Identifying key nodes in multilayer networks based on tensor decomposition

Cite as: Chaos 27, 063108 (2017); https://doi.org/10.1063/1.4985185
Submitted: 15 November 2016 . Accepted: 26 May 2017 . Published Online: 16 June 2017

Dingjie Wang, Haitao Wang, and Xiufen Zou

ARTICLES YOU MAY BE INTERESTED IN

Extracting information from multiplex networks
Chaos: An Interdisciplinary Journal of Nonlinear Science 26, 065306 (2016); https://doi.org/10.1063/1.4953161

Eigenvector centrality of nodes in multiplex networks
Chaos: An Interdisciplinary Journal of Nonlinear Science 23, 033131 (2013); https://doi.org/10.1063/1.4818544

Multilayer motif analysis of brain networks
Chaos: An Interdisciplinary Journal of Nonlinear Science 27, 047404 (2017); https://doi.org/10.1063/1.4979282
Identifying key nodes in multilayer networks based on tensor decomposition

Dingjie Wang,1,2 Haitao Wang,1,2 and Xiufen Zou1,2,a)
1School of Mathematics and Statistics, Wuhan University, Wuhan 430072, China
2Computational Science Hubu Key Laboratory, Wuhan University, Wuhan 430072, China

(Received 15 November 2016; accepted 26 May 2017; published online 16 June 2017)

The identification of essential agents in multilayer networks characterized by different types of interactions is a crucial and challenging topic, one that is essential for understanding the topological structure and dynamic processes of multilayer networks. In this paper, we use the fourth-order tensor to represent multilayer networks and propose a novel method to identify essential nodes based on CANDECOMP/PARAFAC (CP) tensor decomposition, referred to as the EDCPTD centrality. This method is based on the perspective of multilayer networked structures, which integrate the information of edges among nodes and links between different layers to quantify the importance of nodes in multilayer networks. Three real-world multilayer biological networks are used to evaluate the performance of the EDCPTD centrality. The bar chart and ROC curves of these multilayer networks indicate that the proposed approach is a good alternative index to identify real important nodes. Meanwhile, by comparing the behavior of both the proposed method and the aggregated single-layer methods, we demonstrate that neglecting the multiple relationships between nodes may lead to incorrect identification of the most versatile nodes. Furthermore, the Gene Ontology functional annotation demonstrates that the identified top nodes based on the proposed approach play a significant role in many vital biological processes. Finally, we have implemented many centrality methods of multilayer networks (including our method and the published methods) and created a visual software based on the MATLAB GUI, called ENMNFinder, which can be used by other researchers. Published by AIP Publishing. [http://dx.doi.org/10.1063/1.4985185]

I. INTRODUCTION

Complex network research has been shifting away from discovering macroscopic regularities of structure and dynamics to uncovering the role of macroscopic elements acting as nodes in real-world systems.1–5 In the past several decades, designing effective centrality methods to identify essential nodes in complex networks has been an intriguing topic.5–10 Previous studies have shown that centrality measures, such as degree, betweenness, PageRank, hub, authority, and eigenvector centrality, can usefully identify essential proteins and potential drug targets for the survival of the cell,11 allow us to better control the outbreak of epidemics,12 prevent catastrophic outages in power grids,13 drive the network toward a desired state,14–16 improve transport capacity,17 and promote cooperation in evolutionary games.18,19

However, until now, the comprehensive research on centrality methods has focused on single-layer networks whose nodes are linked by a single type of interaction. A variety of real-world complex networks are, in fact, interconnected by different types of interactions between nodes, forming what is known as multilayer networks;1 for example, several types of actors’ relationships, including friendship, vicinity, kinship, and membership in the same cultural society, constitute a multilayer public social network.20 It is apparent that neglecting the multiple relationships of structures between nodes may occasionally result in not fully capturing the details in some real-world systems; this leads to incorrect conclusions about the topological and dynamical properties.21–23 Integrating different sources and the relationships of structures, which vary

Many complex real-world relationships described by different networks are in the majority of cases not independent, for example, the symptoms of complex diseases like allergies, obesity, and cancer depend on the products of multiple interacting genes, forming what is known as multilayer networks. Designing efficient centrality methods to identify the most central agents in multilayer networks is an important means of understanding the topological structure and dynamic processes of multilayer networks. In this paper, based on CANDECOMP/PARAFAC (CP) tensor decomposition, we propose a computational framework to quantify the importance of nodes from large weighted multilayer networks. The experimental results of three real-world multilayer biological networks indicate that the proposed approach is better able to identify actual important nodes than other published methods. Gene Ontology functional annotation indicates that the top nodes identified by the proposed approach play a significant role in many vital biological processes, which also demonstrates that the proposed approach is a useful tool for identifying novel disease genes and potential drug targets. Finally, a visual software, called ENMNFinder, is created to detect the importance of nodes in multilayer networks; this software can be easily used by other researchers.

a)Electronic mail: xzfou@whu.edu.cn.
in terms of scale, to identify the critical nodes of complex systems is a challenging and intriguing topic. In recent years, from the perspective of multilayer networks, several centrality metrics, including the multilayer PageRank centrality,\textsuperscript{24–26} the overlapping degree,\textsuperscript{27} the versatility,\textsuperscript{28} and the multilayer eigenvector,\textsuperscript{29} have been extended to identify essential nodes in real-world complex systems. More specifically, the multilayer PageRank centrality\textsuperscript{24–26} was provided based on the idea of biased random walks, which integrated the effects of the interplay between networks on the centrality of nodes; the overlapping degree centrality\textsuperscript{27} was proposed by aggregating the information of nodes’ neighbors in all layers; the multilayer eigenvector centrality\textsuperscript{29} taken into account the layer structure by means of a directed graph of influences among layers.

As we all know, in the past few decades, the tensor decomposition, an effective tool for the integration and analysis of multi-dimensional data, has been widely applied in many areas of science and engineering.\textsuperscript{30} Recently, the tensorial framework was proposed to study the network descriptors and dynamical processes in temporal and multilayer networks.\textsuperscript{31,32} In particular, based on the tensor representation framework, many well-known centrality measures were extended from single-layer networks to multilayer networks. For example, “versatility” includes Eigenvector and PageRank centralities;\textsuperscript{28} Overlapping degree and multilayer centralities are currently lacking.

This study, we use the fourth-order tensor to represent the multilayer networks and propose a novel computational method to identify the essential nodes based on CANDECOMP/PARAFAC (CP) tensor decomposition,\textsuperscript{30} referred to as EDCPTD centrality. Unlike other methods, the proposed approach integrates the different types of links between nodes to quantify the importance of nodes in multilayer networks using the tensor decomposition technique. The performances of the proposed approach are evaluated based on a comparison of published methods; the essential proteins of three known biological multilayer networks are identified, and their bar chart, receiver operating characteristic (ROC) curves,\textsuperscript{36} and correlation analysis are presented based on different centrality methods.

II. METHODS

A. Multilayer network model and tensor representation

In this paper, we focus on the multilayer networks that consist of a fixed set of nodes connected by different types of links.\textsuperscript{1} A simple multilayer network and corresponding aggregated network\textsuperscript{1,29} are shown in Figs. 1(a) and 1(b). More formally, a multilayer network is a pair \( W = (\Phi, C) \), where \( \Phi = \{G_x : x \in \{1, 2, \ldots, L\}\} \) is a family of graphs \( G_x = (V_x, E_x) \) (called layers of \( W \)) and

\[
C = \{E_{xy} \subseteq V_x \times V_y : x, y \in \{1, 2, \ldots, L\}\},
\]

is the set of interconnections between nodes of different layers \( G_x \) and \( G_y(x \neq y) \). Here, the elements of \( C \) are called crossed layers; \( E_x \) is the set of intralayer connections in the layer \( G_x \); in contrast, \( E_{xy} \) is the set of interlayer connections, where \( V_1 = V_2 = \cdots = V_L = \{v_1, v_2, \ldots, v_N\} \) is a fixed set of \( N \) nodes.

A multilayer network can be represented in tensor form as follows.\textsuperscript{31} Let \( M = (M_{ijk}) \in \mathbb{R}^{N \times L \times N \times L} \) denote the fourth-order weight adjacency tensor of a multilayer network. Each element of \( M \) is defined by

\[
M_{ijk} = \begin{cases} \omega_{ijk}, & \text{if } v_i^1 \rightarrow v_j^k \\ 0, & \text{otherwise}. \end{cases}
\]

Here, \( 1 \leq i, j \leq N, 1 \leq a, b \leq L, \) and \( v_i^1 \) represents the node \( i \) in the layer \( a \); and \( \omega_{ijk} \) represents the weight of the link that node \( i \) in layer \( a \) points to node \( j \) in layer \( b \). Thus, this tensor representation provides a powerful instrument for studying the network descriptors and dynamical processes in interconnected multilayer networks.

B. Notation and preliminaries of the tensor

Before presenting our methods, we first introduce the concept and properties of the tensor.\textsuperscript{30} A tensor is a multi-dimensional array. More formally, an \( n \)-order tensor is an element of the tensor product of \( n \) vector spaces, each of which has its own coordinate system. Thus, the concept “tensor” describes arrays of different sizes, such as a vector \( a \) of order \( N \) is a 1-order tensor, e.g., \( a \in \mathbb{R}^N \), the adjacency matrix \( A \) is a 2-order tensor, e.g., \( A \in \mathbb{R}^{N \times N} \), the above \( M \) is a 4-order tensor, e.g., \( M \in \mathbb{R}^{L \times L \times N \times L} \), etc. In general, an \( I_1 \times I_2 \times \cdots \times I_n \) \( n \)-order array is a tensor in \( \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_n} \).

The inner product of two tensors with the same sizes \( X, Y \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_n} \) is the sum of the products of their entries, i.e.,

\[
(X, Y) = \sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2} \cdots \sum_{i_n=1}^{I_n} X_{i_1 i_2 \cdots i_n} Y_{i_1 i_2 \cdots i_n}.
\]

The norm of a tensor \( X \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_n} \) is the square root of its inner product with itself, i.e.,

\[
\|X\| = \sqrt{(X, X)}.
\]

An \( n \)-order tensor \( X \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_n} \) is rank one if it can be written as the outer product of \( n \) vectors, i.e.,
\[
X = a^{(1)} \circ a^{(2)} \circ \cdots \circ a^{(n)}.
\]

The symbol \( \circ \) represents the vector outer product. This means that each element of tensor is given by \( X_{i_1i_2\cdots i_n} = a_{i_1}^{(1)} a_{i_2}^{(2)} \cdots a_{i_n}^{(n)} \).

Matricization, also known as unfolding or flattening, is the process of reordering the elements of an \( n \)-order tensor into a matrix. The mode-\( m \) matricization of a tensor \( X \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_k} \) is denoted by \( X_{(m)} \) and arranges the mode-\( m \) fibers to be the columns of the resulting matrix. Specifically, tensor element \((i_1, i_2, \ldots, i_k)\) maps to matrix element \((i_m, j)\), where

\[
j = 1 + \sum_{k=1}^{n} (i_k - 1)J_k, \quad \text{with} \quad J_k = \prod_{i=1}^{k-1} I_i.
\]

Given matrices \( A = (a_1, \ldots, a_N) \in \mathbb{R}^{N \times N} \) and \( B = (b_1, \ldots, b_N) \in \mathbb{R}^{N \times N} \), the Kronecker, Khatri-Rao, and Hadamard products, denoted by \( A \otimes B, A \odot B, A * B \), respectively, are defined as follows:

\[
A \otimes B = [a_1 \otimes b_1 \, a_2 \otimes b_2 \, \cdots \, a_N \otimes b_N],
\]

\[
A \odot B = \begin{bmatrix}
a_{11}B & \cdots & a_{1N}B \\
\vdots & \ddots & \vdots \\
a_{N1}B & \cdots & a_{NN}B
\end{bmatrix},
\]

\[
A * B = \begin{bmatrix}
a_{11}b_{11} & \cdots & a_{1N}b_{1N} \\
\vdots & \ddots & \vdots \\
a_{N1}b_{N1} & \cdots & a_{NN}b_{NN}
\end{bmatrix}.
\]

C. Mathematical models of EDCPTD centrality in multilayer networks

The tensor representation provided a powerful instrument for studying the network descriptors and dynamical processes in multilayer networks. Tensor decomposition, an effective tool for data integration and analysis, has been widely applied in many areas of science and engineering. In this paper, we use the fourth-order tensor to represent multilayer networks and consider four roles of each node and each layer in the multilayer networks, namely, the authority and hub of nodes, and the authority and hub of layers, respectively. Based on the above tensor representation framework, we apply the CANDECOMP/PARAFAC (CP) tensor decomposition to obtain the most significant factors, which are called principal singular vectors and produce quadruplets of vectors to obtain the hub and authority scores for all nodes, and hub and authority scores for all layers in multilayer networks. Furthermore, by integrating these four measures, we provide a novel centrality measure, called EDCPTD centrality, which is used to quantitatively evaluate the importance of nodes in multilayer networks. In the following, we provide a mathematical model and algorithm for calculating EDCPTD centrality.

Based on the above discussion, we know that a fourth-order adjacency tensor \( M \) can be used to represent a wealth of complicated relationships among nodes in multilayer networks. Based on CP tensor decomposition, the four-order adjacency tensor \( M \) can be written as the sum of a finite number of rank-one tensors, and the formula is given by

\[
M \approx \sum_{r=1}^{R} \lambda_r a_r^{(1)} \circ a_r^{(2)} \circ a_r^{(3)} \circ a_r^{(4)},
\]

where \( \lambda_r = (\lambda_1, \ldots, \lambda_K) \) and \( \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_K \) are called the singular values of \( M \), and \( R \) is a desired approximation rank.

Here, \( a_r^{(1)}, a_r^{(3)} \in \mathbb{R}^N, a_r^{(2)}, a_r^{(4)} \in \mathbb{R}^L \) are called the singular vectors of the corresponding \( \lambda_r \). Elementwise, Eq. (6) is written as

\[
M_{i,j} \approx \sum_{r=1}^{R} \lambda_r \left( a_r^{(1)} \right)_i \left( a_r^{(2)} \right)_j \left( a_r^{(3)} \right)_i \left( a_r^{(4)} \right)_j,
\]

where \( 1 \leq i,j \leq N, 1 \leq a, b \leq L \).

As shown in steps 2 and 3 of Fig. 2, for multilayer networks, in the principal quadruplet \( \{a_{1}^{(1)}, a_{1}^{(2)}, a_{1}^{(3)}, a_{1}^{(4)}\} \) of the largest singular \( \lambda_1 \) of the adjacency tensor \( M \), the absolute values of two vectors \( a_1^{(1)}, a_1^{(3)} \in \mathbb{R}^N \) provide hub and authority scores of all nodes, and the absolute values of an additional two vectors \( a_1^{(2)}, a_1^{(4)} \in \mathbb{R}^L \) provide the hub and authority scores of all layers in multilayer networks, respectively. It is worth noting that when \( L = 1 \), i.e., the network is a single-layer structure, the above method in fact is the usual hyperlink-induced topic search (HITS) algorithm of single-layer networks. In other words, our method provides a more general framework from the perspective of a multilayer networked structure, which gives not only the hub and authority scores of each node but also the hub and authority scores of each layer.

Furthermore, by integrating these four measures \( \{a_{1}^{(1)}, a_{1}^{(2)}, a_{1}^{(3)}, a_{1}^{(4)}\} \), we propose a new centrality method, referred to as EDCPTD (Essential nodes Determining based on CP Tensor Decomposition) centrality, to evaluate the importance of nodes in multilayer networks. The mathematical formula is given by

\[
H_i = \frac{1}{L} \sum_{z=1}^{L} H_i^{(z)} (1 \leq i \leq N),
\]

where \( H_i^{(z)} = |(a_{i}^{(1)})_z| + |(a_{i}^{(2)})_z| + |(a_{i}^{(3)})_z| + |(a_{i}^{(4)})_z| \) denotes the EDCPTD centrality of the node \( i \) in the layer \( z \). \( H_i (1 \leq i \leq N) \), calculated by Eq. (7), denotes the EDCPTD centrality of node \( i \), which integrates four centralities of node \( i \) in all layers. A larger EDCPTD centrality indicates a relatively greater importance for each node within the whole network. The flow diagram describing the computation of EDCPTD centrality in multilayer networks is shown in Fig. 2.

D. The algorithm for calculating EDCPTD centrality

Suppose that a multilayer network is given adjacency tensor \( M \in \mathbb{R}^{N \times L \times N \times L} \) and the desired approximation rank \( R \). To obtain the EDCPTD centrality for each node, our goal
is to compute a CP decomposition with \( R \) components that best approximates \( M \), i.e.,

\[
\min_{\hat{M}} \| M - \hat{M} \|, \tag{8}
\]

with \( \hat{M} = \sum_{r=1}^{R} \lambda_r a_r^{(1)} \odot a_r^{(2)} \odot a_r^{(3)} \odot a_r^{(4)} \). We denote that \( A^{(n)} = (a_1^{(n)}, \ldots, a_R^{(n)}) \) as factor matrices, which refer to the combination of the vectors from the rank-one components. Here, we denote

\[
\sum_{r=1}^{R} \lambda_r a_r^{(1)} \odot a_r^{(2)} \odot a_r^{(3)} \odot a_r^{(4)} \triangleq \left[ \lambda, A^{(1)}, A^{(2)}, A^{(3)}, A^{(4)} \right].
\]

Then, Eq. (8) can be expressed by

\[
\min_{\lambda, A^{(n) \circ \{1, \ldots, 4\}}} \| M - \left[ \lambda, A^{(1)}, A^{(2)}, A^{(3)}, A^{(4)} \right] \|. \tag{9}
\]

A more popular method for solving the CP factorization as shown in Eq. (9) is the alternating least squares (ALS) algorithm. The main idea of the ALS is to solve each matrix factor in turn and leave all others fixed. Without loss of generality, we assume that \( V \equiv A^{(n)} \) are first solved and leave all other matrix factors \( A^{(m)} (m \neq n) \) to be fixed. Then, Eq. (9) can be converted into the following optimization problem:

\[
\min_{V \in \mathbb{R}^{n \times r}} \| M - \left[ \lambda, A^{(1)}, \ldots, A^{(n-1)}, V, A^{(n+1)}, \ldots, A^{(4)} \right] \|. \tag{10}
\]

According to the matricization of the tensor, the above equation can be expressed in matrix form as

\[
\min_{\Lambda \in \mathbb{R}^{s}, V \in \mathbb{R}^{n \times r}} \| M_{(n)} - V \Lambda T^{(n)} \|, \tag{11}
\]

where \( T^{(n)} \triangleq A^{(4)} \odot \cdots \odot A^{(n+1)} \odot A^{(n-1)} \odot \cdots \odot A^{(1)} \) and \( \Lambda = \text{diag} \{ \lambda_1, \lambda_2, \ldots, \lambda_R \} \). Here, \( M_{(n)} \) represents the mode-\( n \) matricization of the tensor \( M \), and the symbol \( \text{"\odot\"} \) represents the Khatri-Rao product of the matrix. For a detailed introduction regarding these notations, see Sec. II B and Ref. 30. Equation (10) is a classic least square problem, and the optimal solution is easily computed as

\[
V \Lambda = M_{(n)} T^{(n)} U^1,
\]

where \( U = (A^{(4)})^T A^{(4)} \ast \cdots \ast (A^{(n+1)})^T A^{(n+1)} \ast (A^{(n-1)})^T A^{(n-1)} \ast \cdots \ast (A^{(1)})^T A^{(1)} \) and \( U^1 \) represents the pseudoinverse of the matrix \( U \). Here, the symbol \( \ast \) represents the Hadamard product of the matrix; for a detailed introduction, see Sec. II B and Ref. 30. We define that

\[
Z = \left[ [A^{(1)}, \ldots, A^{(n-1)}], U^1, A^{(n+1)}, \ldots, A^{(4)} ] \right],
\]

which is of size \( N \times \cdots \times I_{n-1} \times r \times I_{n+1} \times \cdots \times L \). Then, we have

\[
V \Lambda = M_{(n)} Z_{T}^{(n)} \in \mathbb{R}^{I_{n} \times r}.
\]

The algorithm of computing EDCPTD centrality in multilayer networks is presented in Algorithm 1. Here the ALS
algorithm of CP tensor decomposition is obtained by using MATLAB Tensor Toolbox Version 2.6.40

ALGORITHM 1: Computing EDCPTD centrality.

1: Input: Adjacency tensor $M \in \mathbb{R}^{N \times L \times N \times L}$ and desired rank $R > 0$.
2: Output: $\lambda \in \mathbb{R}^N$, $H = \{H_1, H_2, \ldots, H_N\} \in \mathbb{R}^N$.
3: Method:
4: for $n = 1, 2, \ldots, 4$ do
5: Initialize $A^{(0)}$;
6: end for
7: while not converged do
8: for $n = 1, 2, \ldots, 4$ do
9: $V = M_n X^{(n)} Y^{(n)} \in \mathbb{R}^N$;
10: where $X^{(0)} = A_1 \odot \cdots \odot A^{(n-1)} \odot A^{(n)}$, $Y^{(0)} = B_1 \odot \cdots \odot B^{(n-1)} \odot B^{(n)}$, $B^{(0)} = A^{(0)} A^{(0)}$, $(1 \leq i \leq 4)$,
11: where $B^{(0)} = A^{(0)} A^{(0)}$;
12: for $r = 1, 2, \ldots, R$ do
13: $H_i = \sum_{a=1}^{N} ||(a_1^{(0)})_i, (a_2^{(0)})_i ||^2 + ||(a_2^{(0)})_i, (a_2^{(0)})_i ||^2$.
14: end for
15: end for
16: end for
17: end while
18: for $j = 1, 2, \ldots, N$ do
19: $H_j = \frac{1}{N} \sum_{a=1}^{N} ||(a_1^{(0)})_j, (a_2^{(0)})_j ||^2 + ||(a_2^{(0)})_j, (a_2^{(0)})_j ||^2$.
20: end for
21: return $\lambda \in \mathbb{R}^N$, $H = \{H_1, H_2, \ldots, H_N\} \in \mathbb{R}^N$.

III. EXPERIMENTAL RESULTS

A. An illustrate example

To compare the different behaviors of the proposed method and other methods based on an aggregated single-layer structure,29 we provide a simple multilayer social network as an example, which is shown in Fig. 3. Figure 3(a) includes six individuals connected by two different social relationships, i.e., the friendship and kinship layers. The corresponding aggregated network is shown in Fig. 3(b). Figure 3(c) shows the ranking results for four different cases. EDCPTD centrality is used for the multilayer network, and eigenvector centrality for the other three cases.

In this simple multilayer social network, we observe that nodes 2 and 6 play the most central roles in the multilayer structure because they bridge together two different types of relationships. In other words, these two nodes could be able to spread information from one layer to the other in the interconnected multilayer social network. However, in the aggregated network, both node 3 and node 4 are identified as the most central roles. In layer 1 and layer 2 in the multilayer social network in Fig. 3(a), node 4 and node 3 are identified to be the most central nodes, respectively. These results indicate that calculating the centrality of nodes in each network of the multilayer structure separately or aggregating the information with a single network may lead to misleading results. The proposed centrality measure can effectively capture the importance of nodes in real networks.

B. Experimental data descriptions

To illustrate the different behaviors of the proposed method and other methods, computational analysis is performed using three real-world biological networks including the yeast landscape multilayer network (YLMN),41,42 H3N2 inflammatory multilayer network (HIMN),23 and lung cancer multilayer network (LCMN).43 The structural properties of the three tested multilayer networks are represented as shown in Table I.

The YLMN from Saccharomyces cerevisiae (the data are from Refs. 41 and 42) consist of 4458 proteins connected by four types of layer interactions: positive interactions, negative interactions, positive correlations, and negative correlations. The HIMN and LCMN are constructed by integrating high-throughput data. The HIMN consists of 2400 nodes and 63 558 edges connected by two types of layers, which correspond to the cell-specific normal and inflammatory layers. In addition, the corresponding gene expression profiling dataset of H3N2 is downloaded from the NCBI GEO database (GSE30550),44 and the constructed method of the HIMN is shown in Refs. 2 and 3. The constructed methods of the LCMN are shown in Ref. 43, and the profiling datasets of lung tissue have been downloaded from the NCBI GEO database (GSE10072).44 This network is composed of four layers: the protein-protein interaction (PPI) network, transcription factor (TF) co-targeting network, microRNA co-targeting network, and co-expression network.

TABLE I. The structural properties of the three tested multilayer networks.

<table>
<thead>
<tr>
<th>Name</th>
<th>Nodes</th>
<th>Edges</th>
<th>Layers</th>
<th>Average degree</th>
</tr>
</thead>
<tbody>
<tr>
<td>YLMN</td>
<td>4458</td>
<td>8 473 997</td>
<td>4</td>
<td>475.21</td>
</tr>
<tr>
<td>HIMN</td>
<td>2400</td>
<td>63 558</td>
<td>2</td>
<td>13.24</td>
</tr>
<tr>
<td>LCMN</td>
<td>2853</td>
<td>914 112</td>
<td>4</td>
<td>80.10</td>
</tr>
</tbody>
</table>
To obtain essential nodes of these three biological multilayer networks, for the YLMN and LCMN, by integrating the following databases: DEG,45 GEO,44 MIPS,46 and DAVID,47 we discover that 386 and 434 genes (nodes) are reported as essential genes (nodes) in many vital biological processes. For the HIMN, to obtain essential proteins (nodes), on the one hand, we selected some essential proteins that were verified to have played important biological roles in inflammatory responses,48–52 On the other hand, some proteins are also selected as essential elements based on Gene ontology (GO) annotation, which reported that these genes are vital in inflammatory responses. GO annotation is an effective tool to explain the vital biological functions of genes, and the detailed introduction can be seen in Refs. 53 and 54. Based on Go annotation, here we selected the proteins related to the important biological functions in inflammatory response (GO:0006954, “inflammatory response”) as essential proteins. In summary, in the HIMN, we selected 61 proteins related to the inflammatory response as essential proteins. In our study, these 386, 434, and 61 nodes are considered essential nodes for comparison in the three biological multilayer networks.

C. Comparison with six published centrality methods

To evaluate the performance of EDCPTD centrality, in this section, we create bar charts of the YLMN, HIMN, and LCMN networks. Here, we calculate the number of essential nodes identified by EDCPTD centrality and six other published centrality methods that include three centrality methods of multilayer structures [degree centrality (DC-M), PageRank centrality (PC-M), and eigenvector centrality (EC-M)] and the corresponding three centrality methods of the aggregated single-layer structures (denoted by DC-S, PC-S, and EC-S, respectively).6 The comparisons of the number and the proportion of identified key nodes are presented in Fig. 4 and Table II, respectively. In Table II, the predicted percentage of true essential nodes in each top n% of ranked proteins, denoted by \( Pr(n) \), is defined as follows:

\[
Pr(n) = \frac{TP(n)}{EP},
\]

where \( TP(n) \) is the number of true predicted essential proteins among the top n% of ranked proteins, and \( EP \) is the number of true essential proteins among all proteins.

For the YLMN, as shown in Table II and Figs. 4(a)–4(d), with the top 30% of nodes selected, the EDCPTD centrality can detect 75.13% of true essential nodes. Compared with the PC-M of the multilayer structures, which has the best performance among six published centrality methods in the YLMN, in each top percentage of the YLMN (15%, 20%, 25% and 30%), the prediction accuracy of EDCPTD centrality is improved by 15.79%, 11.94%, 9.54% and 6.62%. Compared with degree centralities (DC-M or DC-S), which are widely used in identifying essential proteins, the EDCPTD centrality performs well in the YLMN. These results indicate that the prediction accuracy of our proposed method is higher than other published methods in the YLMN.

For the HIMN, as shown in Table II and Figs. 4(e) and 4(f), the EC-S and EC-M methods are all worse than the other indexes. The EDCPTD, DC-S, DC-M, PC-S, and PC-M methods have roughly similar performance. Compared with the
DC-M method of the multilayer structures, which has the best performance among six published centrality methods, in each top percentage of HIMN (15%, 20%, 25% and 30%), the prediction accuracy of EDCPTD centrality is improved by 9.09%, 5.26%, 5.00% and 2.27%, respectively. With the top 30% of nodes selected, the EDCPTD centrality can detect 73.77% of true essential nodes. On the whole, the EDCPTD centrality method is a little better than other six centrality methods in the HIMN.

For the LCMN, as illustrated in Table II and Figs. 4(i)–4(l), all the seven centrality measures have roughly similar performance. With the top 15% and 20% of nodes selected, the prediction accuracy of EDCPTD centrality is a little worse than other six centrality methods. But with the top 25% and 30% of nodes selected, the EDCPTD method has the highest prediction accuracy among other six centrality methods.

In summary, for these three real-world biological networks, the bar chart demonstrates that the proposed approach provides a good alternative index to identify real important nodes in real-world multilayer networks.

D. The performance evaluation based on ROC curves

Furthermore, we use ROC curves and the corresponding areas under the curve (AUC) of ROC to evaluate the performance of each method. For a multilayer network with \( N \) nodes, the procedures of ROC analysis are as follows. Suppose the nodes can be classified into two groups, essential and unessential, and we know the realistic classification. For a new method, the nodes are within the range of \([a, b]\). For any threshold value \( T \in [a, b] \), the procedure of ROC analysis is as follows. Suppose the nodes can be classified into two groups, essential and unessential, and we know the realistic classification. For a new method, the nodes are within the range of \([a, b]\). For any threshold value \( T \in [a, b] \), the nodes are classified into two classes. The actual classification is compared with the new classification. The following two equations are used to quantitatively measure the accuracy of the proposed measure.

\[
F_1 = \frac{N_2}{N_2 + N_4}, \quad F_2 = \frac{N_1}{N_1 + N_3},
\]

where \( N_1 \) is defined as the number of true positive nodes, and the nodes are essential in the two classifications. \( N_2 \) denotes the number of false negative nodes, which are considered unessential in the new classification but are actually essential. Analogously, \( N_2 \) and \( N_4 \) are defined as the number of false positive nodes and true negative nodes, respectively. Here, \( F_1 \) and \( F_2 \) are referred to as false and true positive rates, respectively. For a given \( T \in [a, b] \), plotting the corresponding points \( (F_1, F_2) \) in two dimensional coordinate system, we obtain the ROC curves and AUC values (the corresponding areas under the ROC curve), which can reflect the identified accuracy of the new index.

To evaluate the performance of EDCPTD centrality based on ROC curves, computational analysis is performed using the above three real-world biological networks, including HIMN, YLMN, and LCMN. The ROC curves and corresponding AUC values are shown in Fig. 5 and Table III. In Table III, the AUC values of EDCPTD centrality in the HIMN, YLMN, and LCMN are 0.8259, 0.8062, and 0.7005, respectively, and have the highest prediction accuracy in all centrality methods. Figs. 5(a) and 5(b) indicate that in the HIMN, the EDCPTD, DC-M, PC-M, DC-S, and PC-S are able to identify the essential nodes, and the AUC values of these five indexes are 0.8259, 0.7807, 0.7580, 0.7780, and 0.7613, respectively. The EDCPTD method has greater advantages in other six centrality methods. In the YLMN, as shown in Figs. 5(c) and 5(d), all seven centrality methods have roughly similar performance, and AUC values are all above 0.72. The EDCPTD method is a little better than the other methods. In the LCMN, as shown in Figs. 5(e) and 5(f), on the whole, the AUC value of the EDCPTD method is 0.7005, which has the highest prediction accuracy. In particular, when the false positive rate \( F_1 < 0.2 \), the EDCPTD method is a little worse than other six centrality methods. But when the false positive rate \( F_1 > 0.2 \), the EDCPTD method exhibits good prominent prediction accuracy. Together, ROC analysis indicates that EDCPTD centrality is a good alternative index that can be used to identify essential nodes in multilayer networks.

E. Correlation analysis between the EDCPTD and other six published centrality methods

To further quantify the correlation of ranking between the EDCPTD approach and the six other methods, we apply the Kendall Tau coefficient whose definition can be represented as follows. For a set \( \{1, 2, \ldots, N\} \) of \( N \) nodes in a multilayer network, we consider two ranking \( c_1 = (x_1, x_2, \ldots, x_N) \) and \( c_2 = (y_1, y_2, \ldots, y_N) \), then the Kendall Tau coefficient \( \tau \) is defined as

<table>
<thead>
<tr>
<th>Methods</th>
<th>HIMN</th>
<th>YLMN</th>
<th>LCMN</th>
</tr>
</thead>
<tbody>
<tr>
<td>EDCPTD</td>
<td>45.60%</td>
<td>58.29%</td>
<td>68.39%</td>
</tr>
<tr>
<td>DC-S</td>
<td>36.53%</td>
<td>49.22%</td>
<td>60.10%</td>
</tr>
<tr>
<td>PC-S</td>
<td>34.97%</td>
<td>48.45%</td>
<td>58.55%</td>
</tr>
<tr>
<td>EC-S</td>
<td>30.05%</td>
<td>43.01%</td>
<td>51.55%</td>
</tr>
<tr>
<td>DC-M</td>
<td>36.53%</td>
<td>49.22%</td>
<td>60.10%</td>
</tr>
<tr>
<td>PC-M</td>
<td>39.38%</td>
<td>52.07%</td>
<td>62.44%</td>
</tr>
<tr>
<td>EC-M</td>
<td>28.76%</td>
<td>38.34%</td>
<td>49.48%</td>
</tr>
</tbody>
</table>
where \( \bar{K}(c_1, c_2) \) and \( \tilde{K}(c_1, c_2) \) are the number of concordant pairs and discordant pairs, respectively. \( \binom{N}{2} \) is the number of all possible pairs of nodes. Here, any pair of observations \((x_i, y_i)\) and \((x_j, y_j)\), where \( i \neq j \), are said to be concordant if the ranks for both elements agree: that is, if both \( x_i > x_j \) and \( y_i > y_j \); or if both \( x_i < x_j \) and \( y_i < y_j \). They are said to be discordant, if \( x_i > x_j \) and \( y_i < y_j \); or if \( x_i < x_j \) and \( y_i > y_j \). If \( x_i = x_j \) or \( y_i = y_j \), the pair is neither concordant nor discordant. It is apparent that the coefficient \( \tau \) must be in the range \(-1 \leq \tau \leq 1\). A larger value of \( \tau \) indicates a stronger correlation between the ranking sequences of two centrality methods. The numerical results are presented in Table IV and Fig. 6.

To perform the safe significance analysis, we make the null hypothesis using the one-sample t-test\(^{56}\) and calculate the p-value for the observed value of \( \tau \). Then, we calibrate the p-value to a conservative significance by the

\[
\tau(c_1, c_2) = \frac{\bar{K}(c_1, c_2) - \tilde{K}(c_1, c_2)}{\binom{N}{2}},
\]

## Table III. The AUC values of different centrality methods for three multilayer biological networks.

<table>
<thead>
<tr>
<th>Methods</th>
<th>HIMN</th>
<th>YLMN</th>
<th>LCMN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multilayer networks</td>
<td>EDCPTD</td>
<td>0.8259</td>
<td>0.8062</td>
</tr>
<tr>
<td></td>
<td>DC-M</td>
<td>0.7807</td>
<td>0.7599</td>
</tr>
<tr>
<td></td>
<td>PC-M</td>
<td>0.7580</td>
<td>0.7743</td>
</tr>
<tr>
<td></td>
<td>EC-M</td>
<td>0.5968</td>
<td>0.7222</td>
</tr>
<tr>
<td>Aggregating networks</td>
<td>DC-S</td>
<td>0.7780</td>
<td>0.7599</td>
</tr>
<tr>
<td></td>
<td>PC-S</td>
<td>0.7613</td>
<td>0.7533</td>
</tr>
<tr>
<td></td>
<td>EC-S</td>
<td>0.5876</td>
<td>0.7300</td>
</tr>
</tbody>
</table>

## Table IV. The Kendall Tau coefficient between the EDCPTD and other six centrality indices in three biological multilayer networks.

<table>
<thead>
<tr>
<th>Networks</th>
<th>DC-M</th>
<th>EC-M</th>
<th>PC-M</th>
<th>DC-S</th>
<th>EC-S</th>
<th>PC-S</th>
</tr>
</thead>
<tbody>
<tr>
<td>HIMN</td>
<td>0.4602</td>
<td>0.4268</td>
<td>0.3578</td>
<td>0.4609</td>
<td>0.4235</td>
<td>0.3673</td>
</tr>
<tr>
<td>LCMN</td>
<td>0.2427</td>
<td>0.5045</td>
<td>0.1935</td>
<td>0.2500</td>
<td>0.2938</td>
<td>0.2430</td>
</tr>
<tr>
<td>YLMN</td>
<td>0.6365</td>
<td>0.5670</td>
<td>0.6581</td>
<td>0.6342</td>
<td>0.5284</td>
<td>0.5321</td>
</tr>
</tbody>
</table>

FIG. 5. ROC curves of EDCPTD centrality and six other centrality methods in the HIMN, YLMN, and LCMN. (a) and (b) show the ROC curves of EDCPTD and six other centrality methods in the HIMN. (c) and (d) show the ROC curves of EDCPTD and six other centrality methods in the YLMN. (e) and (f) show the ROC curves of EDCPTD and six other centrality methods in the LCMN.

FIG. 6. Bar charts of the Kendall Tau coefficient between the EDCPTD and other six centrality methods include PC-S, EC-S, DC-S, PC-M, EC-M, and DC-M for the YLMN, LCMN, and HIMN.
method of Sellke et al., which are described in the following.

Here, we assume the Kendall coefficient $\tau$ follows the normal distribution $\tau \sim N(\mu, \sigma^2)$, where the mean $\mu$ and variance $\sigma$ are unknown. According to Table IV, we can obtain six dataset $\tau^{(i)} = \left(\tau_1^{(i)}, \tau_2^{(i)}, \tau_3^{(i)}\right)$, $(i = 1, 2, \ldots, 6)$, where $\tau^{(1)}, \ldots, \tau^{(6)}$ are the sets of Kendall coefficient between six centrality indices and EDCPTD in three biological multi-layer networks, respectively. For each dataset $\tau^{(i)} = \left(\tau_1^{(i)}, \tau_2^{(i)}, \tau_3^{(i)}\right)$, the following null hypothesis $H_0$ and alternative hypothesis $H_1$ are to be tested:

$$H_0 : \tau^{(i)} = 0, \text{ versus } H_1 : \tau^{(i)} \neq 0.$$ 

The meaning of $H_0$ is that the centrality method $i$ is not correlated with the EDCPTD method, and $H_1$ represents that the centrality method $i$ is correlated with the EDCPTD method. We apply the one-sample t-test to calculate the p-value. For each test, the corresponding $p$ value can be calculated by

$$p = 2 \left[ 1 - \Phi \left( \frac{\bar{x}^{(i)} - \mu_0}{\bar{s}/\sqrt{n}} \right) \right],$$

where $n, \bar{s},$ and $\bar{x}^{(i)}$ are the sample size, standard deviation, and sample mean corresponding to the test of $\tau^{(i)}$, respectively, and $\Phi$ is the standard normal cumulative distribution function. Based on the method of Sellke et al., we calibrate the p-value to a conservative significance $\alpha(p)$. The calibration $\alpha(p)$ can be calculated by

$$\alpha(p) = \left(1 + \left[-ep \log (p)\right]^{-1}\right)^{-1}. \quad (15)$$

For all six tested centrality methods, the corresponding $p$ value and $\alpha(p)$ are presented in Table V. Here, we select 0.05 as a threshold for statistical significance. According to Table V, we know that the calibration $\alpha(p)$ of all centrality methods is larger than 0.05. The confidence levels of significance analysis for six centrality methods are 68.31%, 91.83%, 61.86%, 69.45%, 79.63%, and 72.47%, respectively, which are far below the accepted standard 95% to reject the null hypothesis. Therefore, the null hypothesis could not be rejected. In other words, the EDCPTD method is not significantly correlated with other six centrality methods.

### F. The enrichment analysis of the identified top nodes based on the EDCPTD centrality

For the HIMN, YLMN, and LCMN, based on the EDCPTD centrality method, the identified top-40 nodes are shown in Table VI. To further evaluate the performance of the EDCPTD centrality, we used DAVID (Database for Annotation, Visualization and Integrated Discovery) to perform functional enrichment analysis of the GO biological processes in the identified top-40 nodes for the HIMN, YLMN and LCMN. The results are presented in Table VII, and the detailed introduction is given in Appendix.

Based on the results of the analysis in Appendix, we demonstrate that the top nodes identified based on the EDCPTD centrality for the LCMN, HIMN, and YLMN play significant roles in many vital biological processes. These

### TABLE V. p-value and $\alpha(p)$ of six centrality methods.

<table>
<thead>
<tr>
<th>Results</th>
<th>DC-M</th>
<th>EC-M</th>
<th>PC-M</th>
<th>DC-S</th>
<th>EC-S</th>
<th>PC-S</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p$ value</td>
<td>0.0593</td>
<td>0.0065</td>
<td>0.0974</td>
<td>0.0562</td>
<td>0.0257</td>
<td>0.0451</td>
</tr>
<tr>
<td>$\alpha(p)$</td>
<td>0.3129</td>
<td>0.0817</td>
<td>0.3814</td>
<td>0.3055</td>
<td>0.2037</td>
<td>0.2753</td>
</tr>
</tbody>
</table>

### TABLE VI. Top-40 nodes ranked by the EDCPTD centrality method in the HIMN, YLMN, and LCMN.

<table>
<thead>
<tr>
<th>Ranking</th>
<th>HIMN</th>
<th>YLMN</th>
<th>LCMN</th>
<th>Ranking</th>
<th>HIMN</th>
<th>YLMN</th>
<th>LCMN</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>MYC</td>
<td>YER0860W</td>
<td>ARHGAP29</td>
<td>21</td>
<td>PCBP2</td>
<td>YER157W</td>
<td>APIS2</td>
</tr>
<tr>
<td>2</td>
<td>SNRPF</td>
<td>YML0949W</td>
<td>AKAP13</td>
<td>22</td>
<td>OAS3</td>
<td>YGR157W</td>
<td>APB21</td>
</tr>
<tr>
<td>3</td>
<td>IFIT3</td>
<td>YFOLO39C</td>
<td>AK2</td>
<td>23</td>
<td>HNRNPR</td>
<td>YHR194W</td>
<td>ARL4C</td>
</tr>
<tr>
<td>4</td>
<td>HNRNPC</td>
<td>YDR448W</td>
<td>BICD2</td>
<td>24</td>
<td>SRSF2</td>
<td>YJR073C</td>
<td>ASCC3</td>
</tr>
<tr>
<td>5</td>
<td>STAT1</td>
<td>YEL036C</td>
<td>APG1</td>
<td>25</td>
<td>OASL</td>
<td>YPL158C</td>
<td>ATPB4</td>
</tr>
<tr>
<td>6</td>
<td>DHX9</td>
<td>YML128C</td>
<td>ADARB1</td>
<td>26</td>
<td>IRF7</td>
<td>YUL236W</td>
<td>AFF1</td>
</tr>
<tr>
<td>7</td>
<td>IFIT2</td>
<td>YLR078C</td>
<td>AKAP11</td>
<td>27</td>
<td>POLR2B</td>
<td>YHR189W</td>
<td>APB2</td>
</tr>
<tr>
<td>8</td>
<td>IFIT1</td>
<td>YDR123C</td>
<td>ARIDB4</td>
<td>28</td>
<td>PRPF8</td>
<td>YMR083W</td>
<td>ARF6</td>
</tr>
<tr>
<td>9</td>
<td>SRSF7</td>
<td>YDR439W</td>
<td>ADAMB1</td>
<td>29</td>
<td>POLR2F</td>
<td>YDR477W</td>
<td>ATP1B1</td>
</tr>
<tr>
<td>10</td>
<td>SRSF1</td>
<td>YIL909W</td>
<td>ARHGFE3</td>
<td>30</td>
<td>MX1</td>
<td>YPL001W</td>
<td>ATPG3</td>
</tr>
<tr>
<td>11</td>
<td>ISG15</td>
<td>YGL100W</td>
<td>APC</td>
<td>31</td>
<td>XAF1</td>
<td>YPL180W</td>
<td>AKAP2</td>
</tr>
<tr>
<td>12</td>
<td>HNRNPC</td>
<td>YER093C</td>
<td>AZIN1</td>
<td>32</td>
<td>SRSF6</td>
<td>YNL330C</td>
<td>ADAMTS13</td>
</tr>
<tr>
<td>13</td>
<td>SRRM1</td>
<td>YLR242C</td>
<td>AKT3</td>
<td>33</td>
<td>OAS2</td>
<td>YEL050C</td>
<td>ANK2</td>
</tr>
<tr>
<td>14</td>
<td>SRSF1</td>
<td>YJR139C</td>
<td>ATP6VC1C</td>
<td>34</td>
<td>IRF9</td>
<td>YUL124C</td>
<td>BACE1</td>
</tr>
<tr>
<td>15</td>
<td>MYC</td>
<td>YLR200W</td>
<td>AKAP7</td>
<td>35</td>
<td>NF1</td>
<td>YLR298C</td>
<td>ANK51A</td>
</tr>
<tr>
<td>16</td>
<td>HNRNPC</td>
<td>YJR118C</td>
<td>ARID4A</td>
<td>36</td>
<td>HDAC1</td>
<td>YIR023W</td>
<td>ATNX7</td>
</tr>
<tr>
<td>17</td>
<td>SFSB1</td>
<td>YL048W</td>
<td>ANK3</td>
<td>37</td>
<td>IRF1</td>
<td>YML125C</td>
<td>APQ4</td>
</tr>
<tr>
<td>18</td>
<td>SNRP70</td>
<td>YLR314C</td>
<td>ARFGEF1</td>
<td>38</td>
<td>IRF2</td>
<td>YDL192W</td>
<td>ANKRD17</td>
</tr>
<tr>
<td>19</td>
<td>SRSF3</td>
<td>YNL216W</td>
<td>ATP10B</td>
<td>39</td>
<td>POLR2H</td>
<td>YML097C</td>
<td>ARFGEF10</td>
</tr>
<tr>
<td>20</td>
<td>SRSF5</td>
<td>YPL213W</td>
<td>ANP32E</td>
<td>40</td>
<td>EGR1</td>
<td>YBL058W</td>
<td>APIS2</td>
</tr>
</tbody>
</table>
TABLE VII. The enrichment analysis of the identified top-40 nodes for the LCMN, HIMN, and YLMN.

<table>
<thead>
<tr>
<th>Functional annotation</th>
<th>P-value</th>
<th>Count</th>
<th>Representative genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>The LCMN network</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATP biosynthetic process</td>
<td>6.19E-05</td>
<td>5</td>
<td>ATPB4, ATP10B, ATP1B1, ATP5G3, ATP6V1C1</td>
</tr>
<tr>
<td>Purine ribonucleoside triphosphate biosynthetic process</td>
<td>9.02E-05</td>
<td>5</td>
<td>ATPB4, ATP10B, ATP1B1, ATP5G3, ATP6V1C1</td>
</tr>
<tr>
<td>Ribonucleoside triphosphate biosynthetic process</td>
<td>9.36E-05</td>
<td>5</td>
<td>ATPB4, ATP10B, ATP1B1, ATP5G3, ATP6V1C1</td>
</tr>
<tr>
<td>Purine nucleoside triphosphate biosynthetic process</td>
<td>9.36E-05</td>
<td>5</td>
<td>ATPB4, ATP10B, ATP1B1, ATP5G3, ATP6V1C1</td>
</tr>
<tr>
<td>Nucleoside triphosphate biosynthetic process</td>
<td>1.05E-04</td>
<td>5</td>
<td>ATPB4, ATP10B, ATP1B1, ATP5G3, ATP6V1C1</td>
</tr>
<tr>
<td>ATP metabolic process</td>
<td>1.18E-04</td>
<td>5</td>
<td>ATPB4, ATP10B, ATP1B1, ATP5G3, ATP6V1C1</td>
</tr>
<tr>
<td>Phosphoprotein</td>
<td>2.36E-04</td>
<td>27</td>
<td>ARHAGP29, ASCC3, ANP32E, \ldots, ARID4B</td>
</tr>
<tr>
<td>Nitrogen compound biosynthetic process</td>
<td>1.06E-03</td>
<td>6</td>
<td>AZIN1, ATPB4, ATP10B, ATP1B1, ATP5G3, ATP6V1C1</td>
</tr>
<tr>
<td>Computationally biased region:Ser-rich</td>
<td>1.64E-03</td>
<td>3</td>
<td>ARHGEF10, ANK3, ANKRD17, AKAP11, ATXN7</td>
</tr>
<tr>
<td>Regulation of Ras protein signal transduction</td>
<td>1.62E-03</td>
<td>5</td>
<td>ARHGEF3, ARHGEF10, ARFGEF1, AKAP13, ARF6</td>
</tr>
<tr>
<td>Splice variant</td>
<td>3.65E-03</td>
<td>25</td>
<td>ARHAGP29, BICD2, AQP4, \ldots, AKAP7</td>
</tr>
<tr>
<td>Small GTPase regulator activity</td>
<td>6.07E-03</td>
<td>5</td>
<td>ARHAGP29, AKAP13, \ldots, ARFGEF1</td>
</tr>
<tr>
<td>The HIMN network</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nuclear mRNA splicing, via spliceosome</td>
<td>1.54E-32</td>
<td>22</td>
<td>SNRPF, HNRNPC, DHX9, \ldots, SRSF1</td>
</tr>
<tr>
<td>RNA splicing, via transesterification reactions with bulged adenosine</td>
<td>1.54E-32</td>
<td>22</td>
<td>SNRPF, HNRNPC, DHX9, \ldots, SRSF1</td>
</tr>
<tr>
<td>mRNA metabolic process</td>
<td>3.21E-24</td>
<td>22</td>
<td>SNRPF, HNRNPC, DHX9, \ldots, SRSF1</td>
</tr>
<tr>
<td>hsa03040: Spliceosome</td>
<td>1.66E-14</td>
<td>13</td>
<td>SNRPF, HNRNPC, SRSF7, \ldots, SRSF1</td>
</tr>
<tr>
<td>Acetylation</td>
<td>3.96E-11</td>
<td>25</td>
<td>MYC, SNRPF, SRSF3, \ldots, SRSF7</td>
</tr>
<tr>
<td>Nitrogen compound biosynthetic process</td>
<td>1.22E-11</td>
<td>17</td>
<td>MYC, SNRNP200, HNRNPC, \ldots, HNRNPR</td>
</tr>
<tr>
<td>Interferon regulatory factor, conserved site</td>
<td>9.87E-07</td>
<td>4</td>
<td>IRF9, IRF7, IRF1, IRF2</td>
</tr>
<tr>
<td>Antiviral defense</td>
<td>7.40E-06</td>
<td>5</td>
<td>ISG15, STAT1, MX1, IRF9, IRF7</td>
</tr>
<tr>
<td>Response to virus</td>
<td>1.93E-04</td>
<td>5</td>
<td>ISG15, STAT1, MX1, IRF9, IRF7</td>
</tr>
<tr>
<td>The YLMN network</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Golgi apparatus</td>
<td>4.09E-04</td>
<td>6</td>
<td>YEL036C, YLR078C, YER157W, \ldots, YIL048W</td>
</tr>
<tr>
<td>Endoplasmic reticulum</td>
<td>1.12E-04</td>
<td>9</td>
<td>YEL036C, YIL090W, YLR242C, \ldots, YLR078C</td>
</tr>
<tr>
<td>Phospholipid biosynthesis</td>
<td>1.16E-04</td>
<td>3</td>
<td>YDR123C, YGR157W, YJR073C</td>
</tr>
<tr>
<td>Isoleucine biosynthesis</td>
<td>1.18E-03</td>
<td>2</td>
<td>YJR139C, YER086W</td>
</tr>
<tr>
<td>Protein localization</td>
<td>3.10E-03</td>
<td>10</td>
<td>YFL039C, YDR439W, YDL192W, \ldots, YML125C</td>
</tr>
<tr>
<td>Protein transport</td>
<td>3.16E-03</td>
<td>9</td>
<td>YFL039C, YDR439W, YDL192W, \ldots, YML125C</td>
</tr>
<tr>
<td>External encapsulating structure organization</td>
<td>3.95E-03</td>
<td>6</td>
<td>YFL039C, YJR118C, YJR073C, YER093C, YPL180W, YLR314C</td>
</tr>
<tr>
<td>Establishment of protein localization</td>
<td>4.61E-03</td>
<td>9</td>
<td>YFL039C, YDR439W, YDL192W, \ldots, YML125C</td>
</tr>
<tr>
<td>Regulation of transcription from RNA polymerase II promoter</td>
<td>4.75E-03</td>
<td>6</td>
<td>YFL039C, YDR123C, YDR448W, YNL330C, YNL236W, YIR023W</td>
</tr>
<tr>
<td>Positive regulation of biosynthetic process</td>
<td>6.15E-03</td>
<td>5</td>
<td>YDR123C, YDR448W, YNL330C, 852088, YNL236W</td>
</tr>
</tbody>
</table>

Further indicate that the EDCPTD centrality is a good alternative index to identify real important nodes. Furthermore, experimental results also demonstrate that the proposed method is useful in identifying novel disease genes and potential drug targets.

IV. VISUAL SOFTWARE FOR IDENTIFYING THE ESSENTIAL NODES OF MULTILAYER NETWORKS

So that other researchers are able to more conveniently identify essential nodes in multilayer networks, we have implemented many centrality methods for multilayer networks (including our method and other published methods) and created a visual software based on the MATLAB GUI, called ENMNFinder (Essential Nodes of Multilayer Networks Finder). The operating instructions and MATLAB file for ENMNFinder are available at http://maths.whu.edu.cn/shizililiang/lj/2016-07-20/6888.html.

V. CONCLUSION AND DISCUSSION

The identification of essential nodes in complex networks, in particular multilayer networks characterized by different types of interactions, is necessary to understand the topological structure and dynamic processes of complex networks. In this paper, a novel comprehensive centrality method based on CP tensor decomposition, denoted as EDCPTD centrality, was proposed to identify essential nodes in interconnected multilayer networks. The EDCPTD centrality integrates the information of different types of links, including the edges among nodes in the same layer and connections between different layers, to identify essential nodes that bridge together different types of relations.

We used a simple multilayer network and three real-world multilayer biological networks, i.e., the HIMN, YLMN, and CCMN networks, to test the performance of the proposed index. By comparing the different behaviors of the proposed method and other aggregated single-layer methods, we demonstrated that neglecting the multiple relationships of structures between nodes may occasionally result in not fully capturing the details presented in real-world systems, even leading to incorrect identification of the most central nodes. Furthermore, the experimental results of the char bar and ROC curves for three real-world biological networks indicated that the EDCPTD centrality method could achieve...
more remarkable prediction accuracy than the six other centrality methods. In particular, we used DAVID to perform functional enrichment analysis of the identified top nodes based on the EDCPTD, which further indicates that the EDCPTD centrality presents good prediction performance. Finally, based on our method and other published methods, we created a visual software based on MATLAB GUI called ENMNFinder, which can be used by other researchers.

In sum, the proposed centrality measure based on tensor representation presents a computational framework for detecting the importance of nodes in multilayer networks. It provides a good foundation for exploring and analyzing complex multilayer networks based on large data sets.

ACKNOWLEDGMENTS

This work was supported by the Major Research Plan of the National Natural Science Foundation of China (No. 91530320) and the Chinese National Natural Science Foundation (No. 61672388).

APPENDIX: THE RESULTS OF ENRICHMENT ANALYSIS IN THREE BIOLOGICAL MULTILAYER NETWORKS

Based on the identified top-40 nodes for the HIMN, YLMN, and LCMN as shown in Table IV, we used DAVID to perform functional enrichment analysis of the GO biological processes. The results are presented in Table V. According to Table V, in the LCMN network, among these essential components, ATP2B4, ATP10B, ATP1B1, ATP5G3, ATP6V1C1 and ARHGAP29 are significantly enriched in many vital biological processes, including the ATP biosynthetic process, purine ribonucleoside triphosphate biosynthetic process, ribonucleoside triphosphate biosynthetic process, and Phosphoprotein process. AZIN1, ATP2B4, ATP10B, ATP1B1, ATP5G3, and ATP6V1C1 are enriched in the nitrogen compound biosynthetic process; ARHGAP29, AKAP13, ARHGEF3, ARHGEF10, and ARFGF1E are enriched in the Small GTPase regulator activity, etc. Geyik et al. reported that ATP2B4 gene expression and the tumor location of patients have a significant relationship. Yanc et al. reported that the altered expression of the gene ATP10B may play a role in the aggressive phenotype seen in cancer. Scotto et al. identified ATP10B as a key target in cancer. Eling et al. indicated that ATP5G3 was significantly reduced in tumor tissues due to patient heterogeneity. Yi et al. showed ARHGAP29 to be a vital cancer-related target gene. In the YLMN network, YEL036C, YLR078C, YIL090W, YGL100W, YER093C, YJR118C, YIL048W, YLR314C, YER157W, YGR157W, YHR194W, YJR073C, and YDR477W are enriched in the basement membrane; YEL036C, YLM128C, YLR078C, YIL090W, YLR242C, YJR118C, YGR157W, and YJR073C are enriched in the endoplasmic reticulum; YFL039C, YER093C, YJR118C, YLR314C, and YJR073C are enriched in the cell wall organization, etc. In the HIMN network, SNRPF, HNRNPC, SRSF7, SRF1, HNRNPK, SF3B1, SNRPN70, SRSF3, SRSF5, SRSF2, PRPF8, and SNRP200 are enriched in the KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway; OASL, PRPF8, SRSF2, HNRNP, OAS3, PCBP2, SRSF5, SRSF3, HNRNPD, SRSF11, SRRM1, HNRNPK, SRSF1, SRSF7, DHX9, HNRNPC, and SNRPF are enriched in the RNA binding processes; ISG15, STAT1, MX1, IRF9, and IRF7 are enriched in antiviral defense and response to virus processes, etc. Ivan et al. reported that IRF1, IRF7, IRF9, STAT1, ISG15, and IFIT3 are vital IFN-related genes in the highly virulent influenza H3N2 infection.
