Relation between structure and size in social networks

Luis López and Miguel A. F. Sanjuán
Nonlinear Dynamics and Chaos Group, Departamento de Ciencias Experimentales e Ingeniería, Universidad Rey Juan Carlos, Tulipán s/n, 28933 Móstoles, Madrid, Spain
(Received 10 August 2001; published 8 February 2002)

In the context of complex network systems, we model social networks with the property that there is certain degradation of the information flowing through the network. We analyze different kinds of networks, from regular lattices to random graphs. We define an *average coordination degree* for the network, which can be associated with a certain notion of efficiency. Assuming that there is a limit to the information a person may handle, we show that there exists a close relationship between the structure of the network and its maximum size.

DOI: 10.1103/PhysRevE.65.036107 PACS number(s): 05.10.−a, 89.65.−s, 89.75.−k

I. INTRODUCTION

The study and characterization of complex systems is a very fruitful research area nowadays with many interesting open problems. Special attention has been paid recently to complex networks, where graph and network analysis plays an important role and is gaining great popularity due to its intrinsic power to reduce a particular system to its simple components and relationships. It is perhaps this reductionism that allows network characterizations to be present in different scientific and technological disciplines such as neurobiology [1–4], the Internet [5], the World Wide Web [6,7], finance [8], etc. Moreover, many physicists have focused their research interests on complex networks, as can be noticed by the number of papers that have appeared in physics literature [9–14] in the past few years. Most of the recent efforts in the understanding of these complex networks are reviewed in Ref. [15].

Among complex networks, social networks appear in a quite natural way, and as any other complex system, they can be analyzed in the framework of graph theory [12,16]. A graph G consists of a nonempty set of elements, called vertices, and a list of unordered pairs of these elements, called edges. If i and j are vertices of G, then an edge of the form (i,j) is said to connect i and j. Many interesting complex systems are built out of simple components that maintain relationships among them. The representation of those systems through a graph is rather straightforward considering each simple component to be a vertex and representing the relationships as edges among them.

Any social structure is composed of different types of elements such as human beings, groups of people, nations, etc., which are linked together following some rules that define the existence and degree of the relationships among them. A very well known example of a social network is the Kevin Bacon game developed by Brett Tjaden and studied thoroughly by Watts [17] in the context of the small world phenomenon. In this model each actor or actress is considered to be a vertex on the graph, two vertices are connected through an edge only if they have ever been in a film together. Another interesting example was developed by Newman [12] when studying the scientific collaboration networks. In this case, each vertex of the resulting graph represents a particular scientific author, two authors are connected when they have coauthored one or more papers together. Other examples are related to the web of human sexual contacts [19] and to the small world phenomenon [20,21].

Using just two ingredients: the elements, represented by the vertices, and the relationships, represented by the edges, it is possible to define any kind of social structure. In this paper we are particularly interested in studying the *collaboration* that emerges in social networks that describe human communities or organizations. We represent any social network by a simple graph where each person is a vertex and the relationships are the edges. For simplicity, we assume some of the restrictions imposed by Watts in his study about the small world phenomenon [17] to describe the relationships. This means that the kinds of edges used in our model are undirectional edges, implying symmetric relationships; unweighted edges, implying that any edge is not assigned any *a priori* strength; simplicity, implying that multiple edges between the same pair of vertices or edges connecting a vertex to itself are forbidden.

The structure of the paper is as follows. First we develop a new model for the spread of information in social communities, taking into account the degradation of the information that exists in a real social scenario. Then we show that, assuming that the amount of information a person may handle is limited, there exists a close relationship between the topology of the network and its maximum size.

II. MODELING SOCIAL ACTIVITY

Traditionally the research in graph theory has been concentrated in modeling the spread of information from one vertex to the rest of a graph considering that the information can travel through edges without degradation in the traveling process. This approach has been very useful for modeling some particular types of phenomena like disease spread in a social community [22–24] or virus infection and error propagation in computer networks [25]. Nevertheless this is not appropriate when trying to model the kinds of processes that take place in collaborative social networks.

Social networks can be of very different natures (an organization, a company, an association, a religious congregation, etc.). Our interest is concentrated on communities where the relationships are established through the interchange of certain *information*. Information here should be understood in a broader sense. For example, in some social communities based on friendship, people may help each other, so the information can take the form of effort, time, money, etc. In companies the relationship might be considered to have a very important professional component so that the information can take place in other different ways.

Experience tells us that the information obtained from a person is extremely dependent on the degree of relationship maintained with that particular individual. As we want to represent social relationships through unweighted undirected edges, we define the degree of the relationship between two persons as the distance, in number of hops, from the vertex representing one person to the other one. Thus, if we are related in first order with a particular person we may easily obtain a lot of information. If we think about a second-order relationship, for example, a friend of one of our friends, the quantity of information we can get is lower than for one of our closest friends, and so on.

In order to create a model for this particular situation we define a quantity that we call the *coordination degree*. The coordination degree measures the ability of the vertices in a graph to interchange information. There are several ways in which we can model this magnitude. One of the easiest is to consider the coordination degree to be exponentially related with the distance between the vertices. In this way, we define the coordination degree γ_{ij} between two vertices i and j as

$$\gamma_{ij}=e^{-\xi d_{ij}}, \tag{1}$$

where d_{ij} is the distance between the two vertices and ξ is a real positive constant, measuring the strength of the relationship, which we call the *coordination strength*.

Quantities similar to the coordination degree have already been discussed in the literature. The most remarkable work in this field is the one by Katz [26], where the author considers the sum of $e^{-\xi d_{ij}}$ over all paths to a particular vertex. This kind of approach is indeed more realistic than ours, because it considers that the information may travel following all possible paths, and not only the shortest paths. Unfortunately, this type of measure can only be expressed in terms of the adjacency matrix of the graph, making the analysis and computations much more complex. For clarity, in this paper we consider, as a reasonable approximation, that the main part of the information travels along the shortest paths.

It is important to remark that a different coordination strength could be considered for each edge of the graph, but this would point to weighted graphs, which we do not consider here. So as a first approximation, ξ is considered to be a constant for each particular graph.

Accepting these assumptions, we can define the *total coordination degree* of a vertex i in a graph as the sum of all the coordination degrees between that particular vertex and the rest,

$$\Gamma_i=\sum_{j=1}^{N} \gamma_{ij}, \tag{2}$$

where N is the order of the graph (the total number of vertices in that particular graph). It is interesting to remark that this definition includes the coordination degree of a node with itself. Following Eq. (1) the coordination degree of a node with itself must have a value of 1, because the distance between the node with itself d_{ii} is zero, and consequently $\gamma_{ii}=e^{-\xi d_{ii}}=e^0=1$. The total coordination degree of a vertex is a measure of the amount of information the vertex is able to receive belonging to that particular network.

In the same way we define the total coordination degree of the graph as the sum of the total coordination degrees of all the vertices belonging to the graph,

$$\Gamma=\sum_{i=1}^{N} \Gamma_i . \tag{3}$$

The total coordination degree of a graph is a measure of the amount of information that is handled in an organization. More interesting than the total coordination degree of a graph is the *average coordination degree* of the graph, which we define as the total coordination degree of the graph divided by its order

$$\overline{\Gamma}=\frac{\sum_{i=1}^{N} \Gamma_i}{N}. \tag{4}$$

This allows us to give an interesting interpretation of the average coordination degree of a graph as a measure of the efficiency of a particular community or organization, because it suggests how much the individual contributes to the community.

As a basic ingredient of our model, it is important to remark on the common perception that the number of close relationships a person may have within a community is necessarily limited to a quite small number, independent of the type of organization. For example, if we think about our circle of friends we can see that the number of people we feel really connected with is normally not higher than five or six. A similar conclusion might be obtained regarding our workplaces; the number of people we can consider to be really coordinated with is not usually higher than half a dozen. This might be the consequence of the fact that establishing close relationships with people is normally very time consuming, and time is a limited resource for every individual.

III. COORDINATION IN DIFFERENT GRAPH STRUCTURES

Once we have defined the average coordination degree, it would be interesting to investigate how the relational structure and size of the organizations can have an influence on it. As the number of graph families is extremely large, we must choose some restrictions to be able to perform an exhaustive analysis over the most characteristic ones. The first con-

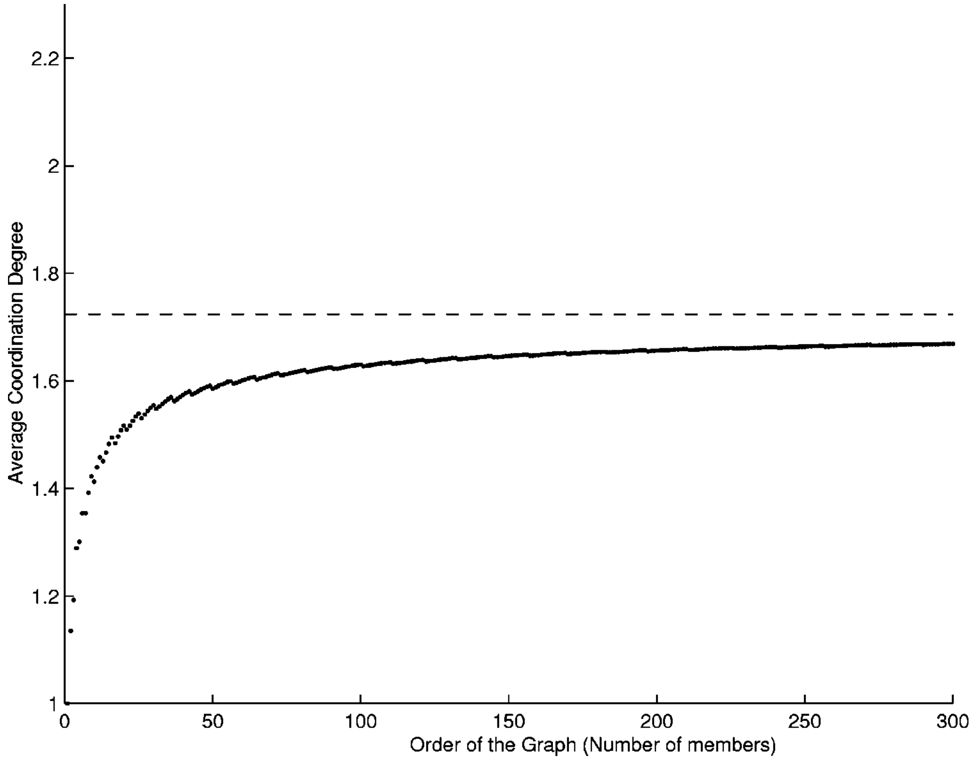

FIG. 1. Average coordination degree $\overline{\Gamma}$ for a regular 2D lattice with $k=4$ and $\xi=2$. The computations have been performed for graph orders varying from $N=1$ to $N=300$. The dashed line represents the upper limit Γ_∞ for this choice of parameters k and ξ. It is important to remark that we have used nonperiodic 2D lattices for the simulations. Notice that for graphs of order less than $N=4$ the degree of the graph k is equal to N.

straint that might be imposed is related with the mean value of the number of relationships a person can have, in other words the mean number of edges starting from a vertex in a particular graph. This is usually called the average degree of the graph. Following the notation by Watts, we represent this magnitude by the letter k. As discussed before, the number of relationships a person may have is limited to a quite small number, hence we may consider only graphs having a small average. For simplicity, in this work, the average degree is considered to be 4, which allows to deal easily with regular two-dimensional (2D) lattices. Analysis with $k=5$ and $k=6$ would give similar conclusions.

Another important point of the model is the coordination strength ξ, which we defined previously in Eq. (1). This parameter measures how the strength of the relationship decreases with the distance between the vertices. It is logical to think that this might depend strongly on the nature of the relationships and on the type of organization and its members. Nevertheless it is out of our objective to discuss the psychological and sociological aspects of this problem. For simplicity, in all the simulations presented in this paper the coordination strength has been fixed to the value of $\xi=2$. Although another different value could have been chosen, this particular one gives reasonable conclusions, because the coordination degree supplied by any first-order relationship is $\Gamma_1=e^{-2}\simeq 0.135$, which means an increase of approximately 13% in the information associated with each particular isolated vertex.

The next step is to determine the different graph structures to be analyzed. A first natural choice is regular lattices, then we consider a 2D lattice of degree $k=4$. If we call η_i the number of neighbors of degree i, for large lattices (assuming the order of the graph to be $N=\infty$) the distribution of η is described in the following way: $\eta_0=1$, $\eta_i=4i$. Then, using Eq. (3), the total coordination degree for a particular vertex i is given by

$$\Gamma_i = 1 + 4\sum_{j=1}^{\infty} j e^{-j\xi}, \tag{5}$$

which can be easily evaluated as the derivative of a geometric progression to the value

$$\Gamma_i = 1 + \frac{4e^{-\xi}}{(1-e^{-\xi})^2}. \tag{6}$$

As we are considering an infinite regular lattice, it can be proven that for all vertices in the graph $\Gamma_i=\overline{\Gamma}$ (remark that Γ_i does not depend on i). We call this value Γ_∞. It can be easily shown that Γ_∞ is an upper limit for the average coordination degree for any 2D lattice with $k=4$. We have carried out extensive numerical simulations over nonperiodic regular lattices that are shown in Fig. 1. This figure shows that as the order of the graph increases, the average coordination degree asymptotically increases towards Γ_∞. A remarkable observation is that after a certain value of the order of the graph the increase of the average coordination degree becomes very small. In terms of the social network we try to describe, this admits an interesting interpretation. The efficiency that might be seen as the average coordination degree does not show a considerable increase once the organization reaches a certain size.

After having analyzed regular lattices, we attempt to study how the average coordination degree changes when the randomness on the graph is increased. Adopting a similar

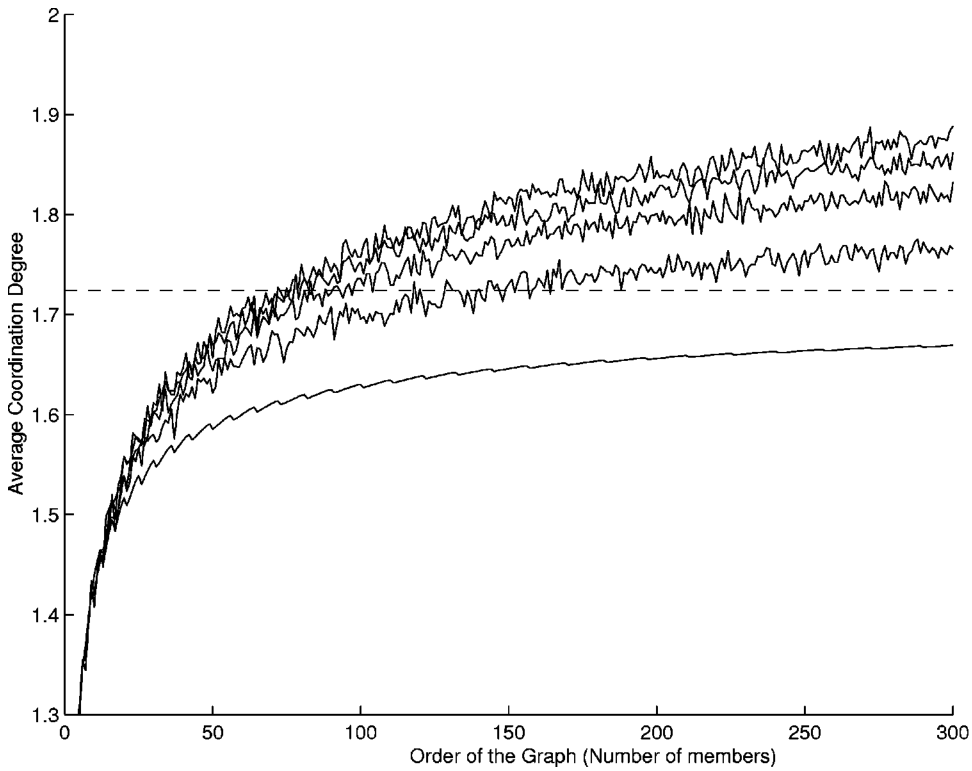

FIG. 2. Average coordination degree $\overline{\Gamma}$ for a pseudoregular 2D lattice with $k=4$ and $\xi=2$. The randomness parameter α takes the values, from the lower curve to the upper curve, 0, 0.25, 0.5, 0.75, and 0.99, respectively. The computations have been performed for graph orders varying from $N=1$ to $N=300$. The dashed line represents the upper limit Γ_∞ for regular 2D lattices. Nonperiodic 2D lattices have been also used here.

approach to that followed by Watts and Strogatz [16,17], we define a family of graphs based on the regular 2D lattice following this simple algorithm.

(i) Choose the order of a particular graph N and build a regular nonperiodic 2D lattice with N vertices called G.

(ii) Choose a parameter α measuring the randomness of the new graph.

(iii) For each edge in G, eliminate that edge with probability α and rewire the vertices involved in a random way. We now obtain a new graph G_α that has the same number or vertices and edges as G, but connected in an irregular way that depends on the value of α.

The question that arises immediately is how the *average coordination degree* depends on α. When $\alpha=0$, $G=G_\alpha$ and the graph is a regular nonperiodic 2D lattice. This situation was described in Fig. 1. If we raise the value of the randomness parameter α the graph becomes more and more random. Numerical simulations for different randomness parameters α have been performed, and we show the results in Fig. 2, where it is shown that the average coordination degree increases when the value of α increases. It is also interesting to remark that for certain values of α, the average coordination degree surpasses the Γ_∞ limit of the regular lattice. In order to offer a better understanding of the role played by the randomness parameter α, we have used a graph with 300 nodes, and made different simulations for various values of α. The result is shown in Fig. 3, from which it can be derived that the variation of $\overline{\Gamma}$ with α follows a monotonic nonlinear function.

As a natural sequence of the previous analysis, we consider now the behavior of the average coordination degree for a completely random graph. For building such a random graph we may consider the traditional Erdös-Rényi model [18]. Starting with N nodes, every pair of nodes is connected with probability p and not connected with probability $1-p$. We must also consider the constraint imposed before, that is, the average degree of the graph must be equal to 4. In this situation, it can be easily proven that

$$p=\frac{4}{N-1}. \tag{7}$$

Using this result we can build a random graph for an arbitrary value of N, assuming an average distribution degree of 4. As we did previously with 2D lattices, we can try to find an upper limit for the average coordination degree of a random graph with $k=4$. It is easy to prove that the maximum value for the average coordination degree will be reached when a perfectly expanding graph is considered. A perfectly expanding graph is one that verifies that starting at any given vertex i, all m-order neighbors are always new unknown nodes (are not neighbors of i in any order smaller than m). For an infinite perfectly expanding graph the average coordination degree can be calculated as

$$\overline{\Gamma}=\sum_{i=1}^{\infty}(ke^{-\xi})^i, \tag{8}$$

which can be proven to be equal to

$$\overline{\Gamma}=\frac{1}{1-ke^{-\xi}}=\Gamma_\infty^R, \tag{9}$$

when $ke^{-\xi}<1$, and ∞ otherwise. In our case, $k=4$ and $\xi=2$, thus Γ_∞^R is a finite upper limit for the value of the coordination for any graph with the same average distribution

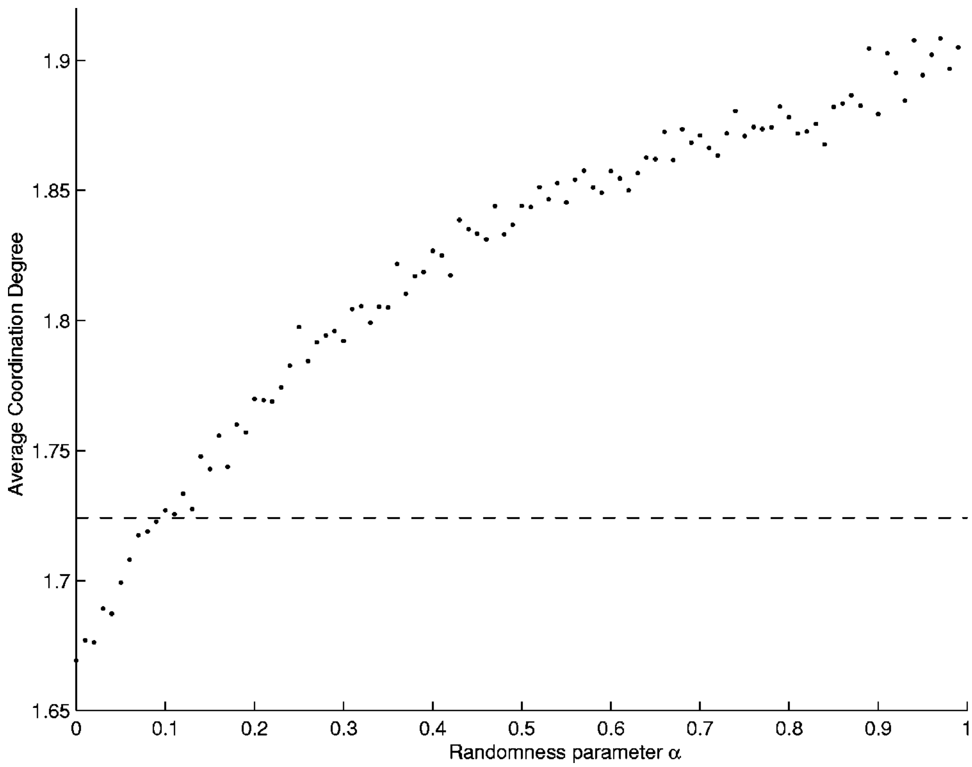

FIG. 3. Average coordination degree $\overline{\Gamma}$ as a function of the randomness parameter α in pseudo-regular 2D lattices with $k=4$ and $\xi=2$. The computations have been performed for a fixed graph of order $N=300$.

degree. In Fig. 4 a simulation performed over 500 different random graphs with $k=4$ is shown. It can be seen that the value of the average coordination degree is always under the limit Γ_∞^R. This result and the results obtained in the analysis of the other types of graphs, give evidence that the same conclusion might be obtained independently of the structure of the graph. Consequently we can conclude that the key point lies in the hypothesis used in our model, concerning the degradation of the information. One interesting feature that can be seen in Fig. 4 is the rather slow convergence of the coordination degree to the asymptote. In fact, it can be proven that it follows a power law with logarithmic corrections. The power law behavior can be immediately recovered by truncating the sum at a fixed n,

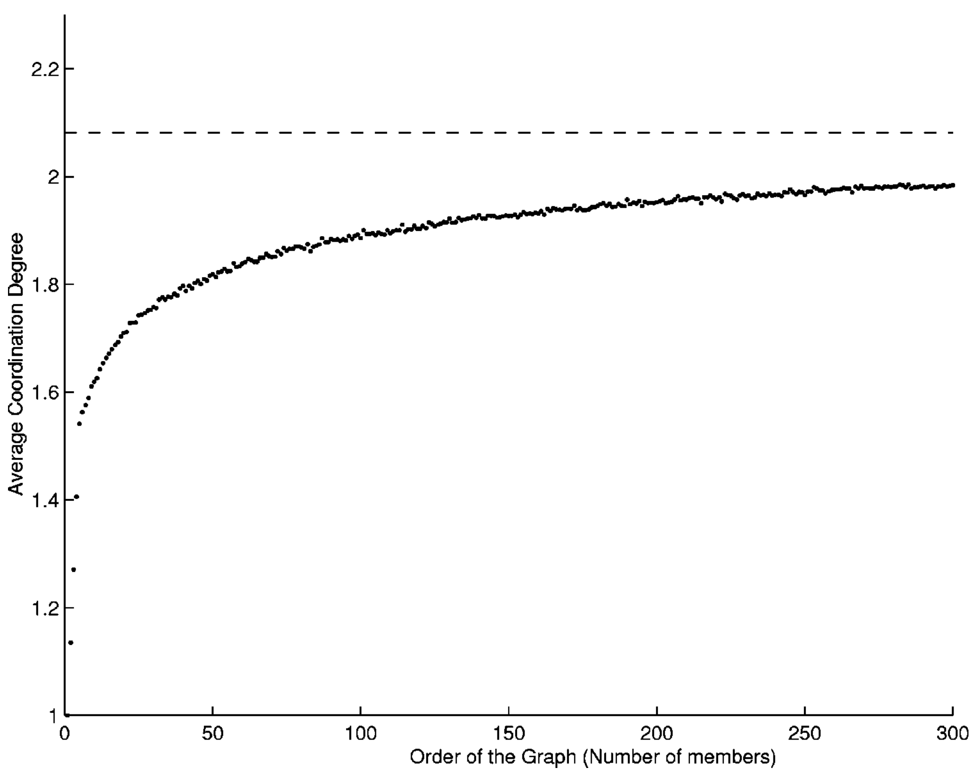

FIG. 4. Average coordination degree $\overline{\Gamma}$ for random graphs with average degree $k=4$ and $\xi=2$. The computations have been performed for graph orders varying from $N=1$ to $N=300$. The dashed line represents the upper limit Γ_∞^R for infinite perfectly expanding trees with $k=4$ and $\xi=2$.

$$\overline{\Gamma}_n=\frac{(ke^{-\xi})^n-1}{ke^{-\xi}}ke^{-\xi}, \tag{10}$$

where n is the diameter of the graph and $n=\log_k N$.

One of the most interesting challenges that this model suggests is the solution of the following problem: given a set of N nodes, an average degree k, and a value for the coordination strength ξ, try to find a topological structure for a graph connecting them so that the average coordination degree for the graph is the maximum of all possible graphs of the same N, k, and ξ. A fast and wrong answer could be to assume the graph to be a k-regular tree. As we know, an infinite tree is a perfect expanding graph, so that the average coordination for such a tree must be equal to the limit Γ_∞^R. Unfortunately when the graph is considered to be finite, the tree is not a perfect expanding graph anymore. There is a set of finite perfectly expanding graphs called Moore graphs [28], which are the most efficient k-regular graphs in the sense that every vertex reaches $k-1$ new vertices at a time, but it has been proven [27] that these kinds of graphs are unrealizable except for some very particular cases. So the question of finding the graph that maximizes the average coordination degree for a given number of nodes still remains open.

IV. THE 150 LIMIT

In relation to the preceding discussion, it is interesting to notice that some scientists propose the existence of a natural limit for the maximum number of members of a social group. Probably the most important work in this direction is the one carried out by the British anthropologist Dunbar [29], who related the size of the neocortex (a part of the brain related to social and language capabilities) and the maximum group size for primates. When applying this relation for the *Homo sapiens*, the group estimate maximum size is 147.8, or roughly 150.

In the anthropological literature the number 150 pops up again and again referencing the maximum size of organizations and groups. According to [30], looking at 21 different hunter-gatherer societies from the Walbiri of Australia, the Tauade of New Guinea, and the Ammassalik of Greenland, to the Ona of Tierra de Fuego; it has been found that the average number of people in their villages was 148.4. The same patterns hold true apparently for religious groups like the Hutterites who have a strict policy that every time a colony approaches 150, they split it in two and start a new one. Another interesting example given in [30] is the American company Gore Associates, a multimillion-dollar high-tech firm based in Newark, Delaware, which is split into independent groups that have never had a size over 150 employees.

Although these arguments seem plausible, they do not give any explanation for the existence of communities of other different sizes. It is evident that not all companies, institutions, or religious congregations are composed of groups of 150 people. It would be easy to find hundreds of examples of organizations with sizes of thousands of people whose members do not show any symptom of information overload. Nevertheless, the proposed existence of a limit to the amount of information a person may deal with in a particular community seems reasonable. The explanation for this paradox may be obtained by analyzing the relationship between the amount of information, the number of members, and the structure of social networks.

One of the most important conclusions that can be extracted from the model proposed in this paper is that the quantity of information an individual is able to receive from the entire network does not vary linearly with the number of members of the community but instead, follows a strongly nonlinear shape that converges to a finite constant value for N sufficiently large. This has a straightforward social interpretation, from an individual perspective, once the value of the average coordination degree is close to the limit, a significant increase in the number of members in the organization does not produce an increase in the average coordination degree. This means that the information received by any individual belonging to that particular network may stay under the Dunbar limit independently of the number of members. Thus, the network may grow indefinitely without any kind of information overload.

On the other hand, we have also seen that the relation between the average coordination degree and the number of members depends strongly on the structure of the organization. It might be possible that for some particular relational structure, the amount of information received by the vertices gets to the Dunbar limit for a particular finite number of members that could be the well-known figure of 150.

In any case, the analysis performed in this paper shows that the size of an organization cannot be only understood in terms of the instrinsic phychological properties of its members as proposed by Dunbar. The relational structure and the properties for the information transfers on the network may also play a definitive role.

V. CONCLUSIONS

Although there are some refinements that must be made to the model, it is important to remark that such a simple model of human relationships is able to explain an important feature of social networks. The same general conclusion can be obtained with any other model just by assuming the number of relationships to be limited and the spread of information as a degrading process.

Even though the model explains a remarkable property of social networks, there are several inconsistencies that need further development and research. Possibly the most serious one is the convergence problem for the upper limit Γ_∞^R. When $ke^{-\xi}>1$, the value of this limit goes to ∞. This may lead to several problems in the interpretation of the model. So it would be interesting to perform some refinements, mainly in modeling the degradation of the information. Another incomplete feature is the characterization of the coordination strength ξ. It has been defined as a constant depending on the nature of the relationship, but a deeper interpretation and analysis would be desirable.

The social models used here have been chosen for their simplicity. They must not be considered as describing real interactions between people in communities and organizations. More realistic models should take account of many different effects such as different orders and strengths of relationships, realistic distribution degrees (scale-free [15], etc.), directed and weighted links, etc.

For the skeptical, this paper can just be read as a different approach to the problem of information spread in networks when there exists a degradation of the information. Moreover, it is original in the sense that its objective is not to find what the structure of a particular social community is, but instead to study how the different structures can influence the collective properties of the group and the particular perception that each individual has of the entire network.

ACKNOWLEDGMENT

This work was supported by the Spanish Ministry of Science and Technology under Project No. BFM2000-0967.

[1] O. Sporns, G. Tononi, and G.M. Edelman, Cereb. Cortex **10**, 127 (2000).

[2] C. Cherniak, J. Neurosci. **14**, 2418 (1994).

[3] D. Golomb and D. Hansel, Neural Comput. **12**, 1095 (2000).

[4] T. B. Achacoso and W. S. Yamamoto, *Neuroanatomy of C. Elegants for Computation* (CRC Press, Boca Raton, FL, 1992).

[5] M. Faloutsos, P. Faloutsos, and C. Faloutsos, Comput. Commun. Rev. **29**, 251 (1999).

[6] L. Adamic, *The Small World Web*, Lecture Notes Computer Science Vol. 1696 (Springer, New York, 1999), pp. 443–452.

[7] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajalopagan, R. Stata, A. Tomkins, and J. Wiener, Comput. Netw. **33**, 309 (2000).

[8] A. Kirman, C. Oddou, and S. Weber, Econometrica **54**, 129 (1986).

[9] R. Albert and A.-L. Barabsi, Phys. Rev. Lett. **84**, 5660 (2000).

[10] A.-L. Barabsi, R. Albert, and H. Jeong, Physica A **281**, 2115 (2000).

[11] M.E.J. Newman, S.H. Strogatz, and D.J. Watts, Phys. Rev. E **64**, 026118 (2001).

[12] M.E.J. Newman, Phys. Rev. E **64**, 016131 (2001); **64**, 016132 (2001); Proc. Natl. Acad. Sci. U.S.A. **98**, 404 (2001).

[13] M.E.J. Newman, Phys. Rev. E **64**, 025102(R) (2001).

[14] L.A.N. Amaral, A. Scala, M. Barthélémy, and H.E. Stanley, Proc. Natl. Acad. Sci. U.S.A. **97**, 11 149 (2000).

[15] R. Albert and A.-L. Barabási, e-print cond-mat/0106096.

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

[17] D. J. Watts, *Small Worlds* (Princeton University Press, Princeton, NJ, 1999).

[18] P. Erdös and A. Rényi, *On the Evolution of Random Graphs* (Mathematical Institute of the Hungarian Academy of Sciences, Budapest, 1960).

[19] F. Liljeros, C.R. Edling, L.A.N. Amaral, H.E. Stanley, and Y. Aberg, Nature (London) **411**, 907 (2001).

[20] S. Milgram, Psychology Today **2**, 60 (1967).

[21] C. Korte and S. Milgram, J. Pers Soc. Psychol. **15**, 101 (1970).

[22] F. Ball, J. Mollison, and G. Scalia-Tomba, Ann. Appl. Probab. **7**, 46 (1997).

[23] M.J. Keeling, Proc. R. Soc. London, Ser. B **266**, 859 (1999).

[24] M. Boots and A. Sasaki, Proc. R. Soc. London, Ser. B **266**, 1933 (1999).

[25] R. Albert, H. Jeong, and A.-L. Barabasi, Nature (London) **406**, 378 (2000).

[26] L. Katz, Psychometrika **18**, 39 (1953).

[27] F. R. K. Chung, *Mathematics of Information Processing, Proceedings of Symposia in Applied Mathematics* (American Mathematical Society, Providence, RI, 1986), Vol. 34, pp. 1–18.

[28] Moore graphs are also called minimal graphs and are distance regular.

[29] R.I.M. Dunbar, Behav. Brain Sci. **16**, 681 (1993).

[30] M. Gladwell, *The Tipping Point. How Little Things Can Make a Big Difference* (Little Brown, Boston, 2000).