Physica D 323–324 (2016) 73–79

Contents lists available at ScienceDirect

Physica D

journal homepage: www.elsevier.com/locate/physd

Random walk centrality in interconnected multilayer networks

Albert Solé-Ribalta, Manlio De Domenico, Sergio Gómez, Alex Arenas *

Departament d'Enginyeria Informàtica i Matemàtiques, Universitat Rovira i Virgili, 43007 Tarragona, Spain

HIGHLIGHTS

- We provide analytical expressions for the centrality of random walks in interconnected multilayer networks.
- We check the theoretical results with extensive Monte Carlo simulations of random walkers in different topologies, and achieve an excellent agreement.
- Our results are useful for the ranking of nodes in multi-categorical systems.

ARTICLE INFO

Article history:
Received 23 June 2015
Received in revised form
8 January 2016
Accepted 8 January 2016
Available online 2 March 2016

Keywords:
Multilayer complex networks
Random walks
Centrality

ABSTRACT

Real-world complex systems exhibit multiple levels of relationships. In many cases they require to be modeled as interconnected multilayer networks, characterizing interactions of several types simultaneously. It is of crucial importance in many fields, from economics to biology and from urban planning to social sciences, to identify the most (or the less) influent nodes in a network using centrality measures. However, defining the centrality of actors in interconnected complex networks is not trivial. In this paper, we rely on the tensorial formalism recently proposed to characterize and investigate this kind of complex topologies, and extend two well known random walk centrality measures, the random walk betweenness and closeness centrality, to interconnected multilayer networks. For each of the measures we provide analytical expressions that completely agree with numerically results.

1. Introduction

It is common practice in many studies involving networks to assume that nodes are connected by a single type of edge that encapsulates all relations between them. In a myriad of applications this assumption oversimplifies the complexity of the system, leading to inaccurate or wrong results. Examples can be found in temporal networks, where neglecting time-dependence washes out the memory of sequences of human contacts in transmission of diseases [1], in co-authorship networks, where neglecting the existence of multiple relationships between actors might alter the topology which may lead to misestimating crucial node's properties [2–7] or in transportation networks where the multilayer topology must be considered to accurately model the dynamics to *a posteriori* predict congested locations [8].

Historically, the term *multiplex* was coined to indicate the presence of more than one relationship between the same actors of a social network [9]. This type of network is well understood in terms of "coloring" (or labeling) the edges corresponding to interactions of different nature. For instance, in a social network the same individual might have connections to other individuals based on financial interests (e.g., color red) and connections with the same or different individuals based on friendship (e.g., color blue). In other real-world systems, like the transportation network of a city, the same geographical position can be part, for instance, of the network of subway or the network of bus routes, simultaneously. In this specific case, an edge-colored graph would not capture the full structure of the network, since information about the cost to *move* from the subway network to the bus route is missing. This cost can be economic or might account for the time required to physically commute between the two layers. It is in this cases where an interconnected multilayer network provides a better representation of the system. Fig. 2 shows an illustration of an interconnected multilayer (Fig. 2(A)) and the classical representation with an aggregated network (Fig. 2(C)). It is evident that a simple projection of the former – mathematically equivalent to sum up the corresponding adjacency matrices of the individual layers – would provide a network where the information about the relation type is lost. On the other hand, an edge-colored graph (Fig. 2(B)) cannot account for interconnections. For further details about the classification of such multilayer networks we refer to [10] and references therein. In the rest of the paper interconnected multilayer networks will be referred in short as multilayer networks.

* Corresponding author. Tel.: +34 977559687.
E-mail address: alexandre.arenas@urv.cat (A. Arenas).

http://dx.doi.org/10.1016/j.physd.2016.01.002
0167-2789/

The remainder of this paper is organized as follows. In Section 2 we briefly describe the tensorial notation, defined in [11], adopted overall the paper. In Section 3 we capitalize on this notation to extend some random walk centrality descriptors, well known in the case of single layer networks, to interconnected multilayer networks. Finally, we discuss our findings in Section 4.

2. Tensorial notation

Edge-colored graphs can be represented by a set of adjacency matrices [12–15]. However, standard matrices, used to represent networks, are limited in the complexity of the relationships that they can capture, i.e., they do not represent a suitable framework in the case of multilayer networks. This is the case of multiple types of relationships – that can also change in time – between nodes. Such a level of complexity can be characterized by considering tensors and algebras of higher order [11].

A great advantage of tensor formalism developed in [11] relies on its compactness. An adjacency tensor can be written using a more compact notation that is very useful for the generalization of network descriptors to multilayer networks. In this notation, a row vector $\mathbf{a} \in \mathbb{R}^N$ is given by a covariant vector a_α ($\alpha = 1, 2, \ldots, N$), and the corresponding contravariant vector a^α (i.e., its dual vector) is a column vector in Euclidean space. A canonical vector is assigned to each node and the corresponding interconnected multi-layer network is represented by a mixed rank-4 adjacency tensor.

However, in the majority of applications, it is not necessary to perform calculations using canonical vectors and tensors explicitly. In this cases, a classical single-layer network can be represented by a rank-2 mixed adjacency tensor W^α_β [11], where the layer information is disregarded. But, in general, systems may exhibit several types of relationships between pairs of nodes and a more general system represented as a multilayer object – in which each type of relationship is represented within a single *layer* α ($\alpha = 1, 2, \ldots, L$) of the network – is required.[1] In these cases, we use an *intra-layer adjacency tensor* for the 2nd-order tensor $W^i_j(\alpha)$ that indicates the relationships between nodes within the *same* layer α and the 2nd-order *inter-layer adjacency tensor* $C^i_j(\alpha\beta)$ to encode information about relationships that incorporate multiple layers.

It has been shown that the mathematical object accounting for the whole interconnected multilayer structure is given by a 4th-order (i.e., rank-4) *multilayer adjacency tensor* $M^{i\alpha}_{j\beta}$. This tensor might be simply thought as a higher-order matrix with four indices. It is the direct generalization of the adjacency matrix in the case of single layer networks and encodes the intensity of the relationship (which may not be symmetric) from a node i in layer α to a node j in layer β [11].

To reduce the notational complexity in the tensorial equations the Einstein summation convention is adopted. It is applied to repeated indices in operations that involve tensors. For example, we use this convention in the left-hand sides of the following equations:

$$A^i_i \equiv \sum_{i=1}^{N} A^i_i, \qquad A^i_j B^j_i \equiv \sum_{i=1}^{N}\sum_{j=1}^{N} A^i_j B^j_i,$$

$$A^{i\alpha}_{j\beta} B^{k\beta}_{i\gamma} \equiv \sum_{i=1}^{N}\sum_{\beta=1}^{L} A^{i\alpha}_{j\beta} B^{k\beta}_{i\gamma},$$

whose right-hand sides include the summation signs explicitly. It is straightforward to use this convention for the product of any number of tensors of any order. In the following, we will use the tth power of rank-4 tensors, defined by multiple tensor multiplications:

$$(A^t)^{i\alpha}_{j\beta} = (A)^{i\alpha}_{j_1\beta_1} (A)^{j_1\beta_1}_{j_2\beta_2} \ldots (A)^{j_{t-1}\beta_{t-1}}_{j\beta}. \tag{1}$$

Repeated indices, such that one index is a subscript and the other is a superscript, is equivalent to perform a tensorial operation known as a *contraction*. Moreover, one should be very careful in performing tensorial calculations. For instance, using traditional notation the product $a^i b^j$ would be a number, i.e., the product of the components of two vectors. However, in our formulation, the same calculation denotes a Kronecker product between two vectors, resulting in a rank-2 tensor, i.e., a matrix.

3. Random walk centrality measures in multilayer networks

In practical applications one is often interested in assigning a global measure of importance to each node. If the system we deal with contains several types of relations between actors we expect that the measures, in some way, consider the importance obtained from the different layers. A simple choice could be to combine the centrality of the nodes – obtained from the different layers independently – according to some heuristic choice. This is a viable solution when there is no interconnection between layers, i.e., in the case of edge-colored graphs [16,17]. However, the main drawback of this approach is that it depends on the choice of the heuristics and thus might not evaluate the actual importance of nodes. Our approach accounts for the higher level of complexity of such systems without relying on external assumptions and naturally extends the well-known centrality measures adopted for several decades in the case of single layer networks.

A random walk is one of the simplest dynamical process that can occur on a network, and random walks can be used to approximate other types of diffusion processes[18,19]. Random walks on networks [18,20,19] have attracted considerable interest because they are both important and easy to interpret. They have yielded important insights on a huge variety of applications and can be studied analytically. For example, random walks have been used to rank Web pages [21] and sports teams [22], optimize searches [23], investigate the efficiency of network navigation [24,25], characterize cyclic structures in networks [26], and coarse-grain networks to highlight meso-scale features such as community structure [27–29]. Another interesting application of random walks is to calculate the centrality of actors in complex networks when there is no knowledge about the full network topology but only local information is available. In such cases, centrality descriptors based on shortest-paths, e.g., betweenness and closeness centrality, should be substituted by centrality notions based on random walks [20,30]. In the following we extend these measures to multilayer networks.

First of all, we define a discrete-time random walk, between two individuals o and d, $o \rightarrow d$, on a multilayer network consisting of L layers and N nodes per layer, as a random sequence of nodes which starts from node o in any layer and finish in node d in any layer where each edge's endpoints are the preceding and following vertices in the sequence. The reasoning behind this definition is that the different node replicas in the different layers correspond to the same individual and so anything traveling between them is independent on the starting and ending layer. Fig. 1 shows and example of a random walk between two nodes in a multilayer network where it is evident the introduction of non-trivial effects because of the presence of inter-layer connections that affects its navigation in the networked system [31].

[1] To avoid confusion, in the following we refer to nodes with Latin letters and to layers with Greek letters, allowing us to distinguish indices that correspond to nodes from those that correspond to layers in tensorial equations.