Identifying Influential Spreaders in Complex Multilayer Networks: A Centrality Perspective

Pavlos Basaras, George Iosifidis, Dimitrios Katsaros, and Leandros Tassiulas, *Fellow, IEEE*

Abstract—Identifying influential spreaders in complex networks is of paramount importance for understanding and controlling the spreading dynamics. A challenging and yet inadequately explored task is to detect such influential nodes in multilayer networks, i.e., networks that encompass different types of connections (e.g., different relationships) among the nodes, hence facilitating a multilayer structure. Our purpose is to devise a method that can accurately detect nodes able to exert strong influence over the multilayer network; the method will be based solely on local knowledge of a network's topology in order to be fast and scalable due to the huge size of the network, and thus suitable for both real-time applications and offline mining. Based on our belief that a strong influencer is a node positioned in a well-connected neighborhood, we propose a series of methods which capture in a single number the rich inter- and intra-layer connectivity of the node. Our simulations showed that the proposed measures can detect effective spreaders in both real and synthetic networks, under various settings and against various competitors.

Index Terms—Centralities, influential spreaders, epidemic spreading, multilayer networks, complex networks

1 Introduction

The study of complex networks [41], i.e., the discipline called network science, is experiencing a blossom in the last decade. A driving reason for this is the abundance of data coming both from online social networks such as Facebook, Twitter, and also from more traditional sources such as the Web, Internet, mobile calling patterns, human interactions and so on. These online traces have enabled the development of algorithms for the analysis of the properties, functioning, and growth of these networks. Among the many problems addressed in the literature of complex networks, the identification of influential spreaders, i.e., the detection of nodes that can affect a large number of other nodes, is of paramount importance in hopes of understanding the spreading dynamics over complex networks, for seed selection in influence maximization problems [25], etc. In simple terms, the problem can be expressed as the need to locate those nodes, who if activated, can effectively propagate information (e.g., rumors, advertisements, products, etc.) to a significantly large subset of network nodes [4], [24], [27], [34], [53], or if isolated, can efficiently mitigate the diffusion of undesired "things" (e.g., a virus) [5], [30].

So far, the literature on this topic—and the study of complex networks in general—has focused on single-layer networks, where the entities (nodes) and their "communication" channels (links) are assumed to belong to the same network. However, the last few years, we are witnessing a phenomenal initiative in the analysis of new kinds of complex networks, where the interacting entities are assumed to belong to more than one network, called *layers*. These networks are termed multiplex [9], multisliced [39], multilevel [52], interdependent [7] or more general, multilayer networks [6], [28]. Online social networks, financial systems, transportation networks are such networks to name a few; more detailed examples can be found in [6], [28]. Research in the realm of multilayer networks investigates topics such as centralities [37], communities [23], growth models [47] and so on. Similarly, the study of spreading processes in multilayer networks has started to attract significant interest, however the field is still developing its basic principles [45]. On the other hand, the literature on developing algorithms for identifying influential spreaders in multilayer networks is yet very narrowed (cf. Section 6). However, the spreading of information, rumors, advertisements, or broadly speaking anything that can be 'shared' through networked populations is rarely isolated into a single network; for instance, information propagation over social networks is taking place in a fashion such that a user decides to share a 'chunk of information' through his/her account, in both Facebook and Twitter.

The identification of influential spreaders in single-layer networks, after the seminal work [27], has concentrated around the idea of 'network decomposition' using concepts such as the k-shell, the k-truss [51], the onion decomposition [21], and so on. All these techniques are iterative and therefore slow; they require knowledge of global network connectivity, in order to locate nodes which are highly connected, hoping that they are also good spreaders. However,

- *P. Basaras is with the Department of Electrical & Computer Engineering, University of Thessaly, Volos 382 21, Greece. E-mail: pbasaras@gmail.com.*
- *G. Iosifidis is with the School of Computer Science and Statistics, Trinity College Dublin 2, Ireland. E-mail: giosifid@gmail.com.*
- *D. Katsaros and L. Tassiulas are with the Department of Electrical Engineering & Yale Institute for Network Science, Yale University, New Haven, CT 06520. E-mail: {d.katsaros, leandros.tassiulas}@yale.edu.*

Manuscript received 9 Mar. 2017; revised 4 Oct. 2017; accepted 4 Oct. 2017. Date of publication 16 Nov. 2017; date of current version 13 Mar. 2019. (Corresponding author: Dimitrios Katsaros.)
Recommended for acceptance by M. Porter.
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TNSE.2017.2775152

all these methods are inapplicable in multilayer networks, because they result in a vector of values for each node [3], i.e., the value of k-shell, or of the k-truss of the node in each layer. Thus, the ranking of nodes using these vectors is not straightforward, unless we define a set of weights for the set of layers and compute a score out of these weights. Apparently, the introduction of artificial weights and computations over them is arbitrary and thus not desirable. An alternative is to address the problem as a "rank aggregation" problem [18], [32], and fuse the ranking lists produced by each value; still, the selection of the fusion algorithm will raise questions about its appropriateness and fairness. On the other hand, the use of centrality measures, such as the shortest-path betweenness centrality, presents the same drawbacks as their counterparts for single-layer networks as analyzed in [4], whereas the use of a PageRank centrality measure adopted for multilayer networks as in [20], has the drawback that its computation requires an artificial ordering among the layers of the complex multilayer network, therefore making this solution to depart from reality. On the other hand, the elegant and mathematically sound generalization of PageRank reported in [15] simply suffers from the computational complexity of the original PageRank, i.e., it is network-wide and iterative, thus time-consuming.

A different line of research on the topic of single-layer influential spreaders detection was described in [4], where the concept of *Power Community Index* (*PCI*) (cf. Definition 1)—and also in [4], [17]—was proposed to detect highly effective spreaders. The proposed method is localized, requiring only local (i.e., two hop) neighborhood information, is fast and proved superior to k-shell. The connectivity of the nodes identified as highly influential spreaders with the aid of PCI is in accordance with the findings of the study [38], which proved analytically that the most effective influential spreaders are those who "...are relatively low-degree nodes surrounded by hierarchical coronas of hubs." In principle, the generalization of the ideas of PCI for multilayer networks would be appropriate, because it would be based on local information of the topology, thus minimizing the computation cost and eliminating the need for having complete knowledge of the entire network state, hence being a good candidate even for real-time applications over massive multilayer complex networks.

This article investigates the problem of identifying influential spreaders over complex multilayer networks, by introducing a family of centrality-like measures tailored for local computation only, and able to locate nodes in dense areas of the multilayer network with many intra- and inter-layer links facilitating the rapid evolution of a diffusion process. The article makes the following contributions:

- It thoroughly investigates the topic of identifying influential spreaders in multilayer networks by maintaining and exploiting the multilayer structure, i.e., without blending and/or weighting—and thus eliminating—the layers as done by [2] (such an approach has already been proven inadequate and inefficient [15]).
- It proposes a family of localized measures that effectively and efficiently address the problem of influentials identification by incorporating multilayer characteristics (existence and density of intra- and inter-layer connections). The proposed methods can be straightforwardly adapted to any type of multilayer network.
- It evaluates the proposed techniques in a wealth of real and semi-synthetic multilayer networks using as competitors all the major high-performing measures, i.e., PageRank, Betweenness, Degree, k-core and their multilayer variations.
- It concludes that one of the proposed methods, namely *mlPCI* is (almost) always the best-performing method irrespectively of the size and characteristics of the investigated complex networks, whereas the traditional ones such as PageRank and Betweenness centrality fail to achieve competitive performance.

TABLE 1
Notation for Multilayer Networks

Notation	Description
G_i	A monoplex network i
V_i	The set of nodes of the monoplex network i
E_i	The set of edges of the monoplex network i
$\mathcal{P}$	A multilayer network
L	The set of layers of the multilayer network
$\mathcal{G}$	A set of monoplex networks: G_i, $i \in (1, N)$
$\mathcal{E}$	A set of edges between different monoplexes
λ_{ii}	Spreading rate at layer i
λ_{ij}	Spreading rate from layer i to j
k_{in}, k_{out}	in-degree, out-degree

The remainder of this paper is organized as follows. Section 6 briefly contemplates related articles. In Section 2 we provide formal definitions and notations for multilayer networks. Section 3 describes and exemplifies the proposed methods, whereas Section 4 outlines the experimentation settings, datasets, competitors and performance measures. In Section 5 results are demonstrated, and finally Section 7 concludes the article.

2 Preliminaries

We are interested in two types of networks, (i) generic multilayer networks, and (ii) multiplex networks. We adopt a graph-theoretic notation and terminology, similar to the one presented in [6]. On the other hand, tensors comprise a similarly powerful, and more compact way to represent multilayer networks; they have been used extensively for the representation of such networks, and for the calculation of centralities and communities in them, e.g., [14], [40]. However, since the measures we introduce in Section 3 make use only of local (around each node) information and they can be very easily described with graph-theoretic terms, we prefer to use the graph-theoretic representation. The rest of the section reviews the notation (Table 1) of multilayer networks and the spreading model.

2.1 Monoplex, Multiplex and Multilayer Networks

A *Single* or *Monoplex* network is represented as a graph $G_i(V_i, E_i)$, where V_i is the set of nodes and E_i is the set of edges which connect those nodes. Edges can be directed or undirected, weighted or unweighted. A *multilayer network* can be described as a combination of graphs, $G_1, G_2, \ldots, G_{|L|}$, and a

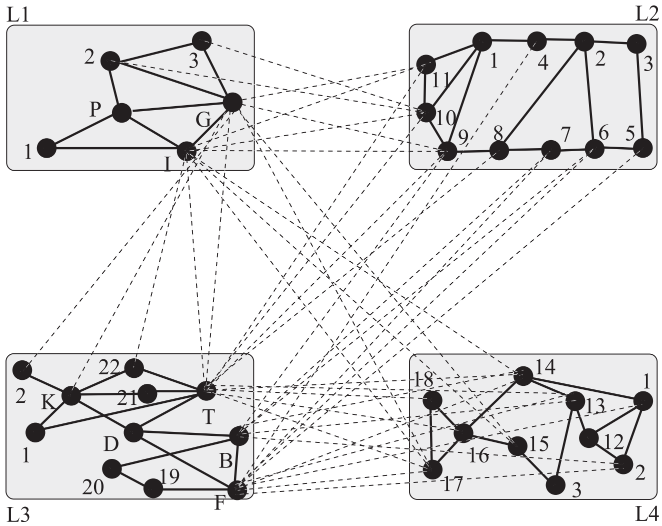

Fig. 1. A multilayer network consisting of four layers L1, L2, L3 and L4. Nodes with the same ID in different layers depict clones of the same node.

set of interconnections between nodes in separate graphs. Edges connecting nodes of a single graph are featured as *intra-edges*, whereas edges connecting nodes of different graphs are notated as *inter-edges*. Formally, we describe a multilayer network as $\mathcal{P}(\mathcal{G},\mathcal{E})$, where $\mathcal{G} = \{G_i; i = 1, 2, .., |L|\}$ is a set of graphs, i.e., the layers of $\mathcal{P}$, and $\mathcal{E} = \{E_{ij} \subseteq V_i \times V_j; i, j \in \{1, 2, \dots, |L|\}, i \neq j\}$ is the set of inter-edges between nodes of different layers, i.e., different graphs. Fig. 1 depicts a four layer multilayer complex network.

Multiplex networks are a special case of multilayer networks, where nodes are clones (counterparts) of themselves in each layer, i.e., $V_1 = V_2 = \cdots = V_{|L|} = V$. For multiplex networks the only inter-connections allowed are between a node and its counterparts in the remaining layers. Formally, $E_{ij} = \{(v, v);\ v \in V\}$ for all $i, j \in \{1, 2, \dots, N\}$ with $i \neq j$.

2.2 Diffusion in Multilayer Networks

Similar to other studies e.g., [54] we use the *Susceptible-Infectious-Recovered* (SIR) model, which models the penetration of a virus (information, product, rumor), and it has three possible states: a) *Susceptible* (*S*), b) *Infectious* (*I*) and c) *Recovered* (*R*):

- *Susceptible* (*S*) state, where a node is vulnerable to infection.
- *Infectious* (*I*) state, where a node tries to infect its susceptible neighbors, and succeeds with probability λ.
- *Recovered* (*R*) state, where a node has recovered and can no longer be infected.

A susceptible (S) node may be a user that is interested in certain information/product. Infectious (I) individuals are those who are already *influenced*, and try to "convince" their susceptible neighbors to follow the same action. Finally, recovered (R) nodes are those nodes who, e.g., have bought the product and can no longer be affected. The diffusion process ends when there are no nodes left in the I state. Hence, *influence* is measured by the number of nodes in the R state at the end of a diffusion process.

In multilayer networks the propagation is expected to diffuse over the different layers at different speeds, i.e., different λ per layer [6], [28]. However the different spreading rates within the various layers is not the only rate that we need to study. Spreading among different layers should also be taken into consideration. Thus, we experience intra-infection probabilities, i.e., infection rate in a single layer i (λ_{ii}), and inter-infection probabilities, i.e., infection rate from a node in layer i to its inter-connection in layer j (λ_{ij}). In multiplex networks nodes are clones in the different layers, hence for this special case $\lambda_{ij} = 1$. In our model, and without loss of generality [4], [24], [27] we assume that an infected source has a single chance to infect its susceptible neighbors, and immediately after it falls to the R state. This is the worst-case scenario to benchmark a method, since the longer a source node is infected the more probable to infect its neighbors. If we allow for (very) long infection periods, then the diffusion process will expand to very large parts of the network (or even to the whole network), irrespectively of the seeding method, the infection probability, the network topology, etc.

Therefore, the question to be answered is *which are those nodes, who if initially activated/incentivised, can trigger a cascade of new adoptions and maximize the spread*.

3 Proposed Methods to Identify Highly Influential Spreaders

Understanding influence in multilayer structures is significantly different from that of monoplex networks; agents (nodes) are subject to different environments which quite naturally have different rules, i.e., ways (paths) to spread information, different spreading rates, etc. Such characteristics introduce new challenges in the domain of influence ranking, and hence new techniques that incorporate those aspects are necessary. In [4] we introduced the *μ-Power Community Index* (*μ-PCI*) of a node, that combines the degree of the focal node with the degree of its direct neighbors. The intuition inferred from the understanding that a node in a dense neighborhood, in principle, can affect a large number of other nodes, i.e., exert strong influence. The proposed technique in addition to its local computation cost, successfully identified influential spreaders. Later in [38], it was proved that such connectivity results in the 'best' influential spreaders.

In our current work, we raise and answer the following question: *can we devise a locally-computed measure, that will characterize a node's vicinity, for their density in both, intra and inter connections?* We believe that identifying nodes with strong connectivity in many layers, will reveal potent entities linked to different connected environments, thus able to exert strong influence over the multilayer network. To put our interest into the test, we devise a number of measures that follow our main idea, and evaluate them in a number of real and semi-synthetic multilayer networks.

3.1 The Family of Multilayer PCI Measures

For the sake of article's self-completeness, we start with the definition of the original measure, i.e., (μ-*PCI*), and then give its multilayer generalizations.

Definition 1 (Power Community Index, μ-PCI [17]). *The μ-PCI index of a node v is the maximum number k, such that there are at least k neighbors of this node with degree larger than or equal to k in the μ-hop neighborhood of v.*

By setting $\mu = 1$, we get a restricted version of the algorithm, namely PCI. PCI coincides with the well-known h-index [22], and therefore μ-PCI generalizes the h-index for single layer networks. PCI is actually a centrality measure, and it was originally used for the purposes of cooperative caching in wireless ad hoc networks. Later in [4] it has been applied to the identification of influential spreaders; similarly, the h-index has been described as a centrality measure [8], [29] and used in the context of influentials [35], [43].

Next, we provide the generalization of PCI (and thus of the h-index) to multilayer networks.

Definition 2 (Minimal-layers PCI, mlPCI$_n$). *The $mlPCI_n$ index of a node v is the maximum number k, such that there are at least k direct neighbors of v with the number of links* towards at least *n layers greater than or equal to k.*

From Fig. 1 with node D as an example: $mlPCI_1(D) = mlPCI_2(D) = mlPCI_3(D) = 3$ and $mlPCI_4(D) = 0$. To combine the distinct n values of $mlPCI_n$ into a single dimension, we propose a simple aggregation. In particular, for a node v we define $mlPCI(v)$ as follows:

$$mlPCI(v) = \sum_n mlPCI_n(v). \quad (1)$$

$mlPCI$ by definition bares no strict limitation with regard to either limited, or large number of layers. The indicator will handle cases where nodes are well connected to all layers, to a few or even just one layer accordingly, which indicates the dynamics of Definition 2. According to $mlPCI$ index, nodes well connected in many layers, i.e., nodes assigned high index scores in the range of the n values, will be better "rewarded" from nodes that are well connected, but, in fewer layers. With this understanding we believe that $mlPCI$ will be a good indicator for the spreading potential of nodes.

Simple aggregation can be considered as a baseline method to combine the different values of $mlPCI_n$. However, since larger n implies connection to more layers, a scaling factor could be used with respect to n in order to handle the vector elements differently. Nonetheless, to devise an appropriate method for handling those values is no trivial task. Several factors need to be taken into consideration and further combined with respect to potentially different characteristics introduced by the different layers, e.g., number of nodes, connectivity, global clustering coefficient, etc. Such characteristics can introduce a different view to $mlPCI_n$ and provide a different ranking for the network nodes. In this article we focus on the simple aggregation introduced in Equation 1, i.e., agnostic to layer characteristics.

Next, we present a set of special cases of Definition 2.

- *Layer-agnostic PCI* ($laPCI$). By ignoring layer information (i.e., ignoring n) in Definition 2, we get a special case of $mlPCI_n$ which we call *Layer-agnostic PCI*, $laPCI$. In Fig. 1, and considering node D as our focal node, the neighbors that contribute to its $laPCI$ index are nodes K, B, F and T with a total of 6, 9, 12 and 16 links respectively in the different layers. Thus we have four neighbors each of which has at least as many links to the different layers, i.e., $laPCI(D) = 4$. $laPCI$ gives credit to a node whose neighbors have many connections in different layer(s), however, it makes no distinction on how those connections are distributed over those layers. This implies that a node may accumulate a large $laPCI$ value by being well connected in a few layers, and at the same time sparsely connected (or even disconnected) to the remaining ones.
- *All-layers PCI* ($alPCI$). We obtain another special case of $mlPCI_n$ by setting n in Definition 2 equal to the number of layers; we call this special case as the *All-layers PCI*, $alPCI$. This approach demands that the neighbors of the focal node have at least k neighbors in *all* layers. Considering node P of Fig. 1, the neighbors that contribute to its $alPCI$ are nodes G and I each of which has at least two links in all layers, thus $alPCI(P) = 2$. $alPCI$ will detect nodes strongly connected to all layers of the multilayer network that we believe is key ingredient for highlighting the most efficient intra- and inter-layer spreaders. However, this measure will be very restrictive for nodes that lack interconnectivity towards all layers. This may be a problem for multilayer networks composed of many layers, where it would be quite difficult to detect many nodes with particularly high *alPCI* index.
- *Layer-symmetric PCI* ($lsPCI$). Finally, by setting $n = k =$ '$numberoflayers$' in Definition 2 we get the so-called *Layer-symmetric PCI*, $lsPCI$. This measure is a combination of three aspects: (a) the inter- and intra-degree of the focal node, (b) the inter and intra-connections of its inter- and intra-neighbors, and (c) the layers; all these are nicely "condensed" into a single value. $lsPCI$ alleviates the strictness of $alPCI$: "to all layers" no longer applies, and can be quite effective when dealing with a large number of layers. For limited number of layers we expect $lsPCI$ to act complementary to other methods, since nodes will be ranked from a limited range of values. In Fig. 1, for node D it applies that $lsPCI(D) = 3$, since nodes B, F and T have at least three links in three layers.

Although we have presented our definitions for undirected networks, their implementation to directed ones is straightforward, i.e., by matching the k attribute to the out-degree of each respective node.

In the next section, we conduct an experimental evaluation of the proposed family of measures providing detailed information about the competitors, the datasets, and the performance measures.

4 Evaluation Settings

4.1 Competitors for Multiplex Networks

Additive PageRank for multiplex networks (addPR). PageRank [31] has been used several times for the identification of influential spreaders [44]. In [20] the original PageRank algorithm is extended for multiplex networks requiring though a "predefined" ordering of the layers. We examine here the so-called *additive Multiplex PageRank*, in which the effect of layer i on layer j is exerted by 'adding' some value to the centrality the nodes have in layer j in proportion to the centrality they have in layer i. Since the authors do not

TABLE 2
A Summary of Competing Methods Evaluated

Multiplex networks	Multilayer networks
aggDeg ≡ aggDeg	
addPR [20]	aggPR [31]
verPR [15]	verPR [15]
verBC [15]	verBC [19]
sumCore [this article]	aggCore [27]
Core [13]	Core [13]

TABLE 3
Multiplex Networks

Networks	N	E	|L|	Type	Nature
Sacchpomb	875	18214	3	Directed	Biological
Drosophila	1364	7267	2	Directed	Biological
Sacchcere	3096	185849	5	Directed	Biological
Homo	3859	77483	3	Directed	Biological
NYClimateMarch	4150	45334	3	Directed	Twitter
MoscowAthletics	4370	33411	3	Directed	Twitter

provide layer ordering methodology, we order layers in decreasing order of their largest eigenvalue. Our choice is driven with respect to the fact that a larger eigenvalue implies faster information dissemination.

Versatility PageRank (verPR) and Versatility Betweenness Centrality (verBC). A fundamentally different flavor in extending PageRank for multiplex networks has been described in [15], which, using a tensorial notation, provides a generalization of the original PageRank for multiplex networks, called the Versatility PageRank. Counting the number of shortest paths that pass through a node (i.e., Betweenness centrality) has been widely used as a competing technique for ranking the influence potential of nodes. In [15] the authors generalize this concept for multiplex networks, describing the Versatility Betweenness. Both techniques are implemented as competitors.

Multiplex k-core percolation methods (Core and sumCore). We include the k-core percolation for multiplex structures [3], [13] in the competitors lists (*Core*). However, in the evaluation we found only limited values for *Core*. This is due to the fact that *Core* will follow the coreness of a node's least connected edge type, regardless of how well connected a node may be in the remaining layers. Thus, we also include a variation of *Core* according to which we calculate the shells for each layer separately and then add those values; we name this version as the *sumCore* index.

Degree centrality for multiplex networks (aggDeg). We employ a straightforward interpretation of degree centrality for multiplex networks, i.e., the aggregation of the intra neighbors of the focal node in all layers; we call it *aggDeg*.

4.2 Competitors for Multilayer Networks

The work presented in [11] proposes a generalization of the k-core algorithm that incorporates λ_{ii} and λ_{ij} within the definition of the technique. However, this is not a characteristic that any method should "know" a priori, and hence, we exclude this method from our list. Also, due to the unique characteristic of multiplex networks, i.e., nodes are clones in the different layers, the Additive PageRank (addPR), presented in the previous section, cannot be applied here. Though, we tested the Versatility PageRank and Versatility Betweenness proposed in [15], and *Core* from [13]. Moreover, in order to provide a complete analysis, we apply the 'traditional' methods, i.e., PageRank, Betweenness centrality, Degree centrality, and k-core by projecting the multilayer network in its aggregated form, implementing in essence the proposals in [2].

4.3 Summary of Competitors

Table 2 summarizes the competitors implemented in this article. Each method's name is comprised by two parts; the latter part discloses the method, e.g., PR stands for 'PageRank', BC stands for 'Betweenness Centrality', Core for 'k-core', 'Deg' for 'Degree', whereas the former part describes the 'flavor' of the method, e.g., 'vers' stands for 'versatility', 'add' stands for 'additive', 'agg' stands for 'aggregated' (i.e., in the aggregated network), 'sum' stands for 'summation' (i.e., summation of values resulting from the calculation of a measure in the different layers).

4.4 Datasets

For the evaluation of the competing methods we used several real and synthetic datasets to compare the algorithms in diverse networked environments.

4.4.1 Real Datasets

Table 3 depicts the basic attributes of the experimented multiplex networks. For more details, readers are referred to: http://deim.urv.cat/~manlio.dedomenico/data.php. We extracted part of the original networks in such a way that all nodes have counterparts in all layers.

4.4.2 Semi-Synthetic Datasets

For synthesizing artificial networks we follow a similar approach with the authors of [11]. Specifically, we consider real monoplex networks from [33], e.g., several Internet peer-to-peer networks, and synthesize their interconnectivity. Table 4 illustrates the real networks used as the different layers of the synthesized multilayer networks. *EgV* corresponds to the largest eigenvalue of each respective network. We generated two types of multilayer networks: (i) a multilayer network composed of layers with similar size, i.e., *Similar Layers Network* (SLN) and (ii) a multilayer network formed of different-sized layers, i.e., *Different Layers Network* (DLN).

The multilayer network of the first type is composed of the networks/layers (3)–(6) (4 similar-sized layers), whereas the second multilayer network is composed of the networks/layers (1)–(3) (i.e., 3 different layers). For the latter case the different networks differ in the number nodes, edges and network type. We present plots about the out-degree

TABLE 4
Layers of Semi-Synthetic Networks

No.	Network	Nodes	Edges	Type	EgV
1.	wiki-Vote	7,115	103,689	social	45.1
2.	cit-HepTh	27,770	352,807	citation	10.8
3.	p2p-Gnutella04	10,876	39,994	p2p	4.4
4.	p2p-Gnutella05	8,846	31,839	p2p	4.3
5.	p2p-Gnutella06	8,717	31,525	p2p	4.7
6.	p2p-Gnutella08	6,301	20,777	p2p	5.1

TABLE 5
Stability of Ranking with Respect to the Average Spreading Power

	Homo			Sacchpomb		
	avg-std	avg	avg+std	avg-std	avg	avg+std
aggDeg	0.9839	0.9859	0.9879	0.9899	0.9869	0.9887
sumCore	0.9162	0.9142	0.9112	0.9781	0.9804	0.9806
verBC	0.7020	0.7013	0.7011	0.8020	0.7972	0.8015
addPR	0.8494	0.8457	0.8421	0.8590	0.8649	0.8602
verPR	0.8495	0.8560	0.8529	0.9498	0.9501	0.9530
Core	0.7713	0.7725	0.7717	0.4363	0.4394	0.4340

The values represent the ratio between the correlation (τ) of a competitor, and the best performing method (i.e., mlPCI).

distribution of these networks in the 'Network properties' section of the Appendix, which can be found on the Computer Society Digital Library at http://doi.ieeecomputersociety.org/10.1109/TNSE.2017.2775152.

4.4.3 Generating Interconnections

Since we make use of real networks to represent the layers of the semi-synthetic multilayer structure, we have to decide how to generate the interconnections among layers. We developed a synthetic multilayer network generator which satisfies the following three needs:

- It can define how many interlinks, i.e., inter-neighbors, a node may have.
- It can define how those links are distributed over the layers.
- It can define how links are distributed in each specific layer.

We apply the Zipfian distribution in our interconnectivity generator. The desired skewness is managed by the parameter $s \in (0, 1)$. The generator uses one Zipfian distribution per parameter of interest:

- $s_{degree} \in (0, 1)$ in order to generate the frequency of appearance of highly interconnected nodes.
- $s_{layer} \in (0, 1)$ in order to choose how frequently a specific layer is selected.
- $s_{node} \in (0, 1)$ in order to choose how frequently a specific node is selected in a specific layer.

Finally, we need to decide the range of values for the different distributions. For s_{layer} and s_{node} the selection is straightforward since all layers and all nodes within a layer must be available options. Note that the different layers are allowed to have different preferences, i.e., skewness towards different network-layers. Following the review of [28] we understand that inter-connections are rarer than the intra-connections. In our simulations, we limit the inter-degree of nodes within (0, $d \cdot log_2 \sum_i V_i$) for all $i = 1, 2, ..N$ layers where $d = 1, 2, 3$ or 4. Hereafter we apply the notation SLN$_d$(s_{degree}, s_{layer}, s_{node}) in order to refer to the generated networks. More algorithmic details and a brief validation of the generator can be found in the Appendix available online of this paper.

4.5 How to Evaluate the Performance

In our experimentation, in order to evaluate the ranking ability of each competitor, we calculated the correlation of the competitors with respect to the spreading power (SP) of each node (i.e., the number of nodes influenced), when initiating the SIR process from this node as the single origin of the diffusion process. The correlation is measured through Kendall's Tau (τ) "b" rank correlation coefficient [26]; the τ value between two equi-sized ranked lists is computed as follows:

$$\tau = \frac{n_c - n_d}{n(n-1)/2}, \tag{2}$$

where n_c is the number of concordant pairs, n_d is the number of discordant pairs, and the denominator is the total number of pairs of n items in the lists. Some more details are provided in the Appendix available online. In order to obtain unbiased results, for each node, the average SP is used over 500 SIR processes.

We found that the average is a proper representative for the following reason: we evaluated the ranking ability of the competitors with respect to the standard deviation of the distribution around the average spreading power of each node. In more detail, all competitors were ranked with respect to: (i) the average spreading power (SP), (ii) the average spreading power minus the standard deviation ($SP - std$), and (iii) the average spreading power plus the standard deviation ($SP + std$), when λ_{ii} is the epidemic probability. Hence for each competitor we obtained three values of τ. We found out that these values differ from each other beyond their third decimal point, as shown in Table 5, where each cell's value is the ratio between the correlation (τ) of a competitor, e.g., Deg, and the technique which scored the largest τ (i.e., $mlPCI$), when using the respective values of SP for two networks, namely Homo and Sacchpomb. Similar results were observed in the remaining networks, and thus, we draw the correlation of each competitor against the average spreading power.

TABLE 6
Experimentation Parameters

Network Type	Rate	Range	Default
Multiplex	$\frac{\lambda_{ii}-\lambda_c}{\lambda_c}$	−0.2 to 0.6	0
Multilayer	$\frac{\lambda_{ii}-\lambda_c}{\lambda_c}$	−0.2 to 0.2	0
	$\frac{\lambda_{ij}-\lambda_c}{\lambda_c}$	−0.3 to 0.3	0
	d	1 to 4	2

4.6 Setting Parameters

Table 6 illustrates an overview of the experimented parameters, range and default values. In our evaluation in multiplex networks we illustrate how the different spreading rates per layer (λ_{ii}) affect the competing methods. Specifically, we compute the epidemic probability λ_c [46] for each layer, and experiment around this value. For example in Fig. 2a, zero in the x-axis sets λ_{ii} of all layers at their respective epidemic thresholds, while -0.2 sets the spreading rate per layer at 20 percent bellow that value etc. Similar notations are used for the semi-synthetic networks, where we also investigate on the impact of the inter spreading rate (λ_{ij}) and on the density of the generated interconnections (d). To decide the spreading rate between the different layers, we calculate the epidemic threshold of the

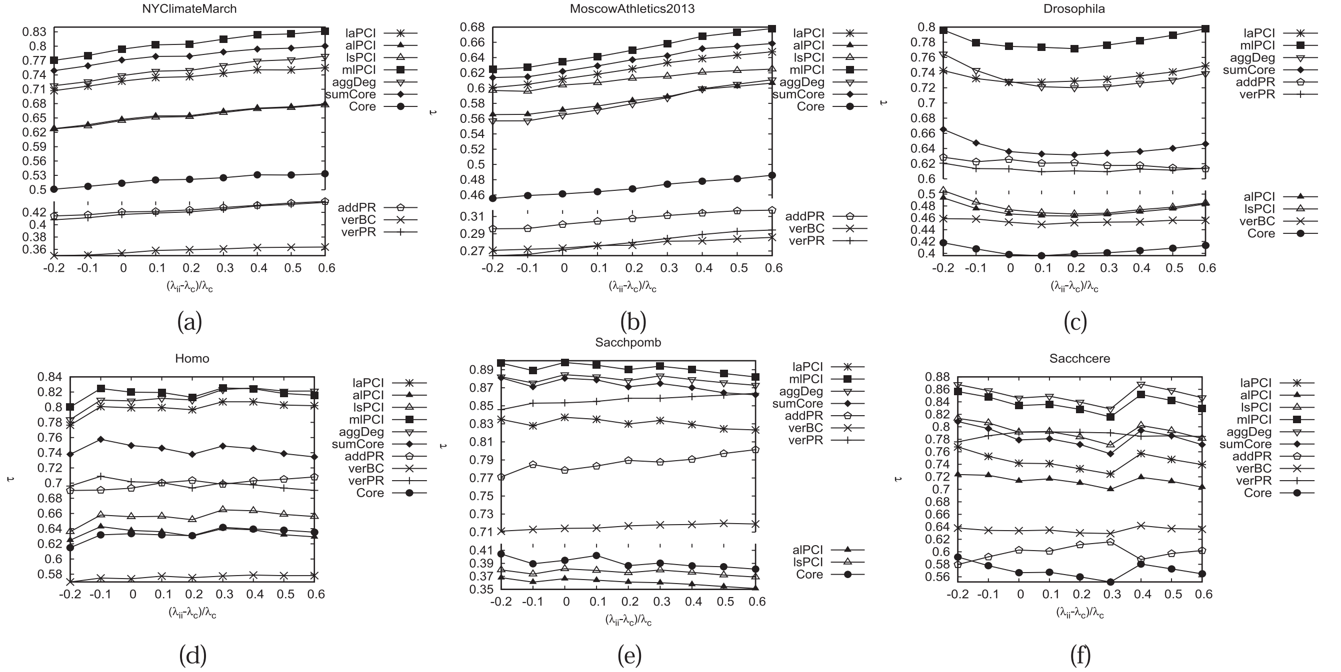

Fig. 2. Rankings capabilities (Kendall's Tau b) of all competing techniques in real multiplex networks with respect to λ_{ii}. It can be observed that all competing algorithms exhibit similar trends, i.e., either increasing or decreasing trend as the intra-spreading probability changes. $mlPCI$ illustrates the largest correlation with influence in almost all networks. While $mlPCI$ shows a relatively stable behavior, i.e., it is (almost) always at the top of the ranking chain, the remaining algorithms do not posses that property as their rank changes in the different networks, e.g., $aggDeg$ is 2nd in Homo and 6th in MoscowAthletics2013.

aggregated network and likewise experiment around this value. We choose to use the same λ_{ij} among all layers in order to give the same "weight" to all interconnections. When evaluating the impact of one parameter, the remaining parameters are set to their default values.

5 Results

5.1 Ranking Influence in Real Networks

In this section we investigate on the performance of the competing techniques in multiplex networks. For our first and most evident observation we elect $mlPCI$ as the most promising technique for the identification of influential spreaders. As illustrated in Fig. 2, $mlPCI$ has the strongest correlation with influence in almost all evaluated scenarios, that is, the largest τ. By combining the connectivity that neighboring nodes posses in the different layers as $mlPCI$ suggests, from just one, to all layers of the multiplex network, we show that the proposed algorithm can take advantage of multiplexity more efficiently than the competing techniques.

In plots (d-f) of Fig. 2, $aggDeg$ performs similarly to $mlPCI$, whereas their in-between performance deviates in (a), (b), and (c). Its worst performance is illustrated in Fig. 2b where the competitor's correlation with influence falls to the fifth place. $aggDeg$ "sees" the network in its aggregated form, i.e., as a monoplex network, and hence disregards a wealth of knowledge regarding the different layers. For instance a node which accumulated most of its $aggDeg$ value from a single layer, is not distinguished from a node of the same index but equally connected to all layers. Nonetheless, these nodes will have different spreading potential. Moreover, although a node with many connections can be an influential one, it is also a misleading characteristic if the node is positioned in the periphery of the network. This claim has been proven for monoplex networks [27], and it was expected to apply in multiplex structures as well.

Focusing on $alPCI$ we observe varying results, i.e., medium performance, as in Fig. 2a or Fig. 2b, or low correlation with influence as illustrated in Fig. 2c or Fig. 2e. At this point we should reminisce that $alPCI$ is a very strict definition which demands connectivity to all layers. Although in terms of spreading capability such characteristic would prove invaluable, in our simulations we found relatively low values for $alPCI$. Fig. 3 illustrates the distribution of $alPCI$ values in the evaluated networks. It can be observed that when we are bound to a poor distribution, i.e., when nodes are not strongly connected to all layers as in Drosophila network (Fig. 2c), we obtain the worst case performance for $alPCI$. Contrary, when nodes are better connected to all

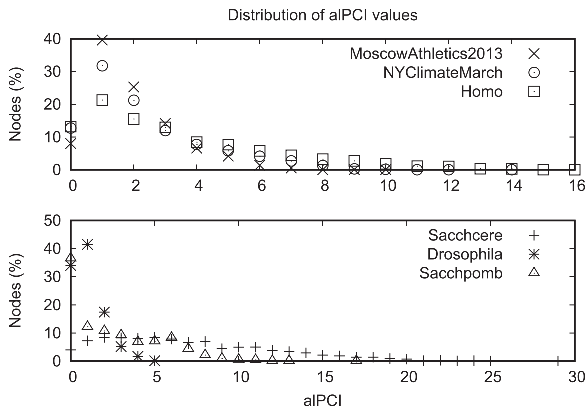

Fig. 3. Distribution of $alPCI$ values for all networks. It can be observed that for most networks the majority of nodes has relatively low $alPCI$ values, whereas the largest indexes are appointed to only a few nodes.

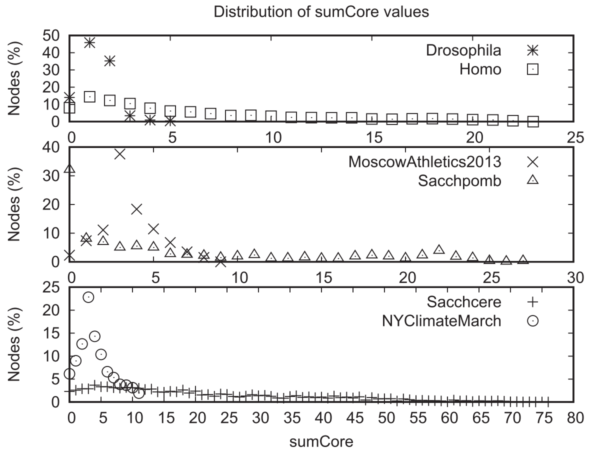

Fig. 4. Distribution of $sumCore$ values for all networks. According to the illustrated distributions, we observe two groups: (Drosophila, MoscowAthletics2013, NYClimateMarch) and (Homo, Sacchpomb, Sacchcere).

layers, the correlation of $alPCI$ with influence increases, e.g., as in the Sacchcere network (Fig. 2f). Of particular importance are Drosophila and Sacchpomb networks where we observe a large portion of network nodes with zero $alPCI$ index. These are the cases where several nodes act only as receivers (not spreading) in a layer, i.e., zero outdegree. Such instances can be related to lurking behaviors in social networks where nodes only "hear" but never spread information [50]. However, $alPCI$ requires spreaders to all layers, hence, by definition these nodes will be "overlooked". Although the above cases contribute negatively in the evaluation of the proposed mechanism, our results show that finding nodes strongly connected to (as) many layers (as possible) is a key factor for the identification of influential spreaders. For $lsPCI$ we also observed variation to its performance. This is due to the relatively low number of layers evaluated (2, 3 or 5), and thus limited range of indexes obtained for ranking the multilayer nodes.

Moving to the evaluation of $sumCore$, we observe that the competitor is ranked second in (a) and (b) of Fig. 2, whereas in Fig. 2e it competes with $aggDeg$ for the second place. However, in the remaining networks the competitor performs differently. From Fig. 4 it can be observed that the largest $sumCore$ values for the Twitter networks are about 10, that is, a large number of nodes distinguished for their influence capabilities from a mere of ten different values (ties are solved via largest $aggDeg$). Although this is a shortcoming shared also by $alPCI$, from Fig. 4 it can be concluded that as we obtain a better distribution for the $sumCore$ values, that is when nodes are ranked more from their $sumCore$ index than their $aggDeg$, the competitor's performance drops, as it is ranked fourth or lower in our simulations e.g., Fig. 2d or Fig. 2f. However, this is an opposite behavior from what we observed for $alPCI$, thus, $sumCore$ cannot be considered a strong indicator for the spreading potential of a node. Furthermore, a relatively poor performance can be observed for $Core$, which can be explained by the fact that the competitor follows the coreness of a node's least connected edge type, regardless of how well connected this node might be in the remaining layers. This characteristic has a negative impact in performance of the technique.

$verPR$ shows an interesting performance. In Figs. 2d, 2e, 2f the technique illustrates a very competitive behavior, i.e., is ranked as 3^{rd} or 4th best method in the ranking chain of the competitors; however, in Figs. 2a, 2b its performance drops. This observation can be attributed to the change in the distribution of in-out neighbors; when these quantities are positively correlated (see Fig. 12 in the Appendix available online), then $verPR$ exhibits very good performance. When compared to the $addPR$, $verPR$'s performance is either similar or significantly higher, e.g., Figs. 2d and 2f respectively. This observation concludes that $verPR$ can identify more effective spreaders than $addPR$.

By definition $addPR$ instructs an ordering of layers where a node gains more centrality in a layer if it is important in previous ones, regardless of the node's ability to attract important nodes in the current layer. Although such attribute can be beneficial for a node when it lacks centrality in a layer, but, is well connected in others, it is also a very restrictive characteristic that requires an optimal selection for the sequence of layers, i.e., the order that layers are being processed, overall, should be beneficial to all nodes of a network. Nonetheless, the decision for such ordering is no trivial task especially as the size (in nodes) and the number of layers increases. But apart from this shortcoming, its relative low performance is explained by the nature of the original PageRank when used for influential detection, which assumes that content spreads randomly in the network that is not valid [44].

$verBC$ inherits the weaknesses of the original betweenness algorithm. As an example, consider a node which is unique for reaching a portion of network nodes in a certain area. Clearly, that node will be part of many shortest paths, hence, it will accumulate a large $verBC$ score. However, if spreading in this area is unfavorable, e.g., nodes are sparsely connected, or the target area reached by this unique node is relatively small, the spreading power of that node will not justify its high $verBC$ score in the ranking process. On the other hand nodes that do not reside in any shortest path will acquire a zero index of $verBC$. Nonetheless such nodes may be (directly) connected to hubs, and thus "indirectly" affect a significant number of network nodes. It is straightforward that in such occasions the performance of the competitor will be negatively affected.

Evidently, the competing algorithms will not be equally influenced from network characteristics, i.e., methods that require global knowledge of the network topology are more depended to network topology than local approaches. For instance, by definition, $verPR$, $addPR$ and $verBC$ will be significantly more influenced than the rest of the competing techniques from the distribution of in-out degree (it is illustrated in Fig. 12 in the Appendix available online), especially when a large number of nodes with low values in either k_{in} or k_{out} are present. To our understanding such characteristics also contribute to their overall significantly lower performance. This is yet another reason for selecting methods that require only local knowledge of the network topology.

Examining the curves of the illustrated results, we observe similar trends for the competing methods, i.e., either increasing or decreasing within a specific range of λ_{ii} values. The observed abrupt changes in τ, as illustrated for example in Fig. 2f for the Sacchcere network from 0.3 to 0.4,

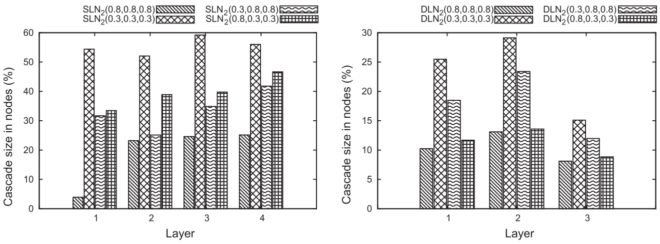

Fig. 5. Maximum cascade size per layer subject to the distribution of interconnections. It can be observed that when all parameters are set to 0.3 the cascade size is maximum, while the opposite occurs, when all parameters are set to 0.8.

or in Fig. 2d, is due to a significant amount of newly influenced nodes with respect to those from the previous λ_{ii} value. In contrast to monoplex networks where spreading is of single dimension, in multiplex networks a node can become influenced because it's counterpart was "reached" in another layer. In other words, although there is an influence rate λ_{ii} per layer, the actual spreading rate can be significantly higher when accounting for multiplexity.

Our evaluation so far strengthens our belief for finding influential spreaders in multilayer networks, by imprinting within the proposed measures the density of inter- and intra-connections in the immediate vicinity of the focal node. $mlPCI$, combines those k neighbors connected in just one layer, those k neighbors residing in two layers and so on up to those k neighbors connected to all layers. It alleviates the shortcoming introduced by $alPCI$ and at the same time can be as restrictive as an application requires by setting our focus to at least as many layers as necessary. In addition to its local computation complexity, $mlPCI$ illustrated the largest correlation with influence in almost all evaluated networks for all respective spreading rates, and is thus our primary selection.

5.2 Ranking Influence in Semi-Synthetic Networks

5.2.1 Interconnections and Influenced Nodes

We start our evaluation by noting the different "rules" that apply for these type of networks with respect to the multiplex structures. First, there are no counterpart nodes, i.e., nodes are different entities, which means that there exists a spreading probability in order to reach nodes in other layers, i.e., λ_{ij}. Furthermore, successfully propagating over an inter-link, only affects one node at one specific layer and not all layers of the multilayer network as in the previous evaluation. The above considerations indicate that we are bound to a significantly different environment, hence, we expect to encounter different results.

First, we examine the effect of the generated interconnections in the diffusion process. We should note that although our generator gives a particular trend on how interconnections are distributed over the layers, the topological characteristics of an inter-neighbor will also play a vital role in the diffusion process. Specifically, an interconnection to a node which resides within a well connected neighborhood will favor the spreading process, whereas the opposite will occur if interlinks are "wasted" over nodes with poor inter/intra connectivity. Fig. 5 illustrates the cascade size per layer in several networks, i.e., the influence exerted by any initially infected node falls within the illustrated range. It can be observed that the way in which interconnections are distributed over the layers plays a major part in the SIR dynamics; as anticipated for SLN_2(0.3,0.3,0.3) and DLN_2 (0.3,0.3,0.3) networks, the cascade size is significantly higher. This is due to the fact that there is no excessive skewness for the inter-degree assigned to the participating nodes (s_{degree}), nor towards which layer those interconnections are guided (s_{layer}), or to the selection of nodes within the target layer (s_{node}). Such configuration will provide a favorable environment for the spreading process, and thus influence a larger portion of network nodes. The opposite scenario is illustrated for SLN_2(0.8,0.8,0.8) and DLN_2 (0.8,0.8,0.8). Similarly, having similar distribution for the inter-degree of nodes, e.g., by setting s_{degree} at 0.3 (or 0.8), and vary in the remaining parameters, shows that increased skewness has negative effect on the percentage of influenced nodes.

5.2.2 Impact of Inter Connections and Intra Diffusion Probability

Fig. 6 illustrates the performance of the competitors in the semi-synthesized networks when evaluating the impact of λ_{ii}. In coherence with our conclusions in real networks, we elect $mlPCI$ as the most promising technique for measuring influence in multilayer networks. It can be observed that $mlPCI$ is at the higher values of τ for almost all spreading rates, however, the ordering for the remaining techniques has changed. Specifically, $laPCI$ can be considered as the second best method, performing almost as good as $mlPCI$ in Figs. 6a, 6g or 6h, and as the next best solution in the remaining networks. $laPCI$ implies k neighbors towards any layers, however, these nodes may reside in many, or, in just one layer. For occasions where the latter holds, and nodes are assigned a large $laPCI$ index, there is strong possibility that an epidemic will arise in the multilayer network, since within these k neighbors, nodes connected to different layers are likely to exist. The same logic applies to nodes with a large $aggDeg$ index as for example in the Wiki-Vote network (details in Fig. 13 in the Appendix available online). The difference between the two measures that discriminates the performance of $laPCI$, is that those k neighbors that form the node's index, is the result of "filtering" that is applied in the focal node's vicinity, that discriminate a highly connected node within a strongly connected neighborhood, from nodes residing in sparser vicinities. This inherent characteristic governs all proposed methods, which in our view enables the proposed techniques to detect more efficient spreaders.

Of particular importance is the performance of $verPR$ in the SLN networks. Apart from the fact that it has increased correlation with influence with respect to its performance in the DLN networks, Fig. 6b illustrates an interesting result, i.e., $verPR$ outperforms $mlPCI$ when λ_{ii} is larger than the epidemic probability. To explain this behavior we need to consider the distribution of the inter-connections. By setting s_{node} at 0.8, we "send" many interconnections to a certain portion of network nodes within the corresponding layer, that is, in terms of $verPR$, specific nodes are inter-pointed by many others. These nodes will accumulate a large $verPR$ index due to their interconnections, thus rendered as efficient cross-layer spreaders detected by $verPR$. It is due to this intrinsic characteristic of the competitor that we observe its efficient ranking

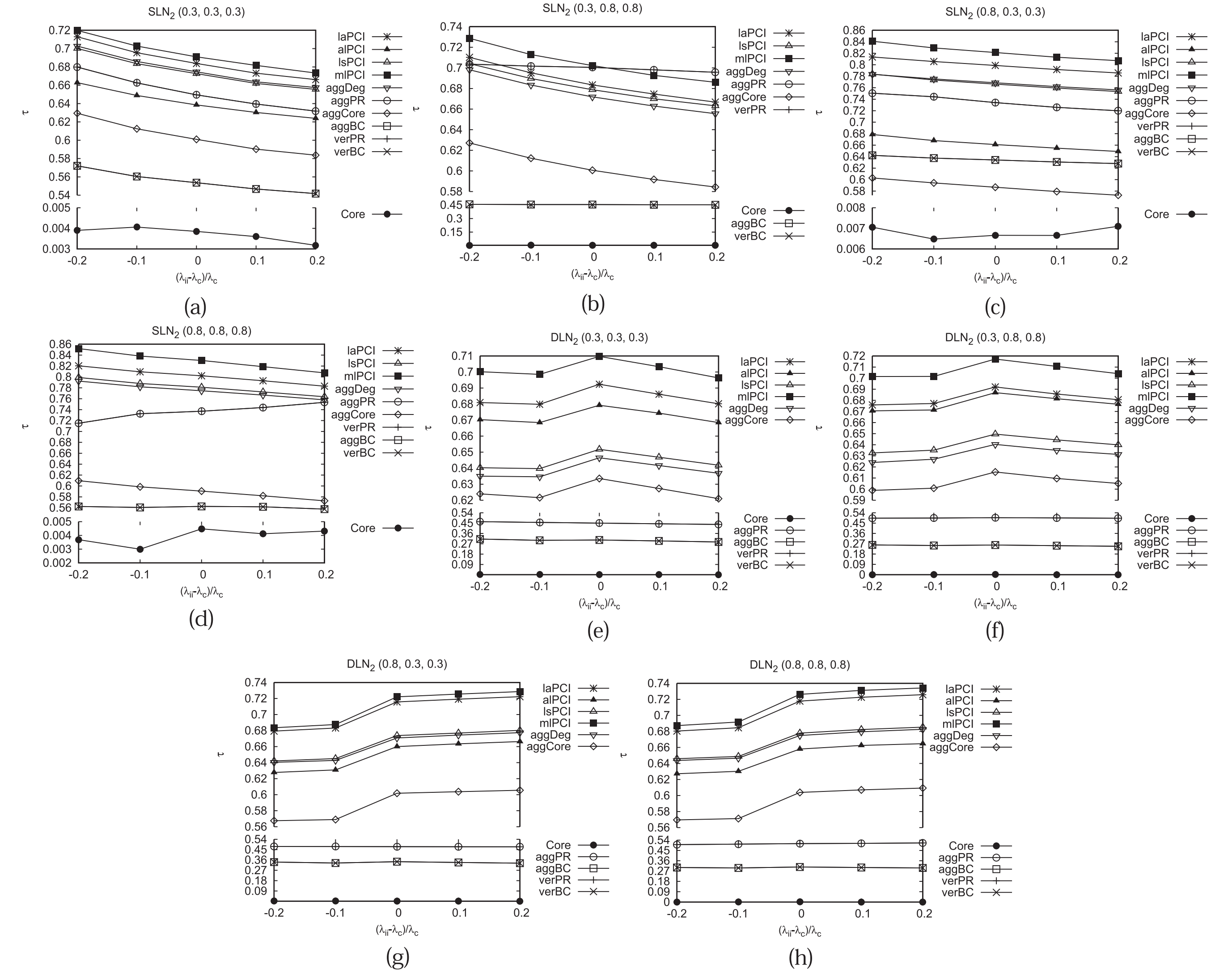

Fig. 6. Rankings capabilities (Kendall's Tau b) of all competing techniques in real networks with synthesized interconnections with respect to λ_{ii}. In coherence to our results in multiplex networks, $mlPCI$ illustrates the largest τ in almost all evaluated scenarios, whereas $laPCI$ can be considered as the 2^{nd} best performing method. As also depicted in Fig. 2, the competitors illustrate similar trends in performance, i.e., increasing or decreasing when λ_{ii} changes. Of particular importance is the performance of $verPR$ in (b) and (d), where due to the distribution of interconnections the competitor shows increased correlation with influence. $Core$ is almost uncorrelated with influence in these networks, because it assigns to almost all network nodes the same index value.

in these specific networks. In Fig. 6d, where interconnections are more constrained with s_{degree} set at 0.8, $verPR$ does not outperform the proposed methods, however, as λ_{ii} increases the distance in their performance decreases. In the DLN networks the performance of $verPR$ is far inferior to all methods with the exception of $verBC$. This gap in performance from the demonstrated results in Figs. 6a, 6b, 6c, and 6d, can be explained by comparing the k_{out} distribution of inter and intra links in each respective network type, i.e., in the latter, there is significant difference in magnitude between the inter and intra neighbors. Evidently from Figs. 11, 12, and 13 (see the Appendix available online), the impact of interconnections in the DLN examples will be considerably smoother, which explains the behavior of the competitor.

In all the experiments concerning multilayer networks, $Core$ seems (almost) uncorrelated to the spreading power of nodes (i.e., almost zero τ). This behavior is explained directly by the definition of the algorithm; nodes would get a $Core$ value different than one, only if they have connections to all layers. This happens only for very few cases in our generated networks, and thus practically all nodes get the same $Core$ value. This results in the phenomenon that we observe. By examining the performance of $aggCore$ we observe varying results, i.e., below the 5th place in the ranking chain of the competitors, e.g., 6th in Fig. 6e and 10th in Fig. 6c. Nonetheless we cannot expect $aggCore$ to be a challenging competitor since it projects all layers in a single dimension and thus neglects the layered structure of the network.

For $verBC$ it is straightforward that the shortcomings discussed in the previous section also apply in the current framework. Generally, when there are fewer paths to the different layers ($s_{degree} = 0.8$), the limited shortest paths work in favor of the competitor that shows a relative increase in performance, e.g., comparing Figs. 6a, 6b, and 6c. However, if either s_{node} or s_{layer} is set to 0.8 we observe decrease in τ as illustrated from Figs. 6a to 6b. It can be concluded that we cannot accurately distinguish the spreading power of nodes by counting

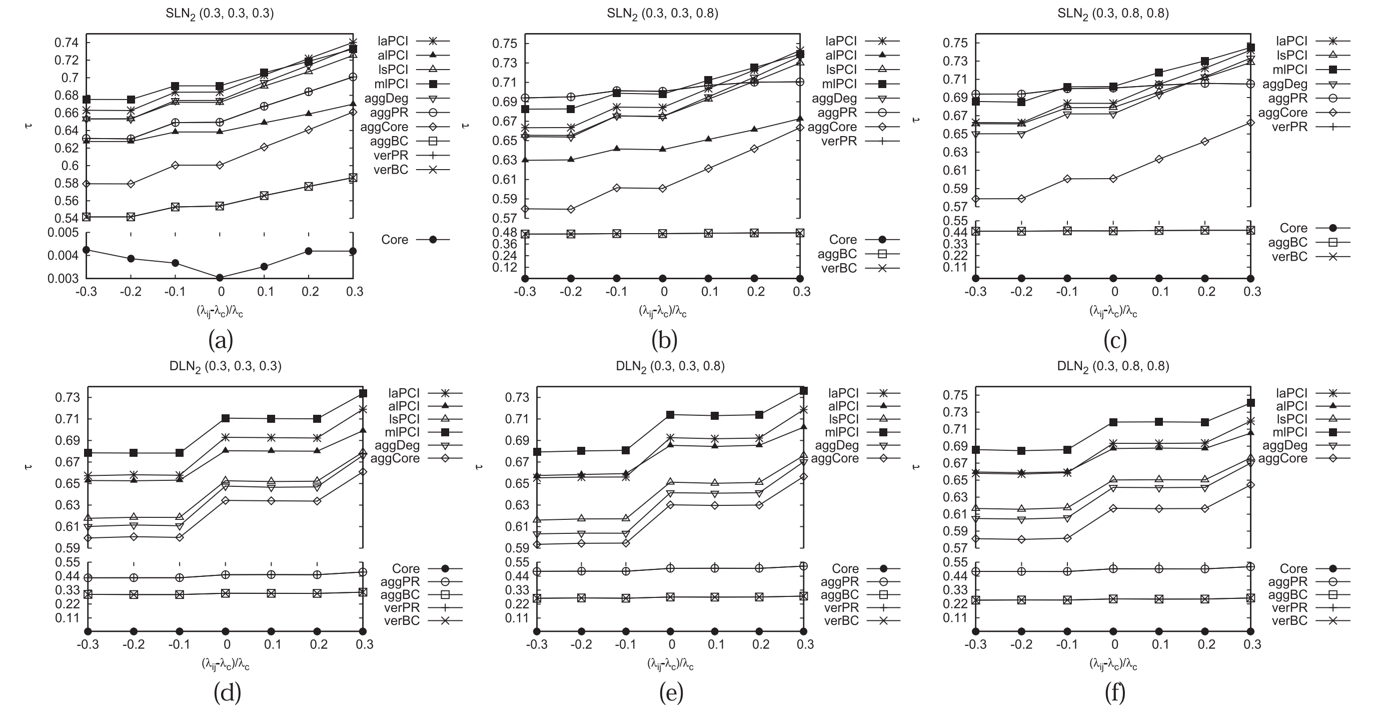

Fig. 7. Rankings capabilities (Kendall's Tau b) of all competing techniques in real networks with synthesized interconnections with respect to λ_{ij}. $mlPCI$ remains at the top of the ranking chain. $verPR$'s performance is better in the SLN networks where interconnections are more dense (when compared to the intra-connections) with respect to the DLN networks, and particularly is at its best when s_{node} or s_{layer} is 0.8. It can be observed that measuring the influence capabilities of a node by counting the number of geodesics that pass through that node ($aggBC$, $verBC$) does not yield competitive results.

the number of shortest paths that pass through them. As described in [15], the performance of $aggPR$ ($aggBC$) coincides with that of $verPR$ ($verBC$).

Setting s_{layer} at 0.8, denotes a possible preference developed between the layers, that is, most interconnections are "guided" from a layer to a specific other(s), while the remaining layers acquire limited inter-links from that particular layer. As illustrated from the results, this parameter has a soft impact to all competitors with the exception of $alPCI$. Particularly in Figs. 6b, 6c, and 6d, due to this setting, $alPCI$ failed to provide an acceptable ranking, since a significant portion of network nodes were assigned zero $alPCI$, or in other words nodes were not inter-linked to *all* layers. In the DLN networks, although less nodes where assigned a zero value, still, the obtained indexes where significantly low and overlapping. For example in DLN_2(0.8,0.8,0.8), most $alPCI$ values were below 6. In these scenarios we can understand the reasons for its questionable performance, however, $alPCI$ can still operate in one more way, i.e., as an additive rank rather than a solo ranking method. This aspect can be related to Figs. 6f, 6g, and 6h where a limited range of $alPCI$ values (ties are solved via the largest $aggDeg$) rank a large number of network nodes, or in other words, nodes are ranked more from their $aggDeg$ index than from their $alPCI$. Such combination, results in distinguishing highly connected nodes that have interlinks to all layers, from those that do not posses that property. $lsPCI$ operates similarly to $alPCI$ since its indexes are limited by the number of layers. Thus, the results illustrated in Fig. 6 is the outcome of ranking nodes according to $lsPCI$, while breaking ties via the largest $aggDeg$ index. Nonetheless we expect that for multilayer networks composed of more layers, $lsPCI$'s efficiency will be distinguished further.

Typically, as λ_{ii} increases above the epidemic probability, the identification of influential spreaders becomes more difficult for any algorithm to detect. This is due to the fact that for large λ_{ii} values, that is, as λ_{ii} deviates significantly from the epidemic probability, an epidemic occurs regardless of the characteristics of the initially infected node [1]. Even if the initially infected node is not an influential one, at broad spreading rates there is high possibility that an influential will be "reached" as the spreading progresses, and thus result in epidemic propagation. Hence true conclusion can only be drawn near the epidemic probability.

It is straightforward to understand that the way interconnections are distributed over the different layers, and to the nodes within those layers, plays a vital role in the diffusion dynamics, and thus, in the performance of the competitors. Hence, for any algorithm in order to be characterized as an efficient technique for the detection of those powerful spreaders, intra and inter connections must be incorporated and combined in the most efficient of ways in order to predict the probability of an epidemic outbreak. Robustness to either limited or increased number of inter-links is also a necessity. Furthermore, it can be concluded that traditional approaches that project the multilayer network to a single dimension cannot predict the actual spreading power of nodes in these complex structures.

5.2.3 Impact of Inter Connections and Inter Diffusion Probability

In Fig. 7, we investigate on how the competitors behave in the increase of the inter-layer spreading probability. To this end we choose to have a favorable distribution regarding the inter-degree of nodes, i.e., $s_{degree} = 0.3$. First, the ranking

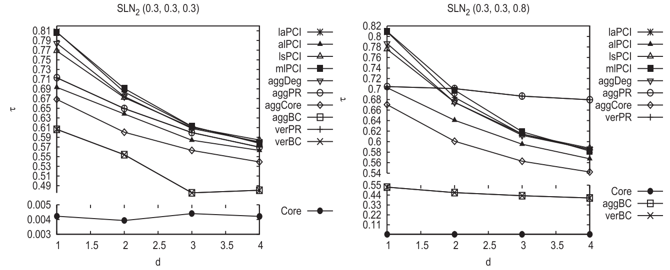

Fig. 8. Increasing in the number of interconnections in the SLN networks. It can be observed that all methods illustrate a decreasing trend as d increases. Setting s_{node} at 0.8 and thus assigning to a specific set of nodes many interconnections, works in favor of $verPR$ which exhibits an exceptional performance in this case.

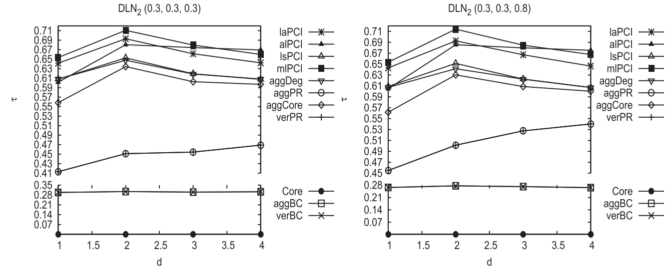

Fig. 9. Increasing in the number of interconnections in the DLN networks. As interconnections increase $alPCI$ yields better results, i.e., from 4th when $d = 1$ to 1st when $d = 4$. Its performance is different from the SLN networks because for the DLN networks, the distribution of inter-k_{out} is still significantly lower (even for $d = 4$) from that of intra-k_{out} (compare Figs. 10 and 11 with Fig. 13 in the Appendix available online) which does not hold for the SLN networks.

obtained from the previous section has remained relatively unchanged. This observation strengthens the evaluation of $mlPCI$ which illustrates a robust behavior to the different spreading rates used in our simulations. Examining the trends of the illustrated curves, it can be observed that all competing methods become more effective as λ_{ij} increases. Focusing on Figs. 7a, 7b, and 7c, we observe that as λ_{ij} increases above the epidemic probability, the distance in performance of $aggDeg$ with $mlPCI$ and $laPCI$ starts to decrease, and coincides at 0.3. However, similarly to our previous discussion, true influential spreaders can only emerge near the epidemic threshold, where we observe that $mlPCI$ has the largest τ compared to the remaining techniques.

In Figs. 7d, 7e, and 7f, the performance of $mlPCI$ is distinct even when λ_{ij} is above the epidemic probability. The basic difference between these networks and those in Figs. 7a, 7b, and 7c lies in the distribution of inter-intra k_{out} (Figs. 11, 12, and 13 in the Appendix available online). Specifically, in the DLN networks, nodes are much more intra-connected in their focal layer than inter-connected to different layers while for the SLN networks, intra and inter connections are more comparable. Hence, for the DLN networks, the inter-connections will have a smoother impact on the spreading dynamics.

Our evaluation so far illustrates that the interplay between the different layers affects the competing algorithms differently. For instance s_{node} at 0.8 affects the performance of $verPR$ positively—also illustrated in the previous section—as depicted for example in Figs. 7b and 7c. $verBC$'s performance decreases when either s_{layer} or s_{node} is set to 0.8, and in fact it is lower, when both parameters are set at 0.8. This observation is most evident in Figs. 7a and 7c.

Similarly to Fig. 6b, due to $s_{layer} = 0.8$, $alPCI$ is unable to rank nodes in the SLN_2(0.3, 0.8, 0.8) network (Fig. 7c). This is due to the fact that nodes are not interconnected towards all layers. Nonetheless, from Figs. 7d, 7e, and 7f, we can observe that even when $alPCI$ ranks nodes with a limited number of different indexes, by breaking ties via the largest $aggDeg$ policy, we obtain a significant improvement in τ.

The above considerations are vital ingredients for building a successful recipe that will detect influential nodes in multilayer networks. It is our belief that all these characteristics must be imprinted within a technique in hopes of understanding and predicting the spreading power of nodes. $mlPCI$ inherently filters a node's near vicinity, i.e., *those "k" neighbors at least "k" connected from just one to all layers of the multilayer network*, which as shown in the majority of the illustrated results, separates it from the rest of the competing algorithms.

5.2.4 Impact of Increasing Interconnections (d)

Our final section illustrates the performance of the competitors as we increase in the density of interconnections (see Figs. 8 and 9). Reminisce that all spreading rates are set to the epidemic probability, however, as d increases, the epidemic probability of the aggregated network decreases, i.e., λ_{ij} decreases. This observation is evident in the SLN networks even at the initial values of d. For instance when s_{degree} is set to 0.3, the largest eigenvalue is about 10, 15, 21 and 27 for $d = 1, 2, 3$ and 4 respectively. Evidently, the increase of the largest eigenvalue, and thus the decrease of the epidemic probability, is confoundedly significant. For $s_{degree} = 0.8$, we observe a smaller increase, e.g., 8.5 for $d = 2$, however such behavior is anticipated due to the distribution of inter-connections. In [46] the authors state that the epidemic probability of the aggregated network is smaller than that of the individual layers. This observation is coherent with our study in the SLN networks (Table 4), however, for the DLN case, where the multilayer network is composed of layers with different number of nodes, edges, degree distribution etc., we found that for $d \leq 2$ the epidemic probability followed the eigenvalue of Wiki-Vote (about 45), that is, the layer with the largest eigenvalue. Even when we increased d up to 4, we did not observe a significant increase, e.g., 47 and 49 for $d = 3$ and 4 respectively. This is due to the large difference in the distribution of k_{out} of the inter and intra neighbors.

In particular, examining Figs. 8 and 9, we observe similar results with our previous discussions. Evidently, the algorithms perform differently in the SLN networks with regard to their performance in the DLN scenarios. The former depicts a decreasing correlation with influence as d increases, with the exception of $verPR$, whereas the latter shows a more complex behavior. At this point we should note that in the SLN networks, the increase of d employs a growing number of interconnections that surpass that of the intra-links for $d > 2$. In terms of $alPCI$, this attribute is not advantageous, since nodes will be indexed for their k neighbors to all layers, thus their rank is bounded to the limits of their intra-connections. On the contrary in the DLN networks which are not governed by such rule, $alPCI$ has increased correlation with influence, performing similar to $mlPCI$ when $d \geq 3$, i.e., when nodes have more connections to all layers.

6 Related Work

The literature in the identification of influential spreaders in multilayer networks is yet limited, and thus, we briefly review several approaches that are a benchmark for the detection of those spreaders in single-layer networks. A large amount of research is focused in the structural properties of a network for detecting influential spreaders, since the topological characteristics of a node affects the diffusion dynamics. To this end, betweenness centrality is often employed to decide the spreading power of a node [10], [27]. The degree centrality is a local and quite effective method, that is often intertwined with a node's spreading potential. The k-core decomposition [27] that discriminates nodes in cores (shells) with respect to their (least) connectivity exploited several shortcoming of the degree centrality and introduced a series of methods which in their turn focused on various drawbacks of k-core. Specifically, in [53] the authors noted that due to the pruning nature of the algorithm, a lot of information regarding the connectivity of a node is completely neglected. Hence they introduced a parameterized method which accounted for the fraction of its discarded degree. In [34] the authors further distinguished the spreading potential of nodes assigned within the same shell, by considering their distance from the core nodes of the network. Later in [24], it was shown that nodes with more connections to the core nodes, reveal more capable spreaders. Other network decomposition methods include the k-truss [36], [51], the onion method [21] etc. Focusing on measures that require local knowledge of a node's characteristics, in [10] the authors aggregated the number of nearest and next nearest neighbors of a node. Similarly in [44] the authors focus only to the nearest neighbors. In [4] another local measure was proposed that distinguishes nodes that reside in dense neighborhoods. Currently, there are various studies investigating epidemics and spreading processes on multilayer networks [12], [16], [45], [46], [54], cascading failures e.g., [48], node ranking e.g., [20], [49], and so on, but very limited work in the identification of influential spreaders. The works most closely related to the current article, i.e., to influentials detection, are those reported in [2], [3], [11], [15] and a detailed critique of them appeared in Section 1; here we briefly mention them. The blending of all layers into a single one and then application of traditional options for influentials detection is proposed in [2]. A generalization of the k-core is proposed in [3] but it results in a vector of values that can not be used in a straightforward manner for detecting effective influential spreaders. In [11], the authors proposed an called KS, which follows the intuition of [24], i.e., aggregates the shell indexes of its neighbors, and moreover combines the intra and inter layer spreading rates. However, to our understanding, incorporating the unknown spreading rates, of (and between) the layers, is not realistic. In [15] very elegant methods based on tensor analysis are proposed. Finally, we need to mention that two of the measures proposed in this work has been used for building connected dominating sets in multilayer ad hoc networks [42].

7 Conclusion

Multilayer complex networks have recently been the focus of intense study in the realm of network science. Real instances of them include transportation networks, online social networks, power networks and so on. Diffusion processes, such as spreading processes, cascading failures, cooperative behavior are significant fields of study. Among them, the identification of influential spreaders is a significant task due to its application in immunization strategies, advertising and so on.

This article investigated the problem of identifying influential spreaders over multilayer complex networks, since we are currently 'embedded' in multiple networks concurrently, e.g., in the case of online networks, we have an account at Facebook, Linkedin, Twitter, etc. and we spread our ideas/product-preferences using all of them. The article explained the lack of proposals so far for carrying out this task, and explained the inadequacy of the corresponding techniques proposed for the same problem in the case of single-layer complex networks because they do not take into account the existence of multiple layers and/or generate solutions that do not allow the straightforward ranking of nodes for selecting the most influentials.

Then, it proposed a family of measures for describing the strategic position of a node within a multilayer network. These measures condense into a single number the connectivity of the node with respect to nodes belonging to the same layer as well as to the rest of the layers. The calculation of these measures requires only information of the connectivity of the surrounding nodes, and not iterative computations with knowledge of the network-wide topology thus making it scalable, and quickly computable. Moreover, this feature makes them suitable both for online (e.g., response to evolving infections) as well as offline mining tasks (e.g., selection of best 'promoters'), due to the huge size of underlying networks.

The experimental evaluation of the proposed methods carried out against all major competitors proposed so far for either single-layer or multilayer networks, i.e., degree, betweenness centrality, PageRank and k-core for single and multilayer/multiplex networks. The complex networks used for the evaluation spanned a wide variety of network structure and size, and a network generator was also developed and used so as to test a wide range of topology characteristics. The final outcome of the evaluation marked $mlPCI$ as the best performing measure for almost each and every dataset used. Its success can be attributed on building on the shortcomings and embedding the benefits of the members of its family proposed in this article; it achieved to summarize the connectivity around a node in a concise and quite accurate way, even though it refrains from examining the whole network topology with time-consuming iterative decomposition procedures.

Acknowledgments

We are indebted to the reviewers for their constructive comments which improved the quality of the article. This work was partially supported by the US Office of Naval Research (ONR) under award N00014-14-1-2190.

References

[1] B. Aditya Prakash, D. Chakrabarti, N. C. Valler, M. Faloutsos, and C. Faloutsos, "Threshold conditions for arbitrary cascade models on arbitrary networks," *Knowl. Inf. Syst.*, vol. 33, no. 3, pp. 549–575, 2012.

[2] M. A. Al-garadi, K. D. Varathan, S. D. Ravana, E. Ahmed, and V. Chang, "Identifying the influential spreaders in multilayer interactions of online social networks," *J. Intell. Fuzzy Syst.*, vol. 31, no. 5, pp. 2721–2735, 2016.

[3] N. Azimi-Tafreshi, J. Gómez-Gardeñes, and S. N. Dorogovtsev, "k-core percolation on multiplex networks," *Phys. Rev. E*, vol. 90, no. 3, 2014, Art. no. 032816.

[4] P. Basaras, D. Katsaros, and L. Tassiulas, "Detecting influential spreaders in complex, dynamic networks," *IEEE Comput.*, vol. 46, no. 4, pp. 26–31, Apr. 2013.

[5] P. Basaras, D. Katsaros, and L. Tassiulas, "Dynamically blocking contagions in complex networks by cutting vital connections," in *Proc. IEEE Int. Conf. Commun.*, 2015, pp. 1170–1175.

[6] S. Boccaletti, et al., "The structure and dynamics of multilayer networks," *Physics Reports*, vol. 544, pp. 1–222, 2014.

[7] S. V. Buldyrev, R. Parshani, P. Gerald, S. H. Eugene, and H. Shlomo, "Catastrophic cascade of failures in interdependent networks," *Nature*, vol. 464, pp. 1025–1028, 2010.

[8] M. G. Campiteli, A. J. Holanda, L. D. H. Soares, P. R. C. Soles, and O. Kinouchi, "Lobby index as a network centrality measure," *Physica A: Statistical Mech. Appl.*, vol. 392, no. 21, pp. 5511–5515, 2013.

[9] A. Cardillo, et al., "Emergence of network features from multiplexity," *Nature Sci. Reports*, vol. 3, 2013, Art. no. 1344.

[10] D. Chen, L. Lu, M. S. Shang, Y. C. Zhang, and T. Zhou, "Identifying influential nodes in complex networks," *Physica A: Statistical Mech. Appl.*, vol. 391, no. 4, pp. 1777–1787, 2012.

[11] Z. Dawei, L. Lixiang, L. Shudong, H. Yujia, and Y. Yixian, "Identifying influential spreaders in interconnected networks," *Physica Scripta*, vol. 89, no. 1, 2014, Art. no. 015203.

[12] M. De Domenico, C. Granell, M. A. Porter, and A. Arenas, "The physics of spreading processes in multilayer networks," *Nature Physics*, vol. 12, pp. 901–906, 2016.

[13] M. De Domenico, M. A. Porter, and A. Arenas, "MuxViz: A tool for multilayer analysis and visualization of networks," *J. Complex Netw.*, vol. 3, no. 2, pp. 159–176, 2015.

[14] M. De Domenico, et al., "Mathematical formulation of multilayer networks," *Phys. Rev. X*, vol. 3, 2013, Art. no. 041022.

[15] M. De Domenico, A. Solé-Ribalta, E. Omodei, S. Gómez, and A. Arenas, "Ranking in interconnected multilayer networks reveals versatile nodes," *Nature Commun.*, vol. 6, 2015, Art. no. 6868.

[16] M. Dickison, S. Havlin, and H. E. Stanley, "Epidemics on interconnected networks," *Phys. Rev. E*, vol. 85, no. 6, 2012, Art. no. 066109.

[17] N. Dimokas, D. Katsaros, L. Tassiulas, and Y. Manolopoulos, "High performance, low overhead cooperative caching for wireless sensor networks," in *Proc. IEEE Int. Symp. World Wireless Mobile Multimedia Netw.*, 2009, pp. 1–9.

[18] R. Fagin, A. Lotem, and M. Naor, "Optimal aggregation algorithms for middleware," in *Proc. ACM SIGMOD-SIGACT-SIGART Symp. Principles Database Syst.*, 2001, pp. 102–113.

[19] C. L. Freeman, "A set of measures of centrality based on betweenness," *Sociometry*, vol. 40, no. 1, pp. 35–41, 1977.

[20] A. Halu, R. J. Mondragón, P. Panzarasa, and G. Bianconi, "Multiplex PageRank," *PLoS ONE*, vol. 8, no. 10, 2013, Art. no. e78293.

[21] L. Hebert-Dufresne, J. A. Grochow, and A. Allard, "Multi-scale structure and topological anomaly detection via a new network statistic: The onion decomposition," *Nature Sci. Reports*, vol. 3, 2016, Art. no. 31708.

[22] J. E. Hirsch, "An index to quantify an individual's scientific research output," *Proc. Nat. Academy Sci.*, vol. 102, no. 46, pp. 16569–16572, 2005.

[23] L. G. S. Jeub, M. W. Mahoney, P. J. Mucha, and M. A. Porter, "A local perspective on community structure in multilayer networks," *Netw. Sci.*, 2017. [Online]. Available: https://doi.org/10.1017/nws.2016.22

[24] B. Joonhyun and K. Sangwook, "Identifying and ranking influential spreaders in complex networks by neighborhood coreness," *Physica A: Statistical Mech. Appl.*, vol. 395, pp. 549–559, 2014.

[25] D. Kempe, E. Tardos, and J. Kleinberg, "Maximizing the spread of influence through a social network," *Theory Comput.*, vol. 11, no. 4, pp. 105–147, 2015.

[26] M. G. Kendall, "A new measure of rank correlation," *Biometrika*, vol. 80, pp. 81–93, 1938.

[27] M. Kitsak, et al., "Identification of influential spreaders in complex networks," *Nature Phys.*, vol. 6, pp. 888–893, 2010.

[28] M. Kivelä, A. Arenas, M. Barthelemy, J. P. Gleeson, Y. Moreno, and M. A. Porter, "Multilayer networks," *J. Complex Netw.*, vol. 2, pp. 203–271, 2014.

[29] A. Korn, A. Schubert, and A. Telcs, "Lobby index in networks," *Physica A: Statistical Mech. Appl.*, vol. 388, no. 11, pp. 2221–2226, 2009.

[30] C. J. Kuhlman, G. Tuli, S. Swarup, M. V. Marathe, and S. S. Ravi, "Blocking simple and complex contagion by edge removal," in *Proc. IEEE Int. Conf. Data Mining*, pp. 339–408, 2013.

[31] A. N. Langville and C. D. Meyer, *Google's Page Rank and Beyond: The Science of Search Engine Rankings*. Princeton, NJ, USA: Princeton Univ. Press, 2006.

[32] A. N. Langville and C. D. Meyer, *Who's #1?: The Science of Rating and Ranking*. Princeton, NJ, USA: Princeton Univ. Press, 2014.

[33] J. Leskovec and A. Krevl, "SNAP datasets: Stanford large network dataset collection," 2014. [Online]. Available: http://snap.stanford.edu/data

[34] J.-G. Liu, Z.-M. Ren, and Q. Guo, "Ranking the spreading influence in complex networks," *Physica A: Statistical Mech. Appl.*, vol. 392, no. 18, pp. 4154–4159, 2013.

[35] L. Lü, T. Zhou, Q.-M. Zhang, and H. E. Stanley, "The H-index of a network node and its relation to degree and coreness," *Nature Commun.*, vol. 7, 2016, Art. no. 10168.

[36] F. D. Malliaros, M.-E. Rossi, and M. Vazirgiannis, "Locating influential nodes in complex networks," *Nature Sci. Reports*, vol. 6, 2016, Art. no. 19307.

[37] X. Meng, "Centrality measures in multilayer networks," Univ. Oxford, Oxford, U.K., 2015.

[38] F. Morone and H. A. Makse, "Influence maximization in complex networks through optimal percolation," *Nature*, vol. 524, pp. 65–68, 2015.

[39] P. J. Mucha and M. A. Porter, "Communities in multislice voting networks," *Chaos: An Interdisciplinary J. Nonlinear Sci.*, vol. 20, no. 4, 2010, Art. no. 041108.

[40] P. J. Mucha, T. Richardson, K. Macon, M. A. Porter, and J.-K. Onnela, "Community structure in time-dependent, multiscale, and multiplex networks," *Sci.*, vol. 328, no. 5980, pp. 876–878, 2010.

[41] M. E. J. Newman, *Netw.: An Introduction*. London, U.K.: Oxford Univ. Press, 2010.

[42] D. Papakostas, P. Basaras, D. Katsaros, and L. Tassiulas, "Backbone formation in military multi-layer ad hoc networks using complex network concepts," in *Proc. IEEE Military Commun. Conf.*, 2016, pp. 842–848.

[43] R. Pastor-Satorras and C. Castellano, "Topological structure and the H index in complex networks," *Phys. Rev. E*, vol. 95, no. 2, 2017, Art. no. 022301.

[44] S. Pei, L. Muchnik, J. A. Andrade, Z. Zheng, and H. A. Makse, "Searching for superspreaders of information in real-world social media," *Nature Sci. Reports*, vol. 4, 2014, Art. no. 5547.

[45] M. Salehi, R. Sharma, M. Marzolla, M. Magnani, P. Siyari, and D. Montesi, "Spreading processes in multilayer networks," *IEEE Trans. Netw. Sci. Eng.*, vol. 2, no. 2, pp. 65–83, Apr.-Jun. 2015.

[46] A. Saumell-Mendiola, M. A. Serrano, and M. Boguñá, "Epidemic spreading on interconnected networks," *Phys. Rev. E*, vol. 86, no. 2, 2012, Art. no. 026106.

[47] E. M. Shahrivar and S. Sundaram, "The strategic formation of multi-layer networks," *IEEE Trans. Netw. Sci. Eng.*, vol. 2, no. 4, pp. 164–178, Oct.-Dec. 2015.

[48] D.-H. Shin, D. Qian, and J. Zhang, "Cascading effects in interdependent networks," *IEEE Netw.*, vol. 28, no. 4, pp. 82–87, Jul-Aug. 2014.

[49] A. Solé-Ribalta, M. De Domenico, S. Gómez, and A. Arenas, "Centrality rankings in multiplex networks," in *Proc. ACM Conf. Web Sci.*, 2014, pp. 149–155.

[50] A. Tagarelli and R. Interdonato, "Understanding lurking behaviors in social networks across time," in *Proc. IEEE/ACM Int.Conf. Advances Social Netw. Anal. Mining*, 2014, pp. 51–55.

[51] J. Wang and J. Cheng, "Truss decomposition in massive networks," *Proc. VLDB Environment*, vol. 5, no. 9, pp. 812–823, 2012.

[52] P. Wang, G. Robins, P. Pattison, and E. Lazega, "Exponential random graph models for multilevel networks," *Social Netw.*, vol. 35, pp. 96–115, 2013.

[53] A. Zeng and C. J. Zhang, "Ranking spreaders by decomposing complex networks," *Physics Lett. A*, vol. 377, no. 14, pp. 1031–1035, 2013.

[54] Y. Zhuang and O. Yagan, "Information propagation in clustered multilayer networks," *IEEE Trans. Netw. Sci. Eng.*, vol. 3, no. 4, pp. 211–224, Oct.-Dec. 2016.

Pavlos Basaras received the MSc degree in computer engineering from the University of Thessaly, Greece. He is working toward the PhD degree in the Department of Electrical and Computer Engineering, the University of Thessaly, Greece. His research focuses on complex networks analysis from the perspective of network science and graph theory; currently, he is interested in multilayer networks and the spreading processes.

George Iosifidis received the Diploma degree in electronics and telecommunications engineering from Greek Air Force Academy in 2000, and the MS and PhD degrees in electrical engineering from the University of Thessaly, Greece, in 2007 and 2012, respectively. He was a post-doctoral researcher with CERTH, Greece, and Yale University. He is currently an ussher assistant professor of future networks with Trinity College, University of Dublin, Ireland, and also a funded investigator with the Telecommunications Research Centre CONNECT, Ireland. His research interests include the broad area of wireless network optimization and network economics.

Dimitrios Katsaros received the PhD degree in informatics from the Aristotle University of Thessaloniki, Greece in 2004. He is a visiting assistant professor in the Department of Electrical Engineering, Yale University, and assistant professor in the Department of Electrical and Computer Engineering, the University of Thessaly, Greece. His research interests include distributed systems and algorithms.

Leandros Tassiulas (S89–M91–SM06–F07) received the PhD degree in electrical engineering from the University of Maryland, College Park, MD, in 1991. He has held faculty positions with Polytechnic University, NY, the University of Maryland, College Park, and the University of Thessaly, Greece. He is currently the John C. Malone professor of Electrical Engineering and member of the Institute for Network Science, Yale University, New Haven, CT. His research interests include the field of computer and communication networks with emphasis on fundamental mathematical models and algorithms of complex networks, architectures and protocols of wireless systems, sensor networks, novel Internet architectures, and experimental platforms for network research. His most notable contributions include the max-weight scheduling algorithm and the back-pressure network control policy, opportunistic scheduling in wireless, the maximum lifetime approach for wireless network energy management, and the consideration of joint access control and antenna transmission management in multiple antenna wireless systems. His research has been recognized by several awards, including the IEEE Koji Kobayashi Computer And Communications Award (2016), the Inaugural INFOCOM 2007 Achievement Award for fundamental contributions to resource allocation in communication networks, the INFOCOM 1994 and 2017 Best Paper Awards, the National Science Foundation (NSF) Research Initiation Award (1992), the NSF CAREER Award (1995), the Office of Naval Research Young Investigator Award (1997), and the Bodossaki Foundation Award (1999). He is a fellow of the IEEE.