Models of social networks based on social distance attachment

Marián Boguñá,[1] Romualdo Pastor-Satorras,[2] Albert Díaz-Guilera,[1] and Alex Arenas[3]
[1]*Departament de Física Fonamental, Universitat de Barcelona, Av. Diagonal 647, 08028 Barcelona, Spain*
[2]*Departament de Física i Enginyeria Nuclear, Universitat Politècnica de Catalunya, Campus Nord B4, 08034 Barcelona, Spain*
[3]*Departament d'Enginyeria Informàtica i Matemàtiques, Universitat Rovira i Virgili, 43007 Tarragona, Spain*
(Received 14 April 2004; published 22 November 2004)

We propose a class of models of social network formation based on a mathematical abstraction of the concept of social distance. Social distance attachment is represented by the tendency of peers to establish acquaintances via a decreasing function of the relative distance in a representative social space. We derive analytical results (corroborated by extensive numerical simulations), showing that the model reproduces the main statistical characteristics of real social networks: large clustering coefficient, positive degree correlations, and the emergence of a hierarchy of communities. The model is confronted with the social network formed by people that shares confidential information using the Pretty Good Privacy (PGP) encryption algorithm, the so-called web of trust of PGP.

DOI: 10.1103/PhysRevE.70.056122 PACS number(s): 89.75.−k, 87.23.Ge, 05.70.Ln

I. INTRODUCTION

Social networks are a paradigm of the complexity of human interactions [1,2]. From a reductionist point of view, social networks can be represented in terms of a set of people or social agents related in pairs between them by a set of peer-to-peer relationships. This structure can thus be abstracted as a *network* or *graph* [3,4], in which vertices represent social agents, while the edges stand for their mutual relations or interactions. This kind of representation places social networks in the more general context of complex networks [5,6], which have attracted recently a great deal of attention within the statistical physics community. The advantage of this abstraction is that any social organization—from companies to groups of friends—can be represented as a tractable mathematical object: a complex graph. Although this representation allows the statistical characteristization of the topology of large real social networks, a similar reductionism representing the intricate mechanism of social network formation, and its differences from the mechanisms driving the generation of other kinds of networks, is still missing.

Finding out the building rules of social network formation is not an easy task provided the myriad of particulars that influence human interactions. Individuals sharing the same interests, common places, similar ideas or akin objectives, for example, tend to form acquaintances. There is a large tradition in sociology (see Ref. [7] and references therein) proposing models of random interactions between social agents that play a certain game, and that use learning and rationality to evolve within its social structure. Although this approach provides valuable information about the way social networks evolve, a question remains open: Could a simple "nonagent based" mathematical model generate the topological structure that ensues from the observation of social networks? It has been suggested that a particular kind of social networks, the so-called collaboration networks (e.g., researchers linked among them by the fact of having coauthored a scientific paper) could owe their topological properties to the fact that the proper social network is actually the projection of a bipartite graph [6] defined in terms of collaborators and acts of collaboration [8]. However, such an explanation fails for other kinds of social networks, such as acquaintance networks [9] (which cannot be associated to any particular collaboration act).

To fill this gap, in the present paper we study a class of social network models based on the concept of social distance. Social distance quantifies in a mathematical way a quite intuitive concept: the degree of closeness or acceptance that an individual or group feels towards another individual or group in a social network. Individuals establish thus social connections with a probability decreasing with their relative social distance (properly defined in a characteristic social space). The result of our models are networks showing most of the topological properties exhibited by their real counterparts.

II. STATISTICAL CHARACTERIZATION OF SOCIAL NETWORKS: EMPIRICAL RESULTS

The recent literature reflects the fact that the structure of social networks is essentially different to the structure of other kind of networks found in nature, such as biological networks, food webs, the World-Wide Web or the Internet [5,6,8,10,11], in at least three specific issues: transitivity of the relationships between peers (clustering), correlations between the number of acquaintances (vertex degree) of peers, and the presence of a community structure with patterns closely resembling the fractal organization present in many natural phenomena [12,13]. Provided this information, any model intended to generate a network consistent with the topological structure of social networks should reproduce these essential characteristics.

The statistical characteristics of social networks could be summarized in the following three points.

Large clustering. The number of transitive relationships between peers in social networks is notably large. The fraction of transitive relations can be measured by means of the clustering coefficient [10], that is defined as follows: The

clustering coefficient c_i of a vertex i is given by the ratio between the number of triangles connected to that vertex e_i, and the total number of possible triangles including it, i.e., $c_i = 2e_i / k_i(k_i - 1)$, where k_i is the number of connections of vertex i (its degree). The clustering coefficient of the network is defined as the average of c_i over all the vertices in the network, $\langle c \rangle = \Sigma_i c_i / N$, where N is the size of the network. Additional information can be obtained from the average clustering coefficient of vertices of degree k, $\overline{c}(k)$ [14], which is related to the presence of modular structures in the network [15]. Recently, it has been shown that a large value of the average clustering coefficient in graphs can be mostly accounted for by a simple random network model in which edges are placed at random, under the constraint of a fixed degree distribution $P(k)$ (defined as the probability that a vertex is connected to k neighbors, i.e., has degree k) [16,17]. For networks with a scale-free degree distribution of the form $P(k) \sim k^{-\gamma}$, this random construction can yield noticeable values of the clustering coefficient for finite networks, indicating that, in this case, the clustering could be a merely topological property. This construction, however, cannot explain the large clustering coefficient observed in social networks with a bounded, non-scale-free degree distribution [9] that is the fingerprint of social networks represented as unipartite graphs [37,38]. Common sense suggests that this specific way of making acquaintances described as "the friends of my friends are my friends" is embedded in the primary mechanism of social network formation.

Positive degree correlations. It has been recognized [18,19] that real networks show degree correlations, in the sense that the degrees at the end points of any given edge are not independent. In particular, this feature can be quantitatively measured by computing the average degree of the nearest neighbors of a vertex of degree k, $\overline{k}_{nn}(k)$ [18]. In this sense, nonsocial networks exhibit *disassortative mixing* [19], implying that highly connected vertices tend to connect to vertices with small degree, and vice versa. This property translates in a decreasing $\overline{k}_{nn}(k)$ function. Social networks, on the other hand, display a strong *assortative mixing*, with high degree vertices connecting preferably to highly connected vertices, a fact that is reflected in an increasing $\overline{k}_{nn}(k)$ function. It has been pointed out [20] that, for finite networks, disassortative mixing can be obtained from a purely random model, by just imposing the condition of having no more than one edge between vertices. This observation implies that negative correlations can find a simple explanation in random connectivity models; explanation that, on the other hand, does not apply to social networks, which must be driven by different organizational principles that favor the formation of groups based on similar connectivity. Analogously, an alternative explanation for collaboration networks [8] does not apply in the more general context of acquaintance networks

Community structure. Social networks possess a complex community structure [12,21,22], in which individuals typically belong to groups or communities, with a high density of internal connections and loosely connected among them, that on their turn belong to groups of groups and so on, giving raise to a hierarchy of nested social communities of practice showing in some cases a self-similar structure [12]. Several authors [8,12,22] have advocated this last property, the presence of a community structure, as the very distinguishing feature of social networks, responsible for the rest of the properties that differentiate those from nonsocial networks. There have been some attempts to model this community aggregation assuming a hierarchical representation of the world [23], however, up to now there are no models from which this particular organization results as an emergent property.

To present an example of the above properties, in the next section we will analyze in detail a real example of a large scale social network: the PGP web of trust.

III. THE WEB OF TRUST OF PGP

The Internet provides the largest publicly communication space ever known, where billions of messages are interchanged between peers every day. Given this enormous flow of information, privacy can be forged quite easily. To overcome this inconvenience several encryption algorithms, aimed to maintain privacy between peers, have been developed. The most popular of these algorithms is the Pretty-Good-Privacy algorithm (PGP) [25,26]. This algorithm makes use of a pair of keys, one of them to encrypt the message, and its counterpart to decrypt the message. Both keys are generated in such a way that it is computationally infeasible to deduce one key from the other. However, the breakthrough of the PGP algorithm is the way in which these keys are used: Everybody can generate its own pair of keys, one of them is published worldwide whereas the other is kept in secret. To establish a private communication using PGP one must get the public key of the target peer and encrypt the message to be sent using this key. The receiver uses his secret key to decrypt the message and in this way privacy is ensured. Provided that everyone can generate a PGP key by himself, if anybody wants to know if a given key belongs really to the person stated in the key, he has to verify that. This is very easy if you know the person who created the key, but it is difficult if you do not know that person at all. This is known as the authentication of the public key problem. A solution to this problem passes through a "signing procedure" where a person signs the public key of another, meaning that she trusts the other person is who she claims to be. This procedure generates a web of peers that have signed public keys of another based on trust, and this is the so-called web of trust of PGP [24].

In this paper, we analyze the web of trust as it was on July 2001, when it comprised 191 548 keys and 286 290 signatures. Since we are mainly interested in the social character of the web of trust we only consider bidirectional signatures, i.e., peers who have mutually signed their keys. This filtering process guarantees mutual knowledge between connected peers and makes the PGP network a reliable proxy of the underlying social network. An extended analysis of the directed PGP network can be found in Ref. [27]. After the filtering process, we are left with an undirected network of 57 243 vertices and average degree $\langle k \rangle = 2.16$. The giant component (GC) of this network, i.e., the largest connected sub-

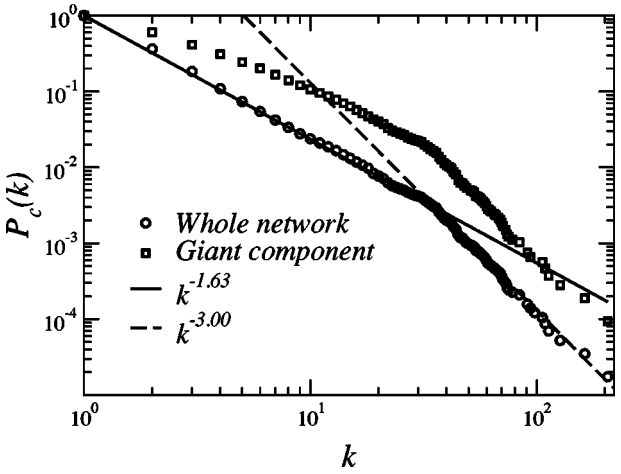

FIG. 1. Cumulative degree distribution of the whole PGP network (circles) and the giant component (squares). As can be seen, there is a region with a power law decay followed by a cutoff for degrees $k>40$.

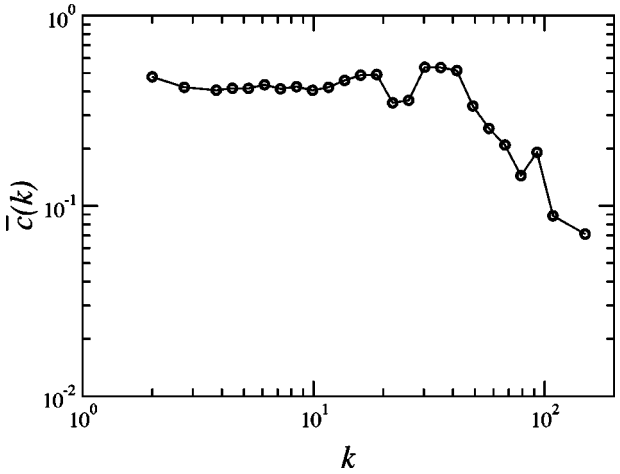

FIG. 2. Average clustering coefficient as a function of the degree k of the PGP network.

network, comprises 10 680 vertices and its average degree is $\langle k\rangle_{\mathrm{GC}}=4.55$.

The interest of the PGP network is twofold: First, it is a web based on trust, and the comprehension of trust networks is, nowadays, crucial to understand the complexity of the information society. Second, unlike collaboration networks, this web is one of the largest reported *nonbipartite* graphs one can build from large databases in social sciences. The consideration of this web of trust as a benchmark for the evaluation of the proposed social model is, thus, fully justified.

Our analysis will focus on the main properties of social networks discussed at the introduction: degree distribution, clustering, correlations, and community aggregation of the web of trust. To begin with, we analyse the degree distribution of the PGP network. Figure 1 shows the cumulative degree distribution [defined as $P_c(k)=\Sigma_{k'=k}^{\infty}P(k')$], for both the whole network and the giant component. In the case of the whole network it is clearly visible a power law decay for the degree distribution $P(k)$, with an exponent ~ 2.6 for small degrees, $k<40$, and a crossover towards another power law with a higher exponent, ~ 4, for large values of the degree. This change of the exponent indicates that, in contrast to many technological networks or social collaboration networks [5], the PGP is not a scale-free network but has a bounded degree distribution.

As discussed above, clustering is one of the distinctive features of social networks. This is also the case of the PGP network, which shows a large clustering coefficient, $\langle c\rangle = 0.4$. In Fig. 2, we plot the clustering coeficient as a function of the degree $\bar{c}(k)$, that is, the average clustering of vertices of degree k. Dispite the short range of values of k shown by this plot, due to the limited size of the network and the bounded nature of the degree distribution, we can observe that $\bar{c}(k)$ is a nearly independent function of the degree. This is surprisingly in contrast to many real networked systems in which it has been shown that $\bar{c}(k)$ is a decreasing function of the degree [15].

In Fig. 3 we analyze the correlations of the PGP network, as measured by the average degree of the nearest neighbors of the vertices of degree k, $\bar{k}_{nn}(k)$. In the range of degree values available in the plot, we observe that this is a growing function, corresponding to a network with assortative mixing. Remarkably, the function $\bar{k}_{nn}(k)$ has an approximately linear behavior, at least for not very large values of k.

Finally, we focus on the community structure of the web of trust. To this purpose, we use the algorithm proposed by Girvan and Newman (GN) [22] to identify communities in complex networks. The performance of this algorithm relies on the fact that edges connecting different communities have high betweenness (a centrality measure of vertices and edges of the network [28], that is defined as the total number of shortest paths among pairs of vertices of the network that pass through a given vertex or edge [29]). The algorithm recursively identifies and cuts the edges with the highest be-

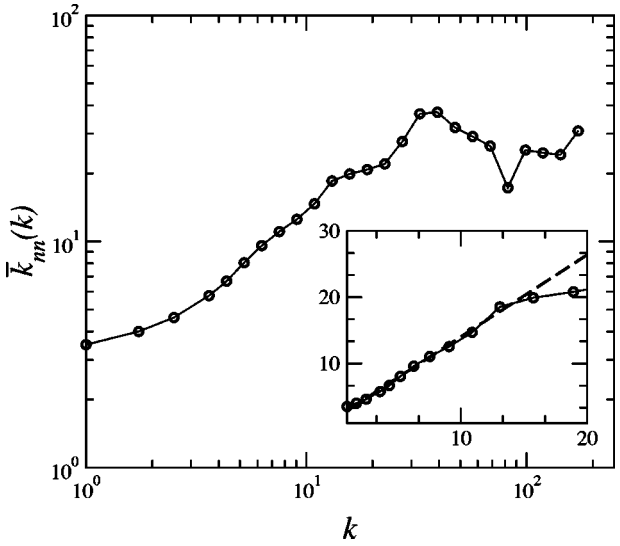

FIG. 3. Average degree of the nearest neighbors as a function of the degree k for the PGP network. This function has an approximately linear behavior for small values of k (see inset).

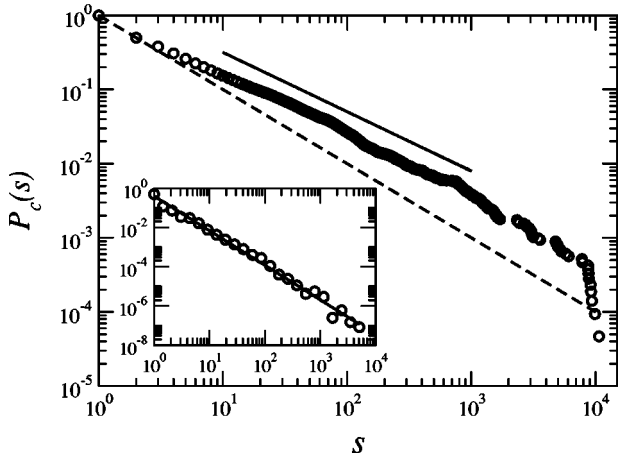

FIG. 4. Cumulative community size distribution of the giant component of the PGP network. The solid line has slope 0.8, indicating a community size distribution of the form $P(s) \sim s^{-1.8}$. The dashed line shows, as comparison, a power law of exponent s^{-2}. Inset: Binned community size distribution. The solid line has slope 1.8, in agreement with the slope measured from the cumulative distribution.

tweenness, splitting the network until the single vertex level. The information of the entire process can be encoded into the binary tree generated by the splitting procedure. The advantage of using the binary tree representation is twofold, since it gives information about the different communities—which are the branches of the tree—and, at the same time, unravels the hierarchy of such communities. In Fig. 4 we explore the scaling of the probability of having a community of size $s, P(s)$, by plotting the cumulative distribution $P_c(s) = \Sigma_{s'=s}^{\infty} P(s')$ as a function of s in log-log scale, for the giant component of the PGP network. The use of cumulative distributions instead of binned distributions is useful to reduce statistical noise of data and, as we shall see, it does not alter the results of the paper. The plot in Fig. 4 reveals a scale-free community hierarchy of the form $P(s) \sim s^{-\delta}$, with an exponent $\delta = 1.8$ for over more than 3 decades (the inset of Fig. 4 shows the binned community size distribution, which leads to the same value for the exponent δ). This result is, indeed, quite remarkable since, in a recent paper [30,31], it has been argued that any treelike representation should lead to a size distribution with exponent 2. This suggests that the exponent $\delta = 1.8$ should find a more complex explanation than that proposed in the above-mentioned work.

IV. A CLASS OF MODELS BASED ON SOCIAL DISTANCE

To explain the above-mentioned properties in an acquaintance (i.e., nonbipartite) network, we propose a class of models based on the concept of *social distance*. The intuitive notion that individuals establish acquaintance or friendship links whenever they feel *close* in some sense, leads to the notion of a social distance between individuals. This social distance will rule the establishment of relations, in such a way that individuals at short distances will have a large probability of being related, while individuals at large distances will be connected with low probability.

To provide a mathematical realization of this concept, we consider a class of models of social networks in which each vertex (individual) has associated a location in a certain social space [23], whose coordinates account for the different characteristics that define their relative social location with respect to the rest of the individuals. Individuals establish social connections (acquaintances) with a probability decreasing with their relative social distance, that is defined on the metric social space. As we will see in the following, for general forms of the connecting probability, the model yields networks of acquaintances with a nonvanishing clustering coefficient as the number of individuals increases, plus general assortative (positive) correlations. For a certain range of connectivity probabilities, moreover, the model reproduces a community structure with self-similar properties. The model we propose reproduces the hierarchical world model proposed in Ref. [23] (see also [32]). Our approach differs, however, in the fact that hierarchies are not defined *a priori*, but they emerge as a result of the construction process.

Our model is defined as follows: Let us consider a set of N disconnected individuals which are randomly placed within a social space, $\mathcal{H}$, according to the density $\rho(\vec{h})$, where the vector $\vec{h}_i \equiv (h_i^1, \ldots, h_i^{d_{\mathcal{H}}})$ defines the position of the ith individual and $d_{\mathcal{H}}$ is the dimension of $\mathcal{H}$ [33,34]. Each subspace of $\mathcal{H}$ (defined by the different coordinates of the vector $\vec{h}$) represents a distinctive social feature, such as profession, religion, geographic location, etc. and, in general, it will be parametrized by means of a continuous variable with a domain growing with the size of the population. This choice is justified by the fact that there are not two identical individuals and, thus, increasing the number of individuals also increases the diversity of the society. Even though it is not strictly necessary for our further development, we also assume that different subspaces are uncorrelated and, therefore, we can factorize the total density as $\rho(\vec{h}) = \Pi_{n=1}^{d_{\mathcal{H}}} \rho_n(h^n)$. Assuming again the independence of social subspaces, we assign a connection probability between any two pairs of individuals, $\vec{h}_i$ and $\vec{h}_j$, given by

$$r(\vec{h}_i, \vec{h}_j) = \sum_{n=1}^{d_{\mathcal{H}}} \omega_n r_n(h_i^n, h_j^n), \tag{1}$$

where ω_n is a normalized weight factor measuring the importance that each social attribute has in the process of formation of connections. The key point of our model is the concept of social distance across each subspace [23]. We assume that given two nodes i and j with respective social coordinates $\vec{h}_i$ and $\vec{h}_j$, it is possible to define a set of distances corresponding to each subspace, $d_n(h_i^n, h_j^n) \in [0, \infty)$, $n = 1, \ldots, d_{\mathcal{H}}$. Moreover, we expect that the probability of acquaintance decreases with social distance. Therefore, we propose a connection probability

$$r_n(h_i^n, h_j^n) = \frac{1}{1 + [b_n^{-1} d_n(h_i^n, h_j^n)]^{\alpha_n}}, \tag{2}$$

where b_n is a characteristic length scale (that, eventually, will control the average degree) and $\alpha_n > 1$ is a measure of ho-

mophily [35], that is, the tendency of people to connect to similar people.

The degree distribution $P(k)$ of the network can be computed using the conditional probability $g(k|\vec{h})$ (propagator) that an individual with social coordinates $\vec{h}$ has k connections [34]. We can thus write

$$P(k) = \int \rho(\vec{h}) g(k|\vec{h}) dh, \tag{3}$$

where dh stands for the measure element of space $\mathcal{H}$. The propagator $g(k|\vec{h})$ can be easily computed using standard techniques of probability theory [34], leading to a binomial distribution

$$g(k|\vec{h}) = \binom{N-1}{k} \left(\frac{\overline{k}(\vec{h})}{N-1}\right)^k \left(1 - \frac{\overline{k}(\vec{h})}{N-1}\right)^{N-1-k}, \tag{4}$$

where $\overline{k}(\vec{h})$ is the average degree of individuals with social coordinate $\vec{h}$. For uncorrelated social subspaces, this average degree takes the form

$$\overline{k}(\vec{h}) = (N-1) \sum_{n=1}^{d_{\mathcal{H}}} \omega_n \int \rho_n(h'^n) r_n(h^n, h'^n) dh'^n. \tag{5}$$

In the case of a sparse network—constant average degree—the propagator takes a Poisson form [34] and the degree distribution can simply be written as

$$P(k) = \frac{1}{k!} \int \rho(\vec{h}) [\overline{k}(\vec{h})]^k e^{-\overline{k}(\vec{h})} dh. \tag{6}$$

Therefore, if the population is homogeneously distributed in the social space, the degree distribution will be bounded, in agreement with the observations made in several real social systems [9,12,36,39].

The clustering coefficient is defined as the probability that two neighbors of a given individual are also neighbors themselves. Following Ref. [34], we first compute the probability that an individual with social vector $\vec{h}$ is connected to an individual with vector $\vec{h}', p(\vec{h}'|\vec{h})$. This probability reads

$$p(\vec{h}'|\vec{h}) = (N-1) \frac{\rho(\vec{h}') r(\vec{h}, \vec{h}')}{\overline{k}(\vec{h})}. \tag{7}$$

Given the independent assignment of edges among individuals, the clustering coefficient of an individual with vector $\vec{h}$ is [34]

$$c(\vec{h}) = \int \int p(\vec{h}'|\vec{h}) r(\vec{h}', \vec{h}'') p(\vec{h}''|\vec{h}) dh' dh'', \tag{8}$$

and the average clustering coefficient is simply given by

$$\langle c \rangle = \int \rho(\vec{h}) c(\vec{h}) dh. \tag{9}$$

Our model, as defined above, describes a general class of models which might be useful for the modeling of different social networks. In the next section we will analyze the simplest element of this class and we will show that this oversimplified model is, nevertheless, able to reproduce qualitatively the main characteristics of real social networks.

V. EXAMPLE OF A SOCIAL NETWORK MODEL

To test the results of our model, we consider the simplest case of a single social feature, i.e., $d_{\mathcal{H}}=1$ [40]. As we will see, even in this case our model presents several nontrivial properties, that are the signature of real social networks. Considering the space $\mathcal{H}$ to be the one-dimensional segment $[0, h_{\max}]$, we assign individuals a random, uniformly distributed, position, i.e., $\rho(h)=1/h_{\max}$. In this way, the density of individuals in the social space is given by $\delta=N/h_{\max}$. The distance between individuals is defined as $d(h_i, h_j) \equiv |h_i - h_j|$. The top panel of Fig. 5 shows some typical examples of networks generated with our model, for different values of the parameter α.

The model, as defined above, is homogeneous in the limit $h_{\max} \gg 1$, which means that all the vertex properties will eventually become independent of the social coordinate h. Therefore, the average degree can be calculated as $\langle k \rangle = \lim_{h_{\max}\to\infty} \overline{k}(h=h_{\max}/2)$, which leads to

$$\langle k \rangle = \frac{2\delta b \pi}{\alpha \sin \pi/\alpha}. \tag{10}$$

Thus, for fixed δ, we can construct networks with the same average degree and different homophily parameter α, by changing b according to the previous expression. The clustering coefficient can be computed by means of Eq. (8), yielding

$$\langle c \rangle = \frac{\alpha^2}{4\pi^2} f(\alpha) \sin^2 \frac{\pi}{\alpha}, \tag{11}$$

where

$$f(\alpha) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{dxdy}{(1+|x|^\alpha)(1+|x-y|^\alpha)(1+|y|^\alpha)}. \tag{12}$$

Figure 6 shows the perfect agreement between simulations of the model compared to the theoretic value Eq. (11), computed by numerical integration. We observe that the clustering coefficient vanishes when $\alpha \to 1$, that is, for weakly homophyllic societies, and converges to a constant value $\langle c \rangle = 3/4$ when $\alpha \to \infty$ [41], which corresponds to a strongly homophilyc society. The inset of Fig. 6 shows the clustering coefficient as a function of the degree, $\overline{c}(k)$, for several values of the homophily parameter. We observe that $\overline{c}(k)$ is, in all cases, independent of the vertex degree in agreement with the empirical measures on the PGP network.

Regarding the degree correlations, at first sight one could conclude that, since the network is homogeneous in the social space $\mathcal{H}$, the resulting network is free of any correlations. However, numerical simulations of the average degree of the nearest neighbors as a function of the degree, $\overline{k}_{nn}(k)$, show a linear dependence on k and, consequently, assortative mixing by degree (see Fig. 7). This counterintuitive result is

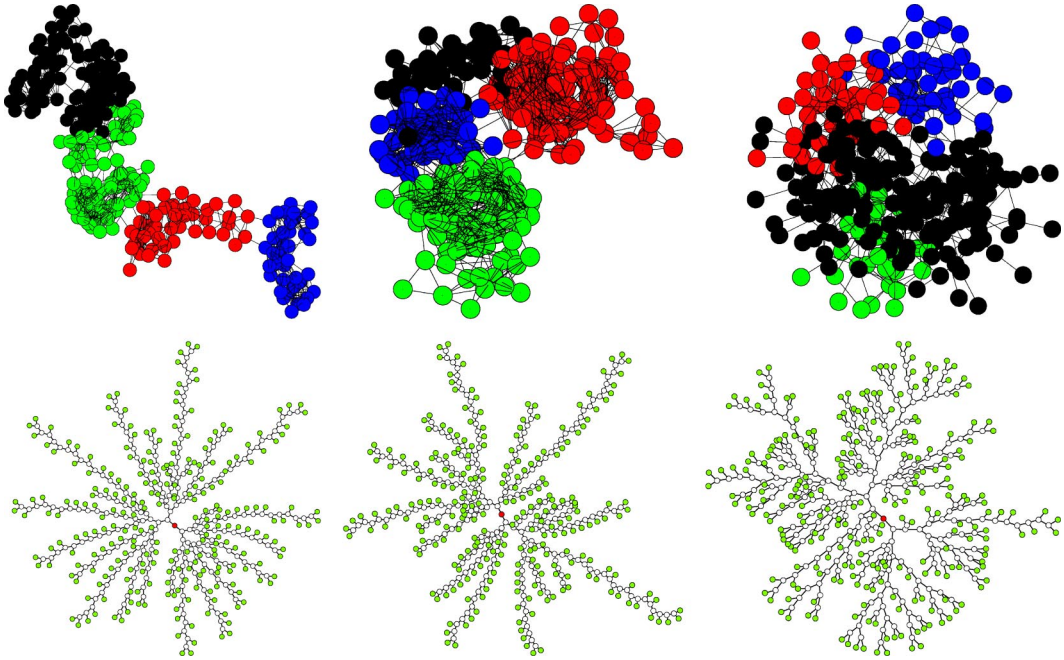

FIG. 5. (Color online) Top panel: Examples of typical networks generated for an average degree $\langle k\rangle=10, N=250, \delta=2$, and different values of the parameter α. Bottom panel: Binary trees representing the community structure of the corresponding networks (see text). Solid (green) circles are the original vertices of the network whereas hollow circles stand for the communities generated by the GN algorithm.

a consequence of the fluctuations of the density of individuals in the social space. Indeed, if individuals are placed in the space $\mathcal{H}$ with some type of randomness, they will end up forming clusters (communities) of close individuals, strongly connected among them. Therefore, an individual with large degree will most probably belong to a large cluster, and consequently its neighbors will have also a high degree.

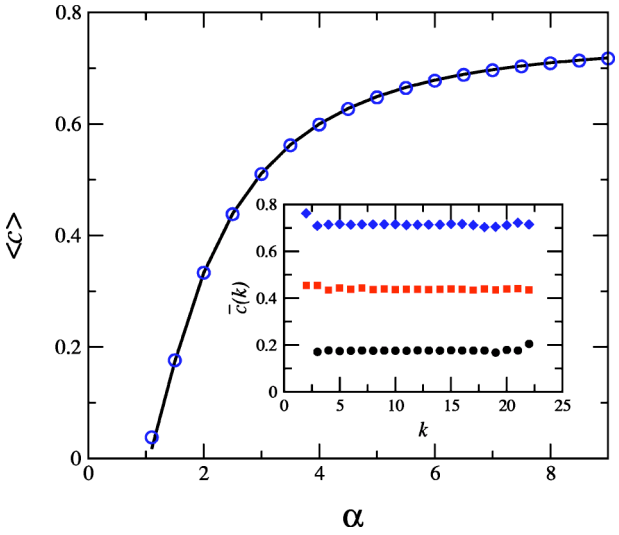

FIG. 6. (Color online) Main plot: Clustering coefficient $\langle c\rangle$ for $d_{\mathcal{H}}=1$ as a function of α and fixed average degree, $\langle k\rangle=10$. The solid line corresponds to the theoretical value Eq. (11) and symbols are simulation results. Inset: Average clustering coefficient as a function of the degree k for $d_{\mathcal{H}}=1$ for different values of α (from bottom to top, $\alpha=1.5$, 2.5, and 3.5). In all cases, the size of the network is $N=10^5$.

In more real situations, however, fluctuations in the density of individuals will, presumably, be originated by more complex processes or even induced by deterministic constraints, leading to complex patterns within the social space. This complex distribution will, in general, alter the shape of the degree distribution but not the assortative character of the network. Indeed, assortative mixing by degree appears as a common feature in any model of network formation driven by distance attachment. It is worth mentioning that the small range of degrees shown in Figs. 6 and 7 is a direct consequence of the boundedness of the degree distribution. However, we expect the same behavior in models with more heavy tailed distributions.

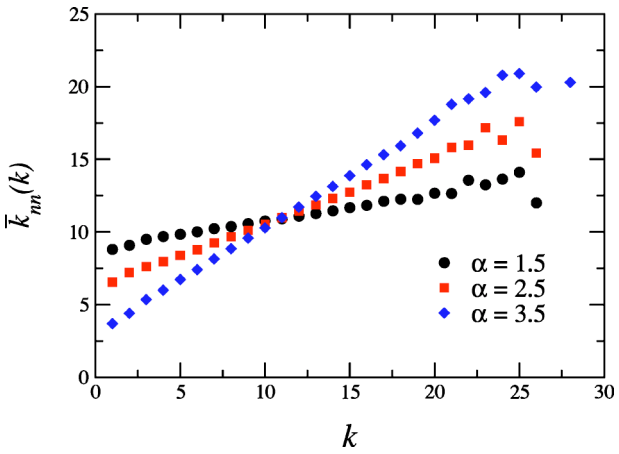

FIG. 7. (Color online) Average nearest neighbors degree for $d_{\mathcal{H}}=1$ as a function of k, for different values of α. In all cases, the size of the network is $N=10^5$.

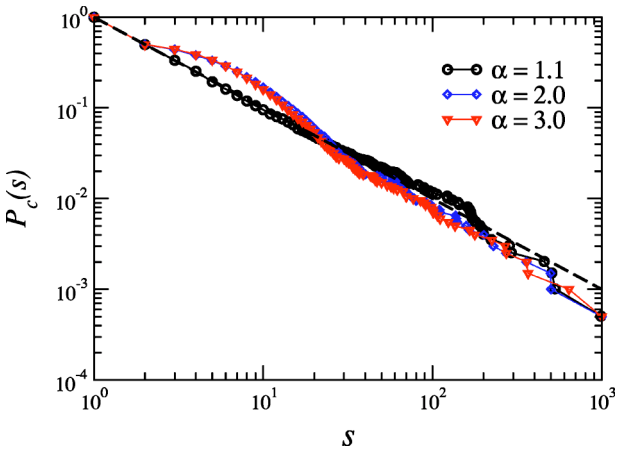

FIG. 8. (Color online) Cumulative size distribution obtained using the GN algorithm for different values of α. As $\alpha \to 1$ the network becomes a perfectly hierarchical network characterized by a power law community size distribution, $P(s) \sim s^{-2}$ (dashed line). In all the cases the size of the network is $N=1000$.

Finally, we analyze the community structure of our model. The top panel of Fig. 5 shows three typical networks, in which the first four communities identified by the GN algorithm have been highlighted in different colors. The binary trees corresponding to these networks are shown in the bottom panel. As α grows, the network eventually becomes a chain of clusters connected by a few edges. In contrast, as α approaches 1 the network is more and more interconnected and develops a hierarchical structure. This hierarchical structure can be quantified by means of the cumulative distribution of community sizes, $P_c(s)$, in which the community size s is defined as the number of individuals belonging to each offspring during the splitting procedure [12]. Figure 8 shows $P_c(s)$ for $\alpha=1.1$, 2 and 3. When $\alpha \sim 1$, the size distribution approaches to $P(s) \sim s^{-2}$, reflecting the hierarchical structure of the network. For higher values of α the hierarchy is still preserved for large community sizes whereas for small sizes there is a clear deviation as a consequence of clusters of highly connected individuals which form indivisible communities, breaking thus the hierarchical structure at low levels. These clusters are identified in the binary tree as the long branches with many leaves at the end of the tree.

VI. CONCLUSIONS

To sum up, in this paper we have presented a model of social network based on the concept of social distance between the elements (individuals) in a social network. The model exhibits, even in its simplest formulation, a nonzero clustering coefficient in the thermodynamic limit, assortative degree mixing, and a hierarchical (self-similar) community structure. The origin of these properties can be traced back to the very presence of communities, due to the fluctuations in the position of individuals in social space. Our approach offers an explanation of a real acquaintance network, such as the PGP web of trust, and opens thus new views for a further understanding of the structure of complex social networks.

ACKNOWLEDGMENTS

This work has been supported by DGES of the Spanish Government, Grant No. BFM2000-0626 and EC-FET Open Project No. IST-2001-33555. R.P.-S. acknowledges financial support from the Ministerio de Ciencia y Tecnología (Spain), and from the Departament d'Universitats, Recerca i Societat de la Informació, Generalitat de Catalunya (Spain). M.B. acknowledges financial support from the Ministerio de Ciencia y Tecnología through the Ramón y Cajal program.

[1] S. Wasserman and K. Faust, *Social Network Analysis: Methods and Applications* (Cambridge University Press, Cambridge, England, 1994).
[2] J. Scott, *Social Network Analysis: A Handbook* (Sage Publications, London, 2000).
[3] B. Bollobás, *Modern Graph Theory* (Springer-Verlag, New York, 1998).
[4] F. Harari and R. Norman, *Graph Theory as a Mathematical Model in Social Science* (University of Michigan Press, Ann Arbor, 1953).
[5] R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).
[6] S. N. Dorogovtsev and J. F. F. Mendes, *Evolution of Networks: From Biological Nets to the Internet and WWW* (Oxford University Press, Oxford, 2003).
[7] R. Skyrms and R. Pemantle, Proc. Natl. Acad. Sci. U.S.A. **97**, 9340 (2000).
[8] M. E. J. Newman and J. Park, Phys. Rev. E **68**, 036122 (2003).
[9] L. A. N. Amaral, A. Scala, M. Barthélémy, and H. E. Stanley, Proc. Natl. Acad. Sci. U.S.A. **97**, 11149 (2000).
[10] D. J. Watts and S. H. Strogatz, Nature (London) **393**, 440 (1998).
[11] R. Pastor-Satorras and A. Vespignani, *Evolution and Structure of the Internet: A Statistical Physics Approach* (Cambridge University Press, Cambridge, England, 2004).
[12] R. Guimera, L. Danon, A. Diaz-Guilera, F. Giralt, and A. Arenas, Phys. Rev. E **68**, 065103 (2003).
[13] A. Arenas, L. Danon, A. Diaz-Guilera, P. Gleiser, and R. Guimera, Eur. Phys. J. B **38**, 373 (2004).
[14] A. Vázquez, R. Pastor-Satorras, and A. Vespignani, Phys. Rev. E **65**, 066130 (2002).
[15] E. Ravasz and A.-L. Barabási, Phys. Rev. E **67**, 026112 (2003).
[16] M. E. J. Newman, in *Handbook of Graphs and Networks: From the Genome to the Internet*, edited by S. Bornholdt and H. G. Schuster (Wiley-VCH, Berlin, 2003), pp. 35–68.
[17] M. E. J. Newman, S. H. Strogatz, and D. J. Watts, Phys. Rev. E **64**, 026118 (2001).
[18] R. Pastor-Satorras, A. Vázquez, and A. Vespignani, Phys. Rev. Lett. **87**, 258701 (2001).
[19] M. E. J. Newman, Phys. Rev. Lett. **89**, 208701 (2002).
[20] S. Maslov, K. Sneppen, and A. Zaliznyak, "Pattern detection

in complex networks: Correlation profile of the Internet," cond-mat/0205379.
[21] J. R. Tyler, D. M. Wilkinson, and B. A. Huberman, "E-mail as spectroscopy: Automated discovery of community structures within organizations," cond-mat/0303264.
[22] M. Girvan and M. E. J. Newman, Proc. Natl. Acad. Sci. U.S.A. **99**, 7821 (2002).
[23] D. J. Watts, P. S. Dodds, and M. E. J. Newman, Science **296**, 1302 (2002).
[24] Public data about the web of trust of PGP can be found at www.dtype.org.
[25] S. Garfinkel, *PGP: Pretty Good Privacy* (O'Reilly and Associates, Cambridge, 1994).
[26] W. Stallings, BYTE **20**, 161 (1995).
[27] X. Guardiola *et al.*, "Micro- and Macro-structure of Trust Networks," cond-mat/0206240.
[28] L. C. Freeman, Sociometry **40**, 35 (1977).
[29] U. Brandes, J. Math. Sociol. **25**, 163 (2001).
[30] G. Caldarelli, C. C. Cartozo, P. De Los Rios, and V. D. P. Servedio, Phys. Rev. E **69**, 035101 (2004).
[31] P. De Los Rios, Europhys. Lett. **56**, 898 (2001).
[32] J. Kleinberg, in *Advances in Neural Information Processing Systems 14*, edited by T. G. Dietterich, S. Becker, and Z. Ghahramani (MIT Press, Cambridge, MA, 2002).
[33] G. Caldarelli, A. Capocci, P. D. L. Rios, and M. A. Muñoz, Phys. Rev. Lett. **89**, 258702 (2002).
[34] M. Boguñá and R. Pastor-Satorras, Phys. Rev. E **68**, 036112 (2003).
[35] M. McPherson, L. Smith-Lovin, and J. M. Cook, Annu. Rev. Sociol. **27**, 415 (2001).
[36] H. Ebel, L.-I. Mielsch, and S. Bornholdt, Phys. Rev. E **66**, 035103 (2002).
[37] G. Bianconi and A.-L. Barabási, Europhys. Lett. **54**, 436 (2001).
[38] Note that collaboration networks, emerging from cumulated bipartite graphs, can show a scale-free distribution; e.g., see Ref. [37].
[39] This form of the degree distribution is due to the particular connection probability Eq. (2) considered. More complex forms can yield different degree distributions, even with scale-free behavior.
[40] Dimensions $d_{\mathcal{H}} > 1$ provide analogous results, M. Boguñá, R. Pastor-Satorras, A. Díaz-Guilera, and A. Arenas (in preparation).
[41] $f(\alpha \to \infty)$ can be exactly computed by noticing that the functions within the integrals approach, in this limit, to step functions.