Invasion Percolation and Critical Transient in the Barabási Model of Human Dynamics

A. Gabrielli and G. Caldarelli

SMC, INFM-CNR, Dipartimento di Fisica, University "La Sapienza", Piazzale A. Moro 2, 00185-Rome, Italy
Istituto dei Sistemi Complessi CNR, via dei Taurini 19, 00185-Rome, Italy
(Received 12 February 2007; published 14 May 2007)

We introduce an exact probabilistic description for $L = 2$ of the Barabási model for the dynamics of a list of L tasks. This permits us to study the problem out of the stationary state and to solve explicitly the extremal limit case where a critical behavior for the waiting time distribution is observed. This behavior deviates at any finite time from that of the stationary state. We study also the characteristic relaxation time for finite time deviations from stationarity in all cases showing that it diverges in the extremal limit, confirming that these deviations are important at all time.

Queuing theory is a very important branch of probability with fundamental applications in the study of different human dynamics [1]. Its capability in explaining and modeling complex behaviors of such activities has potentially important economic consequences. An example is the prediction and organization of "queues" in hi-tech communications. Queue stochastic dynamics is traditionally modeled as a homogeneous Poisson processes [2]. As a result, the time interval between different events is predicted to be exponentially distributed. Actually, different experimental analyses [3,4] have shown that for various human activities the distribution of waiting times is better fitted by Pareto-like heavy tail distributions. In order to reproduce such a behavior Barabási [5] has introduced a simple model for a task list with a heavy tail distribution of waiting times in its stationary state and in a particular extremal limit. In this Letter we analyze the same model out of the stationary state introducing an exact step-by-step method of analysis through which we find the exact waiting time distribution in the same extremal limit. We show that in this limit the stationary state is reached so slowly that the whole dynamics is described by the transient.

In the Barabási model the list consists at any time in a constant number L of tasks to execute. Each task has a random priority index x_i ($i = 1, \ldots, L$) extracted by a probability density function (PDF) $\rho(x)$ independently one of each other. The dynamical rule is the following: with probability $0 \leq p \leq 1$ the most urgent task (i.e., the task with the highest priority) is selected, while with complementary probability $(1 - p)$ the selection of the task is done completely at random. The selected task is executed and removed from the list; it is then replaced by a fresh new task with random priority extracted again from $\rho(x)$. For $p = 0$ the selection of the task at each time step is completely random. Therefore, the waiting time distribution for a given task is $P(\tau) = (1/L)(1 - 1/L)^\tau$ and decays exponentially fast to zero with time constant $\tau_0 \simeq L$. For $p = 1$ instead the dynamics is deterministic selecting at each time step the task in the list with the highest priority. Because of this extremal nature, the statistics of the dynamics for $p = 1$ does not depend on $\rho(x)$. It has been

shown [5,6] that the dynamics for $0 < p < 1$ reaches a stationary state characterized by a waiting time distribution $P(\tau)$ which is proportional to τ^{-1} with a p-dependent upper cutoff and amplitude. For $p \rightarrow 1$ the cutoff diverges, but the amplitude vanishes, so that in this limit from one side criticality (i.e., divergence of the characteristic waiting time τ_0) is approached, but from the other $P(\tau)$ loses sense due to the vanishing of its amplitude. From simulations this behavior does not look to depend on L, and the exact analytic solution for the stationary state has been given for $L = 2$ in [6].

Here we propose a different approach to the problem for $L = 2$ able to give a complete description of the task list dynamics also out of the stationary state. In this way we find the exact waiting time distribution for $p = 1$ which is characterized by power law tails, but with a different exponent from that found in [6] for the stationary state. We also show that in this extremal limit the finite time deviation from the trivial stationary state has diverging relaxation time. Hence the waiting time distribution at any time is completely determined by these deviations and differs from the stationary one.

The method we use here is called run time statistics (RTS) [7] and it has been introduced originally to study the invasion percolation (IP) in dimension $d = 2$ [8,9] and related models [10]. For $p = 1$ the present task list model can be exactly mapped into IP in $d = 1$ which can be so formulated: consider a linear chain of throats [see Fig. 1(a)], each of which is given, independently of the others, a random number x_i extracted from the PDF $\rho(x)$ and representing its capillarity (inversely proportional to the diameter). At $t = 0$ the invader fluid occupies only the central throat. At each time step the growth interface is composed by the two nonoccupied throats in contact with the invaded region, and the invaded region grows by occupying the interface throat with the maximal capillarity (i.e., minimal diameter). Consequently the interface is updated by eliminating from it the selected throat and including the next one in the direction of growth. This is exactly equivalent to the task list problem for $p = 1$ and $L = 2$, with the set of the already executed tasks given by the invaded
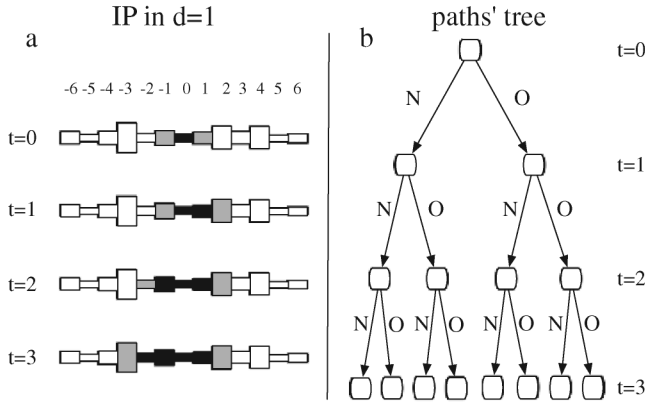
208701-1

IP in d=1

a

paths' tree

b

FIG. 1.   (a) First four time steps of an IP dynamics in $d = 1$. At the initial time step only the central throat is occupied (black). At each time step the occupied region (black) grows by invading the throat of the growth interface (gray) with the maximal capillarity. The growth interface is consequently updated by removing the just occupied throat and including the next unoccupied one in the direction of growth. (b) Tree representation of all the possible selection paths (i.e., realizations) of the task list dynamics or alternatively of IP in $d = 1$. At each time step either the fresh new task (N) either the old one (O) is selected. Any possible dynamical history of length t consists of a continuous sequence of arrows connecting the vertex at height $t = 0$ with a node at height t.

region, and the task list given by the two throats of interface. For $0 < p < 1$ the task list problem would correspond to a sort of IP at finite temperature [11] which introduces a source of noise in the problem permitting random nonextremal growths. The quantity $(1 - p)$ is a measure of such thermal noise which vanishes for $p = 1$ and becomes maximal for $p = 0$. We note also that when the task list size is not fixed but increases in time, the same kind of mapping should apply relating the task list dynamics to IP on a sort of random tree with randomly varying coordination number. This could explain the waiting time exponent $-3/2$ found in [12].

   In general, the basic ingredient to perform averages over the possible realizations of the dynamics is the statistical weight of a growth path (i.e., selection sequence). For growth processes with quenched disorder it is very difficult to factorize this statistical weight into the product of probabilities of the composing elementary steps as these single-step probabilities depend upon all the past history of the process: dynamics in quenched disorder present usually strong memory effects. To overcome this difficulty, RTS gives a step-by-step procedure to write the exact evolution of the probabilities of the single steps conditional to the past history and the related conditional PDF of the random variables attached to the growing elements ("task priorities" in this case). Given two tasks with respective priorities x and y, we call $\eta(x, y)$ the probability conditional to these values to execute the task with priority x. The form of η is determined by the selection rule of the model. Given the definition of the Barabási model and being $\theta(x)$ the

Heaviside step function, we have:

$$\eta(x, y) = p\theta(x - y) + \frac{1 - p}{2}. \tag{1}$$

If we now suppose that at the tth time step x and y are statistically independent and have, respectively, the PDFs conditional to the past history of the process $p(x, t)$ and $q(y, t)$, we can write the probability of selecting the task with priority x conditional to the past history as

$$\mu(t) = \int_0^1 \int_0^1 dx\,dy\,p(x, t)\,q(y, t)\,\eta(x, y). \tag{2}$$

When the selection is done, the selected task is removed from the list and replaced by a fresh new task with a random priority extracted from $\rho(x)$. The other task remains instead in the list. We indicate the first fresh task with N (as "new"). The PDF of its priority conditional to the past history is simply $\rho(x)$ as it is new. The second task is consequently indicated with O (as "old"), and we call the PDF of its priority, conditional to the past history of the task list, as $p_O(x, t)$. Given our particular form (1) of $\eta(x, y)$, $p_O(x, t)$ is, for $p > 0$, more concentrated on the small values of x than $\rho(x)$. At $t = 0$ the initial condition is $p_O(x, 0) = \rho(x)$. Since at each time step the task N is new, the priorities x and y of N and O are statistically independent, and their joint probability factorizes into the product $\rho(x)p_O(y, t)$. Any selection path of the task list lasting τ steps can be represented as a time ordered string of τ letters N and O (e.g., $NONNONO\ldots$). In particular, the statistical weight of the selection path $NN\ldots N$ composed by τ subsequent events N gives the probability that the waiting time of the task O, from the beginning of the list dynamics, is at least τ. In terms of IP in $d = 1$ this path represents a growth avalanche in one single direction starting at $t = 0$ and lasting at least τ steps. We can rewrite Eq. (2) for both the cases in which the task N or O are selected at time t:

$$\mu_N(t) = \int_0^1 \int_0^1 dx\,dy\,\rho(x)\,p_O(y, t)\,\eta(x, y),$$
$$\mu_O(t) = 1 - \mu_N(t) = \int_0^1 \int_0^1 dx\,dy\,p_O(x, t)\,\rho(y)\,\eta(x, y). \tag{3}$$

For both selection events, the conditional PDFs of the priorities are consequently updated including this last step in the past history conditioning probabilities. As explained above, the conditional PDF of the new task replacing the selected one is $\rho(x)$. Instead, the conditional PDF $p_O(x, t + 1)$ at time $(t + 1)$ of the just unselected task, still in the list, is different in the two cases above. If the task N is selected, the task O remains O also at the next time step. We use conditional probability to include also the memory of the last selection step:

$$p_O(x, t + 1) = \frac{1}{\mu_N(t)} p_O(x, t) \int_0^1 dy\,\rho(y)\,\eta(y, x). \tag{4}$$

If instead O is selected at time t, it is removed, and the task N at time t becomes the task O at $(t + 1)$:

$$p_O(x, t + 1) = \frac{1}{\mu_O(t)} \rho(x) \int_0^1 dy p_O(y, t) \eta(y, x). \quad (5)$$

The whole set of all possible selection paths can be represented as a non-Markovian binary branching process whose realizations tree is represented in Fig. 1(b). The initial node (top vertex) represents the initial situation with two tasks with completely random priorities with PDF $\rho(x)$. From each node there is a bifurcation of possible choices: either task N or O. Therefore, each node of the tree represents the task list at the end of the time ordered selection path connecting directly the top vertex with the given node and is characterized by path-dependent conditional PDF p_O and probabilities for the next bifurcation μ_N and μ_O which can be evaluated through the step-by-step RTS procedure. The exact statistical weight of each selection path on the tree is given by the product of the probabilities μ along the path. Therefore, the RTS provides a full mathematical description of the task list dynamics. Note that such dynamics in this model is a binary branching process with memory, in the sense that the probabilities of a configuration at a given time depends for $p > 0$ on all the past history of the list. This memory effect is maximized for $p = 1$ when the dynamics becomes deterministic and extremal.

A fundamental quantity in this class of dynamics is the average priority "histogram" [9,11] that is the statistical distribution of the priorities of the task list at a given time t averaged over all the selection paths: $h(x, t) = [\rho(x) + \langle p_O(x, t) \rangle_t]/2$. Hence, the evolution of $h(x, t)$ is directly given by that of $\rho_1(x, t) = \langle p_O(x, t) \rangle_t$. The equation for its time evolution can be found by observing that at each binary branching starting from a node at time t of the tree, we can say that with probability $\mu_N(t)$ the priority conditional PDF $p_O(x, t)$ updates as in Eq. (4) and with probability $\mu_O(t)$ as in Eq. (5), i.e.,

$$\pi_O(x, t + 1; t) = p_O(x, t) \int_0^1 dy \rho(y) \eta(y, x) + \rho(x)$$
$$\times \int_0^1 dy p_O(y, t) \eta(y, x),$$

where $\pi_O(x, t + 1; t)$ is the conditional PDF of the priority of the task O at time $(t + 1)$ conditional to the history only up to time t. By applying this average from the first time step it is simple to show that:

$$\rho_1(x, t + 1) = \rho_1(x, t) \int_0^1 dy \rho(y) \eta(y, x) + \rho(x)$$
$$\times \int_0^1 dy \rho_1(y, t) \eta(y, x). \quad (6)$$

This is exactly the basic equation used in [6] to study the task list dynamics in the stationary state, i.e., when $\rho_1(x, t + 1) = \rho_1(x, t)$. We show in the following that, however, in the limit $p \to 1$ the stationary state is reached only very slowly and that the waiting time distribution is determined by the finite time deviation from the stationary

state at all time. This waiting time distribution is again a power law, but with a different exponent with respect to that found in [6]. We now solve exactly the RTS for the case $p = 1$. For our calculation we use here $\rho(x) = 1$ for $x \in [0, 1]$ as in [6] because the path statistics, as aforementioned, does not depend on $\rho(x)$. Hence Eqs. (3) above take the simple form:

$$\mu_N^e(t) = \int_0^1 dx (1 - x) p_O^e(x, t),$$

$$\mu_O^e(t) = 1 - \mu_N^e(t) = \int_0^1 dx x p_O^e(x, t),$$

where the superscript "e" stands for "extremal." Analogously Eqs. (4) and (5) for $p = 1$ read, respectively,

$$p_O^e(x, t + 1) = \frac{1}{\mu_N^e(t)} (1 - x) p_O^e(x, t),$$

$$p_O^e(x, t + 1) = \frac{1}{\mu_O^e(t)} \int_x^1 dx p_O^e(x, t).$$

Using $p_O^e(x, 0) = \rho(x)$ in the above equations, one finds that, for any selection path, μ_N^e, μ_O^e, and p_O^e becomes:

$$\mu_N^e(t) = \frac{t + 1}{t + 2}; \qquad \mu_O^e(t) = \frac{1}{t + 2} \quad (7)$$

$$p_O^e(x, t) = (t + 1)(1 - x)^t. \quad (8)$$

Since $p_O^e(x, t)$ is independent of the considered path, $\rho_1(x, t)$ for $p = 1$ coincides with it. Note that $\mu_N^e(t \to \infty) \to 1$ and $\rho_1(x, t) = p_O^e(x, t \to \infty) \to \delta(x - 0^+)$ [where $\delta(x)$ is the Dirac delta function]; i.e., in the infinite time limit the new fresh task is always selected as the old one has vanishing priority with probability one. The fact that both the μ^e's and the p^e's at time t are the same for each selection path of length t is a feature of the $p = 1$ case. This is not the case for $0 < p < 1$ where instead the conditional selection probability and priority PDFs at time t depend on which specific selection path is considered. We now analyze the consequences of Eq. (7). It permits to find the waiting time distribution of a given task entered the list at time t_0. From Eq. (7) we find that for $p = 1$ the waiting time is $\tau = 0$ with probability $P(\tau = 0; t_0) = (t_0 + 1)/(t_0 + 2)$. The probability that it is still waiting after $\tau \geq 1$ steps, i.e., at time $t_0 + \tau$, is

$$W(\tau, t_0) = \frac{1}{t_0 + 2} \prod_{t'=1}^{\tau - 1} \frac{t_0 + t' + 1}{t_0 + t' + 2} = \frac{1}{t_0 + \tau + 1},$$

which is the probability of the path $ONN \ldots N$ with one O event at t_0 and $(\tau - 1)$ subsequent N events. Hence the probability that the waiting time is exactly $\tau \geq 1$ is

$$P(\tau; t_0) = \frac{W(\tau; t_0)}{t_0 + \tau + 2} = \frac{1}{(t_0 + \tau + 1)(t_0 + \tau + 2)}, \quad (9)$$

which is the statistical weight of the selection path $ONN \ldots NO$ with one O event at t_0, $(\tau - 1)$ subsequent
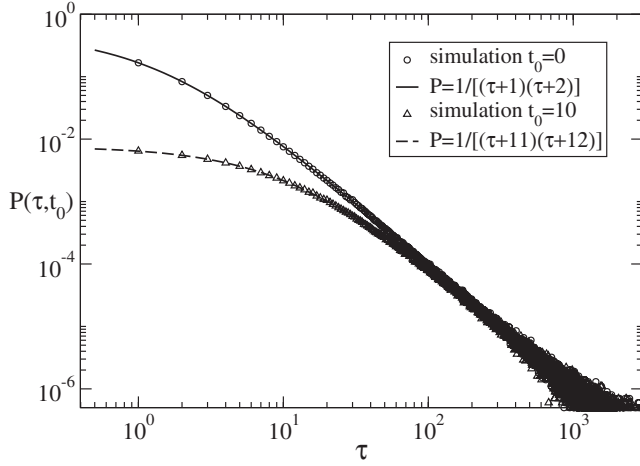
FIG. 2. Waiting time distribution $P(\tau, t_0)$ in the task list dynamics with $N = 2$ and $p = 1$ and initial time $t_0 = 0, 10$. Numerical data for each t_0 have been averaged over 10^7 realizations. As shown, it is very well fitted by Eq. (9).

N events and a final O event. Note that this corresponds in IP to the statistical weight of an avalanche starting a time t_0 and lasting τ steps. The probability $P(\tau; t_0)$ decreases as τ^{-2} for $\tau \gg t_0$ (a behavior confirmed by numerical simulations—see Fig. 2). Therefore, the waiting time distribution for a task entered at time t_0 is normalizable in τ, but with diverging mean value. This behavior is different from the power law $P(\tau) \sim \tau^{-1}$ found in [6] for the stationary state for $0 < p < 1$, which, however, disappears for $p \to 1$ as its amplitude vanishes in this limit. For the opposite limit $t_0 \gg \tau \gg 1$ one can write $P(\tau; t_0) \simeq t_0^{-2}(1 - 2\tau/t_0)$, which decreases as t_0^{-2} with t_0. This gives the rate of approach in the initial time t_0 to the trivial stationary state $P(\tau; t_0 \to \infty) = 1$ or 0, respectively, if $\tau = 0$ or $\tau \geq 1$. This rate is very slow and there is no characteristic time after which one can say that the stationary state is attained in terms of τ dependence.

In order to study the approach to the stationary state for $p \to 1$, let us analyze in detail Eq. (6). First of all we rewrite this equation using the explicit form (1) of $\eta(x, y)$ for this model and $\rho(x) = 1$ for $x \in [0, 1]$:

$$\rho_1(x, t + 1) = \rho_1(x, t)\left[p(1 - x) + \frac{1 - p}{2}\right]$$
$$+ p \int_x^1 dy\rho_1(y, t) + \frac{1 - p}{2}. \quad (10)$$

We now put $\rho_1(x, t) = \rho_1^{(s)}(x) + \delta\rho_1(x, t)$, where $\rho_1^{(s)}(x)$ is the stationary solution found in Eq. (4) of [6]:

$$\rho_1^{(s)}(x) = \frac{1 + p}{1 - p} \frac{1}{[1 + \frac{2p}{1-p}x]^2} \quad (11)$$

and $\delta\rho_1(x, t)$ is the finite time deviation from it. For $p \to 1$ the PDF $\rho_1^{(s)}(x) \to \delta(x - 0^+)$ and it coincides with Eq. (8) for $t \to \infty$. Since $\rho_1(x, t)$ and $\rho_1^{(s)}(x)$ are both normalized to unity, we have $\int_0^1 dx\delta\rho_1(x, t) = 0$. Therefore, as a first or-

der approximation we put $\int_x^1 dy\delta\rho_1(y, t) \simeq -x\delta\rho_1(x, t)$. Taking also the continuous time approximation $[\delta\rho_1(x, t + 1) - \delta\rho_1(x, t)] \simeq d\delta\rho_1(x, t)/dt$, we can rewrite Eq. (10) in terms of $\delta\rho_1(x, t)$ as

$$\frac{d\delta\rho_1(x, t)}{dt} \simeq -\left(\frac{1 - p}{2} + 2px\right)\delta\rho_1(x, t). \quad (12)$$

Hence $\delta\rho_1(x, t)$ decays exponentially in time with an x-dependent time constant inversely proportional to $[(1 - p)/2 + 2px]$. For $p < 1$, at each x the perturbation decays exponentially fast and the stationary state is attained, while for $p \to 1$ the time constants become proportional to $1/x$ and the perturbation in the region around $x = 0$ relaxes very slowly. But from Eq. (11) it is exactly in this region that for $p \to 1$ all the measure $\rho^{(s)}(x)$ is concentrated. This confirms our previous conclusion that for $p \to 1$ the stationary state is very slowly attained and finite time deviation from it plays a fundamental role in determining the rate of decrease of the waiting time distribution.

In this Letter we have studied an interesting queuing model of task list dynamics introduced by Barabási. Through a statistical method called RTS, we are able to give a complete probabilistic description of the dynamics even out of stationarity. We find that for $0 < p < 1$ finite time deviations from stationarity relaxes exponentially fast and, consequently, the dynamics is well described by the stationary state. However, for $p \to 1$ the stationary state becomes trivial and finite time deviations relax so slowly that the task list dynamics has to be described as an intrinsically nonstationary dynamics. This is characterized by power law waiting time distributions with a characteristic exponent which is different from the one found [6] in the stationary state for $0 < p < 1$.

[1] H. C. Tijms, *A First Course in Stochastic Models* (Wiley, New York, 2003).
[2] L. Breuer and D. Baum *An Introduction to Queueing Theory and Matrix-Analytic Methods* (Springer, New York, 2005).
[3] J. G. Oliveira and A.-L. Barabasi, Nature (London) **437**, 1251 (2005).
[4] A. Vázquez *et al.*, Phys. Rev. E **73**, 036127 (2006).
[5] A.-L. Barabási, Nature (London) **435**, 207 (2005).
[6] A. Vázquez, Phys. Rev. Lett. **95**, 248701 (2005).
[7] M. Marsili, J. Stat. Phys. **77**, 733 (1994); A. Gabrielli *et al.*, J. Stat. Phys. **84**, 889 (1996).
[8] D. Wilkinson and J. F. Willemsen, J. Phys. A **16**, 3365 (1983).
[9] R. Cafiero *et al.*, Phys. Rev. E **54**, 1406 (1996).
[10] M. Felici *et al.*, Phys. Rev. Lett. **86**, 1896 (2001).
[11] A. Gabrielli, G. Caldarelli, and L. Pietronero, Phys. Rev. E **62**, 7638 (2000).
[12] G. Grinstein and R. Linsker, Phys. Rev. Lett. **97**, 130201 (2006).