Eur. Phys. J. B **50**, 237–242 (2006)
DOI: 10.1140/epjb/e2006-00107-0

THE EUROPEAN PHYSICAL JOURNAL B

A generalized approach to complex networks

L. da F. Costa[a] and L.E.C. da Rocha[b]

Institute of Physics of São Carlos, University of São Paulo, Caixa Postal: 369, CEP: 13560-970, São Carlos SP, Brazil

Received 29 September 2005 / Received in final form 23 November 2005
Published online 12 April 2006 –

Abstract. This work describes how the formalization of complex network concepts in terms of discrete mathematics, especially mathematical morphology, allows a series of generalizations and important results ranging from new measurements of the network topology to new network growth models. First, the concepts of node degree and clustering coefficient are extended in order to characterize not only specific nodes, but any generic subnetwork. Second, the consideration of distance transform and rings are used to further extend those concepts in order to obtain a signature, instead of a single scalar measurement, ranging from the single node to whole graph scales. The enhanced discriminative potential of such extended measurements is illustrated with respect to the identification of correspondence between nodes in two complex networks, namely a protein-protein interaction network and a perturbed version of it.

PACS. 02.70.Rr General statistical methods – 02.10.Ox Combinatorics; graph theory – 89.75.Hc Networks and genealogical trees

1 Introduction

One of the unavoidable consequences of the fast pace of developments in the new area of complex networks [1–4] is that, while many impressive and relevant concepts and perspectives have been identified and well-developed, some interesting issues have received relatively little attention. Despite the major advances achieved by using powerful tools from theoretical physics (e.g. [1,4]), relatively little attention has been given to the treatment of complex networks in terms of discrete mathematics and mathematical morphology, which are themselves well-established investigation fields. Developed mainly by J. Serra and collaborators [5], the area of mathematical morphology is aimed, through strict mathematical formalization, at representing and analyzing the geometrical and topological features of discrete mathematical structures, especially regular lattices such as those underlying digital images. Mathematical morphology is strongly founded on the discrete operations of complement, dilation and erosion, which can be composed in order to obtain a whole series of new operators with specific properties. At the same time, previous developments by L. Vincent and H. Heijmans [6,7] have shown how the mathematical morphology framework can be extended to graphs, allowing not only the precise mathematical representation and manipulation of those general structures, but also the immediate access to the wealthy of existing results from mathematical morphology. The present article reports on how the application of discrete mathematics, especially mathematical morphology [5] and distance-oriented concepts [6–8], bears the potential not only for formalization, but also to obtain a series of new concepts and results. In particular, by considering the dilations of subnetworks of a network Γ and extending the concepts of numbers of neighbors [9–12] and hierarchical node degree [13], we show that the traditional concepts of node degree and clustering coefficient [1,4] can be generalized in two important ways. First, the concept of subnetwork dilation paves the way to generalize the degree and clustering coefficient to any subnetwork of Γ, and not only their specific nodes as adopted in the complex network literature. Such a concept therefore allows us to speak of the degree of subgraphs of special interest, such as cycles, sets of hubs, or the maximum spanning tree of a given complex network. Second, the consideration of a series of subsequent dilations, together with the respectively induced distance transform and rings, allow the further extension of the degree and clustering coefficient so that a signature, instead of the single scalar traditional measurements, is obtained which can provide information about the network connectivity from the node to the whole graph scales. The potential of such hierarchical extensions for discriminating the connectivity around each node (or subgraph) can be readily appreciated by considering the fact that several nodes in a complex network will have the same degree and clustering coefficient, but very few nodes will share such values calculated for a series of subsequent neighborhoods. Such an interesting feature of the generalized measurements is illustrated in the present article with respect to protein-protein interaction networks. A perturbed version of this

[a] e-mail: luciano@if.sc.usp.br
[b] e-mail: rocha@if.sc.usp.br

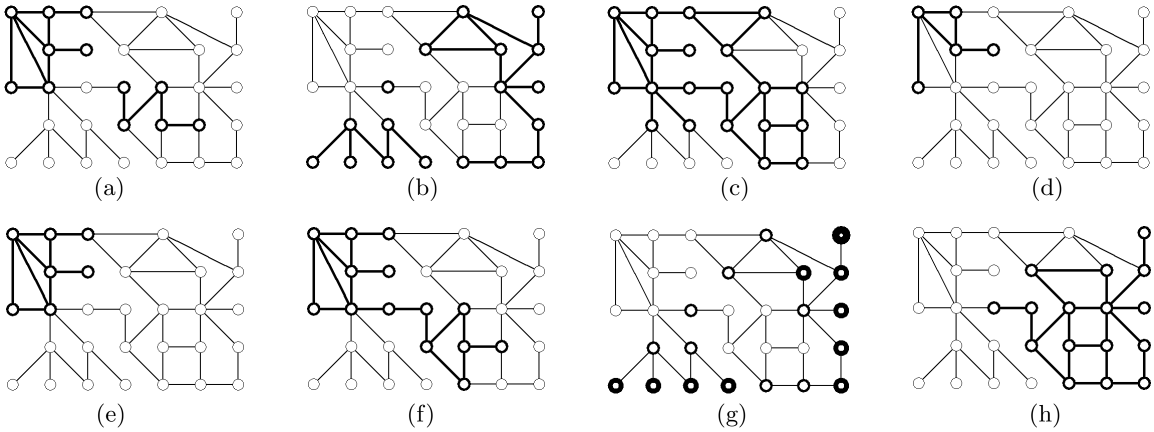

Fig. 1. Original network containing a subnetwork (a), identified by wider-border nodes, and its respective complement (b), dilation (c), erosion (d), opening (e), closing (f), and distance transform (g), with the distances identified by the node border width. The generalized Voronoi tessellation of the two connected components of the subnetwork in (a) is shown in (h).

network is obtained by a random rewiring and the hierarchical measurements are compared so that the most similar nodes are identified.

2 Basic concepts

A *network* Γ without multiple edges is a discrete structure composed of a set of *nodes* $V(\Gamma)$ and a set $E(\Gamma)$ of *edges* (u,v) established between specific pairs of nodes of $V(\Gamma)$, so that the network Γ is represented as $\Gamma=(V,E)$. As we consider undirected networks without loops, it follows that $(u,v) \iff (v,u)$ and $(u,u) \notin E(\Gamma)$. Such a network can be conveniently represented in terms of its respective *adjacency matrix* K such that each edge (u,v) is represented by making $K(u,v)=K(v,u)=1$, while the absence of edge is indicated by zero value. A *subnetwork* ξ of Γ is any network such that $V(\xi)\subseteq V(\Gamma)$ and $E(\xi)\subseteq\{(u,v)|(u,v)\in E(\Gamma) \, \text{and} \, u,v\in V(\xi)\}$. Figure 1a illustrates a network Γ and one of its many subnetworks ξ, identified by the wider-border nodes and wider edges. Particularly interesting subnetworks of a network Γ include its hubs, outmost nodes (i.e. nodes with low degree), as well as its cycles. Special cases of subnetworks of Γ include the empty network $(V=\emptyset, E=\emptyset)$, where $\emptyset$ stands for the empty set, networks containing an isolated node u $\Gamma_u=(V=\{u\in V(\Gamma)\}, E=\emptyset)$, and the own original network Γ. The *complement* of a subnetwork ξ of Γ is the subnetwork ξ'_Γ of Γ such that $V(\xi'_\Gamma)=\{u|u\in V(\Gamma) \, and \, u\notin V(\xi)\}$ and $E(\xi'_\Gamma)=\{(u,v)|(u,v)\in E(\Gamma) \, and \, u,v\in V(\xi'_\Gamma)\}$. Figure 1b illustrates the complement ξ'_Γ of ξ in Γ. A subnetwork is *connected* if any of its nodes can be reached from any of its other nodes. Two subnetworks ζ and ξ of Γ are connected if it is possible to reach a node of ξ from a node of ζ, and vice-versa. The maximal connected subnetworks, in the sense of including the largest number of nodes, of a network are called *connected components*. The subnetwork in Figure 1a is not connected but contains two connected components. The *degree* of a node u of Γ, hence $k(u)$, corresponds to the number of edges attached to that node. The *degree* of a subnetwork ξ of Γ, hence $k(\xi)$, is defined as the number of edges implied by the dilation of ξ, i.e. those edges connecting ξ to the rest of Γ. For instance, the degree of the subnetwork ξ in Figure 1a is 12. The *outmost set* of a subnetwork ξ of Γ is the set of nodes $\Omega(\xi)$ which have unit degree. For simplicity's sake, such nodes are henceforth referred to as *outnodes*. The *1-neighborhood* of a node u of Γ, henceforth represented as $n_1(u)$, is the set of nodes of Γ which are attached to u, plus node u. This concept can be immediately extended to express the *neighborhood of a subnetwork* ξ of Γ, given as the set of nodes of Γ which are connected to ξ plus the nodes in $V(\xi)$.

3 Complex network morphology

The *dilation* of a subnetwork ξ of Γ is defined as the subnetwork $\delta(\xi)$ of Γ having $V(\delta(\xi))=n_1(\xi)$ as its set of nodes while its set of edges include the edges of Γ found between the nodes in $n_1(\xi)$. The *erosion* of ξ, represented as $\varepsilon(\xi)$, is a subnetwork of Γ which can be defined as the complement of the dilation of ξ'_Γ, i.e. $\varepsilon(\xi)=(\delta(\xi'_\Gamma))'_\Gamma$. Observe that the dilation or erosion of Γ yields Γ as result. Figures 1c and 1d ilustrates the dilation and erosion of the subnetwork ξ in a, respectively. Observe that the erosion eliminated one of the connected components of ξ. Generally, $\delta(\varepsilon(\xi))\neq\varepsilon(\delta(\xi))$, i.e. the dilation can not be used to undo an erosion, and vice-versa. Observe also that a subnetwork ξ is necessarily contained or equal to its respective dilation, while the erosion of a subnetwork ξ is necessarily contained or equal to itself. The *d-dilation* of a subnetwork ξ is defined as the subnetwork obtained by dilating d times the subnetwork ξ, i.e.:

$$\delta_d(\xi)=\underbrace{\delta(\delta(\ldots(\xi)\ldots))}_{d}. \tag{1}$$

Similarly, the *d-erosion* can be defined as:

$$\varepsilon_d(\xi)=\underbrace{\varepsilon(\varepsilon(\ldots(\xi)\ldots))}_{d}. \tag{2}$$

Observe that $\delta_d(\xi)$ converges to Γ as d is increased, while $\varepsilon_d(\xi)$ converges to the empty network under similar circumstances. We also have that $\delta_{d=i+j}(\xi) = \delta_i(\delta_j(\xi)) = \delta_j(\delta_i(\xi))$ and $\varepsilon_{d=i+j}(\xi) = \varepsilon_i(\varepsilon_j(\xi)) = \varepsilon_j(\varepsilon_i(\xi))$. The *d-degree* of a subnetwork ξ is defined as the degree of the d-dilation of the network ξ, i.e., $k_d(\xi) = k(\delta_d(\xi))$. It is possible to use combinations of dilations and erosions of a subnetwork ξ in order to obtain new operators such as the *opening* and *closing* of ξ, which are defined as $\alpha(\xi) = \delta(\varepsilon(\xi))$ and $\omega(\xi) = \varepsilon(\delta(\xi))$, respectively. Figures 1e and 1g illustrate, respectively, the opening and closing of the subnetwork ξ in a. Observe that the closing of ξ had as an effect the connection of the two components of that subnetwork, filling the gap between those subnetworks. The opening and closing operations are idempotent, in the sense that $\alpha(\alpha(\xi)) = \alpha(\xi)$ and $\omega(\omega(\xi)) = \omega(\xi)$. It is also interesting to define the *d-opening* of a subnetwork ξ, henceforth represented as $\alpha_d(\xi)$, corresponding to d erosions followed by d dilations. The *d-closing* of ξ, represented as $\omega_d(\xi)$ can be defined in a similar fashion. The latter operator is useful for investigating the progressive merging of subnetworks of Γ in terms of increasing values of d. Particularly, interesting information about the network structure can be provided by the evolution of the number of connected subnetworks, starting from a specific set χ of subnetworks (e.g. the network 3-cycles), in terms of a sequence of d-openings (or closings) performed for increasing values of d.

4 Distances, distance transforms, parallels and rings

Several important features of a complex network are related to the concept of *distance*. If ζ and ξ are any subnetworks of Γ, the (minimal) distance between the respective set of nodes $V(\zeta)$ and $V(\xi)$, hence $D(V(\zeta), V(\xi))$, can be defined as the value of d for which some node u of ζ becomes included into $V(\delta_d(\xi))$. It can be verified that $D(V(\zeta), V(\xi)) = D(V(\xi), V(\zeta))$. Observe that $D(V(\zeta), V(\zeta)) = 0$. In particular, the distance between a node u and a subnetwork ξ is given as $D(\{u\}, V(\xi))$. The *distance transform* of a subset of nodes χ of Γ is the mapping which assigns $D(\{u\}, \chi)$ to every node $u \in V(\Gamma)$, including those in ξ. Figure 1g illustrate the distance transform of the subnetwork in a, with the distance values (4 values for this subnetwork, i.e. $d = 0, 1, 2,$ and 3) expressed in terms of the nodes border widths. Given a subnetwork ξ of Γ, the subnetwork ϱ defined by the set $V(p)$ of nodes such that $D(V(\varrho), V(\xi)) = d$ and the set of those edges of Γ connecting nodes in $V(\varrho)$ is called the *d-parallel* of ξ, henceforth represented as $P_d(\xi)$. The parallels of the subnetwork ξ in Figure 1a correspond to the set of nodes with the same width in g plus the respective interconnecting edges. The number of nodes and edges in a d-parallel of ξ are henceforth represented as $n\{P_d(\xi)\}$ and $e\{P_d(\xi)\}$. Similarly, it is interesting to define the *rs-ring* of ξ, hence $R_{rs}(\xi)$, which corresponds to the union of the respective parallels of ξ for distances $d = r$ to s plus the edges of Γ interconnecting such parallels. The number of nodes and edges in a *rs*-ring of g are henceforth represented as $n\{R_{rs}(\xi)\}$ and $e\{R_{rs}(\xi)\}$. Observe that a d-parallel therefore is the particular case of the *rs*-ring for $d = r = s$. Another interesting possibility is to use the above introduced distance concepts in order to obtain the generalized Voronoi tessellation of subnetworks, as illustrated in Figure 1h with respect to the two connected components in the subnetwork in Figure 1a. The above definitions allow the concept of *clustering coefficient* [1,4] to be generalized to parallels and rings of any subnetwork. The *rs-clustering coefficient* of a subnetwork ξ of Γ, henceforth represented as $cc_{rs}(\xi)$, can be defined as the number of edges in the respective *rs*-ring subnetwork, divided by the total of possible edges between the nodes in that ring, i.e.:

$$cc_{rs}(\xi) = \frac{2e\{R_{rs}(\xi)\}}{n\{R_{rs}(\xi)\}(n\{R_{rs}(\xi)\} - 1)}. \tag{3}$$

5 Hierarchical measurements for single nodes

The concepts discussed above can be naturally extended to a single node and to an edge, respectively, whether the subgraph contains the node alone and whether the subgraph contains the edge and both nodes connected by that edge [14,15]. Using the concept of rings considered in the last section, henceforth the subgraph ζ composed of the ring $R_d(u)$ is defined as the *hierarchical level* related to the subgraph ξ composed of the single node u, such that, the *hierarchical number of nodes* $n_d(u)$ (or $n\{R_d(u)\}$) is given as the number of nodes at *hierarchical distance* d from de reference node u, i.e., the number of nodes in the ring $R_d(u)$. Hence, the *hierarchical degree* $k_d(u)$ is defined as the number of edges between the nodes in the ring R_d and R_{d+1}, such that the *hierarchical number of edges* among the nodes in the ring $R_d(u)$ is $e_d(u)$ (or $e\{R_d(u)\}$). The *hierarchical clustering coefficient* $cc_d(u)$ is written using equation 3.

$$cc_d(u) = \frac{2e_d(u)}{n_d(u)(n_d(u) - 1)}. \tag{4}$$

At last, other hierarchical measurements can be derived from the above definitions and each of them will be dealt with in the following:

- *Convergence ratio* ($C_d(u)$): measures the ratio between the hierarchical node degree of node u at hierarchical distance $d-1$ and the hierarchical number of nodes in the ring $R_d(u)$.

$$C_d(u) = \frac{k_{d-1}(u)}{n_d(u)}. \tag{5}$$

- *Intra-ring degree* ($A_d(u)$): the average among the degrees of the nodes in the ring $R_d(u)$.

$$A_d(u) = \frac{2e_d(u)}{n_d(u)}. \tag{6}$$

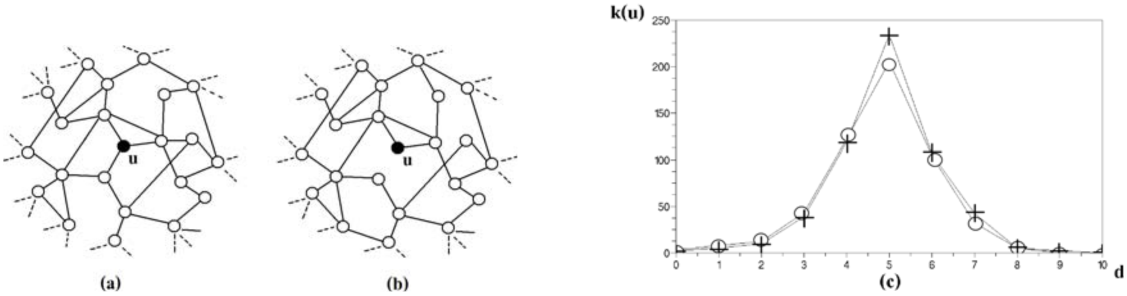

Fig. 2. (a) non-perturbed and (b) perturbed portion of a network; (c) hierarchical degrees $k_d(u)$ of the reference node in black (○ = non-perturbed case; + = perturbed version).

Table 1. Summary of the hierarchical measurements considered in this work.

Hier. number of nodes in the ring $R_d(u)$	$\mathbf{n_d(u)}$
Hier. number of edges among the nodes in the ring $R_d(u)$	$\mathbf{e_d(u)}$
Hier. degree of node u at distance d	$\mathbf{k_d(u)}$
Hier. clust. coeff. of u at hier. level d	$\mathbf{cc_d(u)}$
Convergence ratio of u at hier. level d	$\mathbf{C_d(u)}$
Intra-ring node degree of u at distance d	$\mathbf{A_d(u)}$
Inter-ring node degree of u at distance d	$\mathbf{E_d(u)}$
Hier. common degree of node u at d	$\mathbf{H_d(u)}$

– *Inter-ring degree* ($E_d(u)$): the average of the number of connections between each node in ring $R_d(u)$ and those in $R_{d+1}(u)$.

$$E_d(u) = \frac{k_d(u)}{n_d(u)}. \tag{7}$$

– *Hierarchical common degree* ($H_d(u)$): the average node degree among the nodes in $R_d(u)$, considering all edges connected to nodes in the ring.

$$H_d(u) = \frac{2e_d(u) + k_{d-1}(u) + k_d(u)}{n_d(u)}. \tag{8}$$

A summary of these hierarchical measurements are presented in Table 1.

6 Evaluation of discriminative power

Given a complex network, it is possible to organize several selected measurements of its topology into a *feature vector* $\vec{\mu}$ (e.g. [15]), which therefore provides a quantitative description of properties of the network. Multivariate statistical methods (e.g. [8]) can then be applied in order to separate such vectors into clusters or to identify the category of the network. In a similar fashion, it is possible to assign a feature vector to individual nodes of the network, so that they can be characterized and organized into classes. Although simple measurements such as the node degree and clustering coefficients can be used for this purpose, they are generally not enough for a discriminative characterization of nodes at the individual level because several nodes in a large network will have identical values of such measurements. The hierarchical extensions of the node degree and clustering coefficient, combined with the ancillary hierarchical measurements described in this work, account for substantially enhanced discrimination of the local properties of the connectivity around each node, therefore diminishing the degeneracy of the description. In other words, several nodes may have the same immediate node degree, but it is rather unlikely that they will also share other hierarchical degrees. At the same time, nodes which do present similar connectivity patterns along the hierarchies can be clustered into meaningful classes by considering feature vectors composed of hierarchical measurements. In order to illustrate the above possibilities, we considered a *S. cerevisiae* protein-protein interaction network Γ [16] containing $N = 1922$ nodes and without self-connections and isolated nodes. A perturbed version of this network was obtained by randomly rewiring the edges with probability p. More specifically, each link is removed with probability p, in which case a new link is established between a pair of randomly selected nodes. Nodes in these two networks are then characterized in terms of several combinations of the hierarchical measurements discussed in this work. Figure 2 illustrates a node and its most immediate neighborhood before a and after b perturbation of connectivity, as well as the respective hierarchical node degree signatures c. In cases such as that illustrated in this figure, the hierarchical signature provides a more stable characterization of the neighboring connectivity of node than would be obtained by using traditional measurements such as the node degree and clustering coefficient.

In order to quantify the discriminative power of such measurements, we repeatedly selected a node u from the original network and identified among all nodes v of the perturbed network the node which leads to the smallest Euclidean distance between the respective measurements, i.e., the distance between respective feature vectors, $\|\vec{\mu_u} - \vec{\mu_v}\|$, is minimal. In case these two nodes are verified to indeed correspond one another (recall that the identity of the nodes is guaranteed because the perturbed network is derived from the original network by rewiring), we understand there has been a correct identification. Among the several combinations of measurements, considering varying hierarchical levels, the best results were

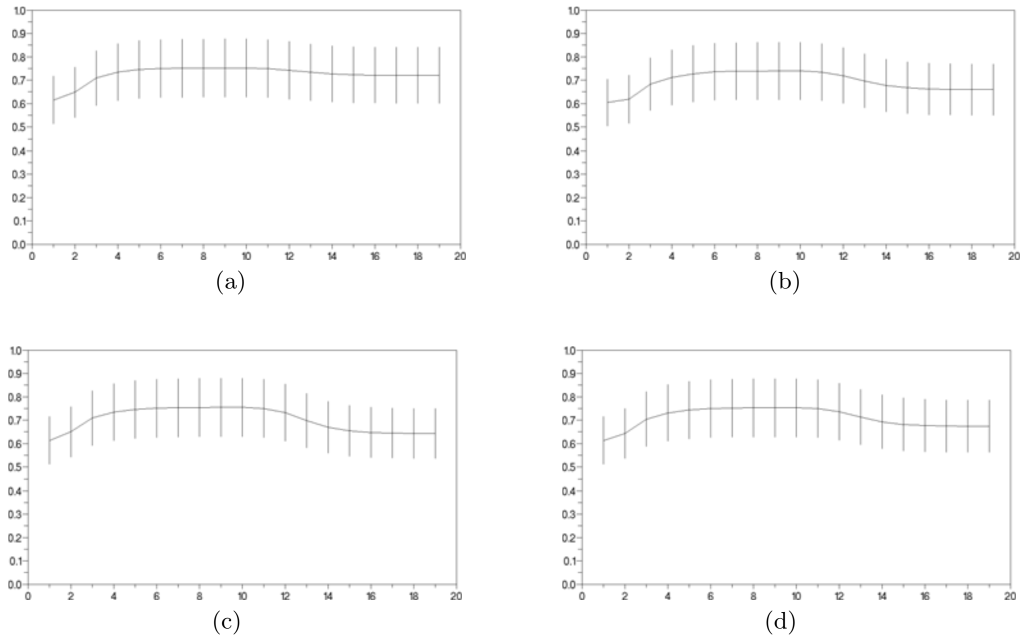

Fig. 3. Four hierarchical measurements combinations for the feature vector: (a) A_d and E_d; (b) E_d and H_d; (c) C_d, A_d and H_d and (d) A_d and H_d.

obtained for pairwise combinations of A_d, E_d, H_d and C_d, particularly the four situations shown in Figure 3. The diagrams in this figure depict the average ± standard deviation of the percentage of correctly identified nodes by using the identified pairs of measurements up to the hierarchical levels identified in the x-axis (varying from 1 to 19). A number of interesting features can be identified from such results. First, it is interesting to notice that the average of correct identifications undergoes the three following regimes: (i) increases along the 4 or 5 initial hierarchies; (ii) stays nearly constant until about 12 hierarchical levels; and (iii) then decreases steadily. This behavior is observed for all graphs in Figure 3. The average performance increase in (i) is a direct consequence of the fact that more information about the network connectivity around each node is taken into account. The performance plateau and decrease taking place after 4 or 5 hierarchical levels are considered for the measurements reflects a degeneration in the discriminative power of the measurements caused by the fact that most of the nodes have been considered at such hierarchical depths. Similar performances are observed for the four cases illustrated in Figure 3, with slightly better results being achieved for the measurement combinations in a and d. The standard deviations values tend to follow the mean, with higher variations observed along the plateaux. It is interesting to note that it would also be possible to consider traditional clustering algorithms [8] in order to organize the nodes into clusters sharing hierarchical properties.

7 Concluding remarks

This article has addressed several issues regarding the generalization of complex networks measurements. First, we have shown for the first time that complex networks and their properties can be formalized in terms of mathematical morphology, allowing the definition of a series of measurements such as the generalized versions of the node degree and clustering coefficient, as well as the possibility to use other features from mathematical morphology so as to investigate further the structure of specific subnetworks. Second, we have emphasized the importance of identifying and studying the properties of subnetworks of special interest — including the set of hubs, out-nodes and 3-cycles, and shown that a particularly comprehensive study of such subnetworks can be obtained by taking into account a whole series of neighborhoods, as allowed by the novel proposed concepts of generalized degrees and clustering coefficient. While the new set of measurements extended to take into account subgraphs and hierarchies can be used to derive new network growth schemes and characterize and classify different types of networks, they also present power for enhanced discrimination between individual nodes. The latter has been illustrated for the first time in this article with respect to protein-protein interaction networks. More specifically, we have shown that the ability to identify correspondence between nodes in two versions of a network (in the case of our example the original and perturbed networks) tend to increase by considering measurements taking into account multiple hierarchical levels. This is a consequence of the fact that the use of more hierarchical levels allows the measurements to reflect in a less degenerate way the network connectivity around each node. At the same time, we have shown that such enhanced discriminative power tends, with the incorporation of additional hierarchical depths, to reach a plateau and then to decrease. The possibilities for future works include the application of the introduced concepts to community finding, characterization of resilience to attack, and extensions to measurements aimed at characterizing the assortative properties of networks.

Luciano da F. Costa is grateful to FAPESP (process 99/12765-2), CNPq (308231/03-1), NIH Fogarty, and the Human Frontier Science Program (HFSP RGP39/2002) for financial support. Luis Enrique Correa da Rocha is grateful to Human Frontier Science Program for his undergraduate grant.

References

1. R. Albert, A.L. Barabási, Rev. Mod. Phys. **74**, 47 (2002)
2. S.N. Dorogovtsev, J.F.F. Mendes, Adv. Phys. **51**, 1079 (2002)
3. S.N. Dorogovtsev, J.F.F. Mendes, *Evolution of Networks, From Biological Nets to Internet and WWW* (Oxford University Press, 2003)
4. M.E.J. Newman, SIAM Rev. **45**, 167 (2003)
5. J. Serra, *Image Analysis and Mathematical Morphology* (Academic Press, 1984)
6. L. Vincent, Signal Proc. **16**, 365 (1989)
7. H.J.A.M. Heijmans, P. Nacken, A. Toet, L. Vincent, J. Vis. Comm. Image Repr. **3**, 24 (1992)
8. L. da F. Costa, R.M.C. Jr, *Shape Analysis and Classification: Theory and Practice* (CRC Press, Boca Raton, 2001)
9. M. Faloutsos, P. Faloutsos, C. Faloutsos, SIAM Rev. **45**, 167 (1999)
10. C. Faloutsos, P. Faloutsos, M. Faloutsos, Comp. Comm. Rev. **29**, 251 (1999)
11. M.E.J. Newman, e-print `arXiv:cond-mat/0111070` (2001)
12. R. Cohen, S. Havlin, S. Mokryn, D. Dolev, T. Kalisky, Y. Shavitt, e-print `arXiv:cond-mat/0305582` (2003)
13. L. da F. Costa, Phys. Rev. Lett. **93**, 098702 (2004)
14. L. da F. Costa, e-print `arXiv:cond-mat/0412761` (2005)
15. L. da F. Costa, F. Rodrigues, G. Travieso, P.R.V. Boas, e-print `arXiv:cond-mat/0505185` (2005)
16. E. Sprinzak, H. Margalit, J. Mol. Biol. **311**, 681 (2001)