Yule processes with rare mutation and their applications to percolation on $b$-ary trees.

G. Berzunza*

September 11, 2014

Abstract

We consider supercritical Bernoulli bond percolation on a large $b$-ary tree, in the sense that with high probability, there exists a giant cluster. We show that the size of the giant cluster has non-gaussian fluctuations, which extends a result due to Schweinsberg [16] in the case of random recursive trees. Using ideas in the recent work of Bertoin and Uribe Bravo [5], the approach developed in this work relies on the analysis of the sub-population with ancestral type in a system of branching processes with rare mutations, which may be of independent interest. This also allows us to establish the analog result for scale-free trees. In addition we obtain a weak limit theorem for the sizes of the next largest clusters.

Key words and phrases: Random tree, branching process, percolation, giant cluster, fluctuations.

1 Introduction and main result

Consider a tree of large but finite size $n$ and perform Bernoulli bond percolation with parameter $p_n \in (0, 1)$ that depends on the size of that tree. So each edge is removed with probability $1 - p_n$ and independently of the other edges, inducing a partition of the set of vertices into connected clusters. We are interested in the supercritical regime, in the sense that with high probability, there exist a giant cluster of size comparable to $n$, and its complement has also a size of order $n$. In fact, it has been shown recently in [4] that for fairly general families of trees, the supercritical regime corresponds to parameters of the form $p_n = 1 - c/\ell(n)$, where $\ell(n)$ is an estimate of the height of a typical vertex in the structure.

In the case of the uniform random recursive trees (i.e. trees on an ordered set of vertices where the smallest vertex serves as the root, and the sequence of vertices along any branch

*Institut für Mathematik, Universität Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland; e-mail: gabriel.berzunza@math.uzh.ch
from the root to a leaf is increasing) it easily seen that $\ell(n) = \ln n$, so choosing the percolation parameter so that

$$p_n = 1 - \frac{c}{\ln n},$$

(1)

where $c > 0$ is fixed, corresponds to the supercritical regime. More precisely, if $\Gamma_n$ denotes the size of the largest cluster, then $\Gamma_n \sim e^{-c} n$ in probability when $n \to \infty$. This can be viewed as the law of large numbers for the giant cluster, and it is then natural to investigate its fluctuations. Schweinsberg [16] (see also Bertoin [3] for an alternative approach) has shown that in this particular case, the fluctuations are non-Gaussian. Specifically

$$(n^{-1} \Gamma_n - e^{-c}) \ln n - ce^{-c} \ln \ln n \Rightarrow -ce^{-c}(Z + \ln c),$$

(2)

where $\Rightarrow$ means convergence in law as $n \to \infty$ and the variable $Z$ has the continuous Luria-Delbrück distribution, i.e. its characteristic function is given by

$$E(e^{i\theta Z}) = \exp \left(-\frac{\pi}{2} |\theta| - i\theta \ln |\theta| \right), \quad \theta \in \mathbb{R}.$$

The main purpose of this work is to investigate the case of large random $b$-ary recursive trees ($b \geq 2$). The process to build a $b$-ary recursive tree starts at $n = 1$ from the tree $T_1$ with one internal vertex (which corresponds to the root) and $b$ external vertices. Then, we suppose that $T_n$ has been constructed for some $n \geq 1$ that is a tree with $n$ internal vertices and $(b-1)n + 1$ external ones (also called leaves). Then choose an external vertex uniformly at random and replace it by an internal vertex to which $b$ new leaves are attached. In the case $b = 2$, the algorithm yields a so-called binary search tree (see for instance Mahmoud [14], Drmota [8]). We consider that the size of the tree is the number of internal vertices.

Then we perform a Bernoulli bond percolation with parameter given by (1) on a random $b$-ary recursive tree of size $n$. Just as in the case of the random recursive trees, this choice for the percolation parameter corresponds to the supercritical regime. Roughly speaking, since the $b$-ary recursive trees have also logarithm height, $\ell(n) = (b \ln n)/(b-1)$ (see Javanian and Vahidi-Asl [12]), one can verify that percolation then produces a giant cluster with size

$$C_0^{(p)} \sim e^{-\frac{b}{b-1} c} n \quad \text{in probability.}$$

We now state the central result of this work, which shows that the fluctuations of the giant cluster in the case of the $b$-ary recursive trees are also described by the continuous Luria-Delbrück distribution. We stress that this distribution was further observed in relation with a random algorithm for the isolation of the root, in the context of uniform random recursive tree by Iksanov and Möhle [11], and for random binary search tree by Holmgren [10].

**Theorem 1.** Set $\beta = b/(b-1)$, and assume that the percolation parameter $p_n$ is given by (1). Then as $n \to \infty$, there is the weak convergence

$$(n^{-1} C_0^{(p)} - e^{-\beta c}) \ln n - \beta c e^{-\beta c} \ln \ln n \Rightarrow -\beta c e^{-\beta c} Z_{c, \beta}$$
where

\[ Z_{c,\beta} = Z - \kappa_{\beta} + \ln \left( \frac{\beta}{\beta - 1} \right) \]  

(3)

with \( Z \) has the continuous Luria-Delbrück distribution,

\[ \kappa_{\beta} = 1 - \frac{1}{\beta} + \frac{1}{\beta} \sum_{k=2}^{\infty} \frac{(\beta)_k (-1)^k}{k! k - 1}, \]  

(4)

and \((x)_k = x(x-1)\cdots(x-k+1)\), for \( k \in \mathbb{N} \) and \( x \in \mathbb{R} \), is the Pochhammer function. In particular, for \( b = 2 \), i.e. for the binary search tree case, \( \kappa_2 = 1 \).

It should be noted the close similarity with the result for uniform recursive trees. It is remarkable that the normalizing functions and the limit in Theorem 1 only depend on the parameter \( \beta = b/(b-1) \) through some constants. Observe that the left-hand side of (2) is the same as in Theorem 1 for \( \beta = 1 \); however the expressions (3) and (4) are not defined for \( \beta = 1 \)!

The basic idea of Schweinsberg [16], for establishing the result (2) for uniform recursive trees relies on the estimation of the rate of decrease of the number of blocks in the Bolthausen-Sznitman coalescent, using the construction due to Goldschmidt and Martin [9] of the latter in terms of uniform recursive trees. On the other hand, the alternative approach of Bertoin [3] makes use on the remarkable coupling due to Iksanov and Möhle [11] connecting the Meir and Moon algorithm for the isolation of the root, with a certain random walk in the domain of attraction of the completely asymmetric Cauchy process. These approaches depend crucially on the splitting property (see Section 3.1 in Bertoin [2]) which fails for the \( b \)-ary recursive trees. We thus have to use a different argument, although some guiding lines are similar to [3].

Essentially, we consider a continuous time version of the growth algorithm of the \( b \)-ary tree which bears close relations to Yule processes. The connection between recursive trees and branching processes are well known, we make reference to Chauvin, et. al. [6] for the binary search trees and Bertoin and Uribe Bravo [5] for the case of scale-free trees. In this way, we adapt the recently strategy of [5]. Roughly speaking, incorporating percolation to the algorithm yields systems of branching processes with mutations, where a mutation event corresponds to disconnecting a leaf from its parent, and simultaneously replacing it by an internal vertex to which \( b \) new leaves are attached. Each percolation cluster size can then be thought of as a sub-population with some given genetic type. Hence the problem is reduced to study the fluctuations of the size of the sub-population with the ancestral type, which corresponds to the number of internal vertices connected to the root cluster.

The work is organized as follows. In Section 2 we introduce the system of branching processes with rare mutations. We investigate the fluctuations of the size of the sub-population with the ancestral type, when the total population goes to infinity and the mutation parameter \( 1 - p_n \) satisfies (11). Then in Section 3 we make the link with percolation on \( b \)-ary recursive trees in order to prove Theorem 1. We briefly show in Section 4.2 that the present approach also applies to study the fluctuations of the size of the giant cluster for percolation on scale-free trees. Finally,
we point out that the approach introduced by Bertoin and Uribe Bravo in \cite{[5]} to analyze the size of the next largest percolation clusters can be adapted without difficulty to the case of $b$-ary recursive trees.

2 Yule process with rare mutations

The purpose of this section is to introduce a system of branching process with rare mutations, which is quite similar to the one considered in \cite{[5]}, although there are also some key differences (in particular, death may occur causing the extinction of sub-populations). Then we focus on estimating the size of the sub-population with the ancestral type, when the total population in the system grows and the mutation parameter depends on the size of the latter.

We consider a population in which each individual is either a clone (ancestral type) of a mutant. A mutant individual lives for an exponential time of parameter 1 and gives birth at its death to $b$ children of the same genetic type as their parent. Similarly, a clone individual lives for an exponential time of parameter 1, and gives birth at its death to $b$ children. With probability $p \in (0,1)$, all children are clones, and with probability $1-p$, all children are mutants of the same new genetic type.

More precisely, the evolution of the population system is described by the process $Z^{(p)} = (Z^{(p)}(t) : t \geq 0)$, where

$$Z^{(p)}(t) = (Z_0^{(p)}(t), Z_1(t), \ldots), \quad \text{for } t \geq 0,$$

is a collection of nonnegative variables which represents the current size of the sub-population with type $i$. At the initial time, $Z_i^{(p)}(0) = 0$ for $i \geq 1$, and $Z_0^{(p)}(0) = z > 0$ which is the size of the ancestral population (i.e. clone). Formally, we take $Z^{(p)}$ to be a pure-jump Markov chain whose transitions are described as follows. When at state $z = (z_i : i \geq 0)$, our process jumps to a state $\hat{z} = (\hat{z}_i : i \geq 0)$ where $\hat{z}_j = z_j$ for $j \neq k$ and $\hat{z}_k = z_k + (b - 1)$ at rate

$$\begin{cases} 
  pz_0 & \text{if } k = 0, \\
  z_k & \text{if } k \neq 0.
\end{cases}$$

This corresponds to a reproduction event in the sub-population with type $k$. Otherwise, the process jumps from $z$ to $\hat{z} = (\hat{z}_i : i \geq 0)$ at rate $(1-p)z_0$ where, if $k$ is the first index such that $z_k = 0$, then $\hat{z}_0 = z_0 - 1$, $\hat{z}_k = b$, and $\hat{z}_j = z_j$ for $j \neq 0, k$. This corresponds to a mutation event of the sub-population with the ancestral type.

The process of the total size of the population in the system

$$Z(t) = Z^{(p)}_0(t) + \sum_{i \geq 1} Z_i(t), \quad t \geq 0,$$

is distributed as a Yule process, where each individual lives for an exponential time of parameter 1 and gives birth at its death to $b$ children, which then evolve independently one of the other
according to the same dynamics as their parent, no matter the choice of \( p \). Clearly, the process of the size of the sub-population with the ancestral type \( Z_0^{(p)} \) is a continuous time branching process, with reproduction law given by the distribution of \( b\epsilon_p \), where \( \epsilon_p \) stands for a Bernoulli random variable with parameter \( p \). Moreover, if for \( i \geq 1 \), we write
\[
a_i^{(p)} = \inf\{t \geq 0 : Z_i(t) > 0\},
\]
for the birth time of the sub-population with type \( i \), then each process
\[
(Z_i(t - a_i^{(p)}) : t \geq a_i^{(p)})
\]
is a branching process with the same reproduction law as \( Z \) starting from \( b \) for \( i \neq 0 \). Indeed, the different populations present in the system (i.e., those with strictly positive sizes) evolve independently one of the other. The following statement is just a formal formulation of the previous observation which should be plainly from the construction of \( Z^{(p)} \); it is essentially Lemma 1 in [5].

**Lemma 1.** The processes \((Z_i(t - a_i^{(p)}) : t \geq a_i^{(p)})\) for \( i \geq 1 \) form a sequence of i.i.d. branching process with the same law as \( Z \) but with starting value \( b \). Further, this sequence is independent of that of the birth-times \((a_i^{(p)})_{i \geq 1}\) and the process \( Z_0^{(p)} \) of the sub-population with ancestral type.

We are now ready to present the main result of this section. We henceforth assume that the parameter \( p = p_n \) is given by (1) and for simplicity, we write \( p \) rather than \( p_n \), omitting the integer \( n \) from the notation. Take \( \sigma \in \mathbb{N} \) and consider the time
\[
\tau(n) = \inf\{t \geq 0 : Z(t) = \sigma n + 1\},
\]
when the total population has size \( \sigma n + 1 \) (the reason for considering the process until the total population size reaches \( \sigma n + 1 \) rather than simply \( n \) is only technical, in order to establish Theorem [1]). The size of the sub-population with the ancestral type at this time is given by
\[
G_\sigma^{\sigma} := Z_0^{(p)}(\tau(n)).
\]

**Theorem 2.** Set \( \beta = b/(b - 1) \). As \( n \to \infty \), there is the weak convergence
\[
(n^{-1}G_\sigma^{\sigma} - \sigma\beta e^{-\beta c}) \ln n - \sigma\beta c e^{-\beta c} \ln \ln n \Rightarrow -\sigma\beta c e^{-\beta c} \left( Z_{c,\beta} + 1 - \frac{1}{\beta} + \ln \sigma(\beta - 1) \right),
\]
where \( Z_{c,\beta} \) is the random variable defined in (3).

We stress that this result also allows us to deduce the fluctuations of the number of mutants in the total population, since this quantity is given by \( \sigma n + 1 - G_\sigma^{\sigma} \).

The rest of this section is devoted to the proof of Theorem 2. Our guiding line is similar to that in [3]. Broadly speaking, we divide the study of the fluctuations in two well-defined phases. The crucial point is to obtain a precise estimate of the number \( \Delta_n \) of mutants when the total population of the system attains the size \( \lceil \ln^4 n \rceil \); this can be viewed as the germ of the fluctuations of \( \sigma n + 1 - G_\sigma^{\sigma} \). Then, we resume the growth of the system from size \( \lceil \ln^4 n \rceil \) to the size \( \sigma n + 1 \) and observe that the sub-population with the ancestral type grows essentially regularly. We point out that even though the study of these two phases plays a key role in [3], the tools developed here to deal with each phase are much different from those used there.
2.1 The germ of fluctuations

In this first phase, we observe the growth of the system of branching processes until the time

\[ \tau(\ln^4 n) = \inf\{t \geq 0 : Z(t) = \lfloor \ln^4 n \rfloor\}, \]

which is when the size of the total size of the population reaches \( \lfloor \ln^4 n \rfloor \), and our purpose in this section is to estimate precisely the number \( \Delta_n \) of mutants in the total population at this time, that is

\[ \Delta_n = \lfloor \ln^4 n \rfloor - Z_0^{(p)}(\tau(\ln^4 n)). \]

We stress that the threshold \( \ln^4 n \) is somewhat arbitrary, and any power close to 4 of \( \ln n \) would work just as well. However, as is remarked by Bertoin in [3], it is crucial to choose a threshold which is both sufficiently high so that fluctuations are already visible, and sufficiently low so that one can estimate the germ with the desired accuracy.

We start by setting down the key results that lead us to the main result of this section, in order to give an easier articulation of the argument. In this direction, it is convenient to introduce the number \( \Delta_n^0 \) of mutants at time

\[ \tau_0(\ln^4 n) = \inf\{t \geq 0 : Z_0^{(p)}(t) = \lfloor \ln^4 n \rfloor\}, \]

which is when the size of the sub-population with the ancestral type reaches \( \lfloor \ln^4 n \rfloor \), i.e.

\[ \Delta_n^0 = Z(\tau_0(\ln^4 n)) - \lfloor \ln^4 n \rfloor. \]

This will be useful as the distribution of \( \Delta_n^0 \) is easier to estimate than that of \( \Delta_n \). Then, we establish the following limit theorem in law that relates the fluctuations of \( \Delta_n^0 \) with the continuous Luria-Delbrück variable \( Z \).

**Proposition 1.** As \( n \to \infty \), there is the weak convergence

\[ \frac{\Delta_n^0}{\ln^3 n} - 3\beta c \ln \ln n \Rightarrow \beta c \left( Z_{c,\beta} + 1 - \frac{1}{\beta} + \ln(\beta - 1) \right) \]

where \( Z_{c,\beta} \) is the random variable defined in [3].

As we are interested in estimate the number \( \Delta_n \) of mutants in the total population at time \( \tau(\ln^4 n) \), and we know the behavior of \( \Delta_n^0 \), the purpose of the next lemma is to point out that these two quantities are close enough when \( n \to \infty \). We need to introduce the notation:

\[ A_n = B_n + o(f(n)) \text{ in probability}, \]

where \( A_n \) and \( B_n \) are two sequences of random variables and \( f : \mathbb{N} \to (0, \infty) \) a function, to indicate that \( |A_n - B_n|/f(n) \to 0 \) in probability when \( n \to \infty \).

**Lemma 2.** We have

\[ \Delta_n = \Delta_n^0 + o(\ln^3 n) \text{ in probability}. \]
It then follows from Proposition 1 that $\Delta_n$ and $\Delta_0^n$ have the same asymptotic behavior. Specifically:

**Corollary 1.** As $n \to \infty$, there is the weak convergence

$$\frac{\Delta_n}{\ln^3 n} - 3\beta c \ln \ln n \Rightarrow \beta c \left( Z_{c,\beta} + 1 - \frac{1}{\beta} + \ln(\beta - 1) \right)$$

where $Z_{c,\beta}$ is the random variable defined in (3).

The above result will be sufficient for our purpose. We prepare the ground for the proofs of Proposition 1 and Lemma 2.

Recall that we wish to study the behavior of the number $\Delta_0^n$ of mutants at time $\tau_0(\ln^4 n)$, which is easier than that of $\Delta_n$, thanks to Lemma 1. In words, at time $\tau_0(\ln^4 n)$ there is an independence property between the mutant sub-populations, and the process that counts the number of mutation events, that allows us to express $\Delta_0^n$ as a random sum of independent Yule processes. Clearly, the above is not possible to effectuate at time $\tau(\ln^4 n)$ due to the lack of independence within the sub-populations.

Formally, we start by writing

$$M(t) = \max\{i \geq 1 : Z_i(t) > 0\}$$

for the number of mutations that have occurred before time $t \geq 0$. Lemma 1 ensures that $M$ is independent of the processes $(Z_i(t - a_i^{(p)}) : t \geq a_i^{(p)})$ for $i \geq 1$. In addition, we note that the jump times of $M$ are in fact $a_1^{(p)} < a_2^{(p)} < \cdots$. This enables us to express the total mutant population at time $t$ as,

$$Z_m(t) = \sum_{i=1}^{M(t)} Z_i(t - a_i^{(p)}),$$

and we are thus interested in

$$\Delta_0^n = Z_m(\tau_0(\ln^4 n)).$$

We now turn our attention to study the fluctuations of $\Delta_0^n$ through the analysis of its characteristic function. In this direction, we will be mainly interested in the following feature of $Z_m(t)$.

**Lemma 3.** We have for $t \geq 0$ and $\theta \in \mathbb{R}$.

i) The characteristic function of $Z(t)$ started from $Z(0) = b$,

$$\varphi_t(\theta) = \mathbb{E} [e^{i\theta Z(t)} | Z(0) = b] = \left( \frac{e^{i\theta(b-1)} e^{-(b-1)t}}{1 - e^{i\theta(b-1)} + e^{i\theta(b-1)} e^{-(b-1)t}} \right)^{\frac{1}{t}}.$$

\[7\]
ii) We have
\[ E[e^{i\theta Z_m(t)}] = E \left[ \exp \left( (1-p) \int_0^t Z_0^{(p)}(t-s) (\varphi_s(\theta) - 1) \, ds \right) \right]. \] (7)

Proof: Recall that the processes \( Z_i(t - a_i^{(p)} : t \geq a_i^{(p)}) \) for \( i \geq 1 \) are i.i.d. branching process with the same law as \( Z \) but with starting value \( b \). Then according to the page 109 in Chapter III of Athreya and Ney \[1\], their characteristic function is given by the expression (6).

We now observe from the dynamics of \( Z^{(p)} \) that the counting process \( M \) has jumps at rate \((1-p)Z_0^{(p)}\). Moreover, conditionally on \( Z_0^{(p)} \), the process \( Z_m \) is a non homogeneous filtered Poisson process whose characteristic function can be written in terms of the characteristic function of \( Z_i \). By extending equation (5.43) of Parzen \[15\] slightly to allow the underlying Poisson process to be non homogeneous, we obtain
\[ E \left[ e^{i\theta Z_m(t)} \mid Z_0^{(p)}(s) : 0 \leq s \leq t \right] = \exp \left( (1-p) \int_0^t Z_0^{(p)}(s) (\varphi_{t-s}(\theta) - 1) \, ds \right), \]
for \( t \geq 0 \) and \( \theta \in \mathbb{R} \). Our claim follows after taking expectation on both sides of the equation and make a simple change of variables.

We recall some important properties of the branching processes \( Z \) and \( Z_0^{(p)} \), which will be useful later on. The process \[ W(t) := e^{-(b-1)t}Z(t), \quad t \geq 0 \]
is a nonnegative square-integrable martingale which converges a.s. and in \( L^2(\mathbb{P}) \), and we write \( W(\infty) \) for its terminal value. Furthermore \( W(\infty) > 0 \) a.s. since \( Z \) can not become extinct (we also pointed out that \( Z \) never explodes a.s.). Similarly, the process \[ W_0^{(p)}(t) = e^{-(bp-1)t}Z_0^{(p)}(t), \quad t \geq 0 \]
is a martingale which terminal value is denoted by \( W_0^{(p)}(\infty) \). In addition, following the same argument as Lemma 3 in \[5\] we have

**Lemma 4.** For every \( z > 0 \), it holds that
\[ \lim_{p \to 1, t \to \infty} E_z \left[ \sup_{s \geq t} \left| W_0^{(p)}(s) - W(\infty) \right|^2 \right] = 0 \]
where we write \( \mathbb{P}_z \) for the distribution of the branching process \( Z \) started from \( z \).

We next estimate the characteristic function of \( Z_m(t) \) given in \[7\], but we still need some additional notation. For \( t \geq 0 \),
\[ I^{(p)}(t) = (1-p) \int_0^t Z_0^{(p)}(t-s)(\varphi_s(u) - 1) \, ds \]
and
\[ I_m^{(p)}(t) = (1 - p) W_0^{(p)}(\infty) e^{(b-1)t} \int_0^t e^{-(b-1)s} (\varphi_s(u) - 1) ds, \]
where \( u = \theta / (\beta c \ln^3 n) \) for \( \theta \in \mathbb{R} \) and \( \beta = b/(b - 1) \).

**Lemma 5.** We have
\[ \lim_{n \to +\infty} \left( I^{(p)}(\tau_0(\ln^4 n)) - I_m^{(p)}(\tau_0(\ln^4 n)) \right) = 0 \quad \text{in probability.} \]

**Proof:** Define the function
\[ I_a^{(p)}(t) = (1 - p) W_0^{(p)}(\infty) e^{(bp-1)t} \int_0^t e^{-(bp-1)s} (\varphi_s(u) - 1) ds, \quad t \geq 0, \]
which is simply obtained by replacing \( b \) by \( bp \) in the exponential terms of \( I_m^{(p)}(t) \). We first prove that
\[ \lim_{n \to +\infty} \left( I^{(p)}(\tau_0(\ln^4 n)) - I_a^{(p)}(\tau_0(\ln^4 n)) \right) = 0 \quad \text{in probability.} \tag{8} \]

In this direction, we observe from the triangle inequality that
\[
\left| I^{(p)}(\tau_0(\ln^4 n)) - I_a^{(p)}(\tau_0(\ln^4 n)) \right|
\leq (1 - p) e^{(bp-1)\tau_0(\ln^4 n)} \int_0^{\tau_0(\ln^4 n)} \left| W_0^{(p)}(\tau_0(\ln^4 n) - s) - W_0^{(p)}(\infty) \right| \left| \varphi_s(u) - 1 \right| e^{-(bp-1)s} ds. \tag{9}
\]

Next define
\[ A^{(p)} := \frac{3}{2(bp - 1)} \sup_{s \geq 0} e^{(bp-1)s/3} \left| W_0^{(p)}(s) - W_0^{(p)}(\infty) \right|, \]
and since Lemma 2 in [5] shows that \( A^{(p)} \) is bounded in \( L^2(\mathbb{P}) \), we have by the Markov inequality that
\[ \lim_{n \to +\infty} \left( \ln^{-\frac{2}{3}} n \right) A^{(p)} = 0 \quad \text{in probability.} \tag{10} \]

We set \( t_n = (b - 1)^{-1} \ln \ln n \) and recall that \( \varphi_t(u) \) fulfills \([3]\). On the one hand, using that \( |e^{ix} - 1| \leq 2 \) for \( x \in \mathbb{R} \), we have
\[ |\varphi_t(u) - 1| \leq 2. \]

Then,
\[
(1 - p) e^{(bp-1)\tau_0(\ln^4 n)} \int_{\tau_0(\ln^4 n) - t_n}^{\tau_0(\ln^4 n)} \left| W_0^{(p)}(\tau_0(\ln^4 n) - s) - W_0^{(p)}(\infty) \right| \left| \varphi_s(u) - 1 \right| e^{-(bp-1)s} ds
\leq 2(1 - p) \int_0^{t_n} \left| W_0^{(p)}(s) - W_0^{(p)}(\infty) \right| e^{(bp-1)s} ds
\leq 2(1 - p) \left( \ln^{-\frac{2}{3}} n \right) A^{(p)}. \tag{11}
\]
On the other hand, since $|e^{ix} - 1| \leq |x|$ for $x \in \mathbb{R}$, we get from (6)

$$|\varphi_t(u) - 1| \leq b|u|e^{(b-1)t},$$

and thus,

$$(1 - p)e^{(bp-1)\tau_0(\ln^4 n)} \int_0^{\tau_0(\ln^4 n) - t_n} |W_0^{(p)}(\tau_0(\ln^4 n) - s) - W_0^{(p)}(\infty)||\varphi_s(u) - 1|e^{-(bp-1)s}ds$$

$$\leq (1 - p)b|u|e^{(b-1)\tau_0(\ln^4 n)} \int_{t_n}^{\tau_0(\ln^4 n)} |W_0^{(p)}(s) - W_0^{(p)}(\infty)|e^{-(b-1)s}e^{(bp-1)s}ds$$

$$\leq 2(1 - p)\left(\ln^{-\frac{1}{3}} n\right)b|u|A^{(p)}e^{(b-1)\tau_0(\ln^4 n)}. \quad (12)$$

Recall that $u = \theta/(\beta c \ln^3 n)$, then combining (9), (11), (12) and the fact $p = p_n$ is given by (11), it follows that

$$|I^{(p)}(\tau_0(\ln^4 n)) - \overline{I^{(p)}}(\tau_0(\ln^4 n))| \leq 2\left(c + (b - 1)|\theta| (\ln^{-4} n) e^{(b-1)\tau_0(\ln^4 n)}\right)(\ln^{-\frac{1}{3}} n)A^{(p)}.$$

So, we deduce (8) from (11), and Lemma 4 which ensures that

$$\lim_{n \to \infty} (\ln^{-4} n) e^{(b-1)\tau_0(\ln^4 n)} = \frac{1}{W(\infty)} \text{ in probability},$$

since $Z_0^{(p)}(\tau_0(\ln^4 n)) = [\ln^4 n]$. Finally, using similar arguments, we can establish that

$$\lim_{n \to \infty} \overline{I^{(p)}}(\tau_0(\ln^4 n)) - \overline{I^{(p)}}(\tau_0(\ln^4 n)) = 0 \text{ in probability},$$

and our claim follows.

$$\square$$

We observe, that thanks to (6), the integral $I^{(p)}_m$ can be computed explicitly.

**Lemma 6.** We have for $t \geq 0$,

$$\int_0^t e^{-(b-1)s}(\varphi_s(u) - 1)ds = \frac{1 - e^{iu(b-1)}}{(b-1)e^{iu(b-1)}} \left(\beta \ln(1 - e^{iu(b-1)} + e^{iu(b-1)}e^{-(b-1)t}) + \kappa_{b,u}(t)\right),$$

where

$$\kappa_{b,u}(t) = \sum_{k=2}^{\infty} \frac{(\beta)_k}{k!} \left(\frac{e^{iu(b-1)} - 1}{k-1}\right)^{k-1} \left(1 - \frac{1}{(1 - e^{iu(b-1)} + e^{iu(b-1)}e^{-(b-1)t})^{k-1}}\right), \quad (13)$$

and $(\cdot)_k$ is the Pochhammer function.
Proof: Define the function
\[
 f(\lambda) = \int_0^t e^{-(b-1)r} \left( \frac{e^{-\lambda(b-1)}e^{-(b-1)r}}{1 - e^{-\lambda(b-1)} + e^{-\lambda(b-1)}e^{-(b-1)r}} \right)^\beta - 1 \right) dr, \quad \lambda \geq 0.
\]

Setting \( x = 1 - e^{-\lambda(b-1)} + e^{-\lambda(b-1)}e^{-(b-1)r} \) and \( y_\lambda = e^{-\lambda(b-1)} \) for convenience, we have that
\[
 f(\lambda) = \frac{1}{(b-1)y_\lambda} \int_{1-y_\lambda+y_\lambda e^{-(b-1)t}}^1 \left( \frac{x + y_\lambda - 1}{x} \right)^\beta - 1 \right) dx.
\]

Moreover, using a well-known extension of Newton’s binomial formula, we get
\[
 f(\lambda) = \frac{1}{(b-1)y_\lambda} \sum_{k=1}^{\infty} \frac{(\beta)_k}{k!} (y_\lambda - 1)^k \int_{1-y_\lambda+y_\lambda e^{-(b-1)t}}^1 x^{-k} dx,
\]
where \((\cdot)_k\) is the Pochhammer function. Note that the series converges absolutely since \(\beta > 0\) and \(|y_\lambda - 1|/x \leq 1\), for \(1 - y_\lambda + y_\lambda e^{-(b-1)t} \leq x \leq 1\). Then straightforward calculations yields
\[
 f(\lambda) = \frac{1 - y_\lambda}{(b-1)y_\lambda} \left( \beta \ln(1 - y_\lambda + y_\lambda e^{-(b-1)t}) + \kappa'_{b,\lambda}(t) \right),
\]
where
\[
 \kappa'_{b,\lambda}(t) = \sum_{k=2}^{\infty} \frac{(\beta)_k}{k!} \frac{(y_\lambda - 1)^{k-1}}{k-1} \left( 1 - \frac{1}{(1 - y_\lambda + y_\lambda e^{-(b-1)t})^{k-1}} \right).
\]

One can note that the function \( f \) allows an analytic extension to \( \{ \lambda \in \mathbb{C} : \text{Re} \lambda \geq 0 \} \). Then, taking into account the principal branch of the complex logarithm, we conclude that
\[
 f(-iu) = \int_0^t e^{-(b-1)r} \left( \frac{e^{iu(b-1)}e^{-(b-1)r}}{1 - e^{iu(b-1)} + e^{iu(b-1)}e^{-(b-1)r}} \right)^\beta - 1 \right) dr.
\]

Therefore our assertion follows, observing that \( \kappa'_{b,\lambda}(t) = \kappa_{b,u}(t) \) when \( \lambda = -iu \).

\[\square\]

We are now able to establish Proposition 1.

Proof of Proposition 1: Fix \( \theta \in \mathbb{R} \) and define \( m_n = \beta c \ln^2 n \). It follows from the identity (5) and Lemma 3 that the characteristic function of \( m_n^{-1} \Delta_0^n - \ln m_n \) is given by
\[
 \mathbb{E} \left[ e^{i\theta(m_n^{-1} \Delta_0^n - \ln m_n)} \right] = \mathbb{E} \left[ \exp \left( I^{(p)}(\tau_0(\ln^4 n)) - i\theta \ln m_n \right) \right].
\]

Recall that by Lemma 3 we have
\[
 \lim_{n \to \infty} \left( I^{(p)}(\tau_0(\ln^4 n)) - I^{(p)}_m(\tau_0(\ln^4 n)) \right) = 0 \quad \text{in probability.}
\]
Then, we must verify that
\[
\lim_{n \to \infty} \left( \tilde{I}_m(p)(\tau_0(\ln^4 n)) - i \theta \ln m_n \right) = -i \theta \left( \kappa_\beta - 1 + \frac{1}{\beta} \right) - i \theta \ln |\theta| - \frac{1}{2} \pi |\theta| \quad \text{in probability.} \quad (14)
\]

In this direction, we define \( y_u = e^{iu(b-1)} \) for convenience and recall also from Lemma 6 that
\[
\tilde{I}_m(p)(\tau_0(\ln^4 n)) = (1 - p)W_0(p)(\infty)e^{(b-1)\tau_0(\ln^4 n)} \frac{1 - y_u}{(b-1)y_u} \left( \beta \ln(1 - y_u + y_u e^{-(b-1)\tau_0(\ln^4 n)}) + \kappa_{b,u}(\tau_0(\ln^4 n)) \right),
\]
where \( \kappa_{b,u}(\tau_0(\ln^4 n)) \) is defined in (13) and \( u = \theta / (\beta c \ln^3 n) \).

We know from Lemma 4 that
\[
\lim_{n \to \infty} e^{-(b-1)\tau_0(\ln^4 n)}Z(\tau_0(\ln^4 n)) = \lim_{n \to \infty} e^{-(bp-1)\tau_0(\ln^4 n)}Z_0(p)(\tau_0(\ln^4 n)) = W(\infty) \quad \text{in probability},
\]
and also we note that \( W_0(p)(\infty) \sim W(\infty) \) in probability (see equation (4) in [5]). Then thanks to (1), we deduce that
\[
(1 - p)W_0(p)(\infty)e^{(b-1)\tau_0(\ln^4 n)} \sim c \ln^3 n \quad \text{in probability}.
\]

On the other hand,
\[
y_u = 1 + O \left( \frac{1}{m_n} \right) \quad \text{and} \quad m_n(1 - y_u) = -i \theta + O \left( \frac{1}{m_n} \right).
\]

Then, it is easy to deduce
\[
\lim_{n \to \infty} \left( \tilde{I}_m(p)(\tau_0(\ln^4 n)) - i \theta \ln m_n \right) = -i \theta \ln(-i \theta) - i \theta \left( \kappa_\beta - 1 + \frac{1}{\beta} \right) \quad \text{in probability},
\]
which implies (14). We conclude that the Fourier transform of \( m_n^{-1} \Delta_n - \ln m_n \) converges pointwise as \( n \) tends to infinity to the continuous function
\[
\theta \mapsto \exp \left( -i \theta \left( \kappa_\beta - 1 + \frac{1}{\beta} \right) - i \theta \ln |\theta| - \frac{1}{2} \pi |\theta| \right),
\]
and then our claim follows for the continuity theorem for Fourier transforms.

\[\square\]

We now turn our attention to the proof of Lemma 2.

We imagine first that we begin our observation of the system of branching processes with rare mutations once it has reached the size \( \lfloor \ln^4 n \rfloor \), that is, from the time \( \tau(\ln^4 n) \). We thus write \( Z' = (Z'(t) : t \geq 0) \) for the process of the total size of the population started from
\( Z'(0) = \lfloor \ln^4 n \rfloor \), which has the same law of the Yule process \( Z \) described in Section 2. We introduce the time

\[
\tau'(\ln^4 n) = \inf\{t \geq 0 : Z'(t) = Z(\tau_0(\ln^4 n))\},
\]

at which it hits \( Z(\tau_0(\ln^4 n)) \). Equivalently, the time needed to have a population with the ancestral type of size \( \lfloor \ln^4 n \rfloor \). We shall first estimate this quantity.

**Lemma 7.** We have

\[
\tau'(\ln^4 n) = o(1) \quad \text{in probability.}
\]

**Proof:** We know that

\[
\lim_{n \to \infty} e^{-(b-1)\tau(\ln^4 n)} Z(\tau(\ln^4 n)) = W(\infty) \quad \text{in probability.}
\]

On the other hand from Lemma 4 we have

\[
\lim_{n \to \infty} e^{-(b-1)\tau_0(\ln^4 n)} Z_0^{(p)}(\tau_0(\ln^4 n)) = W(\infty) \quad \text{in probability.}
\]

By definition of \( \tau(\ln^4 n) \) and \( \tau_0(\ln^4 n) \), we have \( Z(\tau(\ln^4 n)) = Z_0^{(p)}(\tau_0(\ln^4 n)) = \lfloor \ln^4 n \rfloor \), hence

\[
e^{-(b-1)\tau(\ln^4 n)} \sim \frac{W(\infty)}{\ln^4 n} \quad \text{and} \quad e^{-(b-1)\tau_0(\ln^4 n)} \sim \frac{W(\infty)}{\ln^4 n} \quad \text{in probability}
\]

and, we observe that thanks to (1), \( \tau_0(\ln^4 n) \sim 4(b-1)^{-1} \ln \ln n \). Therefore, our claim follows from the identity

\[
\tau'(\ln^4 n) = \tau_0(\ln^4 n) - \tau(\ln^4 n).
\]

\[\square\]

We observe that the population at time \( \tau(\ln^4 n) \) when we start our observation consists in \( \Delta_n \) mutants and \( \lfloor \ln^4 n \rfloor - \Delta_n \) individuals of the ancestral type. Then, focusing on the sub-population with the ancestral type, we write \( Z_0^{(p)} = (Z_0^{(p)}(t) : t \geq 0) \) for the process that counts the number of individuals with the ancestral type, which has the same law of \( Z_0^{(p)} \) but starting from \( Z_0^{(p)}(0) = \lfloor \ln^4 n \rfloor - \Delta_n \). We recall that

\[
W'(t) := e^{-(b-1)t} Z'(t) \quad \text{and} \quad W_0^{(p)}(t) := e^{-(bp-1)t} Z_0^{(p)}(t), \quad t \geq 0
\]

are nonnegative square-integrable martingales which converge a.s. and in \( L^2(\mathbb{P}) \).

**Proof of Lemma 2:** An application of Doob’s inequality (see, e.g., Equation (6) in [7]) shows for all \( \eta > 0 \) that

\[
\lim_{n \to \infty} \mathbb{P}\left(\left| e^{-(b-1)\tau'(\ln^4 n)} Z'(\tau'(\ln^4 n)) - Z'(0) \right| > \eta \ln^3 n \right) = 0
\]

13
and using the fact that $Z_0^{(p)}(0) \leq \lfloor \ln^4 n \rfloor$, we also get

$$\lim_{n \to \infty} \mathbb{P}\left( \left| e^{-(bp-1)\tau'(\ln^4 n)} Z_0^{(p)}(\tau'(\ln^4 n)) - Z_0^{(p)}(0) \right| > \eta \ln^3 n \right) = 0.$$  

Then, since $\Delta_n = Z'(\tau'(\ln^4 n)) - \lfloor \ln^4 n \rfloor$, $\Delta_n = \lfloor \ln^4 n \rfloor - Z_0^{(p)}(0)$, and $Z'(0) = \lfloor \ln^4 n \rfloor$, one readily gets

$$\Delta_n = Z'(\tau'(\ln^4 n)) \left( e^{-(b-1)\tau'(\ln^4 n)} - 1 \right) - Z_0^{(p)}(\tau'(\ln^4 n)) \left( e^{-(bp-1)\tau'(\ln^4 n)} - 1 \right) + o(\ln^3 n)$$

in probability. We next note from Lemma 7 that

$$Z_0^{(p)}(\tau'(\ln^4 n)) \left( e^{(1-p)\tau'(\ln^4 n)} - 1 \right) = o(\ln^3 n) \quad \text{in probability},$$

which yields

$$\Delta_n = \left( W'(\tau'(\ln^4 n)) - W_0^{(p)}(\tau'(\ln^4 n)) \right) \left( 1 - e^{(b-1)\tau'(\ln^4 n)} \right) + o(\ln^3 n)$$

in probability. The claim of Lemma 2 follows from Lemma 7 and the computation of the second moment of $W'(\tau'(\ln^4 n))$ and $W_0^{(p)}(\tau'(\ln^4 n))$ (see Lemma 2 and its proof in [5]), showing that

$$W'(\tau'(\ln^4 n)) - W_0^{(p)}(\tau'(\ln^4 n)) = o(\ln^3 n) \quad \text{in probability.}$$

\[\square\]

### 2.2 The spread of fluctuations

The purpose here is to resume the growth of the system of branching processes with rare mutation from the size $\lfloor \ln^4 n \rfloor$ to the size $\sigma n + 1$ and observe that the germ of the fluctuations $\Delta_n$ spreads regularly. In this direction, we proceed similarly as the last part of the preceding section.

We recall that $Z'$ denotes the process of the total population started from $Z'(0) = \lfloor \ln^4 n \rfloor$. We consider

$$\tau'(n) = \inf\{ t \geq 0 : Z'(t) = \sigma n + 1 \},$$

the time needed for the total population to reach size $\sigma n + 1$. So that, in the notation of Theorem 2

$$G_\sigma^n = Z_0^{(p)}(\tau'(n)),$$

where as the previous section, we write $Z_0^{(p)}$ for the process that counts the number of individuals with the ancestral type starting from $Z_0^{(p)}(0) = \lfloor \ln^4 n \rfloor - \Delta_n$.

We have now all the ingredients to establish Theorem 2.
Proof of Theorem 2: Again from the estimate of Equation (6) in [7], we get for all \( \eta > 0 \) that
\[
\lim_{n \to \infty} P \left( \left| (\sigma n + 1) e^{-(b-1)r'(n)} - \ln^4 n \right| > \eta \ln^3 n \right) = 0 \tag{15}
\]
and using the fact that \( Z_0^{(p)}(0) \leq \lfloor \ln^4 n \rfloor \), we also get
\[
\lim_{n \to \infty} P \left( \left| e^{-(bp-1)r'(n)} Z_0^{(p)}(\tau'(n)) - Z_0^{(p)}(0) \right| > \eta \ln^3 n \right) = 0.
\]
We deduce that
\[
G_{\sigma} = e^{(bp-1)r'(n)}(\ln^4 n - \Delta_n) + o\left( \frac{n}{\ln n} \right) \text{ in probability.}
\]
Skorokhod’s representation theorem enables us to assume that the weak convergence in Corollary 1 holds almost surely and then we have
\[
G_{\sigma} = e^{(bp-1)r'(n)}(\ln^4 n - \beta c \ln^3 n \left( 3 \ln \ln n + \left( Z_{c,\beta} + 1 - \frac{1}{\beta} + \ln(\beta - 1) \right) \right)) + o\left( \frac{n}{\ln n} \right)
\]
in probability. We next note from (15) that
\[
e^{(bp-1)r'(n)} = \sigma e^{-\beta c} \frac{n}{\ln^3 n} + \sigma \beta c e^{-\beta c} \frac{n}{\ln^3 n} (4 \ln \ln n - \ln \sigma) + o\left( \frac{1}{\ln n} \right)
\]
in probability. It follows that
\[
G_{\sigma} = \sigma e^{-\beta c} + \sigma \beta c e^{-\beta c} \frac{\ln \ln n}{\ln n} - \sigma \beta c e^{-\beta c} \frac{n}{\ln n} \left( Z_{c,\beta} + 1 - \frac{1}{\beta} + \ln(\beta - 1) \right) + o\left( \frac{n}{\ln n} \right)
\]
in probability, which completes the proof.

3 Proof of Theorem 1

Our approach is based in the introduction of a continuous version of the construction of a \( b \)-ary recursive tree that enables us to superpose Bernoulli bond percolation dynamically in the tree structure. We begin at time 0 from the tree with just one internal vertex which corresponds to the root having \( b \) external vertices. Once the random tree with size \( n \geq 1 \) has been constructed, we equip each of the \((b-1)n + 1\) external vertices with independent exponential random variables \( \zeta_i \) of parameter 1. Then, the next vertex \( n + 1 \) is attached after a waiting time equal to \( \min_{i \in \{1, \ldots, (b-1)n+1\}} \zeta_i \) to one of the external vertices chosen uniformly random. We observe that \( \min_{i \in \{1, \ldots, (b-1)n+1\}} \zeta_i \) is exponentially distributed with parameter \((b-1)n + 1\).

We denote by \( T(t) \) the tree which has been constructed at time \( t \geq 0 \), and by \( |T(t)| \) its size, i.e. the number of internal vertices. The process of the size \( (|T(t)| : t \geq 0) \) is clearly Markovian and if we define
\[
\gamma(n) = \inf\{t \geq 0 : |T(t)| = n\}, \quad n \geq 1,
\]
then \( T(\gamma(n)) \) is a version of the \( b \)-ary recursive tree of size \( n \), \( T_n \). However for our purpose it will be more convenient work with the process \( Y \) defined by

\[
Y(t) = (b - 1)|T(t)| + 1, \quad t \geq 0
\]

with starting value \( Y(0) = b \). It should be clear that \( Y \) is a Yule process as described at Section 2, i.e. it has jumps of size \( b - 1 \) and unit birth rate per unit population size. We also point out that the process \( Y \) gives us the number of external vertices on the tree.

Next superpose Bernoulli bond percolation with parameter \( p = p_n \) defined in (11) to the growth algorithm in continuous time of the \( b \)-ary recursive tree. We draw an independent Bernoulli random variable \( \epsilon_p \) with parameter \( p \), each time an edge is inserted. If \( \epsilon_p = 1 \), the edge is left intact, otherwise we cut this edge at its midpoint. This disconnects the tree into connected clusters which motivates the following. We write \( T^{(p)}(t) \) for the resulting combinatorial structure at time \( t \). So, the percolation clusters of \( T(t) \) are the connected components by a path of intact edges of \( T^{(p)}(t) \). This way to introduce percolation was developed by Bertoin and Uribe Bravo [5].

Let \( T_0^{(p)}(t) \) be the subtree that contains the root. We write \( H_0^{(p)}(t) \) for the number of half-edges pertaining to the root cluster at time \( t \). So that, its number of external vertices is given by

\[
Y_0^{(p)}(t) = (b - 1)|T_0^{(p)}(t)| + 1 - H_0^{(p)}(t).
\]

**Remark 1.** We stress that is possible to make a similar construction in order to study the rest percolation clusters. We discuss this in the Section 4.2.

We are now be able to observe the connection with the system of branching processes with rare mutations described in the preceding section. It should be plain from the construction that the size of the root-cluster at time \( t \), i.e. \( Y_0^{(p)}(t) \), of \( T(t) \) after percolation with parameter \( p \), coincides with the number of individuals with the ancestral type \( Z_0^{(p)}(t) \) in the system \( Z^{(p)} \) of branching processes with rare mutations of Section 2. In fact, we already mentioned that the process \( Y \) has the same random evolution as the process of the total size in the system \( Z \).

Recall that the algorithm of construction of a \( b \)-ary recursive tree is run until the time

\[
\gamma(n) = \inf\{t \geq 0 : |T(t)| = n\} = \inf\{t \geq 0 : Y(t) = (b - 1)n + 1\}
\]

when the structure has size \( n \). Then, the size \( C_0^{(p)} \) of the percolation cluster containing the root when the tree has \( n \) vertices satisfies

\[
C_0^{(p)} = |T_0^{(p)}(\gamma(n))|.
\]

In addition, it should be plain that

\[
Y_0^{(p)}(\gamma(n)) = (b - 1)C_0^{(p)} + 1 - H_0^{(p)}(\gamma(n)),
\]

coincides with the number of individuals with the ancestral type in the branching system \( Z^{(p)} \), at time when the total population reaches the size \( (b - 1)n + 1 \), i.e. \( C_0^{b-1} \), according to the notation of Theorem 2. Hence in order to establish Theorem 1 it is sufficient to get an estimate of the number of half-edges pertaining to the root-subtree at time \( \gamma(n) \).
Lemma 8. We have
\[ \lim_{n \to \infty} \frac{\ln n}{n} H_0^{(p)}(\gamma(n)) = ce^{-\beta c} \quad \text{in probability}. \]

Proof: We observe that the processes
\[ H_0^{(p)}(t) - (1 - p) \int_0^t Y_0^{(p)}(s)ds \quad \text{and} \quad Y_0^{(p)}(t) - (bp - 1) \int_0^t Y_0^{(p)}(s)ds, \quad t \geq 0 \]
are martingales. Thus,
\[ L^{(p)}(t) := H_0^{(p)}(t) - \frac{1 - p}{bp - 1} Y_0^{(p)}(t), \quad t \geq 0 \]
is also a martingale. Observe that since \( p = p_n \) satisfies (11), for \( n \) large enough such that \( 2/(b + 1) \leq p \leq 1 \), its jumps \( |L^{(p)}(t) - L^{(p)}(t-)| \) have size at most \( b \). Since there are at most \( n \) jumps up to time \( \gamma(n) \), the bracket of \( L^{(p)} \) can be bounded by \( [L^{(p)}]_{\gamma(n)} \leq b^2 n \). Hence we have
\[ \lim_{n \to \infty} \mathbb{E} \left( \left| \frac{\ln n}{n} L^{(p)}(\gamma(n)) \right|^2 \right) = 0. \quad (17) \]

On the other hand, we know from Lemma 8 that
\[ \lim_{n \to \infty} e^{-\beta c} Y_0^{(p)}(\gamma(n)) = \lim_{n \to \infty} e^{-(bp - 1)\gamma(n)} Y_0^{(p)}(\gamma(n)) = W(\infty) \quad \text{in probability} \]
which implies that
\[ e^{-(b-1)\gamma(n)} \sim \frac{W(\infty)}{(b-1)n} \quad \text{in probability}. \]

Then we conclude that
\[ \lim_{n \to \infty} \frac{Y_0^{(p)}(\gamma(n))}{n} = (b-1)e^{-\beta c} \quad \text{in probability}. \]
and the result follows readily from (17), the above limit and the fact that \( 1 - p = o(1) \).

Therefore, from the identity
\[ C_0^{(p)} = \frac{Y_0^{(p)}(\gamma(n)) - 1 + H_0^{(p)}(\gamma(n))}{(b-1)}, \]
Theorem 2 applies to \( Y_0^{(p)}(\gamma(n)) \) and Lemma 8 yields the result of Theorem 1.

4 Complements

We conclude this work by showing that the approach developed here can be applied for instance, to study the analogous problem for scale-free random trees. Also, we briefly give the result about the next largest percolation clusters in the case of \( b \)-ary recursive trees.
4.1 Scale-free random trees

The scale-free random trees are a family of random trees that grow following a preferential attachment algorithm. Typically, fix a parameter $a \in (-1, \infty)$, and start for $n = 1$ from the tree $T_1^{(a)}$ on $\{0, 1\}$ which has a single edge connecting 0 and 1. Suppose that $T_n^{(a)}$ has been constructed for some $n \geq 1$, and for every $i \in \{0, \ldots, n\}$, denote by $d_n(i)$ the degree of the vertex $i$ in $T_n^{(a)}$. Then conditionally given $T_n^{(a)}$, the tree $T_{n+1}^{(a)}$ is built by adding an edge between the new vertex $n+1$ and a vertex $v_n$ in $T_n^{(a)}$ chosen at random according to the law

$$
\mathbb{P}(v_n = i|T_n^{(a)}) = \frac{d_n(i) + a}{2n + a(n + 1)}, \quad i \in \{0, \ldots, n\}.
$$

Clearly, the preceding expression defines a probability measure since the sum of the degrees of a tree with $n + 1$ vertices equals $2n$. Note also that when one let $a \to \infty$ the algorithm yields an uniform recursive tree since $v_n$ becomes uniformly distributed on $\{0, \ldots, n\}$.

Following Bertoin and Uribe Bravo [5], we consider a continuous time version of the growth algorithm with preferential attachment which, as the case of the $b$-ary recursive trees, we can interpret in terms of a Yule process in continuous space. More precisely, we start at 0 from the algorithm with preferential attachment which, as the case of the $b$-ary recursive trees, we can

$Y$ is a Yule process in continuous state starting at $0:\mathbb{R}$, i.e., its number of vertices. It should be plain that if we define

$$
\gamma^{(a)}(n) = \inf\{t \geq 0 : |T^{(a)}(t)| = n + 1\},
$$

then $T^{(a)}(\gamma^{(a)}(n))$ is a version of a scale-free tree of size $n + 1$, $T_n^{(a)}$. Furthermore, the process $Y^{(a)}$ defined by

$$
Y^{(a)}(t) = (2 + a)|T^{(a)}(t)| - 2, \quad t \geq 0,
$$

is a Yule process in continuous state starting at $Y^{(a)}(0) = 2a + 2$, that has only jumps of size $2 + a$, and with unit birth rate per unit population size.

Then, incorporating Bernoulli bond percolation to the algorithm as we have done here in the case of the $b$-ary recursive tree, yields similar systems of branching processes with rare mutations introduced in Section 2; see Section 3 in [5] for details. Therefore, one can easily adapt the approach developed here to study the fluctuations of the percolation cluster containing the root in the case of scale-free trees. Specifically, set $\alpha = (1 + a)/(2 + a)$ and denote by $\Gamma^{(a)}$ the cluster containing the root after performing percolation of parameter $p_n$ given by (2) on a scale-free tree of size $n$. We state the following analog of Theorem 1.

**Theorem 3.** Set $\alpha = (1 + a)/(2 + a)$, and assume that the percolation parameter $p_n$ is given by (4). Then as $n \to \infty$, there is the weak convergence

$$
(n^{-1}\Gamma^{(a)} - e^{-\alpha c})\ln n - \alpha ce^{-\alpha c}\ln \ln n \Rightarrow -\alpha ce^{-\alpha c}Z_c',
$$

where $Z_c'$ is a standard exponential random variable.
where
\[ Z'_{c,\alpha} = Z - \kappa'_\alpha + \ln (\alpha c) \]
with \( Z \) the continuous Luria-Delbrück distribution,
\[ \kappa'_\alpha = \frac{1}{\alpha} - 1 + \frac{1}{\alpha} \sum_{k=2}^{\infty} \frac{(\alpha)_k (-1)^k}{k! (k-1)} \]
and \((\cdot)_k\) is the Pochhammer function.

Informally, we have interpreted the construction of random \( b \)-ary recursive trees and scale-free random trees as a population model and percolation as rare mutations. In this setting, for certain population models with rare mutations, it has been argued that the number of mutants (or equivalently the number of vertices disconnects from the root) might have a distribution close to the Luria-Delbrück law (see Kemp [13]), then the appearance of the continuous Luria-Delbrück variable \( Z \) is not surprising. However, it is important to point out that the constant \( \kappa_\beta \) of Theorem 1 and \( \kappa'_\alpha \) in Theorem 3 are different, and in fact their nature depend of the underlying population model; see Proposition 1. In the case of the random recursive trees this constant is simply zero.

4.2 The next largest clusters

Finally, we by briefly turning our attention to the size of the percolation clusters which do not contain the root. We write
\[ C_1^{(p)} \geq C_2^{(p)} \geq \cdots \]
for the ordered sequence of the sizes of the clusters.

Proposition 2. Set \( \beta = b/(b-1) \), and assume that the percolation parameter \( p_n \) is given by [1]. Then for every fixed integer \( i \geq 1 \),
\[ \left( \frac{\ln n}{n} C_1^{(p)}, \ldots, \frac{\ln n}{n} C_i^{(p)} \right) \]
converges in distribution towards
\[ (x_1, \ldots, x_i) \]
where \( x_1 > x_2 > \cdots \) denotes the sequence of the atoms of a Poisson random measure on \((0, \infty)\) with intensity
\[ \beta ce^{-\beta c}x^{-2}dx. \]
Equivalenty, \( 1/x_1, 1/x_2 - 1/x_1, \ldots, 1/x_j - 1/x_{j-1} \) are i.i.d exponential variables with parameter \( \beta ce^{-\beta c} \).
This result is a directly consequence of the approach developed by Bertoin and Uribe Bravo about the asymptotic behavior of branching process subject to rare mutations. Precisely, we recall the continuous version of Bernoulli bond percolation on \( b \)-ary recursive trees introduced in Section 3. Recall also the notation \( T_p(t) \) for the resulting combinatorial structure at time \( t \), after superpose Bernoulli bond percolation. Then, let \( (T_i(p)(t) : i \geq 0) \) be the sequence of subtrees at time \( t \), where the enumeration in increasing order corresponds to their birth times, and with the convention that \( T_j(p)(t) = \emptyset \), when the number of edges that have been cut at time \( t \) is less than \( j \). We write \( H_i(p)(t) \) for the number of half-edges pertaining to the \( i \)-th subtree at time \( t \). So that, the number of external vertices in the root cluster is given by

\[
Y_0(p)(t) = (b - 1)|T_0(p)(t)| + 1 - H_0(p)(t),
\]
as we have already seen, and similarly the number of external vertices of the \( i \)-th subtree is

\[
Y_i(p)(t) = (b - 1)|T_i(p)(t)| + 2 - H_i(p)(t), \quad i \geq 1,
\]
for \( t \geq 0 \). Then, we clearly see that

\[
Y(t) = \sum_{i \geq 0} Y_i(p)(t),
\]
where the process \( Y \) is defined in (16), represents the number of external vertices on all the tree. So that, relating the process \( Y_p = (Y_j(p) : j \geq 0) \) with a nice system of branching processes with rare mutations, one can easily follow the steps of the proof of Theorem 1 in [5]. Details are left to the interested reader.

We stress the similarity with the analogous result for the families of random recursive trees (see Theorem 1 in [4]) and scale-free random trees (see Theorem 1 in [5]). Essentially, the sizes of the next largest percolation clusters after a normalization are distributed according to a Poisson random measure with the same intensity upon a factor that depends of the nature of the tree.

Acknowledgements. I would like to thank Jean Bertoin, who introduced me the problem and gave guidance throughout the work.

References

[1] Athreya, K. B. and Ney, P. E. *Branching processes*. Dover Publications, Inc., Mineola, NY, 2004. Reprint of the 1972 original [Springer, New York; MR0373040].

[2] Bertoin, J. Almost giant clusters for percolation on large trees with logarithmic heights. *J. Appl. Probab.*, 50(3):603–611, 2013.

[3] Bertoin, J. On the non-gaussian fluctuations of the giant cluster for percolation on random recursive trees. *Electron. J. Probab.*, 19:no. 24, 1–15, 2014.
[4] Bertoin, J. Sizes of the largest clusters for supercritical percolation on random recursive trees. *Random Structures Algorithms*, 44(1):29–44, 2014.

[5] Bertoin, J. and Uribe Bravo, G. Supercritical percolation on large scale-free random trees. *To appear in Ann. Appl. Probab.*, 2012.

[6] Chauvin, B., Klein, T., Marckert, J.-F., and Rouault, A. Martingales and profile of binary search trees. *Electron. J. Probab.*, 10:no. 12, 420–435, 2005.

[7] de La Fortelle, A. Yule process sample path asymptotics. *Electron. Comm. Probab.*, 11:193–199, 2006.

[8] Drmota, M. *Random trees*. Springer. NewYork, Vienna, 2009.

[9] Goldschmidt, C. and Martin, J. B. Random recursive trees and the Bolthausen-Sznitman coalescent. *Electron. J. Probab.*, 10:no. 21, 718–745, 2005.

[10] Holmgren, C. Random records and cuttings in binary search trees. *Combin. Probab. Comput.*, 19(3):391–424, 2010.

[11] Iksanov, A. and Möhle, M. A probabilistic proof of a weak limit law for the number of cuts needed to isolate the root of a random recursive tree. *Electron. Comm. Probab.*, 12:28–35, 2007.

[12] Javanian, M. and Vahidi-Asl, M. Q. Depth of nodes in random recursive $k$-ary trees. *Inform. Process. Lett.*, 98(3):115–118, 2006.

[13] Kemp, A. W. Comments on the Luria-Delbrück distribution. *J. Appl. Probab.*, 31(3):822–828, 1994.

[14] Mahmoud, H. M. *Evolution of random search trees*. John Wiley, New York, 1992.

[15] Parzen, E. *Stochastic processes*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1999. Reprint of the 1962 original.

[16] Schweinsberg, J. Dynamics of the evolving Bolthausen-Sznitman coalescent. *Electron. J. Probab.*, 17:no. 91, 50, 2012.