Families and clustering in a natural numbers network

Gilberto Corso
International Center for Complex Systems and Departamento de Biofísica, Centro de Biociências, Universidade Federal do Rio Grande do Norte, Campus Universitário, 59072 970 Natal, RN, Brazil

(Received 8 September 2003; revised manuscript received 24 November 2003; published 18 March 2004)

We develop a network in which the natural numbers are the vertices. The decomposition of natural numbers by prime numbers is used to establish the connections. We perform data collapse and show that the degree distribution of these networks scales linearly with the number of vertices. We explore the families of vertices in connection with prime numbers decomposition. We compare the average distance of the network and the clustering coefficient with the distance and clustering coefficient of the corresponding random graph. In case we set connections among vertices each time the numbers share a common prime number the network has properties similar to a random graph. If the criterion for establishing links becomes more selective, only prime numbers greater than p_l are used to establish links, where the network has high clustering coefficient.

DOI: 10.1103/PhysRevE.69.036106 PACS number(s): 89.75.Fb, 89.75.Hc, 02.10.De

I. INTRODUCTION

The new century has started with a strong development in network theory and specially in small-world networks. Three ingredients are necessary to define a small-world network: a sparse network, small distance, and high clustering coefficient [1,2]. Many examples of these networks have been analyzed in diverse fields as computation [3,4], linguistic [5,6], biology [7,8], economics [9,10], and social phenomena [11,12]. In addition to the interest in the description of these particular phenomena, small-world networks are promising elements to compose a general theory of complex systems.

Nowadays the two major research lines in networks are the search for small-world networks in nature and the investigation of theoretic models that explain the properties of such networks. The first line was pointed out in the first paragraph. The second line is dominated by the evolving network models [13,14] and the study of phase order transitions in networks [15,16]. This work is situated at an intermediary place between both lines. We characterize a network using recent techniques of statistical physics but instead of looking for real data in the world we construct a network using some properties of the set of natural numbers.

We use as the keystone in the construction of number networks the fundamental theorem of number theory which says that each natural number has a unique decomposition in prime factors. It means, for a number a and prime numbers p_j, there is a unique product

$$a = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_m^{\alpha_m}, \tag{1}$$

where the exponent α_j is the multiplicity of the prime p_j. The simplest way to define a link between two vertices (numbers) is the following: each time two numbers have a prime in common in Eq. (1) they have a connection. We shall see that only this criterion is not good enough to construct networks with small clustering coefficient.

This paper has a twofold objective: present a network model based on number properties and characterize this network using the approach of modern complex network analyses. In Sec. II we show the model M and its extensions M_l, and we drive our attention to the degree distribution of the network, the set of families, and the search of invariant quantities. In Sec. III we analyze the distance and the clustering coefficient for these networks. In Sec. IV we give our final remarks.

II. THE MODEL

In this section we present the standard model M and its extensions M_l. We use the set of natural numbers as vertices and an arithmetic property establishes the connections. The connections have neither weight nor direction.

A. The basic model M

In model M the criterion for the existence of a connection between two vertices is the following: there is a link between two numbers a and b if they share a common divisor. In other words, if a and b have a common prime number p_j in decomposition (1) a link is established.

Figure 1 shows a simple realization of this rule for number of vertices $N=16$. All the even numbers are interconnected (they share the prime $p=2$). Besides the divisors of 3, 5, etc., are also interconnected. The most connected numbers for this N are 6 and 12 because they are linked to all the

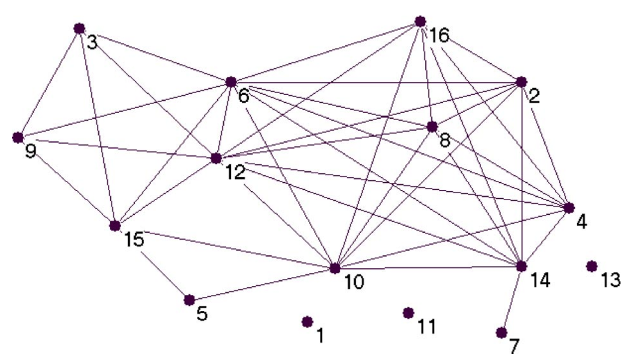

FIG. 1. The network of the model M for number of vertices $N=16$. The natural numbers are connected according to the prime number decomposition. Two numbers have a connection if they share a common prime in the decomposition.

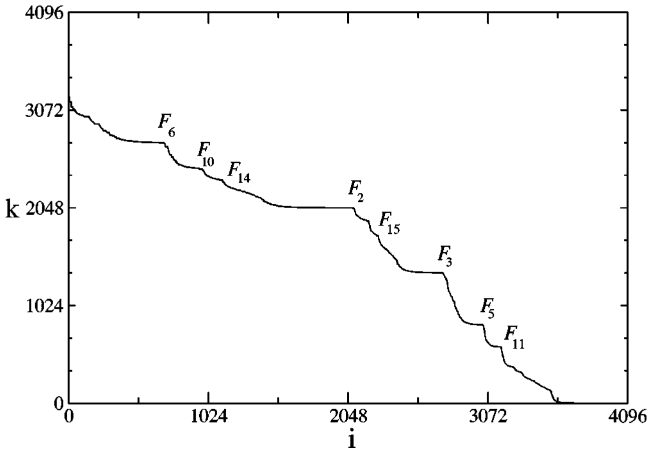

FIG. 2. The connectivity k vs index i for the data of network of model M. It is used $N=4096$. The main families F_k are indicated in the figure.

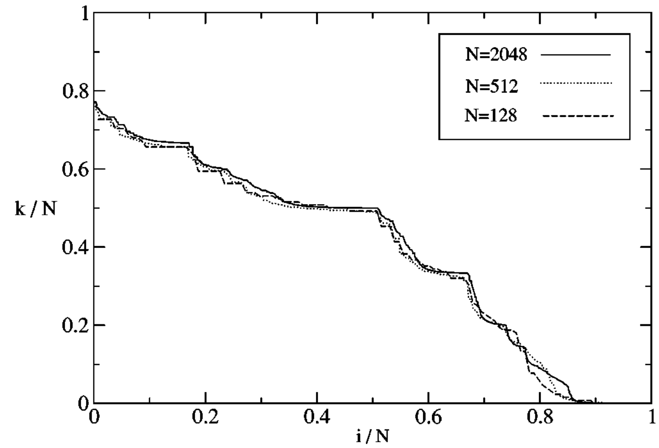

FIG. 3. The normalized connectivity k/N vs normalized index i/N for data of model M. It is used $N=128$, 512, and 2048, as indicated in the figure.

even numbers and all the multiples of 3.

The first point we explore in the network M is the degree distribution; it means the connectivity k of the vertex i for vertices ordered according to the connectivity. Figure 2 displays k versus i. In this figure we have $N=2^{12}=4096$ vertices. The vertex i with maximum number of links is $i=2310=2\times3\times5\times7\times11$ which is connected with all the even numbers and with the multiples of 3, 5, 7, and 11. In general, the most connected vertex in a network M of N vertices corresponds to the maximum number $i=p_1p_2\cdots p_m\leqslant N$ where the p_j correspond to the first primes (which are the most connected numbers). In the opposite side of the graphic there are the prime numbers p_j that have $k=0$; these numbers satisfy the relation $2p_j\leqslant N$ (if p_j do not fulfill this relation it will be connected to the node 2 and it would not have $k=0$). The general view of the degree distribution shown in Fig. 2 is the following. Half of the vertices, which correspond to the even numbers, are connected among themselves, so they have $k\geqslant N/2$. Otherwise, most of odd numbers have $k<N/2$ because they are, in general, not connected to all even numbers. The frontier between these two sets is indicated by the large plateau that starts roughly at $i=N/2$.

The plateau in the middle of Fig. 2 corresponds to the family of multiples of the number 2. In the beginning of this plateau we find the numbers 2^{α} for $1\leqslant\alpha\leqslant12=\log_2 N$. These numbers are connected with all the even numbers and only with them. For these numbers $k=N/2$. The smooth tail that comes for $i<N/2$ is formed by the numbers of type $2^{\alpha}p_j\leqslant N$ where p_j is a weak connected prime. These numbers are connected with all the even numbers and with the connections of p_j that are just a few.

The second largest plateau is related to the multiples of 3 (indicated F_3 in Fig. 2). One-third of the N natural numbers are multiples of 3 and have connectivity larger than all the other odd numbers. In the beginning of this plateau we find the numbers of the form 3^{β} where $1\leqslant\beta\leqslant\log_3 N$. This plateau is smaller than the former one because there are less multiples of 3 than 2. For these numbers $k=N/3$; it means they are connected only with the multiples of 3. The tail of this plateau is formed by the numbers $3^{\beta}p_j\leqslant N$, for p_j weakly connected. There are also plateaus visible in Fig. 2 corresponding to the primes 5, 7 and 11. For these plateaus we have in the coordinate axis $k=N/5$, $N/7$, and $N/11$, respectively.

Another plateau indicated in the figure is formed by the numbers that share the two primes 2 and 3; it means the numbers that are multiples of 6. This plateau starts with numbers of the form $2^{\alpha}3^{\beta}\leqslant N$. If the network were weighted, k of this plateau would be the sum of k of 2 and 3 plateaus. As the network does not count multiple links the corresponding value of k is smaller than the cited sum. We call F_2 the even plateau, F_{10} the plateau generated by the numbers that are divisible only by 2 and 5, and so on. The most important family plateaus F_2, F_3, F_5, F_7, F_{10}, F_{14}, and F_{15} are indicated in Fig. 2.

Figure 2 shows two distinct regions. Above $i=N/2$ the degree distribution is dominated mainly by plateaus of prime numbers of the form $p_j^{\alpha_j}$ and their respective tails. Below this number there are only plateaus composed by the combination of prime numbers of the form $p_j^{\alpha_j}p_m^{\alpha_m}$ and their tails. In fact in the limit of $N\to\infty$ an infinite number of plateaus would appear in the curve whose relative sizes are determined by Eq. (1). We conjecture that in this limit a fractal distribution will appear.

We have $k=N/2$ for the beginning of the plateau F_2, and in general $k=N/p_k$ for the families of primes F_k. This relation suggests an interesting scale property: the degree scales linearly with N. We verify numerically this fact in Fig. 3 where we plot the normalized connectivity (k/N) versus the normalized index (i/N) for $N=128$,512, and 2048. This figure verifies by simulation that the connectivity of the model M scales linearly with the number of vertices.

B. The extension M_l of the model M

In the standard number decomposition, Eq. (1), there is the possibility of including, or not, the factor 1 because all the numbers are trivially divisible by unity. If the number 1 is included in the decomposition all the numbers would share

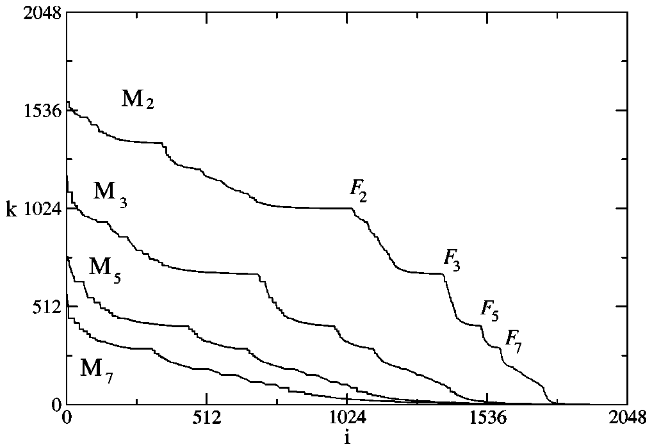

FIG. 4. The connectivity k vs index i for the models M_2, M_3, M_5, and M_7 as indicated in the figure. It is used $N=2048$.

the same common divisor and, as a consequence, the network will become trivial: all the vertices will be interconnected. In the same way as we excluded the number 1 as a divisor in the criterion for establishing links, we could also exclude the number 2. This idea suggests an alternative procedure to define a network model for natural numbers.

The network model M_{p_l} is the following. The vertices are again the natural numbers and the connections are set using Eq. (1), but we take into account only connections of primes $p_j \geqslant p_l$. In this sense the former model M is in fact M_2 because the links are established once there is a common factor p_j such that $p_j \geqslant 2$. In this way, the network M_3 takes into account the primes $p_j = 3,5, \ldots$ to establish links, but not the prime 2.

Figure 4 shows the degree distribution (k versus i in order of decreasing connectivity) for M_2, M_3, M_5, and M_7 as indicated in the figure. We use in the figure $N=2^{11}$. The curve of M_2 is the same curve shown in Fig. 2. The curve of M_3 is similar to M_2, but the largest even plateau F_2 is absent. The largest plateau of M_3 starts at $i=N/3$, because all the one-third of the most connected numbers are multiples of 3. We observe that the largest plateau of M_3 has the same k of F_3 of network M_2; in fact in both cases the plateaus are formed by the F_3 family. The curve M_5 does not show the plateaus corresponding to the families F_2 and F_3 as expected; and in this case the largest plateau is formed by the F_5 family. The curve of M_7 follows the same tendency. As a general trend the curves of degree distribution of M_l become smoother for increasing l because they have less connections related to important prime numbers and, as a consequence, they present less plateaus. The criterion for establishing links in the model M_l becomes more restrictive as l increases because the number of connections decreases. It is interesting that the average connectivity $\langle k \rangle = 2n/N$ normalized by N tends to a constant as $N \to \infty$ (n is the number of connections in the network). In Table I we show $\langle k \rangle/N$ for the networks M_l.

The first conclusion we take from the table is that the model M is not sparse; it means the network does not fulfill the condition $\langle k \rangle \ll N$. Therefore we have to be cautious to

TABLE I. $\langle k \rangle/N$ and $\overline{C}$ for the network number models M_2, M_3, M_5, and M_7.

M_l	M_2	M_3	M_5	M_7
$\langle k \rangle/N$	0.45	0.20	0.09	0.05
$\overline{C}$	1.78	3.65	8.22	14.5

compare properties of such graphs with usual complex graphs in the physics literature. The reason for $\langle k \rangle \propto N$ is rooted in the way connections are established in the network. Each time a new even number is added it is connected with, at least, $N/2$ vertices, and a multiple of 3 with $N/3$ vertices. This fact illustrates that in the average $\langle k \rangle$ increases linearly with N.

We also analyze the scaling properties of the model M_l. Figure 5 shows the normalized connectivity (k/N) versus the normalized index (i/N) for the model M_5. This curve is similar to the one of Fig. 3; here we also use $N=128,512$ and 2048. The data collapse performed in the figure indicates that the network M_5 scales with N. In fact, this same tendency is observed for all networks M_l analyzed. This behavior is related, as before, with the plateaus of prime numbers p_j whose connectivity scales with N. The fact that the degree distribution of the networks M_l scales with N suggests the existence of magnitudes that are independent of N. This is the case of $\langle k \rangle/N$ and this is also the case of the clustering coefficient as we shall see in the following section.

III. CLUSTERING COEFFICIENT AND NETWORK DISTANCE

In this section we characterize the network models M_l using the distance d and the clustering coefficient C. One of our objectives in this work is to differentiate M_l from random networks; it means networks whose distribution of links

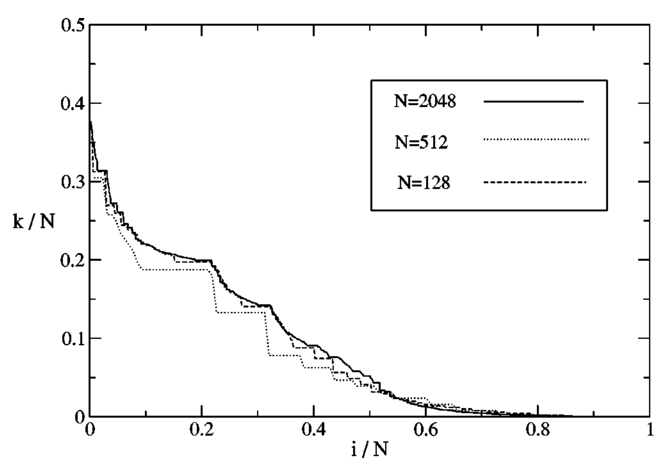

FIG. 5. The normalized connectivity k/N vs normalized index i/N for the data of model M_5. It is used $N=128$, 512, and 2048 as indicated in the figure.

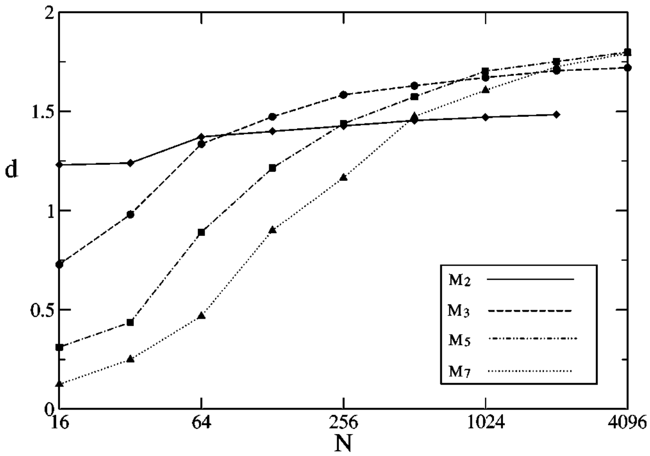

FIG. 6. The distance d against network size N, for the models M_2, M_3, M_5, and M_7 as indicated in the figure.

among vertices follow a Poisson distribution. Therefore we compare d and C of M_l with d and C of the random network associated; it means the random network with the same number of vertices and connections.

The distance of a network is defined as the average distance between all the two vertices of the network. The clustering coefficient is a global parameter $C = \Sigma_N c_i / N$ which is based on the local clustering coefficient c_i. For each vertex i the respective c_i is defined as the normalized number of connection among its first neighbors. The parameter C measures the average interconnection of the network. Using the example of social networks of acquaintances, c_i measures how much the friends of a person (vertex i) are friends among them. For a major treatment on this topic see Ref. [2].

The distance of the random network associated, d_{rand}, is estimated by simulation. We note that, because the graph is not sparse, it is not valid that $d \propto \ln(N)$. In fact, for non-sparse graphs the distance is almost always 2 [17]. Otherwise, the clustering coefficient of the random graph associated, C_{rand}, is analytically estimated [2] and depends only on the number of vertices N and the number of connections n. For the random graph the clustering coefficient is $C_{rand} = \langle k \rangle / N$. We call $\overline{d} \equiv d/d_{rand}$ and $\overline{C} \equiv C/C_{rand}$ as the normalized distance and clustering coefficient.

Figure 6 shows the distance d against the network size N for the data of the models M_2, M_3, M_5, and M_7. The graphic is in log-linear form because of the large interval used in N. The data confirm the prevision for nonsparse graphs that the distance is around 2. We estimate $\overline{d}$ for the networks M_2, M_3, M_5, and M_7 in the range $2^5 \leqslant N \leqslant 2^{12}$ and find that $0.5 < \overline{d} < 0.9$. The general tendency is that $\overline{d}$ slowly increases with N. In addition, l large in models M_l implies smaller $\overline{d}$. This last fact is expected since for large l the connections of the network are more selective and organized.

The analysis of $\overline{C}$ for M_2, M_3, M_5, and M_7 shows that it increases with N until $N \simeq 2^6$ and stabilizes around a constant value. In Table II $\overline{C}$ for several N for the model M_5 is shown; the other models M_l follow a similar trend.

TABLE II. $\overline{C}$ for several network sizes N for the model M_5.

N	32	64	128	256	512	1024
$\overline{C}$	6.95	8.33	8.18	8.19	8.28	8.36

The data point to a constant $\overline{C}$ in the limit $N \to \infty$. The size invariance of $\overline{C}$ is compatible with the size invariance of the degree distribution. The best values of $\overline{C}$ for M_2, M_3, M_5, and M_7 are shown in Table I. We observe in this table that $\overline{C}$ for the network M_l increases with l. The clustering coefficient increases as the criterion for establishing connections in the network becomes more selective. In fact, for a constant N, the number of connection n in the M_l model decreases for high l (see Table I). As a result $C_{rand} = \langle k \rangle / N = 2n/N^2$ decreases in contrast with estimated C that remains almost constant.

IV. FINAL REMARKS

In this work we propose a network model in which the natural numbers are the vertices and the connections are based on their decomposition by prime numbers. Using this criterion we develop a nonsparse network ($\langle k \rangle \sim \ll N$) which has a distance of the order of 2. If we consider that all the prime numbers in the decomposition set a link the network formed is similar to a random graph because the high number of connections implies a small clustering coefficient. If the criterion for establishing links becomes more selective, only prime numbers greater than 3, or 5, are used to establish links where the network has a large clustering coefficient.

We perform data collapse on the data and verify that the networks studied have a degree distribution that is invariant with the number of vertices N. The general view of the degree distribution is a funny discontinuous curve with plateaus of all sizes. These plateaus are generated by the families of numbers that share the same prime numbers in their decomposition.

An important class of networks are the scale-free ones. This concept is mainly used to distinguish between networks that have exponential and power-law degree distributions. In the exponential case most of the vertices have a typical connectivity inside a range defined by the exponent of the exponential function. On the one side, a power-law case has connections of all orders; it does not have a set of vertices with a typical connectivity. The number network M_l does not have a smooth degree distribution because of the plateaus formed by the families of prime numbers F_k, therefore it is not possible to fit the degree by a smooth curve. On the other side, due to families F_k, this network has vertices with all orders of connectivity corresponding to all sizes of primes and their combinations. In this broad sense the network M_l can be called a scale-free network.

This work unfolds an alternative perspective in the study of complex networks. Instead of search for real networks in nature we explore deterministic mathematical networks that show small distance and high clustering coefficient. Despite the present network is non being sparse it is a promising laboratory in the study of degree distribution and cluster families. In a future work we intend to explore some theoretical developments of this problem: an evolving network algorithm for M_l and an analysis of phase transition in this model.

ACKNOWLEDGMENTS

The author thanks Marcos Vinicius C. Henriques, Sidney Redner, Liacir S. Lucena, for the fertile exchange of ideas. The author also acknowledges the financial support of Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq)-Brazil and the use of the following software for network analysis: http://vlado.fmf.uni-lj.si/pub/networks/pajek/.

[1] D.J. Watts and S.H. Strogatz, Nature (London) **393**, 440 (1998).
[2] R. Albert and A-L Barabási, Rev. Mod. Phys. **74**, 47 (2002).
[3] M. Faloutsos, P. Faloutsos, and C. Faloutsos, Comput. Commun. Rev. **29**, 251 (1999).
[4] R. Albert, H. Jeong, and A.-L. Barabási, Nature (London) **401**, 378 (1999).
[5] A.E. Motter, A.P.S. de Moura, Y-C. Lai, and P. Dasgupta, Phys. Rev. E **65**, 065102(R) (2002).
[6] R. Ferrer i Cancho and R.V. Solé, Proc. R. Soc. London, Ser. B **268**, 2261 (2001).
[7] H. Jeong, S.P. Mason, Z.N. Oltvai, and A.-L. Barabási, Nature (London) **411**, 41 (2001).
[8] J.M. Montoya and R.V. Solé, J. Theor. Biol. **214** (2002).
[9] G. Corso, L.S. Lucena, and Z.D. Thomé, Physica A **324**, 430 (2003).
[10] W. Souma, Y. Fujiwara, and H. Aoyama, Physica A **324**, 396 (2003).
[11] M.E.J. Newman, Phys. Rev. E **64**, 016131 (2001).
[12] F. Liljeros, C.R. Edling, Luis A. Nunes Amaral, H.E. Stanley, and Y. Aberg, Nature (London) **411**, 907 (2001).
[13] P.L. Krapivsky and S. Redner, Comput. Netw. **39**, 261 (2002)
[14] S.N. Dorogovtsev and J.F.F. Mendes, Adv. Phys. **51**, 1079 (2002).
[15] R. Cohen, K. Erez, D. Ben-Avraham, and S. Havlin, Phys. Rev. Lett. **85**, 4626 (2000).
[16] D.S. Callaway, M.E.J. Newman, S.H. Strogatz, and D.J. Watts, Phys. Rev. Lett. **85**, 5468 (2000).
[17] B. Bollobas, *Random Graphs* (Academic Press, London, 1985).