_\ell)_{\ell \geq 1}$ and $(M_\ell)_{\ell \geq 1}$ are both non-decreasing with ℓ .

Theorem 5. *Consider the coloring process on $\mathcal{T}^{(\ell)}$ with a blue seed at the root and red seeds at the leaves constructed with the coupling described above. Then the coloring process converges a.s. in the local topology to the random coloring of \mathcal{T} obtained by coloring the genealogical component of the root in blue and all other components in red. Furthermore, the sequence $(M_\ell)_{\ell \geq 1}$ converges a.s. to infinity and satisfies*

$$\lim_{\ell \rightarrow \infty} \mathbb{P}(M_\ell = \ell - 1) = 1. \quad (5)$$

The sequence $(m_\ell)_{\ell \geq 1}$ converges a.s. to a finite random variable m_∞ such that

$$\mathbb{P}(m_\infty = 1) = \inf_{\ell \geq 1} \mathbb{P}(m_\ell = 1) > 0. \quad (6)$$

The a.s. convergence of the coloring process to the limit coloring of \mathcal{T} described above is a soft result that follows from the coupling and the description of the genealogical graph \mathcal{G} given in Theorem 1. The more interesting (and more difficult) part of the theorem is (5), which, somewhat counterintuitively, states that the blue region not only grows to infinity but, in fact, reaches maximum height $\ell - 1$ with

high probability as ℓ tends to infinity. This result is in stark contrast with the behavior of the coloring process in Euclidean space, where it is known that, with positive probability, the blue region does not reach the boundary⁶.

Statement (6) of Theorem 5 shows that, although the distance from the root to the red seeds increases with the height of the tree, the coloring remains non-trivial around the root. Again, this phenomenon contrasts with the Euclidean setting: if one considers the Poissonian coloring in a ball with red boundary conditions and a blue seed at the origin, then letting the radius tend to infinity (together with an appropriate scaling) leads to a trivial limit in which the coloring becomes entirely blue. On the infinite tree \mathcal{T} , however, the limit remains non-trivial because the red seeds do not disappear; instead, they are pushed to infinity, *i.e.*, to the ideal boundary of the tree. This phenomenon is reminiscent of situations arising in hyperbolic geometry, where infinite-volume limits of random structures may naturally live on the ideal boundary of the space. A related example is the recently introduced model of ideal Poisson–Voronoi tessellations [5].

Two-seed coloring. We now study the coloring process on the finite tree with two seeds. Fix another vertex $o' \in \mathcal{T}$ and for $\ell > d(o, o')$, consider the coloring process on $\mathcal{T}^{(\ell)}$ with a blue seed at the root and a red seed at o' . As usual, we denote by $\tilde{B}^{(\ell)}$ (resp. $\tilde{R}^{(\ell)}$) the set of blue (resp. red) vertices. We describe the local limit in distribution of the process $(\tilde{B}^{(\ell)}, \tilde{R}^{(\ell)})$ in terms of the genealogical graph associated with the coloring process on the infinite tree \mathcal{T} with two seeds. We first introduce some notation.

Let $\tilde{\mathcal{G}}$ denote the genealogical graph of the coloring process on the infinite tree \mathcal{T} with two seeds at o and o' . Theorem 3 states that the decomposition of $\tilde{\mathcal{G}}$ into connected components satisfies

$$\tilde{\mathcal{G}} = \bigsqcup_{i \in I} \tilde{\mathcal{C}}_i \sqcup \tilde{\mathcal{U}}_o \sqcup \tilde{\mathcal{U}}_{o'},$$

where $\tilde{\mathcal{U}}_o$ (resp. $\tilde{\mathcal{U}}_{o'}$) is the genealogical component of o (resp. o'), and each genealogical component $\tilde{\mathcal{C}}_i$ is an infinite oriented tree with asymptotic direction $\text{dir } \tilde{\mathcal{C}}_i$. Observe that, for any asymptotic direction in $\partial\mathcal{T}$, we can determine whether it is asymptotically closer to o , closer to o' , or asymptotically at equal distance from o and o' .

Now, consider the following coloring procedure:

- Color all vertices of $\tilde{\mathcal{U}}_o$ in blue and all vertices of $\tilde{\mathcal{U}}_{o'}$ in red.
- For the clusters $\tilde{\mathcal{C}}_i$ such that their asymptotic direction $\text{dir } \tilde{\mathcal{C}}_i$ is asymptotically closer to o (resp. to o'), color all their vertices in blue (resp. red).
- For each cluster $\tilde{\mathcal{C}}_i$ such that its asymptotic direction $\text{dir } \tilde{\mathcal{C}}_i$ is at equal distance from o and o' , pick an independent Bernoulli random variable X_i with parameter $\frac{1}{2}$ and color the cluster in blue if $X_i = 1$ and red otherwise.

This procedure yields a random coloring process of \mathcal{T} in two colors; we denote by $\chi_{o'}$ its distribution.

Theorem 6. *Consider the coloring process on the finite tree $\mathcal{T}^{(\ell)}$ with two seeds located at o and o' . Then the coloring $(\tilde{B}^{(\ell)}, \tilde{R}^{(\ell)})$ on $\mathcal{T}^{(\ell)}$ converges in distribution in the local topology to $\chi_{o'}$ on \mathcal{T} .*

In fact, if we construct the process $(\tilde{B}^{(\ell)}, \tilde{R}^{(\ell)})$ and $\tilde{\mathcal{G}}$ using the same sequence of arrival times $(\tau_v)_{v \in \mathcal{T}}$, then the previous result can be strengthened to yield almost sure convergence of $(\tilde{B}^{(\ell)}, \tilde{R}^{(\ell)})$ outside the clusters whose asymptotic directions are at equal distance from o and o' (but not for the cluster at equal distance). In particular, we also obtain the following proposition.

Proposition 7. *Let o' be a vertex of \mathcal{T} such that $d(o, o')$ is odd. For $\ell > d(o, o')$, consider the coloring process on the finite tree $\mathcal{T}^{(\ell)}$ with two seeds located at o and o' , and assume that all processes are constructed using the same sequence of arrival times. Then $(\tilde{B}^{(\ell)}, \tilde{R}^{(\ell)})_{\ell > d(o, o')}$ converges almost surely in the local topology to some random sets $(\tilde{B}^{(\infty)}, \tilde{R}^{(\infty)})$. Moreover:*

- (a) $\tilde{B}^{(\infty)} \cap \{v \in \mathcal{T} : d(v, o) > d(v, o')\}$ is a.s. finite;

⁶The open question mentioned at the beginning of the introduction is whether this occurs with probability 1, see open Question 1.4 in [3] and the discussion preceding Corollary 1 in [10] for more details.

(b) $\tilde{R}^{(\infty)} \cap \{v \in \mathcal{T} : d(v, o) < d(v, o')\}$ is a.s. finite.

The assumption that $d(o, o')$ is odd is required here to ensure that no vertex of \mathcal{T} can be at the same distance from o and o' . This implies that the color of a vertex v eventually stabilizes according to whether it belongs to a cluster of $\tilde{\mathcal{G}}$ whose asymptotic direction is closer to o or to o' .

We will, in fact, also show that if $d(o, o')$ is even, the sequence $(\tilde{B}^{(\ell)}, \tilde{R}^{(\ell)})_{\ell > d(o, o')}$ still converges outside the set consisting of vertices that are at equal distance from o and o' , and the two finiteness statements above remain valid. Thus, from a macroscopic perspective, (a) and (b) tell us that the coloring of $\mathcal{T}^{(\ell)}$ with two seeds is “almost” deterministic: apart from finitely many vertices, all those that are closer to o are colored blue, while those that are closer to o' are colored red. Simulations also suggest that the same phenomenon occurs for a similar coloring process on the hyperbolic plane.

Outline of the paper

In Section 2, we give the precise construction of the ancestral path $\Gamma(v)$ and of an associated ancestral process $(V_n(v), W_n(v))$. This includes the uniform increasing indexings used to break ties and an equivalent construction in terms of the inter-arrival times of a lower record process. We then derive the preliminary estimates on this process, in particular the growth estimates for the jumps and for the height of the ancestral path (Proposition 14 and Corollary 15).

Section 3 is devoted to the proof of Theorem 1. We first prove that every genealogical component has an asymptotic direction. We then use Proposition 17, which shows that an ancestral path may avoid an entire descendant subtree with positive probability, to prove that \mathcal{G} has infinitely many components and that every descendant sector contains the direction of some component. The fact that two distinct components have distinct directions is proved by introducing a coupled exploration of two ancestral paths, encoded by a lower 2-record process. Finally, we prove that every vertex has only finitely many descendants by a combinatorial estimate on possible descendant paths.

In Section 4, we prove Theorem 3. The argument is based on a coupling between the genealogical graph without seeds and the genealogical graph with a finite set of seeds: the ancestral path of a vertex is changed exactly when its exploration discovers one of the seeds. This observation transfers most of the structure obtained in the seedless case and reduces the remaining work to characterizing when a seed has infinite degree in its genealogical component.

Section 5 returns to finite trees and proves the local-limit results. We first record the local convergence of genealogical graphs under the natural coupling. We then treat the two-seed coloring, proving Theorem 6 and Proposition 7 by reading the limiting color of a vertex from the terminal seed or the asymptotic direction of its infinite-volume ancestral path. The last part of the section studies the coloring with a blue seed at the root and red seeds at the leaves: the local convergence and the statement (6) follow from the infinite-volume results with seeds. The remaining estimate (5), which concerns the behavior of the coloring near the leaves, requires different arguments; it is proved separately in Section 6, where we follow the coloring dynamics forward in time and compare the recursive growth of blue vertices with a supercritical Galton–Watson process.

2 The genealogical process

2.1 Construction of the genealogical process

A central object in the study of the geometry of the genealogical graph is the ancestral path $\Gamma(v)$ of a vertex v , as defined in (2). To analyze this process, it is convenient to keep track not only of $\Gamma(v)$ but also of the set of vertices of \mathcal{T} that need to be examined in order to determine $\Gamma(v)$. We denote this set by $W_\infty(v)$. This subsection gives a precise construction of $\Gamma(v)$ and $W_\infty(v)$. We first introduce the notion of a *uniform increasing indexing*, which gives us a rule to break ties when several possible ancestors of a vertex are available.

Definition 8. *Let $G = (V, E)$ be a connected, locally finite graph, and let $v \in V$. We say that a random sequence $(z_j)_{j \geq 1}$ is a uniform increasing indexing of the vertices of G centered at v if it satisfies the following properties:*

- (i) (indexing of V) $\{z_j : j \geq 1\} = V$ and $z_j \neq z_i$ for $j \neq i$;
- (ii) (increasing) if $j < i$, then $d(v, z_j) \leq d(v, z_i)$;
- (iii) (uniformity) for every $r \geq 0$, the order of the vertices at distance r from v induced by $(z_j)_{j \geq 1}$ is uniform among all possible orderings and independent of the orderings induced at other distances.

Note that (iii) makes sense because we consider locally finite graphs, so the set of vertices at distance r from v is finite for every $r \geq 0$. In particular, a uniform increasing indexing of the vertices of \mathcal{T} centered at v can be constructed by first sampling a uniform random ordering of the vertices at distance 1 from v , then a uniform random ordering of the vertices at distance 2 from v , and so on, independently for each distance.

Recall that $(\tau_u)_{u \in \mathcal{T}}$ denotes the arrival times associated with the vertices of \mathcal{T} . In this section, we consider the model without seeds; hence, the variables $(\tau_u)_{u \in \mathcal{T}}$ are i.i.d. with an atomless distribution. For each vertex $w \in \mathcal{T}$, let $(z_j^w)_{j \geq 1}$ denote a uniform increasing indexing of \mathcal{T} centered at w , independent of the family $(\tau_u)_{u \in \mathcal{T}}$, and such that the families $((z_j^w)_{j \geq 1})_{w \in \mathcal{T}}$ are independent. Using this collection of random variables, we define the direct ancestor $p(v)$ of v as follows: we examine the arrival times of the vertices of \mathcal{T} in the order given by the sequence $(z_j^v)_{j \geq 1}$, and we stop as soon as we find a vertex u such that $\tau_u < \tau_v$. We then declare that $p(v)$ is this vertex u . More formally, we set

$$J(v) := \inf\{j \geq 1 : \tau_{z_j^v} < \tau_v\}, \quad p(v) := z_{J(v)}^v.$$

In this construction, the family of random variables $(\tau_u)_{u \in \mathcal{T}}$ uniquely determines the distance between v and its direct ancestor $p(v)$, while the family of random indexings is used to break ties when several possible ancestors are available. Moreover, because the ordering is uniform on each sphere centered at v , this choice is uniform among all closest such vertices; therefore, this definition of the direct ancestor coincides with the definition given in the introduction.

For $v \in \mathcal{T}$, we can generalize this construction to define the entire ancestral path $\Gamma(v) = (V_n)_{n \geq 1}$ of v by induction. We simultaneously construct a sequence of sets $(W_n(v))_{n \geq 1}$: the set $W_n(v)$ records the vertices whose arrival times have been revealed in order to determine (V_1, \dots, V_n) .

First, set

$$V_1 := V_1(v) := v, \quad W_1 := W_1(v) := \{v\}.$$

Assume now that (V_1, \dots, V_n) and W_n have been constructed, where V_n is the $(n-1)$ -th ancestor of v , and W_n records the vertices revealed up to this point. Assume moreover that, for every $u \in W_n$, we have $\tau_u \geq \tau_{V_n}$. To determine the direct ancestor of V_n , we examine the vertices in the order given by the increasing indexing centered at V_n and set

$$J_n := \inf\{j \geq 1 : \tau_{z_j^{V_n}} < \tau_{V_n}\} = \inf\{j \geq 1 : z_j^{V_n} \notin W_n \text{ and } \tau_{z_j^{V_n}} < \tau_{V_n}\}.$$

The two definitions coincide since, for every $u \in W_n$, we have $\tau_u \geq \tau_{V_n}$. Thus, by the construction of the direct ancestor, $z_{J_n}^{V_n}$ is precisely equal to $p(V_n)$. We therefore set

$$V_{n+1} := z_{J_n}^{V_n}, \quad W_{n+1} := W_n \cup \{z_j^{V_n} : j \leq J_n\}. \tag{7}$$

By construction, W_{n+1} records all vertices whose arrival times have been examined in order to determine (V_1, \dots, V_{n+1}) , and for every $u \in W_{n+1}$, we have $\tau_u \geq \tau_{V_{n+1}}$.

Finally, define

$$\Gamma(v) := (V_n)_{n \geq 1}, \quad W_\infty := W_\infty(v) := \bigcup_{n \geq 1} W_n,$$

so that $\Gamma(v)$ is the ancestral path of v , and W_∞ is the set of vertices whose arrival times have been ultimately revealed. The process $(V_n(v), W_n(v))_{n \geq 1}$ will play a central role throughout the paper, and we therefore introduce the following terminology.

Definition 9. For $v \in \mathcal{T}$, the process $(V_n(v), W_n(v))_{n \geq 1}$ constructed by the procedure described above is called the ancestral process of v . Its first coordinate $(V_n(v))_{n \geq 1}$ is called the ancestral path of v .

For a fixed v , it will be convenient to consider a process $(\hat{V}_n(v), \hat{W}_n(v))_{n \geq 1}$ with the same distribution as the ancestral process of v , but defined directly in terms of the inter-arrival times $(I_n)_{n \geq 1}$ of a generic lower record process associated with an i.i.d. atomless sequence⁷. For each $w \in \mathcal{T}$, let $(z_j^w)_{j \geq 1}$ be a uniform increasing indexing of \mathcal{T} centered at w , independent of $(I_n)_{n \geq 1}$, and such that the families $((z_j^w)_{j \geq 1})_{w \in \mathcal{T}}$ are independent. Set

$$\hat{V}_1 := \hat{V}_1(v) := v, \quad \hat{W}_1 := \hat{W}_1(v) := \{v\}.$$

By induction, assuming that $(\hat{V}_1, \dots, \hat{V}_n)$ and \hat{W}_n have been constructed, define \hat{J}_n as the first index such that we explore I_{n+1} new vertices around \hat{V}_n that do not belong to \hat{W}_n :

$$\hat{J}_n := \inf \left\{ j \geq 1 : \#\{i \leq j : z_i^{\hat{V}_n} \notin \hat{W}_n\} = I_{n+1} \right\}. \quad (8)$$

Then set

$$\hat{V}_{n+1} := z_{\hat{J}_n}^{\hat{V}_n}, \quad \hat{W}_{n+1} := \hat{W}_n \cup \{z_j^{\hat{V}_n} : j \leq \hat{J}_n\}. \quad (9)$$

We note that this construction uses only the inter-arrival times $(I_n)_{n \geq 1}$ of a generic lower record process, rather than the arrival times themselves. Nevertheless, as the next lemma states, it has the same distribution as the ancestral process.

Lemma 10. The process $(\hat{V}_n(v), \hat{W}_n(v))_{n \geq 1}$ defined above has the same distribution as the ancestral process $(V_n(v), W_n(v))_{n \geq 1}$. Moreover, for every $n \geq 1$, we have

$$I_n = \#\left(\hat{W}_n(v) \setminus \hat{W}_{n-1}(v)\right), \quad \text{with the convention } \hat{W}_0(v) := \emptyset.$$

Proof. The result follows directly by induction from the construction. Conditionally on the past up to step n , the unexplored arrival times, read in the random indexing around V_n , form a fresh i.i.d. atomless sequence. Since τ_{V_n} is the current lower record among the revealed times, the number of new vertices inspected before finding the next ancestor has the law of I_{n+1} , and the last inspected vertex is chosen through the same uniform increasing indexing. This is exactly the recursive construction of (\hat{V}_n, \hat{W}_n) . The identity for I_n is also immediate, with $I_1 = 1$ corresponding to $\hat{W}_1 = \{v\}$. \square

In the rest of the paper, whenever we study a single ancestral process, we will usually favor the “record-based” construction above and omit the hats, so we will simply write $(V_n(v), W_n(v))_{n \geq 1}$ for $(\hat{V}_n(v), \hat{W}_n(v))_{n \geq 1}$.

2.2 Preliminary results on the genealogical process

We now present some preliminary results for the ancestral process. We use the construction of the previous section from the inter-arrival times $(I_k)_{k \geq 1}$ and the random indexings $((z_j^w)_{j \geq 1})_{w \in \mathcal{T}}$ through Equations (8) and (9). Since \mathcal{T} is transitive, it will be enough to study the process starting from $v = o$ and, for the sake of readability, we will simply write $(V_n, W_n)_{n \geq 1}$ for $(V_n(o), W_n(o))_{n \geq 1}$.

⁷One concrete realization of $(I_n, n \geq 1)$ is the following. Let $(X_i)_{i \geq 1}$ be i.i.d. with an atomless distribution, and set $T_0 = 0, T_1 = 1$, and, for $n \geq 1$, $T_{n+1} := \inf\{k > T_n : X_k < X_{T_n}\}$. Then set $I_n := T_n - T_{n-1}$ for $n \geq 1$. The law of $(I_n, n \geq 1)$ does not depend on the choice of the distribution of $(X_i)_{i \geq 1}$.

For $n \geq 2$, define R_n as the graph distance between V_n and V_{n-1} :

$$R_n := d(V_n, V_{n-1}). \quad (10)$$

Note that R_n is a function of V_{n-1}, W_{n-1}, I_n given by

$$R_n = \inf\{r \geq 0 : \#\{u \in \mathcal{T} \setminus W_{n-1}, d(V_{n-1}, u) \leq r\} \geq I_n\}. \quad (11)$$

Using this equation, we can give a precise estimate of R_n in terms of I_n . Recall that \mathcal{T}_u denotes the descendant subtree rooted at u (for $u \neq o$, it is the connected component of $\mathcal{T} \setminus \{u\}$ that does not contain the root o).

Lemma 11. *For any $n \geq 1$, we have the following properties:*

- W_n is a connected subset of \mathcal{T} .
- $\mathcal{T}_{V_n} \cap W_n = \{V_n\}$.

Moreover, for any $n \geq 2$, we have

$$\frac{I_n}{K+1} \leq K^{R_n} \leq KI_n. \quad (12)$$

Proof. Let us first note that, by construction, $V_n \in W_n$ for all $n \geq 1$. Let us now prove by induction on n that W_n is connected and $\mathcal{T}_{V_n} \cap W_n = \{V_n\}$. This holds for $n = 1$. Assume it holds for $n - 1$. Let $(z_j^{n-1})_{j \geq 1}$ be the subsequence of $(z_j^{V_{n-1}})_{j \geq 1}$ obtained by keeping only the terms that do not belong to W_{n-1} , and set

$$Z_l := W_{n-1} \cup \{z_j^{n-1} : j \leq l\},$$

so that $Z_0 = W_{n-1}$ and $Z_{I_n} = W_n$. Since $V_{n-1} \in W_{n-1}$ and $z_j^{V_{n-1}}$ is an increasing indexing centered at V_{n-1} , we see by induction on l that Z_l is a connected subset for all $0 \leq l \leq I_n$ and so W_n is connected. Moreover, $W_n = Z_{I_n} \cup \{V_n\}$ and Z_{I_n} is a connected subset which does not contain V_n . Hence, it is contained in one of the connected components of $\mathcal{T} \setminus \{V_n\}$. Since the root o is in Z_{I_n} , the connected component of $\mathcal{T} \setminus \{V_n\}$ containing Z_{I_n} is necessarily the one containing the root. Hence, we get

$$Z_{I_n} \cap \mathcal{T}_{V_n} = \emptyset,$$

and so

$$\mathcal{T}_{V_n} \cap W_n = \{V_n\}.$$

Let us now prove (12). On the one hand, using (11), we have

$$\begin{aligned} I_n &\leq \#\{u \in \mathcal{T} \setminus W_{n-1}, d(V_{n-1}, u) \leq R_n\} \\ &\leq \#\{u \in \mathcal{T} \setminus \{V_{n-1}\}, d(V_{n-1}, u) \leq R_n\} \\ &= \sum_{i=0}^{R_n-1} (K+1) \cdot K^i \leq (K+1)K^{R_n}. \end{aligned}$$

This gives the lower bound. The upper bound is trivial when $R_n = 1$. We therefore assume $R_n \geq 2$, and bound I_n from below using again (11) combined with the fact that $\mathcal{T}_{V_{n-1}} \cap W_{n-1} = \{V_{n-1}\}$ to conclude that

$$\begin{aligned} I_n &> \#\{u \in \mathcal{T} \setminus W_{n-1}, d(V_{n-1}, u) < R_n\} \\ &\geq \#\{u \in \mathcal{T}_{V_{n-1}} \setminus \{V_{n-1}\}, d(V_{n-1}, u) < R_n\} \\ &\geq K^{R_n-1}. \end{aligned}$$

□

The process $(I_n)_{n \geq 1}$ is the inter-arrival-time process of a classical record process. Such processes have been well studied. We will use the following result, which is an easy consequence of the classical asymptotic behavior of record processes.

Proposition 12 (see Arnold [2]). *Let $(I_n)_{n \geq 1}$ be the inter-arrival times of a record process of an i.i.d. sequence with diffuse distribution. Then*

(i) *We have*

$$\lim_{n \rightarrow \infty} \frac{\log I_n}{n} = 1 \text{ a.s.}$$

(ii) *For any $n \geq 1$ and $i_2, \dots, i_n \geq 1$,*

$$\mathbb{P}(I_2 = i_2, \dots, I_n = i_n, 2^k \leq I_k \leq 3^k \text{ for all } k > n) > 0.$$

(iii) *For any $c \geq 0$ and $n \geq 1$, we have*

$$\mathbb{P}(\forall m \geq 0, I_{n+m+1} \geq c \mid I_1, \dots, I_n) \geq \mathbb{P}(\forall m \geq 0, I_{m+2} \geq c) > 0 \text{ a.s.}$$

Proof. The first part of the proposition is a classical result (see, for example, [2], p. 28). For the second point, one uses the Markov description of the lower record process: any finite admissible sequence of inter-arrival times has positive probability, and after such a finite prescription the future process is again a lower record process started from the current record value. Since $2 < e < 3$, point (i) implies that, for some sufficiently large deterministic N , the tail event $\{2^k \leq I_k \leq 3^k \text{ for all } k > N\}$ has positive probability; the finitely many constraints for $n < k \leq N$ can then be imposed with positive probability.

Finally, the last point follows from the standard Markov description of the record-time chain: conditionally on I_1, \dots, I_n , the future gaps are stochastically lower bounded by a fresh record inter-arrival process started from time 1. Point (ii) then implies that $\mathbb{P}(\forall m \geq 0, I_{m+2} \geq c) > 0$. \square

We now introduce the filtration $(\mathcal{F}_n)_{n \geq 1}$ given by

$$\mathcal{F}_n := \sigma((I_k)_{k \geq 1}, ((z_j^{V_i})_{j \geq 1})_{i < n}). \quad (13)$$

Notice that \mathcal{F}_n contains the whole sequence of inter-arrival times $(I_k)_{k \geq 1}$, not only the first n of them.

Lemma 13. *For any $n \geq 1$, V_n, W_n and R_{n+1} are \mathcal{F}_n -measurable. Moreover, conditionally on \mathcal{F}_n , the vertex V_{n+1} is uniformly distributed on the set of vertices $u \in \mathcal{T} \setminus W_n$ such that $d(V_n, u) = R_{n+1}$.*

Proof. The measurability assertion follows by induction. Indeed, V_1 and W_1 are deterministic, and if V_n and W_n are \mathcal{F}_n -measurable, then R_{n+1} is \mathcal{F}_n -measurable by (11), since all the variables $(I_k)_{k \geq 1}$ are included in \mathcal{F}_n . Once the indexing centered at V_n is added, the definitions (8) and (9) give V_{n+1} and W_{n+1} . The conditional distribution of V_{n+1} follows from the fact that $(z_j^{V_n})_{j \geq 1}$ is independent of \mathcal{F}_n and is chosen uniformly among the increasing indexings centered at V_n . \square

Proposition 14. *For any $\varepsilon \in (0, 1)$ and $n \geq 1$,*

$$\mathbb{P}(|V_{n+1}| < |V_n| + R_{n+1}(1 - \varepsilon) \mid \mathcal{F}_n) \leq 2K^{-\frac{\varepsilon R_{n+1}}{2}}.$$

Proof. First note that the proposition is trivial for $n = 1$ since $V_1 = o$ and so $|V_2| = R_2$. Assume now that $n > 1$. Conditionally on \mathcal{F}_n , V_{n+1} is uniform on the set of vertices $u \in \mathcal{T} \setminus W_n$ such that $d(V_n, u) = R_{n+1}$. Hence,

$$\mathbb{P}(|V_{n+1}| < |V_n| + R_{n+1}(1 - \varepsilon) \mid \mathcal{F}_n) = \frac{\#\{u \in \mathcal{T} \setminus W_n, d(V_n, u) = R_{n+1}, |u| < |V_n| + R_{n+1}(1 - \varepsilon)\}}{\#\{u \in \mathcal{T} \setminus W_n, d(V_n, u) = R_{n+1}\}}. \quad (14)$$

First, recalling that $\mathcal{T}_{V_n} \cap W_n = \{V_n\}$, we get

$$\#\{u \in \mathcal{T} \setminus W_n, d(V_n, u) = R_{n+1}\} \geq \#\{u \in \mathcal{T}_{V_n}, d(V_n, u) = R_{n+1}\} = K^{R_{n+1}}.$$

Now let u be such that $d(u, V_n) = R_{n+1}$ and $|u| < |V_n| + R_{n+1}(1 - \varepsilon)$. Given two vertices u, w , we denote by $w \wedge u$ their common ancestor of maximal height. Using that $2|w \wedge u| = |u| + |w| - d(u, w)$, we get

$$2|V_n \wedge u| = |V_n| + |u| - d(u, V_n) < 2|V_n| - \varepsilon R_{n+1},$$

and so

$$|V_n \wedge u| < |V_n| - \frac{\varepsilon R_{n+1}}{2}.$$

If $|V_n| < \lceil \varepsilon R_{n+1}/2 \rceil$, this is impossible and the set under consideration is empty. Otherwise, let w be the vertex on the path from the root to V_n such that $|w| = |V_n| - \lceil \varepsilon R_{n+1}/2 \rceil$. We get $d(w, u) = d(u, V_n) - d(V_n, w) = R_{n+1} - \lceil \frac{\varepsilon R_{n+1}}{2} \rceil$. Thus,

$$\begin{aligned} & \#\{u \in \mathcal{T} \setminus W_n : d(V_n, u) = R_{n+1}, |u| < |V_n| + R_{n+1}(1 - \varepsilon)\} \\ & \leq \#\left\{u \in \mathcal{T} : d(w, u) = R_{n+1} - \left\lceil \frac{\varepsilon R_{n+1}}{2} \right\rceil\right\} \leq 2K^{(1-\varepsilon/2)R_{n+1}}. \end{aligned}$$

Plugging this in (14), we conclude that

$$\mathbb{P}(|V_{n+1}| < |V_n| + R_{n+1}(1 - \varepsilon) \mid \mathcal{F}_n) \leq \frac{2K^{(1-\varepsilon/2)R_{n+1}}}{K^{R_{n+1}}} = 2K^{-\frac{\varepsilon R_{n+1}}{2}}.$$

□

Corollary 15. *The following limits hold a.s.:*

$$\lim_{n \rightarrow \infty} \frac{R_n}{n} = \frac{1}{\log K} \quad \text{and} \quad \lim_{n \rightarrow \infty} \frac{|V_n|}{n^2} = \frac{1}{2 \log K}.$$

Proof. The limit of R_n/n is a direct consequence of Lemma 11 and Proposition 12. Moreover, for any fixed $\varepsilon \in (0, 1)$, this linear growth of R_n makes the upper bounds in Proposition 14 summable. Hence, by the conditional Borel-Cantelli lemma, there exists a.s. N such that, for $n \geq N$,

$$|V_{n+1}| \geq |V_n| + R_{n+1}(1 - \varepsilon). \tag{15}$$

Thus, for all sufficiently large n ,

$$R_{n+1}(1 - \varepsilon) \leq |V_{n+1}| - |V_n| \leq R_{n+1}.$$

Using the asymptotic

$$\sum_{k=2}^n R_k \sim \frac{n^2}{2 \log K},$$

and letting ε tend to 0 yields

$$\lim_{n \rightarrow \infty} \frac{|V_n|}{n^2} = \frac{1}{2 \log K} \quad \text{a.s.}$$

□

3 Study of \mathcal{G} without seeds

The aim of this section is to prove Theorem 1. We do not follow the order of the statements. Instead, the proof is organized around the geometry of ancestral paths: we first show that each component has an asymptotic direction, then use local constructions to produce many components, and finally prove that distinct components have distinct directions and that descendant sets are finite. More precisely, we proceed as follows:

- Item 3: we prove that each connected component of \mathcal{G} admits a limiting direction;
- Items 1 and 4: we prove that \mathcal{G} has infinitely many connected components, that for every subtree \mathcal{T}_v there exists a component whose limiting direction belongs to $\partial \mathcal{T}_v$, and we derive a lower bound on the number of components intersecting the sphere of radius n ;
- Item 5: we show that two distinct connected components almost surely have distinct limiting directions;
- Item 2: we prove that every vertex almost surely has finitely many descendants.

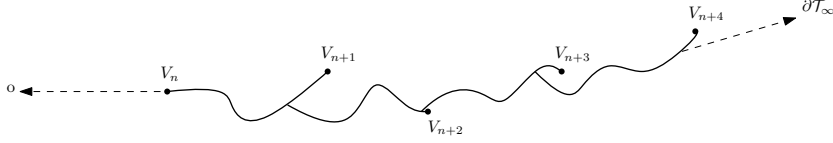


Figure 3: Illustration of the geometry of an ancestral path for large n .

3.1 Proof of Theorem 1, Item 3

We begin by proving that each connected component of \mathcal{G} has a limiting direction. Since \mathcal{T} is vertex-transitive, it is sufficient to prove this property for the connected component containing the root.

Lemma 16. *Let $\Gamma(o) = (V_1, V_2, \dots)$ be the ancestral path of the root. Almost surely, for all sufficiently large n ,*

$$|V_{n+1} \wedge V_n| > |V_n \wedge V_{n-1}|,$$

where $V_n \wedge V_{n-1}$ denotes the common ancestor of V_n and V_{n-1} of maximal height.

Proof. Recall that the ancestral path of the root $\Gamma(o) = (V_1, V_2, \dots)$ is defined in Section 2. Recall also the notation $R_n := d(V_n, V_{n-1})$. Note that

$$2|V_n \wedge V_{n+1}| = |V_n| + |V_{n+1}| - d(V_n, V_{n+1}).$$

Corollary 15 states that R_n/n tends a.s. to $1/\log K$, so in particular, for n large enough, we have

$$R_n > \frac{R_{n+1}}{2}.$$

Moreover, we have proved (see (15)) that, for n large enough, it holds that

$$|V_{n+1}| \geq |V_n| + \frac{R_{n+1}}{2}.$$

Thus, a.s., for n large enough, we have

$$2|V_n \wedge V_{n+1}| = |V_n| + |V_{n+1}| - R_{n+1} \geq 2|V_n| - \frac{R_{n+1}}{2} \geq 2|V_{n-1}| + R_n - \frac{R_{n+1}}{2} > 2|V_{n-1}|.$$

Hence

$$|V_n \wedge V_{n+1}| > |V_n \wedge V_{n-1}|.$$

□

Proof of Theorem 1, Item 3. Any connected component of \mathcal{G} is an oriented tree and $u, v \in \mathcal{T}$ are in the same connected component if and only if their ancestral paths coalesce. So, to prove Item 3 of Theorem 1, it is sufficient to find an infinite ray $z = (z_i)_{i \geq 1} \in \partial\mathcal{T}$ such that, for every $i \geq 1$, the vertices of $\Gamma(o)$ eventually belong to \mathcal{T}_{z_i} .

Note that $|V_n \wedge V_{n+1}| > |V_n \wedge V_{n-1}|$ implies that $V_n \wedge V_{n+1} \in \mathcal{T}_{V_n \wedge V_{n-1}}$ (see Figure 3 for an illustration of a path satisfying this condition). Thus, by the previous lemma, there exists an infinite ray $z = (z_i)_{i \geq 1}$ that contains the vertices $V_n \wedge V_{n+1}$ for large n . By construction, for every $i \geq 1$, the vertices of $\Gamma(o)$ eventually belong to \mathcal{T}_{z_i} and so the ancestral path of the root has a limiting direction given by z .

□

3.2 Proof of Theorem 1, Items 1 and 4

We prove here that \mathcal{G} has infinitely many connected components and, in order to control the number of components in each direction of the space, we prove that, with positive probability, a given neighbor of the root is never explored by the ancestral process of the root.

Proposition 17. *Let $v \in \mathcal{T}$ be a neighbor of the root. Let (V_n, W_n) be the ancestral process of the root and $W_\infty := \cup_{n \geq 1} W_n$. Then*

$$\mathbb{P}(v \notin W_\infty) = \mathbb{P}(\mathcal{T}_v \cap W_\infty = \emptyset) > 0.$$

This proposition shows that, with positive probability, it is not necessary to reveal the arrival times of the vertices in \mathcal{T}_v in order to determine the ancestral line of the root. In particular, on this event, the ancestral path of o is unaffected by the presence of a seed anywhere in the subtree \mathcal{T}_v .

We use the following lemma.

Lemma 18. *Let $v \in \mathcal{T}$ be a neighbor of the root and let $n_0 \geq 7$. Let $w_1 := o$, let w_2 be a neighbor of the root distinct from v , and, for $3 \leq i \leq n_0$, let w_i be a vertex of $\mathcal{T}_{w_{i-1}}$ such that $|w_i| = 2|w_{i-1}|$. Set*

$$\mathcal{A} := \{V_i = w_i \text{ and } v \notin W_i \text{ for } 1 \leq i \leq n_0\}.$$

$$\mathcal{B} := \left\{ \forall n > n_0, \frac{n}{2} \leq \log(K)R_n \leq 2n \text{ and } |V_n| \geq |V_{n-1}| + \frac{R_n}{2} \right\}.$$

Then, we have

$$\mathcal{A} \cap \mathcal{B} \subset \{v \notin W_\infty\}.$$

Proof. The event \mathcal{A} ensures that W_{n_0} does not contain v , that $V_{n_0} \in \mathcal{T}_{w_2}$, and that $|V_{n_0}| = 2^{n_0-2} \geq 32$. Assume now that \mathcal{B} also holds and let us prove by induction that $v \notin W_n$ and $V_n \in \mathcal{T}_{w_2}$ for all $n \geq n_0$. For $i \geq n_0 + 1$, the event \mathcal{B} gives

$$|V_i| \geq |V_{i-1}| + \frac{i}{4 \log K},$$

which yields, for $n \geq n_0$,

$$|V_n| \geq |V_{n_0}| + \sum_{i=n_0+1}^n \frac{i}{4 \log K} \geq 2^{n_0-2} + (n(n+1) - n_0(n_0+1)) \frac{1}{8 \log K}.$$

Thus, using that $R_{n+1} \log K \leq 2(n+1)$, we get

$$|V_n| - R_{n+1} \geq 2^{n_0-2} + (n(n+1) - n_0(n_0+1) - 16(n+1)) \frac{1}{8 \log K}.$$

For $n \geq n_0 \geq 7$, the last expression is bounded below by

$$2^{n_0-2} - \frac{16(n_0+1)}{8 \log K} \geq 2^{n_0-2} - \frac{2(n_0+1)}{\log 2} > 0.$$

Since $R_{n+1} = d(V_n, V_{n+1})$ and $W_{n+1} \setminus W_n \subset B(V_n, R_{n+1})$, the inequality $R_{n+1} < |V_n|$ and the induction hypothesis $V_n \in \mathcal{T}_{w_2}$ imply that $B(V_n, R_{n+1}) \subset \mathcal{T}_{w_2}$. This implies in particular that $v \notin W_{n+1} \setminus W_n$ and $V_{n+1} \in \mathcal{T}_{w_2}$, and so, by induction, that $v \notin W_\infty$. \square

Proof of Proposition 17. Choose $n_0 \geq 7$ large enough so that, for every $n > n_0$,

$$(K+1)e^{n/2} \leq 2^n \leq 3^n \leq \frac{1}{K}e^{2n}.$$

It remains to prove that $\mathbb{P}(\mathcal{A} \cap \mathcal{B}) > 0$, where \mathcal{A} and \mathcal{B} are defined in Lemma 18 with this choice of n_0 . Recall the definition of the filtration $\mathcal{F}_n := \sigma((I_k)_{k \geq 1}, ((z_j^{V_i})_{j \geq 1})_{i < n})$ given in (13) and note that $\mathcal{A} \in \mathcal{F}_{n_0}$ and has positive probability. Set $\mathcal{B}_{n_0} := \Omega$ and, for $N \geq n_0 + 1$, define

$$\mathcal{B}_N := \left\{ \forall n_0 < n \leq N, \frac{n}{2} \leq \log(K)R_n \leq 2n \text{ and } |V_n| \geq |V_{n-1}| + \frac{R_n}{2} \right\}.$$

By monotone convergence, $\mathbb{P}(\mathcal{B} | \mathcal{F}_{n_0}) = \lim_{N \rightarrow \infty} \mathbb{P}(\mathcal{B}_N | \mathcal{F}_{n_0})$. Moreover, by the choice of n_0 and using Lemma 11, we see that

$$2^n \leq I_n \leq 3^n \implies (K+1)e^{\frac{n}{2}} \leq I_n \leq \frac{1}{K}e^{2n} \implies e^{\frac{n}{2}} \leq K^{R_n} \leq e^{2n} \implies \frac{n}{2} \leq \log(K)R_n \leq 2n.$$

Then, for $N \geq n_0 + 1$, using that \mathcal{B}_{N-1} and R_N are \mathcal{F}_{N-1} -measurable, we write

$$\begin{aligned}
\mathbb{P}(\mathcal{B}_N \mid \mathcal{F}_{n_0}) &= \mathbb{E}(\mathbf{1}_{\mathcal{B}_{N-1}} \mathbf{1}_{\{\frac{N}{2} \leq \log(K)R_N \leq 2N\}} \mathbb{E}(\mathbf{1}_{\{|V_N| \geq |V_{N-1}| + \frac{R_N}{2}\}} \mid \mathcal{F}_{N-1}) \mid \mathcal{F}_{n_0}) \\
&\geq \mathbb{E}(\mathbf{1}_{\mathcal{B}_{N-1}} \mathbf{1}_{\{\frac{N}{2} \leq \log(K)R_N \leq 2N\}} (1 - 2K^{-\frac{R_N}{4}}) \mid \mathcal{F}_{n_0}) \\
&\geq \mathbb{E}(\mathbf{1}_{\mathcal{B}_{N-1}} \mathbf{1}_{\{\frac{N}{2} \leq \log(K)R_N \leq 2N\}} (1 - 2K^{-\frac{N}{8 \log K}}) \mid \mathcal{F}_{n_0}) \\
&\geq (1 - 2K^{-\frac{N}{8 \log K}}) \mathbb{E}(\mathbf{1}_{\mathcal{B}_{N-1}} \mathbf{1}_{\{2^N \leq I_N \leq 3^N\}} \mid \mathcal{F}_{n_0}) \\
&= (1 - 2K^{-\frac{N}{8 \log K}}) \mathbf{1}_{\{2^N \leq I_N \leq 3^N\}} \mathbb{P}(\mathcal{B}_{N-1} \mid \mathcal{F}_{n_0}).
\end{aligned}$$

Hence, we get

$$\mathbb{P}(\mathcal{B} \mid \mathcal{F}_{n_0}) \geq \left(\prod_{n > n_0} (1 - 2K^{-\frac{n}{8 \log K}}) \right) \mathbf{1}_{\{\forall n > n_0, 2^n \leq I_n \leq 3^n\}} := c \mathbf{1}_{\{\forall n > n_0, 2^n \leq I_n \leq 3^n\}},$$

for some $c > 0$. Finally, we write

$$\mathbb{P}(\mathcal{A} \cap \mathcal{B}) = \mathbb{E}(\mathbf{1}_{\mathcal{A}} \mathbb{P}(\mathcal{B} \mid \mathcal{F}_{n_0})) \geq c \mathbb{E}(\mathbf{1}_{\mathcal{A}} \mathbf{1}_{\{\forall n > n_0, 2^n \leq I_n \leq 3^n\}}).$$

Since the event \mathcal{A} only depends on the value of I_k for $k \leq n_0$ and on the indexing $(z_j^{V_k})_{j \geq 1}$ for $k < n_0$, using Proposition 12, we conclude that $\mathbb{P}(\mathcal{A} \cap \mathcal{B}) > 0$, which, in view of Lemma 18, implies that $\mathbb{P}(v \notin W_\infty) > 0$. \square

Proof of Theorem 1, Items 1 and 4. We now prove that \mathcal{G} has infinitely many connected components. Since we need to consider the ancestral paths of several vertices at the same time, we use the construction of $W_n(v)$ given in (7) in Section 2.1 (and not the one given by (9)). Since \mathcal{T} is transitive, Proposition 17 implies that, for $v \in \mathcal{T} \setminus \{o\}$, if \overleftarrow{v} denotes the father of v for the tree structure \mathcal{T} , then

$$c := \mathbb{P}(\overleftarrow{v} \notin W_\infty(v)) > 0.$$

Let us define

$$n_0(v) := \sup\{n \geq 1 : \overleftarrow{v} \notin W_n(v)\} \in \mathbb{N} \cup \{\infty\}.$$

We define $W'(v)$ as the set of vertices explored until we explore \overleftarrow{v} , that is

$$W'(v) = \begin{cases} W_{n_0(v)}(v) \cup \{z_j^{V_{n_0(v)}} : j \leq j_0\} & \text{if } n_0(v) < \infty, \\ W_\infty(v) & \text{if } n_0(v) = \infty. \end{cases},$$

where j_0 is the unique index where $z_{j_0}^{V_{n_0(v)}} = \overleftarrow{v}$.

Let us note that the random sets $W'(v)$ for $v \in S(o, r)$ are independent. Indeed, by construction, $W'(v)$ only depends on the arrival times τ_w and the random indexing $(z_j^w)_{j \geq 1}$ for $w \in \mathcal{T}_v$ and we have $\mathcal{T}_v \cap \mathcal{T}_{v'} = \emptyset$ if v, v' are two distinct vertices of $S(o, r)$. Moreover, we have

$$\{\text{The set } W'(v) \text{ is infinite}\} = \{n_0(v) = \infty\} = \{\overleftarrow{v} \notin W_\infty(v)\} \subset \{\Gamma(v) \text{ remains in } \mathcal{T}_v\}.$$

In particular, the number N_r of connected components of \mathcal{G} intersecting the sphere $S(o, r)$ is lower bounded by the number of vertices $v \in S(o, r)$ such that $n_0(v) = \infty$. Since these events are i.i.d. for $v \in S(o, r)$ and have positive probability c , we get

$$\liminf_{r \rightarrow \infty} \frac{N_r}{\#S(o, r)} \geq c \text{ a.s.}$$

This proves Item 1 of Theorem 1. Furthermore, the same argument shows that, for every $v \in \mathcal{T} \setminus \{o\}$, there exist infinitely many vertices $u \in \mathcal{T}_v$ such that $n_0(u) = \infty$. In particular, there exists a $u \in \mathcal{T}_v$ such that $\Gamma(u)$ remains in $\mathcal{T}_u \subset \mathcal{T}_v$ and therefore the asymptotic direction of the genealogical component containing u is in $\partial \mathcal{T}_v$. This proves Item 4 of Theorem 1. \square

3.3 Proof of Theorem 1, Item 5

In this section, we show that two distinct connected components cannot share the same asymptotic direction. In particular, for any $u, v \in \mathcal{T}$, either $\Gamma(u)$ and $\Gamma(v)$ ultimately coalesce or their asymptotic distance tends to infinity. We first prove the following lemma concerning the ancestral process of the root.

Lemma 19. *Let $(V_n, W_n)_{n \geq 1}$ be the ancestral process of the root. Let $\mathcal{H}_n = \sigma(V_k, W_k, k \leq n)$ be the natural filtration of the exploration. There exists a constant $c > 0$ such that, for any $n \geq 1$,*

$$\mathbb{P}(\forall m \geq 0, V_{n+m} \in \mathcal{T}_{V_n} \mid \mathcal{H}_n) \geq c \text{ a.s.}$$

Note that this result implies that for infinitely many n , $(V_{n+m})_{m \geq 0}$ remains in \mathcal{T}_{V_n} . In particular, we recover that *dir* $\Gamma(o)$ exists a.s. (Item 3 of Theorem 1).

Proof. For $u \in \mathcal{T}$, define the function h_u on \mathcal{T} by

$$h_u(w) := \begin{cases} d(u, w) & \text{if } w \in \mathcal{T}_u \\ -1 & \text{otherwise} \end{cases}.$$

Hence, for $w \in \mathcal{T}_u$, $h_u(w)$ is the height of w in \mathcal{T}_u . As before, we construct the process $(V_n, W_n)_{n \geq 1}$ from the inter-arrival times of a record process $(I_k)_{k \geq 1}$ and random indexings $((z_j^v)_{j \geq 1}, v \in \mathcal{T})$. Recall that $(\mathcal{F}_n)_{n \geq 1}$ denotes the filtration given by

$$\mathcal{F}_n := \sigma((I_k)_{k \geq 1}, ((z_j^{V_i})_{j \geq 1})_{i < n}).$$

Recall also that the distance $R_{k+1} = d(V_{k+1}, V_k)$ and (V_k, W_k) are \mathcal{F}_k -measurable, and that, knowing \mathcal{F}_k , V_{k+1} is uniform on $\{u \in \mathcal{T} \setminus W_k : d(V_k, u) = R_{k+1}\}$. Since $W_k \cap \mathcal{T}_{V_k} = \{V_k\}$, we also have

$$\frac{\#(S(V_k, R_{k+1}) \cap \mathcal{T}_{V_k} \setminus W_k)}{\#(S(V_k, R_{k+1}) \setminus W_k)} \geq \frac{\#(S(V_k, R_{k+1}) \cap \mathcal{T}_{V_k})}{\#(S(V_k, R_{k+1}))} = \frac{K}{K+1}.$$

Thus, for $n \geq 1$ and $m \geq 0$, we find that

$$\begin{aligned} \mathbb{P}(h_{V_n}(V_{n+m+1}) = h_{V_n}(V_{n+m}) + R_{n+m+1} \mid \mathcal{F}_{n+m}) &\geq \mathbb{P}(V_{n+m+1} \in \mathcal{T}_{V_{n+m}} \mid \mathcal{F}_{n+m}) \mathbf{1}_{\{V_{n+m} \in \mathcal{T}_{V_n}\}} \\ &\geq \frac{K}{K+1} \mathbf{1}_{\{V_{n+m} \in \mathcal{T}_{V_n}\}}. \end{aligned}$$

Let us now show that for $i \geq 1$, on the event $\{V_{n+m} \in \mathcal{T}_{V_n}\}$, we have

$$\mathbb{P}(h_{V_n}(V_{n+m+1}) \leq h_{V_n}(V_{n+m}) - i \mid \mathcal{F}_{n+m}) \leq K^{-i}.$$

We can assume that $h_{V_n}(V_{n+m}) - i \geq -1$ since the result is trivially true for larger i . If $i \leq h_{V_n}(V_{n+m}) = d(V_{n+m}, V_n)$, let z be the unique vertex on the path $[V_n, V_{n+m}]$ such that $d(V_{n+m}, z) = i$. If $i = h_{V_n}(V_{n+m}) + 1$, let z be the father of V_n for the tree structure (see Figure 4 for an illustration of the notation). On the event $h_{V_n}(V_{n+m+1}) \leq h_{V_n}(V_{n+m}) - i$, observe that $z \in [V_{n+m+1}, V_{n+m}]$ and so

$$V_{n+m+1} \in S(z, R_{n+m+1} - i) \setminus \mathcal{T}_{z \rightarrow V_{n+m}},$$

where $\mathcal{T}_{z \rightarrow V_{n+m}}$ denotes the connected component of $\mathcal{T} \setminus \{z\}$ containing V_{n+m} . Thus,

$$\mathbb{P}(h_{V_n}(V_{n+m+1}) \leq h_{V_n}(V_{n+m}) - i \mid \mathcal{F}_{n+m}) \leq \frac{\#(S(z, R_{n+m+1} - i) \setminus \mathcal{T}_{z \rightarrow V_{n+m}})}{\#(S(V_{n+m}, R_{n+m+1}) \cap \mathcal{T}_{V_{n+m}})} = K^{-i}.$$

Now, fix $M \geq 1$ such that the law μ on \mathbb{Z} defined by

$$\mu(\{M\}) = \frac{K-1}{K} \quad \text{and} \quad \mu((-\infty, -i]) = K^{-i} \text{ for } i \geq 1 \quad (16)$$

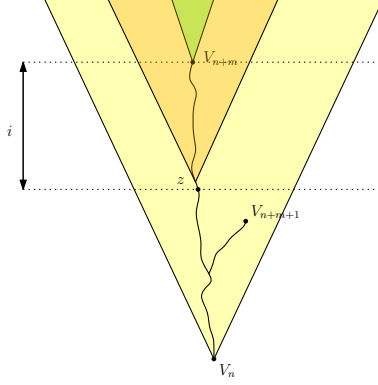


Figure 4: Illustration of the notation used in the proof of Lemma 19. In yellow: the tree \mathcal{T}_{V_n} . In orange: the tree $\mathcal{T}_{z \rightarrow V_{n+m}}$.

has positive expectation. Then, according to Lemma 11, on the event $\{I_{n+m+1} \geq K^{M+1}, m \geq 0\}$, we get $R_{n+m+1} \geq M$ for $m \geq 0$ and so, if $h_{V_n}(V_{n+m}) \geq 0$, we get the stochastic domination:

$$\mathcal{L}(h_{V_n}(V_{n+m+1}) - h_{V_n}(V_{n+m}) \mid \mathcal{F}_{n+m}) \geq_{st} \mu.$$

Let $(Y_k)_{k \geq 0}$ be an i.i.d. sequence with distribution μ . By induction, we get

$$\mathbb{P}(\forall m \geq 0, h_{V_n}(V_{n+m}) \geq 0 \mid \mathcal{F}_n) \geq \mathbb{P}(\forall m \geq 0, \sum_{k=0}^m Y_k \geq 0) 1_{\{I_{n+m+1} \geq K^{M+1}, m \geq 0\}}.$$

Recalling that $h_{V_n}(V_{n+m}) \geq 0$ if and only if $V_{n+m} \in \mathcal{T}_{V_n}$ and $\mathcal{H}_n \subset \mathcal{F}_n$, we get

$$\mathbb{P}(\forall m \geq 0, V_{n+m} \in \mathcal{T}_{V_n} \mid \mathcal{H}_n) \geq \mathbb{P}(\forall m \geq 0, \sum_{k=0}^m Y_k \geq 0) \mathbb{P}(I_{n+m+1} \geq K^{M+1}, m \geq 0 \mid \mathcal{H}_n).$$

The first probability on the right-hand side is positive since $(Y_k)_{k \geq 0}$ are i.i.d. random variables with positive expectation. For the second probability, we have

$$\begin{aligned} \mathbb{P}(I_{n+m+1} \geq K^{M+1}, m \geq 0 \mid \mathcal{H}_n) &= \mathbb{P}(I_{n+m+1} \geq K^{M+1}, m \geq 0 \mid I_1, \dots, I_n) \\ &\geq \mathbb{P}(I_{m+2} \geq K^{M+1}, m \geq 0) > 0, \end{aligned}$$

where we used Proposition 12, Item (iii) for the inequality. \square

In order to prove Item 5 of Theorem 1, which states that two distinct connected components of \mathcal{G} cannot share the same asymptotic direction, we need to consider the ancestral lines of two distinct vertices at the same time. To this end, we introduce a coupled exploration process $(U_n, V_n, W_n)_{n \geq 1}$ that simultaneously constructs the ancestral lines $\Gamma(u)$ and $\Gamma(v)$ of two vertices u, v in a specified order extending the construction of the genealogical process of Section 2.1. This coupled exploration process can, once again, be encoded in terms of a classical record process.

As in Section 2.1, we proceed by induction. Set $(U_1, V_1) := (u, v)$ and $W_1 := \{u, v\}$. Assume that U_n, V_n and W_n have been constructed. Then, there are three cases: if $\tau_{V_n} > \tau_{U_n}$, we consider the increasing uniform indexing $(z_j^{V_n})_{j \geq 1}$ associated to V_n and let

$$J_n := \inf\{j \geq 1 : \tau_{z_j^{V_n}} < \tau_{V_n}\} = \inf\{j \geq 1 : z_j^{V_n} \notin W_n \setminus \{U_n\}, \tau_{z_j^{V_n}} < \tau_{V_n}\}.$$

We then declare $z_{J_n}^{V_n}$ to be the direct ancestor of V_n , and we set

$$V_{n+1} := z_{J_n}^{V_n}, \quad U_{n+1} := U_n, \quad W_{n+1} := W_n \cup \{z_j^{V_n} : j \leq J_n\}.$$

It may happen that $V_{n+1} = U_n = U_{n+1}$, which means that the ancestral lines of v and u have just coalesced. In the second case $\tau_{V_n} < \tau_{U_n}$, we perform a symmetric construction interchanging the roles of

V_n and U_n . Finally, if at some time step n we have $\tau_{V_n} = \tau_{U_n}$, then $V_n = U_n$; the paths have previously merged, and we simply continue with the single exploration process described in the previous section. To summarize this in words: the joint ancestral process of u and v is constructed in the same manner as in Section 2.1, and, at each step, we choose to explore the ancestral line of the vertex with the currently largest arrival time. In particular, keeping V_1 and then the indices $k \geq 2$ such that $V_k \neq V_{k-1}$ recovers the ancestral line $\Gamma(v)$ of v (and a similar statement holds for $\Gamma(u)$).

Just as for the single genealogical exploration process of Section 2.1, it is convenient to encode the joint exploration process $(U_n, V_n, W_n)_{n \geq 1}$ in terms of a record process. But now we need to consider the inter-arrival times $(I'_n, n \geq 1)$ of a *lower 2-record process*⁸ instead of a simple record process. We also need a sequence $(\beta_i)_{i \geq 1}$ of i.i.d. uniformly distributed random variables on $\{u, v\}$ and a family $((z_j^w)_{j \geq 1}, w \in \mathcal{T})$ of independent uniform increasing indexings. We assume that all these families of random variables are independent. We define by induction

$$(\hat{U}_1, \hat{V}_1) := (u, v), \quad \hat{W}_1 := \{u, v\}.$$

Assume that $(\hat{U}_n, \hat{V}_n, \hat{W}_n)$ has been constructed and that $\hat{U}_n \neq \hat{V}_n$. If $\beta_n = u$, define \hat{J}_n as the first index such that we explore I'_{n+1} new vertices around \hat{U}_n that do not belong to \hat{W}_n :

$$\hat{J}_n := \inf \left\{ j \geq 1 : \#\{i \leq j : z_i^{\hat{U}_n} \notin \hat{W}_n\} = I'_{n+1} \right\}. \quad (17)$$

Let \hat{L}_n denote the index of \hat{V}_n in the increasing indexing around \hat{U}_n , *i.e.*, such that $z_{\hat{L}_n}^{\hat{U}_n} = \hat{V}_n$.

- If $\hat{J}_n < \hat{L}_n$, then set $\hat{U}_{n+1} = z_{\hat{J}_n}^{\hat{U}_n}$, $\hat{V}_{n+1} = \hat{V}_n$ and $\hat{W}_{n+1} = \hat{W}_n \cup \{z_j^{\hat{U}_n} : j \leq \hat{J}_n\}$.
- Otherwise, set $\hat{U}_{n+1} = \hat{V}_{n+1} = \hat{V}_n$ and $\hat{W}_{n+1} = \hat{W}_n \cup \{z_j^{\hat{U}_n} : j \leq \hat{L}_n\}$.

If $\beta_n = v$, the construction is the same with the roles of \hat{U} and \hat{V} interchanged. This constructs the process $(\hat{U}_n, \hat{V}_n, \hat{W}_n)$ up to the first step n (possibly infinite) such that $\hat{U}_n = \hat{V}_n$.

Lemma 20. *Consider the two processes defined above. Let $\sigma := \inf\{n \geq 1, U_n = V_n\}$ and $\hat{\sigma} := \inf\{n \geq 1, \hat{U}_n = \hat{V}_n\}$. Then $(U_n, V_n, W_n)_{n \leq \sigma}$ and $(\hat{U}_n, \hat{V}_n, \hat{W}_n)_{n \leq \hat{\sigma}}$ have the same distribution. Moreover, on the event that $\hat{\sigma} > n$, we have*

$$I'_n = \#\left(\hat{W}_n \setminus \hat{W}_{n-1}\right), \quad \text{with the convention } \hat{W}_0 := \emptyset.$$

Proof. We argue by induction before the coalescence time. On the event $\{\sigma > n\}$, the variables τ_{U_n} and τ_{V_n} are the two smallest elements of $\{\tau_w : w \in W_n\}$. The next step explores the endpoint with the larger of these two arrival times, and new vertices are revealed until either the other endpoint is met or a new vertex has arrival time smaller than $\max(\tau_{U_n}, \tau_{V_n})$. Thus, if one ignores the possible coalescence with the other endpoint, the number of new vertices that have to be inspected is exactly an inter-arrival time of a lower 2-record process.

It remains to identify the random choice of which endpoint is explored at each step. For a lower 2-record process, conditionally on the current second record value, say b , the current minimum and the new observation which falls below b are independent and have the same law: they are distributed as the arrival-time law conditioned to be smaller than b . Since this law is atomless, the new observation is smaller than the current minimum with probability 1/2, and larger with probability 1/2. Thus, after each 2-record, the line carrying the larger of the two active arrival times is equally likely to be the u -line or the v -line. The same conditional independence also gives that these choices are i.i.d. and independent of the inter-arrival times and of the uniform increasing indexings. The recursive rules in the construction of $(\hat{U}_n, \hat{V}_n, \hat{W}_n)$ reproduce exactly this mechanism, with the variables (I'_n) giving the 2-record inter-arrival times and the variables (β_n) choosing which endpoint is currently the larger one. If the other

⁸This means that a new record is reached when we observe a value that is smaller than the second smallest value observed so far. More precisely, let $(X_i)_{i \geq 1}$ be i.i.d. with an atomless distribution, set $T_0 := 0, T_1 := 2$, and, for $n \geq 1$, $T_{n+1} := \inf\{k > T_n : X_k < \min_2\{X_1, \dots, X_{T_n}\}\}$, where \min_2 denotes the second smallest element. Then (T_n) is the sequence of record times of the lower 2-record process, and the differences $I'_n := T_n - T_{n-1}$ are the inter-arrival times. This corresponds to the $k = 2$ case of the general k -record process; see [6, 2].

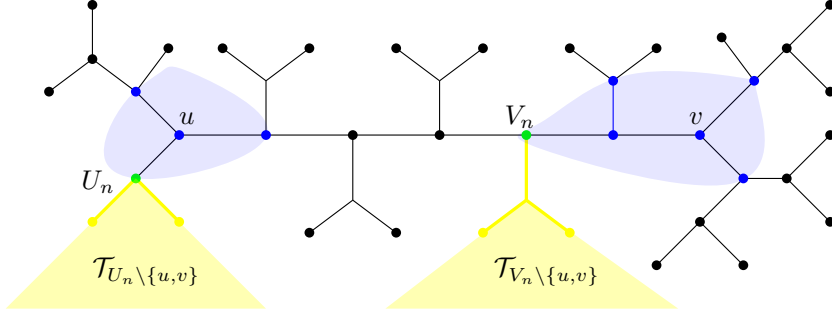


Figure 5: Illustration of the double exploration process on a binary tree. The set \hat{W}_n is represented in blue. The subtrees $\mathcal{T}_{\hat{U}_n \setminus \{u,v\}}$ and $\mathcal{T}_{\hat{V}_n \setminus \{u,v\}}$ are in yellow.

endpoint is encountered before the next 2-record, both constructions coalesce. This proves the equality in distribution up to the coalescence time. Finally, on $\{\hat{\sigma} > n\}$ no coalescence has occurred before time n , and the n -th increment of the explored set is precisely the n -th 2-record inter-arrival time; in particular $I'_1 = 2 = \#\hat{W}_1$ and, for $n \geq 2$, $I'_n = \#(\hat{W}_n \setminus \hat{W}_{n-1})$. \square

Proof of Theorem 1, Item 5. One easily checks that, for any $n \geq 1$, \hat{W}_n either has two connected components in \mathcal{T} , one containing u and the other containing v , or consists of a unique connected component containing both u and v .

For $w \in \mathcal{T}$, define $\mathcal{T}_{w \setminus \{u,v\}}$ to be the subtree of \mathcal{T} composed of $\{w\}$ and all the connected components of $\mathcal{T} \setminus \{w\}$ that contain neither u nor v . Observe that $\mathcal{T}_{w \setminus \{u,v\}}$ is a regular $(K+1)$ -ary tree except at w , whose degree is either $K-1$ or K depending on whether u or v are in the same component of $\mathcal{T} \setminus \{w\}$ or not (see Figure 5 for an illustration).

Since each step of the double exploration is performed from the active endpoint with the larger arrival time, the same argument as in the simple exploration process shows that

$$\hat{W}_n \cap \mathcal{T}_{\hat{U}_n \setminus \{u,v\}} = \{\hat{U}_n\} \quad \text{and} \quad \hat{W}_n \cap \mathcal{T}_{\hat{V}_n \setminus \{u,v\}} = \{\hat{V}_n\},$$

and that the sets $\mathcal{T}_{\hat{U}_n \setminus \{u,v\}}$ and $\mathcal{T}_{\hat{V}_n \setminus \{u,v\}}$ are disjoint as soon as $\hat{U}_n \neq \hat{V}_n$. Define the filtrations

$$\begin{aligned} \mathcal{F}'_n &:= \sigma((I'_k)_{k \geq 1}, (\beta_k)_{k \geq 1}, ((z_j^{\hat{U}_i})_{j \geq 1})_{i < n}, ((z_j^{\hat{V}_i})_{j \geq 1})_{i < n}) \\ \mathcal{H}'_n &:= \sigma((\hat{U}_k, \hat{V}_k, \hat{W}_k), k \leq n), \end{aligned}$$

and the event A_n that, starting from time n , the two explorations remain in their respective sectors:

$$A_n := \left\{ \forall m \geq 0, \hat{U}_{n+m} \in \mathcal{T}_{\hat{U}_n \setminus \{u,v\}}, \hat{V}_{n+m} \in \mathcal{T}_{\hat{V}_n \setminus \{u,v\}} \right\}.$$

Fix $M \geq 1$ such that the law μ defined in (16) has positive expectation. Using an argument similar to the proof of Lemma 19, we now find that

$$\mathbb{P}(\hat{\sigma} < \infty \text{ or } A_n \mid \mathcal{F}'_n) \geq \mathbb{P}\left(\forall m \geq 1, \sum_{k=0}^m Y'_k > 0\right)^2 \mathbf{1}_{\{I'_{n+m+1} \geq K^{M+1}, m \geq 0\}},$$

where $(Y'_k)_{k \geq 0}$ are independent random variables⁹ such that

$$\mathbb{P}(Y'_0 = M) = \frac{K-1}{K+1} \quad \text{and} \quad \mathbb{P}(Y'_0 = -1) = \frac{2}{K+1},$$

and, for $k \geq 1$, Y'_k has law μ defined by (16). We have $\mathbb{P}(\forall m \geq 0, \sum_{k=0}^m Y'_k \geq 0) > 0$ because $\mathbb{P}(Y'_0 = M) > 0$ and the subsequent random walk has increment law μ with positive drift. The proof of Proposition 12 adapts directly to the lower 2-record process and gives

$$\mathbb{P}(I'_{n+m+1} \geq K^{M+1}, m \geq 0 \mid \mathcal{H}'_n) \geq \mathbb{P}(I'_{m+2} \geq K^{M+1}, m \geq 0) > 0.$$

⁹ Y'_0 has this particular distribution because the vertex \hat{U}_n may have degree $K-1$ in the tree $\mathcal{T}_{\hat{U}_n \setminus \{u,v\}}$.

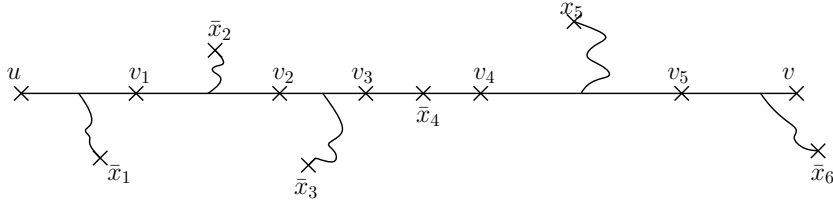


Figure 6: Illustration of a sequence $(\bar{x}_1, \dots, \bar{x}_6) \in A(u, v, 6, r_1, \dots, r_6)$.

Thus, we deduce the existence of $c' > 0$ such that, for any $n \geq 1$, a.s.

$$\mathbb{P}(\hat{\sigma} < \infty \text{ or } A_n \mid \mathcal{H}'_n) \geq c'.$$

Finally, a standard restart argument at the successive exit times from the current sectors gives that

$$\mathbb{P}(\hat{\sigma} < \infty \text{ or } A_n \text{ for some } n) = 1.$$

By Lemma 20, the same dichotomy holds for the original coupled exploration process. So either $\sigma < \infty$ and the two ancestral paths $\Gamma(u)$ and $\Gamma(v)$ eventually coalesce, or they eventually remain in disjoint subtrees, and therefore $\text{dir } \Gamma(u) \neq \text{dir } \Gamma(v)$. \square

3.4 Proof of Theorem 1, Item 2

Let $u \in \mathcal{T}$. Recall that $p(u)$ denotes the direct ancestor of u in the genealogical graph (*i.e.*, the vertex from which u inherits its color) and that $D(u)$ stands for the set of descendants of u (*i.e.*, the set of vertices whose color inheritance can be traced back to u). Rigorously, this means that, given $v \in D(u)$, there exists a finite sequence of vertices (x_0, x_1, \dots, x_n) such that

- (i) $x_0 = u$ and $x_n = v$;
- (ii) for $0 \leq i \leq n-1$, $p(x_{i+1}) = x_i$.

Points (i) and (ii) imply in particular that

- (a) $\tau_u < \tau_{x_1} < \dots < \tau_{x_{n-1}} < \tau_v$;
- (b) for $0 \leq i \leq n-1$, the ball $B(x_{i+1}, d(x_i, x_{i+1}) - 1)$ contains no vertices with arrival time smaller than $\tau_{x_{i+1}} > \tau_u$.

In this section, we prove that $D(u)$ is a.s. finite. To do so, we upper bound the probability that points (a) and (b) occur. The difficulty is that the balls in point (b) may overlap heavily. We therefore extract below a subsequence $(x_{i_1}, \dots, x_{i_m})$ and integers r_1, \dots, r_m such that the balls $B(x_{i_j}, r_j - 1)$ are pairwise disjoint, contain no colored vertex at time τ_u , and still control the total displacement from u to v .

To make this idea precise, we use the following notation: for $u, v \in \mathcal{T}$, $m \geq 1$ and positive integers r_1, \dots, r_m , we define $A(u, v, m, r_1, \dots, r_m)$ to be the set of sequences of vertices $(\bar{x}_1, \dots, \bar{x}_m)$ for which one can find vertices (v_1, \dots, v_{m-1}) on the path $[u, v]$ such that, setting $v_0 := u$, the vertices v_0, v_1, \dots, v_{m-1} are ordered by strictly increasing distance from u and we have

- $d(\bar{x}_i, v_i) = d(\bar{x}_i, v_{i-1}) = r_i$ for $1 \leq i \leq m-1$,
- $d(\bar{x}_m, v_{m-1}) = r_m$ and $d(\bar{x}_m, v) \leq r_m$.

See Figure 6 for an illustration.

Lemma 21. *Let $v \in D(u) \setminus \{u\}$ and let $(x_0 := u, x_1, \dots, x_n := v)$ be such that $p(x_{i+1}) = x_i$ for every $0 \leq i \leq n-1$. Then there exist $m \geq 1$, integers $r_1, \dots, r_m \geq 1$, and indices $1 \leq i_1 < \dots < i_m \leq n$ such that*

- $(x_{i_1}, \dots, x_{i_m}) \in A(u, v, m, r_1, \dots, r_m)$.

- At time τ_u , the union of balls $\bigcup_{j=1}^m B(x_{i_j}, r_j - 1)$ does not contain a colored vertex.

Proof. We construct inductively, for $j \geq 1$, a subsequence $(x_{i_1}, \dots, x_{i_j})$, a sequence of positive integers (r_1, \dots, r_j) and a sequence of vertices (v_1, \dots, v_{j-1}) such that

- (i) The vertices v_1, \dots, v_{j-1} lie on the path $]u, v[$ and are ordered by increasing distance from u .
- (ii) The vertices v_1, \dots, v_{j-1} witness that $(x_{i_1}, \dots, x_{i_j}) \in A(u, x_{i_j}, j, r_1, \dots, r_j)$.
- (iii) At time τ_u , the union of balls $\bigcup_{h=1}^j B(x_{i_h}, r_h - 1)$ contains no colored vertex.

Let us start with $j = 1$. Removing the $K + 1$ edges around u , we obtain $K + 1$ disjoint subtrees of \mathcal{T} . We denote by $\mathcal{T}_{u \rightarrow v}$ the one containing the vertex v . Observe that $x_1 \in \mathcal{T}_{u \rightarrow v}$. Indeed, set

$$I := \inf\{i \geq 1, x_i \in \mathcal{T}_{u \rightarrow v}\},$$

which is well defined since $x_n = v \in \mathcal{T}_{u \rightarrow v}$ and hence $I \leq n$. Moreover, since x_I inherited its color from x_{I-1} , this implies that $d(x_I, x_{I-1}) \leq d(x_I, u)$. So, either $x_{I-1} = u$, *i.e.*, $x_1 \in \mathcal{T}_{u \rightarrow v}$, or $x_{I-1} \in \mathcal{T}_{u \rightarrow v}$, but this would contradict the definition of I . Therefore $I = 1$. We set

$$i_1 = 1 \quad r_1 := d(u, x_1).$$

Of course, we have

$$(x_{i_1}) \in A(u, x_{i_1}, 1, r_1).$$

Since x_1 inherits its color from $u = x_0$, it implies that the ball $B(x_{i_1}, r_1 - 1)$ contains no colored vertices at time τ_u . If $d(x_1, v) \leq r_1$, we set $m = 1$ and the sequence (x_1) satisfies the condition of Lemma 21.

Assume now that $(x_{i_1}, \dots, x_{i_j})$, a sequence of positive integers (r_1, \dots, r_j) and a sequence of vertices (v_1, \dots, v_{j-1}) have been constructed such that (i), (ii), (iii) hold. If $d(v, x_{i_j}) \leq r_j$, then we set $m = j$ and we have as required

- $(x_{i_1}, \dots, x_{i_m}) \in A(u, v, m, r_1, \dots, r_m)$.
- At time τ_u , the union of balls $\bigcup_{h=1}^m B(x_{i_h}, r_h - 1)$ does not contain a colored vertex.

Otherwise, $d(v, x_{i_j}) > r_j$ and, with the convention $v_0 := u$, we define v_j as the unique vertex on $]v_{j-1}, v[$ such that $d(x_{i_j}, v_j) = r_j$. Removing the $K + 1$ edges around v_j , we obtain $K + 1$ disjoint subtrees of \mathcal{T} and we denote by $\mathcal{T}_{v_j \rightarrow v}$ the one containing the vertex v . Let

$$i_{j+1} := \inf\{i > i_j, x_i \in \mathcal{T}_{v_j \rightarrow v}\} \quad r_{j+1} := d(v_j, x_{i_{j+1}}).$$

As before, note that $x_n = v \in \mathcal{T}_{v_j \rightarrow v}$ so i_{j+1} is well defined and $i_{j+1} \leq n$. Moreover, note that

$$d(x_{i_{j+1}}, x_{i_{j+1}-1}) \geq d(x_{i_{j+1}}, v_j) = r_{j+1}.$$

Since $x_{i_{j+1}}$ inherited its color from $x_{i_{j+1}-1}$, this implies that the ball $B(x_{i_{j+1}}, r_{j+1} - 1)$ contains no colored vertex at time τ_u . Moreover, by construction $(x_{i_1}, \dots, x_{i_{j+1}}) \in A(u, x_{i_{j+1}}, j + 1, r_1, \dots, r_{j+1})$. Together with the induction hypothesis, this gives (i)–(iii) at rank $j + 1$. Since the indices i_j are strictly increasing and are bounded by n , the induction stops after finitely many steps. \square

Proof of Theorem 1, Item 2. Fix $u, v \in \mathcal{T}$ with $u \neq v$ and let $\ell := d(u, v)$. To lighten the notation, we use \bar{x}_m to denote a sequence of vertices $(\bar{x}_1, \dots, \bar{x}_m)$ and \mathbf{r}_m to denote a sequence of positive integers (r_1, \dots, r_m) . Let us also define the event

$$\mathcal{B}(\bar{\mathbf{x}}_m, \mathbf{r}_m) := \{\text{At time } \tau_u, \bigcup_{j=1}^m B(\bar{x}_j, r_j - 1) \setminus \{\bar{x}_1, \dots, \bar{x}_m\} \text{ does not contain a colored vertex}\}.$$

Using the previous lemma, we know that if $v \in D(u)$, there exist $m \geq 1$, $\mathbf{r}_m \geq 1$, $\bar{\mathbf{x}}_m \in A(u, v, m, \mathbf{r}_m)$ such that

- $\tau_{\bar{x}_1} < \dots < \tau_{\bar{x}_m}$;
- $\mathcal{B}(\bar{\mathbf{x}}_m, \mathbf{r}_m)$ occurs.

For fixed $\bar{\mathbf{x}}_m$, conditionally on τ_u , the event $\mathcal{B}(\bar{\mathbf{x}}_m, \mathbf{r}_m)$ only involves arrival times outside $\{\bar{x}_1, \dots, \bar{x}_m\}$, whereas the event $\{\tau_{\bar{x}_1} < \dots < \tau_{\bar{x}_m}\}$ only involves the arrival times of $\bar{x}_1, \dots, \bar{x}_m$. These two events are therefore independent, and the latter has probability $1/m!$. Thus, we get

$$\begin{aligned} \mathbb{P}(v \in D(u) \mid \tau_u) &\leq \sum_{m \geq 1} \sum_{\mathbf{r}_m \geq 1} \sum_{\bar{\mathbf{x}}_m \in A(u, v, m, \mathbf{r}_m)} \mathbb{P}(\tau_{\bar{x}_1} < \dots < \tau_{\bar{x}_m}, \mathcal{B}(\bar{\mathbf{x}}_m, \mathbf{r}_m) \mid \tau_u) \\ &= \sum_{m \geq 1} \sum_{\mathbf{r}_m \geq 1} \sum_{\bar{\mathbf{x}}_m \in A(u, v, m, \mathbf{r}_m)} \frac{1}{m!} \mathbb{P}(\mathcal{B}(\bar{\mathbf{x}}_m, \mathbf{r}_m) \mid \tau_u). \end{aligned} \quad (18)$$

Let us note that if $\bar{\mathbf{x}}_m \in A(u, v, m, \mathbf{r}_m)$, the balls $(B(\bar{x}_j, r_j - 1))_{1 \leq j \leq m}$ are pairwise disjoint. Indeed, for $i < j$, the ball $B(\bar{x}_i, r_i - 1)$ is contained in $\mathcal{T}_{v_i \rightarrow u}$ whereas $B(\bar{x}_j, r_j - 1)$ is contained in $\mathcal{T}_{v_i \rightarrow v}$. Moreover, we have the rough lower bound $\sharp B(\bar{x}_i, r_i - 1) \geq K^{r_i - 1}$ so we get

$$\sharp \left(\bigcup_{j=1}^m B(\bar{x}_j, r_j - 1) \setminus \{\bar{x}_1, \dots, \bar{x}_m\} \right) \geq \sum_{i=1}^m K^{r_i - 1} - m.$$

This yields

$$\mathbb{P}(\mathcal{B}(\bar{\mathbf{x}}_m, \mathbf{r}_m) \mid \tau_u) \leq \mathbb{P}(\tau > \tau_u \mid \tau_u) \sum_{i=1}^m K^{r_i - 1} - m. \quad (19)$$

We now give an upper bound on the number of sequences $\bar{\mathbf{x}}_m \in A(u, v, m, \mathbf{r}_m)$. First note that

$$\ell = d(u, v) \leq 2 \sum_{i=1}^m r_i.$$

Hence

$$\sharp A(u, v, m, \mathbf{r}_m) = 0 \quad \text{if } 2 \sum_{i=1}^m r_i < \ell. \quad (20)$$

Otherwise, by definition, \bar{x}_1 must lie in the branch from u toward v , so, since $d(\bar{x}_1, u) = r_1$, there are at most K^{r_1} choices for \bar{x}_1 . Once the vertex \bar{x}_1 is chosen, v_1 is entirely determined. The same argument applies recursively from v_i toward v . Thus, using that $d(\bar{x}_2, v_1) = r_2$, we obtain that there remain at most K^{r_2} choices for \bar{x}_2 and so on. This yields

$$\sharp A(u, v, m, \mathbf{r}_m) \leq K^{\sum_{i=1}^m r_i}. \quad (21)$$

As already remarked, the law of the arrival times does not matter (as long as it is atomless), so we may assume here that $(\tau_w)_{w \in \mathcal{T}}$ are i.i.d. uniform in $[0, 1]$. Combining (18), (19), (20) and (21), we get

$$\begin{aligned} \mathbb{P}(v \in D(u) \mid \tau_u) &\leq \sum_{m \geq 1} \sum_{\substack{\mathbf{r}_m \geq 1 \\ 2 \sum_{i=1}^m r_i \geq \ell}} K^{\sum_{i=1}^m r_i} \frac{1}{m!} (1 - \tau_u)^{\sum_{i=1}^m K^{r_i - 1} - m} \\ &= \sum_{m \geq 1} \sum_{k \geq \ell/2m} \sum_{\substack{\|\mathbf{r}_m\|_\infty = k \\ 2\|\mathbf{r}_m\|_1 \geq \ell}} \frac{1}{m!} \prod_{i=1}^m \left(K^{r_i} (1 - \tau_u)^{K^{r_i - 1} - 1} \right). \end{aligned}$$

Summing this upper bound over all the vertices $v \in \mathcal{T}$, we get

$$\begin{aligned} \mathbb{E}[\sharp D(u) \mid \tau_u] &= 1 + \sum_{v \in \mathcal{T} \setminus \{u\}} \mathbb{P}(v \in D(u) \mid \tau_u) \\ &\leq 1 + (K + 1) \sum_{\ell \geq 1} K^\ell \sum_{m \geq 1} \sum_{k \geq \ell/2m} \sum_{\substack{\|\mathbf{r}_m\|_\infty = k \\ 2\|\mathbf{r}_m\|_1 \geq \ell}} \frac{1}{m!} \prod_{i=1}^m \left(K^{r_i} (1 - \tau_u)^{K^{r_i - 1} - 1} \right) \\ &\leq 1 + (K + 1) \sum_{\ell \geq 1} \sum_{m \geq 1} \sum_{k \geq \ell/2m} \sum_{\substack{\|\mathbf{r}_m\|_\infty = k \\ 2\|\mathbf{r}_m\|_1 \geq \ell}} K^{2\|\mathbf{r}_m\|_1} \frac{1}{m!} \prod_{i=1}^m \left(K^{r_i} (1 - \tau_u)^{K^{r_i - 1} - 1} \right) \\ &= 1 + (K + 1) \sum_{\ell \geq 1} \sum_{m \geq 1} \sum_{k \geq \ell/2m} \sum_{\substack{\|\mathbf{r}_m\|_\infty = k \\ 2\|\mathbf{r}_m\|_1 \geq \ell}} \frac{1}{m!} \prod_{i=1}^m \left(K^{3r_i} (1 - \tau_u)^{K^{r_i - 1} - 1} \right). \end{aligned}$$

Note that for any $\tau_u \in (0, 1)$, the function $z \mapsto K^{3z}(1 - \tau_u)^{K^{z-1}-1}$ is bounded on $[0, \infty)$ by a constant $C \geq 1$ (which depends on τ_u). Therefore, we get

$$\begin{aligned}
\mathbb{E}[\sharp D(u) \mid \tau_u] &\leq 1 + (K+1) \sum_{\ell \geq 1} \sum_{m \geq 1} \sum_{k \geq \ell/2m} \sum_{\|\mathbf{r}_m\|_\infty = k} \frac{C^m}{m!} \left(K^{3k}(1 - \tau_u)^{K^{k-1}-1} \right) \\
&\leq 1 + (K+1) \sum_{k \geq 1} \sum_{m \geq 1} \sum_{\ell \leq 2mk} \frac{(kC)^m}{m!} \left(K^{3k}(1 - \tau_u)^{K^{k-1}-1} \right) \\
&= 1 + 2(K+1) \sum_{k \geq 1} \sum_{m \geq 1} k \frac{(kC)^m}{(m-1)!} \left(K^{3k}(1 - \tau_u)^{K^{k-1}-1} \right) \\
&\leq 1 + 2(K+1) \sum_{k \geq 1} k^2 C (e^C K^3)^k (1 - \tau_u)^{K^{k-1}-1} < \infty.
\end{aligned}$$

This proves that $\mathbb{E}[\sharp D(u) \mid \tau_u] < \infty$, and hence $D(u)$ is finite a.s. Since \mathcal{T} is countable, the conclusion holds simultaneously for every $u \in \mathcal{T}$. \square

Remark 22. *The argument of the proof above can be refined to give a faster-than-polynomial tail for the size of the descendant cluster and prove, in particular, that $\mathbb{E}[(\sharp D(u))^p \mid \tau_u] < \infty$ a.s. for every p .*

4 The genealogical graph with seeds

This section is devoted to the proof of Theorem 3. Most statements of the theorem are consequences of Theorem 1 together with the preliminary results obtained in Section 2. The main remaining work is to determine when a seed has infinitely many children.

Consider a finite set $S := \{\tilde{s}_1, \dots, \tilde{s}_M\}$ of seeds. We construct the genealogical graph without seeds, \mathcal{G} , and the one with seeds, $\tilde{\mathcal{G}}$, on the same probability space using:

- the same arrival times τ_v for all $v \notin S$,
- the same random indexing $(z_j^v)_{j \geq 1}$ for every $v \in \mathcal{T}$.

For $v \in \mathcal{T}$, we still denote by $\Gamma(v) = (V_i(v))_{i \geq 1}$ its ancestral path in \mathcal{G} , and by $W_n(v)$ the set of vertices explored to determine $V_n(v)$. We similarly denote by $\tilde{\Gamma}(v) = (\tilde{V}_i(v))_{i \geq 1}$ the ancestral path of v in $\tilde{\mathcal{G}}$. For $v \notin S$, we have the following dichotomy:

- If $W_\infty(v) \cap S = \emptyset$, then $\tilde{\Gamma}(v) = \Gamma(v)$, meaning that the ancestral path $\tilde{\Gamma}(v)$ of v is infinite or, equivalently, that this vertex does not belong to the genealogical component of a seed in $\tilde{\mathcal{G}}$.
- Otherwise, let $n_0 := \inf\{n \geq 1 : W_n(v) \cap S \neq \emptyset\}$. Then $\tilde{\Gamma}(v) = (V_1, \dots, V_{n_0-1}, s)$, where s is the first seed encountered. In this case, v belongs to the genealogical component of the seed s in $\tilde{\mathcal{G}}$ and its ancestral path $\tilde{\Gamma}(v)$ is finite.

Moreover, if v is not a seed, the set of descendants $\tilde{D}(v)$ of v in $\tilde{\mathcal{G}}$ is a subset of $D(v)$, the set of descendants of v in \mathcal{G} . We deduce that, if we write the decomposition of \mathcal{G} into its connected components as

$$\mathcal{G} = \bigsqcup_{i \in I} \mathcal{C}_i,$$

At this stage, some components of \mathcal{G} may a priori be entirely captured by seed components. Thus, the decomposition of $\tilde{\mathcal{G}}$ into connected components satisfies

$$\tilde{\mathcal{G}} = \bigsqcup_{i \in \tilde{I}} \tilde{\mathcal{C}}_i \sqcup \bigsqcup_{j=1}^M \tilde{\mathcal{U}}_j$$

where

- $\tilde{I} \subset I$ and $\tilde{\mathcal{C}}_i$ is an infinite oriented subtree of \mathcal{C}_i .

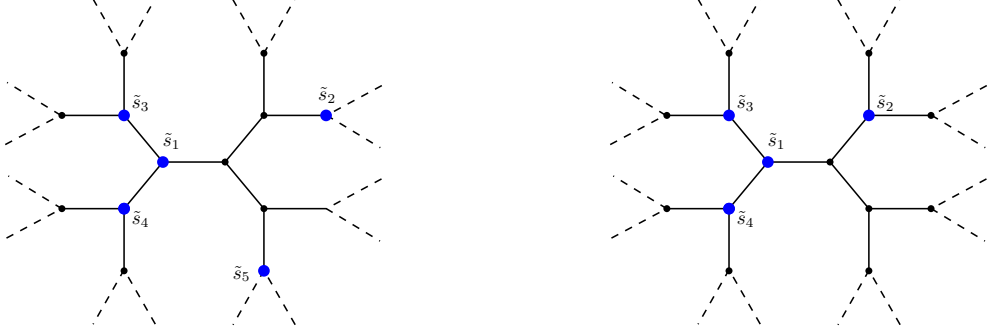


Figure 7: Examples of seed configurations on the binary tree. Seeds are shown in blue. On the left, the seed \tilde{s}_1 has almost surely a finite set of descendants. On the right, the seed \tilde{s}_1 has almost surely infinite degree in $\tilde{\mathcal{G}}$.

- For each $j \in \{1, \dots, M\}$, $\tilde{\mathcal{U}}_j$ is an oriented tree rooted at \tilde{s}_j . Moreover, for $v \in \tilde{\mathcal{U}}_j$, $v \neq \tilde{s}_j$, we have $\mathbb{E}[\#\tilde{D}(v) \mid \tau_v] \leq \mathbb{E}[\#D(v) \mid \tau_v] < \infty$. In particular, we have

$$\tilde{\mathcal{U}}_j \text{ is finite} \iff \tilde{s}_j \text{ has finite degree in } \tilde{\mathcal{U}}_j \quad \text{a.s.}$$

In order to complete the proof of Theorem 3, it remains to show that

- (i) For every $i \in I$, there exists $v \in \mathcal{C}_i$ such that $\Gamma(v)$ does not end at a seed, *i.e.*,

$$\tilde{I} = I \tag{22}$$

- (ii) We have

$$\mathbb{P}(\tilde{s}_j \text{ has infinite degree in } \tilde{\mathcal{U}}_j) = \begin{cases} 1 & \text{if } \{y \in \mathcal{T} : d(y, \tilde{s}_j) = \min_{i \leq M} d(y, \tilde{s}_i)\} \text{ is infinite} \\ 0 & \text{otherwise} \end{cases} \tag{23}$$

Proof of (22). Fix $i \in I$ and $v \in \mathcal{C}_i$, and let $\Gamma(v) := (V_n)_{n \geq 1}$ be its ancestral path in the genealogical graph without seeds. We prove that, for n large enough, $W_\infty(V_n) \cap S = \emptyset$, and hence $V_n \in \tilde{\mathcal{C}}_i \neq \emptyset$.

Recall that R_{n+1} denotes the distance between V_n and V_{n+1} . By Corollary 15, it holds that, almost surely,

$$\log(K)R_n \leq 2n \quad \text{and} \quad |V_{n+1}| \geq \frac{n^2}{3 \log K} \quad \text{for } n \text{ large enough.}$$

Using that the set of seeds S is finite, we deduce that, for N large enough,

$$S \cap \bigcup_{n \geq N} B(V_n, R_{n+1}) = \emptyset.$$

Since $W_\infty(V_N) \subset \bigcup_{n \geq N} B(V_n, R_{n+1})$, this implies that $W_\infty(V_N) \cap S = \emptyset$, and therefore $V_N \in \tilde{\mathcal{C}}_i$. \square

Proof of (23). For $v \in \mathcal{T} \setminus S$, recall that $\tilde{p}(v)$ denotes the direct ancestor of v in $\tilde{\mathcal{G}}$. Let

$$\tilde{D}_1(\tilde{s}_j) := \{v \in \mathcal{T} : \tilde{p}(v) = \tilde{s}_j\}$$

denote the set of direct descendants of the seed \tilde{s}_j in $\tilde{\mathcal{G}}$. Let

$$\mathcal{T}' := \{u \in \mathcal{T} : d(u, \tilde{s}_j) = \min_{i \leq M} d(u, \tilde{s}_i)\}.$$

Note that $\tilde{D}_1(\tilde{s}_j) \subset \mathcal{T}'$ because if $v \notin \mathcal{T}'$, then there exists another seed closer to v than \tilde{s}_j , so it cannot happen that $\tilde{p}(v) = \tilde{s}_j$. Hence

$$\mathcal{T}' \text{ finite} \implies \mathbb{P}(\tilde{s}_j \text{ has infinite degree in } \tilde{\mathcal{U}}_j) = 0.$$

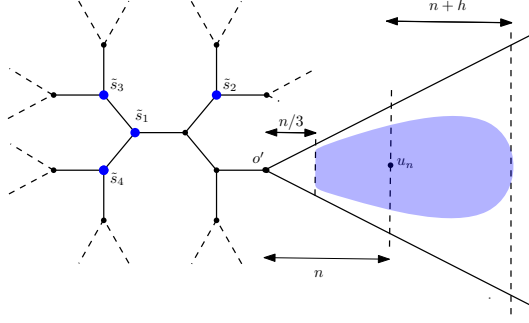


Figure 8: Illustration of the notation used in the proof of (23) for the seed \tilde{s}_1 in the configuration on the right of Figure 7. On the event \mathcal{E}_n , no vertex in the blue region has been colored before u_n . For n large enough, if this event occurs, the ancestor of u_n is either \tilde{s}_1 or \tilde{s}_2 , each with probability $1/2$.

This shows the second case of (23). Let us now assume that \mathcal{T}' is infinite. Note that \mathcal{T}' is a subtree of \mathcal{T} and, because it is infinite, there exists a vertex $o' \in \mathcal{T}$ such that $\mathcal{T}_{o'} \subset \mathcal{T}'$ and we can furthermore assume w.l.o.g. $\tilde{s}_j \notin \mathcal{T}_{o'}$. We prove that

$$\mathcal{T}_{o'} \cap \tilde{D}_1(\tilde{s}_j) \text{ is infinite a.s.} \quad (24)$$

Let $h := d(\tilde{s}_j, o')$. For $n \geq 1$, let u_n be the vertex with minimal arrival time among the vertices v of $\mathcal{T}_{o'}$ such that $d(o', v) = n$. Define the event

$$\mathcal{E}_n := \left\{ \tau_{u_n} \leq \tau_v \text{ for all } v \text{ such that } d(v, u_n) \leq n+h \text{ and } d(v, o') \geq \frac{n}{3} \right\}.$$

Note that $d(u_n, \tilde{s}_j) = n+h$, so the ancestor of u_n is at distance at most $n+h$ from u_n . Hence, if \mathcal{E}_n occurs, then

$$\tilde{p}(u_n) \in \{v \in \mathcal{T}_{o'} : d(o', v) < n/3\} \cup \{v \in \mathcal{T} : d(o', v) \leq h\}.$$

We now prove the following two facts:

- (i) Infinitely many events \mathcal{E}_n occur almost surely.
- (ii) $\tilde{p}(u_n) \notin \{v \in \mathcal{T}_{o'} : d(o', v) < n/3\}$ for all n large enough almost surely.

On the one hand, we note that

$$\#\left\{v : d(v, u_n) \leq n+h \text{ and } d(v, o') \geq \frac{n}{3}\right\} \leq \#B(u_n, n+h) \leq (K+1)K^{n+h}.$$

Since τ_{u_n} is the minimum of K^n i.i.d. random variables, the bound above yields

$$\mathbb{P}(\mathcal{E}_n) \geq \frac{K^n}{K^n + (K+1)K^{n+h}} = \frac{1}{1 + (K+1)K^h} =: c_0 > 0.$$

Moreover, for $n > 3h$, the event \mathcal{E}_n only depends on the arrival times of the vertices v of $\mathcal{T}_{o'}$ such that

$$\frac{n}{3} \leq d(o', v) < 3n.$$

Thus, the events $(\mathcal{E}_{9n})_{n > 3h}$ are independent and the Borel–Cantelli lemma implies that (i) holds.

On the other hand, a similar computation using that $\#\{v : d(o', v) < n/3\} \leq (K+1)K^{\lfloor n/3 \rfloor}$ shows that

$$\mathbb{P}(\min\{\tau_v : d(o', v) < n/3\} \leq \tau_{u_n}) \leq \frac{(K+1)K^{n/3}}{K^n + (K+1)K^{n/3}} \leq (K+1)K^{-2n/3}.$$

Hence, the Borel–Cantelli lemma now implies that

$$\min\{\tau_v : d(o', v) < n/3\} > \tau_{u_n}$$

for all n large enough. This implies that the ancestor of u_n cannot belong to $\{v \in \mathcal{T}_{o'} : d(o', v) < n/3\}$, so (ii) holds.

Let us now conclude that (24) holds. Using (i) and (ii), we find that, for infinitely many n ,

$$\tilde{p}(u_n) \in \{v \in \mathcal{T} : d(o', v) \leq h\}.$$

Since τ_{u_n} tends to 0 almost surely, it follows that, ultimately, $\tilde{p}(u_n)$ must be a seed, *i.e.*

$$\tilde{p}(u_n) \in \{s \in S : d(o', s) = d(o', \tilde{s}_j)\} \quad \text{infinitely often a.s.}$$

Finally, by the independence of the tie-breaking indexings at the distinct vertices u_n , each seed in this finite set is chosen infinitely often; in particular, \tilde{s}_j has infinitely many direct descendants a.s. \square

5 Local limits

We now come back to the case of finite trees and study the local limits of the random coloring in different settings. Recall that $\mathcal{T}^{(\ell)} := \mathcal{T} \cap B(o, \ell)$ denotes the finite tree of height ℓ rooted at o .

Let s_1, \dots, s_m be a (possibly empty) set of seeds. We always assume ℓ is sufficiently large so that $s_i \in \mathcal{T}^{(\ell)}$ for all i . We can construct the genealogical graph $\tilde{\mathcal{G}}$ on \mathcal{T} and the genealogical graph $\tilde{\mathcal{G}}^{(\ell)}$ on the finite tree $\mathcal{T}^{(\ell)}$ (both with seeds s_1, \dots, s_m) on the same probability space using the natural coupling: we use the same arrival times $(\tau_v)_{v \in \mathcal{T}}$ and the same uniform increasing indexings $((z_j^w)_{j \geq 1}, w \in \mathcal{T})$ to break ties (and we simply ignore the vertices not in $\mathcal{T}^{(\ell)}$ when constructing $\tilde{\mathcal{G}}^{(\ell)}$).

By definition, the genealogical graph is constructed by a local exploration around each vertex of the tree, which makes it nicely behaved with respect to the notion of local convergence. More precisely, we have the straightforward result, stated here for trees but valid for any locally finite connected graph:

Proposition 23. *We have*

$$\tilde{\mathcal{G}}^{(\ell)} \xrightarrow{\ell \rightarrow \infty} \tilde{\mathcal{G}}$$

for the topology of local convergence. In other words, for any $N \geq 0$ we have

$$\tilde{\mathcal{G}} \cap B(o, N) = \tilde{\mathcal{G}}^{(\ell)} \cap B(o, N) \quad \text{for all } \ell \text{ large enough a.s.}$$

where $\tilde{\mathcal{G}} \cap B(o, N)$ (resp. $\tilde{\mathcal{G}}^{(\ell)} \cap B(o, N)$) stands for the genealogical graph $\tilde{\mathcal{G}}$ (resp. $\tilde{\mathcal{G}}^{(\ell)}$) restricted to the set of vertices $B(o, N)$.

Proof. We just need to prove that any vertex v (that is not a seed) has the same direct ancestor in $\tilde{\mathcal{G}}^{(\ell)}$ and $\tilde{\mathcal{G}}$ for ℓ large enough. Since both graphs are constructed with the same arrival times and the same increasing indexing, this is true as soon as the exploration around v does not reach the boundary of $\mathcal{T}^{(\ell)}$, *i.e.*, as soon as $\ell \geq d(o, v) + d(v, \tilde{p}(v))$, where $\tilde{p}(v)$ denotes the direct ancestor of v in $\tilde{\mathcal{G}}$. \square

While the genealogical graph is continuous with respect to the local topology, the coloring itself does not have this property because the color of a vertex is related to the connectedness of vertices in the genealogical graph and therefore to the existence of an infinite ancestral path, which is a non-local notion. In the rest of this section, we study the limits of the coloring in two particular cases:

- (a) Two fixed seeds colored blue and red, respectively.
- (b) A single blue seed at the root and a red seed on each leaf of $\mathcal{T}^{(\ell)}$.

5.1 Local limit with two seeds (Theorem 6)

We fix another vertex o' and consider the random coloring on $\mathcal{T}^{(\ell)}$ with a blue seed at the root o and a red seed at o' . This gives a partition of the vertices of $\mathcal{T}^{(\ell)}$ into the blue set $\tilde{B}^{(\ell)}$ and the red set $\tilde{R}^{(\ell)}$, which are precisely the vertices of the two connected components of the genealogical graph $\tilde{\mathcal{G}}^{(\ell)}$ rooted respectively at o and o' .

As in the previous sections, we denote by $\tilde{\Gamma}(v) = (\tilde{V}_i)_{i \leq n(v)}$ the ancestral path of v in $\tilde{\mathcal{G}}$. According to Theorem 3, this path may be infinite ($n(v) = \infty$) or finite ($n(v) < \infty$) and, in the latter case, it ends

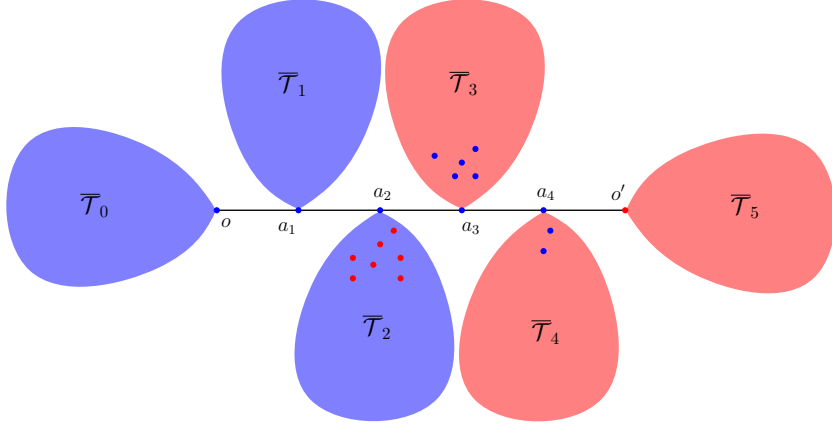


Figure 9: Illustration of the trees $(\bar{\mathcal{T}}_i)_{0 \leq i \leq d}$ for $d = 5$. The two-seed coloring process on $\mathcal{T}^{(\ell)}$ converges locally to the coloring shown in the figure. All vertices of $\bar{\mathcal{T}}_0$ (resp. $\bar{\mathcal{T}}_5$) are almost surely blue (resp. red). Moreover, $\bar{\mathcal{T}}_1$ and $\bar{\mathcal{T}}_2$ contain only finitely many red vertices, whereas $\bar{\mathcal{T}}_3$ and $\bar{\mathcal{T}}_4$ contain only finitely many blue vertices.

either at o or o' . Similarly, we denote by $\tilde{\Gamma}^{(\ell)}(v) = (\tilde{V}_i^{(\ell)})_{i \leq n_\ell(v)}$ the ancestral path of v in $\tilde{\mathcal{G}}^{(\ell)}$. In this case, the path is necessarily finite and $\tilde{V}_{n_\ell(v)}^{(\ell)} \in \{o, o'\}$ defines the color of the vertex v .

Let (a_0, a_1, \dots, a_d) denote the vertices on the path $[o, o']$, ordered by increasing distance from o . Thus $a_0 = o$, $a_d = o'$, and $d(o, a_i) = i$. Let

$$\bar{\mathcal{T}}_i := \mathcal{T}_{a_i \setminus \{o, o'\}}$$

be the subtree of \mathcal{T} composed of $\{a_i\}$ and all the connected components of $\mathcal{T} \setminus \{a_i\}$ that contain neither o nor o' (see Figure 9 for an illustration of $\bar{\mathcal{T}}_i$). The family $(\bar{\mathcal{T}}_i)_{0 \leq i \leq d}$ forms a partition of \mathcal{T} . Moreover, a vertex $v \in \bar{\mathcal{T}}_i$ is closer to o than to o' if $i < d/2$, at equal distance from both if $i = d/2$, and closer to o' if $i > d/2$.

Lemma 24. *Let $v \in \mathcal{T}$. We have*

(i) *If $\tilde{\Gamma}(v)$ is finite, then for ℓ large enough,*

$$\tilde{\Gamma}^{(\ell)}(v) = \tilde{\Gamma}(v).$$

(ii) *If $\tilde{\Gamma}(v)$ is infinite, then, for any vertex $w \in \text{dir } \tilde{\Gamma}(v)$ and for all ℓ large enough, the penultimate vertex of $\tilde{\Gamma}^{(\ell)}(v)$ is in \mathcal{T}_w , i.e.,*

$$\tilde{V}_{n_\ell(v)-1}^{(\ell)} \in \mathcal{T}_w.$$

Proof. Item (i) is a direct consequence of the previous proposition, applied to a ball containing the finite path $\tilde{\Gamma}(v)$. Assume now that $\tilde{\Gamma}(v)$ is infinite, and fix a vertex w on the ray $\text{dir } \tilde{\Gamma}(v)$. If $w = o$, there is nothing to prove. We therefore assume that $w \neq o$. There exists $N \geq 1$ such that $\tilde{V}_n \in \mathcal{T}_w$ for every $n \geq N$. Set

$$A_w := B(w, d(w, o)) \setminus (\mathcal{T}_w \cup \{o, o'\}).$$

Since A_w is finite and $\tau_{\tilde{V}_n} \rightarrow 0$, by possibly increasing N we may assume that $\tau_{\tilde{V}_N} < \min\{\tau_u : u \in A_w\}$.

By the previous proposition, applied to a ball containing $(\tilde{V}_1, \dots, \tilde{V}_N)$, the first N vertices of $\tilde{\Gamma}^{(\ell)}(v)$ coincide with those of $\tilde{\Gamma}(v)$ for all sufficiently large ℓ . Let $Y^{(\ell)}$ denote the last non-seed vertex of the path $\tilde{\Gamma}^{(\ell)}(v)$ that belongs to \mathcal{T}_w . Since $\tilde{V}_N \in \mathcal{T}_w$, it implies that $Y^{(\ell)}$ appears at or after that vertex in the ancestral path:

$$\tau_{Y^{(\ell)}} \leq \tau_{\tilde{V}_N} < \min\{\tau_u : u \in A_w\}.$$

Consequently, the direct ancestor of $Y^{(\ell)}$ in $\tilde{\mathcal{G}}^{(\ell)}$ cannot belong to A_w . If this ancestor were not a seed, then, by the definition of $Y^{(\ell)}$, it could not belong to \mathcal{T}_w either. It would then belong to $\mathcal{T} \setminus (\mathcal{T}_w \cup A_w \cup \{o, o'\})$, and hence would be farther from $Y^{(\ell)}$ than the seed o , which is impossible by the definition of

the direct ancestor. Thus this direct ancestor belongs to $\{o, o'\}$, and hence $Y^{(\ell)}$ is the penultimate vertex of $\tilde{\Gamma}^{(\ell)}(v)$, *i.e.*

$$\tilde{V}_{n_\ell(v)-1}^{(\ell)} \in \mathcal{T}_w.$$

□

Proof of Proposition 7. We assume here that $d = d(o, o')$ is odd and show almost-sure convergence of the coloring under the natural coupling. Furthermore, we establish that the limiting color of a vertex can be read from the ancestral path in $\tilde{\mathcal{G}}$: either from its terminal seed, if it is finite, or from the “side” of its asymptotic direction, if it is infinite. This follows directly from the following observations. Fix $v \in \mathcal{T}$; we have the following alternatives:

- If $\tilde{\Gamma}(v)$ is finite and ends at o (resp. o'), then the same property holds for $\tilde{\Gamma}^{(\ell)}(v)$ for ℓ large enough and so $v \in \tilde{B}^{(\ell)}$ (resp. $v \in \tilde{R}^{(\ell)}$) for ℓ large enough.
- Assume now that $\tilde{\Gamma}(v)$ is infinite, with $\text{dir } \tilde{\Gamma}(v) \in \partial\overline{\mathcal{T}}_i$ and $i < d/2$. Choose a vertex w on the ray $\text{dir } \tilde{\Gamma}(v)$ such that $\mathcal{T}_w \subset \overline{\mathcal{T}}_i$. By the previous lemma, for ℓ large enough, $\tilde{V}_{n_\ell(v)-1}^{(\ell)} \in \mathcal{T}_w \subset \overline{\mathcal{T}}_i$. Thus, we have $d(\tilde{V}_{n_\ell(v)-1}^{(\ell)}, o) < d(\tilde{V}_{n_\ell(v)-1}^{(\ell)}, o')$ and so necessarily $\tilde{V}_{n_\ell(v)}^{(\ell)} = o$, *i.e.*, $v \in \tilde{B}^{(\ell)}$.
- For a similar reason, $v \in \tilde{R}^{(\ell)}$ for ℓ large enough if $\text{dir } \tilde{\Gamma}(v) \in \partial\overline{\mathcal{T}}_i$ with $i > d/2$.

These facts show the a.s. convergence stated above and the description of the limiting blue and red sets $\tilde{B}^{(\infty)}, \tilde{R}^{(\infty)}$ as

$$\tilde{B}^{(\infty)} := \{v \in \mathcal{T} : \tilde{\Gamma}(v) \text{ ends at } o \text{ or } \tilde{\Gamma}(v) \text{ is infinite and } \text{dir } \tilde{\Gamma}(v) \in \cup_{i < d/2} \partial\overline{\mathcal{T}}_i\},$$

$$\tilde{R}^{(\infty)} := \{v \in \mathcal{T} : \tilde{\Gamma}(v) \text{ ends at } o' \text{ or } \tilde{\Gamma}(v) \text{ is infinite and } \text{dir } \tilde{\Gamma}(v) \in \cup_{i > d/2} \partial\overline{\mathcal{T}}_i\}.$$

To complete the proof of Proposition 7, it remains to show that

$$\tilde{B}^{(\infty)} \cap \{v \in \mathcal{T} : d(v, o) > d(v, o')\} \text{ is a.s. finite.}$$

i.e., only finitely many vertices of $\overline{\mathcal{T}}_i$ for $i > d/2$ have an ancestral path $\tilde{\Gamma}(v)$ which ends at o or is infinite with asymptotic direction $\text{dir } \tilde{\Gamma}(v) \in \partial\overline{\mathcal{T}}_j$ for some $j < d/2$. Let v be such a vertex. Then, there exists $z \in \tilde{\Gamma}(v)$ such that

$$z \in \cup_{i > d/2} \overline{\mathcal{T}}_i \quad \text{and} \quad \tilde{p}(z) \in \cup_{i < d/2} \overline{\mathcal{T}}_i.$$

Using the fact that $d(\tilde{p}(z), z) \leq d(z, o')$, we get that $\tilde{p}(z) \in B(o', d-1) \setminus \{o'\}$. Thus

$$\tilde{B}^{(\infty)} \cap \{v \in \mathcal{T} : d(v, o) > d(v, o')\} \subset \bigcup_{u \in B(o', d-1) \setminus \{o'\}} \tilde{D}(u),$$

where we recall that $\tilde{D}(u)$ denotes the set of descendants of u in the graph $\tilde{\mathcal{G}}$. We conclude using the fact that $B(o', d-1) \setminus \{o'\}$ is a finite set, together with Theorem 3, which states that $\tilde{D}(u)$ is a.s. finite for $u \notin \{o, o'\}$. The second finiteness statement follows by symmetry. □

Remark 25. *When $d(o, o')$ is even, the natural coupling does not give an a.s. local limit in general. Indeed, for a vertex whose infinite ancestral path has asymptotic direction in $\partial\overline{\mathcal{T}}_{d/2}$, the penultimate vertex of $\tilde{\Gamma}^{(\ell)}(v)$ eventually lies in the middle subtree, where the two seeds are at the same distance. As ℓ grows, infinitely many fresh tie-breakings between o and o' are therefore used, and the color of such a vertex does not stabilize, but takes both colors infinitely often. This is why Theorem 6 proved below is stated as a convergence in distribution, with an additional Bernoulli choice for each middle-direction component.*

Proof of Theorem 6. It suffices to show that for any finite collection of vertices (v_1, \dots, v_k) , the following properties hold:

- (i) If $\tilde{\Gamma}(v_i)$ is finite and ends at o (resp. o'), or if $\text{dir } \tilde{\Gamma}(v_i) \in \partial\overline{\mathcal{T}}_j$ with $j < d/2$ (resp. $j > d/2$), then $v_i \in \tilde{B}^{(\ell)}$ (resp. $\tilde{R}^{(\ell)}$) for all sufficiently large ℓ .
- (ii) If $\text{dir } \tilde{\Gamma}(v_i) \in \partial\overline{\mathcal{T}}_{d/2}$, define

$$\mathcal{I} := \{i' \leq k : \tilde{\Gamma}(v_{i'}) \text{ coalesces with } \tilde{\Gamma}(v_i)\}.$$

Then, for all sufficiently large ℓ , the vertices $(v_{i'})_{i' \in \mathcal{I}}$ receive the same color in the two-seed coloring process on $\mathcal{T}^{(\ell)}$, and this color is blue or red with probability 1/2 each. Moreover, this color is asymptotically independent of the colors of $(v_{i'})_{i' \notin \mathcal{I}}$.

Property (i) was established in Proposition 7 when $d = d(o, o')$ is odd. The proof given there in fact applies verbatim when d is even and $j \neq d/2$. We now prove (ii). We assume that d is even, since otherwise there is nothing to prove. Also, to avoid cumbersome notation, we only treat the case of two vertices v_1 and v_2 ; the general case follows by the same argument.

Assume that $\text{dir } \tilde{\Gamma}(v_1) \in \partial\overline{\mathcal{T}}_{d/2}$. Suppose first that $\tilde{\Gamma}(v_1)$ and $\tilde{\Gamma}(v_2)$ coalesce. Then, for all sufficiently large ℓ , the truncated ancestral paths $\tilde{\Gamma}^{(\ell)}(v_1)$ and $\tilde{\Gamma}^{(\ell)}(v_2)$ also coalesce. Consequently, v_1 and v_2 receive the same color in the two-seed coloring process on $\mathcal{T}^{(\ell)}$. By Lemma 24, the penultimate vertex of $\tilde{\Gamma}^{(\ell)}(v_1)$ is in $\overline{\mathcal{T}}_{d/2}$ for all sufficiently large ℓ . Since this vertex is at equal distance from o and o' , it is connected to either seed with probability 1/2 each. This proves (ii) in the case of coalescing paths.

Assume now that $\tilde{\Gamma}(v_1)$ and $\tilde{\Gamma}(v_2)$ do not coalesce. We show that, for all sufficiently large ℓ , the paths $\tilde{\Gamma}^{(\ell)}(v_1)$ and $\tilde{\Gamma}^{(\ell)}(v_2)$ are disjoint, except possibly for their last vertex. By Item 5 of Theorem 1, we have

$$\text{dir } \tilde{\Gamma}(v_1) \neq \text{dir } \tilde{\Gamma}(v_2).$$

(We assume here that $\tilde{\Gamma}(v_2)$ is infinite; the finite case is even simpler.) Hence, there exist vertices $w_1 \in \overline{\mathcal{T}}_{d/2}$ and $w_2 \in \mathcal{T}$ such that \mathcal{T}_{w_1} and \mathcal{T}_{w_2} are disjoint, while

$$\text{dir } \tilde{\Gamma}(v_1) \in \partial\mathcal{T}_{w_1}, \quad \text{dir } \tilde{\Gamma}(v_2) \in \partial\mathcal{T}_{w_2}.$$

By Lemma 24, for all sufficiently large ℓ ,

$$\tilde{V}_{n_\ell(v_1)-1}^{(\ell),1} \in \mathcal{T}_{w_1}, \quad \tilde{V}_{n_\ell(v_2)-1}^{(\ell),2} \in \mathcal{T}_{w_2},$$

where $(\tilde{V}_n^{(\ell),i})_{n \geq 1}$ denotes the ancestral path of v_i in $\mathcal{T}^{(\ell)}$. Since \mathcal{T}_{w_1} and \mathcal{T}_{w_2} are disjoint, the paths $\tilde{\Gamma}^{(\ell)}(v_1)$ and $\tilde{\Gamma}^{(\ell)}(v_2)$ are disjoint, except possibly at their final vertex. As before, the fact that $\text{dir } \tilde{\Gamma}(v_1) \in \partial\overline{\mathcal{T}}_{d/2}$ implies that the penultimate vertex of $\tilde{\Gamma}^{(\ell)}(v_1)$ is at equal distance from o and o' . Hence, the color of v_1 is blue or red with probability 1/2 each, depending on the increasing ordering attached to $\tilde{V}_{n_\ell(v_1)-1}^{(\ell),1}$. Since this ordering is independent of the ancestral path of v_2 (hence of the color of v_2), it follows that

$$\lim_{\ell \rightarrow \infty} \mathbb{P}(v_1 \in \tilde{B}^{(\ell)} \mid v_2 \in \tilde{B}^{(\ell)}) = \lim_{\ell \rightarrow \infty} \mathbb{P}(v_1 \in \tilde{B}^{(\ell)} \mid v_2 \in \tilde{R}^{(\ell)}) = \frac{1}{2}.$$

This proves (ii) and completes the proof of the theorem. \square

5.2 Local limit of a tree with seeds at the root and the leaves (Theorem 5)

We now consider the coloring process on the finite tree $\mathcal{T}^{(\ell)}$, with a blue seed at the root o and a red seed at each leaf of $\mathcal{T}^{(\ell)}$. As usual, we let $B^{(\ell)}$ and $R^{(\ell)}$ denote the final sets of blue and red vertices, respectively. We also consider the genealogical graph $\tilde{\mathcal{G}}^o$ on \mathcal{T} with a single seed located at o . These processes are constructed on the same probability space using the same arrival times $(\tau_w)_{w \in \mathcal{T}}$ and the same random orderings $((z_j^w)_{j \geq 1})_{w \in \mathcal{T}}$ used to break ties.

As already observed in the introduction, if a vertex v is colored blue in $\mathcal{T}^{(\ell)}$, then it is also colored blue in all subsequent $\mathcal{T}^{(\ell')}$ for every $\ell' > \ell$. Hence,

$$B^{(\ell)} \subset B^{(\ell')} \quad \text{for every } \ell < \ell'$$

and, as $\ell \rightarrow \infty$, the set $B^{(\ell)}$ converges a.s. to $\tilde{\mathcal{U}}_o$, the set of vertices of \mathcal{T} whose ancestral path ends at the root o , *i.e.*, the genealogical component of $\tilde{\mathcal{G}}^o$ containing the root. This proves the local-limit statement in Theorem 5. Furthermore, recalling the notation

$$m_\ell := \min\{|v| : v \in R^{(\ell)}\}$$

for the minimum height of the red vertices, we find that

$$m_\infty := \lim_{\ell \rightarrow \infty} m_\ell = \min\{|v| : v \notin \tilde{\mathcal{U}}_o\}.$$

Now, by Theorem 3, the random variable m_∞ is almost surely finite. Moreover, Proposition 17 implies that

$$\mathbb{P}(m_\infty = 1) > 0.$$

This proves statement (6) of Theorem 5.

We now turn to the study of

$$M_\ell := \max\{|v| : v \in B^{(\ell)}\}.$$

Again, we have

$$M_\infty := \lim_{\ell \rightarrow \infty} M_\ell = \max\{|v| : v \in \tilde{\mathcal{U}}_o\},$$

and, by Item 2(b) of Theorem 3, the component $\tilde{\mathcal{U}}_o$ is almost surely infinite. Consequently, $M_\infty = \infty$ a.s. This completes the part of Theorem 5 which follows from local convergence and from the infinite-volume genealogical picture. It remains only to prove the last part of Theorem 5, namely that, with high probability, the blue vertices reach the leaves of $\mathcal{T}^{(\ell)}$ as $\ell \rightarrow \infty$, *i.e.*,

$$\lim_{\ell \rightarrow \infty} \mathbb{P}(M_\ell = \ell - 1) = 1, \tag{25}$$

which is exactly the estimate (5), restated here for convenience. This estimate is of a different nature: it concerns the behavior of the coloring near the boundary of $\mathcal{T}^{(\ell)}$, which escapes to infinity, and therefore cannot be read from the convergence of any fixed neighborhood of the root.

6 The blue vertices reach the leaves

This last section is devoted solely to the proof of (25), and hence completes the proof of Theorem 5. It is somewhat standalone: rather than reconstructing colors through backward genealogical explorations, as in the previous sections, we follow the evolution of the coloring process forward in time.

The proof uses a recursive argument on finite descendant subtrees and a coupling with a supercritical Galton–Watson process. For this reason, it is convenient to work with a generic finite rooted K -ary tree T instead of $\mathcal{T}^{(\ell)}$. The only discrepancy between the two trees is at the global root, which has $K + 1$ children in $\mathcal{T}^{(\ell)}$ and only K children in T ; this first-generation difference will play no role, so we will work with T from now on to avoid having to distinguish the case of the root vertex.

We use the following notation throughout the proof.

- T denotes a finite rooted K -ary tree, with root o and height $|T|$. We write ∂T for its set of leaves, and $|v|$ for the height of a vertex $v \in T$.
- For a vertex $v \in T$, we denote by T_v the subtree of T rooted at v , consisting of all its descendants. Thus T_v is itself a rooted K -ary tree of height $|T| - |v|$. Whenever this notation is applied to a rooted subtree, heights are understood relative to the root of that subtree.
- For two disjoint subsets $B_0, R_0 \subset T$, we consider the coloring process on T with blue seeds in B_0 and red seeds in R_0 . The vertices in $B_0 \cup R_0$ have arrival time 0, while the variables $(\tau_u)_{u \notin B_0 \cup R_0}$ are i.i.d. exponential random variables with parameter 1. For $t \geq 0$, we denote by B_t (resp. R_t) the set of vertices colored blue (resp. red) up to time t .
- Finally, X_1^T, X_2^T, \dots denote the vertices of $T \setminus \{o\}$ whose heights are strictly smaller than $|T|/2$, ordered according to their coloring times in increasing order.

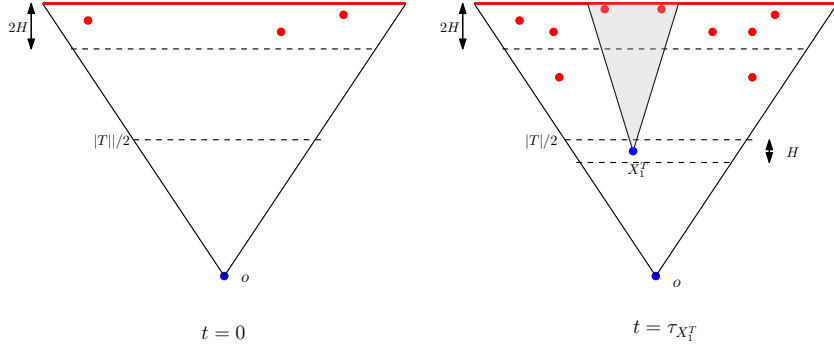


Figure 10: Illustration of the initial condition in Proposition 26 and of the event \mathcal{A}_1 , with $C = 3$. If \mathcal{A}_1 occurs, then at time $t = \tau_{X_1^T}$, at most C colored non-leaf vertices lie in the subtree $T_{X_1^T}$, and they are all close (at distance at most $2H$) from its leaves.

Note that, once v has been colored, it separates the subtree T_v from the rest of the tree: conditionally on the state $(B_{\tau_v} \cap T_v, R_{\tau_v} \cap T_v)$, the evolution of the coloring inside T_v after time τ_v depends only on the clocks $(\tau_w)_{w \in T_v}$ and on the increasing orderings $((z_j^w)_{j \geq 1})_{w \in T_v}$, and is therefore independent of the exterior.

The event \mathcal{A}_i defined below plays the role of a “good event” for the recursive step: it says that X_i^T becomes a blue vertex near the middle of the tree, while the part of its descendant subtree that has already been colored remains confined to a bounded layer near the leaves, which enables the recursion to be applied in $T_{X_i^T}$. More precisely, for $C, H \geq 0$ and $i \geq 1$, we define the event $\mathcal{A}_i := \mathcal{A}_i(T, C, H, B_0, R_0)$ by

$$\mathcal{A}_i := \left\{ \begin{array}{l} X_i^T \text{ is colored blue,} \\ |T|/2 - H < |X_i^T| < |T|/2, \\ \text{at time } (\tau_{X_i^T})^-, T_{X_i^T} \text{ contains at most } C \text{ colored non-leaf vertices,} \\ \text{and all these vertices are at distance at most } 2H \text{ from the leaves} \end{array} \right\}. \quad (26)$$

6.1 Coupling with a supercritical Galton–Watson process

The next proposition is the key ingredient in the proof. It states that with high probability one can find two disjoint descendant subtrees in which the good event \mathcal{A}_i occurs. Thanks to the independence observation mentioned previously, the coloring can then be restarted in these subtrees, which will allow us to compare the growth of the blue region with a supercritical Galton–Watson process.

Proposition 26. *Recall the definition of the events \mathcal{A}_i given in (26). Let $\eta > 0$. There exist $H, C, n_0 \geq 0$ such that, for all $|T| \geq n_0$, and for any choice of vertices w_1, \dots, w_C of height at least $|T| - 2H$, starting from the initial blue seed $B_0 = \{o\}$ and red seeds $R_0 = \partial T \cup \{w_1, \dots, w_C\}$, we have*

$$\mathbb{P}(\exists i < j, T_{X_i^T} \cap T_{X_j^T} = \emptyset \text{ and } \mathcal{A}_i \cap \mathcal{A}_j) \geq 1 - \eta. \quad (27)$$

See Figure 10 for an illustration of the initial condition and of the event \mathcal{A}_1 .

We first explain how Proposition 26 implies (25) and hence completes the proof of Theorem 5, before proving the proposition itself.

Proof of (25) using Proposition 26. Set

$$\mathcal{C}_{i,j} := \mathcal{C}_{i,j}(T, C, H, B_0, R_0) := \{T_{X_i^T} \cap T_{X_j^T} = \emptyset \text{ and } \mathcal{A}_i \cap \mathcal{A}_j\}.$$

Proposition 26 asserts that, for suitable choices of (T, C, H, B_0, R_0) , with high probability there exist indices $i < j$ such that $\mathcal{C}_{i,j}$ occurs.

Fix $\eta \in (0, 1/4)$ and choose $H, C, n_0 \geq 0$ such that (27) holds. By possibly increasing n_0 , we may assume that $C < K^{n_0 - 2H - 1}$ and $n_0/2 - 2H > 2H$. Consider the following process $(\mathcal{V}_l)_{l \geq 0}$ taking values in the subsets of T :

- $\mathcal{V}_0 = \{o\}$. If there exist $i < j$ such that $\mathcal{C}_{i,j}(T, C, H, \{o\}, \partial T)$ occurs, we set $\mathcal{V}_1 := \{X_i^T, X_j^T\}$ and call X_i^T, X_j^T the children of o . Otherwise, we set $\mathcal{V}_1 = \emptyset$.

By definition of the event \mathcal{A}_i , for any vertex $v \in \mathcal{V}_1$, at time $(\tau_v)^-$, the subtree T_v contains at most C colored vertices that are not leaves and have height at least $|T_v| - 2H$. Note that these vertices must be red, since they are at distance at most $2H$ from the leaves and at distance at least $|T|/2 - 2H \geq n_0 - 2H > 2H$ from any blue vertex. Hence $B_{\tau_v} \cap T_v = \{v\}$. Moreover, for $v_1 \neq v_2$ in \mathcal{V}_1 , we have $T_{v_1} \cap T_{v_2} = \emptyset$. Thus, conditionally on $(B_{\tau_v}, R_{\tau_v})_{v \in \mathcal{V}_1}$, the processes $((B_t \cap T_v)_{t \geq \tau_v}, (R_t \cap T_v)_{t \geq \tau_v})$, $v \in \mathcal{V}_1$ are independent.

- Assume that $\mathcal{V}_l = \{U_1^{(l)}, \dots, U_{\#\mathcal{V}_l}^{(l)}\}$. For each $v \in \mathcal{V}_l$, consider the coloring process on T_v for $t \geq \tau_v$. By induction, these processes are independent for $v \in \mathcal{V}_l$. Each of them starts from an initial configuration where only the root v is blue, while the red vertices are the leaves together with at most C additional vertices, each located at distance at most $2H$ from the leaves.

If there exist $i < j$ such that $\mathcal{C}_{i,j}(T_v, C, H, \{v\}, R_{\tau_v} \cap T_v)$ occurs, we declare $\{X_i^{T_v}, X_j^{T_v}\}$ to be the children of v . We then define \mathcal{V}_{l+1} as the union of the children of all vertices $v \in \mathcal{V}_l$.

By induction, we have that if $v \in \mathcal{V}_l$, then

$$\frac{|T|}{2^l} \leq |T_v| \leq \frac{|T|}{2^l} + 2H.$$

Let $L \in \mathbb{N}$ be such that $\frac{|T|}{2^L} \geq n_0 > \frac{|T|}{2^{L+1}}$ that is,

$$L := \left\lfloor \frac{\log(|T|/n_0)}{\log 2} \right\rfloor.$$

Thus, for every $v \in \mathcal{V}_L$, we have

$$n_0 \leq |T_v| \leq 2n_0 + 2H.$$

Using Proposition 26, we see that $(\#\mathcal{V}_l)_{l \leq L}$ stochastically dominates a Galton–Watson process $(Z_l)_{l \leq L}$ starting from $Z_0 = 1$ with offspring distribution

$$\mu(0) = \eta, \quad \mu(2) = 1 - \eta.$$

Since $\eta < 1/4$, this Galton–Watson process is supercritical. It survives with probability $(1 - 2\eta)/(1 - \eta)$, and for l large enough,

$$\mathbb{P}(Z_l > (2(1 - 2\eta))^l \mid Z_l > 0) \geq 1 - \eta.$$

Hence, if $|T|$ is large enough, we obtain

$$\mathbb{P}(\#\mathcal{V}_L \geq (2(1 - 2\eta))^L) \geq \mathbb{P}(Z_L > (2(1 - 2\eta))^L) \geq 1 - 2\eta.$$

Set $\beta := \frac{\log(2(1 - 2\eta))}{\log 2} > 0$ so that

$$(2(1 - 2\eta))^L = 2^{\beta L} \geq \left(2^{\frac{\log(|T|/n_0)}{\log 2} - 1}\right)^\beta = \left(\frac{|T|}{2n_0}\right)^\beta.$$

Therefore,

$$\mathbb{P}\left(\#\mathcal{V}_L \geq \left(\frac{|T|}{2n_0}\right)^\beta\right) \geq 1 - 2\eta.$$

This shows that, with probability at least $1 - 2\eta$, there exist at least $(|T|/2n_0)^\beta$ disjoint subtrees of height between n_0 and $2n_0 + 2H$ such that, in each of them, the coloring process colors the root v blue and, at the time when it becomes blue, the subtree contains at most C red vertices that are not leaves and have height at least $|T_v| - 2H$. Each of these subtrees then evolves independently.

Using the facts that $C < K^{n_0-2H-1}$ and $|T_v| \geq n_0$, we obtain that the number of vertices in T_v at height $|T_v| - 2H$ is strictly larger than C . Therefore, at time τ_v , there necessarily exists a path from the root v to a leaf consisting entirely of uncolored vertices (except, of course, for the root and the leaf). Hence, the probability that no blue vertex reaches distance one from the leaves is strictly less than one. Since the sizes of these $(|T|/2n_0)^\beta$ trees are uniformly bounded, we in fact obtain a uniform upper bound $\alpha(n_0, H, C) < 1$ for the probability that a given tree has no blue vertex at distance one from the leaves.

Thus, writing M_T for the maximal height reached by a blue vertex in the coloring process on T , we obtain

$$\mathbb{P}(M_T < |T| - 1) \leq 2\eta + \alpha(n_0, H, C)^{(|T|/2n_0)^\beta},$$

which can be made arbitrarily small as $|T| \rightarrow \infty$, provided that η is chosen sufficiently small. \square

Remark 27. *The proof that the blue vertices reach the leaves presented here is rather intricate, and the definition of the event \mathcal{A}_i used to run the recursion may seem overcomplicated at first, but this complexity is necessary. For instance, requiring that the whole subtree $T_{X_i^T}$ be uncolored at time $(\tau_{X_i^T})^-$ would be too restrictive and prevent the recursion from being applied. Indeed, after one step of the construction, the descendant subtrees naturally inherit a bounded number of red vertices close to their leaves. The event \mathcal{A}_i allows precisely this boundary layer and keeps the class of admissible initial configurations stable under the recursion.*

6.2 Existence of a blue vertex at height around $|T|/2$

In order to prove Proposition 26, we first establish a weaker version by showing that there exists N such that, with high probability, one of the first N non-root vertices of height smaller than $|T|/2$ to be colored is colored blue and has an almost uncolored subtree in front of it. This is the content of Proposition 28 below, which will later be strengthened to Proposition 26.

Proposition 28. *Let $\eta > 0$. Then, there exist $H, C, N, n_0 \geq 0$ such that, if $|T| \geq n_0$, then for any choice of vertices w_1, \dots, w_C of height greater than or equal to $|T| - 2H$, starting from $B_0 = \{o\}$ and $R_0 = \partial T \cup \{w_1, \dots, w_C\}$, we have*

$$\mathbb{P}(\forall 1 \leq i \leq N, \mathcal{A}_i^c) \leq \eta.$$

Proof of Proposition 28. For $v \in T$ and $r, h \geq 0$, we denote

$$B(v, r) := \{w \in T : d(v, w) \leq r\} \quad \text{and} \quad B_{\geq h}(v, r) := \{w \in T : d(v, w) \leq r \text{ and } |w| \geq h\},$$

the ball centered at v of radius r and the set of vertices in this ball whose height is at least h , respectively. For $v_1, v_2 \in T$, we denote by $v_1 \wedge v_2$ their most recent common ancestor. Recall that $R_0 = \partial T \cup \{w_1, \dots, w_C\}$ denotes the set of initial red vertices, and that X_1^T, X_2^T, \dots enumerates the vertices of $T \setminus \{o\}$ whose heights are strictly smaller than $|T|/2$, ordered according to their coloring times. Since T is fixed throughout this section, we will simplify the notation and write X_1, X_2, \dots instead of X_1^T, X_2^T, \dots . We now introduce the following events.

$$\begin{aligned} \mathcal{E}_1 = \mathcal{E}_1(T, N, H, C, R_0) &:= \{\forall i \neq j \leq N, |X_i \wedge X_j| < |T|/8\}, \\ \mathcal{E}_2 = \mathcal{E}_2(T, N, H, C, R_0) &:= \{\forall i \leq N, |X_i| > |T|/2 - H\}, \\ \mathcal{E}_3 = \mathcal{E}_3(T, N, H, C, R_0) &:= \{\forall i \leq N, \forall k \leq C, X_i \notin B(w_k, |T|/2 + H)\}, \\ \mathcal{E}_4 = \mathcal{E}_4(T, N, H, C, R_0) &:= \{\forall i \leq N, \mathcal{A}_i^c\} \cap \mathcal{E}_1 \cap \mathcal{E}_2 \cap \mathcal{E}_3. \end{aligned}$$

Thus, we have

$$\mathbb{P}(\forall i \leq N, \mathcal{A}_i^c) \leq \mathbb{P}(\mathcal{E}_1^c) + \mathbb{P}(\mathcal{E}_2^c) + \mathbb{P}(\mathcal{E}_3^c) + \mathbb{P}(\mathcal{E}_4).$$

The first three terms on the right-hand side are easy to bound.

Indeed, to bound $\mathbb{P}(\mathcal{E}_1^c)$, observe that X_i and X_j are distributed as two distinct vertices chosen uniformly at random from the complete K -ary tree of height $\lceil |T|/2 - 1 \rceil$, excluding the root. It is easy

to see that the height of the most recent common ancestor of such vertices is stochastically dominated by a geometric random variable with parameter $(K-1)/K$. Hence, we obtain

$$\mathbb{P}(\mathcal{E}_1^c) \leq N^2 \mathbb{P}(|X_1 \wedge X_2| \geq |T|/8) \leq N^2 \frac{1}{K^{|T|/8}}. \quad (28)$$

Concerning $\mathbb{P}(\mathcal{E}_2^c)$, and still using that X_i is uniformly distributed on the complete K -ary tree of height $\lceil |T|/2 - 1 \rceil$, excluding the root, together with a simple counting argument on the number of vertices of height smaller than $|T|/2 - H$, we obtain

$$\mathbb{P}(\mathcal{E}_2^c) \leq NK^{-H}. \quad (29)$$

For $\mathbb{P}(\mathcal{E}_3^c)$, recall that by assumption, for any $k \leq C$, $|w_k| \geq |T| - 2H$, whereas for any $i \geq 1$ we have $|X_i| \leq \lceil |T|/2 - 1 \rceil$. Let z_k denote the ancestor of w_k at height $\lceil |T|/2 - 1 \rceil$. Then

$$X_i \in B(w_k, |T|/2 + H) \implies X_i \in B(z_k, 3H).$$

Hence,

$$\mathbb{P}(\mathcal{E}_3^c) \leq CN \mathbb{P}(X_1 \in B(z_1, 3H)) \leq CN \frac{K^{3H}}{K^{|T|/2-2}}, \quad (30)$$

again using that X_1 is uniformly distributed on the complete K -ary tree of height $\lceil |T|/2 - 1 \rceil$ (excluding the root).

It remains to upper bound $\mathbb{P}(\mathcal{E}_4)$, which requires more work. In the following, for a vertex v and a time $t > 0$, we say that the clock of v has rung before time t if $0 < \tau_v < t$. We introduce the following two events:

- (i) $\mathcal{F}_i := \{\text{At time } (\tau_{X_i})^-, \text{ strictly more than } C \text{ clocks have rung in } T_{X_i}\}.$
- (ii) $\mathcal{O}_i := \{\text{A clock has rung in } B_{\geq |T|/2}(X_i, |X_i|) \text{ before time } (\tau_{X_i})^-\}.$

Lemma 29. *Assume that $H < |T|/4$. Then we have*

$$\mathcal{E}_4 \subset \{\forall i \leq N, \mathcal{O}_i \cup \mathcal{F}_i\} \cap \mathcal{E}_1 \cap \mathcal{E}_2.$$

Proof. Recalling the definition of \mathcal{A}_i given in (26), we have $\mathcal{A}_i := \mathcal{X}_i \cap \mathcal{Y}_i \cap \mathcal{Z}_i$, where

$$\begin{aligned} \mathcal{X}_i &:= \{|X_i| > |T|/2 - H\}, \\ \mathcal{Y}_i &:= \{X_i \text{ is colored blue}\}, \\ \mathcal{Z}_i &:= \left\{ \text{At time } (\tau_{X_i})^-, T_{X_i} \text{ contains at most } C \text{ colored vertices that are not leaves,} \right. \\ &\quad \left. \text{and each of them is at distance at most } 2H \text{ from the leaves} \right\}. \end{aligned}$$

We prove that

$$\mathcal{F}_i^c \cap \mathcal{O}_i^c \cap \mathcal{E}_1 \cap \mathcal{E}_2 \cap \mathcal{E}_3 \subset \mathcal{A}_i,$$

by showing that the event on the left-hand side implies \mathcal{X}_i , \mathcal{Y}_i , and \mathcal{Z}_i .

We have $\mathcal{E}_2 \subset \mathcal{X}_i$. Assume now that $\mathcal{O}_i^c \cap \mathcal{E}_1 \cap \mathcal{E}_2 \cap \mathcal{E}_3$ occurs. Observe that for $i \neq j \leq N$,

$$d(X_i, X_j) = |X_i| + |X_j| - 2|X_i \wedge X_j| > |X_i| + \frac{|T|}{2} - H - \frac{|T|}{4} > |X_i|.$$

Hence, we have $X_j \notin B(X_i, |X_i|)$. By definition of the sequence $(X_j)_{j \leq N}$, it follows that at time $(\tau_{X_i})^-$ no clock in $B(X_i, |X_i|)$ at height smaller than $|T|/2$ has rung. Moreover, since \mathcal{O}_i^c holds, no clock in $B(X_i, |X_i|)$ at height at least $|T|/2$ has rung either. Therefore, no clock in $B(X_i, |X_i|)$ has rung before time $(\tau_{X_i})^-$. Using now \mathcal{E}_3 and the fact that $|X_i| < |T|/2$, we deduce that $B(X_i, |X_i|)$ contains no vertex of R_0 . Hence, at time $(\tau_{X_i})^-$, the ball $B(X_i, |X_i|)$ contains no red vertex but does contain the root, which is blue. Consequently, X_i becomes blue, and we obtain

$$\mathcal{O}_i^c \cap \mathcal{E}_1 \cap \mathcal{E}_2 \cap \mathcal{E}_3 \subset \mathcal{Y}_i.$$

Moreover, on the event $\mathcal{O}_i^c \cap \mathcal{E}_1 \cap \mathcal{E}_2$, we have $|X_i| > |T|/2 - H$ and no clock has rung in $B(X_i, |X_i|)$ before time τ_{X_i} . In particular, any clock that has rung in T_{X_i} before time τ_{X_i} must correspond to a vertex at height at least $2|X_i| \geq |T| - 2H$. Hence, by definition of \mathcal{F}_i , on the event $\mathcal{F}_i^c \cap \mathcal{O}_i^c \cap \mathcal{E}_1 \cap \mathcal{E}_2$, at most C vertices have rung in T_{X_i} before time τ_{X_i} , and all of them lie at distance at most $2H$ from the leaves. Using additionally \mathcal{E}_3 , which ensures that the initial red vertices in T_{X_i} are only leaves, we obtain

$$\mathcal{F}_i^c \cap \mathcal{O}_i^c \cap \mathcal{E}_1 \cap \mathcal{E}_2 \cap \mathcal{E}_3 \subset \mathcal{Z}_i.$$

Combining the previous observations, we deduce that

$$\mathcal{F}_i^c \cap \mathcal{O}_i^c \cap \mathcal{E}_1 \cap \mathcal{E}_2 \cap \mathcal{E}_3 \subset \mathcal{A}_i,$$

which implies

$$\mathcal{A}_i^c \cap \mathcal{E}_1 \cap \mathcal{E}_2 \cap \mathcal{E}_3 \subset \mathcal{F}_i \cup \mathcal{O}_i.$$

Hence,

$$\mathcal{E}_4 \subset \left\{ \forall i \leq N, \mathcal{F}_i \cup \mathcal{O}_i \right\} \cap \mathcal{E}_1 \cap \mathcal{E}_2.$$

This concludes the proof of the lemma. \square

Let us now derive an upper bound on the probability of $\{\forall i \leq N, \mathcal{F}_i \cup \mathcal{O}_i\} \cap \mathcal{E}_1 \cap \mathcal{E}_2$. Recall that τ_v denotes the coloring time of a vertex v . We define the σ -algebra

$$\mathcal{H} := \sigma(\{\tau_v : |v| < |T|/2\}),$$

so that $\mathcal{E}_1, \mathcal{E}_2, \mathcal{E}_3 \in \mathcal{H}$. For $i \leq N$, we denote by τ_i the coloring time of X_i , i.e. $\tau_i = \tau_{X_i}$. Note that τ_i is \mathcal{H} -measurable. We observe that for any $i \leq N$ and any $u \in B_{\geq |T|/2}(X_i, |X_i|)$, we have

$$2|u \wedge X_i| = |u| + |X_i| - d(u, X_i) \geq |u| \geq \frac{|T|}{2}.$$

Moreover, if $u \in T_{X_i}$, then $|u \wedge X_i| = |X_i|$. Hence, assuming $H < |T|/4$, on the event \mathcal{E}_2 we obtain that for all $u \in B_{\geq |T|/2}(X_i, |X_i|) \cup T_{X_i}$,

$$|u \wedge X_i| > \frac{|T|}{8}.$$

In particular, on $\mathcal{E}_1 \cap \mathcal{E}_2$, for any $i \neq j \leq N$, the sets $B_{\geq |T|/2}(X_i, |X_i|) \cup T_{X_i}$ and $B_{\geq |T|/2}(X_j, |X_j|) \cup T_{X_j}$ are disjoint. Therefore, conditionally on \mathcal{H} and on $\mathcal{E}_1 \cap \mathcal{E}_2$, the events $\mathcal{F}_i \cup \mathcal{O}_i$, $i \leq N$, are independent. Besides, we have

$$\#B_{\geq |T|/2}(X_i, |X_i|) \leq (K+1)K^{|X_i|} \quad \text{and} \quad \#T_{X_i} \leq K^{|T|/2+H+1}.$$

Hence, conditionally on \mathcal{H} and on the event $\mathcal{E}_1 \cap \mathcal{E}_2$, the time until the first clock rings in $B_{\geq |T|/2}(X_i, |X_i|)$, namely

$$\min\{\tau_v : v \in B_{\geq |T|/2}(X_i, |X_i|)\},$$

is stochastically lower bounded by an exponential random variable with parameter $(K+1)K^{|X_i|}$. Moreover, the number of clocks that have rung before time τ_i in T_{X_i} is stochastically upper bounded by a binomial random variable with parameters $(K^{|T|/2+H+1}, 1 - e^{-\tau_i})$. Let e_h denote an exponential random variable with parameter $(K+1)K^h$ and let Y_t be a binomial random variable with parameters $(K^{|T|/2+H+1}, 1 - e^{-t})$. We obtain

$$\begin{aligned} \mathbb{P}(\mathcal{E}_4 \mid \mathcal{H}) &\leq \mathbb{1}_{\{\mathcal{E}_1 \cap \mathcal{E}_2\}} \prod_{i=1}^N \mathbb{P}(e_{|X_i|} \leq \tau_i \text{ or } Y_{\tau_i} \geq C \mid \mathcal{H}) \\ &= \mathbb{1}_{\{\mathcal{E}_1 \cap \mathcal{E}_2\}} \prod_{i=1}^N \left(1 - \mathbb{P}(e_{|X_i|} > \tau_i \mid \mathcal{H}) \mathbb{P}(Y_{\tau_i} < C \mid \tau_i)\right). \end{aligned}$$

We write, for a fixed $A > 0$,

$$\mathbb{P}(Y_{\tau_i} < C \mid \tau_i) \geq \mathbb{1}_{\{\tau_i \leq AK^{-|T|/2}\}} \mathbb{P}(Y_{AK^{-|T|/2}} < C).$$

We set $q_{A,C,|T|} := \mathbb{P}(Y_{AK^{-|T|/2}} < C)$ so that

$$\mathbb{P}(\mathcal{E}_4 \mid \mathcal{H}) \leq \mathbb{1}_{\{\mathcal{E}_1 \cap \mathcal{E}_2\}} \prod_{i=1}^N \left(1 - \exp(-(K+1)K^{|X_i|}\tau_i) q_{A,C,|T|} \mathbb{1}_{\{\tau_i \leq AK^{-|T|/2}\}} \right).$$

Let m denote the largest integer strictly smaller than $|T|/2$, and let T_m^* be the set of vertices of $T \setminus \{o\}$ whose height is at most m . We have

$$\#T_m^* = \frac{K(K^m - 1)}{K - 1}.$$

Note that X_1, \dots, X_N are N distinct vertices of T_m^* chosen uniformly at random. Moreover, the random variables $(\tau_i)_{i \leq N}$ and $(X_i)_{i \leq N}$ are independent, and $(\tau_i - \tau_{i-1})_{i \leq N}$ are independent exponential random variables with parameters $\#T_m^* - i$. We obtain, for any $i \geq 1$ and $h < m$,

$$\mathbb{P}(|X_i| = m - h) = \frac{K^{m-h}}{\#T_m^*} = \frac{(K-1)K^{m-h}}{K(K^m - 1)} \geq \frac{1}{2K^h}.$$

Moreover, for any $h_i \geq 1$, we have

$$\begin{aligned} \mathbb{P}((|X_1|, \dots, |X_N|) = (h_1, \dots, h_N)) &\leq \left(\frac{\#T_m^*}{\#T_m^* - N} \right)^N \prod_{i=1}^N \mathbb{P}(|X_i| = h_i) \\ &\leq \left(1 - \frac{N}{K^m} \right)^{-N} \prod_{i=1}^N \mathbb{P}(|X_i| = h_i) \\ &\leq \left(1 - \frac{N}{K^{|T|/2-1}} \right)^{-N} \prod_{i=1}^N \mathbb{P}(|X_i| = h_i). \end{aligned}$$

Thus, we get

$$\begin{aligned} \mathbb{P}(\mathcal{E}_4 \mid (\tau_i)_{i \leq N}) &\leq \sum_{h_1, \dots, h_N \geq 1} \mathbb{P}((|X_1|, \dots, |X_N|) = (h_1, \dots, h_N)) \prod_{i=1}^N \left(1 - e^{-\tau_i(K+1)K^{h_i}} q_{A,C,|T|} \mathbb{1}_{\{\tau_i \leq AK^{-|T|/2}\}} \right) \\ &\leq \left(1 - \frac{N}{K^{|T|/2-1}} \right)^{-N} \sum_{h_1, \dots, h_N \geq 0} \prod_{i=1}^N \left(\mathbb{P}(|X_i| = h_i) (1 - e^{-\tau_i(K+1)K^{h_i}} q_{A,C,|T|} \mathbb{1}_{\{\tau_i \leq AK^{-|T|/2}\}}) \right) \\ &\leq \left(1 - \frac{N}{K^{|T|/2-1}} \right)^{-N} \prod_{i=1}^N \mathbb{E} \left(1 - e^{-\tau_i(K+1)K^{|X_i|}} q_{A,C,|T|} \mathbb{1}_{\{\tau_i \leq AK^{-|T|/2}\}} \mid (\tau_i)_{i \leq N} \right) \\ &= \left(1 - \frac{N}{K^{|T|/2-1}} \right)^{-N} \prod_{i=1}^N \left(1 - \mathbb{E} \left(e^{-\tau_i(K+1)K^{|X_i|}} \mid (\tau_i)_{i \leq N} \right) q_{A,C,|T|} \mathbb{1}_{\{\tau_i \leq AK^{-|T|/2}\}} \right) \\ &\leq \left(1 - \frac{N}{K^{|T|/2-1}} \right)^{-N} \prod_{i=1}^N \left(1 - \left(\sum_{h=0}^L \frac{1}{2K^h} e^{-\tau_i(K+1)K^{m-h}} \right) q_{A,C,|T|} \mathbb{1}_{\{\tau_i \leq AK^{-|T|/2}\}} \right) \end{aligned}$$

valid for any $L < m$ (recall that m is the largest integer strictly smaller than $|T|/2$).

We now use that the random variables $(\tau_i - \tau_{i-1})_{i \leq N}$ are independent exponential random variables with parameters $\#T_m^* - i$. If $i \leq \#T_m^*/2$, we have

$$\#T_m^* - i \geq \frac{\#T_m^*}{2} \geq K^{m-1} \geq K^{|T|/2-2}.$$

Thus, if $N \leq K^m/2$, the sequence $(K^{|T|/2-2}(\tau_i - \tau_{i-1}))_{i \leq N}$ is stochastically smaller than a sequence $(\varepsilon_i)_{i \leq N}$ of i.i.d. exponential random variables with parameter 1. Let $\sigma_i = \sum_{k=1}^i \varepsilon_k$. We get

$$\begin{aligned} \mathbb{P}(\mathcal{E}_4) &\leq \left(1 - \frac{N}{K^{|T|/2-1}} \right)^{-N} \mathbb{E} \left(\prod_{i=1}^N \left(1 - \left(\sum_{h=0}^L \frac{1}{2K^h} e^{-\tau_i(K+1)K^{|T|/2-h}} \right) q_{A,C,|T|} \mathbb{1}_{\{\tau_i \leq AK^{-|T|/2}\}} \right) \right) \\ &\leq \left(1 - \frac{N}{K^{|T|/2-1}} \right)^{-N} \mathbb{E} \left(\prod_{i=1}^N \left(1 - \left(\sum_{h=0}^L \frac{1}{2K^h} e^{-\sigma_i(K+1)K^{2-h}} \right) q_{A,C,|T|} \mathbb{1}_{\{\sigma_i \leq AK^{-2}\}} \right) \right). \end{aligned}$$

We now use the following lemma, whose proof is postponed until the end of this section.

Lemma 30. *Let $\eta > 0$. Then, there exist integers $N, L > 0$ such that if $(\varepsilon_i)_{i \leq N}$ are i.i.d. exponential random variables with parameter 1 and $\sigma_i = \sum_{k=1}^i \varepsilon_k$, we have*

$$\mathbb{E} \left(\prod_{i=1}^N \left(1 - \sum_{h=0}^L \frac{1}{2K^h} e^{-\sigma_i(K+1)K^{2-h}} \right) \right) \leq \eta/5. \quad (31)$$

We finish the proof of Proposition 28 assuming that Lemma 30 holds. Fix $\eta > 0$ and choose N, L such that (31) holds. Since for each $i \leq N$, $1_{\{\sigma_i \leq AK^{-2}\}}$ tends a.s. to 1 as A tends to infinity, by the dominated convergence theorem, there exists A large enough such that

$$\mathbb{E} \left(\prod_{i=1}^N \left(1 - 1_{\{\sigma_i \leq AK^{-2}\}} \sum_{h=0}^L \frac{1}{2K^h} e^{-\sigma_i(K+1)K^{2-h}} \right) \right) \leq \eta/4.$$

Using (29), fix now H such that

$$\mathbb{P}(\mathcal{E}_2^c) \leq N \frac{1}{K^H} \leq \eta/3.$$

Recall that

$$q_{A,C,|T|} := \mathbb{P}(Y_{AK^{-|T|/2}} < C) \text{ where } Y_{AK^{-|T|/2}} \sim \text{Bin}(K^{|T|/2+H+1}, 1 - e^{-AK^{-|T|/2}}).$$

Thus, we get that

$$1 - q_{A,C,|T|} = \mathbb{P}(Y_{AK^{-|T|/2}} \geq C) \leq \frac{\mathbb{E}(Y_{AK^{-|T|/2}})}{C} \leq \frac{AK^{H+1}}{C}.$$

This shows that, for fixed A and H , we can find C large enough such that $q_{A,C,|T|}$ is close to 1. Using again the dominated convergence theorem, we deduce the existence of a C such that

$$\mathbb{E} \left(\prod_{i=1}^N \left(1 - \left(\sum_{h=0}^L \frac{1}{2K^h} e^{-\sigma_i(K+1)K^{2-h}} \right) q_{A,C,|T|} 1_{\{\sigma_i \leq AK^{-2}\}} \right) \right) \leq \eta/3,$$

and so, for this choice of N, H, C , we obtain $\mathbb{P}(\mathcal{E}_4) \leq \eta/2$ if $|T|$ is large enough. Combining this with (28) and (30), we conclude that for this choice of H, C, N and for $|T|$ large enough,

$$\mathbb{P}(\forall 1 \leq i \leq N, \mathcal{A}_i^c) \leq \eta.$$

This is exactly the statement of Proposition 28. \square

Proof of Lemma 30. Let us note that σ_i is the sum of i i.i.d. exponential random variables; thus σ_i/i tends a.s. to 1 as i tends to infinity. Hence, we can find $D > 0$ such that

$$\mathbb{P}(\exists i \geq 1, K^2(K+1)\sigma_i \geq Di) \leq \eta/10.$$

With this choice for D , we get

$$\mathbb{E} \left(\prod_{i=1}^N \left(1 - \sum_{h=0}^L \frac{1}{2K^h} e^{-\sigma_i(K+1)K^{2-h}} \right) \right) \leq \eta/10 + \prod_{i=1}^N \left(1 - \sum_{h=0}^L \frac{1}{2K^h} e^{-DiK^{-h}} \right).$$

Define

$$\Pi_{L,N} := \prod_{i=1}^N \left(1 - \sum_{h=0}^L \frac{1}{2K^h} e^{-DiK^{-h}} \right)$$

and

$$\Pi_\infty := \prod_{i=1}^\infty \left(1 - \sum_{h=0}^\infty \frac{1}{2K^h} e^{-DiK^{-h}} \right).$$

Since $\Pi_{L,N}$ tends to Π_∞ when L, N tend to infinity, it only remains to show that $\Pi_\infty = 0$, as this will imply that, for L and N large enough, $\Pi_{L,N} \leq \eta/10$, as required. Indeed, we have

$$\log \Pi_\infty = \sum_{i=1}^{\infty} \log \left(1 - \sum_{h=0}^{\infty} \frac{1}{2K^h} e^{-DiK^{-h}} \right) \leq - \sum_{i=1}^{\infty} \sum_{h=0}^{\infty} \frac{1}{2K^h} e^{-DiK^{-h}},$$

and

$$\sum_{h=0}^{\infty} \sum_{i=1}^{\infty} \frac{1}{2K^h} e^{-DiK^{-h}} = \sum_{h=0}^{\infty} \frac{1}{2K^h} \frac{e^{-DK^{-h}}}{1 - e^{-DK^{-h}}} = \infty,$$

since

$$\lim_{h \rightarrow \infty} \frac{1}{2K^h} \frac{e^{-DK^{-h}}}{1 - e^{-DK^{-h}}} = \frac{1}{2D}.$$

This yields that $\Pi_\infty = 0$, and so $\Pi_{L,N}$ tends to 0 as N and L tend to infinity. \square

6.3 Proof of Proposition 26

We first prove a two-window version of Proposition 28: with high probability, one can find a good vertex among the first N candidates and another one among the candidates with indices between $N + 1$ and N' . The final step, carried out afterwards, will be to show that these two good vertices may be chosen with disjoint descendant subtrees, as required in Proposition 26.

Proposition 31. *Let $\eta > 0$. Then, there exist $H, C, N, N', n_0 \geq 0$ such that, if $|T| \geq n_0$, then for any choice of vertices w_1, \dots, w_C of height greater than or equal to $|T| - 2H$, starting from $B_0 = \{o\}$ and $R_0 = \partial T \cup \{w_1, \dots, w_C\}$, we have*

$$\mathbb{P}(\forall 1 \leq i \leq N, \mathcal{A}_i^c) \leq \eta. \quad (32)$$

$$\mathbb{P}(\forall N < i \leq N', \mathcal{A}_i^c) \leq \eta. \quad (33)$$

Proof. For $N < N'$, we introduce the following events:

$$\begin{aligned} \mathcal{E}'_1 &= \mathcal{E}_1(T, N', H, C, R_0) &:= \{ \forall i \neq j \leq N', |X_i \wedge X_j| < |T|/8 \}, \\ \mathcal{E}'_2 &= \mathcal{E}_2(T, N', H, C, R_0) &:= \{ \forall i \leq N', |X_i| > |T|/2 - H \}, \\ \mathcal{E}'_3 &= \mathcal{E}_3(T, N', H, C, R_0) &:= \{ \forall i \leq N', \forall k \leq C, X_i \notin B(w_k, |T|/2 + H) \}, \\ \mathcal{E}'_4 &= \mathcal{E}'_4(T, N, N', H, C, R_0) &:= \{ \forall i \leq N, \mathcal{A}_i^c \} \cap \mathcal{E}'_1 \cap \mathcal{E}'_2 \cap \mathcal{E}'_3, \\ \mathcal{E}'_5 &= \mathcal{E}'_5(T, N, N', H, C, R_0) &:= \{ \forall N < i \leq N', \mathcal{A}_i^c \} \cap \mathcal{E}'_1 \cap \mathcal{E}'_2 \cap \mathcal{E}'_3. \end{aligned}$$

Note that $\mathbb{P}(\mathcal{E}'_4) \leq \mathbb{P}(\mathcal{E}_4)$. Moreover, arguing exactly as in the proof of Proposition 28, we obtain

$$\mathbb{P}(\mathcal{E}'_5) \leq \left(1 - \frac{N'}{K^{|T|/2-1}} \right)^{-N'} \mathbb{E} \left[\prod_{i=N+1}^{N'} \left(1 - \left(\sum_{h=0}^L \frac{1}{2K^h} e^{-\sigma_i(K+1)K^{2-h}} \right) q_{A,C,|T|} 1_{\{\sigma_i \leq AK^{-2}\}} \right) \right],$$

where $\sigma_i = \sum_{k=1}^i \varepsilon_k$, and $(\varepsilon_k)_{k \leq N'}$ are i.i.d. exponential random variables with parameter 1. Let $\eta > 0$. Choose N, N' , and L with $N' > N$ such that

$$\mathbb{E} \left(\prod_{i=1}^N \left(1 - \sum_{h=0}^L \frac{1}{2K^h} e^{-\sigma_i(K+1)K^{2-h}} \right) \right) \leq \frac{\eta}{5},$$

and

$$\mathbb{E} \left(\prod_{i=N+1}^{N'} \left(1 - \sum_{h=0}^L \frac{1}{2K^h} e^{-\sigma_i(K+1)K^{2-h}} \right) \right) \leq \frac{\eta}{5}.$$

Finally, choosing A, H , and C large enough exactly as in the proof of Proposition 28, we obtain that, for this choice of parameters (N, N', H, C) , both (32) and (33) hold whenever $|T|$ is large enough. \square

Proof of Proposition 26. Let H, C, N, N' , and n_0 be chosen so that Proposition 31 holds. Then, for every $|T| \geq n_0$, with probability at least $1 - 2\eta$, there exist indices $i < N < j$ such that both \mathcal{A}_i and \mathcal{A}_j occur. Recall that X_1, X_2, \dots denote the vertices of $T \setminus \{o\}$ of height smaller than $|T|/2$, ordered according to their coloring times. We have

$$\mathbb{P}(\exists i < j \leq N' : T_{X_i} \cap T_{X_j} \neq \emptyset) \leq N'^2 \mathbb{P}(T_{X_1} \cap T_{X_2} \neq \emptyset) \leq \frac{|T| N'^2}{K^{|T|/2-1}}.$$

Therefore,

$$\mathbb{P}(\forall i < j \leq N', T_{X_i} \cap T_{X_j} = \emptyset) \longrightarrow 1 \quad \text{as } |T| \rightarrow \infty.$$

Combining this estimate with Proposition 31, we obtain that, for $|T|$ sufficiently large,

$$\mathbb{P}(\exists i < j, T_{X_i} \cap T_{X_j} = \emptyset \text{ and } \mathcal{A}_i \cap \mathcal{A}_j) \geq 1 - 3\eta.$$

This completes the proof of Proposition 26. □

Acknowledgements

The authors warmly thank Nicolas Broutin and Guillaume Chapuy for insightful discussions during the early stages of this work.

References

- [1] D. Aldous. Random partitions of the plane via Poissonian coloring and a self-similar process of coalescing planar partitions. *The Annals of Probability*, 46(4):2000 – 2037, 2018.
- [2] B. C Arnold, N. Balakrishnan, and H. N Nagaraja. *Records*. John Wiley & Sons, 2011.
- [3] A-L. Basdevant, G. Blanc, N. Curien, and A. Singh. Fractal properties of the frontier in Poissonian coloring. *ALEA Lat. Am. J. Probab. Math. Stat.*, 20:1565–1586, 2023.
- [4] J. Casse. Siblings in d-dimensional nearest neighbour trees. *Bernoulli*, 31(2):1300–1324, 2025.
- [5] M. d’Achille, N. Curien, N. Enriquez, R. Lyons, and M. Ünel. Ideal Poisson–Voronoi tessellations on hyperbolic spaces. Revised version, 49 pages, 12 figures, March 2023.
- [6] W. Dziubdziela and B. Kopociński. Limiting properties of the k-th record values. *Applicaciones Mathematicae*, 15(2):187–190, 1976.
- [7] L. Lichev and D. Mitsche. New results for the random nearest neighbor tree. *Probability Theory and Related Fields*, 189:229–279, 2024.
- [8] M. D. Penrose and A. R. Wade. Limit theory for the random on-line nearest-neighbor graph. *Random Structures & Algorithms*, 32(2):125–156, 2008.
- [9] M. D. Penrose and A. R. Wade. Random directed and on-line networks. In W. S. Kendall and I. Molchanov, editors, *New Perspectives in Stochastic Geometry*, pages 248–275. Oxford University Press, Oxford, 2009.
- [10] J. Preater. A species of voter model driven by immigration. *Statistics & Probability Letters*, 79(20):2131–2137, 2009.