

Random recursive trees and the elephant random walk

Rüdiger Kürsten*

Institut für Theoretische Physik, Universität Leipzig, P.O. Box 100 920, D-04009 Leipzig, Germany and International Max Planck Research School Mathematics in the Sciences, Inselstraße 22, D-04103 Leipzig, Germany

(Received 16 December 2015; published 7 March 2016)

One class of random walks with infinite memory, so-called elephant random walks, are simple models describing anomalous diffusion. We present a surprising connection between these models and bond percolation on random recursive trees. We use a coupling between the two models to translate results from elephant random walks to the percolation process. We calculate, besides other quantities, exact expressions for the first and the second moment of the root cluster size and of the number of nodes in child clusters of the first generation. We further introduce another model, the skew elephant random walk, and calculate the first and second moment of this process.

DOI: [10.1103/PhysRevE.93.032111](https://doi.org/10.1103/PhysRevE.93.032111)**I. INTRODUCTION**

Anomalous diffusion appears in many physical, biological, and social systems. Examples are the traveling motion of people or money [1], macromolecules in cytoplasm [2], the motion on membranes [3], diffusion in porous materials [4], and diffusion in polymer networks [5], to name only a few. Theoretical models describing such phenomena often include a fractal structure of space [6,7] or incorporate memory effects [8–11].

The elephant random walk (ERW) introduced in [11] is one of the simplest models leading to anomalous diffusion. Since its appearance the ERW and several modifications of it (see, e.g., Refs. [12–18]) have been the subject of intense research. As we will demonstrate in this paper, there is a surprising connection between ERW-like models and percolation on random recursive trees (RRTs) that seems to be unknown in the literature.

Random recursive trees appear naturally in Yule processes [19]. There, an individual reproduces with a constant rate. When a child is born, it evolves independently from its ancestor and they both reproduce with the same rate. The resulting genealogical tree is a random recursive tree. If the possibility of mutations is taken into account in a simple way, percolation clusters of random recursive trees are obtained naturally as genealogical trees, assuming that with some fixed probability a child forms a new species due to mutation.

Versions of Yule processes have been used to study the distribution of words in prose samples, the distribution of scientists by the number of papers, the distribution of cities by population, and the distribution of income [20] and they share certain features with scientific citation networks [21] and preferential attachment models [22]. In recent years there has been active research on bond percolation on random recursive trees from a mathematical point of view (see, e.g., Refs. [23–25]). There have been also studies that deal with the cutting of RRTs [26–28] and with fires on RRTs [29]. Also fragmentation processes induced by deletion of edges [30] or removal of vertices [31] are closely related to bond or site percolation on RRTs.

In this paper we use another formulation of the ERW model that was first investigated in Ref. [17] without mentioning the one-to-one equivalence [32] to the original model [11]. Furthermore, we develop and analyze a generalization of the ERW, the skew elephant random walk (SERW), and we employ another generalization of the ERW that was introduced and analyzed in [13]. The ERW and alike models are understood quite well and we can translate many results into the framework of percolation on random recursive trees.

The paper is organized as follows. In Sec. II we introduce the ERW model and discuss two ways to reformulate it. In Sec. III we introduce the memory tree and memory forest of the ERW as a stochastic coupling between the ERW and percolation on RRTs. In Sec. IV we discuss in detail the Bernoulli bond percolation on RRTs and calculate exact expressions for expectation values of the root cluster size, of the root cluster size squared, of the number of nodes in child clusters of the k th generation, of the number of nodes in first generation child clusters squared, of the probability of the root to be isolated, and of the number of clusters of size one, using results from the ERW, the SERW, or the process introduced in [13]. In Sec. V we use the results of Sec. IV to obtain the root cluster size in certain limiting cases that have been studied in the literature. We prove the mean square convergence in contrast to the weaker convergence in probability, proved in [23,24], which makes our findings on the root cluster size stronger than those previous results. In Sec. VI we discuss how our findings on percolation on RRTs lead to a deeper understanding of ERW-like models. The model introduced in [13] can be reduced to an ERW with a random effective time, which is the root cluster size of a percolation process on a RRT. The appearance of superdiffusion in ERWs can be explained by a single huge memory component of ERWs, which is also equivalent to a root cluster of the percolation process.

II. ELEPHANT RANDOM WALK

In order to make this paper self-contained we will review the main results of Schütz and Trimper [11] in this section. They introduced the elephant random walk. We will also mention the formulation of Kim [17]. He reformulated the ERW model for some parameter regime and reobtained the same results

*kuersten@itp.uni-leipzig.de

as in [11] with the same methods, but he did not mention the one-to-one correspondence between the two models. We will also introduce a reformulation similar to that used in [17], covering the other parameter regime.

In [11] the ERW is introduced as follows. The elephant starts at time $t = 0$ and position $x_0 = 0$ and moves in discrete time on the integers. We denote the position of the elephant at time t by x_t . In each time step the elephant moves a distance σ_t . Hence, the elephant's position satisfies

$$x_{t+1} = x_t + \sigma_{t+1}, \quad (1)$$

where σ_t are random variables, which can take one of the values ± 1 .

In the first step the elephant may move right to $x_1 = +1$, that is, $\sigma_1 = +1$ with probability $q \in [0, 1]$, or it moves left to $x_1 = -1$, that is, $\sigma_1 = -1$ with probability $1 - q$. At any later time t a random number t' is chosen from $\{1, \dots, t - 1\}$ with uniform probability $1/(t - 1)$. Then

$$\sigma_{t+1} = \begin{cases} \sigma_{t'} & \text{with probability } \tilde{p} \\ -\sigma_{t'} & \text{with probability } 1 - \tilde{p}. \end{cases} \quad (2)$$

That means the elephant chooses some random time t' from the past to remember. With probability \tilde{p} it is doing the same as in the past and with probability $1 - \tilde{p}$ it is doing the opposite.

Using the abbreviations

$$\alpha = 2\tilde{p} - 1, \quad \beta = 2q - 1, \quad (3)$$

the first and the second moment have been calculated in [11] resulting in

$$\langle x_t \rangle = \beta \frac{\Gamma(\alpha + t)}{\Gamma(t)\Gamma(\alpha + 1)} \quad \text{for } t > 1, \quad (4)$$

$$\langle x_t^2 \rangle = \frac{t}{2\alpha - 1} \left[\frac{\Gamma(t + 2\alpha)}{\Gamma(t + 1)\Gamma(2\alpha)} - 1 \right]. \quad (5)$$

Consider $\beta = 0$ for simplicity. Then the first moment is zero $\langle x_t \rangle = 0$ and for large times the mean square displacement is approximately [11]

$$\langle x_t^2 \rangle \approx \begin{cases} \frac{t}{3-4\tilde{p}} & \text{for } \tilde{p} < 3/4 \\ t \ln t & \text{for } \tilde{p} = 3/4 \\ \frac{t^{4\tilde{p}-2}}{(4\tilde{p}-3)\Gamma(4\tilde{p}-2)} & \text{for } \tilde{p} > 3/4. \end{cases} \quad (6)$$

This means the system shows normal diffusion for $\tilde{p} < 3/4$ and superdiffusion for $\tilde{p} > 3/4$. Therefore, the ERW is interesting from a theoretician's point of view as it is one of the simplest models that show anomalous diffusion.

For $\tilde{p} > 1/2$ the model can be reformulated as follows. In each time step starting from the second, the elephant remembers some point in the past. Then it needs to decide whether to do the same as or the opposite it did in the past.

For that purpose we can choose a uniform random number r between zero and one. If $r \leq \tilde{p}$ the elephant does the same as it did in the past and if $r > \tilde{p}$ it does the opposite. We will use the notion that the elephant flips a coin that shows heads in the first case and tails in the second.

Now imagine we choose some parameter p such that $0 < p < \tilde{p}$ and we modify the coin such that it shows heads if $r \leq p$ and tails otherwise. If the coin shows heads it is clear that $r \leq p < \tilde{p}$ and the elephant has to do the same as in the

past. If, on the other hand, the coin shows tails the elephant does not know what to do since, although $r > p$, neither of the possibilities $r \leq \tilde{p}$ or $r > \tilde{p}$ can be ruled out.

In that case the elephant needs a second coin. Assume that if $p < r \leq \tilde{p}$ the second coin will show heads and if $r > \tilde{p}$ it will show tails. Hence the elephant will do the same as in the past when the second coin shows heads and it will do the opposite when the second coin shows tails. Then the total probabilities of doing the same as or opposite to the past are still the same:

$$\begin{aligned} \text{Prob}(\sigma_t = \sigma_{t'}) & \\ &= \text{Prob}(\text{1st coin is heads}) \\ &\quad + \text{Prob}(\text{1st coin is tails})\text{Prob}(\text{2nd coin is heads}) \\ &= p + (1 - p) \frac{\tilde{p} - p}{1 - p} = \tilde{p}, \end{aligned} \quad (7)$$

$$\begin{aligned} \text{Prob}(\sigma_t = -\sigma_{t'}) & \\ &= \text{Prob}(\text{1st coin is tails})\text{Prob}(\text{2nd coin is tails}) \\ &= (1 - p) \frac{1 - \tilde{p}}{1 - p} = 1 - \tilde{p}. \end{aligned} \quad (8)$$

We assumed $\tilde{p} > 1/2$, therefore $2\tilde{p} - 1$ is between zero and one. Hence we are free to choose

$$p = 2\tilde{p} - 1. \quad (9)$$

In this case the second coin shows heads and tails with equal probability $1/2$. Thus once the first coin shows tails, the elephant will do the same as or the opposite it has done in the past with equal probability. From this it follows that it will go right or left with equal probability $1/2$ no matter what it has done in the past. Therefore, the elephant can just as well decide to go to the right if the second coin shows heads and to go to the left if the second coin shows tails.

This is exactly what was done in [17]. There the elephant does the same as it did in the past with probability p and with probability $1 - p$ it chooses at random (with equal probability) to go right or left independent of the past. Naturally, the moments obtained in [17] agree with the ones from the original work [11] when the parameter identification (9) is taken into account. (See [32] to see exactly how the processes can be defined on the same probability space.) We will refer to this formulation of the ERW as the ERW1. In Fig. 1 we illustrate the connection between these two formulations.

If $\tilde{p} = 1/2$ the ERW is just a simple random walk where the elephant goes left or right in each time step with equal probability $1/2$ independent of the past. When $\tilde{p} < 1/2$ we can use a similar two-coin formulation as for the ERW1, but with

$$p = 1 - 2\tilde{p}, \quad (10)$$

with the only difference being that the elephant will do the opposite to what it did in the past when the first coin shows heads. That is, it will do the opposite to what it did in the past with probability p and with probability $1 - p$ it will choose at random whether to go right or left. We refer to this formulation as the ERW2.

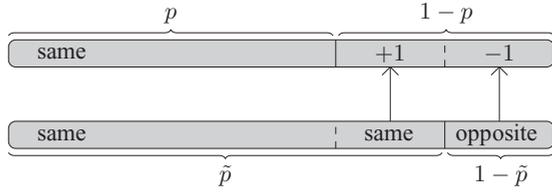


FIG. 1. Illustration of the connection between the formulations ERW and ERW1. The upper bar represents the ERW1 and the lower bar the ERW. There, with probability \tilde{p} the elephant does the same as in the past and with probability $1 - \tilde{p}$ it does the opposite. The event of doing the same is split into two parts, where the first part has probability p and the second part has probability $\tilde{p} - p = 1 - \tilde{p}$. Going to the formulation ERW1, the first part is kept. The second part and the event of doing the opposite have the same probability; considering only these two events, the probability of going right or left is $1/2$ independent of the step that was done in the past. Hence these two events are replaced by the events of going to the right or going to the left in the ERW1, without changing any transition probabilities.

III. MEMORY TREE OF THE ELEPHANT

Consider the formulation ERW1. We are going to draw the memory tree of the elephant. For each time step we will add a node that will be labeled by the number of the time step and a spin. The spin of the node with number t will be σ_t . Hence it is either $+1$ or -1 depending on the direction the elephant moves at time t . For the spin we might for brevity write only $+$ or $-$ instead of $\sigma_t = \pm 1$ and we will justify calling them spins in a moment.

We start by drawing a node for the first step. It will get label 1 and spin $+$ if $\sigma_1 = +1$ and spin $-$ if $\sigma_1 = -1$. We continue to add the second node. As the elephant will remember what it did in the first step, that is, $t' = 1$, we will connect the second with the first node. However, we will delete this edge if the first coin shows tails. In this case the elephant does not really remember but it chooses at random what to do. The second node gets the label 2 and spin $+$ or $-$ depending on the value of σ_2 .

We continue to add nodes for each time step and connect them to the node of the point t' in the past that the elephant is remembering. When the first coin shows tails we delete this edge.

When the elephant made N steps we end up with a graph with N nodes. If we would not have deleted edges the graph would be a tree. With some missing edges it is a forest. We call them the memory tree and the memory forest of the elephant, respectively. In Fig. 2 we show one realization of the ERW1 until time $t = 15$ together with the corresponding memory forest.

There are some interesting observations apparent from the graph. Following some path of the memory tree starting from the root (the node with the label 1), labels are always increasing along the path. That is because we attached nodes with higher labels to previously existing ones with lower labels. The second important observation is that all nodes from one connected component of the memory forest, which we call a cluster, carry the same spin. This has the simple reason that each time

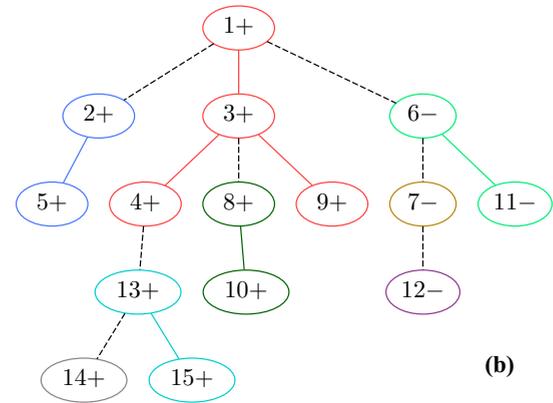
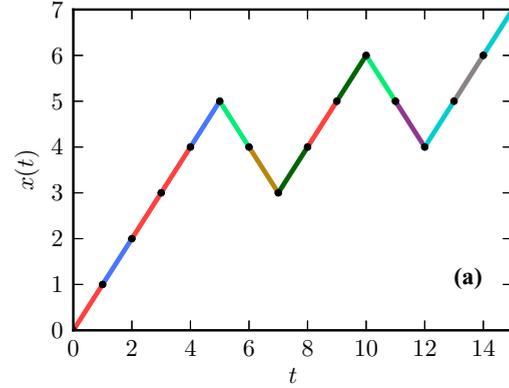


FIG. 2. (a) Trajectory and (b) memory forest of the same realization of the ERW1. Deleted edges are indicated by dashed lines. Time steps, or nodes, that belong to one memory component, that is, one cluster, are drawn with the same color. Here $q = 1/2$ and $p = 0.7$.

an edge was not removed, that is, the first coin showed heads, the elephant did the same as in the past.

At this stage it is interesting to review how randomness comes into the construction of the graph. It enters in three stages. The first stage is the construction of the memory tree. There it is chosen at random to which node the new one should be attached, that is, which time point in the past the elephant chooses to remember. The next stage is the deletion of edges, that is, the tossing of the first coin. The third stage is the flipping of the second coin.

In fact, these three random influences are independent of each other. For instance, it will not affect the result of the first coin flip which point in the past the elephant has chosen to remember. In fact, there is no need to flip the second coin if the first one shows heads. However, it can be flipped anyway without influencing the resulting graph. By the choice of the memory tree and the tossing of the first coin, a forest is constructed and by the flipping of the second coin a spin is assigned to each cluster. It is important to note that the construction of the forest and the assignment of spins are really independent.

We observe furthermore that the position of the elephant after N steps is just the sum of the spins of all nodes. Hence it corresponds to some kind of magnetization of the graph, which is why we used the notion of spins in the first place. Denoting this magnetization of the memory forest of size N

by M_N we have

$$x_{t=N} = M_N = \sum_{i=1}^N \sigma_i. \quad (11)$$

IV. DELETING EDGES OF RANDOM RECURSIVE TREES

Recursive trees are nonplanar trees. That means the order of the child nodes of one node does not matter. Furthermore, a recursive tree of size N is labeled with numbers from one to N , where the node labeled with 1 is called the root of the tree. As already mentioned in the previous section, a recursive tree is defined by the property that all paths starting from the root follow nodes with increasing labels.

We can obtain a RRT of size N in two ways. The first is to choose one of all different recursive trees of size N at random with uniform probability. This is more a definition than a practical way to obtain a random recursive tree. The second procedure is constructive and recursive. In the first step start with the root node. In the following $N - 1$ steps attach the next node (with next highest label) to one of the existing nodes chosen at random with uniform probability.

Random recursive trees have interesting fractal or self-similar properties. If one edge of a RRT is deleted, two independent RRTs emerge. More formally, if we choose a RRT of size N and delete a randomly chosen edge, we obtain a tree that contains the root and another tree. Conditioned on the size of the root-containing tree to be $k < N$, the root-containing tree is a RRT of size k and the other tree is an independent RRT of size $N - k$ [26] (see also [25]).

In percolation theory one investigates usually the clusters of a random graph that is obtained from another graph by deleting either randomly chosen edges (bond percolation) or vertices (site percolation) [33]. Often the original graph is a lattice of dimension d , e.g., \mathbb{Z}^d . Some authors investigated the deletion of vertices [31] or edges [23–25,29,30] of RRTs. If each edge is deleted independently with probability $1 - p$ or kept with probability $p \in [0,1]$, the process is called Bernoulli bond percolation on RRTs. This is exactly what we did to obtain the memory forest of the elephant. We can use the correspondence between the ERW and percolation on RRTs to translate results from one model to the other and vice versa.

We denote the clusters of the memory forest by c_i , where $i = 1, 2, 3, \dots$ gives an order to the clusters. The c_i are chosen such that c_1 denotes the root cluster and $i < j$ if the first node of c_i was attached to the memory tree before the first node of c_j . For technical reasons we might set $c_i = \emptyset$ for all i that are larger than the number of clusters. We further denote the spin that is carried by all nodes of cluster c_i by m_i , that is, $m_i = \pm 1$ or zero if $c_i = \emptyset$. We denote the total magnetization, the sum of all spins, of cluster c_i by \mathcal{M}_i . Since all spins of one cluster are equal, we have $\mathcal{M}_i = m_i |c_i|$, where $|c_i|$ denotes the number of nodes that belong to cluster c_i . For the expectation value of the magnetization of the memory forest we find

$$\begin{aligned} \langle x_{t=N} \rangle &= \langle M_N \rangle = \sum_{i=1}^{\infty} \langle \mathcal{M}_i \rangle = \sum_{i=1}^{\infty} \langle |c_i| \rangle \langle m_i \rangle \\ &= (2q - 1) \langle |c_1| \rangle, \end{aligned} \quad (12)$$

where $\langle \cdot \rangle$ denotes the expectation value over both the construction of the random forest and the assignment of spins. The only contribution comes from the root cluster, since the expectation value of the magnetization of all other clusters is zero, since they carry the spin ± 1 with equal probability $1/2$. Hence we connect the expectation value of the root cluster size with the first moment of the ERW that is known [see Eq. (4)]. This leads to an exact expression for the expectation value of the root cluster size

$$\langle |c_1| \rangle = \frac{\Gamma(N + p)}{\Gamma(p + 1)\Gamma(N)}. \quad (13)$$

We can squeeze out even a little more. We say a cluster is of the k th generation if each path from the root to an arbitrary node of this cluster is interrupted by exactly k deleted edges. We are interested in the number of nodes in clusters of the first generation. Therefore, we modify the coupling between the ERW1 and the percolation on RRTs, that is, we modify the construction of the memory tree. We replace the first coin by a coin that can show heads, tails, or ϵ . We keep the probability $1 - p$ to show tails, the probability to show heads is changed to $p - \epsilon$ for some small parameter $\epsilon > 0$, and the first coin shows ϵ with probability ϵ . The elephant is supposed to remember what it did in the past, when the first coin shows either heads or ϵ such that the movement of the elephant will not be affected by these changes. However, we will change the construction of the memory tree. We delete an edge if the result of the corresponding first coin toss is either tails or ϵ . If the result is heads the edge is kept. We change nothing in the assignment of the spins, that is, the spin of node i will still be σ_i such that the magnetization M_N still corresponds to the position of the elephant $x_{t=N}$. All spins of nodes that belong to one cluster are still equal, but the assignments of spins to the clusters is not independent anymore. More precisely, the spin assigned to a cluster of the first generation will be coupled to the spin of the root cluster with probability $\epsilon/(1 - p + \epsilon)$ and with probability $1 - \epsilon/(1 - p + \epsilon)$ it will be chosen independently. Hence the expectation value of the magnetization is

$$\begin{aligned} \langle M_N \rangle &= (2q - 1) \left[\langle |c_1| \rangle + \frac{\epsilon}{1 - p + \epsilon} \sum_{i, c_i \text{ is 1st generation}} \langle |c_i| \rangle \right. \\ &\quad \left. + \left(\frac{\epsilon}{1 - p + \epsilon} \right)^2 \sum_{j, c_j \text{ is 2nd generation}} \langle |c_j| \rangle + O(\epsilon^3) \right]. \end{aligned} \quad (14)$$

We have deleted each edge with probability $1 - p + \epsilon$ and according to Eq. (13) we have

$$\langle |c_1| \rangle = \frac{\Gamma(N + p - \epsilon)}{\Gamma(p - \epsilon + 1)\Gamma(N)}. \quad (15)$$

Inserting Eqs. (4) and (15) into Eq. (14) and considering the first order in ϵ , that is, subtracting the left-hand side, dividing by ϵ , and taking the limit $\epsilon \rightarrow 0$, we obtain

$$\begin{aligned} &\sum_{i, c_i \text{ is 1st generation}} \langle |c_i| \rangle \\ &= \frac{1 - p}{p} \frac{\Gamma(N + p)}{\Gamma(p)\Gamma(N)} [\Psi_0(N + p) - \Psi_0(p + 1)], \end{aligned} \quad (16)$$

where Ψ_0 denotes the digamma function. Evaluating the terms of order ε^2 in Eq. (14), we can easily obtain an expression for the expectation value of the number of nodes in the second generation child clusters. Analogously, one obtains the expectation value of the number of nodes in the k th generation child clusters for arbitrary $k \geq 1$.

So far we have used the first moment of the ERW to infer properties of clusters of percolation processes on RRTs. In addition, we can use the second moment of the ERW to observe further properties. The second moment corresponds to the expectation value of the square of the magnetization of the memory forest

$$\begin{aligned} \langle x_{t=N}^2 \rangle &= \langle M_N^2 \rangle = \left\langle \sum_{i=1}^{\infty} \mathcal{M}_i^2 \right\rangle + 2 \sum_{j < k} \langle \mathcal{M}_j \mathcal{M}_k \rangle \\ &= \sum_{i=1}^{\infty} \langle m_i^2 |c_i|^2 \rangle + 2 \sum_{j < k} \langle m_j m_k \rangle \langle |c_j| |c_k| \rangle \\ &= \sum_{i=1}^{\infty} \langle |c_i|^2 \rangle = \frac{N}{2p-1} \left(\frac{\Gamma(N+2p)}{\Gamma(N+1)\Gamma(2p)} - 1 \right), \end{aligned} \quad (17)$$

where we used that the spin assignment to the clusters is independent of the cluster structure and of other spins, hence $\langle m_i \rangle = 0$ for all $i > 1$. Therefore, all mixed terms $\langle \mathcal{M}_j \mathcal{M}_k \rangle$ cancel. Furthermore, $m_i^2 = 1$ when $m_i = \pm 1$. In the last line of Eq. (17) we inserted the result for the second moment from Eq. (5) with $\alpha = p$ [cf. Eqs. (3) and (9)]. Thus we found the expectation value of the sum of the cluster sizes squared.

We successfully used the ERW1 to study percolation on RRTs. We can also use the ERW2. The construction of the memory tree and forest remains almost the same apart from differences in the spin assignments. Spins of one cluster are not identical, but spins of neighboring nodes have opposite spins. This is because the elephant always does the opposite when it decides to remember what it did in the past. The spin of the root of each cluster is determined by the toss of the second coin. All other spins of the same cluster follow from the tree structure. In Fig. 3 we show one realization of the ERW2 until time $t = 15$ and the corresponding memory forest.

Given that the root of a cluster c_i has spin $+1$ and the size of the cluster is at least 2, we have $\langle \mathcal{M}_i = 0 \rangle$, which we prove in Appendix A. The same result holds when the root of the cluster has spin -1 . For all clusters except the root cluster, $\langle \mathcal{M}_i = 0 \rangle$ remains true also if the cluster has size 1, since then it carries spin ± 1 with equal probability $1/2$. Hence

$$\begin{aligned} \langle x_{t=N} \rangle &= \langle M_N \rangle = \sum_{i=1}^{\infty} \langle \mathcal{M}_i \rangle = \langle \mathcal{M}_1 \rangle \\ &= \text{Prob}(|c_1| = 1) \langle m_1 \rangle = \text{Prob}(|c_1| = 1) (2q - 1). \end{aligned} \quad (18)$$

Thus we obtain with Eq. (4) the probability that the root is isolated

$$\text{Prob}(|c_1| = 1) = \frac{\Gamma(N-p)}{\Gamma(N)\Gamma(1-p)}. \quad (19)$$

To infer information on the percolation process on RRTs we can consider a modified version of the ERW. As in the ERW, the elephant starts at $x_0 = 0$ and goes right in the first step with probability q and left with probability $1 - q$. In each

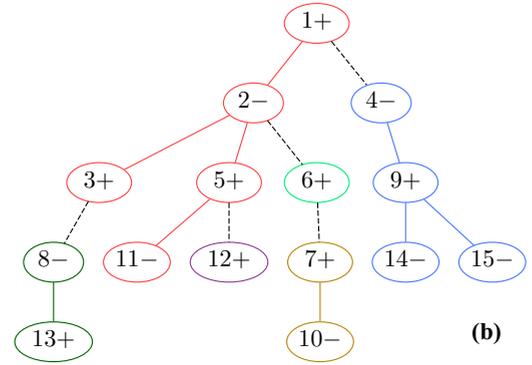
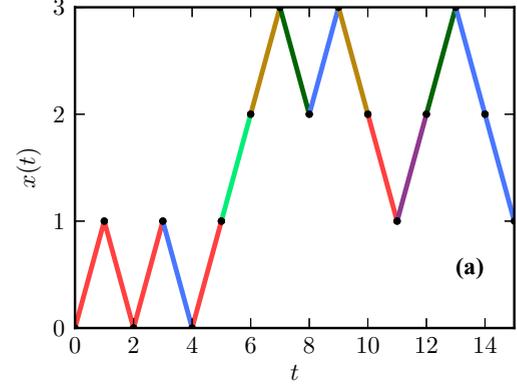


FIG. 3. (a) Trajectory and (b) memory forest of the same realization of the ERW2. Deleted edges are indicated by dashed lines. Time steps, or nodes, that belong to one memory component, that is, one cluster, are drawn with the same color. Here $q = 1/2$ and $p = 0.7$.

following step it chooses a time t' from the past with uniform probability. Then

$$\sigma_{t+1} = \begin{cases} \sigma_{t'} & \text{with probability } \tilde{p} \\ -\sigma_{t'} & \text{with probability } 1 - \tilde{p} - r \\ +1 & \text{with probability } r. \end{cases} \quad (20)$$

That means it either does the same as it did in the past or the opposite, or it goes right, independent on the past. Clearly \tilde{p} , $1 - \tilde{p} - r$, and r must all be probabilities, hence $0 \leq r + \tilde{p} \leq 1$ must be satisfied. This model is only a slight modification of the original ERW that is recovered for $r = 0$. We will refer to this process as the skew elephant random walk (SERW). We can apply the same method as in [11] to calculate the first moment. The probability to go right in step $t + 1$, given the past, is

$$\begin{aligned} \text{Prob}(\sigma_{t+1} = +1 | \sigma_1, \dots, \sigma_t) &= r + \tilde{p} \text{Prob}(\sigma_{t'} = 1 | \sigma_1, \dots, \sigma_t) \\ &+ (1 - \tilde{p} - r) \text{Prob}(\sigma_{t'} = -1 | \sigma_1, \dots, \sigma_t). \end{aligned} \quad (21)$$

Suppose the elephant moves k times to the right and $t - k$ times to the left until time t . Then clearly $x_t = k - (t - k) = 2k - t$ and consequently $k = (x_t + t)/2$. Hence

$$\begin{aligned} \text{Prob}(\sigma_{t'} = +1 | \sigma_1, \dots, \sigma_t) &= \frac{k}{t} = \frac{t + x_t}{2t}, \\ \text{Prob}(\sigma_{t'} = -1 | \sigma_1, \dots, \sigma_t) &= \frac{t - k}{t} = \frac{t - x_t}{2t}. \end{aligned} \quad (22)$$

Inserting this expression into Eq. (21), we find

$$\text{Prob}(\sigma_{t+1} = +1 | \sigma_1, \dots, \sigma_t) = \frac{r+1}{2} + \frac{2\tilde{p} + r - 1}{2} \frac{x_t}{t}. \quad (23)$$

From this expression we can calculate the expectation value of the increment as

$$\begin{aligned} \langle \sigma_{t+1} | \sigma_1, \dots, \sigma_t \rangle &= +1 \text{Prob}(\sigma_{t+1} = +1 | \sigma_1, \dots, \sigma_t) \\ &\quad - 1[1 - \text{Prob}(\sigma_{t+1} = +1 | \sigma_1, \dots, \sigma_t)] \\ &= r + (2\tilde{p} + r - 1) \frac{x_t}{t}. \end{aligned} \quad (24)$$

Averaging over all realizations until time t , we find

$$\langle \sigma_{t+1} \rangle = r + (2\tilde{p} + r - 1) \frac{\langle x_t \rangle}{t}. \quad (25)$$

Taking the expectation value of Eq. (1), we find from this expression

$$\langle x_{t+1} \rangle = r + \left(1 + \frac{\gamma}{t}\right) \langle x_t \rangle, \quad (26)$$

with

$$\gamma = 2\tilde{p} + r - 1. \quad (27)$$

With the initial value $\langle x_1 \rangle = 2q - 1$ the explicit expression for $\langle x_t \rangle$ follows

$$\langle x_t \rangle = \left(2q - 1 - \frac{r}{1 - \gamma}\right) \frac{\Gamma(t + \gamma)}{\Gamma(t)\Gamma(\gamma + 1)} + \frac{rt}{1 - \gamma}. \quad (28)$$

For completeness we will also calculate the second moment. From Eq. (1) we find

$$x_{t+1}^2 = x_t^2 + 2\sigma_{t+1}x_t + \sigma_{t+1}^2. \quad (29)$$

Taking into account that $\sigma_{t+1}^2 = 1$ and using Eq. (24), we find for the second moment

$$\begin{aligned} \langle x_{t+1}^2 \rangle &= 1 + 2r \langle x_t \rangle + \left(1 + \frac{2\gamma}{t}\right) \langle x_t^2 \rangle \\ &= 1 + 2r \left(2q - 1 - \frac{r}{1 - \gamma}\right) \frac{\Gamma(t + \gamma)}{\Gamma(t)\Gamma(\gamma + 1)} \\ &\quad + \frac{2r^2t}{1 - \gamma} + \left(1 + \frac{2\gamma}{t}\right) \langle x_t^2 \rangle, \end{aligned} \quad (30)$$

where we inserted Eq. (28). With the initial condition $x_1^2 = 1$ we find the explicit formula

$$\begin{aligned} \langle x_t^2 \rangle &= \frac{\Gamma(\gamma + t)}{\Gamma(\gamma + 1)\Gamma(t)} [-2r(t + 1)] \left(\frac{2q - 1}{\gamma - 1} + \frac{r}{(\gamma - 1)^2} \right) \\ &\quad + \frac{2\Gamma(2\gamma + t)}{\Gamma(2\gamma + 1)\Gamma(t)} \left(\frac{2r(2q - 1)}{\gamma - 1} + \frac{r^2(3\gamma - 2)}{(\gamma - 1)^2(2\gamma - 1)} \right. \\ &\quad \left. + \frac{\gamma}{2\gamma - 1} \right) + \frac{r^2t^2}{(\gamma - 1)^2} - \frac{t}{2\gamma - 1} \\ &\quad + \frac{r^2t}{(\gamma - 1)^2(2\gamma - 1)}. \end{aligned} \quad (31)$$

Similar to the ERW, we can reformulate also the SERW. For $\tilde{p} = 1 - \tilde{p} - r$ the SERW is just an ordinary random walk where the elephant goes right with probability $\tilde{p} + r$ and left

with probability $1 - \tilde{p} - r$ in all time steps starting from $t = 1$ and independent of the past. For completeness we will also give the SERW1, although we will not use it further. This is the case when $\tilde{p} > 1 - \tilde{p} - r$. In the SERW1 the first step is the same as before and for any later time

$$\sigma_{t+1} = \begin{cases} \sigma_{t'} & \text{with probability } p \\ +1 & \text{with probability } 1 - \tilde{p} \\ -1 & \text{with probability } 1 - \tilde{p} - r. \end{cases} \quad (32)$$

This formulation is equivalent to the SERW when

$$p = 2\tilde{p} + r - 1 = \gamma. \quad (33)$$

The last case, which we call the SERW2, appears for $\tilde{p} < 1 - \tilde{p} - r$ and differs from the previous ones only in the increment, which now reads

$$\sigma_{t+1} = \begin{cases} -\sigma_{t'} & \text{with probability } p \\ +1 & \text{with probability } \tilde{p} + r \\ -1 & \text{with probability } \tilde{p}. \end{cases} \quad (34)$$

This formulation is equivalent to the SERW when

$$p = -2\tilde{p} - r + 1 = -\gamma. \quad (35)$$

Considering the SERW2, we can infer information on the number of clusters of size 1 from its respective memory tree. For each time step in which the elephant is remembering and doing the opposite of the past, the corresponding edge is kept and in the case in which the elephant goes right or left, the corresponding edge is removed. Hence vertices that are connected by an edge carry opposite spin and the spins of vertices that do not belong to the same cluster are independent of each other. As we have argued before, all clusters are random recursive trees themselves and if their size is larger than 1, the expected magnetization is zero independent on the spin of the root (see Appendix A). The expectation value of the spin of the root of each child cluster is $r/(2\tilde{p} + r) = r/(1 - p)$. Hence we find the expectation of the magnetization

$$\begin{aligned} \langle x_{t=N} \rangle &= \langle M_N \rangle = \sum_{i=1}^{\infty} \langle \mathcal{M}_i \rangle = (2q - 1) \text{Prob}(|c_1| = 1) \\ &\quad + \sum_{i=2}^{\infty} \frac{r}{1 - p} \text{Prob}(|c_i| = 1) \\ &= (2q - 1) \frac{\Gamma(N - p)}{\Gamma(N)\Gamma(1 - p)} \\ &\quad + \frac{r}{1 - p} \langle \text{No. of } i > 1 : |c_i| = 1 \rangle \\ &= \left(2q - 1 - \frac{r}{p + 1}\right) \frac{\Gamma(N - p)}{\Gamma(N)\Gamma(1 - p)} + \frac{rN}{1 + p}, \end{aligned} \quad (36)$$

where we have plugged Eq. (19) into the first line and the last line comes from Eq. (28) with $\gamma = -p$ [cf. Eq. (35)]. From this expression we obtain the expectation value of the number of clusters different from the root cluster of size 1,

$$\langle \text{No. of } i > 1 : |c_i| = 1 \rangle = \frac{1 - p}{1 + p} \left[\frac{-\Gamma(N - p)}{\Gamma(N)\Gamma(1 - p)} + N \right]. \quad (37)$$

Adding the probability that the root cluster is of size 1 [see Eq. (19)], we find the expectation value of the number of clusters that are of size 1,

$$\begin{aligned} & \langle \text{No. of } i \geq 1 : |c_i| = 1 \rangle \\ &= \frac{2p}{1+p} \frac{\Gamma(N-p)}{\Gamma(N)\Gamma(1-p)} + N \frac{1-p}{1+p}. \end{aligned} \quad (38)$$

In the following we use a process introduced in [13] that we call the lazy elephant random walk (LERW) due to the fact that in this model the elephant is allowed to sleep, which means that it is allowed to stay at the same position for one time step. In the first step the elephant still moves right with probability q and left with probability $1-q$. The position after all proceeding steps is determined by Eq. (1) with

$$\sigma_{t+1} = \begin{cases} \sigma_{t'} & \text{with probability } p \\ -\sigma_{t'} & \text{with probability } 1-p-s \\ 0 & \text{with probability } s, \end{cases} \quad (39)$$

where t' is randomly chosen from $\{1, \dots, t\}$ and the system parameters s, p must be chosen in such a way that all appearing probabilities are between zero and one.

The first and second moments that have been calculated in [13] are

$$\begin{aligned} \langle x_t \rangle &= (2q-1) \frac{\Gamma(t+2p+s-1)}{\Gamma(t)\Gamma(2p+s)}, \quad (40) \\ \langle x_t^2 \rangle &= \frac{1}{4p+3s-3} \left(\frac{\Gamma(4p+2s+t-2)}{\Gamma(4p+2s-2)\Gamma(t)} - \frac{\Gamma(1-s+t)}{\Gamma(1-s)\Gamma(t)} \right). \end{aligned} \quad (41)$$

For large t the leading behavior is

$$\langle x_t^2 \rangle \approx \begin{cases} \frac{t^{1-s}}{(3-4p-3s)\Gamma(1-s)} & \text{for } 4p+3s-3 < 0 \\ \frac{1}{\Gamma(1-s)} t^{1-s} \ln(t) & \text{for } 4p+3s-3 = 0 \\ \frac{t^{4p+2s-2}}{(4p+3s-3)\Gamma(4p+2s-2)} & \text{for } 4p+3s-3 > 0. \end{cases} \quad (42)$$

Depending on the choice of parameters, we can obtain all exponents between 0 and 2. Thus the model can show subdiffusion as well as superdiffusion. Note that Eq. (6) is recovered for $s=0$.

To obtain the second moment of the root cluster size we set the parameter $s=1-p$. That means the elephant either repeats what it did in the past or sleeps. We can construct again the memory forest of the elephant by keeping the edges whenever the elephant decides to do the same as in the past and by deleting the corresponding edge whenever the elephant decides to sleep. By this construction the spins of the root cluster are all either $+1$ or -1 and the spins of all other clusters are zero. Hence we find for the square of the magnetization

$$\langle x_{t=N}^2 \rangle = \langle M_N^2 \rangle = \langle |c_1|^2 \rangle, \quad (43)$$

and with Eq. (41) we find an exact expression for the second moment of the root cluster size

$$\langle |c_1|^2 \rangle = \frac{1}{p} \left(\frac{\Gamma(2p+N)}{\Gamma(2p)\Gamma(N)} - \frac{\Gamma(p+N)}{\Gamma(p)\Gamma(N)} \right). \quad (44)$$

Using certain ε modifications of the coupling between the LERW and percolation on RRTs similar to the one used before,

one can obtain correlations of the root cluster size with the number of nodes in first generation child clusters

$$\begin{aligned} & \langle |c_1| \langle \text{No. of nodes in 1st generation child clusters} \rangle \rangle \\ &= \frac{1-p}{p} \left[\frac{\Gamma(2p+N)}{\Gamma(2p)\Gamma(N)} \left(-\frac{1}{p} - \Psi_0(2p) + \Psi_0(2p+N) \right) \right. \\ & \quad \left. + \frac{1}{p} \frac{\Gamma(p+N)}{\Gamma(p)\Gamma(N)} \right], \end{aligned} \quad (45)$$

as well as the expectation value of the number of nodes in first generation child clusters squared

$$\begin{aligned} & \langle (\text{No. of nodes in 1st generation child clusters})^2 \rangle \\ &= \frac{(1-p)^2}{p^2} \left[\frac{\Gamma(2p+N)}{\Gamma(2p)\Gamma(N)} \left(\frac{8}{p} + \frac{1}{1-p} + 6\Psi_0(2p) \right. \right. \\ & \quad - 6\Psi_0(2p+N) + 2p[\Psi_0^2(2p) + \Psi_0^2(2p+N)] \\ & \quad \left. - 2\Psi_0(2p)\Psi_0(2p+N) - \Psi_1(2p) + \Psi_1(2p+N) \right) \left. \right. \\ & \quad \left. + \frac{\Gamma(p+N)}{\Gamma(p)\Gamma(N)} \left(-\frac{8}{p} - \frac{1}{1-p} + \frac{p}{1-p} [\Psi_0(p) \right. \right. \\ & \quad \left. \left. - \Psi_0(p+N)] + 2\Psi_0(p) - 2\Psi_0(p+N) \right. \right. \\ & \quad \left. \left. - 2p\Psi_0(p)\Psi_0(p+N) \right) \right], \end{aligned} \quad (46)$$

where Ψ_k denotes the polygamma function. Details of the calculation are given in Appendix B.

V. COMPARISON WITH PREVIOUS RESULTS

The main quantity that was discussed in [23–25] is the size of the root cluster in certain limits, where $N \rightarrow \infty$ and $p \rightarrow 1$. In this regime the root cluster is the largest cluster with high probability. Also the sizes of the next largest clusters were investigated in these works. The result may depend on the precise way the aforementioned limits are assumed. Therefore, we need to state the limit more precisely and there are several ways to do it. One limit that was considered in [23] is

$$N \rightarrow \infty, \quad p(N) = 1 - \lambda / \ln(N). \quad (47)$$

This limit is called supercritical in the literature. The first part of the main result of [23] is that in this limit

$$\frac{1}{N} |c_1| \rightarrow \exp(-\lambda), \quad (48)$$

where the convergence is in probability. We obtained an exact expression for the expectation value of the root cluster size for finite N [see Eq. (13)]. Evaluating the root cluster size divided by N in the limit (47), we find

$$\frac{1}{N} \langle |c_1| \rangle \rightarrow \exp(-\lambda). \quad (49)$$

We can also calculate the variance

$$\begin{aligned} & \lim_{N \rightarrow \infty} \left\langle \left(\frac{1}{N} |c_1| - \exp(-\lambda) \right)^2 \right\rangle \\ &= \lim_{N \rightarrow \infty} \frac{1}{N^2} \langle |c_1|^2 \rangle - \exp(-2\lambda). \end{aligned} \quad (50)$$

As we have also the exact expression of the second moment of the root cluster size [see Eq. (44)] we can calculate the limit

$$\lim_{N \rightarrow \infty} \frac{1}{N^2} \langle |c_1|^2 \rangle = \exp(-2\lambda). \quad (51)$$

Hence the limit (48) is also achieved in mean square convergence, which implies convergence in probability.

The second part of the main result of [23] deals with the sizes of the next largest clusters after the root cluster. It is stated that if C_1, \dots, C_l denote the l next largest clusters, then

$$\left(\frac{\ln N}{N} C_1, \dots, \frac{\ln N}{N} C_l \right) \rightarrow (x_1, \dots, x_l), \quad (52)$$

where the convergence is in distribution in the limit (47) and $x_1 > x_2 > \dots$ denotes the sequence of atoms of a Poisson random measure on $(0, \infty)$ with intensity $\lambda \exp(-\lambda)x^{-2}dx$. This result clearly outruns our results as it gives detailed information over the next largest clusters. However, as it was argued in [24], the next largest clusters are of first generation with high probability. Hence our result (16) on the number of nodes in first generation child clusters is related to it. Considering the limit (47), we find

$$\begin{aligned} \frac{\ln N}{N} \langle \text{No. of nodes in 1st generation child clusters} \rangle \\ \rightarrow \lambda \exp(-\lambda). \end{aligned} \quad (53)$$

Furthermore, we find from Eq. (46) in the limit (47),

$$\begin{aligned} \left(\frac{\ln N}{N} \right)^2 \langle (\text{No. of nodes in 1st generation child clusters})^2 \rangle \\ \rightarrow 2\lambda^2 \exp(-2\lambda), \end{aligned} \quad (54)$$

such that the variance of the scaled quantity $\frac{\ln N}{N} \langle \text{No. of nodes in 1st generation child clusters} \rangle$ converges to $\lambda^2 \exp(-2\lambda)$. That means the scaled number of nodes in first generation child clusters is still fluctuating in the limit (47), which is in accord with [23].

The main results of [24] generalize the findings of [23] as [24] deals with more general limits of the form

$$N \rightarrow \infty, \quad p(N) \rightarrow 1, \quad (55)$$

where (47) is called supercritical, the case $1/\ln N \ll 1 - p(N) \ll 1$ is called weakly supercritical, and the case $0 < 1 - p(N) \ll 1/\ln N$ is called strongly supercritical. The results of [24] deal with the size of clusters of all generations. For the root cluster size the result is

$$\frac{1}{N^{p(N)}} |c_1| \rightarrow 1, \quad (56)$$

where the limit is in distribution. Also in these limits our results (13) and (44) lead to the limit (56) in mean square convergence.

Some authors investigated the cutting of random recursive trees until the root is isolated [26–28]. In this process a randomly chosen edge is removed and afterward only the connected component that contains the root is considered. The removal of edges is repeated until the root is isolated. Our results are not so much connected to this process, however we give the probability that the root is isolated in percolation on

RRTs [see Eq. (19)], which is in some sense a complementary consideration.

VI. PERCOLATION VIEW ON ANOMALOUS DIFFUSION

We are using the analogy between the ERW and percolation on RRTs to attain a deeper understanding of superdiffusion in ERWs. For this purpose the study of the root cluster size is essential. According to Eq. (13), the expectation value of the root cluster size behaves for fixed p and large N as

$$\langle |c_1| \rangle \approx \frac{1}{\Gamma(p+1)} N^p. \quad (57)$$

Considering also the second moment (44), one easily checks that

$$\frac{|c_1|}{N^p} \rightarrow \frac{1}{\Gamma(p+1)} \quad (58)$$

in mean square convergence. Hence, for large N , the leading behavior of the root cluster size is

$$|c_1| \approx \frac{1}{\Gamma(p+1)} N^p. \quad (59)$$

From Eq. (44) we find also the leading behavior of the second moment of the root cluster size for large N ,

$$\langle |c_1|^2 \rangle \approx \frac{1}{p\Gamma(2p)} N^{2p}. \quad (60)$$

We are considering now the ERW in its formulations ERW1 and ERW2. In the ERW2 there is no anomalous diffusion.

In the memory forest of the ERW1 the spins of all nodes belonging to one cluster are equal. According to Eq. (17) we have

$$\langle x_t^2 \rangle = \langle |c_1|^2 \rangle + R, \quad (61)$$

where the remainder is

$$R = \sum_{i=2}^{\infty} \langle |c_i|^2 \rangle. \quad (62)$$

Taking into account the asymptotic second moment of the root cluster size (60) with $t = N$ we find

$$\langle x_t^2 \rangle \approx \frac{1}{p\Gamma(2p)} t^{2p} + R. \quad (63)$$

Hence, for $p > 1/2$ the superdiffusive behavior of the ERW1 can be explained solely by the root cluster size of the memory forest of the ERW1. That means the leading contribution to the second moment comes exclusively from the memory component that contains the first step.

For $p < 1/2$ also the other cluster sizes are important. The root cluster size squared grows sublinearly with the system size with large probability [see Eq. (59)]. Each of the other clusters is likely to be smaller than the root cluster. Thus the leading contribution in (17) does not come from a single cluster, but from all clusters. Since the number of clusters is growing linearly in t with high probability, the second moment also grows linearly in time and hence normal diffusion is observed.

We also want to attain a deeper understanding of subdiffusion in the LERW model [13]. Therefore, we are going to modify the construction of the memory forest once more. Each time the elephant is remembering a time in the past it was not sleeping and it decides to sleep in this step, the corresponding edge is deleted. That means these edges are deleted with probability s . If the elephant is remembering a step in the past when it was sleeping, it will sleep with probability one. However, we will delete the corresponding edge also in this case with probability s . Thus each edge is kept with probability $1 - s$ and deleted with probability s .

We observe that the spins of all nodes that do not belong to the root cluster are zero. These nodes correspond to the steps the elephant is sleeping. On the other hand, the spins of all nodes of the root cluster are nonzero and they are distributed as the steps of ERW with the parameter

$$\tilde{p} = \frac{p}{1-s}. \quad (64)$$

Since the root cluster is an RRT of random size itself we can construct a realization of the LERW at time t as follows. Choose a random time T , which we call the effective time, from a distribution equal to the distribution of the root clusters size for percolation on a RRT of size t and with percolation parameter $1 - s$. Then perform an ERW with the parameter $\tilde{p} = p/(1 - s)$ until time T . This means that the LERW can be seen as an ERW with an effective time T , which is a random variable. For $t \rightarrow \infty$ the leading behavior of T is deterministic, according to Eq. (59),

$$T \approx \frac{1}{\Gamma(2-s)} t^{1-s}. \quad (65)$$

Replacing the time in Eq. (6) by the effective time (65) and taking Eq. (64) into account, we find the leading behavior of the LERW (42).

One can imagine to choose the root cluster not according to the ERW but following some other process. If we choose a ballistic motion, that means all spins of the root cluster are $+1$, with the effective time as the root cluster size, we arrive at the process investigated in [16]. A generalization of the LERW is obtained when the root cluster is chosen according to the SERW.

VII. CONCLUSION

In this paper we revealed a surprising connection between percolation on random recursive trees and the elephant random walk or similar models. We used a stochastic coupling between these models to obtain exact results for the expectation values of certain properties of the percolation process on RRTs. These results allowed us to reobtain certain limiting results on the root cluster size for percolation on RRTs known in the literature. The results of this paper have proved mean square convergence, which implies the previously proved convergence in probability.

All results on percolation on RRTs in this paper rely in principle on recursion relations of expectation values of certain quantities. We want to remark that in principle we do not need to define any random walk to obtain these results. Not

even the assignment of a spin to each node is necessary. However, the calculations become much more intuitive using this formulation and it directly gives an idea how certain quantities can be calculated. Furthermore, it is interesting on its own that percolation on RRTs is related to random walks that show anomalous diffusion.

Reformulations of the ERW model introduced in this paper, as well as the correspondence with RRTs, lead to a deeper understanding of ERW-like models. In particular we gave a demonstrative explanation of subdiffusion present in the LERW by introducing an effective time. We also found that superdiffusion in the ERW is due to one giant memory component only. We further introduced a generalization of the ERW, the skew elephant random walk, for which we calculated the first and the second moment.

The correspondence between ERW-like models and percolation on RRTs gives a powerful tool to solve even more problems that might appear in the study of percolation on RRTs. Furthermore, future work might profit from the equivalence of these models also in the other direction, where one might infer information on ERW-like models from what is known or might be found in the future about percolation on RRTs.

ACKNOWLEDGMENTS

The author thanks the International Max Planck Research School, Mathematics in the Sciences, Leipzig for supporting part of this work with a scholarship. The author thanks Ulrich Behn and Marco Müller for valuable remarks on the manuscript.

APPENDIX A: MAGNETIZATION OF RANDOM RECURSIVE TREES WITH ANTIALIGNED SPINS

We consider RRTs of size N where the root node has spin $\sigma_1 = +1$ and for neighboring nodes i, j the spins have opposite signs $\sigma_i = -\sigma_j$. We prove that the expectation value of the magnetization of the RRT is one for $N = 1$ and zero for $N \geq 2$. The case $N = 1$ is trivial. The case $N = 2$ is also clear since there is only one possible RRT of size 2, for which we have $\sigma_1 = +1$ and $\sigma_2 = -1$. Let us denote the set of all possible RRTs of size N by T_N . Once we have chosen a realization $\tau \in T_N$, then the magnetization $M_N(\tau)$ is already determined. For $N > 2$ we will define a bijective map $f : T_N \rightarrow T_N$ that satisfies

$$M_N(\tau) = -M_N(f(\tau)). \quad (A1)$$

The map f interchanges the child nodes of the first two nodes. That means if in the RRT τ any node $k > 2$ is a child of the root node, it will be a child of the second node in $f(\tau)$ and if any node $k > 2$ is a child of the second node in τ , it will be a child of the root node in $f(\tau)$. All other connections remain unaffected by f . One easily checks that $f(\tau) \in T_N$ and $f(f(\tau)) = \tau$, hence the map is bijective. Since the root node has spin $\sigma_1 = +1$ and the second node has the opposite spin $\sigma_2 = -1$, the property (A1) follows. Let us denote all realizations that lead to a positive, negative, or zero magnetization by $T_{N,+}$, $T_{N,-}$, and $T_{N,0}$, respectively. Then the sets $T_{N,+}$, $T_{N,-}$, $T_{N,0}$ are pairwise

disjoint and their union is

$$T_N = T_{N,+} \cup T_{N,-} \cup T_{N,0}. \quad (\text{A2})$$

The expectation value of the magnetization of the RRT is

$$\begin{aligned} \langle M_N \rangle &= \frac{1}{|T_N|} \sum_{\tau \in T_N} M_N(\tau) \\ &= \frac{1}{|T_N|} \left(\sum_{\tau \in T_{N,+}} M_N(\tau) + \sum_{\tau \in T_{N,-}} M_N(\tau) + \sum_{\tau \in T_{N,0}} M_N(\tau) \right) \\ &= \frac{1}{|T_N|} \left(\sum_{\tau \in T_{N,+}} M_N(\tau) + \sum_{\tau \in T_{N,-}} M_N(\tau) \right) \\ &= \frac{1}{|T_N|} \left(\sum_{\tau \in T_{N,+}} M_N(\tau) + \sum_{\tau \in T_{N,+}} M_N(f(\tau)) \right) \\ &= \frac{1}{|T_N|} \sum_{\tau \in T_{N,+}} [M_N(\tau) + M_N(f(\tau))] = 0, \end{aligned} \quad (\text{A3})$$

where we used that for all $\tau \in T_{N,0}$, $M_N(\tau) = 0$ and that due to the bijectivity of f and the reflection property (A1) we have $T_{N,-} = f(T_{N,+})$. The considerations are of course equivalent if the root has spin $\sigma_1 = -1$. Hence, in this case the expectation value of the magnetization is -1 for $N = 1$ and 0 for $N \geq 2$.

APPENDIX B: SECOND MOMENT OF NUMBER OF NODES IN FIRST GENERATION CHILD CLUSTERS

We are calculating the expectation value of the number of nodes that are in child clusters of the first generation squared. Therefore, we consider the LERW with the parameter $s = 1 - p - \varepsilon$. We further modify the construction of the memory forest a little. Each time the elephant decides to go to sleep or to do the opposite of what it has done in the past the corresponding edge will be deleted. In addition, given that the elephant decides to do the same as it did in the past, the corresponding edge will be deleted as well with probability ε/p . Thus each edge will be deleted with probability $1 - p + \varepsilon$ and it will be kept with probability $p - \varepsilon$. With this setup we find for the magnetization squared

$$\begin{aligned} \langle x_{\tau=N}^2 \rangle &= \langle M_N^2 \rangle \\ &= \langle |c_1|_{p-\varepsilon}^2 \rangle + \frac{2\varepsilon}{1-p+\varepsilon} \sum_{i,c_i \text{ of 1st generation}} \langle |c_i|_{p-\varepsilon}^2 \rangle \\ &\quad + \frac{4\varepsilon^2}{(1-p+\varepsilon)^2} \sum_{i,c_i \text{ of 2nd generation}} \langle |c_i|_{p-\varepsilon}^2 \rangle + O(\varepsilon^3) \\ &= \frac{1}{p-3\varepsilon} \left(\frac{\Gamma(2p-2\varepsilon+N)}{\Gamma(2p-2\varepsilon)\Gamma(N)} - \frac{\Gamma(p+\varepsilon+N)}{\Gamma(p+\varepsilon)\Gamma(N)} \right), \end{aligned} \quad (\text{B1})$$

where the lower index $p - \varepsilon$ indicates that each edge was kept with probability $p - \varepsilon$. We used Eq. (41) in the last line. When plugging Eq. (44) in with the parameter $p - \varepsilon$, we can compare coefficients in different powers of ε . The terms of zeroth order in ε just reproduce (44) for $\varepsilon \rightarrow 0$. Comparing terms linear in

ε we find

$$\begin{aligned} \sum_{i,c_i \text{ of 1st generation}} \langle |c_i|_p^2 \rangle &= \frac{1-p}{p} \left[\frac{1}{p} \frac{\Gamma(2p+N)}{\Gamma(2p)\Gamma(N)} + \frac{\Gamma(p+N)}{\Gamma(p)\Gamma(N)} \right. \\ &\quad \left. \times \left(\Psi_0(p) - \Psi_0(p+N) - \frac{1}{p} \right) \right], \end{aligned} \quad (\text{B2})$$

where Ψ_k denotes the polygamma function.

Replacing p with $p - \varepsilon$ in Eq. (B2) and plugging it in Eq. (B1), we can compare coefficients of ε^2 in Eq. (B1) to find

$$\begin{aligned} \sum_{i,c_i \text{ of 2nd generation}} \langle |c_i|_p^2 \rangle &= \frac{(1-p)^2}{p^2} \left[\frac{1}{p} \frac{\Gamma(2p+N)}{\Gamma(2p)\Gamma(N)} + \frac{\Gamma(p+N)}{\Gamma(p)\Gamma(N)} \right. \\ &\quad \times \left(-\frac{1}{p} + \Psi_0(p) - \Psi_0(p+N) - \frac{p}{2} [\Psi_0^2(p) \right. \\ &\quad \left. + \Psi_0^2(p+N) - 2\Psi_0(p)\Psi_0(p+N) - \Psi_1(p) \right. \\ &\quad \left. \left. + \Psi_1(p+N) \right) \right]. \end{aligned} \quad (\text{B3})$$

Now we consider the original memory forest of the LERW. That means each time the elephant is doing the same as it did in the past, the corresponding edge is kept and otherwise it is deleted. That means it is kept with probability p and deleted with probability $1 - p$. We still consider the sleeping probability $s = 1 - p - \varepsilon$. Then we find for the magnetization squared

$$\begin{aligned} \langle x_i^2 \rangle &= \langle M_N^2 \rangle \\ &= \langle |c_1|^2 \rangle + \frac{\varepsilon}{1-p} \sum_{i,c_i \text{ of 1st generation}} (\langle |c_i|^2 \rangle - 2\langle |c_1||c_i| \rangle) \\ &\quad + \frac{\varepsilon^2}{(1-p)^2} \left(\sum_{i < j, c_i, c_j \text{ of 1st generation}} 2\langle |c_i||c_j| \rangle \right. \\ &\quad \left. + \sum_{i,c_i \text{ of 2nd generation}} \langle |c_i|^2 \rangle \right) \\ &= \frac{1}{p-3\varepsilon} \left(\frac{\Gamma(2p-2\varepsilon+N)}{\Gamma(2p-2\varepsilon)\Gamma(N)} - \frac{\Gamma(p+\varepsilon+N)}{\Gamma(p+\varepsilon)\Gamma(N)} \right). \end{aligned} \quad (\text{B4})$$

Comparing terms of Eq. (B4) that are linear in ε , we find

$$\begin{aligned} \sum_{i,c_i \text{ of 1st generation}} (\langle |c_i|^2 \rangle - 2\langle |c_1||c_i| \rangle) &= \frac{1-p}{p} \left[\frac{\Gamma(2p+N)}{\Gamma(2p)\Gamma(N)} \left(\frac{3}{p} + 2\Psi_0(2p) - 2\Psi_0(2p+N) \right) \right. \\ &\quad \left. + \frac{\Gamma(p+N)}{\Gamma(p)\Gamma(N)} \left(-\frac{3}{p} + \Psi_0(p) - \Psi_0(p+N) \right) \right]. \end{aligned} \quad (\text{B5})$$

From Eqs. (B2) and (B5) we obtain the correlations between the root cluster size and the number of nodes in first generation child clusters (45).

Comparing terms of Eq. (B4) that are quadratic in ε , we find

$$\begin{aligned} & \sum_{i < j, c_i, c_j \text{ of 1st generation}} 2\langle |c_i| |c_j| \rangle + \sum_{i, c_i \text{ of 2nd generation}} \langle |c_i|^2 \rangle \\ &= \frac{(1-p)^2}{p^2} \left[\frac{\Gamma(2p+N)}{\Gamma(2p)\Gamma(N)} \left(\frac{9}{p} + 6\Psi_0(2p) - 6\Psi_0(2p+N) \right. \right. \\ & \quad \left. \left. + 2p[\Psi_0^2(2p) + \Psi_0^2(2p+N) - 2\Psi_0(2p)\Psi_0(2p+N) - \Psi_1(2p) + \Psi_1(2p+N)] \right) + \frac{\Gamma(p+N)}{\Gamma(p)\Gamma(N)} \right. \\ & \quad \left. \times \left(-\frac{9}{p} + 3\Psi_0(p) - 3\Psi_0(p+N) - \frac{p}{2} [\Psi_0^2(p) + \Psi_0^2(p+N) + 2\Psi_0(p)\Psi_0(p+N) - \Psi_1(p) + \Psi_1(p+N)] \right) \right]. \quad (\text{B6}) \end{aligned}$$

Subtracting Eq. (B3) from and adding Eq. (B2) to Eq. (B6), we obtain the expectation value of the number of nodes in first generation child clusters squared. The result is given in Eq. (46). One can continue to calculate correlations and the second moment of the number of nodes in k th generation child clusters, evaluating terms of higher order in ε .

-
- [1] D. Brockmann, L. Hufnagel, and T. Geisel, *Nature (London)* **439**, 462 (2006).
- [2] M. Weiss, M. Elsner, F. Kartberg, and T. Nilsson, *Biophys. J.* **87**, 3518 (2004).
- [3] G. Schütz, H. Schindler, and T. Schmidt, *Biophys. J.* **73**, 1073 (1997).
- [4] D. L. Koch and J. F. Brady, *Phys. Fluids* **31**, 965 (1988).
- [5] I. Y. Wong, M. L. Gardel, D. R. Reichman, E. R. Weeks, M. T. Valentine, A. R. Bausch, and D. A. Weitz, *Phys. Rev. Lett.* **92**, 178101 (2004).
- [6] B. B. Mandelbrot and J. W. Van Ness, *SIAM Rev.* **10**, 422 (1968).
- [7] Y. Gefen, A. Aharony, and S. Alexander, *Phys. Rev. Lett.* **50**, 77 (1983).
- [8] H. Scher and E. W. Montroll, *Phys. Rev. B* **12**, 2455 (1975).
- [9] R. Metzler and J. Klafter, *Phys. Rep.* **339**, 1 (2000).
- [10] R. Metzler and J. Klafter, *J. Phys. A: Math. Gen.* **37**, R161 (2004).
- [11] G. M. Schütz and S. Trimper, *Phys. Rev. E* **70**, 045101 (2004).
- [12] J. C. Cressoni, M. A. A. da Silva, and G. M. Viswanathan, *Phys. Rev. Lett.* **98**, 070603 (2007).
- [13] N. Kumar, U. Harbola, and K. Lindenberg, *Phys. Rev. E* **82**, 021101 (2010).
- [14] J. C. Cressoni, G. M. Viswanathan, A. S. Ferreira, and M. A. A. da Silva, *Phys. Rev. E* **86**, 022103 (2012).
- [15] J. C. Cressoni, G. M. Viswanathan, and M. A. A. da Silva, *J. Phys. A: Math. Theor.* **46**, 505002 (2013).
- [16] U. Harbola, N. Kumar, and K. Lindenberg, *Phys. Rev. E* **90**, 022136 (2014).
- [17] H.-J. Kim, *Phys. Rev. E* **90**, 012103 (2014).
- [18] M. da Silva, G. Viswanathan, and J. Cressoni, *Physica A* **421**, 522 (2015).
- [19] G. U. Yule, *Philos. Trans. R. Soc. London Ser. B* **213**, 21 (1925).
- [20] H. A. Simon, *Biometrika* **42**, 425 (1955).
- [21] Derek J. de Solla Price, *Science* **149**, 510 (1965).
- [22] A.-L. Barabási and R. Albert, *Science* **286**, 509 (1999).
- [23] J. Bertoin, *Random Struct. Algor.* **44**, 29 (2014).
- [24] E. Baur, *Random Struct. Algor.* (2015), doi: 10.1002/rsa.20603.
- [25] E. Baur and J. Bertoin, in *Stochastic Analysis and Applications 2014*, edited by D. Crisan, B. Hambly, and T. Zariwopoulou, Springer Proceedings in Mathematics & Statistics Vol. 100 (Springer, Berlin, 2014), pp. 51–76.
- [26] A. Meir and J. Moon, *Math. Biosci.* **21**, 173 (1974).
- [27] A. Iksanov and M. Möhle, *Electron. Commun. Probab.* **12**, 28 (2007).
- [28] M. Kuba and A. Panholzer, *Online J. Anal. Comb.* **9**, 1 (2014).
- [29] C. Marzouk, *Stoch. Proc. Appl.* **126**, 265 (2016).
- [30] E. Baur and J. Bertoin, *Electron. J. Probab.* **20**, 1 (2015).
- [31] Z. Kalay and E. Ben-Naim, *J. Phys. A: Math. Theor.* **48**, 045001 (2015).
- [32] R. Kürsten, [arXiv:1503.03302](https://arxiv.org/abs/1503.03302).
- [33] G. R. Grimmett, *Percolation*, 2nd ed. (Springer, Berlin, 1999).