

Characterization of complex networks by higher order neighborhood properties

R.F.S. Andrade^{1,a}, J.G.V. Miranda¹, S.T.R. Pinho¹, and T.P. Lobão²

¹ Instituto de Física, Universidade Federal da Bahia Campus Universitário da Federação, 40210-340 - Salvador - BA, Brazil

² Instituto de Matemática, Universidade Federal da Bahia Campus Universitário da Federação, 40130-240 - Salvador - BA, Brazil

Received 20 August 2007

Published online 6 February 2008 – © EDP Sciences, Società Italiana di Fisica, Springer-Verlag 2008

Abstract. A concept of higher order neighborhood in complex networks, introduced previously [Phys. Rev. E **73**, 046101 (2006)], is systematically explored to investigate larger scale structures in complex networks. The basic idea is to consider each higher order neighborhood as a network in itself, represented by a corresponding adjacency matrix, and to settle a plenty of new parameters in order to obtain a best characterization of the whole network. Usual network indices are then used to evaluate the properties of each neighborhood. The identification of high order neighborhoods is also regarded as intermediary step towards the evaluation of global network properties, like the diameter, average shortest path between node, and network fractal dimension. Results for a large number of typical networks are presented and discussed.

PACS. 89.75.Fb Structures and organization in complex systems – 89.75.Hc Networks and genealogical trees – 02.10.Ox Combinatorics; graph theory

1 Introduction

In recent years, complex networks have become one of the major research themes in complex systems. They allow for an interesting interplay between mathematics and physics through the well-established theories of graphs and statistical mechanics [1–3]. Based on concepts developed within these theories, several concepts were developed to characterize the networks.

The huge development in this research area comes along with the proposition of parameters that capture some essential properties that such objects may have. Most of the current works in the literature expresses the quantitative results in terms of a hand full of indices or quantifiers [4–6], like the average number of links per node $\langle k \rangle$, clustering coefficient C , mean minimal distance among the nodes $\langle d \rangle$, diameter D , and the assortativity degree A . Also of relevance are certain relations that express how these quantifiers are composed, like $p(k)$, the probability distribution of nodes with k links, the distribution of each individual node clustering coefficients $C(k)$ with respect to its node degree, and the average degree of the neighbors of a node with k links $\langle k_{nn}(k) \rangle$.

Nevertheless, other parameters can be proposed to account for other properties that are overlooked by the above quoted set. The actual relevance of the information brought by a new parameter depends whether it is, in a

certain sense, *orthogonal* to the set of parameters that had been used so far. In such case, its evaluation contributes either to emphasize similarities among a set of networks, or to uncover differences among them. Therefore, new parameters may indeed become relevant to identify differences between networks that could have been thought to be similar in a more restricted space of characterization parameters. With the help of pertinent correlation analysis among distinct measures, it is possible to decide whether the new parameter provides additional information on the network structure [7]. Also, such analyzes make it easier to identify whether a given network can be casted into one of (or more) some typical network classes that have been systematically studied.

From our point of view, it is important to probe the network with respect not only to the immediate, close neighborhood of a node, but also how all pairs of nodes are related among themselves, from the nearest up to the maximal distance D .

In a previous work [8], we have indicated how the adjacency matrix M of a network R can be used to obtain all higher order neighborhoods of a node in R with the help of Boolean product of matrices. According to the definition introduced there, two nodes in R are neighbors of order $O(\ell)$, $\ell = 1, 2, \dots, D$, when the shortest path connecting them, along links in R , has ℓ steps. For the purpose of completeness, it is interesting to define that each node is neighbor of order $O(0)$ of itself.

^a e-mail: randrade@ufba.br

It is worth mentioning that other authors [9–17] also defend similar positions to ours, having proposed several strategies to investigate network properties that advance beyond the immediate vicinity of a node. However, as it shall be clear in the next sections, our approach is quite different from those just presented in the related literature.

The aim of the current work is to characterize the large structure properties of R by a larger set of measures, which are based on the higher order adjacency matrices $M(\ell)$ describing the corresponding neighborhoods $O(\ell)$. The process of evaluating all $M(\ell)$'s is of importance as it unfolds, in a straightforward way, a large amount of information contained in M . It can be used for the analysis of the higher order neighborhood properties, like the analysis of the spectral density [8]. There we exemplify the consistency and usefulness of the proposed framework, characterizing the eigenvalue spectra of matrices $M(\ell)$ for several types of networks. For each $M(\ell)$, we apply the same procedures used on M to evaluate the quoted set of indices to characterize R . This way, we obtain new measures $\langle k(\ell) \rangle$, $C(\ell)$, $p(k; \ell)$, $A(\ell)$, and the distribution $C(k; \ell)$. The ℓ dependent measures are useful to express topological properties and dynamical processes in R that can only be achieved by moving exactly ℓ steps from each node.

Nevertheless, this framework offers an alternative topological interpretation for our results, based on the fact that the $M(\ell)$'s can also be regarded as describing a set of networks $R(\ell)$. Each of them is formed by the same set of nodes as R , but the set of edges in $R(\ell)$ links pairs of nodes that first reach each other after ℓ steps in R . Within this interpretation, the ℓ dependent measures describe the immediate neighborhood of hypothetical $R(\ell)$ networks.

As already quoted, this framework goes along several recent efforts [9–17] to characterize networks through the connection of a node to its more distant neighbors. Indeed, some of proposed measures, as the set $C(\ell)$, are closely related to similar clustering coefficients introduced to identify whether two distant neighbors of a node are first order neighbors of each other [16]. Measures quantifying how the number of more distant neighbors behave with respect to ℓ , discussed recently in the context of social networks [9,17], are connected to $\langle k(\ell) \rangle$ quoted above.

The exact values of the global network parameters D and $\langle d \rangle$ are readily obtained after the decomposition process. In addition to that, the adopted procedure stores in \overline{M} , valuable information on the neighborhood structure that is required for the evaluation of the fractal dimension d_F of a network, according to a scheme [18] proposed recently.

The rest of this work is so organized: in Section 2 we briefly review the main steps required to describe the higher order networks, and discuss how the quantities obtained in this process can be used to describe and represent other properties of networks. Section 3 is divided into subsections. In each of them, we discuss, respectively, the behavior of $C(\ell)$, $\langle k(\ell) \rangle$, $p(k; \ell)$, $C(k; \ell)$, $A(\ell)$, and d_F for a large number of networks. In this work we consider

networks generated by well defined algorithms, as those proposed by Erdős and Renyi (ER) [19], Barabasi and Albert (BA) [20], and Newman [5]. As this later corresponds to a slightly changed version of the original Watts and Strogatz (WS) [21] algorithm, we denote it by WS-N. We also include deterministic networks, like the Cayley tree with coordination number 3 (CT), the Apollonian network (AN) [22] and two well known structures that have been largely used in the context of spin models in statistical mechanics [23], namely, the Diamond Hierarchical Lattice (DHL) and the Wheatstone Hierarchical Lattice (WHL). Whenever pertinent, we include a comparison of the values we obtained with those for similar parameters proposed in other works to characterize the topology of more distant neighbors already proposed by other authors. Finally, in Section 4, we close the work with final remarks and conclusions.

2 Evaluating the adjacency matrices $M(\ell)$

As explained in the Introduction, each $R(\ell)$ is described by $M(\ell)$, $\ell \in [0, D]$, with the following property:

$$M(\ell)_{ij} = \begin{cases} 1, & \text{if } j \in O_i(\ell) \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

where $O_i(\ell)$ is the set of ℓ -neighbors of node i . Note that this definition also describes $M(0) = I$, the identity matrix. Once the matrix elements of each $M(\ell)$ are either 0 or 1, it is possible to use well known Boolean (B) operations of sum \oplus , subtraction \ominus and multiplication \otimes [24] to successively evaluate all elements of $M(\ell)$, $\ell = 1, 2, \dots, D$.

This can be easily accomplished in a two step process, that can be illustrated when $\ell = 2$. First note that, if the matrix element (i, j) of $\overline{M(2)} = M(1) \otimes M(1)$ is equal to 1, we know that it is possible to go from node i to j by a two-step walk. $\overline{M(2)}$ distinguishes itself from $M(1)^2$, the usual matrix product of $M(1)$ by itself, as it indicates the existence of two-step walks, but not the multiplicity of walks informed by the later. In the sequence, we note that $\overline{M(2)}$ includes more elements 1 than necessary to account for all pairs of nodes for which the shortest path equals to 2. Indeed, $\overline{M(2)}$ has 1's in the whole main diagonal, as well as between nodes i and j that, together with a third node k , are connected in a triangle. The first set of 1's in the diagonal is already described by $M(0)$, while the second is already present in $M(1)$. Taking these observations into account, it is possible to obtain

$$M(2) = (M(1) \oplus \overline{M(2)}) \ominus (I \oplus M(1)), \quad (2)$$

which can be rewritten in a more convenient form as

$$M(2) = (I \oplus M(1)) \otimes M(1) \ominus (I \oplus M(1)). \quad (3)$$

Equation (3) can be easily generalized by finite induction for arbitrary value of ℓ by:

$$M(\ell) = \left(\bigoplus_{g=0}^{\ell-1} M(g) \right) \otimes M(1) \ominus \left(\bigoplus_{g=0}^{\ell-1} M(g) \right). \quad (4)$$

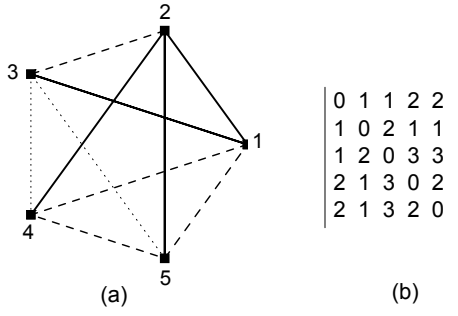


Fig. 1. (a) Small network with 5 nodes and 5 edges (solid lines). Neighbors of second and third order are connected, respectively, by dashed and dotted lines. (b) Network neighborhood matrix $\widehat{\mathbf{M}}$.

In Figure 1a we illustrate the neighborhoods of all orders in a simple network with 5 nodes and diameter $D = 3$. Besides the direct links in solid black lines, we also draw the connections of order $\ell = 2$ and 3.

The evaluation of all $M(\ell)$'s opens the path for a direct evaluation of many network parameters. Many of them are neighborhood dependent and, as far as we know, have not been considered before. On the other hand, the evaluation of some global parameters, which have usually been performed along other methods, can be obtained within this framework in a straight forward way. Let us first note that, once N is finite, for some large enough ℓ_{max} we find $M(\ell) \equiv 0, \forall \ell > \ell_{max}$, so that $D \equiv \ell_{max}$. Next, the knowledge of all $M(\ell)$ allows for the definition of the matrix $\widehat{\mathbf{M}}$, which carries all information on the shortest path between any two vertices i and j along the network. As, according to (1), $M(\ell)_{ij} = 1$ for only one ℓ , the definition

$$\widehat{\mathbf{M}} = \sum_{g=0}^D gM(g), \quad (5)$$

implies that, for all pairs $(i, j) \in O(\ell)$, $\widehat{\mathbf{M}}_{ij} = \ell$. In Figure 1b we show $\widehat{\mathbf{M}}$ for the simple network drawn in Figure 1a. Note that $\widehat{\mathbf{M}}$ is only a bit differently defined from the so called *distance matrix* [25] used in the graph theory; it is reduced to that one in the case of a connected network. With the help of (5), the average minimal path for a node i is easily expressed by $d_i = (\sum_{j=1}^N \widehat{\mathbf{M}}_{i,j}) / (N - 1)$, leading, immediately, to the average $\langle d \rangle$.

Despite the fact that several complex networks have an intrinsic length scale, represented e.g., by $\langle d \rangle$ or D , there have been some attempts to associate fractal dimensions to these objects. The definition for a fractal dimension d_F , cited in the introduction, is deeply related to the concept of $O(\ell)$ neighborhoods. Indeed, once length is measured in R by the number of steps between nodes, scaling arguments lead to

$$L(\ell) \sim \ell^{d_F}. \quad (6)$$

In this expression, $L(\ell)$ counts the number of boxes required to cover all nodes in the network such that: (i) any node is covered only by a single box; (ii) any two nodes i

and j in a given box, are $O(\ell')$ neighbors of each other, with $\ell' \leq \ell$. Thus, the knowledge of the matrices $M(\ell)$ constitutes an important step towards the evaluation of the set of boxes for each value of ℓ .

The evaluation of large scale structures in the network proceeds by considering each $R(\ell)$, represented by $M(\ell)$, as an independent network. Therefore, the parameters $C(\ell)$, $\langle k(\ell) \rangle$, $p(k; \ell)$, $C(k; \ell)$ and $A(\ell)$, which describe the local information on each $R(\ell)$, also provide information on the large scale structures of R . For any value of ℓ and node i , the node degree $k_i(\ell) = \sum_{j=1}^N M(\ell)_{i,j}$ and the node clustering coefficient $C_i(\ell) = \sum_{m \in O_i(\ell)} \sum_{j=1}^N M(\ell)_{i,j} M(\ell)_{m,j} / 2$ are directly expressed in terms of elements of $M(\ell)$, while the other three quantities follow immediately by counting the number of occurrences of nodes with degree $k(\ell)$.

To characterize a data set represented as a network, one has to know whether all points are indeed connected among themselves in a single component, or partitioned into disjoint sub-networks. This important large scale property can be exactly answered provided the set $M(\ell)$ is evaluated. If the network consists of a single component, the quantity Z , defined by

$$Z \equiv \sum_{\ell=1}^D \sum_{j=1}^N M(\ell)_{i,j}, \quad (7)$$

always assumes the value $N(N - 1)$. If $Z < N(N - 1)$, two or more components are present and, in this case, their number and corresponding sizes can be evaluated as follows. First evaluate $\kappa_i = 1 + \sum_{\ell=1}^D k_i(\ell) < N - 1, \forall i$. It counts the number of nodes in the specific component the node i belongs to, but it also indicates possible sizes for any other component. The maximal number of distinct values assumed by κ_i is limited by $(-1 + (1 + 8N)^{1/2}) / 2$. If $\sigma(\kappa)$ represents the number of nodes that share the same value of κ , the number of components of this size is simply $\sigma(\kappa) / \kappa$, what completes the characterization on the partition of the network. The adjacency matrix of a non connected network, by a suitable rearrangement of its nodes, may be reduced to a form of non-zero diagonal blocks, i.e. a direct sum of matrices of smaller order, each block corresponding to a connected component of the network. Because of this, one may just deal with connected networks, as we shall consider herein. We would like to emphasize that, for the sake of simplicity, we will restrict ourselves to present results only for undirected networks, without self links, and parallel links between any two nodes.

3 Neighborhood characterization

In this section we present a characterization of $O(\ell)$ neighborhood for some standard networks. A flavor of this procedure is available in a previous work [8], where we concentrated on the spectral properties of the networks $R(\ell)$. A summary of our most interesting results are depicted in Figures 2–7, where we draw the parameters quoted in the

previous section as function of the neighborhood ℓ . The results were obtained for geometrically grown networks (CT, AN, DHL, WHL), but also for networks generated by precisely defined algorithms (ER, BA and WS-N). They display features that allow to associate them with different complex network scenarios: small world ($\langle d \rangle \lesssim \log N$, “large” C , mostly assortative); scale free ($p(k) \sim k^{-\gamma}$, “small” C , dissortative); random ($p(k) \sim$ Gaussian distribution, “small” C , anassortative) [6]. This investigation indicates the properties of ℓ -neighborhoods for each of these scenarios. We discuss results for networks ranging from $N \sim 100$ up to a maximum of $N = 10\,000$ nodes, the choice of N depending on which aspect must be emphasized. As will be exemplified for some specific situations, our results remain quite independent of the size of the network, provided we scale properly the linking probabilities with N for those networks generated by random algorithms. All of them are prototypes of many other natural and social networks.

3.1 Clustering coefficient

The physical meaning of the coefficients $C(\ell)$ is closely related to that of the usual clustering coefficient. As, for each value of ℓ , the connection among nodes corresponds to a connection through ℓ steps in the original network, $C(\ell)$ measures how the neighbors of higher order of a node are connected among themselves along ℓ step paths. The increase or decrease in the value of $C(\ell)$ with respect to $C(\ell = 1)$ indicates whether the nodes become more likely to build cliques or not. As we shall see in the following, the behavior is strongly dependent on the network features, so that these measures uncover new information about node connectivity that can not be displayed by the single parameter $C(1)$.

Figure 2 shows how $C(\ell)$ behaves for the distinct networks. For the CT, $C(\ell)$ oscillates between zero and finite values for odd and even values of ℓ . This is explained by the fact that, as ℓ grows, the $R(\ell)$ networks assume, alternately, the structures of loop-less trees and Husimi cactuses, which entail a lot of triangles. This network has relatively large diameter, so that this behavior is sustained for several periods. Finite size effects cause a small decrease of the value of $C(\ell)$, what is enhanced when ℓ approaches D . Similar behavior is observed in many other situations, as the regular hyper-cubic lattices of dimension larger than 1, where triangles appear only for even ℓ 's. It is also present in hierarchical structures as DHL. On the other hand, WHL behaves differently only for $\ell = 1$, when $C(1) = 0.57$, due to the presence of triangles. However they are absent for all larger order odd $\ell > 1$, so that an oscillatory pattern between 0 and non-zero values of $C(\ell)$ sets in.

A second common pattern for $C(\ell)$ is that of curve with a well defined maximum, taking finite values only over a finite range. It is found for BA networks generated according to the standard procedure [20], as well as for ER networks when, for instance, $N = 1000$ and linking

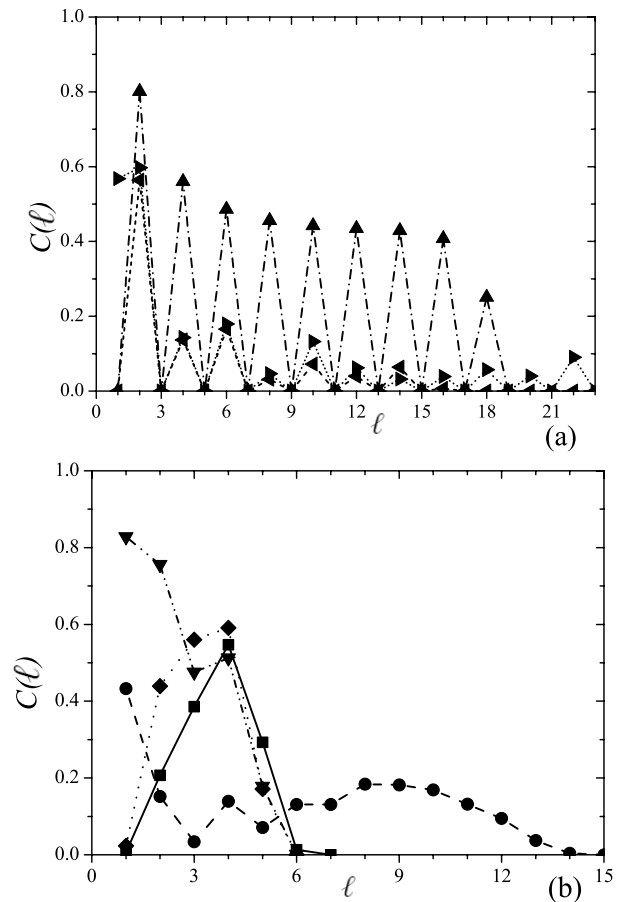


Fig. 2. Dependence of $C(\ell) \times \ell$ for the seven distinct networks. Conventions for (symbol, line) introduced here and used in the rest of this work are as follows: CT (up triangle, dash dot); AN (down triangle, dash dot dot); DHL (left triangle, short dash); WHL (right triangle, short dot); BA (diamond, dash); ER (square, solid); WS-N (circle, dash). The data were obtained for following values of N : (a) 1534 (CT), 684 (DHL), 1564 (WHL). (b) 3283 (AN), 2000 (BA); for ER, $N = 1000$ and $p_a = 0.008$; for WS-N, $N = 1000$, $p_r = 0.2$.

probability $p_a = 0.008$. For the BA networks, this pattern emerges because neither the dominant hubs nor the nodes with lower k are likely to be connected in cliques. However, the hubs induce many lower k 's nodes to form cliques at second and third order neighborhoods, explaining the sharp increase in the value of $C(\ell)$. The decrease in the value of $C(\ell)$ as $\ell \rightarrow D$ reflects the fact that the most part of all pairs of nodes have already been considered. For ER networks, similar arguments apply.

A third kind of pattern has also been observed for networks that are characterized by a large value of $C(1)$ as, e.g., the AN and WS-N networks. In the first case we start from a large value of $C(1) = 0.828$, following an almost monotonic decrease of $C(\ell)$. The same is observed for relatively large values of the rewiring probability p_r in WS-N networks.

We find important to notice, however, that if the behavior for the deterministic networks remain essentially the same as N grows, the same is not observed for WS-N

and ER if we keep N constant and decrease the values for, respectively, p_r and the linking probability p_a . For both situations, oscillatory pattern between non-zero values emerges, indicating a similar picture to those of geometrically constructed networks.

It is worth mentioning that other authors also proposed some generalized extensions of the notion of the clustering coefficient. In [10], it is investigated the issue if two neighbors of a given node r are far from each other with a chosen distance x that do not pass through r . Differently, in [16], the author asks if two nodes, at a chosen distance ℓ of a node r , are neighbors of each other. Of course it is possible to generalize the clustering coefficient definition, considering the possibilities of two nodes at distance ℓ of r being connected by t steps, where $t = 1, 2, \dots, 2\ell$. Our measure $C(\ell)$ corresponds to $t = \ell$, whereas that one in [16] considers $t = 1$. Finally, there are also more ways to interpret the concept of neighborhood of a given node [12]. However, despite the similarities among the distinct measures, our approach seemingly introduces new specificities, which are different from those of the known literature.

3.2 Average degree

The measure $\langle k(\ell) \rangle$ has a more direct physical meaning than the previous one, as it does not involve the connectivity among neighbors, but only counts how the number of neighbors at a given distance ℓ of a node varies with ℓ . For regular lattices, the higher order neighbors $\langle k(\ell) \rangle$ of a site are roughly located on the surface of a hyper-sphere. Their number grows as ℓ^{dim-1} , where dim denotes the Euclidean dimension of the lattice. For the extreme situation of an exact CT, where the number of sites on the surface is of the same order of magnitude as those in the bulk, corresponding to $dim \rightarrow \infty$. This is reflected by an exponential increase of $\langle k(\ell) \rangle$ with ℓ , as shown in Figure 3a.

A similar behavior is expected to be found for many other complex networks, which have a very small diameter. In Figure 3 we illustrate the behavior of $\langle k(\ell) \rangle$ for the same set of networks, showing that exponential increase is present for the majority of them. However, as the network is finite, $\langle k(\ell) \rangle$ must decrease for sufficiently large value of ℓ , making it difficult to assign a precise behavior for the dependence between $\langle k(\ell) \rangle$ and ℓ for these networks. In Figure 3a we find deviations from the exponential increase already when $\ell = 4$ for some networks. The DHL and WHL (Fig. 3b) show a different pattern, consisting of an oscillating period 2 behavior, which is caused mainly by the contribution of the nodes introduced in the last hierarchy. For instance, in the DHL, these nodes are connected only to two other nodes but, for $\ell = 2$ and 3, they can have up to 9 and 4 neighbors. This situation is repeated for larger values of ℓ and also for the WHL.

Results for higher order degree have been considered previously, as in the attempt to express $\langle k(2) \rangle$ as function of $\langle k(1) \rangle$ in the context of egocentric networks [9]. However, the analysis does not go beyond $\ell = 2$. In [17],

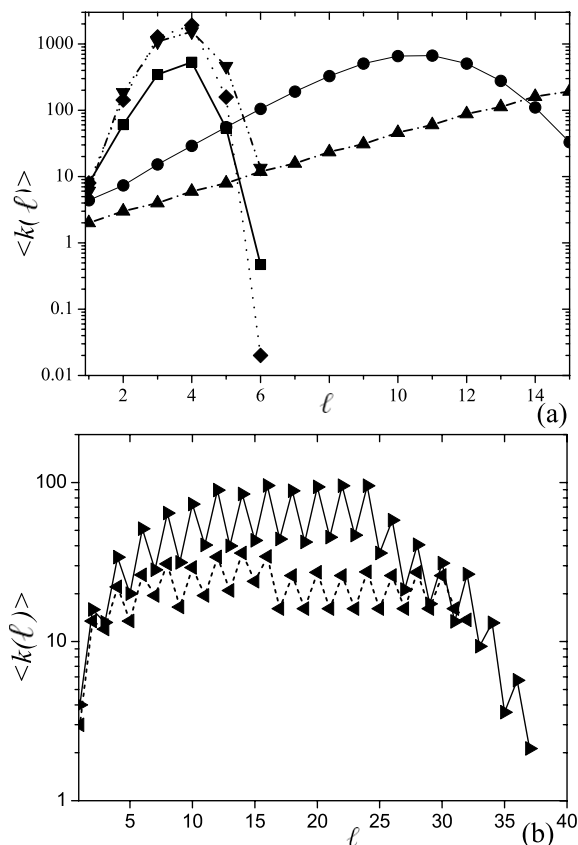


Fig. 3. Dependence of $\langle k(\ell) \rangle \times \ell$ for distinct networks. Symbols and lines follow the same convention as in Figure 2. (a) Values of N are the same used in Figure 2 for CT, AN, BA and ER, with $p_a = 0.008$. For WS-N, $N = 3500$ and $p_r = 0.2$. (b) Same values of N as in Figure 2 for DHL and WHL.

the authors define and evaluate the hierarchical degree of order ℓ , which is related, but not equivalent, to that one resulting in our framework. Nevertheless, the overall dependence of this parameter with respect to ℓ , follow similar behavior as those displayed in Figure 3.

3.3 Node distribution probability

The usual node distribution $p(k; \ell = 1)$, evaluated after the knowledge of the degrees of each individual node, gives a picture about the way the nodes are related to each other, e.g. whether there are nodes with preferential attachment, or whether all of them are, on the average, similar to each other. For general ℓ , $p(k; \ell)$ provides the same information on the number of nodes connected by ℓ steps. As we will see below for our analysis of prototypical networks, the same distribution law can subsist or not for values of ℓ other than 1, what can be useful to the classification of actual networks. Let us note that $p(k; \ell = 1)$ usually displays a bell shaped form for many kinds of networks found in nature, as well as those generated by the ER (Fig. 4a) and WS-N (not shown) algorithms. Social and natural BA networks, the features of which we reproduce by the preferential attachment growth algorithm,

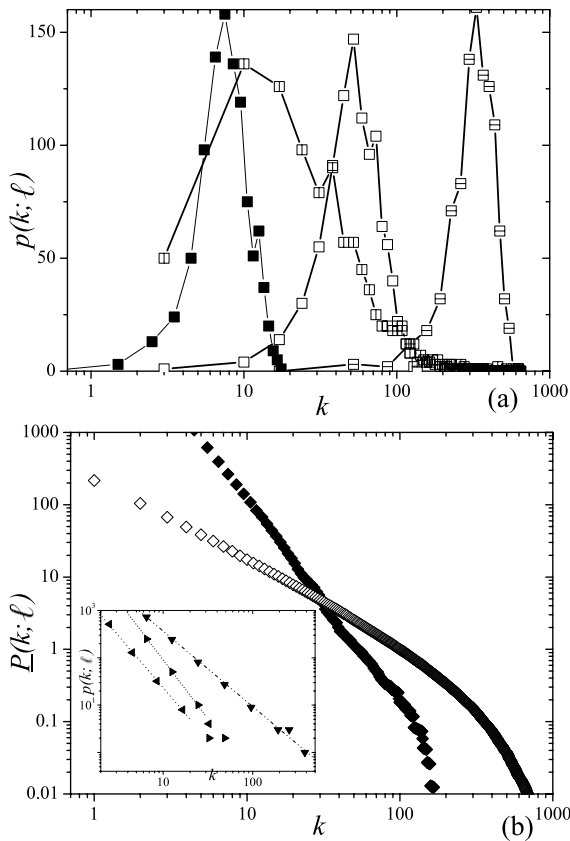


Fig. 4. Dependence of the probability distribution $p(k; \ell)$ and cumulative probability distribution $\underline{P}(k; \ell) \times k$ for distinct networks. (a) Curves for ER with same values of N and p_a used in Figures 2 and 3, when $\ell = 1$ (solid), 2 (hollow), 3 (horizontal dash) and 5 (vertical dash) squares. (b) In the large panel, $\underline{P}(k; \ell)$ follows power law for BA (when $\ell = 1$ (solid) and $\ell = 5 = D - 1$ (hollow)); in the inset, points indicate power law decay for $p(k; \ell = 1)$ in AN, DHL and WHL.

are representative of another possible pattern, characterized by $p(k; 1) \sim k^{-\gamma}$. For the purpose of a clearer picture, there we draw $\underline{P}(k; \ell) = \frac{1}{k} \int_k^\infty p(k'; \ell) dk'$ for the usual BA, reproducing the value $\gamma \simeq 3$, as shown in Figure 4b for $\ell = 1$. In our investigations, we also find similar behavior for the AN, DHL and WHL networks, as shown in the points for $p(k)$ in the inset of this figure. The CT is a trivial case where $p(k; 1)$ reduces to a single point.

For $\ell > 1$, our results show that the bell shaped pattern for $p(k; \ell)$ is reproduced when $\ell > 1$ for the ER and WS-N networks (see Fig. 4a). As shown before, the average node degree increases with ℓ , so that it is natural that the curves for $p(k; \ell)$ are shifted towards larger values of k . The situation changes only for sufficient large $\ell \simeq D$, when finite size effects become relevant. Then, the majority of nodes have already reached their most distant neighbors while others miss only few connections. In this situation, the few non zero contributions to $p(k; \ell \simeq D)$ come for values of $k \sim 1$, so that the peak of the curve is shifted to the region close to the origin, as shown, in Figure 4a, for $\ell = 5$.

For the BA networks, the regions where the higher order $p(k; \ell)$'s receive significant contributions are also pushed to large values of $k(\ell)$. In general, the distributions loose the power law behavior, assuming distinct forms as ℓ increases. Exceptions are provided, e.g., in the example shown for the cumulative distribution $\underline{P}(k; \ell)$ in the large panel of Figure 4b. There we find a very interesting return to a power law distribution in the region of low values of k when $\ell = D - 1$, with an exponent $\gamma \simeq 1.2$. It is important to recall that, once we analyze networks with a finite number of nodes N , the sum over ℓ of individual values $k_i(\ell)$ is bounded to $N - 1$. Thus, for nodes with large values of $k_i(\ell)$ for low ℓ , this measure has a tendency to decrease as ℓ increases and the other way around. Thus, the power law decay for $\ell = D - 1$ observed in Figure 4b is due to a special behavior of the low degree nodes within the BA network. Note that other networks that satisfy the scale free distribution, like AN, WHL and DHL do not show similar behavior. As before, this measure can help identifying whether a given network based on empirical data can be accurately described by those generated by proposed algorithms.

3.4 Hierarchical property

There are several distinct concepts of hierarchical organization, some of them stemming from the network framework, others from geometrical constructions, self similar fractal sets, etc. In this work, apart from using the word ‘‘hierarchical’’ to denote DHL and WHL, we refer to the concept introduced in [26], according to which a network has the hierarchical property if $C_i(1)$, the clustering coefficient of an individual node i , when are drawn as function of the individual node degree $k_i(1)$, shows a power law decrease. Much as observed with the analysis of the node distribution $p(k; \ell)$, evidences of hierarchical character for $R(\ell > 1)$, networks are rare, and are only found when $R(1)$ has already such character.

For the distinct networks we work with, indications of power law behavior for $\ell \geq 1$ are found for the AN and WHL networks, as illustrated in Figures 5a and 5b. Both networks show $C(k; 1) \sim k^{-1}$ asymptotic dependence. On the other hand, DHL does not show this property since, as discussed before, $c_i(\ell = 1) \equiv 0, \forall i$.

We find slight evidences, for $\ell > 1$, of power law dependence in some subsets of points. In the particular case $\ell = 2$, for the AN, several points are aligned, seemingly building short patches that obey power law decrease. This very peculiar distribution of points is recurrent for all values of N that correspond to a full generation of the construction of AN.

For WHL, several points of the curves for $C(k; 1)$ and $C(k; 2)$ coincide exactly, so that a subset of points of $C(k; 2) \times k$ still follows a power law decay. This is exemplary shown in Figure 5b, for the 7th generation (7814 nodes). To explain the presence of this coincident points, we must take into account that, on increasing the generation of WHL, we add points to the region of large values of k of the $C(k; 1)$ curve, while conserving all but

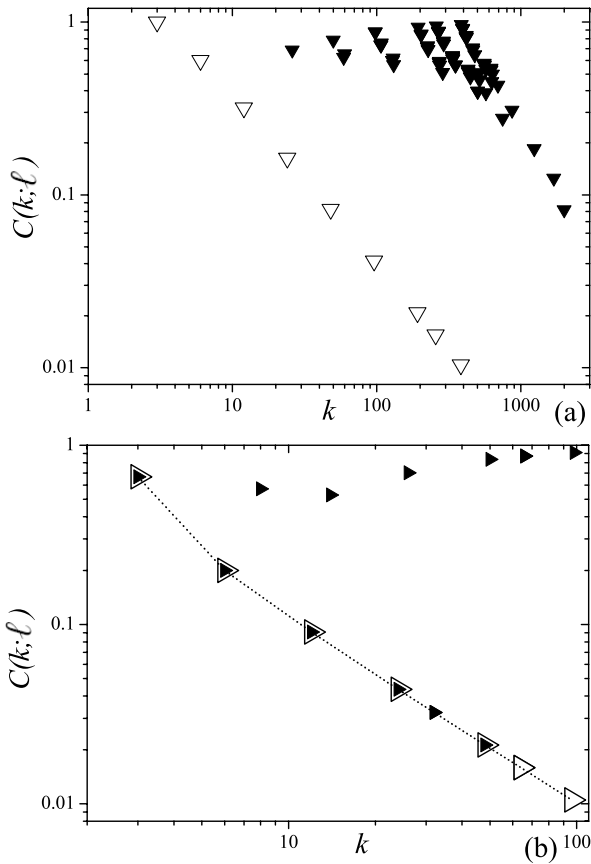


Fig. 5. Data for $C(k; \ell) \times k$ indicating hierarchical arrangement of nodes for AN (a) and WHL (b), respectively with $N = 3283$ and 7814 nodes. Data are drawn for $\ell = 1$ (hollow) and $\ell = 2$ (solid).

one point of the previous generation. In Figure 5b, two newly added points, $k = 64$ and 96 , are absent in the curve for the 6th generation. They correspond to nodes with very large degree, e.g. at the root sites or the intermediate position, where the main bridge connecting two branches is placed. Now, when we look for pairs of second neighbors of the 7th generation, we find that part of them coincides exactly with first neighbors in the 6th generation, so that the $C(k; 2) \times k$ curve of the 7th generation, for this subset, falls on the top the $C(k; 1)$ curve of the 6th generation. The other subset, which contributes to points that fall off the straight line, is formed by pairs of second neighbors that do not correspond to any first neighbors of the previous generation. For larger values of ℓ , we should still observe this truncated pattern only for the even values of ℓ . Indeed, as already discussed before, $C(\ell) = 0$ for odd ℓ 's.

3.5 Assortativity degree coefficient

Several assortativity properties can be assigned to a network [2]. Each of them is quantified by a corresponding coefficient A , which indicates whether the pairs of nodes

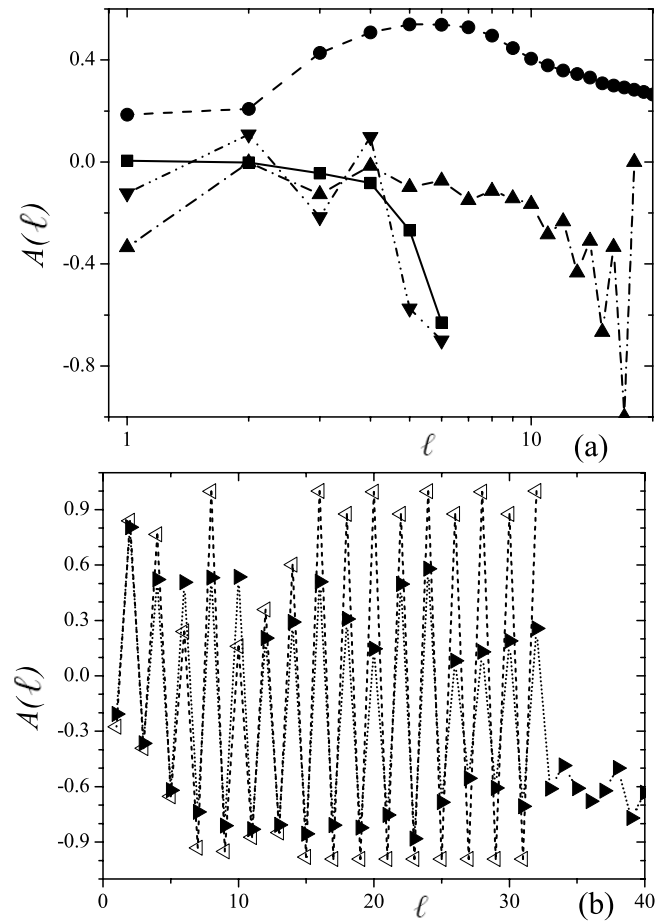


Fig. 6. Dependence of the degree assortativity coefficient $A(\ell)$ with ℓ . (a) Examples of assortative behavior for WS-N network, with $N = 3500$ and $p_r = 0.2$; neutral behavior for ER, with same parameter values as in Figure 2; oscillatory character for AN ($N = 3283$); dissortative behavior for CT ($N = 1534$). (b) Results for DHL and WHL with the same values of N as in Figure 2, indicate $\ell = 2$ periodic oscillations from dissortative to assortative behavior.

directly connected by a link are more likely to behave alike ($A > 0$) or dislike ($A < 0$). The assortativity degree coefficient [27], which takes into account the average degree of the nearest neighbors of a node of degree k represented by $k_{nn}(k)$, probes the degrees of the nodes at each side of a link. Here, we denote by $A(\ell)$ the coefficients that quantify the degree assortativity for the corresponding neighborhoods $O(\ell)$. Each $A(\ell)$ measures whether pairs of nodes, that are ℓ steps apart, are likely to be ℓ -connected to other nodes that have the same ℓ degree $k(\ell)$. As we show below, the assortative character prevails over a large range of values of ℓ for most networks, with the exception of the geometrically constructed AN, DHL and WHL.

The results in Figures 6a, 6b display several kinds of patterns, indicating that the behavior of $A(\ell) \times \ell$ is very sensitive to the type of network. Figure 6a shows that WS-N networks have usual positive assortativity $A(\ell = 1) \simeq 0.2$. This value results from the contribution

of the large majority of sites that are connected to their original neighbors in an ordered structure. On increasing ℓ , $A(\ell)$ remains positive for a large ℓ interval. After this phase, $A(\ell)$ goes through a steep descent to negative values, where it remains until $\ell = D$. This change reveals that the large ℓ neighborhoods completely loose the local character and nodes are overwhelmingly ℓ -connected to nodes with distinct ℓ -degree.

The results for a finite CT, in the same panel, are strongly biased by surface effects. In an infinite tree, all nodes have the same degree, so that $A(\ell = 1) = 1$. However, for any finite tree, a dissortative character is observed, which can be explained as follows: (i) the number of sites added in the last generation has the same order of magnitude of the existing sites; (ii) all of them have a distinct number of neighbors as those added before, the same happening to the nodes that are connected to them. In the evaluation of $A(1)$, nodes with distinct degree on the end of all newly introduced links, contribute negatively, so that these contributions lead to a dissortative character to CT. The same occurs for larger values of ℓ , so that all $A(\ell)$ deviate strongly from the constant value 1 that they assume in an infinite tree. However, the constant value 1 can be recovered, in a finite sized network, if one neglects, successively, the effect of the ℓ th lately added nodes.

For the ER networks, also illustrated in Figure 6a, pure randomness shows neutral behavior, hence $A(1) = 0$. Numerical simulations reproduce this result, which should be valid for several values of ℓ . However, for large ℓ , finite size effects end up by driving $A(\ell)$ to the negative region.

With respect to geometrically grown networks, oscillations between dissortative to assortative character is a common feature for AN, DHL and WHL. In the first situation (Fig. 6a), the amplitude of variation of $A(\ell)$ is not so large and, due to the very small value of D , short lived. On the other hand, for DHL we have very large variations limited only by the extremal values ± 1 (see Fig. 6b). This very peculiar behavior may be explained by noting that, for $\ell = 1$, no node has neighbors with the same degree as itself. On the other hand, when $\ell = 2$, the number of nodes with second order neighbors having the same degree is very large. Thus, oscillations set in, changing the assortative character at each step. For the WHL, very large oscillations are also observed, although they do not reach the extremal values as for DHL. Indeed, in this situation, the presence of the cross bonds lead to a large heterogeneity in the degree of the nodes, what causes a decrease in the amplitudes of oscillations.

3.6 Fractal dimension

As discussed in Section 2, the evaluation of the higher order neighborhoods is an important step in towards covering the network with disjoint boxes, what is required to obtain the network fractal dimension d_F as proposed in [18]. Results for the distinct networks are summarized in Figure 7, where we draw $L(\ell)$, the number of boxes containing nodes which are at maximal distance ℓ of each

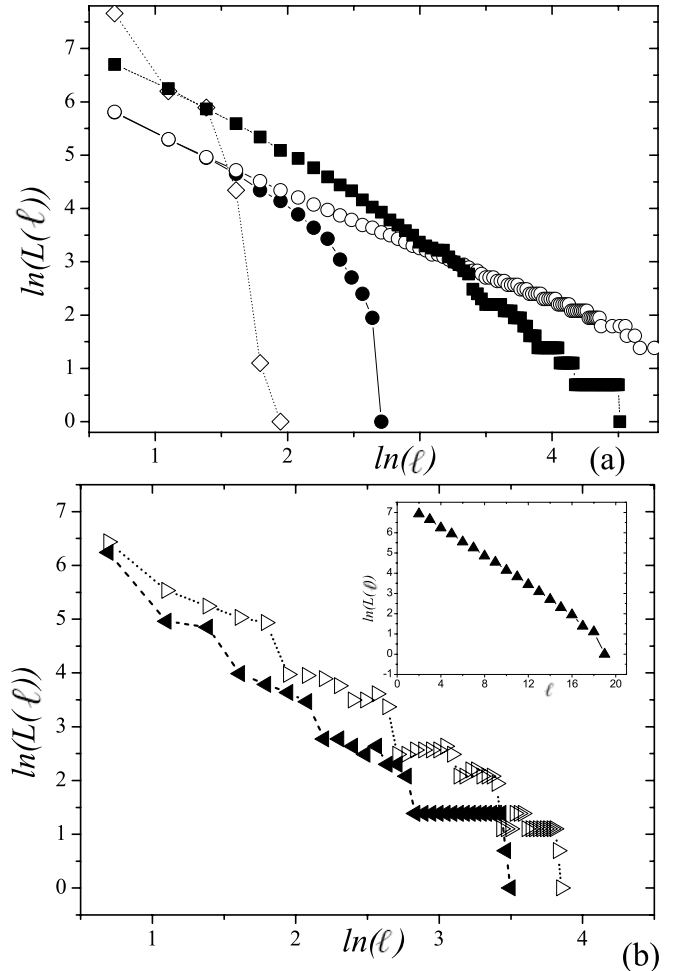


Fig. 7. Dependence of $\ln(L(\ell)) \times \ln(\ell)$. (a) Results for WS-N ($N = 1000, p_r = 0.001$, open circles) and ER ($N = 1432, p_a = 0.0014$) suggest scaling and fractality. On the other hand, no evidence of scaling is found for BA and WS-N, when a larger value of $p_r (= 0.2)$ and same N is used (solid circles). (b) In the large panel, results for DHL and WHL (with same N as in Fig. 2) also suggest scaling. In the inset, results for CT ($N = 1534$), with exponential dependence indicating that both Euclidean and fractal dimensions $\rightarrow \infty$.

other, as function of ℓ , in logarithmic scale. A linear dependence over a significative ℓ interval, indicative of a fractal dimension, can be assigned only for some networks. Our results go along the lines of [18], indicating that AN, BA and CT do not show scaling behavior.

So, we see in Figure 7a that, for WS-N networks, well defined scaling region can be found for very small values of p_r , when we have a large value of D . Despite the fact that WS-N networks show small world features, for this small value of p and of N , the network does not show yet a $\langle d \rangle \sim \log N$ dependence, what would prevent the network from displaying a fractal dimension. On the other hand, for larger values of p_r , the figure shows that it is no longer possible to assign a value for d_F . The same behavior is found for AN and BA networks, where the points do not follow a linear behavior, for any value of nodes in

the network. For ER networks, this is also the typical behavior, unless p_a assumes the critical value, where a giant percolating cluster containing the large majority of nodes emerge. In this case, a fairly well defined scaling region between $L(\ell)$ and ℓ over a significant ℓ interval is noted with $d_F = 1.96$, confirming the results in [18].

In Figure 7b, DHL and WHL also show scaling behavior, although the decrease of $L(\ell)$ proceeds through a series of large size steps. The slope of both curves are similar (~ 1.82), but differ from the values obtained for the usual fractal dimension [28]. We observe, in the inset of Figure 7a, that the CT definitely follows an exponential rather than power law decay. This corresponds to an infinite value for d_F , what is in agreement with the fact that it is not possible to associate a finite dimension to the tree, where the number of sites in the bulk and surface have the same order of magnitude. These results confirm that the definition of d_F is sound, capturing the essential features of quite distinct networks but, as in the case of DHL, values for d_F may differ from those obtained from other definitions.

4 Conclusions

In this work we put forward, in a systematic way, the investigation of the higher order neighborhoods of a complex network R . The basic concepts, that have been introduced in a previous work, are developed in many aspects, opening new paths for the exploration of properties related to structures in a larger scale than that of the immediate vicinity of a node. According to the basic idea, each neighborhood $O(\ell)$, represented by an adjacency matrix $M(\ell)$, is regarded as a network in itself, so that many distinct quantifiers used for network characterization can be used to obtain the properties of each neighborhood. On the other hand, the analysis of each of these networks uncovers large distance properties of the original network.

We were able to reproduce known results for the network indices when $\ell = 1$ and, at the same time, present some very peculiar features for the same parameters when evaluated at larger values of ℓ .

In the discussion of the ℓ dependent quantifiers, we find that many of them behave in an oscillatory way with respect to ℓ . This is the case of $C(\ell)$ for self-similar networks, as CT, DHL, WHL, as well as for hypercubic lattices. For ER and WS-N networks, oscillatory behavior in $C(\ell)$ can also be found, provided linking and rewiring probabilities are small enough. Interesting enough, the values of $\langle k(\ell) \rangle$ oscillate for both DHL and WHL, for all ℓ range. For all other networks, the average node degree $\langle k(\ell) \rangle$ increases exponentially for several networks, but this behavior can be masked, for networks with very small values of D . Hierarchical property and scale-free distribution of nodes are rare events for large values of ℓ . Nevertheless, we present examples of BA, AN and WHL, where corresponding power law distributions for $p(k; \ell)$ and $C(k; \ell)$ are found for $\ell > 1$. Properties related to degree assortativity or disassortativity are likely to remain the same as ℓ increases. One exception, again due to the emergence of oscillatory behavior, refers to the self similar DHL and

WHL. Finally, we use the information obtained from the set $M(\ell)$ to evaluate the fractal dimension of all investigated networks. The obtained results reproduce well into the expected values, among which we quote $d_F \rightarrow \infty$ for CT, smaller values of d_F for less connected networks, and some troubles in finding precise values when the network diameter is very small.

Our results for $C(\ell)$ and $\langle k(\ell) \rangle$ can be compared to reports for similar measures discussed recently in the literature [9,16,17], although the authors do not explore all range $\ell = 1, 2, \dots, D$ nor all network types as we do. Regarding $C(\ell)$, our measure takes into account two neighbors of order ℓ of node r that are also linked to each other by ℓ steps, while the previous definition requires that they are $\ell = 1$ neighbors. On the other hand, $\langle k(\ell) \rangle$ considers only one single connection between the node r and each of its ℓ -th order neighbors, while the hierarchical degree considers all links among nodes of the ℓ th and $\ell + 1$ -th.

We emphasize that these prior results could also be used to evaluate distributions that are correspondent to $p(k; \ell)$ and $C(k; \ell)$ introduced here. However, these issues have not been considered by the authors. Our results indicate that power law dependence for $p(k; \ell)$ can indeed be found for BA networks, when $\ell \simeq 1$, whereas scaling behavior in $C(k; \ell)$ has been found for WHL. Finally, as far as we know, similar analyzes for $A(\ell)$ have not been reported before. The main result of that subsection is that, with exception of DHL and WHL, the same assortative property is preserved for higher order neighborhoods.

As a concluding remark, we would like to reinforce that this series of investigations we opened in this work is very broad. It can be explored in many directions, certainly including the exploration of data from actual networks. We think that the neighborhood identification approach may be helpful also in the evaluation of other suggested measures that can be found in the available literature.

This work was partially supported by CNPq and FAPESB.

References

1. J.L. Gross, J. Yellen, *Handbook of Graph Theory*, Discrete Mathematics and its Applications (CRC Press, Boca Raton, 2000)
2. S.N. Dorogovtsev, J.F.F. Mendes, *Evolution of Networks: From Biological Nets to the Internet and WWW* (Oxford Univ. Press, 2003)
3. M.E.J. Newman, A.-L. Barabási, D.J. Watts, *The Structure and Dynamics of Networks* (Princeton University Press, 2006)
4. R. Albert, A.-L. Barabási, *Rev. Mod. Phys.* **74**, 47 (2002)
5. M.E.J. Newman, *SIAM Rev.* **45**, 167 (2003)
6. S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, D.-U. Hwang, *Phys. Rep.* **424**, 175 (2006)
7. L. da F. Costa, F.A. Rodrigues, G. Traverso, P.R. Villas Boas, *Adv. Phys.* **56**, 167 (2007)
8. R.F.S. Andrade, J.G.V. Miranda, T.P. Lobão, *Phys. Rev. E* **73**, 046101 (2006)
9. M.E. Newman, *Social Networks* **27**, 83 (2003)

10. A. Fronczak, J.A. Holyst, M. Jedynek, J. Sienkiewicz, *Physica A* **316**, 688 (2002)
11. R. Xulvi-Brunet, J.M. Sokolov, *Acta Physica Polonica B* **36**, 1431 (2005)
12. G. Caldarelli, R. Pastor-Santorras, A. Vespignani, *Eur. Phys. J. B* **38**, 183 (2004)
13. R. Cohen, S. Havlin, S. Mokryn, D. Dolev, T. Kalisky, Y. Shavitt (2003), e-print [arXiv:cond-mat/0305582](https://arxiv.org/abs/cond-mat/0305582)
14. L. da F. Costa, *Phys. Rev. E* **70**, 056106 (2004)
15. L. da F. Costa, L.H.C. da Silva, *Eur. Phys. J. B* **50**, 237 (2006)
16. L. da F. Costa, *Phys. Rev. Lett.* **93**, 098702 (2004)
17. L. da F. Costa, F.N. Silva, *J. Stat. Phys.* **125**, 841 (2006)
18. C. Song, S. Havlin, H.A. Makse, *Nature* **433**, 392 (2005)
19. P. Erdős, A. Rényi, *Publ. Math. (Debrecen)*, **6**, 290 (1959)
20. A.-L. Barabási, R. Albert, *Science* **286**, 509 (1999)
21. D.J. Watts, S.H. Strogatz, *Nature* **393**, 440 (1998)
22. J.S. Andrade Jr, H.J. Herrmann, R.F.S. Andrade, L.R. daSilva, *Phys. Rev. Lett.* **94**, 018702 (2005)
23. M. Kaufman, R.B. Griffiths, *Phys. Rev. B* **26**, 5282 (1982)
24. J.E. Whitesitt, *Boolean Algebra and its Applications* (Dover, New York, 1995)
25. F. Harary, *Graph Theory* (Perseus Books Publishing, Cambridge 1995)
26. E. Ravasz, A-L Barabási, *Phys. Rev. R* **67**, 026112 (2003)
27. M.E.J. Newman, *Phys. Rev. Lett.* **89**, 208701 (2002)
28. J.R. Melrose, *J. Phys. A: Math. Gen.* **16**, 3077 (1983)