J. Appl. Prob. **56,** 533–545 (2019)
*doi:*10.1017/jpr.2019.29

ACCESSIBILITY PERCOLATION ON RANDOM ROOTED LABELED TREES

ZHISHUI HU,*
ZHENG LI,* AND
QUNQIANG FENG,* ** *University of Science and Technology of China*

Abstract

The accessibility percolation model is investigated on random rooted labeled trees. More precisely, the number of accessible leaves (i.e. increasing paths) Z_n and the number of accessible vertices C_n in a random rooted labeled tree of size n are jointly considered in this work. As $n \to \infty$, we prove that (Z_n, C_n) converges in distribution to a random vector whose probability generating function is given in an explicit form. In particular, we obtain that the asymptotic distributions of $Z_n + 1$ and C_n are geometric distributions with parameters $\mathrm{e}/(1+\mathrm{e})$ and $1/\mathrm{e}$, respectively. Much of our analysis is performed in the context of local weak convergence of random rooted labeled trees.

Keywords: Accessibility percolation; random tree; increasing path; accessible vertex; local weak convergence

2010 Mathematics Subject Classification: Primary 05C05; 60C05
Secondary 60F05

1. Introduction

Accessibility percolation models were inspired by some recent work in evolutionary biology and were developed in order to answer questions of mathematical interest. Consider a rooted graph G with a unique root o. To each vertex $v \in G$, a continuous random variable, denoted by X_v, is assigned. We always assume that these random variables $\{X_v, v \in G\}$ are independently and identically distributed (i.i.d.). Let $P = ov_1v_2 \cdots v_kv$ be a path from the root to vertex v in G, where $k \geq 0$ is an integer. If the assigned random variables form an increasing sequence, i.e.

$$X_o < X_{v_1} < \cdots < X_{v_k} < X_v,$$

we say that vertex v is *accessible* and the path P is *increasing*.

Interest in increasing paths and accessible vertices was motivated by the classical model for the evolution of an organism involving genetic mutation and selection (see Weinreich *et al.* (2006); Weinreich, Watson, and Chao (2005)). In evolutionary biology literature, the assigned random variables are known as *fitness values*, and increasing paths as *selectively accessible paths* (see, for example, Franke *et al.* (2011)). The existence of increasing paths is also called *accessibility percolation* by Nowak and Krug (2013). For more details on biological motivation, we refer to Roberts and Zhao (2013) and Berestycki, Brunet and Shi (2016).

Received 2 July 2018; revision received 18 March 2019.
* Postal address: Department of Statistics and Finance, School of Management, University of Science and Technology of China, Hefei 230026, China.
** Email address: fengqq@ustc.edu.cn

The phase transition of existence or the numbers of increasing paths in different deterministic graphs have been studied by several authors. For such problems in regular n-ary rooted trees, where each vertex has exactly n children, see Roberts and Zhao (2013) and Chen (2014). Recently, Coletti, Gava, and Rodríguez (2018) considered the existence problem of increasing paths in infinite spherically symmetric trees, where the degrees of vertices grow exponentially as their depth increases. And for the increasing paths from $(0, 0, \dots, 0)$ to $(1, 1, \dots, 1)$ in an n-dimensional binary hypercube $\{0, 1\}^n$ under various assumptions, we refer to Hegarty and Martinsson (2014), Berestycki, Brunet and Shi (2016), and Li (2017).

The purpose of this work is to establish several limiting theorems of the accessibility percolation model on a class of rooted trees, not only for the number of increasing paths but also the cluster size in this percolation. Furthermore, the underlying graphs we consider here are random trees: *random rooted labeled trees* (or *uniform rooted Cayley trees*), which are a classic combinatorial model. According to Cayley's formula, the number of unrooted labeled trees on n vertices is n^{n-2}. Thus, the set of rooted labeled trees on n vertices has cardinality n^{n-1}. It is natural to assume a uniform model on the set of such trees of the same size. That is, we always assume that all n^{n-1} rooted labeled trees on n vertices occur with equal probability. Let T_n be a random rooted labeled tree on n vertices.

In a rooted tree, there is a unique path from the root to any non-root vertex. However, we only consider the paths from the root to leaves when the increasing paths are counted. As a result, the number of increasing paths is equal to that of *accessible leaves* in any rooted tree. Moreover, it is easy to see that the cluster size in accessibility percolation is exactly the number of *accessible vertices* in any rooted graph. We denote by Z_n the number of accessible leaves and C_n the number of accessible vertices in a random rooted labeled tree T_n. The asymptotic distributions of these two random variables are our main concern.

Since we are only concerned with the order of the fitness values, under the assumption of continuity of their distribution, changing the precise distribution will not influence the results. Without loss of generality, we can assume throughout this work that all the fitness values of vertices in a tree are mutually independent and uniformly distributed on the interval $[0, 1]$. It is not hard to see that the probability that a path of length n is increasing (or a vertex with depth n is accessible) is equal to $1/(n+1)!$.

Our main result in this work is stated in the following.

Theorem 1. *Let Z_n and C_n be the numbers of accessible leaves and accessible vertices in a random rooted labeled tree T_n, respectively. Then the probability generating function of the random vector (Z_n, C_n) satisfies that, for any $0 \le s, t \le 1$,*

$$\lim_{n\to\infty} \mathrm{E}[s^{Z_n} t^{C_n}] = \frac{1}{s_1} + \frac{t-1}{t\mathrm{e}^{-s_1} + s_1 - t}, \tag{1}$$

where $s_1 = \mathrm{e}^{-1}(1-s)t + 1$. Consequently, the asymptotic distributions of $Z_n + 1$ and C_n are geometric distributions with parameters $\mathrm{e}/(1+\mathrm{e})$ and $1/\mathrm{e}$, respectively. That is,

$$\lim_{n\to\infty} \mathrm{P}(Z_n = k) = \frac{\mathrm{e}}{(1+\mathrm{e})^{k+1}}, \qquad k = 0, 1, 2, \dots$$

and

$$\lim_{n\to\infty} \mathrm{P}(C_n = k) = \frac{(\mathrm{e}-1)^{k-1}}{\mathrm{e}^{k}}, \qquad k = 1, 2, \dots$$

Throughout, we shall use the following notation. We denote by $|T|$ the size of a tree T, i.e. the number of vertices in T. Let $\mathbf{1}(E)$ denote the indicator of the event E. For probabilistic convergence we use $\xrightarrow{\text{D}}$ and $\xrightarrow{\text{P}}$ to denote convergence in distribution and in probability, respectively.

The rest of the paper is organized as follows. In Section 2 we introduce several definitions and state some basic results. In addition, the numbers of accessible leaves and accessible vertices in two related random trees are also discussed. Section 3 is devoted to the proof of our main theorem stated above.

2. Preliminaries

In order to obtain the asymptotic distributions of the number of accessible leaves Z_n and the number of accessible vertices C_n in a random rooted labeled tree T_n, our analysis is performed in the context of the local weak convergence of such random trees as the size $n \to \infty$. We shall first introduce the basic concept of local weak convergence, and then precisely describe the limiting distribution of random rooted labeled trees.

We also consider in this section the probability measure under the condition that the fitness value of the root $X_o = x \in [0, 1]$ is given, i.e.,

$$\mathrm{P}_x(\,\cdot\,) = \mathrm{P}(\cdot \mid X_o = x),$$

and denote by E_x the expectation with respect to P_x.

2.1. Local weak convergence

We now formally introduce the definition of local weak convergence of random rooted trees.

Let $\mathcal{T}$ denote the set of locally finite rooted trees, considered up to rooted isomorphism. Here, the terminology rooted isomorphism means that there exists an isomorphism $\phi : (T_1, o_1) \to (T_2, o_2)$ for the two rooted trees T_1 and T_2. Then, each vertex in any tree $T \in \mathcal{T}$ has finite degree, and for any two trees $T_1, T_2 \in \mathcal{T}$, we regard them as an identical element in $\mathcal{T}$ if $T_1 \cong T_2$, where '$\cong$' denotes rooted isomorphism. For any tree $T \in \mathcal{T}$ and integer $m \geq 0$, let $T^{(m)}$ denoted the subgraph induced by all vertices whose depths are at most m. That is, $T^{(m)}$ contains all the vertices which are at a distance of at most m from the root, and all edges of T that join these vertices. Then we can make $\mathcal{T}$ into a metric space by endowing it with the distance d_{loc} defined by

$$d_{\text{loc}}(T_1, T_2) = 2^{-\sup\{m \geq 0 : T_1^{(m)} \cong T_2^{(m)}\}}, \qquad T_1, T_2 \in \mathcal{T}.$$

Now let T and $T_n (n = 1, 2, \ldots)$ be random rooted locally finite trees. Then, following Benjamini and Schramm (2001) and Aldous and Steele (2004), if $T_n \xrightarrow{\text{D}} T$ as $n \to \infty$ with respect to this topology, we say T is the *local weak limit* of T_n, or T_n converges weakly in $\mathcal{T}$ to T.

The concept of local weak convergence can be easily extended to random rooted locally finite graphs. For more information on this field, we refer to Aldous and Steele (2004). We shall precisely describe the local weak limit of random rooted labeled trees in the following subsections.

2.2. The branching Poisson tree

The family of *branching Poisson trees* is a subset of Galton–Watson trees (see, for example, Drmota (2009)), where the common offspring distribution is a Poisson distribution. Let $\lambda > 0$ be a given real number. A branching Poisson tree with parameter λ can be generated in a plane as follows. Start with a single vertex (the root); each vertex independently has a number of children distributed as a Poisson distribution with parameter λ. Let PGW(λ) denote the distribution law of the family of branching Poisson trees with parameter λ.

In a rooted tree T, let $W_i(T)$ denote the number of vertices of depth i for any $i \geq 0$. Several properties of PGW(λ) trees are listed in the following.

Lemma 1. *In a random rooted labeled tree T_n, the sequence $\{W_i(T_n), 1 \leq i \leq n-1\}$ has the same joint distribution as $\{W_i(PGW(\lambda)){:}1 \leq i \leq n-1\}$ conditioned on $|PGW(\lambda)| = n$. That is, for any sequence of nonnegative integers $a_1, a_2, \ldots, a_{n-1}$ such that $\sum_{i=1}^{n-1} a_i = n-1$, we have*

$$\mathrm{P}(W_i(T_n) = a_i,\ 1 \leq i \leq n-1) = \mathrm{P}(W_i(PGW(\lambda)) = a_i,\ 1 \leq i \leq n-1 \,|\, |PGW(\lambda)| = n).$$

Proof. See Kolchin (1977), or Theorem 1 in Grimmett (1980). □

Lemma 2. *For all n and $i \geq 1$, we have*

$$\mathrm{E}[W_i^r(PGW(\lambda)) \,|\, |PGW(\lambda)| = n] \leq c\, i^r, \qquad r = 1, 2, \ldots,$$

where c is a constant depending on r only.

Proof. This is an immediate consequence of Theorem 1.13 in Janson (2006). □

We point out that the known results in the above two lemmas are independent of λ, and they are applied in the next section only in the critical case $\lambda = 1$.

For the numbers of accessible leaves and accessible vertices in a PGW(λ) tree, we have the following results, which are of interest in their own right.

Lemma 3. *Let Z_λ and C_λ be the numbers of accessible leaves and accessible vertices in a PGW(λ) tree, respectively. If the fitness value of the root $x \in [0, 1]$ is given, then the conditional probability generating function of (Z_λ, C_λ) is*

$$f_\lambda(x, s, t) = \mathrm{E}_x[s^{Z_\lambda} t^{C_\lambda}] = \frac{s_\lambda t \mathrm{e}^{-\lambda(1-x)s_\lambda}}{t\mathrm{e}^{-\lambda(1-x)s_\lambda} + s_\lambda - t}, \qquad 0 \leq s, t \leq 1, \tag{2}$$

where $s_\lambda := \mathrm{e}^{-\lambda}(1-s)t + 1$.

Proof. Without loss of generality, we may assume that the children of the root o are labeled by $v_1, \ldots, v_N$, where N is a random variable, Poisson distributed with parameter λ. If a graph only contains a single vertex, there does not exist a path. Therefore, we always define $Z_\lambda = 0$ (and $C_\lambda = 1$) if $N = 0$. However, it is more convenient to work with a modified version of Z_λ that does not get affected if $N \geq 1$, but rather only requires to replace its value by 1 if $N = 0$. Denote the modified version of Z_λ by $\widetilde{Z}_\lambda$. We also define $\widetilde{Z}_{v_i}$ as the modified number of accessible leaves in the subtree rooted at v_i. Recall that the fitness values $X_o, X_{v_1}, X_{v_2}, \ldots$ are i.i.d. random variables uniformly distributed on [0, 1]. Then, it is not hard to see that

$$Z_\lambda = \sum_{i=1}^{N} \widetilde{Z}_{v_i} \mathbf{1}(X_{v_i} > X_o)$$

and

$$\widetilde{Z}_\lambda = \sum_{i=1}^{N} \widetilde{Z}_{v_i} \mathbf{1}(X_{v_i} > X_o) + \mathbf{1}(N = 0). \tag{3}$$

https://doi.org/10.1017/jpr.2019.29 Published online by Cambridge University Press

Similarly, we can see that

$$C_\lambda = \sum_{i=1}^{N} C_{v_i} \mathbf{1}(X_{v_i} > X_o) + 1, \tag{4}$$

where C_{v_i} is the number of accessible vertices in the subtree rooted at v_i. Conditioning on the fitness value of the root $X_o = x$, it then follows that the relation between $f_\lambda(x, s, t)$ and $\tilde{f}_\lambda(x, s, t) := \mathrm{E}_x[s^{\widetilde{Z}_\lambda} t^{C_\lambda}]$, the probability generating function of $(\widetilde{Z}_\lambda, C_\lambda)$, is given by

$$f_\lambda(x, s, t) = \tilde{f}_\lambda(x, s, t) + \mathrm{P}(N = 0)(1 - s)t = \tilde{f}_\lambda(x, s, t) + s_\lambda - 1, \tag{5}$$

where $s_\lambda = \mathrm{e}^{-\lambda}(1 - s)t + 1$. If $N = n \geq 1$ is given, it is easy to see that $\{(\widetilde{Z}_{v_i}, C_{v_i}, X_{v_i}),\ i = 1, \ldots, n\}$ are i.i.d. random vectors, and have the same distribution as $(\widetilde{Z}_\lambda, C_\lambda, X_o)$. Hence, by (3) and (4),

$$\begin{aligned} \tilde{f}_\lambda(x, s, t) &= \mathrm{e}^{-\lambda} st + t \sum_{n=1}^{\infty} (\mathrm{E}[s^{\widetilde{Z}_\lambda \mathbf{1}(X_o > x)} t^{C_\lambda \mathbf{1}(X_o > x)}])^n \frac{\lambda^n}{n!} \mathrm{e}^{-\lambda} \\ &= \mathrm{e}^{-\lambda} t \sum_{n=0}^{\infty} \left(\int_x^1 \tilde{f}_\lambda(y, s, t)\, \mathrm{d}y + x \right)^n \frac{\lambda^n}{n!} - \mathrm{e}^{-\lambda}(1 - s)t \\ &= t \exp\left\{ \lambda \int_x^1 \tilde{f}_\lambda(y, s, t)\, \mathrm{d}y + \lambda(x - 1) \right\} - (s_\lambda - 1). \end{aligned} \tag{6}$$

We now fix s and t, and regard $\tilde{f}_\lambda(x, s, t)$ as a univariate function of x and (6) as a functional equation for it. Taking the common logarithm of both sides of (6) yields that

$$\ln(\tilde{f}_\lambda(x, s, t) + s_\lambda - 1) - \ln t = \lambda \int_x^1 \tilde{f}_\lambda(y, s, t)\, \mathrm{d}y + \lambda(x - 1).$$

Further, differentiating both sides of the above equation on x, we obtain

$$\frac{\tilde{f}'_\lambda(x, s, t)}{\tilde{f}_\lambda(x, s, t) + s_\lambda - 1} = \lambda[1 - \tilde{f}_\lambda(x, s, t)]. \tag{7}$$

Note that for any $0 < y < 1$,

$$\begin{aligned} &\int_y^1 \frac{\tilde{f}'_\lambda(x, s, t)}{(\tilde{f}_\lambda(x, s, t) + s_\lambda - 1)(1 - \tilde{f}_\lambda(x, s, t))}\, \mathrm{d}x \\ &\quad = \frac{1}{s_\lambda} \int_y^1 \left(\frac{\tilde{f}'_\lambda(x, s, t)}{\tilde{f}_\lambda(x, s, t) + s_\lambda - 1} + \frac{\tilde{f}'_\lambda(x, s, t)}{1 - \tilde{f}_\lambda(x, s, t)} \right) \mathrm{d}x \\ &\quad = \frac{1}{s_\lambda} \left(\ln \frac{\tilde{f}_\lambda(1, s, t) + s_\lambda - 1}{1 - \tilde{f}_\lambda(1, s, t)} - \ln \frac{\tilde{f}_\lambda(y, s, t) + s_\lambda - 1}{1 - \tilde{f}_\lambda(y, s, t)} \right), \end{aligned}$$

and (6) also indicates the boundary condition $\tilde{f}_\lambda(1, s, t) = 1 + t - s_\lambda$. Thus, by (7), we have

$$\ln t - \ln(s_\lambda - t) - \ln \frac{\tilde{f}_\lambda(y, s, t) + s_\lambda - 1}{1 - \tilde{f}_\lambda(y, s, t)} = \lambda(1 - y)s_\lambda,$$

https://doi.org/10.1017/jpr.2019.29 Published online by Cambridge University Press

from which it follows that, for any $x, s, t \in [0, 1]$,

$$\frac{\tilde{f}_\lambda(x, s, t) + s_\lambda - 1}{1 - \tilde{f}_\lambda(x, s, t)} = \frac{t}{s_\lambda - t} \mathrm{e}^{-\lambda(1-x)s_\lambda}.$$

After simple calculations, one can get

$$\tilde{f}_\lambda(x, s, t) = \frac{t\mathrm{e}^{-\lambda(1-x)s_\lambda} - (s_\lambda - 1)(s_\lambda - t)}{t\mathrm{e}^{-\lambda(1-x)s_\lambda} + s_\lambda - t},$$

from which (2) follows by (5). □

Applying Lemma 3, we can get the probability generating functions of random variables Z_λ and C_λ by setting $t = 1$ or $s = 1$ in the function $f_\lambda(x, s, t)$, and also their distributional properties.

Proposition 1. *Let Z_λ be the number of accessible leaves in a PGW(λ) tree. Then the probability generating function of Z_λ is given by*

$$\mathrm{E}[s^{Z_\lambda}] = \frac{1}{\lambda} \ln \frac{s_\lambda^*}{\mathrm{e}^{-\lambda s_\lambda^*} + s_\lambda^* - 1}, \qquad 0 \le s \le 1, \tag{8}$$

where $s_\lambda^* = \mathrm{e}^{-\lambda}(1 - s) + 1$.

Proof. Letting $t = 1$ in (2) gives that the conditional probability generating function of Z_λ is

$$\mathrm{E}_x[s^{Z_\lambda}] = \frac{s_\lambda^* \mathrm{e}^{-\lambda(1-x)s_\lambda^*}}{\mathrm{e}^{-\lambda(1-x)s_\lambda^*} + s_\lambda^* - 1}, \qquad 0 \le s \le 1.$$

Since the fitness value of the root is uniformly distributed on [0, 1], the integral of this conditional probability generating function over x from 0 to 1 gives the desired result, (8). □

Theoretically, one can get the exact distribution law of Z_λ from Proposition 1. However, we only give the probability of a special event in what follows, since the probability mass function of Z_λ cannot be expressed in a simple form.

Corollary 1. *In a PGW(λ) tree, we have*

$$\mathrm{P}(\textit{there exists an accessible leaf in a PGW}(\lambda)\textit{ tree}) = 1 - \frac{1}{\lambda} \ln \frac{\mathrm{e}^{\lambda} + 1}{\mathrm{e}^{-\lambda \mathrm{e}^{-\lambda}} + 1}.$$

Proof. It follows by Proposition 1 and the fact that the desired probability is $1 - \mathrm{E}[s^{Z_\lambda}]|_{s=0}$. □

The probability that there exists an accessible leaf in a PGW(λ) tree is, of course, a function of λ. Figure 1 illustrates that this probability trends to 0 when $\lambda \to 0$ or $\lambda \to \infty$. Using Mathematica®, we find that it achieves the maximum value 0.22558725 . . . when $\lambda = 1.53366758\ldots$

For the number of accessible vertices in a PGW(λ) tree, we can derive the exact distribution law of C_λ from its probability generating function.

https://doi.org/10.1017/jpr.2019.29 Published online by Cambridge University Press

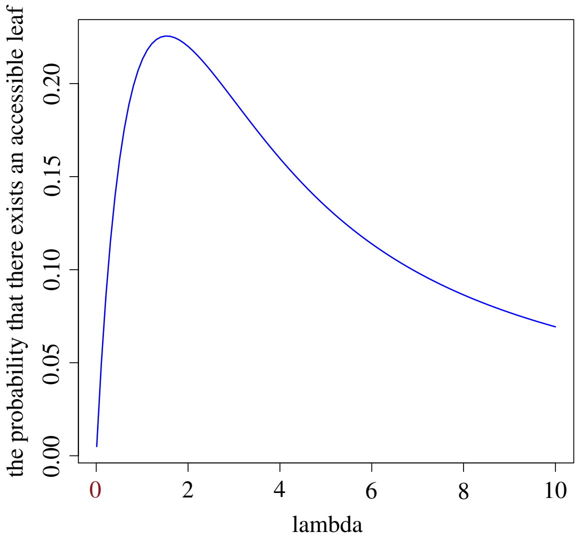

FIGURE 1: The probability that there exists an accessible leaf in a PGW(λ) tree.

Proposition 2. *Let C_λ be the number of accessible vertices in a PGW(λ) tree. Then the random variable C_λ follows the logarithmic series distribution with parameter $1-\mathrm{e}^{-\lambda}$. That is, the probability mass function of C_λ is given by*

$$\mathrm{P}(C_\lambda = k) = \frac{(1-\mathrm{e}^{-\lambda})^k}{\lambda k}, \qquad k = 1, 2, \ldots \tag{9}$$

Proof. Similar to the proof of Proposition 1, we can get the probability generating function of C_λ via its conditional probability generating function, and then we can further derive its exact distribution law in a simple form. In fact, letting $s=1$ in the probability generating function (2) gives that

$$\mathrm{E}_x[t^{C_\lambda}] = \frac{t\mathrm{e}^{-\lambda(1-x)}}{1-t+t\mathrm{e}^{-\lambda(1-x)}}, \qquad 0 \le t \le 1.$$

Therefore, the probability generating function of C_λ is given by

$$\begin{aligned} \mathrm{E}[t^{C_\lambda}] &= \int_0^1 \mathrm{E}_x[t^{C_\lambda}]\,\mathrm{d}x \\ &= \int_0^1 \frac{t\mathrm{e}^{-\lambda(1-x)}}{1-t+t\mathrm{e}^{-\lambda(1-x)}}\,\mathrm{d}x \\ &= -\frac{1}{\lambda}\ln(1-t+t\mathrm{e}^{-\lambda}), \qquad 0 \le t \le 1. \end{aligned} \tag{10}$$

It follows from the Maclaurin series expansion

$$\ln(1+z) = \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k} z^k$$

https://doi.org/10.1017/jpr.2019.29 Published online by Cambridge University Press

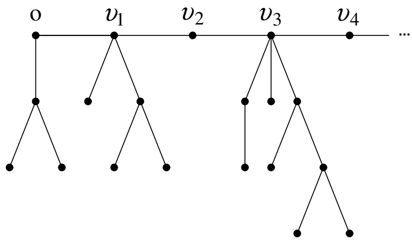

FIGURE 2: An illustration for the infinite random tree with distribution law $\mathrm{PGW}_\infty(1)$.

for any $|z| < 1$ that

$$-\frac{1}{\lambda}\ln(1 - t + t\mathrm{e}^{-\lambda}) = \sum_{k=1}^{\infty}\frac{(1-\mathrm{e}^{-\lambda})^k}{\lambda k}t^k,$$

since $|t(\mathrm{e}^{-\lambda} - 1)| < 1$ for $0 \le t \le 1$ and $\lambda > 0$. Then (9) follows from the definition of the probability generating function and (10). □

2.3. The limiting tree

We now describe an infinite random tree model as follows. Let $P = ov_1v_2\cdots$ be an infinite path with root o. To every vertex in P, an independent $\mathrm{PGW}(\lambda)$ tree is attached as a subtree rooted at it. And we still keep vertex o as the unique root of the whole tree (see Figure 2). Let $\mathrm{PGW}_\infty(\lambda)$ be the distribution law of the family of these infinite random trees; we call such a random tree a $\mathrm{PGW}_\infty(\lambda)$ tree.

The relation between $\mathrm{PGW}_\infty(\lambda)$ trees with the special case $\lambda = 1$ and random rooted labeled trees is given in the following lemma, which plays an essential role in the proof of our main result. Recall that $\mathcal{T}$ is the set of locally finite rooted trees, considered up to rooted isomorphism.

Lemma 4. *The random rooted labeled tree T_n on n vertices converges weakly in $\mathcal{T}$ to the $PGW_\infty(1)$ tree as $n \to \infty$.*

Proof. This result was first formalized and proved in Grimmett (1980). See also Lemma 2.3 in Aldous and Steele (2004). □

We now state the result analogous to Lemma 3 for the numbers of accessible leaves and accessible vertices in a $\mathrm{PGW}_\infty(\lambda)$ tree.

Lemma 5. *Let Z_∞ and C_∞ be the numbers of accessible leaves and accessible vertices in a $PGW_\infty(\lambda)$ tree, respectively. If the fitness value of root $X_o = x \in [0, 1]$ is given, then $g_\lambda(x, s, t)$, the conditional probability generating function of (Z_∞, C_∞), satisfies the equation*

$$g_\lambda(x, s, t) = \mathrm{E}_x[s^{Z_\infty}t^{C_\infty}] = f_\lambda(x, s, t)\bigg(\int_x^1 g_\lambda(y, s, t)\,\mathrm{d}y + x\bigg), \qquad 0 \le s, t \le 1,$$

where $f_\lambda(x, s, t)$ is defined in (2).

Proof. First, we introduce some notation in a $\mathrm{PGW}_\infty(\lambda)$ tree. Let X_v be the fitness value of vertex v. We also denote by Z_{v_1} and C_{v_1} the numbers of accessible leaves and accessible vertices in the (sub)tree rooted at v_1, respectively. Then it is easy to see that

$$Z_\infty = Z_o + Z_{v_1}\mathbf{1}(X_{v_1} > X_o), \qquad C_\infty = C_o + C_{v_1}\mathbf{1}(X_{v_1} > X_o), \tag{11}$$

https://doi.org/10.1017/jpr.2019.29 Published online by Cambridge University Press

where Z_o (C_o) denotes the number of accessible leaves (accessible vertices) only in the PGW(λ) tree rooted at o. It is obvious that the random vector (Z_o, C_o) defined in a $\mathrm{PGW}_\infty(\lambda)$ tree has the same distribution as (Z_λ, C_λ) defined in Lemma 3, and that the random vector $(Z_{v_1}, C_{v_1}, X_{v_1})$ also has the same distribution as $(Z_\infty, C_\infty, X_o)$.

We next consider the situation under the condition that $X_o = x \in [0, 1]$. Note that the random vectors (Z_o, C_o) and (Z_{v_1}, C_{v_1}) are now (conditionally) independent. Then it follows by (11) that

$$\begin{aligned} g_\lambda(x, s, t) &= \mathrm{E}_x[s^{Z_\lambda} t^{C_\lambda}] \mathrm{E}[s^{Z_\infty \mathbf{1}(X_o > x)} t^{C_\infty \mathbf{1}(X_o > x)}] \\ &= f_\lambda(x, s, t) \left(\int_x^1 g_\lambda(y, s, t)\, \mathrm{d}y + x \right), \end{aligned}$$

which completes the proof. □

In principle, one can further derive the corresponding results on the numbers of accessible leaves and accessible vertices for $\mathrm{PGW}_\infty(\lambda)$ trees from Lemma 5, as we have done in Propositions 1 and 2. But now it becomes much more complicated, since neither of the probability generating functions of Z_∞ and C_∞ has a closed form expression for general $\lambda > 0$. Fortunately, as Lemma 4 has indicated, for our main purpose it is sufficient to treat the special case $\lambda = 1$ only. We shall give the explicit results on $\mathrm{PGW}_\infty(1)$ trees in the next section.

3. Proof of main result

For the proof of our main result, we need the following auxiliary lemma.

Lemma 6. *Let $X_m, X_{n,m}, Y, Y_n$ be random vectors in the k-dimensional real space $\mathbb{R}^k$ and $n, m \geq 1$. If $X_{n,m} \xrightarrow{\mathrm{D}} X_m$ as $n \to \infty$ for any fixed m, $X_m \xrightarrow{\mathrm{D}} Y$ as $m \to \infty$, and, for any $\varepsilon > 0$,*

$$\lim_{m\to\infty} \limsup_{n\to\infty} \mathrm{P}(\|X_{n,m} - Y_n\| \geq \varepsilon) = 0,$$

where $\|\cdot\|$ denotes the Euclidean norm on $\mathbb{R}^k$, then $Y_n \xrightarrow{\mathrm{D}} Y$.

Proof. See, for example, Theorem 4.28 in Kallenberg (2002). □

To apply Lemma 6, we introduce another piece of notation as follows. For any $m \geq 1$, let $T_n^{(m)}$ be the subgraph of the random rooted labeled tree T_n induced by all vertices whose depths are at most m.

Proof of Theorem 1. For any rooted tree T, let $Z(T)$ and $C(T)$ denote the numbers of accessible leaves and accessible vertices in it, respectively. Recall that Z_n (C_n) denotes the number of accessible leaves (accessible vertices) in a random rooted labeled tree T_n.

For the difference between Z_n and $Z(T_n^{(m)})$, we have

$$|Z(T_n^{(m)}) - Z_n| \leq \sum_{i=m}^{n-1} \sum_{ov_1\cdots v_i \in \mathcal{A}} \mathbf{1}(X_o < X_{v_1} < \cdots < X_{v_i}),$$

where $\mathcal{A}$ denotes the set of all paths from the root in T_n. Thus,

$$\begin{aligned} \mathrm{E}|Z(T_n^{(m)}) - Z_n| &\leq \sum_{i=m}^{n-1} \mathrm{E}[W_i(T_n)]\, \mathrm{P}(X_o < X_{v_1} < \cdots < X_{v_i}) \\ &= \sum_{i=m}^{n-1} \frac{\mathrm{E}[W_i(T_n)]}{(i+1)!}, \end{aligned} \tag{12}$$

https://doi.org/10.1017/jpr.2019.29 Published online by Cambridge University Press

where $W_i(T_n)$ denotes the number of vertices of depth i in T_n. By Lemmas 1 and 2, we have

$$\mathrm{E}[W_i(T_n)] = \mathrm{E}[W_i(\mathrm{PGW}(1)) \mid |\mathrm{PGW}(1)| = n] \le ci, \tag{13}$$

for some absolute constant c and $1 \le i \le n-1$. By (12) and (13), it follows from Markov's inequality that, for any $\varepsilon > 0$,

$$\begin{aligned}\lim_{m\to\infty} \limsup_{n\to\infty} \mathrm{P}(|Z(T_n^{(m)}) - Z_n| \ge \varepsilon) &\le \lim_{m\to\infty} \limsup_{n\to\infty} \frac{\mathrm{E}|Z(T_n^{(m)}) - Z_n|}{\varepsilon} \\ &\le \lim_{m\to\infty} \sum_{i=m}^{\infty} \frac{\mathrm{E}[W_i(T_n)]}{(i+1)!\varepsilon} \\ &\le \lim_{m\to\infty} \sum_{i=m}^{\infty} \frac{ci}{(i+1)!\varepsilon} \\ &= 0.\end{aligned}$$

Similarly, it is also follows that

$$\lim_{m\to\infty} \limsup_{n\to\infty} \mathrm{P}(|C(T_n^{(m)}) - C_n| \ge \varepsilon) = 0$$

holds for any $\varepsilon > 0$, so that

$$\lim_{m\to\infty} \limsup_{n\to\infty} \mathrm{P}(\|(Z(T_n^{(m)}), C(T_n^{(m)})) - (Z_n, C_n)\| \ge \varepsilon) = 0. \tag{14}$$

By Lemma 4, we can obtain that for any fixed integer $m \ge 1$,

$$(Z(T_n^{(m)}), C(T_n^{(m)})) \xrightarrow{\mathrm{D}} (Z(\mathrm{PGW}_\infty^{(m)}(1)), C(\mathrm{PGW}_\infty^{(m)}(1)))$$

as $n \to \infty$, where $\mathrm{PGW}_\infty^{(m)}(1)$ also denotes the subgraph of the $\mathrm{PGW}_\infty(1)$ tree induced by all vertices whose depths are at most m. An analogous technique to (12)–(14) gives that, as $m \to \infty$,

$$\|(Z(\mathrm{PGW}_\infty^{(m)}(1)), C(\mathrm{PGW}_\infty^{(m)}(1))) - (Z(\mathrm{PGW}_\infty(1)), C(\mathrm{PGW}_\infty(1)))\| \xrightarrow{\mathrm{P}} 0,$$

which implies that

$$(Z(\mathrm{PGW}_\infty^{(m)}(1)), C(\mathrm{PGW}_\infty^{(m)}(1))) \xrightarrow{\mathrm{D}} (Z(\mathrm{PGW}_\infty(1)), C(\mathrm{PGW}_\infty(1)))$$

as $m \to \infty$. Collecting the above results and applying Lemma 6, we have

$$(Z_n, C_n) \xrightarrow{\mathrm{D}} (Z(\mathrm{PGW}_\infty(1)), C(\mathrm{PGW}_\infty(1)))$$

as $n \to \infty$. From Levy's continuity theorem, the probability generating function of (Z_n, C_n) thus converges to that of $(Z(\mathrm{PGW}_\infty(1)), C(\mathrm{PGW}_\infty(1)))$. That is, it follows that as $n \to \infty$, for any $0 \le s, t \le 1$,

$$\lim_{n\to\infty} \mathrm{E}[s^{Z_n} t^{C_n}] = \mathrm{E}[s^{Z(\mathrm{PGW}_\infty(1))} t^{C(\mathrm{PGW}_\infty(1))}]. \tag{15}$$

https://doi.org/10.1017/jpr.2019.29 Published online by Cambridge University Press

Letting $\lambda = 1$ and $x = 0$ in Lemma 5 yields that

$$\begin{aligned} \mathrm{E}[s^{Z(\mathrm{PGW}_\infty(1))} t^{C(\mathrm{PGW}_\infty(1))}] &= \int_0^1 \mathrm{E}_x[s^{Z(\mathrm{PGW}_\infty(1))} t^{C(\mathrm{PGW}_\infty(1))}] \, \mathrm{d}x \\ &= \int_0^1 g(x, s, t) \, \mathrm{d}x \\ &= \frac{g(0, s, t)}{f(0, s, t)}, \end{aligned} \tag{16}$$

where $g(x, s, t) := g_1(x, s, t)$ and

$$f(x, s, t) := f_1(x, s, t) = \frac{s_1 t \, \mathrm{e}^{s_1(x-1)}}{t \, \mathrm{e}^{s_1(x-1)} + s_1 - t}, \tag{17}$$

with $s_1 = \mathrm{e}^{-1}(1 - s)t + 1$.

We next find the exact expression for the probability generating function of the random vector $(Z(\mathrm{PGW}_\infty(1), C(\mathrm{PGW}_\infty(1)))$. By Lemma 5, the relation between $g(x, s, t)$ and $f(x, s, t)$ reduces to

$$g(x, s, t) = f(x, s, t)\left(\int_x^1 g(y, s, t) \, \mathrm{d}y + x \right), \qquad 0 \le x, s, t \le 1. \tag{18}$$

We can also fix s and t, and regard $g(x, s, t)$ and $f(x, s, t)$ as univariate functions of x. Taking the derivative of both sides of (18) with respect to x yields

$$g'(x, s, t) + p(x, s, t) g(x, s, t) = f(x, s, t), \tag{19}$$

which is a standard first-order differential equation for $g(x, s, t)$ with

$$p(x, s, t) := f(x, s, t) - \frac{f'(x, s, t)}{f(x, s, t)} = 2f(x, s, t) - s_1,$$

by (17). With the boundary condition $g(1, s, t) = f(1, s, t) = t$, it is well known that the solution of (19) can be given as

$$g(x, s, t) = \exp\left\{ -\int_1^x p(y, s, t) \, \mathrm{d}y \right\} \left(\int_1^x f(y, s, t) \exp\left\{ \int_1^y p(z, s, t) \mathrm{d}z \right\} \mathrm{d}y + t \right).$$

After straightforward calculations, we get that

$$g(x, s, t) = \frac{t\mathrm{e}^{s_1(x-1)}(t\mathrm{e}^{s_1(x-1)} + s_1^2 x - t - s_1 t(x-1))}{(t\mathrm{e}^{s_1(x-1)} + s_1 - t)^2}. \tag{20}$$

Inserting $x = 0$ into (17) and (20), it thus follows by (16) that, for $0 \le s, t \le 1$,

$$\mathrm{E}[s^{Z(\mathrm{PGW}_\infty(1))} t^{C(\mathrm{PGW}_\infty(1))}] = \frac{1}{s_1} + \frac{t - 1}{t\mathrm{e}^{-s_1} + s_1 - t},$$

which, together with (15), implies that (1) holds.

One can immediately obtain the asymptotic distributions of Z_n and C_n through (1). In fact, letting $t = 1$ in (1) yields that the limiting probability generating function of Z_n is given by

$$\lim_{n\to\infty} \mathrm{E}[s^{Z_n}] = \frac{\mathrm{e}}{1+\mathrm{e}-s} = \sum_{k=0}^{\infty} \frac{\mathrm{e}s^k}{(1+\mathrm{e})^{k+1}},$$

which implies that

$$\lim_{n\to\infty} \mathrm{P}(Z_n = k) = \frac{\mathrm{e}}{(1+\mathrm{e})^{k+1}}, \qquad k = 0, 1, 2, \ldots$$

For C_n, it also follows directly by (1) that

$$\lim_{n\to\infty} \mathrm{E}[t^{C_n}] = \frac{t/\mathrm{e}}{1-t(1-1/\mathrm{e})},$$

which is the probability generating function of the geometric distribution with parameter $1/\mathrm{e}$. That is,

$$\lim_{n\to\infty} \mathrm{P}(C_n = k) = \frac{(\mathrm{e}-1)^{k-1}}{\mathrm{e}^k}, \qquad k = 1, 2, \ldots$$

The proof of Theorem 1 is complete. □

Acknowledgements

We thank the anonymous referee for useful suggestions that greatly improved the presentation of this work. This work is supported by NSFC grant numbers 11671373 and 11771418.

References

[1] Aldous, D. J. and Steele, J. M. (2004). The objective method: Probabilistic combinatorial optimization and local weak convergence. In *Probability on Discrete Structures*, ed. H. Kesten, Vol. 110 of *Encyclopaedia of Mathematical Sciences*. Springer, Berlin, pp. 1–72.
[2] Benjamini, I. and Schramm, O. (2001). Recurrence of distributional limits of finite planar graphs. *Electron. J. Prob.* **6,** 23.
[3] Berestycki, J., Brunet, É. and Shi, Z. (2016). The number of accessible paths in the hypercube. *Bernoulli* **22,** 653–680.
[4] Chen, X. (2014). Increasing paths on N-ary trees. Preprint, arXiv:1403.0843.
[5] Coletti, C. F., Gava, R. J. and Rodríguez, P. M. (2018). On the existence of accessibility in a tree-indexed percolation model. *Physica A* **492,** 382–388.
[6] Drmota, M. (2009). *Random Trees: An Interplay between Combinatorics and Probability*. Springer, Wien.
[7] Franke, J., Klözer, A., de Visser, J. A. G. M. and Krug, J. (2011). Evolutionary accessibility of mutational pathways. *PLoS Comput. Biol.* **7,** e1002134.
[8] Grimmett, G. R. (1980). Random labelled trees and their branching networks. *J. Austral. Math. Soc. (Ser. A)* **30**, 229–237.
[9] Hegarty, P. and Martinsson, A. (2014). On the existence of accessible paths in various models of fitness landscapes. *Ann. Appl. Prob.* **24,** 1375–1395.
[10] Janson, S. (2006). Random cutting and records in deterministic and random trees. *Random Structures Algorithms* **29,** 139–179.
[11] Kallenberg, O. (2002). *Foundations of Modern Probability*, 2nd ed. Springer, Berlin.
[12] Kolchin, V. F. (1977). Branching processes, random trees, and a generalized scheme of arrangements of particles. *Math. Notes Acad. Sci. USSR* **21,** 386–394.
[13] Li, L. (2017). Phase transition for accessibility percolation on hypercubes. *J. Theoret. Prob.* **31,** 2072–2111.
[14] Nowak, S. and Krug, J. (2013). Accessibility percolation on n-trees. *Europhys. Lett.* **101,** 66004.

[15] Roberts, M. I. and Zhao, L. Z. (2013). Increasing paths in regular trees. *Electron. Commun. Prob.* **18,** 87.
[16] Weinreich, D. M., Delaney, N. F., DePristo, M. A. and Hartl, D. M. (2006). Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science* **312,** 111–114.
[17] Weinreich, D. M., Watson, R. A. and Chao, L. (2005). Perspective: Sign epistasis and genetic constraints on evolutionary trajectories. *Evolution* **59,** 1165–1174.

https://doi.org/10.1017/jpr.2019.29 Published online by Cambridge University Press