ALMOST GIANT CLUSTERS FOR PERCOLATION ON LARGE TREES WITH LOGARITHMIC HEIGHTS

JEAN BERTOIN,* *Universität Zürich*

Abstract

This paper is based on works presented at the 2012 Applied Probability Trust Lecture in Sheffield; its main purpose is to survey the recent asymptotic results of Bertoin (2012a) and Bertoin and Uribe Bravo (2012b) about Bernoulli bond percolation on certain large random trees with logarithmic height. We also provide a general criterion for the existence of giant percolation clusters in large trees, which answers a question raised by David Croydon.

Keywords: Random tree; percolation; giant component

2010 Mathematics Subject Classification: Primary 60K35; 05C05

1. Introduction

It is well known that percolation is considerably simpler to study on a tree than on a general graph, thanks to the property of uniqueness of the path connecting two vertices. We refer in particular to Chapter 5 of [9] and the references therein for a number of important and useful results for infinite trees, such as criteria for the existence or absence of infinite percolation clusters. Here, we will be interested in a somewhat different type of question. Specifically, we consider a tree of large but finite size, and perform a Bernoulli bond percolation with a parameter that depends on the size of that tree; our purpose is to investigate the asymptotic behavior of the sizes of the largest clusters for appropriate regimes when the size of the tree goes to ∞.

Our motivation comes from a celebrated result of Erdős and Rényi on the random graph model, which can be phrased informally as follows. With high probability, when $n \gg 1$, Bernoulli bond percolation on the complete graph with n vertices and with parameter $p(n) \sim c/n$ for some fixed $c > 1$ produces a single giant cluster of size close to $\theta(c)n$, where $\theta(c) \in (0, 1)$ is some known constant, while the second, third, etc. largest clusters are almost microscopic, and more precisely have size of only order $\ln n$.

In the first part of this paper, we provide a simple characterization of tree families and percolation regimes which yield giant clusters, answering a question raised by David Croydon. In the second part, we briefly review the main results of [3] and [4] concerning two natural families of random trees with logarithmic heights, namely recursive trees and scale-free trees. In those works, we showed that the next largest clusters are almost giant, in the sense that their sizes are of order $n/\ln n$, and obtained precise limit theorems in terms of certain Poisson random measures. A common feature in the analysis of percolation for these models is that, even though one addresses a static problem, it is useful to consider dynamical versions in which edges are removed, respectively vertices are inserted, one after the other in a certain order as time passes.

2. Giant clusters

We first introduce notation and hypotheses which will have an important role in this section. For a given integer n, we consider a set of $n + 1$ vertices, say $\mathcal{V}_n = \{0, 1, \ldots, n\}$, and a tree structure T_n on $\mathcal{V}_n$. So T_n has n edges, and we should think of 0 as the root of T_n. We perform a Bernoulli bond percolation on T_n with parameter $p(n)$, so that each edge of T_n is kept with probability $p(n)$ and removed with probability $1 - p(n)$, independently of the other edges. The resulting connected components are then referred to as clusters.

We write $C^0_{p(n)}$ for the size of the cluster that contains the root; plainly, $C^0_{p(n)} \leq n + 1$. We say that $C^0_{p(n)}$ is *giant* if $n^{-1} C^0_{p(n)}$ converges in law to some random variable $G \not\equiv 0$, which should be thought of as the asymptotic proportion of vertices pertaining to the root cluster. David Croydon raised the question of finding a simple criterion for $C^0_{p(n)}$ to be giant, depending of course on the nature of T_n and regimes of the percolation parameter $p(n)$; this motivates the following.

For each fixed $n \in \mathbb{N}$, we denote by $V_1, V_2, \ldots$ a sequence of independent and identically distributed (i.i.d.) vertices in $\mathcal{V}_n$ with the uniform distribution. Next, for every $k \in \mathbb{N}$, we write $L_{k,n}$ for the length of the tree T_n reduced to $V_1, \ldots, V_k$ and the root 0, i.e. the minimal number of edges of T_n which are needed to connect 0 and $V_1, \ldots, V_k$. In particular, $L_{1,n}$ should be thought of as the height of a typical vertex in T_n. Let $\ell: \mathbb{N} \to \mathbb{R}_+$ be some function with $\lim_{n \to \infty} \ell(n) = \infty$. We introduce the hypothesis

$$(H_k) \quad \frac{1}{\ell(n)} L_{k,n} \Rightarrow L_k,$$

where '$\Rightarrow$' means weak convergence and L_k is some random variable with values in $\mathbb{R}_+$. We stress that (H_k) can be assumed to hold for different values of k and then only convergence in the sense of one-dimensional distributions is involved.

In several examples, the function ℓ is a logarithm and $L_k \equiv ak$ with a a positive constant. For instance, this happens for some important families of random trees, such as recursive trees, binary search trees, etc.; see [5]. Aldous [1] considered a different class of examples, including the case when T_n is a Cayley tree of size $n + 1$ (i.e. a tree picked uniformly at random amongst the $(n + 1)^{n-1}$ trees on $\mathcal{V}_n$), for which it is known that (H_k) holds with $\ell(n) = \sqrt{n}$ and L_k a chi-variable with $2k$ degrees of freedom.

We now state the central result of this section.

Theorem 1. *For an arbitrary $c \geq 0$, consider the regime*

$$p(n) = 1 - \frac{c}{\ell(n)} + o\left(\frac{1}{\ell(n)}\right). \tag{1}$$

(i) *If (H_k) holds for every $k \in \mathbb{N}$ then we have, in the regime (1),*

$$n^{-1} C^0_{p(n)} \Rightarrow G(c), \tag{2}$$

where $G(c) \not\equiv 0$ is a random variable whose law is determined by its entire moments:

$$\mathbb{E}(G(c)^k) = \mathbb{E}(e^{-cL_k}), \qquad k \in \mathbb{N}. \tag{3}$$

In particular, $\lim_{c \to 0+} G(c) = 1$ in probability.

(ii) *Conversely, suppose that, for every $c > 0$, (2) holds in the regime (1) for some random variable $G(c)$ with values in $[0, 1]$. Suppose further that $\lim_{c \to 0+} G(c) = 1$ in probability. Then (H_k) is fulfilled for every $k \geq 1$, with L_k a nonnegative random variable whose Laplace transform is given by (3).*

Proof. The proof relies on the observation that, for each $k \geq 1$, there is the identity

$$\mathbb{E}(((n+1)^{-1}C^0_{p(n)})^k) = \mathbb{E}(p(n)^{L_{k,n}}). \tag{4}$$

Indeed, recall that $V_1, \ldots, V_k$ are k i.i.d. uniformly distributed vertices, which are independent of the percolation process. This enables us to interpret the left-hand side of (4) as the probability that $V_1, \ldots, V_k$ belong to the percolation cluster containing the root. On the other hand, considering the tree reduced to $V_1, \ldots, V_k$ and the root shows that this same probability can also be expressed in terms of the length $L_{k,n}$ of this reduced tree, as the right-hand side of (4).

The assumption (H_k) entails that, in the regime (1),

$$\lim_{n \to \infty} \mathbb{E}(p(n)^{L_{k,n}}) = \lim_{n \to \infty} \mathbb{E}\left(\exp\left(-\frac{c}{\ell(n)} L_{k,n}\right)\right) = \mathbb{E}(e^{-cL_k}),$$

and then we deduce from (4) that

$$\lim_{n \to \infty} \mathbb{E}(((n+1)^{-1}C^0_{p(n)})^k) = \mathbb{E}(e^{-cL_k}).$$

Thus, if (H_k) holds for every $k \in \mathbb{N}$ then $(n+1)^{-1}C^0_{p(n)}$ converges in law to some variable $G(c)$ with values in $[0, 1]$. More precisely, the law of $G(c)$ is determined by its entire moments $\mathbb{E}(G(c)^k) = \mathbb{E}(e^{-cL_k}) > 0$; this proves (2).

Conversely, as the size of a cluster cannot exceed $n + 1$, (2) implies that in the regime (1), we have, for every integer $k \geq 1$,

$$\lim_{n \to \infty} \mathbb{E}(((n+1)^{-1}C^0_{p(n)})^k) = \mathbb{E}(G(c)^k).$$

From (4) we rewrite this as

$$\lim_{n \to \infty} \mathbb{E}(p(n)^{L_{k,n}}) = \mathbb{E}(G(c)^k).$$

Plugging in the expression for the parameter $p(n)$ given in (1), we easily derive

$$\lim_{n \to \infty} \mathbb{E}\left(\exp\left(-\frac{c}{\ell(n)} L_{k,n}\right)\right) = \mathbb{E}(G(c)^k).$$

Recall the assumption that $\lim_{c \to 0+} G(c) = 1$ in probability, in particular,

$$\lim_{c \to 0+} \mathbb{E}(G(c)^k) = 1.$$

We conclude from Theorem XIII.1.2 of [7, p. 431] that, for each $k \geq 1$, the function $c \mapsto \mathbb{E}(G(c)^k)$ is the Laplace transform of a random variable $L_k \geq 0$, and that (H_k) holds. This completes the proof. $\square$

We next point at an interesting consequence of Theorem 1 to the characterization of the cases for which the proportion of vertices in the root cluster converges in probability to a constant. We consider the situation where the variables L_k appearing in hypothesis (H_k) are of the form

(H'_k) $L_k = \xi_1 + \cdots + \xi_k$,

where $\xi_1, \ldots$ is a sequence of i.i.d. variables in $\mathbb{R}_+$.

Corollary 1. (i) *Suppose that (H_k) and (H'_k) hold for $k = 1, 2$. Then in the regime (1) we have*

$$\lim_{n \to \infty} n^{-1} C^0_{p(n)} = \theta(c) \quad \text{in probability}, \tag{5}$$

where $\theta(c) = \mathbb{E}(e^{-c\xi_1}) > 0$. Furthermore, (H_k) and (H'_k) hold for every $k \geq 1$.

(ii) *Conversely, suppose that, for every $c > 0$, (5) holds in the regime (1) for some function $\theta \colon [0, \infty) \to [0, 1]$ such that $\lim_{c \to 0+} \theta(c) = 1$. Then θ is the Laplace transform of a nonnegative random variable ξ, and (H_k) and (H'_k) are fulfilled for every $k \geq 1$ with $\xi_1, \ldots$ a sequence of i.i.d. copies of ξ.*

Proof. When (H'_k) holds, we have $\mathbb{E}(\exp(-cL_k)) = \theta(c)^k$, with $\theta(c) = \mathbb{E}(e^{-c\xi_1})$. We now see from the proof of Theorem 1 that hypotheses (H_k) and (H'_k) entail that in the regime (1) we have

$$\lim_{n \to \infty} \mathbb{E}((n^{-1} C^0_{p(n)})^k) = \theta(c)^k.$$

In particular, if (H_k) and (H'_k) hold for $k = 1, 2$ then

$$\lim_{n \to \infty} \mathbb{E}((n^{-1} C^0_{p(n)} - \theta(c))^2) = 0,$$

which proves (5).

Conversely, if (5) holds then we can apply Theorem 1(ii) with $G(c) \equiv \theta(c)$. In particular, we know that (H_k) holds for all $k \in \mathbb{N}$. Furthermore, we get $\mathbb{E}(e^{-cL_k}) = \theta(c)^k$, which in turn shows that (H'_k) is fulfilled. This completes the proof.

We now conclude this section by pointing at a simple criterion which ensures that the cluster containing the root is the unique giant component.

Proposition 1. *In the preceding notation, assume that there is the joint weak convergence*

$$\frac{1}{\ell(n)}(L_{1,n}, L_{2,n}) \Rightarrow (L_1, L_2),$$

where (L_1, L_2) is a pair of random variables such that $L_2 - L_1$ has the same law as L_1. Then, for every $c > 0$, in the regime (1) we have

$$\lim_{n \to \infty} n^{-1} C^1_{p(n)} = 0 \quad \text{in probability},$$

where $C^1_{p(n)}$ denotes the size of the largest percolation cluster which does not contain the root 0.

Proof. Recall that V_1 and V_2 denote two independent, uniformly distributed random vertices. Plainly, the probability $\varrho(n)$ that V_1 and V_2 both belong to the same percolation cluster and are disconnected from 0 can be bounded from below by $(n + 1)^{-2} \mathbb{E}(|C^1_{p(n)}|^2)$.

On the other hand, $\varrho(n)$ is bounded from above by the probability that at least one edge of the branch from the root 0 to the branch point $V_1 \wedge V_2$ of V_1 and V_2 has been removed, viz.

$$(n + 1)^{-2} \mathbb{E}(|C^1_{p(n)}|^2) \leq \varrho(n) \leq 1 - \mathbb{E}(p(n)^{d_n(0, V_1 \wedge V_2)}), \tag{6}$$

where d_n denotes the graph distance in T_n.

Next, write

$$L_{2,n} = d_n(0, V_1) + d_n(0, V_2) - d_n(0, V_1 \wedge V_2).$$

Since $L_{1,n} = d_n(0, V_1)$ has the same law as $d_n(0, V_2)$, it follows from our assumption that the sequences $\ell(n)^{-1} d_n(0, V_2)$ and $\ell(n)^{-1}(d_n(0, V_2) - d_n(0, V_1 \wedge V_2))$ converge weakly to the same distribution. This readily implies that

$$d_n(0, V_1 \wedge V_2) = o(\ell(n)) \quad \text{in probability,}$$

and we conclude that the right-hand side of (6) tends to 0 as $n \to \infty$.

3. Almost giant clusters

In this section we turn our attention to the percolation clusters which do not contain the root. We write

$$C_{p(n)}^1 \geq C_{p(n)}^2 \geq \cdots$$

for the sequence of their sizes, ranked in decreasing order. Beware that this convenient notation may be slightly misleading, since $C_{p(n)}^0$ is always the size of the cluster containing the root 0, while, for $i \geq 1$, $C_{p(n)}^i$ is in general not the size of the cluster containing the vertex i. A natural problem is then to determine the asymptotic behavior of this sequence. We first point out that hypothesis (H_k) is insufficient to characterize the latter, by considering three simple examples in which very different behaviors can be observed.

First, imagine that T_n is a star-shaped tree centered at 0, meaning that the root is the unique branching point. Suppose also for simplicity that there are approximately $n^{1-\alpha}$ branches attached to the root, each of size approximately n^α, where $\alpha \in (0, 1)$ is some fixed parameter. Then one readily checks that (H_k) and (H'_k) hold for every $k \geq 1$ with $\ell(n) = n^\alpha$ and $L_k = \xi_1 + \cdots + \xi_k$, where the ξ_i are i.i.d. uniformly distributed on $[0, 1]$. It is further straightforward to see in the regime (1) that we have

$$C_{p(n)}^1 \sim C_{p(n)}^2 \sim \cdots \sim C_{p(n)}^j \sim n^\alpha$$

for every fixed $j \in \mathbb{N}$.

Second, consider the case when T_n is the complete regular d-ary tree with height h, where $d \geq 2$ is some integer. So there are d^j vertices at distance $j = 0, 1, \ldots, h$ from the root and

$$n = n(h) = \frac{d(d^h - 1)}{d - 1}.$$

One readily checks that hypotheses (H_k) and (H'_k) hold for every $k \geq 1$ with $\ell(n) = \ln n$ and $\xi_i \equiv 1/\ln d$. Because the subtree spanned by a vertex at height $j \leq h$ is again a complete regular d-ary tree with height $h - j$, we deduce from the preceding section that in the regime (1), the size $C_{p(n)}^1$ of the largest cluster which does not contain the root is close to

$$\frac{\mathrm{e}^{-c/\ln d} d^{h-\kappa(h)+1}}{d - 1},$$

where $\kappa(h)$ is the smallest height at which an edge has been removed. Recall that there are $d(d^j - 1)/(d - 1)$ edges with height at most j, so the law of $\kappa(h)$ is given by

$$\mathbb{P}(\kappa(h) > j) = p(n)^{d(d^j-1)/(d-1)}, \qquad j = 1, \ldots, h.$$

It follows readily that in the regime (1), the sequence $(\kappa(h) - \ln(h)/\ln(d) : h \in \mathbb{N})$ is tight. We stress however that this sequence *does not converge in distribution* as $h \to \infty$; more precisely, weakly convergent subsequences are obtained provided that the fractional part

$\{\ln(h)/\ln(d)\}$ converges. It follows that the sequence $(\ln(n)n^{-1}C^1_{p(n)}: n = n(h), h \in \mathbb{N})$ is also tight. It does not converge as $h \to \infty$; however, weakly convergent subsequences can be excerpt provided that $\{\ln(h)/\ln(d)\}$ converges.

Third, we recall that in the case of Cayley trees, Pitman [11], [12] showed that, for $1 - p(n) \sim c/\sqrt{n}$ with a fixed $c > 0$, the sequence of the sizes of the clusters ranked in decreasing order and renormalized by a factor $1/n$ converges weakly as $n \to \infty$ to a random mass partition which can be described explicitly in terms of a conditioned Poisson measure. It is interesting to observe that in this situation, the number of giant components is unbounded as $n \to \infty$. We stress that the conditions of Proposition 1 and hypothesis (H'_k) for $k \geq 2$ fail for Cayley trees.

We will now study the asymptotic behavior of the sizes of the largest clusters which do not contain the root for two families of random trees with logarithmic heights, i.e. which fulfill (H_k) with $\ell(n) = \ln n$. In particular, we will point out that in the regime (1), the largest percolation clusters which do not contain the root fail to be giant only by a logarithmic factor.

3.1. Random recursive trees

A tree on an ordered set of vertices is called *recursive* if, when we agree that the smallest vertex serves as the root, then the sequence of vertices along any branch from the root to a leaf is increasing. Recursive trees are sometimes also known as increasing trees in the literature; they arise for instance in computer science as data structures, or as simple epidemic models.

Of course, there is no loss of generality in assuming that the set of vertices is $\mathcal{V}_n = \{0, 1, \dots, n\}$ (and then 0 is the root); however, other ordered sets may arise naturally in this setting as we will see. Each recursive tree on $\mathcal{V}_n$ encodes a permutation of $\{1, \dots, n\}$ in such a way that the subtrees attached to the root 0 correspond to the cycles of the permutation, and this encoding is bijective; see Section 6.1.1 of [5]. In particular, there are $n!$ recursive trees on $\mathcal{V}_n$; we pick one of them uniformly at random and denote it by T_n. In other words, T_n can be viewed as a Cayley tree on $\mathcal{V}_n$, subject to the condition that the sequence of vertices along any branch from the root to a leaf is increasing. We stress that, informally, the conditioning becomes singular as $n \to \infty$. Indeed, the geometry of large Cayley trees and large uniform recursive trees are notoriously different; for instance, the typical height of the former is of order $\sqrt{n}$ while that of latter is only of order $\ln n$.

There is an elementary algorithm for constructing T_n which is closely related to the so-called Chinese restaurant process (see, e.g. Section 3.1 of [12]), and, hence, further points at the connection with uniform random permutations. For every $i = 1, \dots, n$, we pick a vertex U_i uniformly at random from $\{0, \dots, i - 1\}$ and independently of the U_j for $j \neq i$. The random tree induced by the set of edges $\{(i, U_i): i = 1, \dots, n\}$ is then a version of T_n.

Uniform recursive trees fulfill an important *splitting property* which is the key to many of their features. Fix an arbitrary $k \in \{1, \dots, n\}$ and remove the edge between k and its parent U_k. This disconnects T_n into two subtrees, say T and T'. If we denote by $\mathcal{V}$ (respectively $\mathcal{V}'$) the set of vertices of T (respectively of T'), then, conditionally on $\mathcal{V}$ and $\mathcal{V}'$, T and T' are two independent uniform recursive trees with respective sets of vertices $\mathcal{V}$ and $\mathcal{V}'$. This basic property is easy to check, either directly from the definition, or from the Chinese restaurant construction of T_n.

It is easy to verify that the conditions (H_k) are fulfilled for all $k \geq 1$ with $\ell(n) = \ln n$ and $L_k \equiv k$; see Section 6.2.5 of [5]. We conclude from the preceding section that in the regime (1), the cluster containing the root 0 is the unique giant percolation cluster of T_n, and, more precisely, that (5) holds with $\theta(c) = e^{-c}$. The main result of [3] is that the next largest clusters are almost giant, and, more precisely, one has the following weak limit theorem.

Theorem 2. *Let T_n denote a uniform random recursive tree on $\{0, 1, \ldots, n\}$. For every fixed integer j, in the regime (1) with $\ell(n) = \ln n$,*

$$\left(\frac{\ln n}{n} C_{p(n)}^1, \ldots, \frac{\ln n}{n} C_{p(n)}^j \right)$$

converges in distribution towards

$$(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_j),$$

where $\boldsymbol{x}_1 > \boldsymbol{x}_2 > \cdots$ denotes the sequence of the atoms of a Poisson random measure on $(0, \infty)$ with intensity $c e^{-c} x^{-2} \, dx$.

There is an equivalent simple description of the law of the limiting sequence, namely, $1/\boldsymbol{x}_1$, $1/\boldsymbol{x}_2 - 1/\boldsymbol{x}_1, \ldots, 1/\boldsymbol{x}_j - 1/\boldsymbol{x}_{j-1}$ are i.i.d. exponential variables with parameter $c e^{-c}$. In particular, $\boldsymbol{x}_j$ has the same distribution as the inverse of a gamma variable with parameter $(j, c e^{-c})$, and $\lim_{j \to \infty} j \boldsymbol{x}_j = c e^{-c}$ in probability.

The basic idea in [3] for establishing Theorem 2 is to relate percolation on a rooted tree T to a random algorithm for the isolation of its root that was introduced by Meir and Moon. Specifically, following these authors, we can imagine that we remove an edge in T uniformly at random, disconnecting T into two subtrees. We set aside the subtree which does not contain the root and iterate in an obvious way with the subtree containing the root, until the root is finally isolated. Loosely speaking, we can think of this algorithm as a dynamical version of percolation (i.e. edges are now removed one after the other rather than simultaneously), except that each time an edge is removed, the cluster which does not contain the root is instantaneously frozen, in the sense that only edges belonging to the cluster that contains the root can be removed.

The upshot of this point of view is that it enables us to use a coupling due to Iksanov and Möhle [8], which, informally, identifies the sequence of the sizes of the frozen subtrees which arise from the isolation of the root algorithm, with the sequence $\eta_1, \eta_2, \ldots, \eta_k$ of i.i.d. variables with distribution

$$\mathbb{P}(\eta = j) = \frac{1}{j(j+1)}, \qquad j \in \mathbb{N},$$

at least as long as $\eta_1 + \cdots + \eta_k \leq n$. In short, this coupling follows from the splitting property of random recursive trees, and the following remarkable fact observed by Meir and Moon [10]. Imagine that we remove an edge of T_n uniformly at random, and consider the size of the resulting subtree that does not contain the root. Then the latter has the same distribution as η conditioned on $\eta \leq n$.

The coupling of Iksanov and Möhle enables us to use extreme values theory and determine the asymptotic behavior of the sizes of these frozen subtrees, jointly with the steps of the algorithm at which they have appeared. In short, one finds that the largest frozen subtrees have size of order $n / \ln n$ and a precise limit theorem can be given in terms of the atoms of some Poisson random measure. It then remains to de-freeze each of these subtrees by performing an additional Bernoulli percolation with a suitable parameter, to recover the outcome of percolation on T_n. Roughly speaking, each of these frozen subtrees can be viewed conditionally on its size as a uniform recursive tree. As a consequence, the additional percolation produces a single relatively giant component of size again of order $n / \ln n$ and further clusters of smaller size $O(n / \ln^2 n)$. In particular, the largest percolation clusters of T_n which do not contain the root correspond to simple transformations of the frozen subtrees arising from the algorithm of isolation of the root, and their limiting distribution is obtained as the image of some Poisson random measure.

3.2. Scale-free random trees

Scale-free random trees form a one-parameter family of random trees that grow following a preferential attachment algorithm; see [2]. Fix a parameter $\beta \in (-1, \infty)$, and start for $n = 1$ from the unique tree $T_1^{(\beta)}$ on $\mathcal{V}_1 = \{0, 1\}$ which has a single edge connecting 0 and 1. Then suppose that $T_n^{(\beta)}$ has been constructed for some $n \geq 1$, and, for every $i \in \mathcal{V}_n = \{0, \dots, n\}$, denote by $d_n(i)$ the degree of vertex i in $T_n^{(\beta)}$. Conditionally given $T_n^{(\beta)}$, we construct the tree $T_{n+1}^{(\beta)}$ by incorporating the new vertex $n + 1$ to $T_n^{(\beta)}$ and adding an edge between $n + 1$ and a vertex $v_n \in \{0, \dots, n\}$ chosen at random according to the law

$$\mathbb{P}(v_n = i) = \frac{d_n(i) + \beta}{2n + \beta(n + 1)}, \qquad i \in \{0, \dots, n\}.$$

Recall that there is the identity $\sum_{i=0}^{n} d_n(i) = 2n$ (because $T_n^{(\beta)}$ is a tree with n edges), so the preceding indeed defines a probability on $\{0, \dots, n\}$. Note also that when one lets $\beta \to \infty$, then v_n becomes uniformly distributed on $\{0, \dots, n\}$, and the algorithm yields a uniform recursive tree as in the preceding section.

Just as for recursive trees, one can check that the conditions (H_k) are fulfilled for all $k \geq 1$ with $\ell(n) = \ln n$ and $L_k \equiv k(1 + \beta)/(2 + \beta)$; see, for instance, Section 4.4 of [6] in the case $\beta = 0$. Hence we know from Theorem 1 and Proposition 1 that percolation in the regime (1) produces a single giant cluster, and, more precisely, that (2) holds with $G(c) \equiv \theta(c) = e^{-c(1+\beta)/(2+\beta)}$. It has been shown recently in [4] that the asymptotic behavior of the sizes of the largest clusters for percolation on a scale-free tree is similar to that on a random recursive tree.

Theorem 3. *Let $T_n = T_n^{(\beta)}$ denote a random scale free tree on $\{0, 1, \dots, n\}$ with parameter $\beta > -1$. For every fixed integer j, in the regime (1) with $\ell(n) = \ln n$,*

$$\left(\frac{\ln n}{n} C_{p(n)}^1, \dots, \frac{\ln n}{n} C_{p(n)}^j \right)$$

converges in distribution towards

$$(\boldsymbol{x}_1, \dots, \boldsymbol{x}_j),$$

where $\boldsymbol{x}_1 > \boldsymbol{x}_2 > \cdots$ denotes the sequence of the atoms of a Poisson random measure on $(0, \infty)$ with intensity $c e^{-c(1+\beta)/(2+\beta)} x^{-2} \, dx$.

The key splitting property of random recursive trees fails for scale-free random trees, and the approach in [4] for establishing Theorem 3 thus departs significantly from that for Theorem 2. In short, one superposes Bernoulli bond percolation to the growth algorithm with preferential attachment as follows. Each time an edge is inserted, we draw an independent Bernoulli variable ε with parameter $p(n)$. If $\varepsilon = 1$, the edge is left intact; otherwise, we cut this edge in two at its midpoint. The upshot of cutting rather than removing edges is that the former procedure preserves the degrees of vertices, where the degree of a vertex is defined as the sum of the intact edges and half-edges attached to it. This is crucial for running the construction with preferential attachment.

This enables us to adapt a classical idea in this area (see, e.g. [6]), namely to consider a continuous-time version of the growth algorithm with preferential attachment and interpret the latter in terms of a continuous-time branching processes. Roughly speaking, incorporating percolation to the algorithm yields systems of branching processes with rare neutral mutations, where a mutation event corresponds to the insertion of an edge that is cut at its midpoint. Each branching process in the system corresponds to a percolation cluster which grows following a

dynamic with preferential attachment. One has to study carefully the asymptotic behavior of such systems of branching processes with neutral mutations, and then derive Theorem 3.

Acknowledgement

References

[1] ALDOUS, D. J. (1993). The continuum random tree. III. *Ann. Prob.* **21,** 248–289.

[2] BARABÁSI, A.-L. AND ALBERT, R. (1999). Emergence of scaling in random networks. *Science* **286,** 509–512.

[3] BERTOIN, J. (2012a). Sizes of the largest clusters for supercritical percolation on random recursive trees. *Random Structures Algorithms*, 16pp. Available at http://onlinelibrary.wiley.com/doi/10.1002/rsa.20448/abstract.

[4] BERTOIN, J. AND URIBE BRAVO, G. (2012b). Supercritical percolation on large scale-free random trees. Preprint. Available at http://hal.archives-ouvertes.fr/docs/00/76/32/33/PDF/PerkoOnBat.pdf.

[5] DRMOTA, M. (2009). *Random Trees.* Springer Wien New York, Vienna.

[6] DURRETT, R. (2007). *Random Graph Dynamics*. Cambridge University Press.

[7] FELLER, W. (1971). *An Introduction to Probability Theory and Its Applications,* Vol. II, 2nd edn. John Wiley, New York.

[8] IKSANOV, A. AND MÖHLE, M. (2007). A probabilistic proof of a weak limit law for the number of cuts needed to isolate the root of a random recursive tree. *Electron. Commun. Prob.* **12,** 28–35.

[9] LYONS, R. AND PERES, Y. (2013). *Probability on Trees and Networks*. In preparation. Available at http://mypage.iu.edu/~rdlyons/prbtree/prbtree.html.

[10] MEIR, A. AND MOON, J. W. (1974). Cutting down recursive trees. *Math. Biosci.* **21,** 173–181.

[11] PITMAN, J. (1999). Coalescent random forests. *J. Combinatorial Theory A* **85,** 165–193.

[12] PITMAN, J. (2006). *Combinatorial Stochastic Processes* (Lecture Notes Math. **1875**). Springer, Berlin.