Communication in Networks with Hierarchical Branching

A. Arenas,[1] A. Díaz-Guilera,[2] and R. Guimerà[3]

[1]*Departament d'Enginyeria Informàtica, Universitat Rovira i Virgili, Carretera Salou s/n, 43006 Tarragona, Spain*

[2]*Departament de Física Fonamental, Universitat de Barcelona, Diagonal 647, 08028 Barcelona, Spain*

[3]*Departament d'Enginyeria Química, Universitat Rovira i Virgili, Carretera Salou s/n, 43006 Tarragona, Spain*

(Received 5 June 2000)

We present a simple model of communication in networks with hierarchical branching. We analyze the behavior of the model from the viewpoint of critical systems under different situations. For certain values of the parameters, a continuous phase transition between a sparse and a congested regime is observed and accurately described by an order parameter and the power spectra. At the critical point the behavior of the model is totally independent of the number of hierarchical levels. Also scaling properties are observed when the size of the system varies. The presence of noise in the communication is shown to break the transition. The analytical results are a useful guide to forecasting the main features of real networks.

DOI: 10.1103/PhysRevLett.86.3196 PACS numbers: 89.20.Ff, 05.70.Jk, 64.60.–i

Nowadays, many challenging questions have arisen concerning the behavior of complex technological, economical, and social systems [1]. In particular, computer simulations of agents and their interactions (agent-based modeling) have become a widely used tool in our current understanding of their macroscopic behavior [2]. Especially interesting is the study of hierarchical branching in networks because it seems to be the basic structure underlying complex organizational systems. Our interest is focused on the behavior of hierarchical structures formed by elements (or agents) that interact with each other via communication processes. This framework is especially adequate to study, e.g., Internet flow [3–7], traffic networks [8], river networks [9], and even communication flows in organizations [10].

In this Letter, we propose and study a very simple model of communication. The model includes only the basic ingredients present in a communication process between two elements: (i) information packets to be transmitted (delivered) and (ii) communication channels to transmit the packets. Despite its simplicity, the model reproduces the main characteristics of the flow of information packets in a network, and is general enough to allow the study of communication processes in many conditions: for example, different capabilities of agents to transmit packets, and/or heterogeneity in the communication channels (miscommunication, exogenous effects, etc.) represented by introducing disorder. We observe three different behaviors depending on the capability of agents to transmit packets. In particular, for a certain capability, we observe a continuous phase transition between a sparse and a congested regime when the number of packets to deliver reaches a critical value. Near the transition point signs of criticality arise, we find large fluctuations, critical slowing down, and power law behavior of power spectrum of the amount of information flowing in the network, in agreement with reported empirical data [5]. We provide a mean-field estimation of the critical point in good agreement with simulation results and we define analytically an order parameter to characterize the behavior of the system.

The model is defined in the following way: the communication network is mapped onto a lattice where nodes represent the communicating elements (for instance, employees in a company, routers and servers in a computer network, etc.) and the links between them represent communication lines. In particular, we use hierarchical trees as depicted in Fig. 1, although most of the results reported hold when considering that the hierarchical branching is characteristic of the paths that information follows and not of the topology of the network itself. These structures are characterized by two quantities: the branching factor, z, and the number of levels, m. From now on, we will use the notation (z, m) to describe a particular tree.

The dynamics of the model is the following. At each time step t, an information packet is created at every node with probability p. When a new packet is created, a destination node, different from the origin node, is chosen at random in the network. Thus, during the following time steps $t, t+1, \ldots, t+T$, the packet is traveling towards its destination: once the packet reaches this destination node, it is delivered (disappears from the network). The time a packet remains in the network is related not only to the distance between the source node and the target node, but also to the amount of packets in the network. In

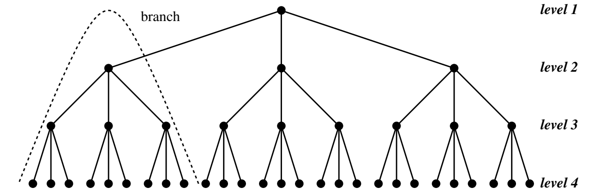

FIG. 1. Typical hierarchical tree structure used for simulations and calculations: in particular, it is a tree $(3, 4)$. Dashed line: definition of *branch*.

particular, at each time step, all the packets move from their current position, i, to the next node in their path, j, with a probability q_{ij}. We define q_{ij}, *quality of communication* between i and j, as

$$q_{ij} = \sqrt{k_{ij} k_{ji}}\,, \tag{1}$$

where $k_{\alpha\beta}$ represents the capability of agent α to communicate with agent β at each time step. For $k_{\alpha\beta}$ we propose

$$k_{\alpha\beta} = \xi_{\alpha\beta} f(n_\alpha)\,, \tag{2}$$

where $\xi_{\alpha\beta}$ is a uniformly distributed random number in the interval $[0, 1]$ representing the effects mentioned above for the directional connection between α and β [11], n_α is the total number of packets currently at node α, and $f(n)$ determines how the capability evolves when the number of packets at a given node increases.

Since any election of $f(n)$ could be valid, we will study the general form

$$f(n) = \begin{cases} 1 & \text{for } n = 0\,, \\ n^{-\gamma} & \text{for } n = 1, 2, 3, \ldots\,, \end{cases} \tag{3}$$

with $\gamma \geq 0$. The average number of packets delivered by a node α to another node β will be proportional to $n_\alpha/(n_\alpha^{\gamma/2} n_\beta^{\gamma/2})$. Assuming the degree of homogeneity derived from the model, $n_\alpha \sim n_\beta$, the former expression reads $n_\alpha^{1-\gamma}$. It is straightforward to recognize three different behaviors corresponding to three different values of γ in the previous formula. For $\gamma > 1$, the number of transmitted packets decreases as n_α grows. For small values of the probability of packet generation per node and time step, p, all the packets are delivered and hence, after a transient, the system reaches a steady state in which the total number of packets, N, fluctuates around a constant value. However, if we increase p at some point the total number of packets will be so large that the network will not be able to handle them, N will increase continuously and, at the end, no packets at all will be delivered to their destination. On the contrary, for $\gamma < 1$, the number of transmitted packets grows as n_α does. Thus, the number of delivered packets increases as N grows until an equilibrium between generated and delivered packets is reached: at this point, N remains constant (except fluctuations). In case $\gamma = 1$, the number of delivered packets is constant irrespective of the number of stored packets (note that this is consistent with simple models of queues [6]). This particular behavior is less obvious and will be treated accurately from the viewpoint of critical systems.

As a first step, let us concentrate on the case $\xi_{ij} = 1$, $\forall i, j$. From simulations, we observe two different regimes and, as in the case $\gamma > 1$, p plays the role of a control parameter. For small values of p, all the packets are delivered, while for large values of p, not all the packets can reach their destination, and N grows in time with no limit. The key point is that, since the number of delivered packets is independent of N, there is always a fraction of problems reaching their destination and the transition to the collapsed regime is continuous. This transition occurs for a critical value of p, p_c, whose exact value depends on the network parameters z and m (see Fig. 2). For values of p smaller than but close to p_c, the steady state is reached but large fluctuations with long correlation times appear.

At the subcritical region, the power spectrum of the total number of packets, $N(t)$, is well fitted by a Lorentzian characterized by a certain frequency, f_c. As we get closer to p_c, we observe that $f_c \to 0$ and the power spectrum becomes $1/f^2$ for the whole range of frequencies. That means that the average time the packets remain in the network grows as we approximate the critical point (critical slowing down). We have also analyzed the power spectrum of the number of packets at individual nodes, $n_i(t)$. The main result is that the power spectrum of $N(t)$ is dominated by the top node which is the most congested: near p_c, the power spectrum for this node is also $1/f^2$. As one goes down in the hierarchy the number of packets diminishes and the power spectra have $1/f^\beta$ tails with β decreasing from 2 to 0 at the lowest level. The last result is consistent with the fact that the bottom agents deliver packets immediately and so $n_i(t)$ is a time series of peaks separated by Poisson distributed time intervals. As it is well known, this kind of series has white noise spectra. We have also checked other topologies [12] and found that in a square lattice with closed boundaries the central sites have $\beta \sim 1.2$ (in agreement with Refs. [5,7]) whereas agents close to the boundaries are less congested and a much lower exponent for the tail ($\beta \sim 0$) is observed.

As happens in other problems in statistical physics [13], the particular symmetry of the hierarchical tree allows a mean-field estimation of the critical point p_c (although these calculations can be performed under more general

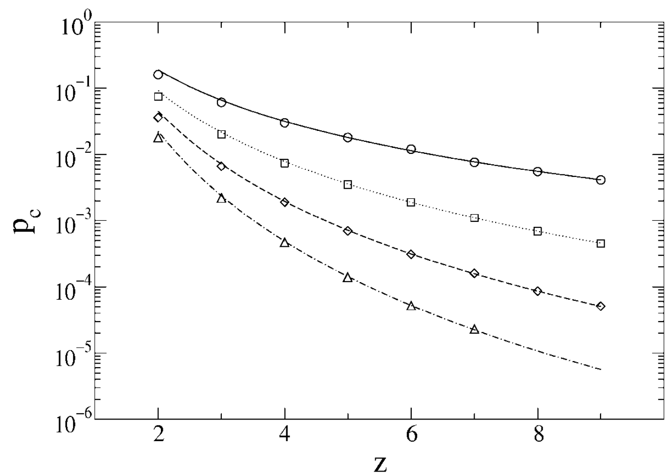

FIG. 2. Comparison between simulated (symbols) and analytical (lines) values for the critical probability of packet generation, p_c as a function of the branching factor z for hierarchies with different numbers of levels: $m = 4$ (circles and solid line), $m = 5$ (squares and dotted line), $m = 6$ (diamonds and dashed line), and $m = 7$ (triangles and dot-dashed line). The error bars are smaller than the symbol size.

conditions [12]). Since in the steady state regime there is no accumulation of packets, the number of packets arriving at the top of the hierarchical structure (level 1) per time unit, n_1^a, is, on average, equal to the number of packets that are created in one branch of the network and have their destination in a different branch (see Fig. 1). Since the origin and the destination of the packets are chosen at random, from purely geometric considerations it is straightforward to estimate this number of packets per unit time as

$$n_1^a = p\left[\frac{z(z^{m-1} - 1)^2}{z^m - 1} + 1\right]. \qquad (4)$$

Within this mean-field approach, it can be easily shown that it is indeed the top node which is the most congested.

On the other hand, in our mean-field calculation q_{12} is the average probability that a given packet moves from a node in the second level to the top node and vice versa, and is given, as a first approximation, by $q_{12} = 1/\sqrt{n_1 n_2}$, where n_1 is the average number of packets at level 1 and n_2 is the average number of packets at each of the z nodes in the second level. Thus the average number of packets leaving the top at each time step will be $n_1^l = n_1 q_{12}$, and the average number of packets going from the z nodes in the second level to the top will be $n_1^a = z\alpha n_2 q_{12}$, where α stands for the fraction of packets in the second level that are trying to go up (some of the packets in level 2 are, of course, trying to go down to level 3).

At the critical point the top agent becomes collapsed and the communications between the first and the second level are much more congested than the communications between levels 2 and 3 so we can assume that $\alpha \approx 1$. At this point, by imposing the steady state condition $n_1^a = n_1^l$ we arrive to the relations $n_1 = zn_2$ and $n_1^a = \sqrt{z}$. Using Eq. (4) we obtain the final expression for p_c:

$$p_c = \frac{\sqrt{z}}{\frac{z(z^{m-1}-1)^2}{z^m-1} + 1}. \qquad (5)$$

Although strictly speaking the condition $\alpha = 1$ provides an upper bound to p_c, Eq. (5) is an excellent approximation for $z \geq 3$, as depicted in Fig. 2.

The critical total number of generated packets, $N_c = p_c S$, with S standing for the size of the system, can be approximated, for large enough values of z and m such that $z^{m-1} \gg 1$, by

$$N_c = \frac{z^{3/2}}{z - 1}, \qquad (6)$$

which is independent of the number of levels in the tree. It suggests that the behavior of the top node is affected only by the total number of packets arriving from each node of the second level, which is consistent with the mean-field hypothesis.

In order to characterize the transition, we introduce an order parameter:

$$\eta(p) = \lim_{t\to\infty} \frac{1}{pS} \frac{\langle \Delta N \rangle}{\Delta t}, \qquad (7)$$

where $\Delta N = N(t + \Delta t) - N(t)$ and $\langle \cdots \rangle$ indicates average over time windows of width Δt. Essentially, this order parameter represents the ratio between undelivered and generated packets at the stationary state. For $p > p_c$, the system collapses, $\langle \Delta N \rangle$ grows linearly with Δt, and thus η is a function of p only. For $p < p_c$, $\langle \Delta N \rangle = 0$ and $\eta = 0$. As observed in Fig. 3, and as may be expected from a properly chosen order parameter, when p is rescaled with p_c, the form of η does not depend on the details of the structure of the network, z and m.

Insofar as η does not depend on the structure of the network, we can study the simplest case $(1, 2)$ in order to obtain an analytical estimation of the order parameter. In this case, the network consist of only 2 nodes, 1 and 2, interchanging packets. Since from symmetry considerations $n_1 = n_2$, the maximum average number of delivered packets per time unit will be $(n_1 + n_2)/\sqrt{n_1 n_2} = 2$. Thus $p_c = 1$ and with the present formulation of the model it is not possible to achieve the supercritical regime. However, it is possible to extend p to be the average number of generated packets per node and time step and then p can be greater than one. In this case, for $p > p_c$ the number of packets delivered per time unit will be 2 while the number of generated packets will be $2p$. Thus

$$\eta = \frac{p - 1}{p}, \qquad (8)$$

in good agreement with simulated values (Fig. 3). In particular, near p_c we have

$$\eta \sim (p - p_c). \qquad (9)$$

Now let us consider the case where ξ_{ij} takes values uniformly distributed in $[0, 1]$. Even for very small values of p, a particular realization of the disorder can provoke

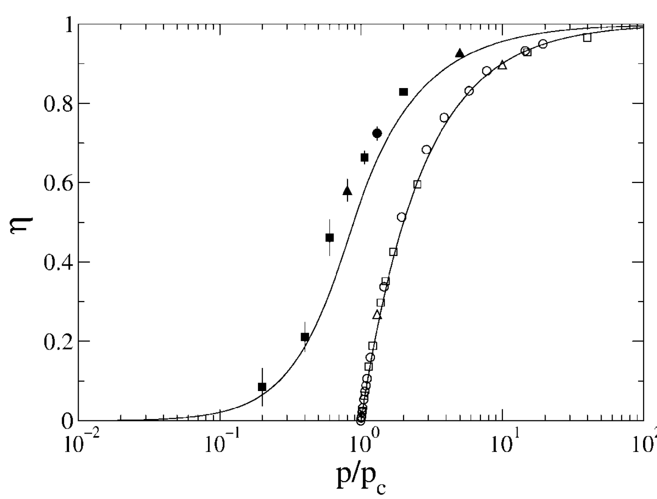

FIG. 3. Behavior of the order parameter in both cases with noise (filled symbols) and without noise (open symbols), for different structures: $(6, 7)$ (circles), $(3, 6)$ (squares), $(5, 4)$ (triangles). The lines represent analytical results obtained for the simplest case $(1, 2)$.

a very weak communication line and the congestion of the whole network. Thus there is no transition controlled by p. However, it is still possible to define the order parameter as in (7), just considering that the average $\langle \cdots \rangle$ has to be taken over time and over disorder realizations. As observed in Fig. 3, the existence of disorder destroys the phase transition acting as a random local magnetic field in a paramagnetic-ferromagnetic transition [14] and other physical systems [15].

Again, it is possible to obtain an analytical expression of the order parameter in the case of two nodes. As in the ordered case, the number of packets generated in a time step will be $2p$. Now, however, for a particular realization, ξ_{12}andξ_{21}, the maximum number of delivered packets will be $2\sqrt{\xi_{12}\xi_{21}}$. Thus, if $\xi_{12}\xi_{21} > p^2$ the system will reach the steady state and the configuration will not contribute to the order parameter, while if $\xi_{12}\xi_{21} < p^2$ the system will collapse and the contribution will be $\eta_{\xi_{12}\xi_{21}} = 1 - \sqrt{\xi_{12}\xi_{21}}/p$.

Thus we can define

$$\eta(p, \xi_{12}, \xi_{21}) = \begin{cases} 0 & \text{for } \xi_{12}\xi_{21} > p^2, \\ 1 - \sqrt{\xi_{12}\xi_{21}}/p & \text{for } \xi_{12}\xi_{21} < p^2, \end{cases} \tag{10}$$

and the order parameter will be given by the average over the random variables:

$$\eta(p) = \int_0^1 d\xi_{12} \int_0^1 d\xi_{21}\, \eta(p, \xi_{12}, \xi_{21}). \tag{11}$$

It is straightforward to obtain the result:

$$\eta(p) = \begin{cases} 1 - 4/(9p) & \text{for } p > 1, \\ (5p^2 - 3p^2 \ln p^2)/9 & \text{for } p < 1. \end{cases} \tag{12}$$

As depicted in Fig. 3, there is reasonable agreement between this analytical expression and the points obtained by simulation, always keeping in mind the simplicity of our approach.

Summarizing, we have studied a simple and general model of communication in a network with hierarchical branching. We have obtained some analytical results defining an order parameter and studying its behavior with respect to the relevant parameters of the model. The behavior of the system at the critical regime shows to be independent of the number of levels in the hierarchy. This phenomenon shows that the main features of information flow in a network with hierarchical branching are determined by the branching of the first level. Although we are in a very tentative stage of the model, we think that this result can help us to understand flow in real networks, where this effect can dominate the global behavior of the system. Another interesting issue is the scaling observed in Fig. 3. From the viewpoint of organizational design, this scaling can be used to forecast the behavior of the organization when increasing or decreasing its size. The inclusion of a quenched randomness accounting for different kinds of interaction is not a hindrance for our theoretical analysis and we give an accurate behavior of the order parameter in this situation. The approach presented here opens a line of research which will follow to study different dynamics and topologies.

The authors gratefully acknowledge F. Giralt, C. J. Pérez, F. Vega, and H. J. Witt for helpful comments. This work has been supported by DGES of the Spanish Government, Grants No. PPQ2000-1339, No. BFM2000-0626, and No. PB96-1025, and EU TMR Grant No. ERBFMRXCT980183. R. G. also acknowledges financial support from the Generalitat de Catalunya.

[1] *The Economy as an Evolving Complex System II,* edited by W. B. Arthur, S. N. Durlauf, and D. A. Lane (Addison-Wesley, Reading, 1997).

[2] R. Axelrod, *The Complexity of Cooperation* (Princeton University Press, Princeton, New Jersey, 1997).

[3] I. Csabai, J. Phys. A **27**, L417 (1994).

[4] A. Y. Tretyakov, H. Takayasu, and M. Takayasu, Physica (Amsterdam) **253A**, 315 (1998).

[5] M. Takayasu, H. Takayasu, and T. Sato, Physica (Amsterdam) **233A**, 824 (1996).

[6] T. Ohira and R. Sawatari, Phys. Rev. E **58**, 193 (1998).

[7] R. V. Solé and S. Valverde, Physica (Amsterdam) **289A**, 595 (2001).

[8] D. Chowdhury, L. Sanken, and A. Schadschneider, Phys. Rep. **329**, 199 (2000), and references therein.

[9] J. R. Banavar, A. Maritan, and A. Rinaldo, Nature (London) **399**, 130 (1999).

[10] R. Radner, Econometrica **61**, 1109 (1993).

[11] We consider $\xi_{\alpha\beta} \neq \xi_{\beta\alpha}$. Conceptually, our election is more adequate for systems where agents determine the efficiency of the communication process (as in organizations) while the symmetric case $\xi_{\alpha\beta} = \xi_{\beta\alpha}$ would be more adequate for systems where the channel does so (as in computer networks traffic). Although the particular results obtained would be different in this case, the qualitative behavior of the system would be the same and similar calculations should be performed.

[12] A. Arenas, A. Diaz-Guilera, and R. Guimerà (unpublished).

[13] D. Stauffer and A. Aharony, *Introduction to Percolation Theory* (Taylor & Francis, London, 1992).

[14] J. Villain, Phys. Rev. Lett. **52**, 1547 (1984).

[15] A. Arenas and C. J. Perez, Physica (Amsterdam) **201A**, 614 (1993).