Computational approach to organizational design

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

[1] *Departament d'Enginyeria Informàtica, Universitat Rovira i Virgili, Tarragona, Spain*

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

[3] *Departament d'Enginyeria Química, Universitat Rovira i Virgili, Tarragona, Spain*

The relevance of information flow and processing in organizations have been stressed. In this work we propose a simple agent-based model of communication in hierarchical networks and study both computationally and analitically its behavior. The results obtained are used to identify efficient organizatinal structures.

I. INTRODUCTION

Nowadays, a lot of attention is paid by physicist to the dynamics of complex social and economic systems. In particular, many challenging questions have arisen concerning the influence of the topology and the interaction processes in the behavior of such systems [1, 2]. Our interest is focused on the behavior of hierarchical structures formed by agents (or element, in general) that interact with each other via communication processes. This framework is especially adequate to study for instance packet flow in computer networks as the Internet [3–5], traffic networks [6], river networks [7] and particularly communication flows in organizations [8–12].

Using Radner's words [8]:

> *The typical U.S. company is so large that a substantial part of its workforce is devoted to information-processing, rather than to "making" or "selling" things in the narrow sense. Although precise definitions and data are not available, a reasonable estimate is that more than one-half of U.S. workers (including managers) do information-processing as their primary activity.*

In this work, we propose and study a very simple model of communication in a hierarchical network [13]. 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 with finite capacity to transmit packets. Despite the simplicity, the model reproduces the main characteristics of the flow of information packets in a network. We observe a phase transition between a sparse and a congested regime when the number of packets to deliver reaches a critical value (for a general reference about phase transitions and critical phenomena see [14]). Near the transition point signs of criticality arise in agreement with reported empirical data [3]. On the other hand, the model is simple enough to allow analytical characterization: we provide a mean field estimation of the critical point in good agreement with simulation results. A first step towards a more realistic model is to include the cost of stablishing communication lines.

In both cases, with and without cost associated to the links, the optimal organizational structures are studied.

II. THE MODEL

The model is defined as follows: the organization is mapped into a lattice where nodes represent the communicating agents (employees) and the links between them represent communication lines. In particular, we use hierarchical trees as depicted if Fig. 1. 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.

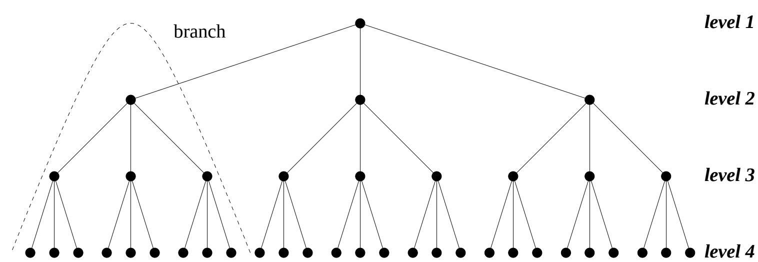

FIG. 1: Typical hierarchical tree structure used for simulations and calculations: in particular, it is a tree $(3, 4)$. Dashed line: definition of *branch*.

The dynamics of the model evolves according to the following scheme. At each time step t, an information packet is created by every agent with probability p. When a new packet is created, a destination agent, different from the origin agent, is chosen at random in the network. Thus, during the following time steps $t, t+1, \dots, t+T$, the packet is traveling towards its destination: once the problem reaches this destination agent, it is delivered and disappears from the network. When an agent receives a packet, she knows whether the destination is to be found somewhere below him. If so, she directs the problem downwards in the right direction. Otherwise, she transmits it upward to the agent overseeing him. Thus, the information packets move towards their destination following the shortest path. The time a packet remains in the network is related not only to the distance between the origin and the destination agents, but also to the amount of packets in the network. In particular, at each time step, all the packets move from their current position, i, to the next agent in their path, j, with a probability q_{ij}. We define q_{ij}, *quality of communication* between agents i and j, as

$$q_{ij} = \sqrt{k_i k_j}. \tag{1}$$

where k_α represents the capability of agent α to communicate at each time step. For k_α we propose:

$$k_\alpha = Q_L(c_\alpha) f(n_\alpha) \tag{2}$$

where c_α is the number of links of agent α, $0 < Q_L(c) \leq 1$ is a cost factor related to these links (note that, the higher the number of links, the smaller Q_L, so Q_L is a monotonically decreasing function of its argument), L is the *linking capability* that tunes the magnitude of this cost, n_α is the total number of packets currently at agent α, and $0 < f(n) \leq 1$ is the function that determines how the capability of a particular agent decreases when the number of information packets to handle grows (again, $f(n)$ is a decreasing function of the argument). For the functions $Q_L(c)$ and $f(n)$ we chose the following cies:

$$Q_L(c) = 1 - \tanh \frac{c}{L}, \tag{3}$$

and

$$f(n) = \begin{cases} 1 & \text{for } \; n = 0 \\ 1/n & \text{for } \; n = 1, 2, 3, \ldots \end{cases} \tag{4}$$

The election of Q_L is completely arbitrary but (3) has two desirable properties: (i) it is a monotonically decreasing strictly positive function and (ii) Q_L decreases linearly for small values of Q_L (compared to L). The parameter L tunes how fast the capability decreases when new links are added: for large values of L (big capabilities) this decreasing is slow while for small L it is very fast (Fig. 2). Different elections of this cost factor would lead to results qualitatively similar to the ones reported in the following sections.

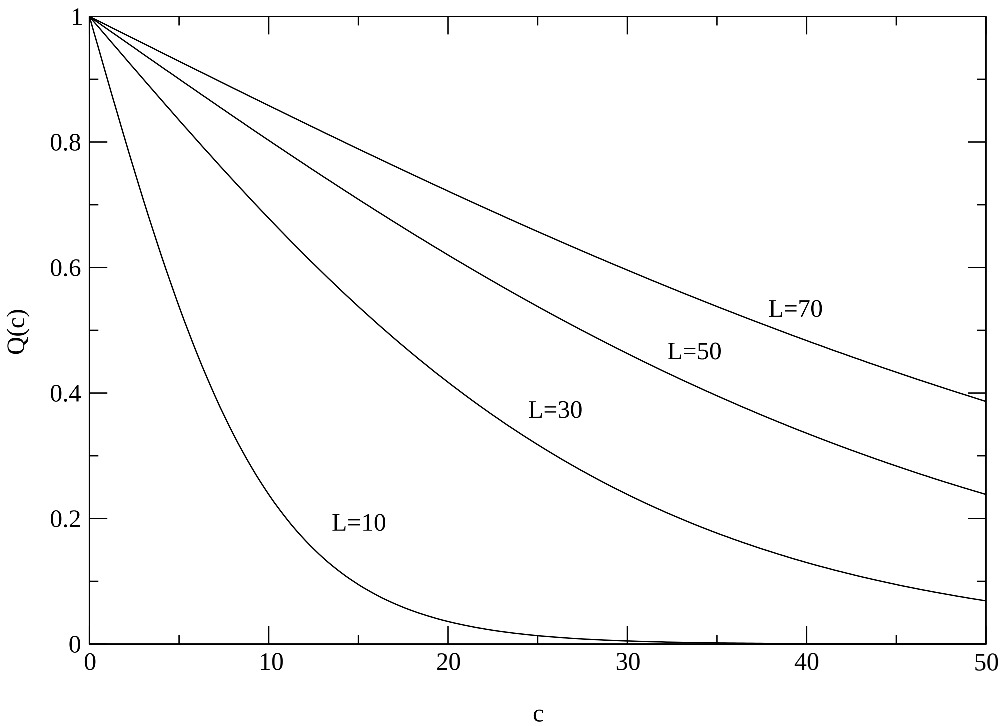

FIG. 2: Behavior of the cost factor for different values of the linking capability L.

The election of $f(n)$ is more delicate but again it seems plausible that, for instance, the capability is reduced to one half when the amount of information to deliver is double. Furthermore, this election is consistent with existing queueing models for information flow in computer systems such as the Internet [5].

III. RESULTS AND DISCUSSION

As a first step, let us concentrate in the simpler case $L \to \infty$, i.e. costless connections. The probability of generating a packet per agent and time unit, p, is an exogenous parameter that controls the behavior of the system. For small values of p, all the packets are delivered and so, after a transient, the system reaches a steady state in which the total number of packets, N, fluctuates around a constant value, i.e. the number of delivered packets is equal, on average, to the number of generated packets. However, for large values of p, not all the packets can be delivered, and N grows in time without limit. The transition between one regime and the other occurs for a critical value of p, p_c. For values of p smaller than but close to p_c, the steady state is reached but large fluctuations with long correlation times appear. These three behaviors can be observed in Fig. 3.

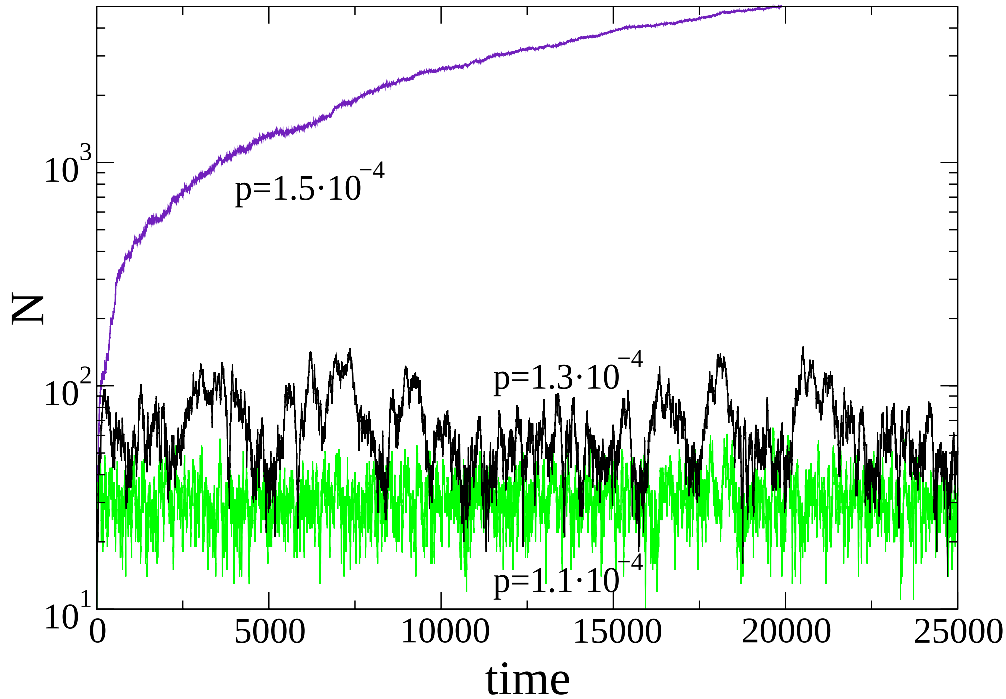

FIG. 3: Evolution of the number of packets in a network for regimes below ($p = 1.1 \cdot 10^{-4}$), above ($p = 1.5 \cdot 10^{-4}$) and near ($p = 1.3 \cdot 10^{-4}$) the critical value, p_c. These results correspond to simulations performed in a tree $(7, 5)$.

It is possible to give an analytical estimation of p_c. Within a mean field approach, it is if we do not consider fluctuations and we assume that the behavior of all the agents in the same level is statistically identical, we arrive to the following expression for p_c (see Appendix A)

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

For values of z and m such that $z^{m-1} \gg 1$ (note that this condition is satisfied even for relatively small values of z and m), this expression can be approximated by

$$p_c \approx z^{3/2-m}. \tag{6}$$

Although strictly speaking (5) (and its approximation (6)) provides an upper bound to p_c, it is an excellent estimation for $z \geq 4$, as can be seen in Fig. 4.

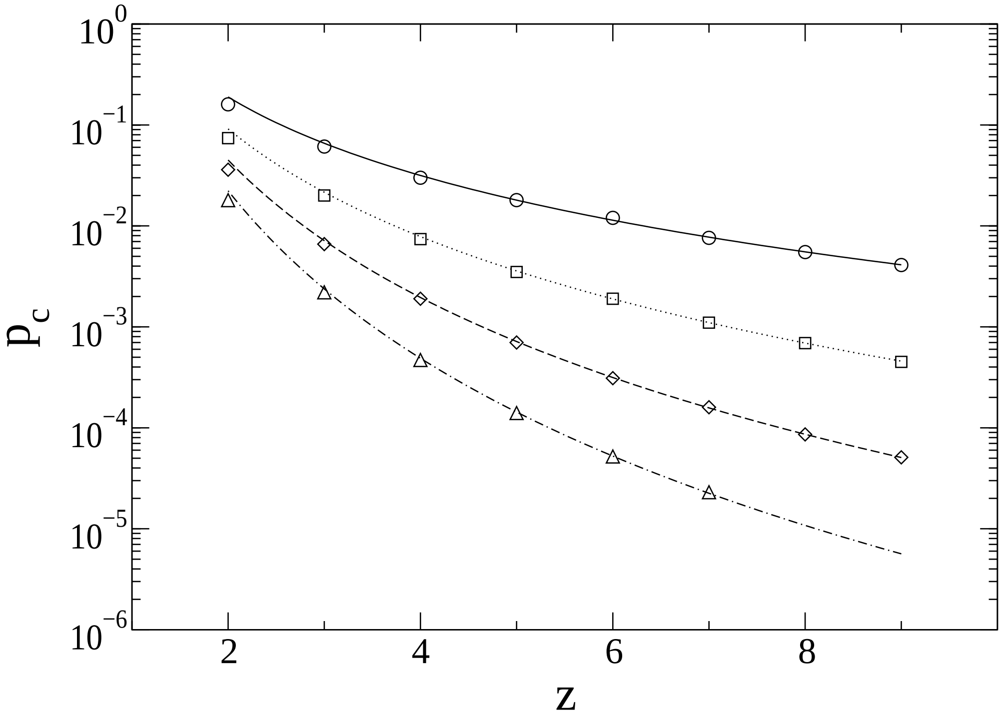

FIG. 4: Comparison between simulated (symbols) and analytical (lines) values for the critical probability of packet generation, p_c as a function of the branching factor z for hierarchies with different number of levels: $m = 4$ (circles and full line), $m = 5$ (squares and dotted line), $m = 6$ (diamonds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

It is interesting to note, from (5), that the maximum number of information packets that can be generated in a time step without collapsing the organization, $N_c = p_c S$, with S standing for the size of the organization, is given by

$$N_c = \frac{\sqrt{z}}{\frac{z(z^{m-1}-1)^2}{z^m-1}+1}\frac{z^m-1}{z-1} \approx \frac{z^{3/2}}{z-1} \tag{7}$$

again with the same approximation as in (6). Thus the total number of packets a network can deal with does not depend on the number of hierarchical levels. This fact is verified by simulations. Furthermore N_c is a monotonically increasing function of z, suggesting that, fixed the number of agents in the organization, S, the optimal organizational structure, understood as the structure with higger capacity to handle information, is the flattest one, with $m = 2$ and $z = S - 1$.

However, from a practical point of view this structure is not possible: an organization with 10,000 employees, for instance, cannot be organized in only two hierarchical levels, since it is impossible to mantain such a enormous number of communication lines. Thus, it is necessary to introduce the cost for establishing links in order to get a more realistic picture of the problem. In this case, following arguments analogous to that used in the case of costless connections, we can arrive to the following expression for p_c:

$$p_c = \frac{\sqrt{zQ_L(z)Q_L(z+1)}}{\frac{z(z^{m-1}-1)^2}{z^m-1} + 1}. \tag{8}$$

Again, for z and m such that $z^{m-1} \gg 1$, the maximum number of packets that can be generated per time step without collapsing the system is independent of m, and is given by

$$N_c \approx \frac{z^{3/2}(Q_L(z)Q_L(z-1))^{1/2}}{z-1}. \tag{9}$$

As can be seen from Fig. 5, the scenario that arises with the introduction of the cost factor is much more interesting. Now, the cost term (which is a decreasing function) compete with the behavior we have found for the critical number of generated packets, N_c, in the case of costless connections. Thus, a maximum typically arises in N_c. This maximum is related to an optimum value of z, z^*, which in turn defines an optimal organizational structure different from the trivial $m = 2$ and $z = S - 1$.

IV. CONCLUSION

A very simple model for dealing with the problem of communication and information flow in organizations has been introduced. The model considers agents, organized in a hierarchical tree-like structure, which interchange information packets (that can be understood in the most general sense) following a simple set of rules: we define a communication capability for each agent which is decreased if the quantity of information to handle grows. We observe that the system can show two qualitatively different behaviors. When the amount of information to handle is below a certain critical value, all the packets are delivered to its destination. However, when the amount of packets to deliver reaches this critical threshold value, the agents in the organization get collapsed and a certain amount of information never reaches its destination.

The simplicity of the model allows an estimation of the critical values, i.e. for a given hierarchical structure it is possible to calculate the maximum amount of information packets that can be generated at each time step without collapsing the network. For the simplest case in which no

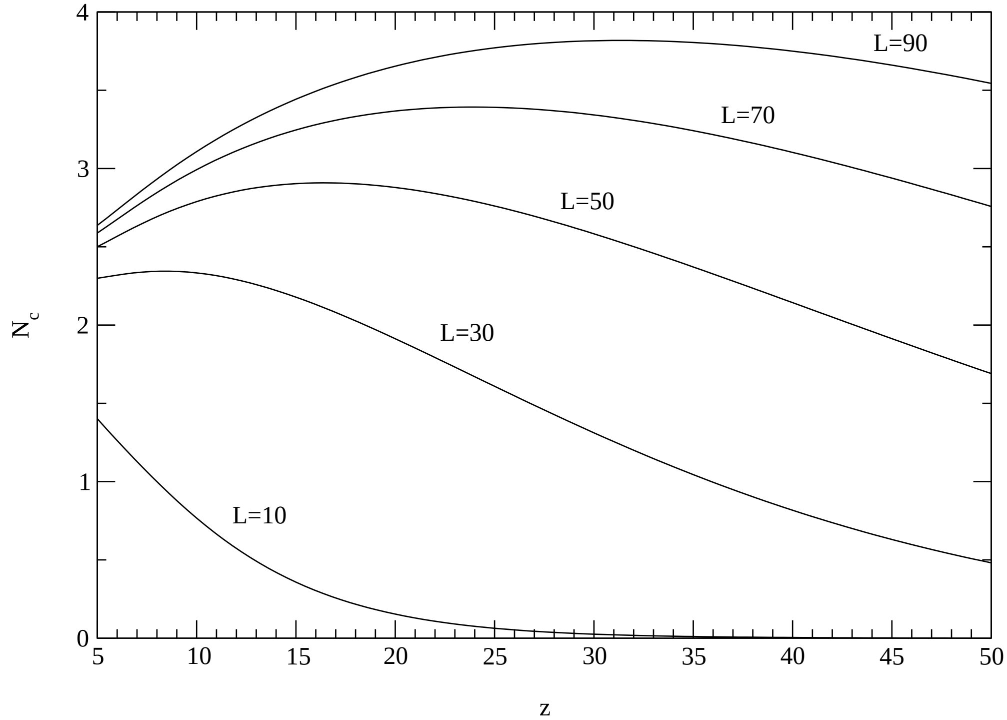

FIG. 5: Maximum number of packets that can be generated in an organization per time unit without collapsing it, plotted as a function of z. Different curves correspond to different values of the linking capability, L.

cost is associated to the existence of communication lines, two main features are observed: (i) the maximum number of packets per time unit the organization can deal with does not depend on the number of levels in the hierarchical structure and (ii) this critical number of packets is a monotonically increasing function of the branching factor (span of control), thus suggesting that, for a fixed size of the organization, the optimal organizational structure is the flattest one, with only two levels. Optimality is defined as maximum capacity to deal with information.

A different scenario arises when the more realistic situation of costly connections is considered by introducing a cost factor in the definition of agents capability. Although as in the previous case, the maximum amount of information the organization is able to handle does not depend on the number of levels, it is not a monotonically increasing function of the branching factor. Thus the flattest structure is not the best in general. Actually, the steepness of the optimal organization structure is tuned by the *intensity* of the cost factor. As may be expected, the higher the cost of the connections, the steeper the optimal structure and vice versa.

The extension of our model to more general conditions and topologies is easy and new challenging situations arise. The introduction of *lateral connections* or shortcuts in our starting point hierarchical structure, for instance, seems to be helpful in achieving better performances. We think that the approach presented here opens a promising line of research which will follow to study more complicated dynamics and topologies.

APPENDIX A: CALCULATION OF THE CRITICAL PROBABILITY OF PACKET GENERATION

As happens in other problems is statistical physics [15], the particular symmetry of the hierarchical tree allows a mean field estimation of the critical point p_c. Since in the steady state regime there is no accumulation of packets, the number of packets arriving to 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:

$$n_1^a = p\left(\frac{z\left(z^{m-1}-1\right)^2}{z^m-1}+1\right). \tag{A1}$$

Within this mean field approach, it can be easily shown that this top agent 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 an agent in the second level to the top agent 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 one and n_2 is the average number of packets at each of the z agents 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 agents 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, the communications between the first and the second level are much more congested than the communications between levels 2 and 3 and we can assume that $\alpha \approx 1$. With this and 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 equation (A1) we obtain the final expression for p_c:

$$p_c = \frac{\sqrt{z}}{\frac{z(z^{m-1}-1)^2}{z^m-1} + 1} \quad \text{(A2)}$$

REFERENCES

[1] D.J. Watts and D.H. Strogatz, Nature **393**, 440 (1998).

[2] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).

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

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

[5] T. Ohira and R. Sawatari, Phys. Rev. E **58** (1), 193 (1998).

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

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

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

[9] B. Watkins, Santa Fe Institue Working Paper 98-06-055 E (1999).

[10] A. Creti, CEPR Conference: Information Processing Organizations, Brussels (1999).

[11] T. van Zandt, Contemporary Economic Issues, Vol 4: Economic design and behavior, MacMillan Press Ltd., 125 (1999).

[12] A. Zanatta and F. Luna, Procedings of the WHEIA'99, Genoa (1999).

[13] A. Arenas, A. Diaz-Guilera, R. Guimerà, (submited).

[14] H.E. Stanley, *Introduction to phase transitions and critical phenomena*, Oxford University Press (1987).

[15] D. Stauffer, A. Aharony, *Introduction to percolation theory*, Taylor & Francis (1992).