Universal tree structures in directed polymers and models of evolving populations

Éric Brunet,* Bernard Derrida,† and Damien Simon‡
Laboratoire de Physique Statistique, École Normale Supérieure, 24, rue Lhomond, 75231 Paris Cedex 05, France
(Received 10 June 2008; published 2 December 2008)

By measuring or calculating coalescence times for several models of coalescence or evolution, with and without selection, we show that the ratios of these coalescence times become universal in the large size limit and we identify a few universality classes.

DOI: 10.1103/PhysRevE.78.061102 PACS number(s): 05.40.−a, 75.50.Lk, 89.75.Hc, 87.23.Kg

Random trees appear in many contexts in biology, mathematics, and physics. In evolutionary biology, they represent the genealogies of reproducing populations. In physics, random trees appear in many systems such as diffusion limited aggregation (DLA) [1], coarsening, river networks [2,3], diagrams in perturbation theory, ultrametric structure of pure states in mean-field spin glasses [4,5], directed polymers in a random medium [6,7], shocks in one-dimensional turbulence [8–10], etc.

From a mathematical point of view, one of the simplest examples of random trees is Kingman's coalescent [11,12]: it describes the coalescence tree of particles, where each pair of particles has a probability δt of coalescing into a single particle during every infinitesimal time interval δt. The random tree structures of Kingman's coalescent are identical to the genealogies obtained in simple mean field models of neutral evolution such as the Wright-Fisher model [13,14]. In such models, each individual of a population of fixed size N at a given generation gives birth to a random number of offspring and the population at the next generation is obtained by choosing N survivors at random among all these offspring. If one follows the evolution over a large enough number of generations for the initial condition to be forgotten, a steady state is reached where the statistics of the genealogical tree of a large population are identical to those of Kingman's coalescent.

Other random trees have been considered in the mathematical literature, such as the Λ coalescents [15–17], which generalize Kingman's coalescent and describe a wider class of mean-field coalescence models [18]. In the Λ coalescent, each subset of k particles among n particles has a probability $\lambda_{n,k}\delta t$ of coalescing into a single particle during an infinitesimal time δt. As a set of n particles can be considered as a subset of a larger set of $n+1$ particles, the rates $\lambda_{n,k}$ have to satisfy some consistency relations: the coalescence of k particles in the subset of size n happens in two cases: either these k particles coalesce in the set of size $n+1$ (rate $\lambda_{n+1,k}$) or they coalesce together with the $(n+1)$-th particle (rate $\lambda_{n+1,k+1}$). Therefore

$$\lambda_{n,k} = \lambda_{n+1,k} + \lambda_{n+1,k+1}. \tag{1}$$

This recursion leads to the following general expression for the coalescence rates [15,16]:

$$\lambda_{n,k} = \int_0^1 x^{k-2}(1-x)^{n-k}\Lambda(x)dx, \tag{2}$$

where Λ is some positive measure on the interval [0,1]. With these notations, Kingman's coalescent corresponds to $\Lambda(x) = \delta(x)$. Another particular case, which has been studied, in the context of spin glasses, is the Bolthausen-Sznitman coalescent [19] for which $\Lambda(x)=1$. Trees in the Kingman's coalescent and in the Bolthausen-Sznitman coalescent have different statistical properties.

In order to compare different models of physical or biological systems which generate random trees and to try to identify universality classes, we consider here simple quantities characteristic of these random tree structures. For a tree with a large number of end points, we define T_p as the distance one has to go up into the tree to find the most recent common ancestor of p given points (see Fig. 1).

For models of evolving populations, the distance T_p is the age of the most recent common ancestor of p individuals chosen at random in the population. In general, it depends both on the generation at which these p individuals live, but also on the choice of the p individuals, even in the limit of very large trees. This double source of fluctuations for the T_p is reminiscent of what happens in mean field spin glasses [5]: as for the overlaps in Parisi's theory, the distribution of the T_p remains broad even when the size of the population becomes very large [5,20].

For a given model, one can try to determine averages $\langle T_p \rangle$ or moments $\langle (T_p)^k \rangle$ of these times T_p (the averages are taken over all the branches of the tree, i.e., over all the population at a given generation, and over all the random trees, i.e., over all the generations in the language of models of evolution). In recent works [21,22], it was noticed that for a large class of mean field models of evolution with selection, the ratios of these average times $\langle T_p \rangle$ take, for a large population, simple universal values indicating that the genealogical trees

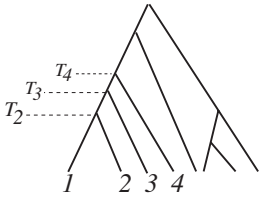

FIG. 1. The times T_p are the ages of the most common ancestors of p individuals chosen at random.

*Eric.Brunet@lps.ens.fr
†Bernard.Derrida@lps.ens.fr
‡Damien.Simon@lps.ens.fr

are distributed according to the statistics of the Bolthausen-Sznitman coalescent. Therefore, at the mean-field level and for a large size of the population, two universality classes seem to emerge for models of evolution: Kingman's trees in the case of neutral evolution for which

$$\frac{\langle T_3\rangle}{\langle T_2\rangle}=\frac{4}{3},\quad \frac{\langle T_4\rangle}{\langle T_2\rangle}=\frac{3}{2},\quad \frac{\langle T_2^2\rangle}{\langle T_2\rangle^2}=2,\quad \frac{\langle T_3^2\rangle}{\langle T_2\rangle^2}=\frac{26}{9}, \tag{3}$$

and Bolthausen-Sznitman's trees in the case of selection

$$\frac{\langle T_3\rangle}{\langle T_2\rangle}=\frac{5}{4},\quad \frac{\langle T_4\rangle}{\langle T_2\rangle}=\frac{25}{18},\quad \frac{\langle T_2^2\rangle}{\langle T_2\rangle^2}=2,\quad \frac{\langle T_3^2\rangle}{\langle T_2\rangle^2}=\frac{11}{4}. \tag{4}$$

The goal of the present work is to try to measure these coalescence ratios for other models of evolution, in particular to analyze the effect of spatial fluctuations, and to argue that directed polymers in a random medium are in the same universality classes as evolution models in presence of selection.

The paper is organized as follows. In Sec. I we consider, at the mean field level or in finite dimension, coalescence models which are equivalent, as we will see, to neutral models of evolution. Above two dimensions of space, the coalescence trees have the same statistics [23] as in mean field with coalescence times given by Eq. (3), whereas in one dimension, they lead to a different universality class for which we compute the ratios of coalescence times. In Sec. II, we consider the trees of optimal paths in the problem of directed polymers in a random medium. Our numerical results will show that at the mean field level, the trees satisfy the Bolthausen-Sznitman statistics Eq. (4), whereas the ratios of coalescence times vary with dimension as expected by the known universality classes of the problem.

I. COALESCENCE AND MODELS OF NEUTRAL EVOLUTION

A. Kingman's coalescent

Kingman's coalescent [11,12] is a mean-field model of coalescing particles: during each infinitesimal time interval δt every pair of particles has a probability of coalescing into a single particle. Therefore if one starts with p particles, there is a random waiting time τ_p until a coalescence event occurs when these p particles become $p-1$ particles. Then there is another random time τ_{p-1} until a pair among these $p-1$ particles coalesce (and one is left with $p-2$ particles), and so on. The times τ_k are independent and distributed according to exponential distributions

$$\rho_k(\tau_k)=\frac{k(k-1)}{2}\exp\left(-\frac{k(k-1)}{2}\tau_k\right) \tag{5}$$

and the time T_p for p particles chosen at random to coalesce is given by

$$T_p=\tau_p+\tau_{p-1}+\cdots+\tau_3+\tau_2. \tag{6}$$

This allows one to easily recover the values of Eq. (3). In fact the whole generating functions of the times T_p can be calculated as

$$\langle e^{\lambda T_p}\rangle=\prod_{k=2}^{p}\frac{k(k-1)}{k(k-1)-2\lambda}. \tag{7}$$

In particular one can notice that the time T_2 has an exponential distribution.

B. Wright-Fisher model

The Wright-Fisher model [13,14] is one of the simplest neutral models of an evolving population. It describes a population of constant size N with nonoverlapping generations and asexual reproduction. At each generation, all the population is replaced by N new individuals with the following rule: each individual at a given generation has its parent randomly chosen among the N individuals at the previous generation. If one goes backward in times, the lineage of an individual performs a random walk on a fully connected graph of N sites. Following the lineages of p individuals is the same as following p coalescing random walks on this fully connected graph. Since the random walks are independent, the statistics of coalescence times can be easily calculated [11,12]: for p individuals chosen at random at generation g, the time T_p is the age of their most recent common ancestor, i.e., T_p is the number of time steps for the p random walkers on the fully connected graph to coalesce. At each generation in the past, two distinct lineages have a probability $1/N$ of merging, thus T_2 scales as the size N of the population. For fixed $p>2$, the probability that a pair of lineages coalesce is $1/N$ whereas multiple coalescences occur with higher powers of $1/N$ for large N. One can then neglect these multiple coalescences and

$$T_p\simeq N(\tau_p+\tau_{p-1}+\cdots+\tau_3+\tau_2), \tag{8}$$

where the times τ_k are distributed according to Eq. (5), implying that the statistics of the times T_p are exactly the same as in Kingman's coalescent Eq. (3).

C. Coalescing random walks in finite dimension

We are now going to look at coalescing random walks on an hypercube of $N=L^d$ sites in dimension d with periodic boundary conditions. We consider the continuous time case, where during infinitesimal time interval δt, each walker on the hypercube has a probability δt of hopping to each of its neighboring sites, and whenever two walkers occupy the same site, they instantaneously coalesce into a single walker. If $T_2(\vec{r})$ is the coalescence time between two walkers at a distance $\vec{r}$ apart, its evolution is

$$T_2(\vec{r})=\begin{cases}\delta t+T_2(\vec{r}) & \text{with probability } 1-4d\,\delta t,\\ \delta t+T_2(\vec{r}+\vec{e}_i) & \text{with probability } 2\,\delta t,\end{cases} \tag{9}$$

where $\vec{e}_i$ is one of the $2d$ unit vectors on the hypercubic lattice.

It is clear that the distance between the two walkers performs a random walk and that T_2 is simply the first time that this distance vanishes. This is of course a very well known first passage problem [24,25] which can be solved easily (it

reduces to the inversion of a Laplacian): the generating function of $T_2(\vec{r})$ satisfies for $\delta t \ll 1$ and for $\vec{r} \neq 0$

$$\langle e^{\lambda T_2(\vec{r})} \rangle = e^{\lambda \delta t} \left[(1 - 4d\delta t) \langle e^{\lambda T_2(\vec{r})} \rangle + 2\delta t \sum_{i=1}^{2d} \langle e^{\lambda T_2(\vec{r}+\vec{e}_i)} \rangle \right], \tag{10}$$

where $\langle \cdots \rangle$ denotes an average over all the random walks. At $\vec{r}=\vec{0}$, it satisfies the boundary condition

$$\langle e^{\lambda T_2(\vec{0})} \rangle = 1. \tag{11}$$

For $\vec{r} \neq \vec{0}$, one can rewrite Eq. (10) as

$$\lambda \langle e^{\lambda T_2(\vec{r})} \rangle + 2 \sum_{i=1}^{2d} [\langle e^{\lambda T_2(\vec{r}+\vec{e}_i)} \rangle - \langle e^{\lambda T_2(\vec{r})} \rangle] = 0 \tag{12}$$

and this can be easily solved in Fourier space to give

$$\langle e^{\lambda T_2(\vec{r})} \rangle = A(\lambda) \sum_{n_1=0}^{L-1} \cdots \sum_{n_d=0}^{L-1} \frac{\exp \frac{2i\pi \vec{n} \cdot \vec{r}}{L}}{\lambda + 4\Sigma_{i=1}^{d} \left(\cos \frac{2\pi n_i}{L} - 1 \right)}, \tag{13}$$

where the constant $A(\lambda)$ is fixed by the condition of Eq. (11).

Starting with two particles at random positions on the lattice and averaging over these two positions leads to

$$\langle e^{\lambda T_2} \rangle = \left[1 + \sum_{\vec{n} \neq 0} \frac{\lambda}{\lambda + 4\Sigma_{i=1}^{d} \left(\cos \frac{2\pi n_i}{L} - 1 \right)} \right]^{-1}. \tag{14}$$

This implies that

$$\langle T_2 \rangle = \sum_{\vec{n} \neq 0} \frac{1}{4\Sigma_{i=1}^{d} \left(1 - \cos \frac{2\pi n_i}{L} \right)}, \tag{15}$$

$$\langle T_2^2 \rangle = 2 \left[\sum_{\vec{n} \neq 0} \frac{1}{4\Sigma_{i=1}^{d} \left(1 - \cos \frac{2\pi n_i}{L} \right)} \right]^2 + 2 \sum_{\vec{n} \neq 0} \left[\frac{1}{4\Sigma_{i=1}^{d} \left(1 - \cos \frac{2\pi n_i}{L} \right)} \right]^2. \tag{16}$$

For large L, it is well known [24,25] that Eq. (15) gives

$$\langle T_2 \rangle \propto \begin{cases} L^2 & \text{for } d = 1, \\ L^2 \ln L & \text{for } d = 2, \\ L^d & \text{for } d > 2. \end{cases} \tag{17}$$

On the other hand, one can show that the second term in the right-hand side of Eq. (16) grows as L^d in dimension $d>4$ and as L^4 in dimension $d<4$. Therefore for $d \geqslant 2$, the ratio $\langle T_2^2 \rangle / \langle T_2 \rangle^2$ goes to 2 when $L \to \infty$, as in the mean-field case Eq. (3).

In fact it has been proved [23,26] that in $d \geqslant 2$ (and for large L) the whole genealogies of p individuals (averaged over all their positions) are given by the Kingman coalescent, up to the rescaling (17). This implies that the distribution of the time T_2 is exponential and that the moments of all the coalescing times T_p have their mean fied values above the upper critical dimension $d=2$ which is the well known upper critical dimension of coalescing random walks [27].

D. Coalescing random walks in one dimension

In dimension $d<2$, the two terms in the right-hand side of Eq. (16) are comparable, and the ratio $\langle T_2^2 \rangle / \langle T_2 \rangle^2$ no longer converges to 2.

In dimension $d=1$ the calculation of all the moments of the times T_p is rather straightforward. First one can easily solve Eq. (12) for periodic boundary conditions with condition Eq. (11) and one gets

$$\langle e^{\lambda T_2(r)} \rangle = \frac{\left(\frac{4 - \lambda + \sqrt{\lambda^2 - 8\lambda}}{4} \right)^{L/2-r} + \left(\frac{4 - \lambda + \sqrt{\lambda^2 - 8\lambda}}{4} \right)^{r-L/2}}{\left(\frac{4 - \lambda + \sqrt{\lambda^2 - 8\lambda}}{4} \right)^{L/2} + \left(\frac{4 - \lambda + \sqrt{\lambda^2 - 8\lambda}}{4} \right)^{-L/2}}. \tag{18}$$

For large L, this becomes a scaling function of λL^2 and of r/L

$$\langle e^{\lambda T_2(r)} \rangle \simeq \frac{\cos \frac{(L-2r)\sqrt{\lambda}}{2\sqrt{2}}}{\cos \frac{L\sqrt{\lambda}}{2\sqrt{2}}} = \frac{\sin \frac{r\sqrt{\lambda}}{\sqrt{2}} + \sin \frac{(L-r)\sqrt{\lambda}}{\sqrt{2}}}{\sin \frac{L\sqrt{\lambda}}{\sqrt{2}}} \tag{19}$$

and, averaging over r, one gets

$$\langle e^{\lambda T_2} \rangle \simeq \frac{2\sqrt{2}}{L\sqrt{\lambda}} \tan \frac{L\sqrt{\lambda}}{2\sqrt{2}}, \tag{20}$$

which shows that the distribution of T_2 is no longer exponential.

One can write down the equations satisfied by the generating functions of the times T_p. For large L and $\lambda = O(L^{-2})$ the solution is

$$\langle e^{\lambda T_p(r_1,\ldots,r_p)} \rangle \simeq \sum_{k=1}^{p} \frac{\sin(r_k \sqrt{\lambda/2})}{\sin(L\sqrt{\lambda/2})}, \tag{21}$$

where the r_k are the distances between consecutive particles along the ring (one has of course $r_1 + \cdots + r_p = L$). In particular, for $p=2$, $r_1=r$ and $r_2=L-r$, one recovers Eq. (19). Averaging Eq. (21) over all the positions of the p particles on the ring leads to

$$\langle e^{\lambda T_p} \rangle \simeq p(p-1) \int_0^1 dx \frac{\sin(Lx\sqrt{\lambda/2})}{\sin(L\sqrt{\lambda/2})} (1-x)^{p-2}. \tag{22}$$

From Eq. (22) one can then obtain all the moments of $\langle T_p \rangle$. For example, one has

$$\langle T_p \rangle \simeq \frac{(p-1)(p+4)}{12(p+1)(p+2)} L^2 \tag{23}$$

and one can show

$$\frac{\langle T_3 \rangle}{\langle T_2 \rangle} = \frac{7}{5}, \quad \frac{\langle T_4 \rangle}{\langle T_2 \rangle} = \frac{8}{5}, \quad \frac{\langle T_2^2 \rangle}{\langle T_2 \rangle^2} = \frac{12}{5}, \quad \frac{\langle T_3^2 \rangle}{\langle T_2 \rangle^2} = \frac{124}{35}, \tag{24}$$

in contrast with Eqs. (3) and (4).

One could repeat the calculations which lead to Eqs. (15)–(17) and Eq. (24) for models of coalescence on other lattices or with more general jumping rates. As long as the motion of the coalescing particles remains diffusive, one would recover the same values Eq. (3) or Eq. (24) for the statistics of the trees.

E. Neutral evolution in finite dimension

One can try to generalize the Wright-Fisher model to the finite-dimensional case, for example, by considering an hypercube with a finite population of fixed size m on each lattice site, and the case where each individual chooses its parent in the previous generation with a probability p on the same lattice site and with probability $1-p$ on one of the neighboring sites. The study of the genealogies in this case is obviously the same problem as following the coalescences of the lineages which perform random walks on this lattice. Therefore in dimension $d=2$ and above, the trees are given by the statistics Eq. (3) of Kingman's coalescent whereas in dimension $d=1$ they will be in the universality class Eq. (24) of coalescing random walks in one dimension.

II. DIRECTED POLYMERS IN A RANDOM MEDIUM

Directed polymers in a random medium is one of the simplest examples of a strongly disordered system [7,28–30]. It describes directed paths in a random energy landscape. In its zero-temperature version, the problem reduces to finding the optimal path, i.e., the path of minimal energy in this random energy landscape. The optimal paths starting at the same point but arriving at different points give rise to a tree structure, that we try to characterize in this section by measuring the coalescence times T_p.

A directed polymer in dimension $d+1$ is a line extending in one of the directions (traditionnaly called "time," and which we represent as the vertical direction in Fig. 2 and Fig. 3) with some random excursions in the d other transverse directions (see Fig. 2). We consider here directed polymers on a lattice which is infinite in the "time" direction but finite and periodic in the d transverse directions. In each time section, there are $N=L^d$ sites located on a d-dimensional hypercube of linear size L with periodic boundary conditions. Each site in a given time section is connected to $M=2^d$ sites in the previous time section (and it is also connected to M other sites in the next time section). The way each site is connected is shown for dimension $1+1$ in Fig. 2. In higher dimensions, we generalized the lattice of Fig. 2 in the following way: let $\vec{x}=(x_1,x_2,\ldots,x_d)$ be the transverse coordinates of a given site; the x_i are integers at even times and half-integers at odd times, and the $M=2^d$ potential parent sites of $\vec{x}$ have coordinates $(x_1 \pm 1/2, x_2 \pm 1/2, \ldots, x_d \pm 1/2)$ in the previous time section.

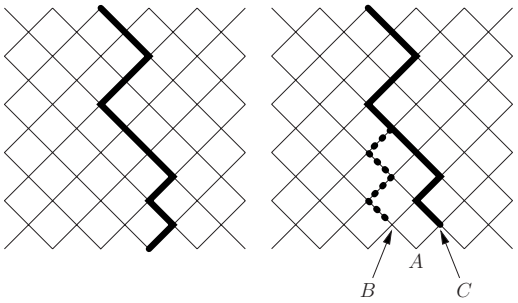

FIG. 2. (Left) a directed polymer in dimension $1+1$. The "time" direction is vertical. (Right) A directed polymer arriving at A comes either from B or from C, whichever is more energetically favorable: In the example shown, the coalescence time of the directed polymers arriving at B and C is 4.

We consider also a mean-field version (Fig. 3), where there is no spatial structure in the transverse directions: A time section consists of a set of N sites, and each of them is connected to M sites chosen at random among the N sites of the previous time section, where M might be any number between 2 and N.

We assume that each link (AB) between two connected sites A and B carries a random energy $\epsilon_{(AB)}$. The energy E of the polymer is then the sum of all the energies $\epsilon_{(AB)}$ of the visited links.

We choose an origin where the polymer starts, and for any given site A on the lattice, we call E_A the minimal energy of the polymer over all the possible directed paths connecting this origin to A. At zero temperature, the directed polymer chooses the path which minimizes its energy and one has the simple recursion relation

$$E_A = \min(E_B + \epsilon_{(AB)}, E_C + \epsilon_{(AC)}, \ldots), \tag{25}$$

where $B, C, \ldots$, are the M potential parent sites of site A.

For any pair of sites A and A' in the same time section, we define their coalescence time (see Fig. 2) as the number of up steps during which the two optimal paths arriving at A and A' differ (we suppose that the origin of the directed polymers is at a remote enough time in the past for the paths to coalesce). In a similar way, we define the coalescence times of any group of p different sites as the maximal coalescence time of any pair within the p sites. All these quan-

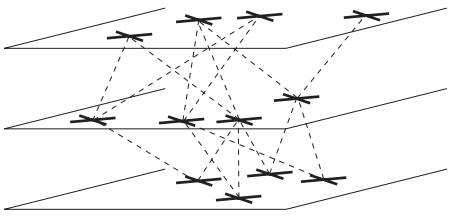

FIG. 3. Directed polymer in mean field. At a given time, each of the $N=4$ sites is connected to $M=2$ random sites at the previous time.

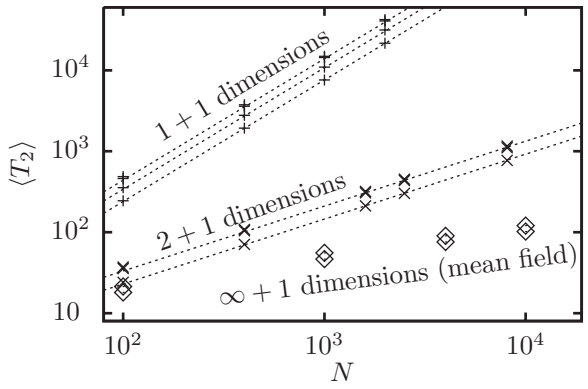

FIG. 4. Averaged coalescence time $\langle T_2 \rangle$ of two individuals for several models of directed polymers in dimensions $1+1$ and $2+1$, and in mean-field, as a function of the number N of sites in each time section. The data are compared to the prediction $\langle T_2 \rangle \propto N^{1/(d\nu)}$ in dotted lines for dimensions $1+1$ and $2+1$. Note that, by chance, two out of the four models in dimension $1+1$ and two out of the three models in dimension $2+1$ have nearly the same prefactor and their data are undistinguishable.

tities depend on the chosen sites and on the realization of the disorder and, as in the previous section, we note by $\langle \cdots \rangle$ the average over the choice of sites and the disorder. In this section, we consider the averaged coalesence time $\langle T_p \rangle$ and the averaged square of the coalesence time $\langle T_p^2 \rangle$ of p sites.

We have simulated four models in dimension $1+1$; from top to bottom on Fig. 4: on the lattice of Fig. 2 with a discrete distribution of ϵ with values $\epsilon=0$ or $\epsilon=1$ with probabilities $1/2$, on the lattice of Fig. 2 with a uniform distribution of ϵ in $[0,1]$, on the lattice of Fig. 2, with negative values of ϵ, distributed according to $\rho(\epsilon)=e^{\epsilon}\theta(-\epsilon)$, on a square lattice where each site is connected to $M=3$ parents (just above itself, on its right and on its left) where ϵ takes positive values, with an exponentially decreasing distribution: $\rho(\epsilon)=e^{-\epsilon}\theta(+\epsilon)$. In dimension $2+1$, we have simulated three models all on the lattice with $M=4$ ancestors described above; from top to bottom on Fig. 4: with an exponentially increasing distribution $\rho(\epsilon)=e^{+\epsilon}\theta(-\epsilon)$, with a uniform distribution of ϵ in $[0,1]$ with an exponentially decreasing distribution $\rho(\epsilon)=e^{-\epsilon}\theta(+\epsilon)$.

Finally, we have simulated two models in mean-field with a uniform distribution of ϵ in $[0,1]$ and either $M=2$ or $M=4$ random ancestors for each site ($M=2$ is above $M=4$ in Fig. 4). Our data for all these models are plotted together with the same symbol for each dimension to emphasize the universality of our results.

To measure the T_p's, the conceptualy simplest way is to update a $N \times N$ matrix containing for all pairs (i,j) of individuals the time $T_2(i,j)$ of their most commun ancestor. Indeed, for an arbitrary number p of individuals, one has $T_p(i_1,\ldots,i_p)=\max[T_2(i_1,i_2),T_2(i_1,i_3),\ldots,T_2(i_1,i_p)]$, so that the matrix of the T_2's contains all the relevant information. Updating this matrix at each time step is easy: The T_2 of two different sites is one plus the T_2 of their parents, and the T_2 of a site with itself is zero. Because updating at each time step a $N \times N$ matrix is time consuming, we used a more sophisticated method [21], where we keep track of the genealogical tree of all the sites at a current time: there are of course N sites at the current time, and at most $N-1$ nodes, where a node is a site from previous times which is the most recent common ancestor of two sites at the current time. At each time step, updating the whole tree takes a time linear in N, and averaging the T_p over all the choices of p individuals takes also a time linear in N, as one simply has to recursively walk down the tree from its root and count for each node the number of times it is the most recent commun ancestor of p sites in the current time. This algorithm is described in more details in Ref. [21].

For each data point, we have run one long simulation and averaged our results over all the time steps once the steady state was reached. This is equivalent to averaging over many independent realizations if we run a simulation for a time much longer than the correlation time, which we estimated to be of the order of magnitude of $\langle T_2 \rangle$. All of our simulations were at least 20 000 times longer than $\langle T_2 \rangle$.

In Fig. 4, we plot the coalescence time $\langle T_2 \rangle$ as a function of the system size. For directed polymers on a lattice which is infinite both in the time direction and in the d transverse directions, the transverse displacement of the optimal path scales as t^{ν}, where t is the length of the directed polymer and ν is a universal exponent [7] equal to $\nu_{1+1}=2/3$ in dimension $1+1$ and $\nu_{2+1} \approx 0.624$ in dimension $2+1$. In our setup, with a finite lattice of linear size L in the transverse directions, this scaling can only hold as long as $t < T_{\mathrm{corr}}$ with $T_{\mathrm{corr}}^{\nu}=L=N^{1/d}$. This time T_{corr} is the correlation time on the scale of which the system forgets its initial condition. Moreover, if we consider several sites and the optimal paths arriving at these sites, these paths coalesce on a time scale of the order of T_{corr}, as can be seen in Fig. 4.

In mean-field with a finite number M of potential ancestors per site, there is no notion of distance in the transverse directions, and the exponent ν is meaningless. We therefore expect a different scaling. The problem of zero-temperature mean-field directed polymers can be formulated [31] as a noisy Fisher-Kolmogorov-Petrovsky-Pisconov (Fisher-KPP) like equation [32,33]. Recently, a phenomenological theory of coalescence trees in models of Fisher-KPP fronts suggested [21,34,35] that the coalescence time in such models should be of order $T_{\mathrm{corr}} \propto (\ln N)^3$. On Fig. 4, one can see that the data seem to have a slower growth than a power law, but the values of N we simulated here are too small to check the $(\ln N)^3$ prediction. Better simulations on a closely related model are presented in Ref. [21], where the $(\ln N)^3$ scaling appears clearly.

We now turn to the ratios of coalescence times. Figure 5 shows the ratios $\langle T_3 \rangle / \langle T_2 \rangle$ and $\langle T_4 \rangle / \langle T_2 \rangle$ as a function of the system size for all the models we study (four models in dimension $1+1$, three in dimension $2+1$, and two in mean-field). Numerically, all the symbols for a given dimension overlap of large N: these ratios seem to depend only on the dimension, and not on the distribution $\rho(\epsilon)$ of the bond energies, nor on the shape of the lattice. The results in mean-field are compatible with the prediction that for an infinitely large system in the Fisher-KPP front equation class [21], the genealogical tree converges to a Bolthausen-Sznitman coalescent, with ratios given by Eq. (4). In dimensions $1+1$ and $2+1$, our numerical results indicate clearly that we have tree

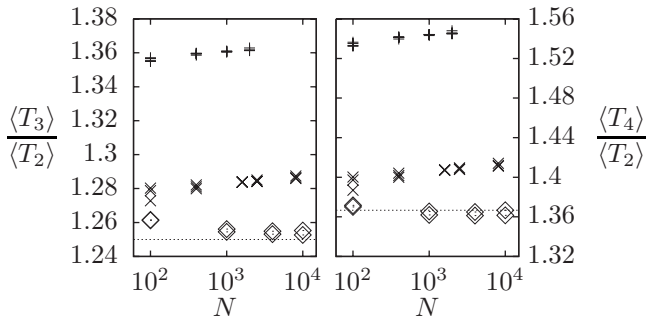

FIG. 5. Ratios of coalescence times for directed polymers at zero temperature as a function of the size N of the system in, from top to bottom, dimension 1+1, dimension 2+1, and meanfield. The dotted line represents the prediction Eq. (4) for meanfield in the limit of infinite size. (Left) ratios $\langle T_3\rangle/\langle T_2\rangle$. (Right) ratios $\langle T_4\rangle/\langle T_2\rangle$.

statistics different from the Bolthausen-Sznitman coalescent, and also different from the Kingman coalescent for which $\langle T_3\rangle/\langle T_2\rangle$ would be 3/2 as in Eq. (3).

On Fig. 6, we show the ratios $\langle T_2^2\rangle/\langle T_2\rangle^2$ and $\langle T_3^2\rangle/\langle T_2\rangle^2$. Here, the situation is less clear: the symbols for the different models do not superpose and the ratios do not seem to have converged [in particular, the mean-field ratios are rather far from the prediction (4)]. For some reason we do not understand, it seems that the $\langle T_p^2\rangle/\langle T_2\rangle^2$ need much larger values of N to converge to their final values than the $\langle T_p\rangle/\langle T_2\rangle$. We already observed a similar phenomenon on an exactly solvable related model [21]. We also measured the ratios $\langle T_N\rangle/\langle T_2\rangle$, where T_N is the age of the most recent common ancestor of the whole population and found these ratios to be close to 1.93 in dimensions 1+1 and 2+1, while it diverges in meanfield [21].

A. Long tail distributions

In the directed polymer problem, it is known that the scaling regime is modified when the distribution $\rho(\epsilon)$ of the energies of the bonds decays as a power law $\rho(\epsilon) \propto |\epsilon|^{-\alpha}$ for large negative ϵ: when $\alpha < \alpha_c$ with $\alpha_c \simeq 7$, the directed polymer in dimension 1+1 has an anomalous scaling [7,36] and the exponent ν depends on α. We have measured the coalescence times in dimension 1+1 for a distribution of energies given by

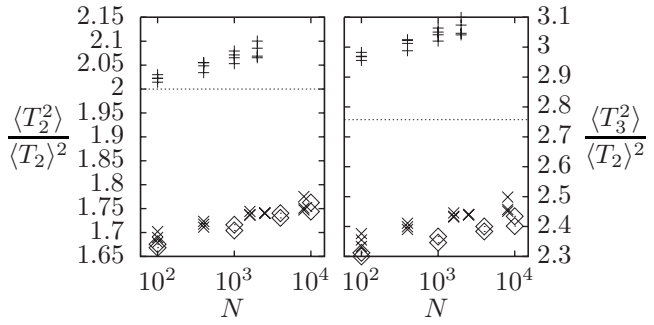

FIG. 6. Ratios of moments of the coalescence times for directed polymers at zero temperature as a function of the size N of the system in, from top to bottom, dimension 1+1, dimension 2+1, and meanfield. The dotted line represents the prediction Eq. (4) for meanfield in the limit of infinite size. (Left) ratios $\langle T_2^2\rangle/\langle T_2\rangle^2$. (Right) ratios $\langle T_3^2\rangle/\langle T_2\rangle^2$.

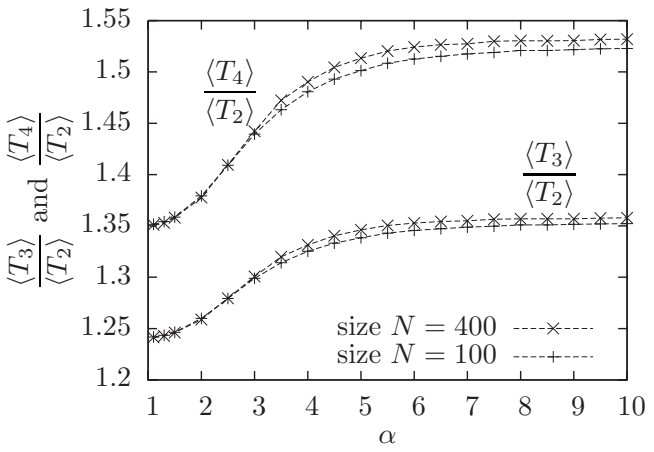

FIG. 7. Ratios $\langle T_3\rangle/\langle T_2\rangle$ and $\langle T_4\rangle/\langle T_2\rangle$ as a function of the exponent α appearing in the noise Eq. (26), for two different system sizes in dimension 1+1.

$$\rho(\epsilon) = \frac{A(\alpha)}{(1+|\epsilon|)^{\alpha}}, \tag{26}$$

with $\alpha > 1$ and $A(\alpha)$ such that $\rho(\epsilon)$ is normalized, for sizes $N=100$ and $N=400$. The ratios $\langle T_3\rangle/\langle T_2\rangle$ and $\langle T_4\rangle/\langle T_2\rangle$ are presented in Fig. 7. We observe that, for large α, these ratios converge towards the universal values shown on Fig. 5, while for $\alpha \to 1^+$, they seem to converge close to, respectively, 1.24 and 1.35.

As we expect $\langle T_2\rangle$ to scale like $N^{1/\nu(\alpha)}$, it is possible to obtain a rough estimate of the exponent $\nu(\alpha)$ from the only two datapoints at sizes $N=100$ and $N=400$. This estimate is shown in Fig. 8.

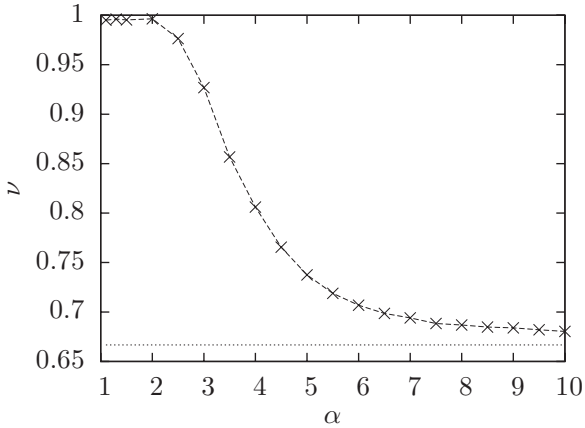

FIG. 8. Estimate of the exponent ν of the directed polymer as a function of the exponent α appearing in the distribution of ϵ in Eq. (26). This exponent has been evaluated from the formula $\ln(4)/\ln[\langle T_2(N=400)\rangle/\langle T_2(N=100)\rangle]$. The universal value $\nu_{1+1} = 2/3$ for distributions decaying fast enough is also shown.

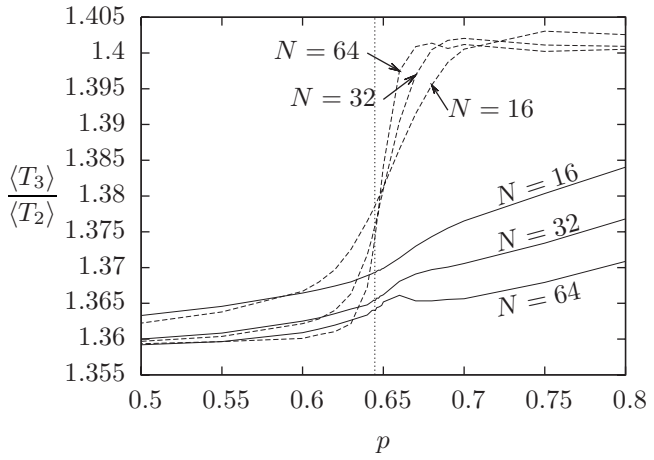

FIG. 9. Ratios $\langle T_3\rangle/\langle T_2\rangle$ as a function of p for the distribution of ϵ of Eq. (27) in dimension 1+1. The dashed lines correspond to the simplest procedure of choosing with equal probabilities one of the potential parent sites realizing the minimal energy, and the plain lines represent the results using the weights Ω which corresponds to the $T\to 0^+$ limit of finite temperature directed polymers. The vertical dotted line indicates the directed percolation threshold on the same lattice.

The exponent $\nu(\alpha)$ seems to converge toward the universal value $\nu_{1+1}=2/3$ for large α, while it seems to be 1 for $1<\alpha\lesssim 2$. As with previous numerical studies [7], our results are not precise enough to determine accurately the critical α_c above which $\nu=2/3$.

B. Discrete distributions

We are now going to discuss the case where the energies of the bonds take discrete values. In this case, it may happen in Eq. (25) that there are several paths coming from different potential parent sites in the previous time section with the same minimal energy, and the question is, of course, which path should be selected as the parent site.

The simplest idea is to choose randomly at each time step with equal probabilities one of the parent site with the lowest energy. With this procedure, we have run numerical simulations in dimension 1+1 for several sizes with a binary noise for the energies ϵ of the bonds

$$\epsilon = \begin{cases} 0 & \text{with probability } p, \\ 1 & \text{with probability } 1-p, \end{cases} \tag{27}$$

for several values of p. Our results for the ratio $\langle T_3\rangle/\langle T_2\rangle$ as a function of p are shown in Fig. 9 as dotted lines. As p varies, we observe a crossover between two values: for small p, $\langle T_3\rangle/\langle T_2\rangle\approx 1.36$ as for directed polymers in dimension 1+1 when the distribution of energies is continuous (see Fig. 5) and, for large p, $\langle T_3\rangle/\langle T_2\rangle\approx 1.4$ which corresponds to the coalescence of random walks in dimension 1, as in Eq. (24). The crossover between the two regions becomes sharper as L increases, which suggests a phase transition. The critical value of p is very consistent with the known threshold 0.6447 for directed percolation on the same lattice [37]. Thus, the system behaves similar to the neutral model when the $\epsilon=0$ bonds percolate.

Instead of choosing with equal probabilities which bond the polymer follows when they are energetically equivalent, there is an alternative procedure which corresponds to taking the limit $T\to 0^+$ in the problem of directed polymers at a finite temperature T. At finite temperature, we keep track for each site A of the partition function Z_A of a polymer arriving on A. Assuming that the site A has $M=2$ potential parent sites B and C, we have the recursion Eq. (25)

$$Z_A = Z_{A\leftarrow B} + Z_{A\leftarrow C}, \tag{28}$$

where $Z_{A\leftarrow B}=Z_B\exp(-\beta\epsilon_{(AB)})$ is the partition function of a directed polymer arriving on A via the site B and where $\beta=1/T$. The probability that a polymer reaching A comes from B is given by

$$(\text{probability the polymer comes from } B) = \frac{Z_{A\leftarrow B}}{Z_A}. \tag{29}$$

At very low temperature, the partition function is dominated by the lowest energy paths

$$Z \approx \Omega e^{-\beta E}, \tag{30}$$

where E is the minimal energy and Ω the number of ways that this energy E can be obtained, so that Eq. (28) reads, at low temperature,

$$\Omega_A e^{-\beta E_A} \approx \Omega_B e^{-\beta E_{A\leftarrow B}} + \Omega_C e^{-\beta E_{A\leftarrow C}}, \tag{31}$$

where $E_{A\leftarrow B}=E_B+\epsilon_{(AB)}$ is the minimal energy of the path arriving at A through B. If $E_{A\leftarrow B}<E_{A\leftarrow C}$, then the first term in the right-hand side of Eq. (31) dominates and we obtain $E_A=E_{A\leftarrow B}$ and $\Omega_A=\Omega_B$. Furthermore, from Eq. (29), the chosen path comes from B. On the other hand, if $E_{A\leftarrow B}=E_{A\leftarrow C}$, both terms in Eq. (31) have the same order of magnitude and we obtain $E_A=E_{A\leftarrow B}=E_{A\leftarrow C}$ and $\Omega_A=\Omega_B+\Omega_C$. Then, from Eq. (29), the probability that the directed polymer comes from B is Ω_B/Ω_A.

In this way, we not only choose the optimal energy but we also keep track of entropy effects. We have run numerical simulations with the same parameters as above but with this new procedure. The ratios $\langle T_3\rangle/\langle T_2\rangle$ are shown on Fig. 9 in plain lines. For small values of p, both procedures yield the same results. For larger p, however, the difference is striking, and the phase transition seems to have disappeared: on both sides of the percolation threshold the data seem to be in the same universality class (as they converge to ≈ 1.36).

III. CONCLUSION

In this paper, we have presented analytical and numerical results showing the existence of universality classes in the tree structures which appear in several models of evolution and in directed polymers (see Table I for a summary). Without selection, the genealogies of neutral models like the Wright-Fisher model or coalescing random walks are described above the critical dimension $d_c=2$ by the Kingman coalescent. For $d=1$ the universality class is different: we have obtained the distribution Eq. (22) of the ages T_p of the most recent common ancestor of p individuals.

For directed polymers in a random medium at zero temperature, the same coalescence times T_p have been measured

TABLE I. Universal ratios and order of magnitudes of coalescence times for models of evolution with and without selection, and directed polymers in a random medium. We could not reach large enough system sizes to give a reliable numerical prediction for the ratios $\langle T_p^2\rangle/\langle T_p\rangle^2$.

	$\frac{\langle T_3\rangle}{\langle T_2\rangle}$	$\frac{\langle T_4\rangle}{\langle T_2\rangle}$	$\frac{\langle T_N\rangle}{\langle T_2\rangle}$	$\frac{\langle T_2^2\rangle}{\langle T_2\rangle^2}$	$\frac{\langle T_3^2\rangle}{\langle T_2\rangle^2}$	$\langle T_2\rangle$
Kingman coalescent						1
Coalescing random walks in $d>2$	$\frac{4}{3}$	$\frac{3}{2}$	2	2	$\frac{26}{9}$	$\propto N$
Neutral models of evolution in $d>2$						$\propto N$
Neutral model in $d=2$						$\propto N \ln N$
Neutral model in $d=1$	$\frac{7}{5}$	$\frac{8}{5}$	2	$\frac{12}{5}$	$\frac{124}{35}$	N^2
Bolthausen-Sznitman's coalescent						1
Models of evolution with selection (or directed polymers at zero temperature in mean field)	$\frac{5}{4}$	$\frac{25}{18}$	∞	2	$\frac{11}{4}$	$\propto (\ln N)^3$
Models of evolution with selection in dimension 2 (or directed polymers at zero temperature in dimension 2+1)	≈ 1.29	≈ 1.42	≈ 1.93	?	?	$\propto N^{1/(2\nu_{2+1})} \approx N^{0.80}$
Models of evolution with selection in dimension 1 (or directed polymers at zero temperature in dimension 1+1)	≈ 1.36	≈ 1.55	≈ 1.93	?	?	$\propto N^{3/2}$

numerically. In the mean field case, their values are compatible with Bolthausen-Sznitman's coalescent, which is already known to appear in spin glasses [19] and in branching random walks with a selection mechanism keeping the size constant [21,22]. In low dimension (at least $d=1$ and $d=2$), the coalescence times belong to different universality classes. It would be interesting to predict analytically the values of $\langle T_3\rangle/\langle T_2\rangle$ and $\langle T_4\rangle/\langle T_2\rangle$ measured numerically in Fig. 5 for fast decaying distributions of ϵ as well as the ones obtained in Fig. 7 for power-law distributions of ϵ with exponent $\alpha \simeq 1^+$. In the mean-field case, it would also be interesting to know if the replica method can be used in order to determine the coalescence times.

The simulations presented in this paper deal only with directed polymers at $T=0$. Directed polymers exhibit a phase transition for $d>2$ as the temperature increases [38]. We expect the tree statistics to change at T_c from the universality class of directed polymers at zero temperature to the universality class of coalescing random walks.

The construction of the minimal energy path for directed polymers can be related to spatial models in presence of selection. In population dynamics, selection can be taken into account through a parameter, called the fitness or the adaptability, which characterizes the ability of an individual to survive and reproduce [39–43]. Individuals with a higher fitness have a higher probability of having a descendance. This parameter is transmitted from parents to offspring up to fluctuations due to mutations. An analogy can be drawn between the minimal energy of a directed polymer arriving on a site, and minus the fitness of an individual living on a site. In presence of local selection, a spatial model of population could therefore be formulated as follows: on each site there would be one (or a finite number m of individuals); at each generation, each individual would branch into k offspring with mutated fitnesses. These offspring diffuse and, under the effect of selection, only the best (or the m best) individual(s) on each site would be kept. Because of the similarity of such spatial models of population dynamics in presence of selection with the directed polymers, we expect these models to belong to the same universality classes.

We performed preliminary simulations on such a spatial model of evolution with selection in dimension 1+1 with $m=5$ individuals per site. Our results for the ratios $\langle T_3\rangle/\langle T_2\rangle$ and $\langle T_4\rangle/\langle T_2\rangle$ coincide with those of directed polymers.

[1] T. A. Witten and L. M. Sander, Phys. Rev. Lett. **47**, 1400 (1981).
[2] M. Takayasu and H. Takayasu, *Nonequilibrium Statistical Mechanics in One Dimension* (Cambridge University Press, Cambridge, 1997), Chap. 9.
[3] M. Cieplak, A. Giacometti, A. Maritan, A. Rinaldo, I. Rodriguez-Iturbe, and J. Banavar, J. Stat. Phys. **91**, 1 (1998).
[4] E. Bolthausen and A.-S. Sznitman, Commun. Math. Phys. **197**, 247 (1998).
[5] M. Mézard, G. Parisi, and M. Virasoro, *Spin Glass Theory and Beyond* (World Scientific, Singapore, 1987).
[6] M. Kardar, G. Parisi, and Y.-C. Zhang, Phys. Rev. Lett. **56**, 889 (1986).
[7] T. Halpin-Healy and Y.-C. Zhang, Phys. Rep. **254**, 215 (1995).
[8] C. Giraud, Commun. Math. Phys. **1**, 67 (2001).
[9] Z.-S. She, E. Aurell, and U. Frisch, Commun. Math. Phys. **148**, 623 (1992).
[10] Y. G. Sinai, Commun. Math. Phys. **148**, 601 (1992).
[11] J. Kingman, Stochastic Proc. Appl. **13**, 235 (1982).
[12] J. Kingman, J. Appl. Probab. **19A**, 27 (1982).
[13] S. Wright, Genetics **16**, 97 (1931).
[14] R. Fisher, *The Genetical Theory of Natural Selection* (Clarendon Press, Oxford, 1930).
[15] S. Sagitov, J. Appl. Probab. **36**, 1116 (1999).
[16] J. Pitman, Ann. Probab. **27**, 1870 (1999).
[17] J. Schweinsberg, Electron. J. Probab. **5**, 1 (2000).
[18] M. Birkner, J. Blath, M. Capaldo, A. M. Etheridge, M. Möhle, J. Schweinsberg, and A. Wakolbinger, Electron. J. Probab. **10**, 303 (2005).
[19] E. Bolthausen and A.-S. Sznitman, Commun. Math. Phys. **197**, 247 (1998).
[20] B. Derrida and L. Peliti, Bull. Math. Biol. **53**, 355 (1991).
[21] É. Brunet, B. Derrida, A. H. Mueller, and S. Munier, Phys. Rev. E **76**, 041104 (2007).
[22] É. Brunet, B. Derrida, A. H. Mueller, and S. Munier, Europhys. Lett. **76**, 1 (2006).
[23] V. Limic and A. Sturm, Electron. J. Probab. **11**, 363 (2006).
[24] E. W. Montroll, Proc. Symp. Appl. Math. **16**, 193 (1964).
[25] E. W. Montroll and G. H. Weiss, J. Math. Phys. **6**, 167 (1965).
[26] J. T. Cox, Ann. Probab. **17**, 1333 (1989).
[27] L. Peliti, J. Phys. A **19**, L365 (1986).
[28] M. Kardar, Nucl. Phys. B **290**, 582 (1987).
[29] T. Emig and M. Kardar, Nucl. Phys. B **604**, 479 (2001).
[30] M. Kardar and Y.-C. Zhang, Phys. Rev. Lett. **58**, 2087 (1987).
[31] É. Brunet and B. Derrida, Phys. Rev. E **70**, 016106 (2004).
[32] R. A. Fisher, Proc. Annu. Symp. Eugen. Soc. **7**, 355 (1937).
[33] A. Kolmogorov, I. Petrovsky, and N. Piscounov, Bull. Univ. État Moscou A **1**, 1 (1937).
[34] É. Brunet and B. Derrida, Phys. Rev. E **56**, 2597 (1997).
[35] É. Brunet, B. Derrida, A. H. Mueller, and S. Munier, Phys. Rev. E **73**, 056126 (2006).
[36] Y.-C. Zhang, Physica A **170**, 1 (1990).
[37] I. Jensen, J. Phys. A **29**, 7013 (1996).
[38] C. Monthus and T. Garel, Eur. Phys. J. B **53**, 39 (2006).
[39] R. E. Snyder, Ecology **84**, 1333 (2003).
[40] M. Kloster and C. Tang, Phys. Rev. Lett. **92**, 038101 (2004).
[41] M. Kloster, Phys. Rev. Lett. **95**, 168701 (2005).
[42] T. Antal, K. B. Blagoev, S. A. Trugman, and S. Redner, J. Theor. Biol. **248**, 411 (2007).
[43] B. Derrida and D. Simon, Europhys. Lett. **78**, 60006 (2007).