Generalized connectivity between any two nodes in a complex network

Francisco Aparecido Rodrigues[*]
Departamento de Matemática Aplicada e Estatística, Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo-Campus de São Carlos, Caixa Postal 668, 13560-970, São Carlos, SP, Brazil

Luciano da Fontoura Costa[†]
Instituto de Física de São Carlos, Universidade de São Paulo, Av. Trabalhador São Carlense 400, Caixa Postal 369, CEP 13560–970 São Carlos, São Paulo, Brazil
and National Institute of Science and Technology for Complex Systems, Rua Dr. Xavier Sigaud-150 Urca, Rio de Janeiro, Brazil

(Received 6 March 2009; revised manuscript received 2 November 2009; published 22 March 2010)

This article focuses on the identification of the number of paths with different lengths between pairs of nodes in complex networks and how these paths can be used for characterization of topological properties of theoretical and real-world complex networks. This analysis revealed that the number of paths can provide a better discrimination of network models than traditional network measurements. In addition, the analysis of real-world networks suggests that the long-range connectivity tends to be limited in these networks and may be strongly related to network growth and organization.

DOI: 10.1103/PhysRevE.81.036113 PACS number(s): 89.75.Hc, 79.60.Ht, 45.70.Vn, 89.75.Fb

I. INTRODUCTION

A large number of natural and artificial complex systems can be represented and modeled in terms of networks involving interacting components. Such interactions can range from signaling between cells (e.g., [1]) to social contacts (e.g., [2]). Indeed, complex networks theory has been considered in a wide range of investigations including neuronal connections, protein-protein interactions, economy, and internet communication [3], to cite just a few possibilities.

The characterization of network structure is one of the fundamental aspects of complex networks research because the modeling, simulation and classification of networks all depend strongly on accurate descriptions of the respective topology [4,5]. In order to quantify the properties of the diverse types of networks, a large set of network measurements has been developed [4]. Many of these features are related to the immediate links between each pair of nodes. Indeed, several of the measurements currently employed in order to characterize network structure—such as degree, clustering coefficient and shortest path length—are ultimately related to short-range pairwise interconnectivity [4]. However, it is also important to resort to longer range interaction between nodes in order to achieve more comprehensive description, characterization, and modeling of complex structures.

Long range interactions can be defined in terms of paths and walks. Walks are given by a sequence of nodes and edges. Paths, on the other hand, are a special type of walks described by sequence of nodes and edges without repetition. The use of walks and paths of different lengths for network characterization is not new and has been preliminarily explored in other works. For instance, walks of different lengths have been used to characterized different network topologies (e.g., [6]). Middendorf *et al.* [7] proposed a sound statistical analysis to characterize and classify networks according to walks of different lengths. They defined the concept of "words" to establish walks related measurements to quantify different network properties. In addition to walks, paths have been used for network topology description [8] and analysis of criticality and phase transition in grids and networks [9]. The average shortest path length (or geodesic distance) between a pair of nodes is obtained by taking into account the shortest distance between every possible pairs of nodes. Some works have used distance matrices, containing minimum shortest path lengths, in order to enhance the characterization of networks [10–12] and identification of isomorphisms [11,12]. Nevertheless, the isolated consideration of the local connectivity measurements and shortest distance matrix results in incomplete network characterization, since important information about network structure is not taken into account. For instance, the alternative paths between pairs of nodes whose lengths are larger than the shortest path are completely overlooked by more traditional network analysis, which consider just the shortest distances. Thus, two networks presenting the same degree and the same shortest path distributions, but though with different alternative paths organization, can be characterized as being identical by many of the traditional approaches in complex networks research, which is clearly inappropriate. Also, alternative paths can provide additional information about network resilience, once they generally reinforce connections, providing alternative routes, and maximizing the flow. More traditional robustness analysis taking into account just the local connectivity and measurements related to the shortest paths, such as betweenness centrality [13], also do not take into account the richer interconnectivity structure provided by longer alternative paths. Shavitt and Singer [8] analyzed the alternative paths in networks and suggested measurements related to the quality of backup and alternative path centrality. They showed that social structures do not necessarily depend on the most central nodes, as expected, and that nodes with medium centrality measurements are ultimately crucial for efficient routing in the internet.

[*] francisco@icmc.usp.br
[†] luciano@if.sc.usp.br

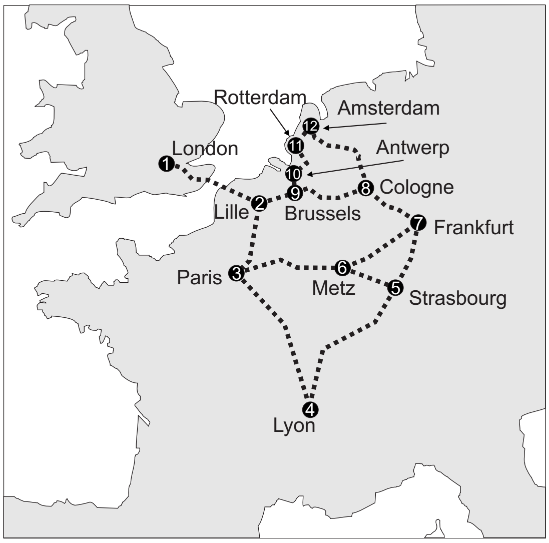

FIG. 1. The European high-speed rail network connecting some of the main cities of northern central Europe. While the traditional shortest distance approach takes into account just the path of length three between London and Lyon, the other seven alternative paths are overlooked. However, the alternative paths are still fundamental for network topology and can be associated to important dynamics such as traffic jamming and resilience [5].

The comprehensive characterization of pairwise connectivity clearly requires more general approaches, such as the consideration of multiscale interactions extending from the immediate link to long-range connectivity scales. The term multiscale refers to the varying topological scales which are progressively taken into account around the nodes. Most of the traditional network measurements consider just the first interconnectivity scale, i.e., immediate neighbor connectivity. In addition to immediate-connection measurements and limited long-range information such as the shortest paths, the identification of alternative paths of any length can enhance the network characterization, providing a more complete description of network topology. Measurements taking into account the successive shortest path lengths from a reference node (concentric neighborhoods) have been proposed in the literature in terms of hierarchical or concentric representations [4,14–17].

The further generalization of the concepts of connectivity and interaction in order to account for larger portions and scales of networks, requires the identification of alternative paths between pairs of nodes, as illustrated in Fig. 1. Let us suppose we are interested in the pairwise interconnection between London (UK) and Lyon (France), which is particularly important for those people wanting to travel by express train between those two cities, as shown in Fig. 1. If we consider the shortest path route, just the path of length three between those two cities is taken into account, while the other seven alternative paths are completely ignored. Nevertheless, such paths are still fundamental for network communication and resilience. For instance, if the connection London-Paris-Lyon is blocked at any part other than from London to Lille, the passengers can always take alternative routes. The importance of the identification of paths in networks has also been substantiated with respect to the functionality of the cardiovascular system [10].

The current work focus on exploring the properties of paths of different lengths in a more systematic way, especially from the point of view of providing valuable information about the structure of complex networks. We report a comprehensive approach to generalize the concept of pairwise connectivity through the quantification of the distribution of paths of different lengths between pairs of nodes. The potential of such a framework is illustrated with respect to network characterization (theoretical models and real-world networks) as well as investigations about network community organization. Helped by optimal multivariate statistical methods, we characterize the relationships between the topologies of six theoretical models and discuss the achieved discriminability. In order to illustrate the variation of the generalized connectivity in real-world networks, we report and discuss results obtained with respect to: (i) the US highway network, (ii) the neural *C. elegans* network [18], (iii) the cat cortical network [19] and (iv) a food web of a broadleaf forest in New Zealand [20]. In addition, we characterize the network modular structure (community) considering respective generalized connectivity matrices. The projection of the network nodes considering an optimal multivariate statistical method resulted in nodes belonging to the same communities being projected nearby, forming clusters of points. In summary, the main contributions of this work are (i) characterization and classification of different complex network structures by considering the number o alternative paths between vertices and optimal multivariate statistical techniques, (ii) analysis of the number of paths distribution in four real-world networks, and (iii) analysis of the relationship between alternative path distributions and community organization.

In next sections, we provide the basic concepts related to network models, paths between nodes, principal component analysis (PCA), canonical variable analysis and network discriminability. An optimal algorithm to find the number of paths between pair of nodes is also provided. The illustration of the potential of the proposed methodology with respect to theoretical and real-world networks are presented and discussed subsequently.

II. BASIC CONCEPTS AND METHODOLOGY

An undirected network can be represented by its adjacency matrix A, whose elements a_{ij} are equal to one whenever there is a connection between the nodes i and j, or equal to zero otherwise. The number of connections of a given node i is called its degree k_i, while the clustering coefficient cc_i, is defined as $cc_i = 2n_i/(k_i - 1)k_i$, where n_i is the number of connections between the neighbors of i [18]. The number of paths with length $h = 1, \ldots, H$ (H is the length of largest considered path) between each pair of nodes can be expressed in terms of the three-dimensional matrix $R = R(h,i,j)$ (see Fig. 2), so that each matrix $R_h(i,j)$, belonging to the set R, gives the total number of paths of length h extending from node i to node j [observe that $R_1(i,j) = A(i,j)$]. These matrices will always be symmetric for undirected networks. The set of matrices R therefore conveys comprehensive information

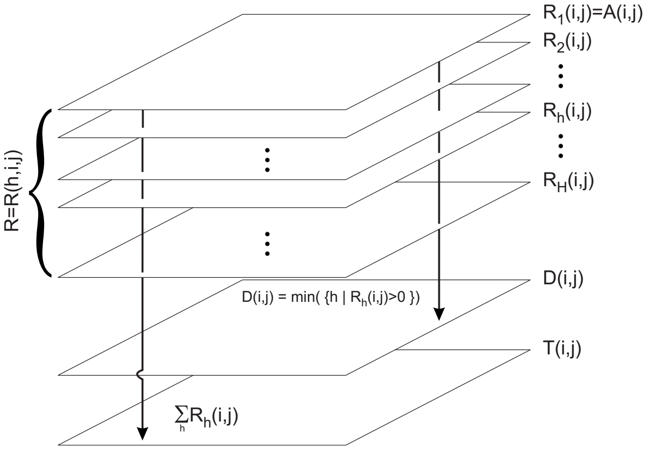

FIG. 2. Networks can be characterized in terms of the three-dimensional matrix $R=R(h,i,j)$, which provides a more comprehensive description of the network structure than the traditional adjacency $[A(i,j)=R_1(i,j)]$ and shortest paths length matrices $\{D(i,j)=min[h|R_h(i,j)\rangle 0]\}$. The matrix T, which provides the total number of paths between every pair of nodes i and j, can be obtained by summing the elements of the matrices $R_h(i,j)$.

about the generalized connectivity between any pair of nodes, providing valuable additional information about the network structure. In addition, the shortest path distance matrix D can be derived from such matrices by taking the value of h respective to the minimum value of the elements along all matrices R obtained for all possible h (see Fig. 2), i.e., $D(i,j)=min[h|R_h(i,j)>0]$. Therefore, the matrices A and D are special cases of the set of matrices R. In this way, the set of matrices $R(h,i,j)$ constitutes a generalization of connectivity matrices previously applied to complex network characterization. The matrix T, which is obtained by summing the elements along the set R, gives the number of paths of lengths $h=1,\ldots,H$ between every pair of nodes. As such, this matrix quantifies all alternative paths between pair of nodes and can be used, for instance, in analysis of network resilience.

An illustration of the several connectivity approaches that can be applied in order to characterize the network in Fig. 1 is provided in Figs. 3 and 4. Figures 3(a) and 3(b) show the traditionally adopted matrices of adjacency and the shortest path length distances, respectively. While the adjacency matrix A indicates the immediate connectivity between pairs of nodes, the shortest path lengths matrix D contains the number of edges along the shortest paths between each pair of nodes. On the other hand, the matrices in Fig. 4 are rarely (if ever) considered in the literature and express other types of pairwise interactions between the nodes. In such a figure, the matrices R_2, R_3, R_4, and R_5 express the number of paths of lengths $h=2$, $h=3$, $h=4$, and $h=5$ between each possible pair of nodes in the network in Fig. 1, respectively. Observe that these matrices make explicit important information which cannot be easily inferred from any of the two previous matrices, A and D. For instance, while matrix D only indicates that the shortest distance between Strasbourg (France) and Antwerp (Belgium) is equal to four—without providing any information about the number of paths with this length—the matrix R_5 shows that there are four paths of length five between those same locations. Similarly, while the matrix D shows that the value of the shortest distance between Lyon (France) and Metz (France) is equal to two, the matrix R_2, indicates the existence of two paths of length two, and the matrix R_3, of one path of length three, all viable alternative connections between these two cities in the case of eventual disruption of the shortest path. The other matrices provide information about even longer alternative paths, of eventual interest for a tourist who wants to visit several nearby places. Therefore, the set of matrices R can provide valuable additional information about the network structure, leading to more accurate network characterization, classification, and modeling.

(a)
$$A = \begin{array}{r|cccccccccccc|} & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 \\ \hline 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 3 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 4 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 5 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 6 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 7 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 8 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 \\ 9 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 10 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 11 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 12 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \end{array}$$

(b)
$$D = \begin{array}{r|cccccccccccc|} & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 \\ \hline 1 & 0 & 1 & 2 & 3 & 4 & 3 & 4 & 3 & 2 & 3 & 4 & 4 \\ 2 & 1 & 0 & 1 & 2 & 3 & 2 & 3 & 2 & 1 & 2 & 3 & 3 \\ 3 & 2 & 1 & 0 & 1 & 2 & 1 & 2 & 3 & 2 & 3 & 4 & 4 \\ 4 & 3 & 2 & 1 & 0 & 1 & 2 & 2 & 3 & 3 & 4 & 5 & 4 \\ 5 & 4 & 3 & 2 & 1 & 0 & 1 & 1 & 2 & 3 & 4 & 4 & 3 \\ 6 & 3 & 2 & 1 & 2 & 1 & 0 & 1 & 2 & 3 & 4 & 4 & 3 \\ 7 & 4 & 3 & 2 & 2 & 1 & 1 & 0 & 1 & 2 & 3 & 3 & 2 \\ 8 & 3 & 2 & 3 & 3 & 2 & 2 & 1 & 0 & 1 & 2 & 2 & 1 \\ 9 & 2 & 1 & 2 & 3 & 3 & 3 & 2 & 1 & 0 & 1 & 2 & 2 \\ 10 & 3 & 2 & 3 & 4 & 4 & 4 & 3 & 2 & 1 & 0 & 1 & 2 \\ 11 & 4 & 3 & 4 & 5 & 4 & 4 & 3 & 2 & 2 & 1 & 0 & 1 \\ 12 & 4 & 3 & 4 & 4 & 3 & 3 & 2 & 1 & 2 & 2 & 1 & 0 \end{array}$$

FIG. 3. The (a) adjacency and (b) distance matrices, respective to the network in Fig. 1.

(a)
$$R_2 = \begin{array}{r|cccccccccccc|} & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 \\ \hline 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 3 & 1 & 0 & 0 & 0 & 2 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 4 & 0 & 1 & 0 & 0 & 0 & 2 & 1 & 0 & 0 & 0 & 0 & 0 \\ 5 & 0 & 0 & 2 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 6 & 0 & 1 & 0 & 2 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 7 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 8 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ 9 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 \\ 10 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 11 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 12 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 \end{array}$$

(b)
$$R_3 = \begin{array}{r|cccccccccccc|} & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 \\ \hline 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 2 & 0 & 0 & 0 & 0 & 2 & 0 & 2 & 0 & 0 & 0 & 1 & 1 \\ 3 & 0 & 0 & 0 & 1 & 1 & 1 & 2 & 2 & 0 & 1 & 0 & 0 \\ 4 & 1 & 0 & 1 & 0 & 1 & 1 & 2 & 1 & 1 & 0 & 0 & 0 \\ 5 & 0 & 2 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 1 \\ 6 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 1 & 2 & 0 & 0 & 1 \\ 7 & 0 & 2 & 2 & 2 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 8 & 1 & 0 & 2 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ 9 & 0 & 0 & 0 & 1 & 1 & 2 & 0 & 0 & 0 & 0 & 1 & 1 \\ 10 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 \\ 11 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 12 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \end{array}$$

(c)
$$R_4 = \begin{array}{r|cccccccccccc|} & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 \\ \hline 1 & 0 & 0 & 0 & 0 & 2 & 0 & 2 & 0 & 0 & 0 & 1 & 1 \\ 2 & 0 & 0 & 0 & 1 & 2 & 2 & 2 & 1 & 0 & 0 & 1 & 1 \\ 3 & 0 & 0 & 0 & 1 & 0 & 1 & 2 & 2 & 1 & 0 & 1 & 2 \\ 4 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 3 & 1 & 1 & 0 & 1 \\ 5 & 2 & 2 & 0 & 1 & 0 & 0 & 1 & 0 & 3 & 1 & 1 & 1 \\ 6 & 0 & 2 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 2 & 1 & 1 \\ 7 & 2 & 2 & 2 & 1 & 1 & 1 & 0 & 0 & 1 & 1 & 1 & 0 \\ 8 & 0 & 1 & 2 & 3 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 9 & 0 & 0 & 1 & 1 & 3 & 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 10 & 0 & 0 & 0 & 1 & 1 & 2 & 1 & 0 & 1 & 0 & 1 & 0 \\ 11 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 1 & 0 & 1 \\ 12 & 1 & 1 & 2 & 1 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 \end{array}$$

(d)
$$R_5 = \begin{array}{r|cccccccccccc|} & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 \\ \hline 1 & 0 & 0 & 0 & 1 & 2 & 2 & 2 & 1 & 0 & 0 & 1 & 1 \\ 2 & 0 & 0 & 1 & 2 & 1 & 2 & 1 & 3 & 1 & 1 & 0 & 1 \\ 3 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 2 & 1 & 2 & 3 \\ 4 & 1 & 2 & 0 & 0 & 0 & 0 & 1 & 1 & 3 & 1 & 2 & 3 \\ 5 & 2 & 1 & 1 & 0 & 0 & 0 & 0 & 3 & 1 & 4 & 2 & 0 \\ 6 & 2 & 2 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 2 & 3 & 1 \\ 7 & 2 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 3 & 1 & 0 & 1 \\ 8 & 1 & 3 & 1 & 1 & 3 & 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 9 & 0 & 1 & 2 & 3 & 1 & 1 & 3 & 1 & 0 & 0 & 0 & 0 \\ 10 & 0 & 1 & 1 & 1 & 4 & 2 & 1 & 0 & 0 & 0 & 0 & 0 \\ 11 & 1 & 0 & 2 & 2 & 2 & 3 & 0 & 0 & 0 & 0 & 0 & 0 \\ 12 & 1 & 1 & 3 & 3 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \end{array}$$

(e)
$$R_5 = \begin{array}{r|cccccccccccc|} & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 \\ \hline 1 & 0 & 0 & 0 & 1 & 2 & 2 & 2 & 1 & 0 & 0 & 1 & 1 \\ 2 & 0 & 0 & 1 & 2 & 1 & 2 & 1 & 3 & 1 & 1 & 0 & 1 \\ 3 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 2 & 1 & 2 & 3 \\ 4 & 1 & 2 & 0 & 0 & 0 & 0 & 1 & 1 & 3 & 1 & 2 & 3 \\ 5 & 2 & 1 & 1 & 0 & 0 & 0 & 0 & 3 & 1 & 4 & 2 & 0 \\ 6 & 2 & 2 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 2 & 3 & 1 \\ 7 & 2 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 3 & 1 & 0 & 1 \\ 8 & 1 & 3 & 1 & 1 & 3 & 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 9 & 0 & 1 & 2 & 3 & 1 & 1 & 3 & 1 & 0 & 0 & 0 & 0 \\ 10 & 0 & 1 & 1 & 1 & 4 & 2 & 1 & 0 & 0 & 0 & 0 & 0 \\ 11 & 1 & 0 & 2 & 2 & 2 & 3 & 0 & 0 & 0 & 0 & 0 & 0 \\ 12 & 1 & 1 & 3 & 3 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \end{array}$$

FIG. 4. The matrices containing the number of paths of length (a) $h=2$, (b) $h=3$, (c) $h=4$, and (d) $h=5$ between each pair of nodes in the network in Fig. 1.

A. Algorithm for identification of number of paths

Many algorithms have been developed in order to find shortest paths in networks [21–23]. The number of paths can be determined by simple matrix manipulations. For instance, the number of paths of length 2 and 3 can be obtained as

$$R_2(i,k) = \sum_j A_{ij}A_{jk}(1 - \delta_{ik}) \tag{1}$$

and

$$R_3(i,l) = \sum_{j,k} A_{ij}A_{jk}A_{kl}(1 - \delta_{ik})(1 - \delta_{il})(1 - \delta_{jl}), \tag{2}$$

respectively, where $\delta_{i}j$ is the Kronecker delta. Nevertheless, this algorithm if of order $O(N^{h+1})$, where h is the length of the path. In this work, we proposed a faster algorithm, Algorithm 2, which allows the identification not only of the shortest paths, but also of all the alternative paths between a reference node i and all the other nodes in a network. Such algorithm is optimal in the sense that every path is determined without waste of calculation. It can be applied to direct and undirected networks. The operations $push(a)$ and $pop(a)$ place and remove the data a into a stack, respectively. Though this deterministic algorithm is optimal, it may require long periods of time depending on the type of network, its size, average degree, as well as the total number of steps H required. Stochastic algorithms such as that described in [24,25] can be considered for estimations in such cases. The execution of such an algorithm from all nodes on the network yields the set of matrices R.

The algorithm is based on agents moving through the network. One of the neighbors of the current node i, node j, is selected and the agent moves to this node. The remainder neighbors and their respective distance from node i are stored into a stack. Next, one of the neighbors of j is selected, and the agent moves to this node. Note that node i is not taken into account in this step, since it has already been visited during the walk. The agent continues to move until no more movements are possible. Then, one node at the stack is removed and the agent is placed in it in order to start the movement from this node. The process stops when all possible movements have been taken into account. The number of paths between each pair of nodes are calculated at each step and stored in the matrix R_h.

B. Decorrelation of measurements and dimensionality reduction

1. Principal component analysis

Because of the relatively high dimensionality of the path measurements, especially as a consequence of their parameterization with h, as well as the already observed strong correlations along h, it becomes important to consider means for obtaining effective projections of the measurements (dimensionality reduction) so as to visualize the network and node separations. This can be optimally performed through the method known as PCA.

PCA can be defined as the orthogonal projection of the original data onto a lower dimensional space, called the principal subspace, such that the variance of the projected data is maximized along its first axes [26]. Indeed, PCA can be understood as a rotation of the axes of the original variable coordinate system to new orthogonal axes in order to makes the new axes coincide with the directions of maximum variation of the original variables [27]. In practice, PCA consists initially of finding the eigenvalues and eigenvectors of the sample covariance matrix [28]. So, let each of Q observations (e.g., a node, a pair of nodes, or network), henceforth represented as $v=\{1,2,\ldots,Q\}$, be characterized in terms of M respective features or measurements each, represented in terms of the feature vector $\vec{f}_v$ [each element $f_v(i)$, $i \in \{1,2,\ldots,M\}$, of this vector corresponds to one measurement of the observation v]. For instance, we can consider the number of paths between each node i and all other nodes in the network. In this case, each node presents a feature vector with N elements. In cases where the number of features is large, it is possible to optimally reduce their dimensionality M by removing the correlations between them. This important dimensional reduction transformation can be easily implemented by using the PCA methodology (e.g., [4,27]).

Let the covariance between each pair of measurements i and j be given as

$$C(i,j) = \frac{1}{Q-1}\sum_{v=1}^{Q}[f_v(i) - \mu_i][f_v(j) - \mu_j], \tag{3}$$

where μ_i is the average of $f_v(i)$ over the Q observations, i.e.,

$$\mu_i = \frac{1}{Q}\sum_{v=1}^{Q} f_v(i). \tag{4}$$

The covariance matrix between these measurements is defined as $C=[C(i,j)]$, with dimension $M \times M$. Let the eigenvalues of C, sorted in decreasing order, be represented as λ_i, $i=1,2,\ldots,M$, with respective eigenvectors $\vec{v_i}$. By stacking such eigenvectors, it is possible to obtain the matrix

$$G = \begin{bmatrix} \uparrow & \uparrow & \cdots & \uparrow \\ \vec{v_1} & \vec{v_2} & \cdots & \vec{v_m} \\ \uparrow & \uparrow & \cdots & \uparrow \end{bmatrix}, \tag{5}$$

which defines the stochastic linear transformation known as Karhunen-Loève transform [4,27]. Now, the new feature vectors can be obtained from the original measurement vectors $\vec{f}$ by making

$$\vec{g} = G\vec{f}. \tag{6}$$

The variances of the new measurements in $\vec{g}$ are provided by the respective eigenvalues. In case where the measurements are correlated, most of their variances will be concentrated along the first elements of $\vec{g}$, which is guaranteed by the fact that the PCA completely decorrelates the original measurements. Indeed, the PCA is optimal with respect to concentrating the variation along the first axes while completely decorrelating all the original measurements. Therefore, it is possible to reduce the dimensionality of the features vectors by disregarding in the matrix in Eq. (5) all eigenvectors associated to eigenvalues smaller than a given threshold, or by taking only the R first eigenvectors. The resulting variables, which are fully uncorrelated linear combinations of the original measurements, concentrate the variance of the overall data and therefore represent a particularly meaningful characterization of the distribution of the original observations.

In this work, the matrix R_h is the feature (or measurement) matrix, whose rows represent the features of each node. The importance of PCA lies in the fact that the number of paths of different sizes tend to be highly correlated one another for subsequent values of h. From the matrix R_h, it covariance matrix as well as respective eigenvalues and eigenvectors are obtained. Selecting two or three of the eigenvectors corresponding to the highest eigenvalues and multiplying these eigenvectors by the original R_h matrix, it is possible to visualize the nodes into an optimally projected two-dimensional (2D) or three-dimensional space. Indeed, the principal axis i is given by the dot product of the i-th eigenvector by the original attribute matrix R_h.

2. Canonical variable analysis

Canonical variable analysis is a generalization of the principal component analysis [29]. Indeed, while PCA is a nonsupervisioned technique, the canonical variable analysis needs class information. Canonical analysis can be used to project the features of networks into a space so as to maximize the separation between the network models [30]. In this way, canonical analysis is suitable to visualize classes of networks into a two or three dimensional space, which allows to quantify the differences and similarities between them. In order to perform canonical analysis, it is necessary to construct a matrix which quantifies the variation inside each class of observations, and a second matrix, which quantifies the variation among these categories. If we consider C classes, types of network models, each one identified as C_i, $i=1,\ldots,C$ with N_i elements, and that each observation n is represented by its respective feature vector $\vec{x}_n=[x_n(1),x_n(2),\ldots,x_n(M)]^T$ composed by a set of M measurements, the intraclass scatter matrix is defined as

$$S_{\text{intra}}=\sum_{i=1}^{C}\sum_{n\in C_i}(\vec{x}_n-\langle\vec{x}\rangle_i)(\vec{x}_n-\langle\vec{x}\rangle_i)^T, \tag{7}$$

while the interclass scatter matrix is given as

$$S_{\text{inter}}=\sum_{i=1}^{C}N_i(\langle\vec{x}\rangle_i-\langle\vec{x}\rangle)(\langle\vec{x}\rangle_i-\langle\vec{x}\rangle)^T, \tag{8}$$

where $\langle\vec{x}\rangle_i$ corresponds to the average vector of all variables (measurements) for the class i and $\langle\vec{x}\rangle$ is the general average vector of all variables for all classes.

By computing the eigenvectors of the matrix $S_{\text{intra}}^{-1}S_{\text{inter}}$ and selecting those corresponding to highest absolute value eigenvalues, $\lambda_1,\ldots,\lambda_M$, it is possible to project the set of variables into a M-dimensional space, i.e., the canonical projection for a given observation n is obtained by

$$\vec{X}_n=\Gamma^T\vec{x}_n, \tag{9}$$

where

$$\Gamma=\begin{bmatrix}\uparrow & \uparrow & \cdots & \uparrow\\ \vec{\gamma}_1 & \vec{\gamma}_2 & \cdots & \vec{\gamma}_M\\ \uparrow & \uparrow & \cdots & \uparrow\end{bmatrix}, \tag{10}$$

corresponds to the eigenvectors of the matrix $S_{intra}^{-1}S_{inter}$. The eigenvectors $\vec{\gamma}_i$ are associated to the eigenvalue λ_i and they are arranged in Γ in increasing order according to their respective eigenvalues. Such a projection is guaranteed to maximize the interclass dispersion while minimizing the intraclass dispersion, leading to optimal separation between the objects belonging to the distinct categories. The overall quality of the separation between all categories can be quantified in terms of the following coefficient:

$$q=\text{tr}(S_{intra}^{-1}S_{inter}). \tag{11}$$

The higher the value of q, the best the separation between the categories [30].

III. CHARACTERIZATION OF THEORETICAL NETWORK MODELS

The following six different types of theoretical network models are considered in this article. The Erdős-Rényi (ER) random graphs [31] are obtained by connecting N initially isolated nodes with constant probability p. The traditional preferential attachment rule [32] is used to obtain the scale-free Barabási-Albert (BA) networks. Such a model can be understood as a particular case of the Krapivsky *et al.* [33] complex network model, which applies a nonlinear preferential attachment rule to establish connections during network growth—the probability of connection is defined as $\mathcal{P}_{i\rightarrow j}=k_j^{\alpha}/\Sigma_u k_u^{\alpha}$, where α is the nonlinear exponent. Observe that $\alpha=1$ yields the BA model. In the Watts-Strogatz (WS) small-world model, each connection in a linear lattice is rewired with probability p [18]. Geographical networks (GN) are obtained by starting with N nodes distributed uniformly along a two-dimensional space and connecting them according to distance, i.e., the probability to connect two nodes i and j is given by $P_{ij}=\lambda\exp(-\lambda d_{ij})$, where λ is a parameter to adjust the network degree and d_{ij} is the Euclidean distance between i and j. Such a model was introduced by Waxman in order to model the Internet topology [34]. Knitted networks (KT) [24] can be obtained by generating random sequences of nodes and connecting them sequentially (without repetition). The number of generated sequences depends on the network average connectivity. This network is particularly regular with respect to several of its topological and dynamical properties [24,25]. In the current work, all these networks are grown with parameter sets so as to have the same number N of nodes and average degree as similar as possible.

In order to visualize the network distribution and separation (discriminability), the set of feature vectors corresponding to the measurements of each network can be projected onto a 2D space of by canonical variable analysis. In the current work, we take into account as original measurements the mean, standard deviation, kurtosis and skewness of each matrix R_h. In this case, if we consider a maximum of H distances, we have a set of $4H$ measurements, and each network is represented by a feature vector $\vec{v}_n = \{\mu_1, \sigma_1, \kappa_1, \omega_1, \mu_2, \sigma_2, \kappa_2, \omega_2 \ldots, \mu_H, \sigma_H, \kappa_H, \omega_H\}$, where μ_h, σ_h, κ_h, and ω_h stand for the average, standard deviation, kurtosis, and skewness of the values in the matrix R_h, respectively. The network projections obtained by the canonical variable analysis reflect the network similarities in terms of their respective feature vectors. Indeed, models that are mapped nearby in the projected space present similar topological properties.

IV. RESULTS AND DISCUSSION

Our first experimental investigation concentrates in the characterization and discrimination between the topologies of six different complex networks theoretical models, namely: (i) the random graphs of ER [31], (ii) the small-world network model of WS [18], (iii) the geographical model of Waxman (GN) [34], (iv) the scale-free model of BA [32], (v) the nonpreferential attachment model of Krapivsky *et al.* (NL) [33] and (vi) the knitted network model of Costa (KT) [24]. We obtained the four first statistical moments, namely the mean (μ), standard deviation (σ), kurtosis (κ) and skewness ω of the matrices R_h for $h=2,\ldots,5$, for each network model realization. In this way, each generated network is represented in terms of a vector with 16 elements, *i.e.* the network n is represented by the respective vector $\vec{v}_n = \{\mu_1, \sigma_1, \kappa_1, \omega_1, \mu_2, \sigma_2, \kappa_2, \omega_2 \ldots, \mu_5, \sigma_5, \kappa_5, \omega_5\}$, where μ_h, σ_h, κ_h, and ω_h stand for the average, standard deviation, kurtosis and skewness of the values in the matrix R_h, respectively. We generated 25 network realizations for each model, considering $N=500$, and, after standardization of the feature vectors [35], they were projected into the 2D space by applying the canonical variable analysis methodology.

Networks generated by each model must present the same number of nodes and edges. The comparison of networks with different numbers of nodes and connections is artificial and can result in biased results. To obtain a significant comparison, the networks must present comparable (ideally the same) number of nodes and connections, since the differences between two networks should be reflected only in the organization of the connections. To compare networks of different size, it is necessary to consider the z score, as present in the following.

In Fig. 5, each of the respective types of networks generated by these models is represented by a respective independent cluster of points (networks share similar topological properties and therefore have similar measurements), which indicates a clear separation between each network theoretical model. Networks presenting similar structures, such as ER and WS networks, are projected nearby. Therefore, the networks are basically discriminated according to the organization of the connections, which is reflected in the distribution of the number of paths. While the networks generated by preferential attachment rule (BA and NL, orange and light-blue points) are organized at the right side of the projection, the most regular models (KT and ER, gray and blue points) are found at the left side. In addition, the network models that generate networks with more regular structure tend to present the smallest cloud dispersions (KT and WS, gray and green points). In this way, by providing accurate discriminability between different models, the generalized connectivity approach presents good potential for enhancing network characterization and classification of networks, as well as for establishing relationships between them.

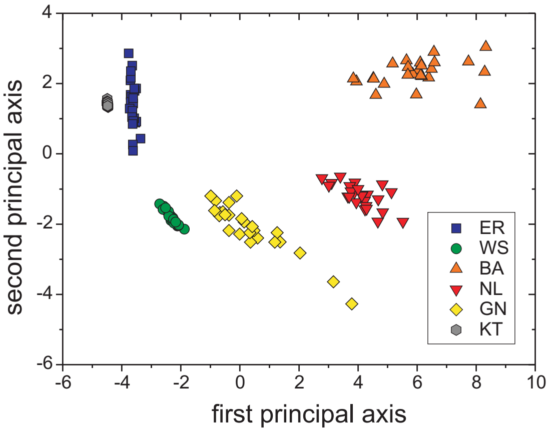

FIG. 5. (Color online) The projection of the networks generated by the ER (square), WS (circle), BA (triangle), NL (upside down triangle), GN (diamond), and KT (hexagon) network models in the two-dimensional space.

We compared the discrimination between theoretical models obtained by taking into account the number of paths with that achieved by using traditional network measurements. In the latter case, we calculated the mean degree, average clustering coefficient, average shortest path length, central point dominance, mean betweenness centrality, and assortativity coefficient from the same set of networks which were considered in the number of paths analysis. All these measurements are described in [4]. Figure 6 presents the obtained projection. The discrimination between the classes can be quantified in terms of the coefficient presented in Eq. (11). In this way, while the moments of the number of paths provided a discrimination coefficient equal to $q=608$, the tradi-

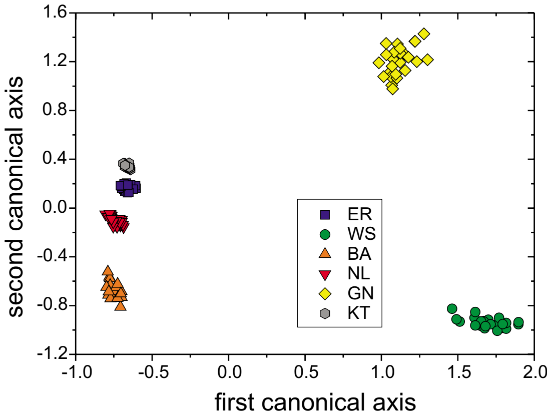

FIG. 6. (Color online) The projection of the networks generated by the ER (square), WS (circle), BA (triangle), NL (upside down triangle), GN (diamond), and KT (hexagon) network models in the two-dimensional space by taking into account traditional network measurements, i.e., mean degree, average clustering coefficient, average shortest path length, central point dominance, mean betweenness centrality, and assortativity coefficient.

TABLE I. The z scores and the average number of paths (indicated between parenthesis) obtained for the real-world networks investigated in this work.

Network	N	$\langle k\rangle$	Z_2 ($\langle R_2\rangle$)	Z_3 ($\langle R_3\rangle$)	Z_4 ($\langle R_4\rangle$)	Z_5 ($\langle R_5\rangle$)	Z_6 ($\langle R_6\rangle$)
Food web	78	3.1	0.002 (0.05)	−0.087 (0.03)	−0.12 (0)	−0.13 (0)	−0.11 (0)
Cortical network	53	15.5	−0.027 (5)	−0.043 (85)	−0.06 (1400)	−0.08 (21800)	−0.11 (331000)
Neural network	297	7.9	0.026 (0.30)	0.045 (3)	0.01 (25)	−0.05 (210)	−0.10 (1750)
US Highway	284	6.0	0 (0.02)	−0.048 (2)	−0.06 (13)	−0.06 (100)	−0.06 (680)

tional measurements resulted in a coefficient equal to $q=439$. Note that the larger this coefficient, the better the separation between classes. So, the number of paths provide a better discrimination than the network measurements. At the same time, the groups obtained by using the path-based measurements are substantially more compact than those produced by the traditional measurements, which further corroborate the better discriminative power of the former type of measurements.

We also illustrate the potential of the identification of alternative paths with respect to the following real-world networks: (i) US highway network, (ii) neural *C. elegans* network [18], (iii) cat cortical network [19], and (iv) a food web of a broadleaf forest in New Zealand [20]. Details about these networks are given in Table I. Since these networks present different number of nodes and connections, they cannot be directly compared—note that the number of paths for the cortical network is higher than for the other networks, which is a direct consequence of its higher average node degree. In this way, we considered the z score in order to characterize the distribution of paths. The z score is calculated as [36]

$$Z_h = \frac{\mu_h - \mu_{random}}{\sigma_{random}}, \tag{12}$$

where μ_h is the average number of paths of length h in the real network, and μ_{random} and σ_{random} are the average and standard deviation of the number of paths in the respective randomized network ensemble, which were generated by the configuration model and present the same degree distribution as the respective real-world network [37]. The obtained results for the four networks are presented in Table I. It is interesting to note that just the neural network of the nematode *C. elegans*, which is the only case of a nervous system completely mapped at the level of neurons and chemical synapses [18], presents larger number of paths of lengths $h=2$, 3, and 4 than the randomized counterparts. For $h>4$, the randomized versions present higher number of paths. This suggests that connections of length 2, 3, and 4 could be more important for allowing proper dynamics in the C. elegans network. The highest difference for $h=3$ suggests that the evolution of the neuronal organization in this species tended to favor the alternative connections of length 3, while avoiding longer range connections. On the other hand, in case of the food web, the cortical network and the US highway, the z scores decrease with h, which indicates that such networks tend to present smaller number of paths of length $h>2$ than their randomized versions. Particularly, since food web tend to present a small number of trophic levels, there are no paths of length $h>4$, while the randomized version can display longer path sizes. Indeed, the small network diameter is a direct consequence of the energy transmission between trophic levels [38]. In the case of the highway network, the fact that the randomized versions tended to present larger number of paths than the respective real-world version is a direct consequence of the fact that the connections in geographical highway network are likely to be constrained by the adjacency between neighboring localities.

We also investigated the distribution of alternative paths in modular networks. Since pairs of nodes belonging to a same community tend to be more strongly connected, the number of paths between them tends to be relatively large. We investigated the relationship between the number of paths and community structure with respect to the Zachary karate club network and to an artificial modular network of the type which has been widely used as tests for community structure algorithms (e.g., [39]). Note that the consideration of this networks is only to exemplify the relationship between the number of paths and network modular organization.

The karate club network was constructed with the data collected while observing 34 members of a karate club over a period of 2 years and considering friendship between members [40]. From this network, we calculated the respective R_h matrices for $h=1$, 2, and 3. After standardization of the feature vectors, we applied the PCA on the matrices R_h and obtained the projections presented in Fig. 7. Note that each node i presents a respective feature vector corresponding to the line (or row) i of R_h. The best separation of the Zachary karate club network was obtained for $h=2$, with the classification of the nodes into the two clusters corresponding precisely to the actual division of the club members. The case $h=1$, which is equivalent to the traditional adjacency matrix, does not provide an accurate separation of the communities into different clusters. For $h\geq 3$, the separation is worse than for $h=2$ because the network presents a very small average shortest distance ($\ell=2.3$). When taking into account the shortest path matrix, the discriminability also resulted worse than that obtained for R_2. In fact, the best value of h can be determined in terms of the modularity value [41], i.e., the separation can be calculated using different R_h matrices and then choosing the one which results in the largest modularity value.

The relationship between the number of paths and the network modular organization was also investigated with respect to computer-generated networks with a known community structure, as described in [39]. These networks are

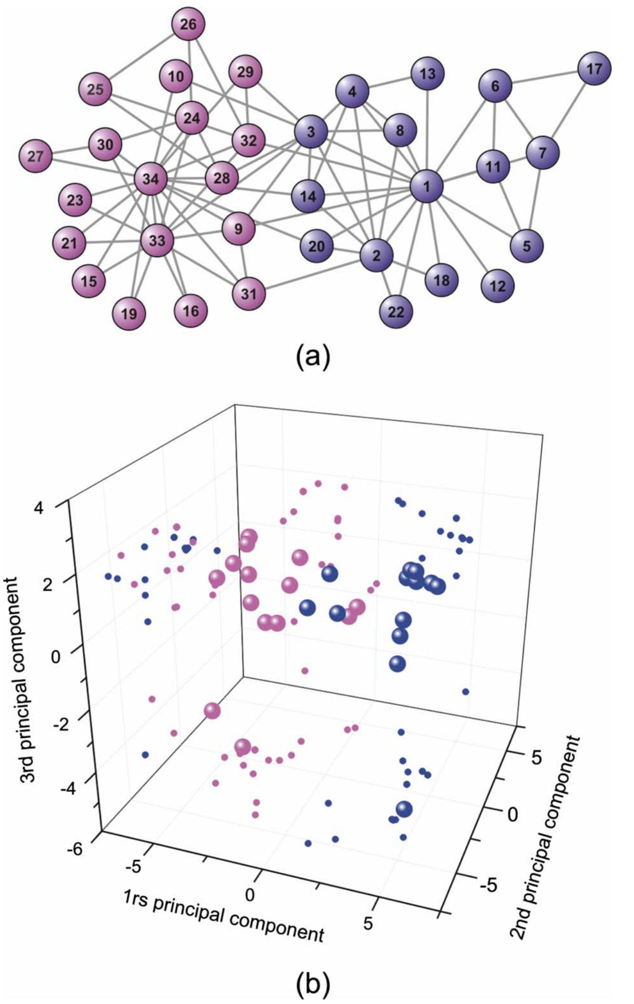

FIG. 7. (Color online) The original separation between the two classes of karate club member (a), and the projection into the three-dimensional space of the generalized matrix R_2.

formed by 128 vertices grouped into four communities of 32 vertices each. Each vertex has z_{in} links to vertices in the same community and z_{out} edges to vertices in other communities. The vertices between communities were distributed uniformly. Figure 8 presents an example of the obtained community networks (z_{out}=6) and the respective projection of the R_2 matrix into three dimensions. As we can see, the communities are perfectly separated, which suggests that short paths tend to connect nodes inside the same community.

V. CONCLUDING REMARKS

The concept of connectivity underlies great part of complex networks research. However, connectivity has typically been understood and quantified in terms either of strictly local measurements between neighboring nodes, such as the local degree, or by considering shortest path lengths. Though more global, the latter feature fails to take into account alternative pathways between pairs of nodes, which are extremely important in influencing the topological properties of the networks. For instance, the presence of more than one path between two nodes tends to increase the connectivity between them and consequently raises their communication robustness under edge disruption.

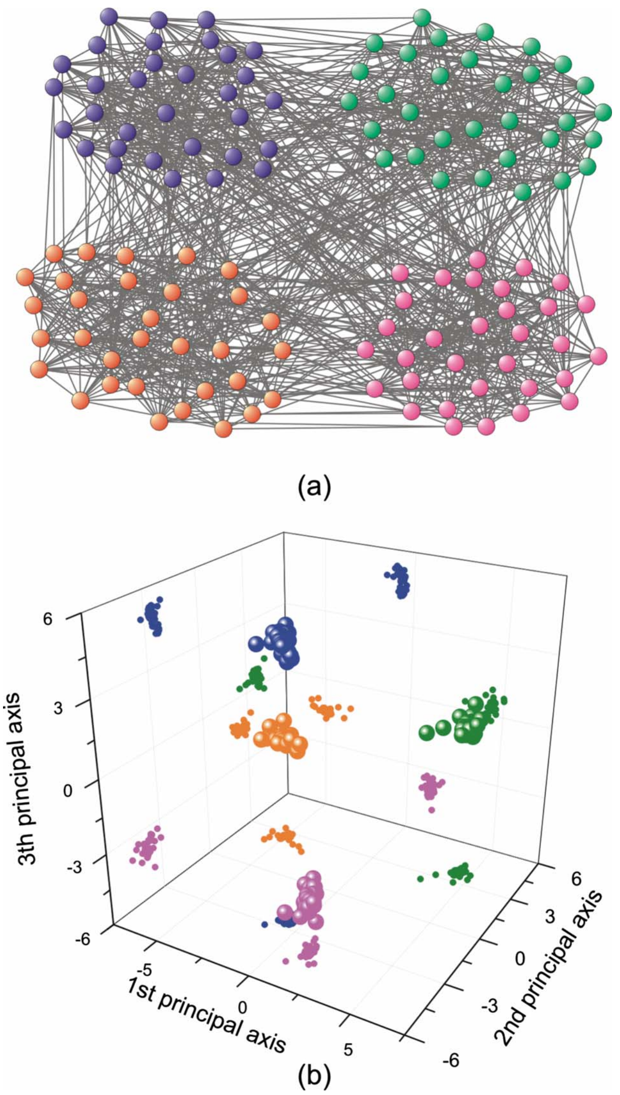

FIG. 8. (Color online) (a) The artificial network containing four communities and (b) the projection of the respective matrix R_2 into the three-dimensional space considering the PCA methodology.

In the current paper, we analyzed the generalized network connectivity, i.e., the consideration of alternative paths of different lengths between each pair of nodes, with respect to the characterization of six theoretical network models and four real-world networks, as well as for investigation of the relationship between the number of paths and community organization. We showed that the consideration of the alternative paths between nodes can provide accurate network topology discriminability while identifying interesting relationships, as observed for the networks generated by the different models. The analysis of real-world networks suggests that the long-range connectivity tends to be limited in those networks and may be strongly related to network evolution and organization. In addition, we studied how the distribution of the number of paths is related to network modular structure. The obtained results indicate that the proposed approach is potentially promising for community identification. In addition, a possibility for future work would be the consideration of pattern recognition approaches to quantify the separation between several types of networks models and therefore provide complex networks taxonomies. In this case, real-world networks can be associated to the most likely theoretical model, as described in [4]. Studies relating the number of paths with network dynamics constitute another promising research possibility.

Algorithm 1 The general algorithm to obtain the number of paths between each pair of nodes.

```
for each node $i$ do
  $h=1$;
  $next$=one of the nonvisited immediate neighbors of $i$;
  $stack.push$(remainder of nonvisited immediate neighbors of $i$, $h$);
  $path.push(next)$;
  $R(next,i,h)=1$;
  while stack not empty or size($path$) $>0$ do
    $curr=next$;
    $ng$=number of nonvisited neighbors of $curr$;
    $h=h+1$;
    if $ng>0$ then
      $next$=one of the nonvisited neighbors of $curr$;
      $stack.push$(remainder of nonvisited immediate neighbors of $curr$, $h$);
      $path.push(next)$;
    else
      $next,h=stack.pop$(one node, $h$);
      $node=-1$;
      while $node \neq next$ do
        $node=path.pop()$;
        Set $node$ as not visited;
      end while
    end if
    $R(next,i,h)=R(next,i,h)+1$;
  end while
end for
```

ACKNOWLEDGMENTS

The authors thank Lucas Antiqueira, Bruno N. Travençolo, and Renato A. Pimentel for comments on a preliminary version of this article. Luciano da F. Costa thanks CNPq (Grant No. 301303/06-1) and FAPESP (Grant No. 05/00587-5) for sponsorship. F.A.R. is grateful to FAPESP (Grant No. 07/50633-9) for financial support.

[1] H. Kitano, Science **295**, 1662 (2002).

[2] D. J. Watts, P. S. Dodds, and M. E. J. Newman, Science **296**, 1302 (2002).

[3] L. F. Costa *et al.*, e-print arXiv:0711.3199.

[4] L. da F. Costa, F. A. Rodrigues, G. Travieso, and P. R. V. Boas, Adv. Phys. **56**, 167 (2007).

[5] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D. U. Hwang, Phys. Rep. **424**, 175 (2006).

[6] E. Estrada and N. Hatano, Phys. Rev. E **77**, 036111 (2008).

[7] E. Ziv, R. Koytcheff, M. Middendorf, and C. Wiggins, Phys. Rev. E **71**, 016110 (2005).

[8] Y. Shavitt and Y. Singer, New J. Phys. **9**, 266 (2007).

[9] H. E. Stanley, Nature (London) **378**, 554 (1995).

[10] M. A. Nicolelis, C. H. Yu, and L. A. Baccala, Comput. Biol. Med. **20**, 379 (1990).

[11] R. F. S. Andrade *et al.*, Eur. Phys. J. B **61**, 247 (2008).

[12] J. P. Bagrow, E. M. Bollt, and J. D. Skufca Europhys. Lett. **81**, 68004 (2008).

[13] A. E. Motter and Y. C. Lai, Phys. Rev. E **66**, 065102(R) (2002).

[14] L. F. Costa, Phys. Rev. Lett. **93**, 098702 (2004).

[15] L. da F. Costa and R. F. S. Andrade, New J. Phys. **9**, 311 (2007).

[16] L. da F. Costa and F. N. Silva, J. Stat. Phys. **125**, 845 (2006).

[17] L. da F. Costa and L. E. C. da Rocha, Eur. Phys. J. B **50**, 237 (2006).

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

[19] O. Sporns and J. D. Zwi, Neuroinformatics **2**, 145 (2004).

[20] N. G. Jaarsma, S. M. De Boer, C. R. Townsend, R. M. Thompson, and E. D. Edwards, N.Z.J. Mar. Freshwater Res. **32**, 271 (1998).

[21] P. Brucker, Computing **10**, 271 (1972).

[22] T. C. Hu, SIAM J. Appl. Math. **15**, 1517 (1967).

[23] D. Lohse, Computing **17**, 93 (1976).

[24] L. da F. Costa, e-print arXiv:0711.2736.

[25] L. da F. Costa, e-print arXiv:0712.0415.

[26] H. Hotelling, J. Educ. Psychol. **24**, 417 (1933).

[27] L. da F. Costa and R. M. Cesar, Jr., *Shape Analysis and Classification: Theory and Practice* (CRC Press, Boca Raton, FL, 2001).

[28] C. M. Bishop, *Pattern Recognition and Machine Learning* (Springer-Verlag, New York, 2006).

[29] N. A. Campbell and W. R. Atchley, Syst. Zool. **30**, 268 (1981).

[30] K. Fukunaga, *Introduction to Statistical Pattern Recognition* (Academic Press, New York, 1990).

[31] P. Erdős and A. Rényi, Publ. Math. Debrecen **6**, 290 (1959).

[32] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).

[33] P. L. Krapivsky, S. Redner, and F. Leyvraz, Phys. Rev. Lett. **85**, 4629 (2000).

[34] B. M. Waxman, IEEE J. Sel. Areas Commun. **6**, 1617 (1988).

[35] The standardization of a random variable consists of subtracting its respective average and dividing by the standard deviation. The resulting transformed random variable necessarily has zero mean and unitary standard deviation [27].

[36] R. J. Larsen and M. L. Marx, *An Introduction to Mathematical Statistics and Its Applications* (Prentice-Hall, Englewood Cliffs, NJ, 1981).

[37] E. A. Bender and E. R. Canfield, J. Comb. Theory, Ser. A **24**, 296 (1978).

[38] M. E. Power, Science **250**, 811 (1990).

[39] M. Girvan and M. E. J. Newman, Proc. Natl. Acad. Sci. U.S.A. **99**, 7821 (2002).

[40] W. W. Zachary, J. Anthropol. Res. **33**, 452 (1977).

[41] M. E. J. Newman, Eur. Phys. J. B **38**, 321 (2004).