Multi-Dimensional HITS Based on Random Walks for Multilayer Temporal Networks

Laishui Lv, Kun Zhang*

School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, Jiangsu, China.

* Corresponding author. Tel.: 18879731910; email: zhangkun@njust.edu.cn

Manuscript submitted May 20, 2019; accepted May 5, 2020.
doi: 10.17706/jcp.15.3.98-105

Abstract: Numerous centrality measures have been established to identify the important nodes in static networks, among them, HITS centrality is widely used as a ranking method. In this paper, we extend the classical HITS centrality to rank nodes in multilayer temporal networks with directed edges. First, we use a sixth-order tensor to represent multilayer temporal network and then introduce random walks in the established sixth-order tensor by constructing six transition probability tensors. Second, we establish tensor equations based on these constructed tensors to obtain six centrality vectors: two for the nodes, two for the layers and two for the time stamps. Besides, we prove the existence of the proposed centrality measure under some conditions. Finally, we experimentally show the effectiveness of the proposed centrality on an synthetic network and a real-world network.

Key words: Multilayer temporal networks, centrality measure, HITS, transition probability tensors.

1. Introduction

Evaluation of nodes centrality is important for many applications, such as link prediction [1] extracting communities in social networks [2]. In the recent years, many ranking methods have been proposed for identifying critical nodes in complex networks, such as eigenvector centrality [3], PageRank [4], and so on.

Eigenvector centrality is one of the most popular ranking algorithm to obtain the key nodes of multilayer networks. In [5], Wang generalized the classical HITS centrality to identify the influential nodes in multilayer biological networks. Tudisco *et al*. [6] established a co-rank method for multilayer networks without interlayer coupling. On the other hand, multidimensional PageRank and HITS centrality based on random walks have been proposed for multilayer networks without interlayer coupling in [7], [8].

Similar to multilayer networks, numerous centrality have been defined for ranking the nodes in temporal networks, such as communicability centrality [9], and closeness [10]. By translating temporal networks into multilayer networks, eigenvector-based ranking method have also been defined in [11].

To the best of our knowledge, none of these ranking algorithms can be directly applied to identify the nodes in networks with multilayer and temporal features (called multilayer temporal networks). Arrigo *et al.* [12] proposed a multi-dimensional HITS centrality (MD-HITS) for multilayer temporal networks by evaluating the Perron eigenvector of a multi-homogeneous map. MD-HITS centrality defines five centrality vectors: two for the nodes, two for the layers and one for the time stamps. However, the relationships between time stamps are neglected.

In this work, we extend the classical HITS centrality to multilayer temporal networks by introducing a random walk in networks, referred to as the HITSRW. To consider relationships between time stamps, we use a sixth-order tensor to represent multilayer temporal networks. We then introduce a random walk in multilayer temporal networks by constructing six transition probability tensors. Based on these six constructed tensors, we propose a set of tensor equations to obtain a new HITS centrality, which we call HITSRW centrality. The proposed centrality defines six centrality vectors: two for the nodes, two for the layers and two for the time stamps. Besides, we prove the existence of HITSRW under some conditions by applying Brouwer fixed point theorem. Finally, we experimentally demonstrate the effectiveness of the proposed ranking algorithm on a synthetic network and a real-world networks.

2. Tensor Representation of Multilayer Temporal Networks

A static monolayer network can be described by a graph $G=(V,E)$ of node set V and edge set E. Moving up in dimensionality, we can define a multilayer network that comprise a fixed set of nodes connected by different types of interrelationships. A multilayer network with L layers and N nodes is represented by a pair $W=(\Phi,C)$, where $\Phi=\{G_1,G_2,\cdots,G_L\}$ is the set of all different layers and

$$C=\left\{E_{\alpha\beta}\subseteq V_\alpha\times V_\beta;\alpha,\beta\in\{1,2,\cdots,L\}\right\},$$

is the set of interconnections between nodes of different layers. Here, each layer is a monolayer $G_k=(V,E_k)$, with $V=\{v_1,v_2,\cdots,v_N\}$ (i.e,. each layer has the same N nodes), and E_k is the set of intralayer connections in the layer G_k.

We further define that the multilayer network W is a multi-layer temporal network if the edges are assigned a time label t, within a discrete time window $T=1,\cdots,n_T$. Thus, each edge in this network is identified by two nodes, two layers and the time when the interaction tales place, i.e., node i on layer α points to node j on layer β at time t. In order to consider the relationships between time s-tamps, we use a sixth order tensor $T=(t_{i\alpha tj\beta\tau})\in R^{N\times L\times n_T\times N\times L\times n_T}$ to represent multilayer temporal network and each element of this tensor A is defined by:

$$t_{i\alpha tj\beta\tau}=\begin{cases}\omega_{i\alpha tj\beta\tau}, & if\ v_{i\alpha t}\to v_{j\beta\tau}, t=\tau\\ \rho, & \alpha=\beta, i=j, t=\tau+1\\ 0, & othersise\end{cases}\tag{1}$$

where $1\le i,j\le N, 1\le\alpha,\beta\le L, 1\le t,\tau\le n_T$, $\omega_{i\alpha tj\beta\tau}$ represents the weight of the edge that node i in layer α at time t points to node j in layer β at time τ. The parameter $\rho>0$ is used to tune interactions between time stamps. $v_{i\alpha t}$ represents node i on layer α at time t.

3. Preliminary

Some notations and preliminary knowledge of tensors are given in this section. Let R be the real field, we call $T=(t_{i\alpha tj\beta\tau})\in R^{N\times L\times n_T\times N\times L\times n_T}$ a real (2,2,2)th order $(N\times L\times n_T)$-dimensional tensor, where $1\le i,j\le N, 1\le\alpha,\beta\le L, 1\le t,\tau\le n_T$.

Similar to the definition of matrix-vector product, the tensor-vector product is defined as follows: Given six vectors $x,y\in R^N, z,w\in R^L$ and $u,v\in R^{n_T}$, $[Tyzwuv]_1$ is vector in R^N such that:

$$([Tyzwuv]_1)_i=\sum_{\alpha tj\beta\tau}t_{i\alpha tj\beta\tau}y_j z_\alpha w_\beta u_t v_\tau,\ 1\le i\le N.$$

The $[Txzwuv]_2$,$[Txywuv]_3$, $[Txyzuv]_4$,$[Txyzwv]_5$ and $[Txyzwu]_6$ have similar definitions.

In this work, we introduce a random walk in a nonnegative sixth order tensor T arising from multilayer temporal networks, and study the probability that we will arrive at any particular node, layer and time stamp as a hub or as an authority. By normalizing the elements of T, we construct the following six tensors $H, A, \bar{H}, \bar{A}, R\ \bar{R}$ with respect to hub and authority of nodes, layers and time stamps, respectively:

$$h_{i\alpha tj\beta\tau} = \frac{t_{i\alpha tj\beta\tau}}{\sum_{i=1}^{N} t_{i\alpha tj\beta\tau}},\ a_{i\alpha tj\beta\tau} = \frac{t_{i\alpha tj\beta\tau}}{\sum_{j=1}^{N} t_{i\alpha tj\beta\tau}}, \bar{h}_{i\alpha tj\beta\tau} = \frac{t_{i\alpha tj\beta\tau}}{\sum_{\alpha=1}^{L} t_{i\alpha tj\beta\tau}},$$
$$\bar{a}_{i\alpha tj\beta\tau} = \frac{t_{i\alpha tj\beta\tau}}{\sum_{\beta=1}^{L} t_{i\alpha tj\beta\tau}}, r_{i\alpha tj\beta\tau} = \frac{t_{i\alpha tj\beta\tau}}{\sum_{t=1}^{n_T} t_{i\alpha tj\beta\tau}},\ \bar{r}_{i\alpha tj\beta\tau} = \frac{t_{i\alpha tj\beta\tau}}{\sum_{\tau=1}^{n_T} t_{i\alpha tj\beta\tau}},$$

here $i, j = 1,2,\cdots,N,\ \alpha,\beta = 1,2,\cdots,L,\ t,\tau = 1,2,\cdots,n_T$.

Note that if $t_{i\alpha tj\beta\tau} = 0$ for all $1 \le i \le N$, we set $h_{i\alpha tj\beta\tau} = 1/N$. The same constructions are for the other five tensors. This is similar to the construction of transition probability matrix.

4. The Proposed Centrality and Iterative Algorithm

In this section, we introduce the proposed centrality (referred to as the HITSRW centrality) for the identification of important nodes layers and time stamps in multilayer temporal networks based on random walks.

Suppose that vectors x, z and u are hub centrality values for nodes, layers and time stamps, respectively. Vectors y, w and v are authority centrality values for nodes, layers and time stamps, respectively. Then, these six centrality vectors can be obtained by solving following tensor equations

$$\begin{aligned} x_i &= \textstyle\sum_{\alpha tj\beta\tau} h_{i\alpha tj\beta\tau} y_j z_\alpha w_\beta u_t v_\tau,\ y_j = \sum_{i\alpha t\beta\tau} a_{i\alpha tj\beta\tau} x_i z_\alpha w_\beta u_t v_\tau, \\ z_\alpha &= \textstyle\sum_{itj\beta\tau} \bar{h}_{i\alpha tj\beta\tau} x_i y_j w_\beta u_t v_\tau,\ w_\beta = \sum_{i\alpha tj\tau} \bar{a}_{i\alpha tj\beta\tau} x_i y_j z_\alpha u_t v_\tau, \\ u_t &= \textstyle\sum_{i\alpha j\beta\tau} r_{i\alpha tj\beta\tau} x_i y_j z_\alpha w_\beta v_\tau,\ v_\tau = \sum_{i\alpha tj\beta} \bar{r}_{i\alpha tj\beta\tau} x_i y_j z_\alpha w_\beta u_t, \end{aligned} \tag{2}$$

where $1 \le i, j \le N, 1 \le \alpha, \beta \le L, 1 \le t, \tau \le n_T$. Equivalently, according to the definition of tensor-vector product, (2) can be reformulated as the following tensor equations:

$$\begin{aligned} &[Hyzwuv]_1 = x, [\bar{H}xywuv]_3 = z, [Rxyzwv]_5 = u \\ &[Axzwuv]_2 = y, [\bar{A}xyzuv]_4 = w, [\bar{R}xyzwu]_6 = v \end{aligned} \tag{3}$$

with $x, y \in \Omega_N, z, w \in \Omega_L, u, v \in \Omega_{n_T}$, where $\Omega_N = \{x = (x_1, x_2, \cdots, x_N) \in R^N | x_i \ge 0, 1 \le i \le N, \sum_{i=1}^{N} x_i = 1\}$, $\Omega_L = \{z = (z_1, z_2, \cdots, z_L) \in R^L | z_i \ge 0, 1 \le i \le L, \sum_{i=1}^{L} z_i = 1\}$ and $\Omega_{n_T} = \{u = (u_1, u_2, \cdots, u_{n_T}) \in R^{n_T} | u_i \ge 0, 1 \le i \le L, \sum_{i=1}^{n_T} u_i = 1\}$.

For simplicity, we will drop the bracket and the number ($[\cdot]_1, [\cdot]_2, [\cdot]_3, [\cdot]_4, [\cdot]_5$ and $[\cdot]_6$) for the above tensor-product operations in the following section.

Theorem 1: Suppose $H, A, \bar{H}, \bar{A}, R\ \bar{R}$ are constructed in Section 3, then there exist non-zero non-negative vectors $\bar{x}, \bar{y} \in \Omega_N, \bar{z}, \bar{w} \in \Omega_L, \bar{u}, \bar{v} \in \Omega_{n_T}$ such that $H\bar{y}\bar{z}\bar{w}\bar{u}\bar{v} = \bar{x}$, $\bar{H}\bar{x}\bar{y}\bar{w}\bar{u}\bar{v} = \bar{z}$, $R\bar{x}\bar{y}\bar{z}\bar{w}\bar{v} = \bar{u}$, $A\bar{x}\bar{z}\bar{w}\bar{u}\bar{v} = \bar{y}$, $\bar{A}\bar{x}\bar{y}\bar{z}\bar{u}\bar{v} = \bar{w}$ and $\bar{R}\bar{x}\bar{y}\bar{z}\bar{w}\bar{u} = \bar{v}$.

Proof. The proof is similar to Theorem 2 in [8].

4.1. The Proposed Iterative Algorithm

In order to obtain $\bar{x}, \bar{y}, \bar{z}, \bar{w}, \bar{u}, \bar{v}$ for the centrality scores of nodes, layers and time stamps, in this subsection, we establish an efficient iterative algorithm for computing the tensor equations (3). The iterative algorithm is given below.

Algorithm 1: Computing HITSRW centrality

1: **Input**: Given six transition probability tensors $H, A, \bar{H}, \bar{A}, R\ \bar{R}$, five initial vectors $y_0 \in \Omega_N$, $z_0, w_0 \in \Omega_L$ and $u_0, v_0 \in \Omega_{n_T}$, the tolerance ϵ.

2: **Output**: Six centrality vectors $\bar{x}, \bar{y}$(hub and authority centrality scores of nodes), $\bar{z}, \bar{w}$(hub and authority centrality scores of layers) and $\bar{u}, \bar{v}$ (hub and authority centrality scores of time stamps).

3: Set $t = 1$;

4: Compute $x_t = Hy_{t-1}z_{t-1}w_{t-1}u_{t-1}v_{t-1}$;

5: Compute $y_t = Ax_t z_{t-1}w_{t-1}u_{t-1}v_{t-1}$;

6: Compute $z_t = \bar{H}x_t y_t w_{t-1}u_{t-1}v_{t-1}$;

7: Compute $w_t = \bar{A}x_t y_t z_t u_{t-1}v_{t-1}$;.

8: Compute $u_t = Rx_t y_t z_t w_t v_{t-1}$;

9: Compute $v_t = \bar{R}x_t y_t z_t w_t u_t$;

10: If $\|x_t - x_{t-1}\|_1 + \|y_t - y_{t-1}\|_1 + \|z_t - z_{t-1}\|_1 + \|w_t - w_{t-1}\|_1 + \|u_t - u_{t-1}\|_1 + \|v_t - v_{t-1}\|_1 < \epsilon$, then stop, otherwise set $t = t + 1$ and go to step 4.

Now, we establish the convergence of Algorithm 1 in following theorem, whose proof is similar to Theorem 3 in [8].

Theorem 3: Suppose $H, A, \bar{H}, \bar{A}, R\ \bar{R}$ are constructed in Section 3, let $[\bar{x}, \bar{y}, \bar{z}, \bar{w}, \bar{u}, \bar{v}]$ be a solution to tensor equations(5) and $[x_t, y_t, z_t, w_t, u_t, v_t]$ is generated by Algorithm 1, then we have

$$\lim_{n\to\infty}[x_t, y_t, z_t, w_t, u_t, v_t] = [\bar{x}, \bar{y}, \bar{z}, \bar{w}, \bar{u}, \bar{v}],$$

for a given initial point $[y_0, z_0, w_0, u_0, v_0]$.

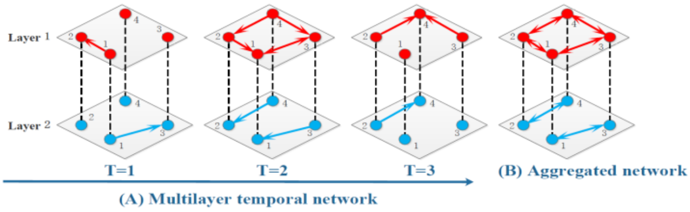
5. Numerical Experiments

In this section, we demonstrate the effectiveness of our proposed centrality by carrying out a number of numerical experiments on a sample multilayer temporal network and a citation network [12] and then make a comparison with the other existing centrality measures to understand the different behaviors. The proposed ranking method was compared with Multi-Dimensional HITS (MD-HITS) [12], PageRank centrality (called T-PageRank) [11], Singular Vector of Tensor (SVT) centrality [5], where MD-HITS gives the hub and authority values of nodes and layers. The value of ρ is set to 1/2 in HITSRW and T -PageRank algorithms.

5.1. Synthetic Multilayer Temporal Network

In this subsection, we compare the different behaviors with T-PageRank and SVT centrality measures by constructing a simple directed multilayer temporal network and the corresponding aggregated network (shown in Fig. 1(A) and Fig. 1(B), respectively). This synthetic network comprises four individual nodes connected by two different layers during three time stamps. Note that the T-PageRank and SVT are used to rank the nodes in aggregated network, because these two centrality measures cannot be directly used to identify the important nodes in multilayer temporal network.

The ranking results are shown in Fig. 1(C) when HITSRW, T-PageRank and SVT centrality measures are applied. Fig. 1(C) reveals that node 4 has the highest hub value. This is not surprising, because node 4 at time layer 2 has the largest number of outdegree and the time layer 2 has a higher hub value (i.e., 0.55). We have also observed that nodes 1, 2, 3 and 4 have the different hub values in the synthetic multilayer temporal network. However, nodes have the same hub score when SVT is applied for aggregated network. This indicates that the use SVT may not be effective when the multilayer temporal network is aggregated into a multilayer network. On the other hand, by analyzing the ranking results of nodes authority, we arrived to the same conclusion (i.e., the use SVT and T-PageRank may not be effective when multilayer temporal network is aggregated).



(C)

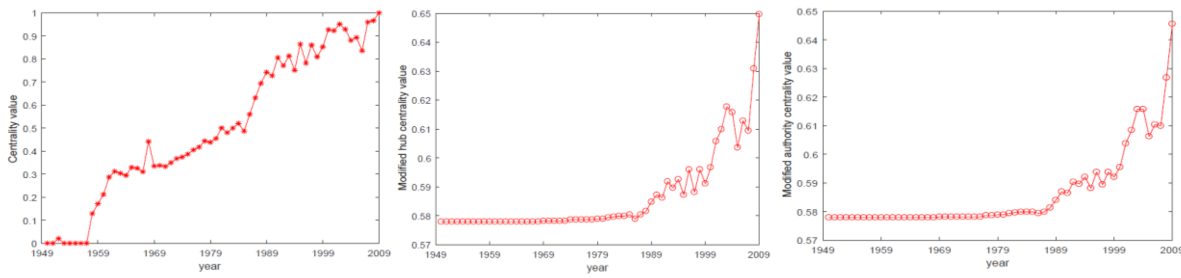
Nodes	Ranking (Hub)		Ranking (Authority)		
	HITSRW	SVT	HITSRW	SVT	T-PageRank
1	2	1	1	1	1
2	4	1	4	1	1
3	3	1	2	1	1
4	1	1	3	1	1

Fig. 1. (A) A simple multilayer temporal network. (B) The corresponding aggregated network. (C) The ranking results of nodes.

5.2. Multilayer Temporal Citation Network

Now, we test out our proposed centrality and the MD-HITS on a large real-world network of scientific publications [12]. This network represents the citations of authors (i.e., nodes) via different journals (i.e., layers) during a period of time (i.e., time stamps) and it contains 592373 nodes, 12604 layers and 63 time stamps. $t_{i\alpha t j\beta\tau}$ is set to 1 if node i authored a paper in journal α at time t cites at least a paper authored by node j in journal β at time τ, otherwise 0.

As there are no objective criteria to rely on, the quality of centrality is hardly quantifiable. Moreover, comparing the ranking results is open to interpretation and need application specific expertise. For example, in [13], protein essentiality was used as a target for centrality measures. For the citation network, it has two important features: (1) the volume of papers published per year has recently considerably increased, (2) research papers are far more easily accessible now than in earlier times, making it easier for researchers to cite each other. We hope that the proposed centrality can capture these features of citation network. We show the centrality values of time stamps in Fig. 2 when HITSRW and MD-HITS are used for citation network. For the HITSRW centrality, the hub and authority centrality values of time stamps (i.e., u_t and v_t, $1 \leq t \leq n_T$) are quite small. In order to visualize the trend of time stamps hub and authority centrality, we enlarge the these value with 1014times(i.e., $u_t = 1014 * u_t$ and $v_t = 1014 * v_t, 1 \leq t \leq n_T$). As we can see, in Fig. 2(b) and Fig. 2(c), the HITSRW can capture the features of citation network that recent years have higher centrality values (hub and authority) than the earlier ones. The MD-HITS centrality also displays the features of citation network (see Fig. 2(a)).

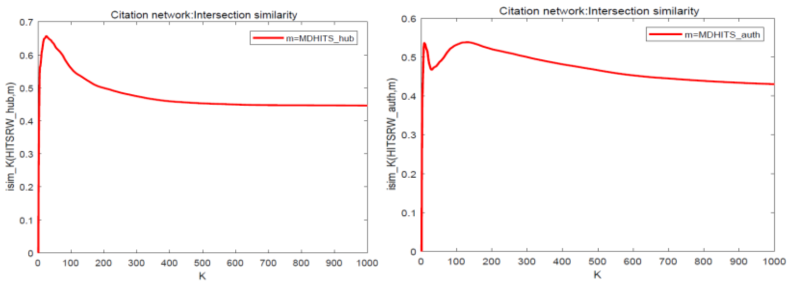


(a) MD-HITS: Time stamps centrality (b) HITSRW: Time stamps hub centrality (c) HITSRW: Time stamps authority centrality

Fig. 2. Centrality values of time stamps in citation network.

We now analyze the linear correlation between the HITSRW and MD-HITS by introducing the Pearsons correlation coefficient. For the nodes hub, the Pearson correlation coefficient value is equal to 0.66 between the HITSRW and MD-HITS. For the nodes authority, the Pearson correlation coefficient value is equal to 0.61 between the HITSRW and MD-HITS. This indicates that there exists a linear correlation between the HITSRW and MD-HITS.

To further compare the ranking results of nodes between HITSRW and MD-HITS centrality by introducing intersection similarity [14]. In Fig. 3, we show the values of intersection similarity versus K when HITSRW and MD-HITS are applied to citation network. As we can see, in Fig. 3(a), the value of intersection similarity between HITSRW and MD-HITS is equal to 0 for nodes hub centrality when $K = 1$. However, it grows quickly and reaches high value (around 0.65). This means that the top ranking of nodes in these ranking methods are different significantly for large values of K Fig. 3(b) shows the similar phenomenon for nodes authority centrality.



(a) Nodes hub centrality (b) Nodes authority centrality

Fig. 3. Intersection similarity between HITSRW and MD-HITS, K=1000.

From the above results, the ranking results of nodes in multilayer temporal citation network obtained by HITSRW is quite different from MD-HITS. This is because the basic ideas of the two ranking methods are different. We define the HITSRW centrality by introducing the random walks in networks and consider the relationships between time stamps. In fact, it is hard to judge whether one centrality outperforms another. Comparing the ranking results is open to interpretation and need application specific expertise.

6. Conclusion

In this paper, we extended the classical HITS centrality to rank the nodes, layers and time stamps in multilayer temporal networks. In order to consider the relationships between time stamps, we used a sixth-order tensor to represent multilayer temporal networks. After the normalization of this tensor elements, we got six transition probability tensors. Then, based on these constructed transition probability tensors, we proposed a ranking model to obtain a new centrality measure, which we called HITSRW centrality. The proposed centrality defines six centrality vectors: two for the nodes (i.e., hub and authority scores for nodes), two for the layers (i.e., hub and authority scores for layers) and two for the time stamps (i.e., hub and authority scores for time stamps). We also proved theoretically the existence of the proposed centrality. The results of the numerical experiments showed the scalability and efficiency of the proposed ranking method.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Laishui LV conducted the research and wrote the paper. Kun Zhang analyzed the data. All authors had approved the final version.

Acknowledgment

This research is funded by the National Natural Science Foundation of China (No.11701097,11661007), Natural Science Foundation of Jiangxi Province (No.20181BAB201007), Research Funds for the Central Universities (No.30918012204), and Postgraduate Research & Practice Innovation Program of Jiangsu Province (No. SJCX18_0149).

References

[1] Liu, H., Hu Z., Haddadi, H., *et al.* (2013). Hidden link prediction based on node centrality and weak ties. *Europhysics Letters, 101(1)*, 18004.

[2] Kanawati, R. (2012). Seed-centric approaches for community detection in complex networks. *Proceedings of the 6th International Conference on Social Computing and Social Media* (pp. 197-208).

[3] Bonacich, P. (1987). Power and centrality: A family of measures. *American Journal of Sociology, 92(5)*, 1170-1182.

[4] Agryzkov, T., Oliver J. L., Tortosa, L., *et al.* (2012). An algorithm for ranking the nodes of an urban network based on the concept of PageRank vector. *Applied Mathematics and Computation, 219(5)*, 2186-2193.

[5] Wang, D., & Zou, X. (2018). A new centrality measure of nodes in multilayer networks under the framework of tensor computation. *Applied Mathematical Modelling, 54*, 46-63.

[6] Tudisco, F., Arrigo, F., & Gautier, A. (2017). Node and layer eigenvector centralities for multiplex networks. *SIAM Journal on Applied Mathematics, 78(2)*, 853-876.

[7] Li, X., Ng, M. K., & Ye, Y. (2011). HAR: Hub, authority and relevance scores in multi-relational data for query search. *Proceeding of the 12th SIAM International Conference on Data Mining* (pp. 1217-1225).

[8] Ng, M. K., Li, X., & Ye, Y. (2012). MultiRank: Co-ranking for objects and relations in multi-relational data, *Proceedings of the 17th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (pp. 141-152).

[9] Grindrod, P., Parsons, M. C., Higham, D. J., & Estrada, E. (2011). Communicability across evolving networks. *Physical Review E, 83*, 046120.

[10] Williams, M. J., & Mirco. M. (2016). Spatio-temporal networks: Reachability centrality and robustness. *Royal Society Open Science, 3(6)*, 160196.

[11] Taylor, D., Myers, S. A., Clauset, A., *et al.* (2015). Eigenvector-based centrality measures for temporal networks. *Physics, 15(1)*, 537-574.

[12] Arrigo, F., & Tudisco, F. (2018). Multi-Dimensional, Multilayer, Nonlinear and Dynamic HITS. Retrieved from http://10.1137/16M1066142

[13] Estrada, E. (2015). Virtual identification of essential proteins within the protein nteraction network of yeast. *Proteomics, 6(1)*, 35-40.

[14] Fagin, R., Kumar, R., & Sivakumar, D. (2003). Comparing top k lists. *SIAM Journal on Discrete Mathematics, 17(1)*, 134-160.

Laishui Lv received the B.S and M.S from the Gannan Normal University. He is currently pursuing the Ph.D. degree with the Nanjing University of Science and Technology, Nanjing, China.

His research interests are focused on artificial intelligence and complex networks.

Kun Zhang received the B.S., M.S., and Ph.D. from the Nanjing University and Science and Technology. Her research interests are focused on big data, pattern recognition, medical engineering, complex networks, network security and trusted computing.