Dynamics of condensation in growing complex networks

Luca Ferretti[1] and Ginestra Bianconi[2]
[1]*SISSA, via Beirut 4, I-34014 Trieste, Italy*
[2]*The Abdus Salam International Center for Theoretical Physics, Strada Costiera 11, 34014 Trieste, Italy*
(Received 10 April 2008; revised manuscript received 24 August 2008; published 6 November 2008)

A condensation transition was predicted for growing technological networks evolving by preferential attachment and competing quality of their nodes, as described by the fitness model. When this condensation occurs, a node acquires a finite fraction of all the links of the network. Earlier studies based on steady-state degree distribution and on the mapping to Bose-Einstein condensation were able to identify the critical point. Here we characterize the dynamics of condensation and we present evidence that below the condensation temperature there is a slow down of the dynamics and that a single node (*not* necessarily the best node in the network) emerges as the winner for very long times. The characteristic time t^* at which this phenomenon occurs diverges both at the critical point and at $T \rightarrow 0$ when new links are attached deterministically to the highest quality node of the network.

DOI: 10.1103/PhysRevE.78.056102 PACS number(s): 89.75.Hc, 89.75.Da, 89.75.Fb

INTRODUCTION

Condensation phenomena [1–7] in complex networks [8–11] are structural phase transitions in which a node grabs a finite fraction of all the links of the network. This phenomenon is of particular interest in the case of technological networks. The maps of the Internet at the Autonomous System Level show that the fraction of nodes connected to the most connected node is increasing in time reaching a share of the order of 10% in recent maps [12]. Also in the World-Wide-Web, the share of webpages linked to Google webpages is of the order of 1%, a large number if one takes into account the size of the World-Wide-Web. Are the Internet and the World-Wide-Web close to a condensation transition? What are the dynamical signatures of a condensation? What are the consequences of a condensation of technological networks? The problem might have relevant implications for the monitoring of technological networks, and there might be important differences between the statistics of lead change below and above [13] the condensation transition. In the following, we are studying these problems focusing on the dynamics of the fitness model [2,14] in which nodes acquire links in proportion to their connectivity and in proportion to their fitness, which indicates the quality of the node. This model has been considered a good stylized model for the Internet [11] and is a good stylized model describing the emergence of high-quality search engines in the World-Wide-Web and shows a condensation phase transition depending on the parameters of the model. The study of the dynamical properties of this model will be able to give some estimates of the characteristic time scale at which a condensate node might emerge in a condensation scenario and will allow us to evaluate, once a condensate node is formed, the probability that a new node would overcome the condensate at later times.

Questions concerning the dynamics of condensation within the fitness model might also shed some light on the relation between the condensation phenomena occurring in this model and the other off-equilibrium condensation phase transitions occurring in complex systems (traffic jam, wealth distribution, urn, and network models) [15–23]. Off-equilibrium phenomena have attracted a large amount of interest in the past ten years. Most of the attention has been focused on the characterization of the steady state of these out-of-equilibrium systems, above the condensation transition [2,3,7,16,17]. On the contrary, the dynamics by which the condensate emerges [17–19,22] has been studied only in the case of urn models and models that can be reduced to the zero-range process [15]. However, the models where the condensation occurs as a structural phase transition of a network are not in general reducible to the zero-range process.

In this paper, we study the dynamics of condensation in the fitness model. We apply the rank statistics for the description of the dynamics of condensation, and we will show that below the condensation phase transition a condensate node emerges only after a characteristic time t^*. This node is then one of the nodes with highest fitness. The probability that late high fitness nodes will overcome the condensate is decaying with time and there is a slow down of the dynamics. At the condensation transition, the characteristic time t^* diverges and the condensation is only marginal, i.e., we have a sequence of highly connected high-fitness nodes, each one overcoming the other and then grabbing a very small fraction of the total links of the network.

FITNESS MODEL AND CONDENSATION PHASE TRANSITION

The fitness model [14] is a growing network model. We start from a finite connected network of N_0 nodes. At each time t_i, a new node i and $m < N_0$ new links are added to the network. To the node i, it is assigned a quenched variable, ε_i ("energy" of the node), drawn from a $g(\varepsilon)$ distribution. The variable ε_i indicates the intrinsic quality of the node (lower "energy," better quality). Each of the m new links that are added to the network at time t_i connects the new i node to a node j chosen with probability

$$\Pi_j(t) = \frac{\eta_j k_j(t)}{\Sigma_\ell \eta_\ell k_\ell(t)}, \tag{1}$$

where we have introduced the fitness η_j associated to node j as

$$\eta_j = e^{-\beta\epsilon_j}. \tag{2}$$

The parameter $\beta = 1/T$ tunes the relevance of the quality of the nodes in the choice of the target node. In the limit case $T \to \infty$, the dynamics of the attachment of new links is only dependent on the connectivity of the nodes and we recover the Barabási-Albert model [24]. For $T \ll 1$, the intrinsic quality of the nodes highly affects the dynamic of the system.

For the $g(\epsilon)$ distribution such that $g(\epsilon) \to 0$ as $\epsilon \to 0$, a phase transition can occur at a critical temperature T_c. Above this critical temperature, every node has an infinitesimal fraction of all the links; below the critical temperature, one node grabs a finite fraction of all the links. These results can be obtained by consideration of the characteristics of the steady state of the model above the phase transition. In fact, the model can be solved in a mean-field approximation by making a self-consistent assumption. In particular, in [14] it is assumed that $Z_t = \Sigma_j \eta_j k_j(t) \to \langle Z_t \rangle \to me^{-\beta\mu}t + O(t^0)$. With this assumption, it can be shown that, at time t, the average connectivity of node i arriving in the network at time t_i grows as a power law $\langle k_i(t) \rangle = m(t/t_i)^{f(\epsilon_i)}$ with $f(\epsilon) = e^{-\beta(\epsilon-\mu)}$ and with the constant μ determined by the self-consistent equation

$$1 = I(\beta,\mu) = \int d\epsilon\, g(\epsilon) \frac{1}{e^{\beta(\epsilon-\mu)} - 1}. \tag{3}$$

In [2], it was shown that the self-consistent equation (3) is equivalent to the equation for the conservation of the total number of links present in the network,

$$2mt = mt + mtI(\beta,\mu). \tag{4}$$

If Eq. (3) has a solution, then the degree distribution of nodes of fitness $\eta = e^{-\beta\epsilon}$ is given by $P_\epsilon(k) \simeq k^{-1/f(\epsilon)-1}$ and $P(k) = \int d\epsilon g(\epsilon) P_\epsilon(k)$ is dominated by the term $P_0(\epsilon) = k^{-\gamma}$ with $\gamma = e^{-\beta\mu} + 1$ but can have logarithmic corrections [14]. Nevertheless, if $I(\beta,0) < 1$, the self-consistent equation (3) cannot be solved. In this case, the self-consistent approach fails and we have the condensation phase transition. A necessary condition for condensation [i.e., for $I(\beta,0) < 1$] is that the distribution $g(\epsilon) \to 0$ for $\epsilon \to 0$. For this type of $g(\epsilon)$ distributions there can be a critical temperature $T_c = 1/\beta_c$ such that $I(\beta,0) < 1 \,\forall\, \beta > \beta_c$. This phase transition is formally equivalent to the Bose-Einstein condensation for noninteracting Bose gases in dimensions $d > 2$. Using the similarity to the Bose-Einstein condensation in Bose gases, in Ref. [2] it was assumed that below the condensation temperature, the equation for conservation of the links can be written as

$$2mt = mt + mtI(\beta,0) + n_0(\beta)mt, \tag{5}$$

where $n_0(\beta)$ indicates the number of links attached to the most connected node. Consistent with Eq. (5), simulation results reported in [2] show that indeed the fitness model undergoes a condensation phase transition at $T = T_c$. An example of $g(\epsilon)$ distributions for which there is a condensation is

$$g(\epsilon) = (\theta + 1)\epsilon^\theta \tag{6}$$

and $\epsilon \in [0,1]$. Here and in the following, we will always consider $g(\epsilon)$ distributed according to Eq. (6). Assuming Eq. (5) and a distribution of the type (6), the fraction of condensed links, similarly to the corresponding result for the Bose gas, will take the form

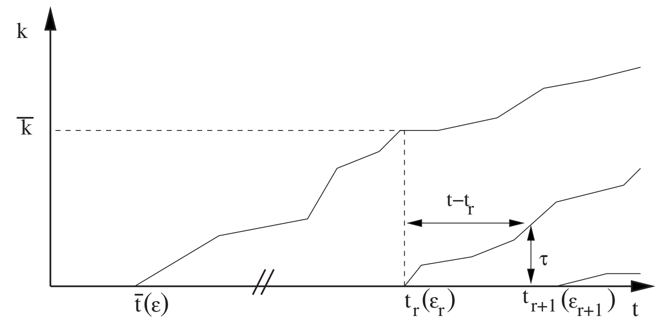

FIG. 1. We plot the connectivities of records as a function of time. The time t_r indicates the time at which a record node with ϵ_r arrives in the network. The effective model only focuses on the competition for links of the most connected nodes with the node of highest fitness.

$$n_0(\beta) = 1 - \left(\frac{T}{T_c}\right)^{\theta+1}. \tag{7}$$

DYNAMICS OF THE FITNESS MODEL

The necessary condition for condensation $g(\epsilon) \to 0$ as $\epsilon \to 0$ is a clear indication that dynamical effects not captured by the stationary-state approaches [2,3,7] will be very important for the emergence of a long-lasting condensate node in the network. Moreover, all considerations regarding the dynamical aspects of the condensation phenomena on the fitness network cannot take advantage of the similarity with the condensation phenomena in the noninteracting Bose gas. In fact, in a Bose gas, the equivalent of $n_0(\beta)$ is the occupation of the state at energy $\epsilon = 0$. On the contrary, in the fitness networks, $n_0(\beta)$ cannot be the connectivity of the fittest node ($\epsilon = 0$). Indeed, in the fitness model, the fittest node appears in the network only in the infinite time limit because the probability $g(\epsilon)$ of having a node of quality ϵ goes to zero as $\epsilon \to 0$. Therefore, the question is as follows: is the condensate node the best node in the network? Once the condensate has appeared in the network, is there a possibility for late high fitness nodes to take over the condensate? To narrow down this general question, we define a specific problem we want to address. First of all, let us introduce the notion of a record. A record $\epsilon_R(t)$ at time t is the node (or equivalently the value of its ϵ) with minimal ϵ (maximal fitness) present in the network. In the fitness model below the condensation transition, all records are good candidates as condensate nodes. Suppose that a record ϵ, arrived in the network at time $\bar{t}$, is a condensate with $\bar{k}$ links at the time t_r where the rth record occurs. We want to know the probability that the rth record of fitness $\eta_r = e^{-\beta\epsilon_r}$ overcomes the condensate node before the subsequent record ϵ_{r+1} enters the network. (See Fig. 1 for a pictorial representation of our problem.)

In order to answer this question, we have to use some results related to record statistics.

RECORDS

The typical value of the record at time t is given by

$$\int_0^{\varepsilon_R} g(x)dx = \frac{1}{t}, \tag{8}$$

indicating that the average number of nodes with $\varepsilon > \varepsilon_R(T)$ in a network of t nodes is less than one. Using Eq. (8) for the distribution (6), we obtain

$$\varepsilon_R(t) \simeq t^{-1/(\theta+1)}. \tag{9}$$

The statistics of records is a field on its own [25]; here we cite only another result that will be used in the following: given the rth record at time t_r, the probability to have the following record at time $t_{r+1} > t$ in the approximation of a continuous time is given by [25]

$$P_R(t_{r+1} > t, t_r) = \frac{t_r}{t}, \tag{10}$$

where $t > t_r$.

EFFECTIVE MODEL

At low temperatures, below the condensation phase transition, the linking probability Π_j [Eq. (1)] of the fitness model will be relevant only for nodes of very high fitness (the series of records) or for the condensate node. Therefore, we consider the extremely simplified effective model of just two competing nodes (the condensate node and the record) disregarding all the other nodes present in the network. This simplification will provide us with a scenario for condensation that we will subsequently compare with simulation results on the original fitness model. At time t_r, there is a node (the condensate) arriving in the network at time $\bar{t}$ with fitness $\bar{\eta} = e^{-\beta\bar{\varepsilon}}$ and $\bar{k}$ links, and a record node with fitness $\eta_r = e^{-\beta\varepsilon_r}$ that is just born and has degree 1. At each time, we distribute one link to one of the two nodes with probability given by expression (1).

The probability $p_t(\tau)$ that the node with fitness η_r has τ links at time t satisfies in the continuous time limit

$$\frac{\partial p_t(\tau)}{\partial t} = \frac{\eta_r(\tau-1)}{Z_{t,\tau-1}} p_t(\tau-1) - \frac{\eta_r \tau}{Z_{t,\tau}} p_t(\tau), \tag{11}$$

where the competition is included in the formula through $Z_{t,\tau} = \bar{\eta}(\bar{k}+t-t_r-\tau) + \eta_r\tau$. We are interested in the dynamics of this effective model until the node with fitness η_r takes over, i.e., until

$$\tau < \tau_C = (\bar{k}+t-t_r)/2. \tag{12}$$

Equation (11) is nontrivial to study and some approximations are necessary. Therefore, we assume that $Z_{t,\tau} \simeq Z_{t,\tau-1}$ (i.e., $\beta\bar{\varepsilon} \ll 1$), which, taking into account the scaling of the records with the time of their appearance, Eq. (8), reduces to the condition

$$\bar{t} > T^{-(\theta+1)}. \tag{13}$$

Furthermore, we assume that $C \simeq Z_{t,\tau}/(\bar{k}+t-t_r)$ can be approximated by a constant for $\tau < \tau_C$. Indeed, for $\tau < \tau_C$ we have that C varies in a narrow interval $C \in (\bar{\eta}, (\bar{\eta}+\eta_r)/2)$. In this approximation, we can rewrite Eq. (11) using the generating function $q(z,t) = \Sigma_{\tau=1}^{\infty} p(\tau,t) z^{\tau}$. The equation becomes

$$C(\bar{k}+t-t_r)\frac{\partial q}{\partial t} = \eta_r z(z-1)\frac{\partial q}{\partial z}, \tag{14}$$

which, if we impose the initial condition $q(z,t_r) = z$, has the solution

$$q(z,t) = \frac{z}{z\left(\frac{\bar{k}+t-t_r}{\bar{k}}\right)^{-\eta_r/C} + (1-z)} \left(\frac{\bar{k}+t-t_r}{\bar{k}}\right)^{-\eta_r/C},$$

and the expression for the probability $p_t(\tau)$ becomes

$$p_t(\tau) = \left(\frac{\bar{k}}{\bar{k}+t-t_r}\right)^{\eta_r/C} \left[1 - \left(\frac{\bar{k}}{\bar{k}+t-t_r}\right)^{\eta_r/C}\right]^{\tau-1}.$$

The probability that, at some time t, $\tau < \tau_C$, i.e., the probability that the record has not become the most connected node of the network, can be calculated,

$$P_1(\tau < \tau_C) \simeq 1 - \exp\left[-\frac{\bar{k}}{2}\left(\frac{\bar{k}}{\bar{k}+t-t_r}\right)^{(\eta_r-C)/C}\right],$$

where in the last step we have assumed that $t_C - t_r \gg 1$. Expression (15) for $P_1(\tau < \tau_C)$ allows us to evaluate the characteristic time t_C at which the record r becomes the most connected node, i.e.,

$$t_C - t_r \simeq \bar{k}^{1+C/(\eta_r-C)}. \tag{15}$$

Therefore, by making use of Eq. (10), we can evaluate the probability P_C^{eff} that the record r has become the most connected node of the network before the appearance of the record $r+1$,

$$P_C^{\mathrm{eff}} = P_R(t_{r+1} > t_C, t_r) = \simeq \frac{1}{\hat{n}\bar{k}^{C/(\eta_r-C)}} = \frac{1}{\hat{n}\bar{k}^{\alpha}}, \tag{16}$$

where $\hat{n} = \bar{k}/t_r$. Taking into account the interval of definition of C, the scaling exponent is $\alpha = C/(\eta_r - C) \in [1/[\beta(\bar{\varepsilon}-\varepsilon_r)], 2/[\beta(\bar{\varepsilon}-\varepsilon_r)]]$. Therefore, if $\bar{\varepsilon} \gg \varepsilon_r$, then

$$\alpha \simeq \frac{1}{\beta\bar{\varepsilon}} \simeq T\bar{t}^{1/(\theta+1)}. \tag{17}$$

In other words, as long as Eq. (13) is satisfied, the exponent α is greater than 1 (i.e., $\alpha > 1$) and it increases with larger $\bar{t}$ times and for higher temperatures T.

If we fix the condition $P_C^{\mathrm{eff}} < 1/B$, we find that a characteristic time for the appearance of a long-lasting condensate node in the effective model is given by

$$t^* = \max\left(T^{-(\theta+1)}, \left(\frac{B}{\hat{n}}\right)^{1/\alpha} \frac{1}{\hat{n}}\right), \tag{18}$$

where $\hat{n} = \bar{k}/t_r$.

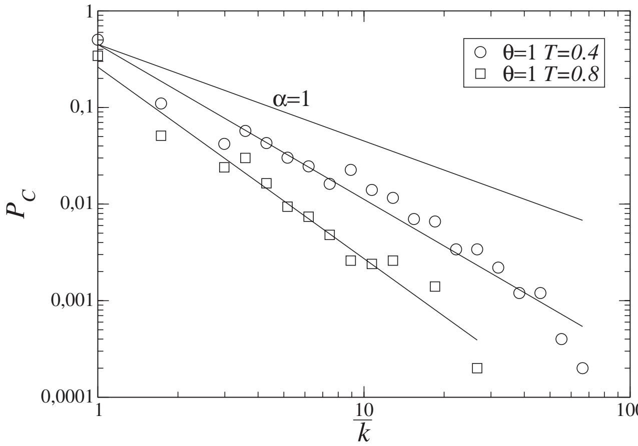

FIG. 2. Probability P_C that a record overcomes the most connected node of the network in the fitness model averaged over 10^4 runs for a time lapse of $t_{\max}=10^5$.

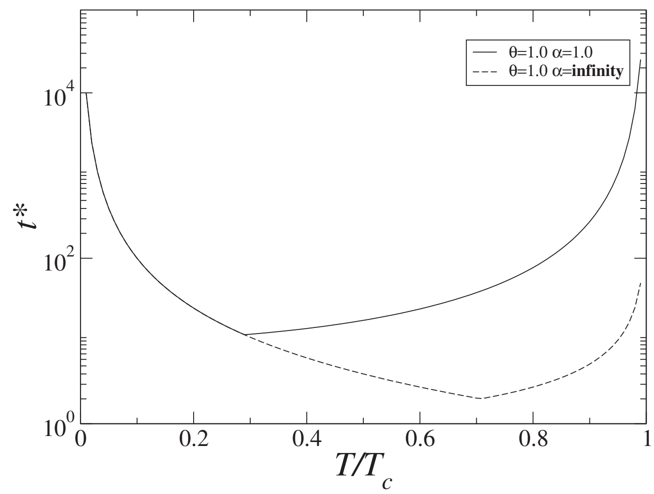

FIG. 3. Characteristic time t^* calculated with the estimation (20) ($B=10$) after which a long-lasting condensate node emerges for the fitness model for $\theta=1$ and $\alpha=1,\infty$.

DYNAMICS OF CONDENSATION

We expect that the effective model, describing the competition of only two nodes, provides an upper bound for the probability $P_C(\bar{k})$ that in the fitness model a new record overcomes the most connected node before the appearance of a subsequent record. We define $P_C(\bar{k})$ as the probability that a record, introduced in a network with a condensate of connectivity $\bar{k}$, is able to overcome the connectivity of the condensate node before the subsequent record. To numerically evaluate $P_C(\bar{k})$, we consider all the records that follow the emergence of a condensate and we evaluate the ratio between the number of positive events and the total number of records that appear in the network when the condensate has $\bar{k}$ links. The data reported in Fig. 2 are taken for $\theta=1$ at different temperatures T below the phase transition. The results shown are statistics taken over 10^4 networks for a time lapse $T^{-(\theta+1)}<t<10^5$. Figure 2 shows clearly that the probability $P_C(\bar{k})$ decays as

$$P_C(\bar{k}) \propto \bar{k}^{-\alpha} \tag{19}$$

with $\alpha \geqslant 1$ and that the exponent α increases with T as predicted by the effective model. Using the effective model estimate (18) for the characteristic time t^* for having a long-lived winner node that condenses, and assuming $\hat{n} \simeq n_0(\beta)$ with $n_0(\beta)$ given by Eq. (7), we get

$$t^* = \max\left\{T^{-(\theta+1)}, B^{1/\alpha}\left[1-\left(\frac{T}{T_c}\right)^{\theta+1}\right]^{-(1+1/\alpha)}\right\}. \tag{20}$$

This characteristic time is depicted in Fig. 3 for different values of α, and the results are consistent with a divergence of t^* at $T\to T_c$ and at $T\to 0$.

In order to evaluate the time life of a condensate for times $t>t^*$, we can use the scaling of the probability $P_C \sim n_0^{-1}\bar{k}^{-\alpha}$ and we can assume that for the condensate node $\bar{k}\simeq n_0 t$. Moreover, if the relevant competition is only between the condensate node and the node with highest fitness, the probability to have the same condensate after S records is given by

$$P_W^{\text{eff}}(S) \simeq \prod_{s=r^\star+1}^{r^\star+S}\left(1-\frac{1}{n_0(n_0 t_s)^\alpha}\right),$$

where t_s are the times of record occurrence after the condensate has become the most connected node in the network (i.e., for $s>r^\star$). The average of this quantity over the distributions of the times of the records $\Pi_s P_R(t_s,t_{s-1})=\Pi_s t_{s-1}/t_s^2$ gives

$$\langle P_W^{\text{eff}}\rangle(S) \simeq \exp\left\{-\left[1-\frac{1}{(1+\alpha)^S}\right]\frac{1+\alpha}{\alpha n_0(n_0 t_{r^\star})^\alpha}\right\}. \tag{21}$$

Therefore, for long times, we expect $\langle P_W^{\text{eff}}\rangle(S)\to$ const as $S\to\infty$. Consequently, we can conclude that for times $t>t^*$, the condensate node is typically not the best node of the network, nevertheless it dominates the network for very long times. Only for $T\to T_c$ and $T\to 0$ does the time needed to have long-lasting condensate diverge, and in the network there is a fair competition and a succession of high fitness nodes on which condensation occurs.

CONCLUSIONS

In conclusion, we have studied the dynamics of condensation in growing network models within the fitness model. We show that below the condensed phase transition, there is a characteristic time t^*. For times $t<t^*$, the network dynamics is dominated by a succession of high fitness nodes with a finite fraction of the links. Above the characteristic time t^*, a long-lasting winner takes over acquiring a finite fraction of all the links and slowing down the dynamics. The characteristic time t^* is diverging at the critical point of the condensation phase transition and in the limit $T\to 0$. These results

may have strong implications for the monitoring of technological networks in which we expect a fair competition lasting for very long times only if it is close to the condensation transition.

ACKNOWLEDGMENTS

We acknowledge fruitful discussions with Claude Godréche, Jean Marc Luck and Marcello Mamino.

[1] P. L. Krapivsky, S. Redner, and F. Leyvraz, Phys. Rev. Lett. **85**, 4629 (2000).
[2] G. Bianconi and A.-L. Barabási, Phys. Rev. Lett. **86**, 5632 (2001).
[3] G. Ergün and G. J. Rodgers, Physica A **303**, 261 (2002).
[4] P. L. Krapivsky and S. Redner, Comput. Netw. **39**, 261 (2002).
[5] S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin, Nucl. Phys. B **653**, 307 (2003).
[6] S. N. Dorogovtsev, J. F. F. Mendes, A. M. Povolotsky, and A. N. Samukhin, Phys. Rev. Lett. **95**, 195701 (2005).
[7] C. Borgs, J. Chayes, C. Daskalakis, and S. Roch, in *Annual ACM Symposium on the Theory of Computation* (ACM, New York, 2007).
[8] R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).
[9] S. N. Dorogovtsev and J. F. F. Mendes, *Evolution of Networks: From Biological Nets to the Internet and WWW* (Oxford University Press, Oxford, 2003).
[10] S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes, Rev. Mod. Phys. **80**, 1275 (2008).
[11] R. Pastor-Satorras and A. Vespignani, *Evolution and Structure of the Internet: A Statistical Physics Approach* (Cambridge University Press, Cambridge, 2004).
[12] See the Internet map published by Marc Newman at http://www-personal.umich.edu/mejn/netdata/ extracted from the University of Oregon Route Views Project data of 22-July 2006.
[13] P. L. Krapivsky and S. Redner, Phys. Rev. Lett. **89**, 258703 (2002).
[14] G. Bianconi and A.-L. Barabási, Europhys. Lett. **54**, 436 (2001).
[15] M. R. Evans and T. Hanney, J. Phys. A **38**, R195 (2005).
[16] P. Bialas, Z. Burda, and J. Johnston, Nucl. Phys. B **493**, 505 (1997).
[17] M. R. Evans, J. Phys. A **30**, 5669 (1997).
[18] C. Godréche and J. M. Luck, J. Phys. A **38**, 7215 (2005).
[19] C. Godréche and J. M. Luck, Eur. Phys. J. B **23**, 473 (2001).
[20] J. P. Bouchaud and M. Mézard, Physica A **282**, 536 (2000).
[21] J. D. Noh, G. M. Shim, and H. Lee, Phys. Rev. Lett. **94**, 198701 (2005).
[22] J. D. Noh, Phys. Rev. E **72**, 056123 (2005).
[23] L. Borgacz, Z. Burda, W. Janke, and B. Waclaw, Chaos **17**, 026112 (2007).
[24] A.-L. Barábasi and R. Albert, Science **286**, 509 (1999).
[25] V. B. Nevzorov, *Records: Mathematical Theory* (American Mathematical Society, Providence, 2001).