

基于网络超链接信息熵的节点重要性 序结构演化建模分析

胡 钢¹, 牛 琼¹, 许丽鹏¹, 卢志宇¹, 过秀成²

(1. 安徽工业大学管理科学与工程学院, 安徽马鞍山 243032; 2. 东南大学交通学院, 江苏南京 210096)

摘 要: 动态复杂网络在时空演化过程中, 网络节点重要性层内交互关系和层间耦合关系可以更为准确对时序网络节点序结构演化进行分析. 本文提出基于网络超链接信息熵的节点重要性序结构演化模型. 分析时序网络层内节点超链接信息熵重要性排序结果, 得到时序网络节点相邻时间层与跨时间层节点重要性排序模型. 节点超链接信息熵总结相邻时间层与跨时间层节点相似性耦合效应. 通过 SIR (Susceptible Infected Recovered) 模型检验节点传播效率进行实证网络仿真, 结果与经典时序网络模型相比, 本文模型 Kendall's τ 值在各时间层均有提高, 最高为 11.310%.

关键词: 动态时序网络; 超链接信息熵; 节点重要性; 序结构; 层内相似; 层间耦合

中图分类号: TP393 **文献标识码:** A **文章编号:** 0372-2112(2022)11-2638-07

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.12263/DZXB.20211307

The Model to Analyses of Node Importance Order Structure Evolution Based on Network Hyperlink Information Entropy

HU Gang¹, NIU Qiong¹, XU Li-peng¹, LU Zhi-yu¹, GUO Xiu-cheng²

(1. School of Management Science and Engineering, Anhui University of Technology, Maanshan, Anhui 243032, China;

2. School of Transportation, Southeast University, Nanjing, Jiangsu 210096, China)

Abstract: In the time evolution of dynamic complex networks, the intra-layer interaction relationship and inter-layer coupling relationship of the importance of network nodes can be more accurately analyzed for the evolution of the node order structure of the time ordered networks. In this paper, a model of node importance order structure evolution based on network hyperlink information entropy is proposed. Through the results of the importance ranking of node hyperlink information entropy within the temporal network layer, the importance ranking model of nodes in the adjacent temporal layer and across temporal layers of the temporal network is analyzed. The entropy of node hyperlink information summarizes the coupling effect of similarity between adjacent temporal layers and inter-temporal layers. An empirical network simulation was conducted to check the node propagation efficiency by SIR (Susceptible Infected Recovered) model. Compared with the classical temporal network model, the Kendall's τ value of this model was improved in all time layers, with the highest value of 11.310%.

Key words: dynamic temporal networks; hyperlink information entropy; node importance; ordered structure; intra-layer similarity; interlayer coupling

1 引言

研究分析时序网络节点动态交互过程、网络节点属性特征分析和网络节点重要性辨识, 能够更加准确地刻画现实网络中如社交通讯、电力输送、疾病传播对人类行为的影响等^[1-5]复杂系统的交互关系. 静态网络

中节点重要性排序理论有多种, 如度中心性^[6]、介数中心性^[7]、紧密度中心性^[8]、特征向量中心性^[9, 10]等. 曹玖新等^[11]引入 K-核分解概念将节点层次性划分, 根据影响半径参数找出影响力节点合集, 利用启发式核覆盖算法对网络节点重要性进行排序. Liu 等^[12]分析节点及

收稿日期: 2021-09-23; 修回日期: 2022-01-08; 责任编辑: 覃怀银

基金项目: 国家自然科学基金(No.71772002); 安徽省自然科学基金(No.2108085MG236); 安徽省高校自然科学基金项目(No.KJ2021A0385); 安徽省高校研究生科学研究项目(No.YJS20210356)

其邻居节点的重要性,提出邻域中心性重要性评价指标,确定网络最具影响力传播者.

时序网络不同于静态网络,网络节点交互信息与关联关系伴随时间演进会非规律性重现、消失及再生.因此,时序网络节点重要性辨识与静态网络节点重要性辨识理论方法相比存在较大差异.时序网络中相似性度量是静态网络相似性度量理论的推广.静态网络中相似性度量方法有共同邻居指标^[13]、Adamic-Adar指标^[14]、资源分配指标^[15]等.Taylor等^[16]考虑使用多层耦合网络分析时序网络按层内关系和层间关系建立超邻接矩阵,定义基于特征向量中心性指标和节点重要性随时间波动的评判指标,其中SAM(Supra-Adjacency Matrix)指标中可调参数的使用对层间节点重要性排序结果有明显影响.杨剑南等^[17]结合多层耦合网络提出基于节点层间相似性的超邻接矩阵时序网络节点重要性识别方法,与经典的SAM方法相比,作者使用邻居拓扑重叠系数对层间关系进行度量,进行相邻层间相似性计算.胡钢等^[18]提出基于时序网络层间同构率动态演化的超邻接矩阵建模重要节点辨识方法,解决时序网络跨层关系对节点重要性的影响.

基于以上分析,本文提出基于时序网络超链接信息熵的节点重要性序结构演化(Ordered structure-based Supra-Adjacency Matrix, OSAM)模型,利用节点在时间层网络中超链接信息熵排序结果,得到节点在不同时间层内的排序变化,分析节点活跃性随时间变化的不同时间层路径演化轨迹,迭代计算时序网络节点活跃重要性程度.通过对现实网络进行SIR传播试验检测OSAM方法的有效性.

2 理论基础

2.1 时序网络符号描述

时序网络的拓扑结构带有时间属性,可用三元组 $e_{ij} = (i, j, t)$ 表述,表示节点 i 和节点 j 在时间 t 时刻发生一次交互.将时序网络整个观察期 $[0, m]$ 按相等的时间间隔切分为 T 个时间窗口,每个时间窗口的大小为 $\tau = m/T$,时序网络被切分为 T 个离散有序的时间网络层 G_1, G_2, \dots, G_T .

2.2 时序网络多属性定义描述

节点邻接信息熵系数源自文献[19]提出网络节点重要性排序相关系数,获取节点交互关系,计算节点邻接度和概率函数,进而求得节点邻接信息熵,网络中节点信息熵重要性排序表示为

节点邻接度:

$$Q_i = \sum_{\omega \in \Gamma_i} k_{\omega} \quad (1)$$

其中, k_{ω} 为节点 v_i 的度值, Γ_i 为节点 v_i 的邻居节点集合.

节点概率函数:

$$p_i = \frac{k_i}{Q_i}, j \in \Gamma_i \quad (2)$$

其中, k_i 为节点 v_i 的度值.

节点邻接信息熵:

$$H_i = - \sum_{j \in \Gamma_i} (p_i \log_2 p_i) p_j \quad (3)$$

其中, p_i 为节点 v_i 的概率函数, p_j 为节点 v_j 的概率函数, Γ_i 为节点 v_i 邻居节点的集合, Γ_j 为节点 v_j 邻居节点的集合.

2.3 经典SAM时序网络模型分析

文献[16]将时序网络通过层内交互关系和层间耦合关系综合表示,提出超邻接矩阵SAM模型具体表示如下:

$$A = \begin{bmatrix} A^{(1)} & \omega I & \mathbf{0} & \dots \\ \omega I & A^{(2)} & \omega I & \ddots \\ \mathbf{0} & \omega I & A^{(3)} & \ddots \\ \vdots & \ddots & \ddots & \ddots \end{bmatrix} \quad (4)$$

其中, A 表示经典时序网络模型; $A^{(1)}, A^{(2)}, \dots, A^{(T)}$ 表示层内连接关系,依次位于SAM的对角线上,表示有序时间层网络;定义 $a_{ij}(t)$ 为邻接矩阵 $A^{(1)}$ 中的元素,则 $a_{ij}(t) = 0$ 表示在时间层网络 G_t 中节点 i 与节点 j 无连接, $a_{ij}(t) = 1$ 表示在时间层网络 G_t 中节点 i 与节点 j 有连接关系; ωI 表示层间连接关系,其中 ω 为可调参数, I 为 $N \times N$ 单位矩阵. SAM模型节点相邻层间耦合关系使用统一参数 $\omega \in [0.1, \dots, 1]$ 进行调整.

文献[17]提出基于层间相似性SSAM(Similarity-based Supra-Adjacency Matrix)时序网络模型,使用邻居拓扑重叠系数来确定层间连接关系,减少可调参数使用对时序网络层间耦合关系的影响.节点 i 在时间层网络 G_t, G_{t+1} 间的邻居拓扑重叠系数具体形式如下:

$$C_i^{(t,t+1)} = \frac{\sum_j a_{ij}(t) a_{ij}(t+1)}{\sqrt{\left[\sum_j a_{ij}(t) \right] \left[\sum_j a_{ij}(t+1) \right]}} \quad (5)$$

改进后基于层间相似性SSAM时序网络具体表现形式如下:

$$A = \begin{bmatrix} A^{(1)} & C^{(1,2)} & \mathbf{0} & \dots \\ C^{(1,2)} & A^{(2)} & C^{(2,3)} & \ddots \\ \mathbf{0} & C^{(2,3)} & A^{(3)} & \ddots \\ \vdots & \ddots & \ddots & \ddots \end{bmatrix} \quad (6)$$

文献[18]提出基于层间同构率的ISAM(Isomorphism rate based Supra-Adjacency Matrix)时序网络模型,定义动态时序网络综合逼近关系系数,在计算网络节点层间关系与跨层关系时考虑网络节点相邻层与节

点跨层的时序演化效应,具体表现形式如下:

$$A = \begin{bmatrix} A^{(1)} & \mathbf{Z}^{(1,2)} & \mathbf{Z}^{(1,3)} & \dots \\ \mathbf{Z}^{(1,2)} & A^{(2)} & \mathbf{Z}^{(2,3)} & \ddots \\ \mathbf{Z}^{(1,3)} & \mathbf{Z}^{(2,3)} & A^{(3)} & \ddots \\ \vdots & \ddots & \ddots & \ddots \end{bmatrix} \quad (7)$$

其中, $\mathbf{Z}^{(1,2)}, \mathbf{Z}^{(2,3)}, \dots$ 表示相邻层间逼近关系, $\mathbf{Z}^{(1,3)}, \dots$ 表示非相邻层之间逼近关系; $\mathbf{Z}^{(1,2)}$ 为 $N \times N$ 的对角矩阵, 即 $\mathbf{Z}^{(1,2)} = \text{diag}(z_1^{(1,2)}, z_2^{(1,2)}, \dots, z_N^{(1,2)})$, $z_i^{(1,2)}$ 即为节点层间逼近关系系数, 表征节点 i 的层间同构率.

3 时序网络节点重要性排序模型构建

节点邻接信息熵为节点在静态网络中重要性排序提供准确描述, 延伸至时序网络中, 节点在相邻层间和跨层间交互关系随时间中断和重连, 静态网络无法描述网络节点交互过程. 如图 1 所示, 节点 1 在时序网络中的第一层内与节点 2 存在交互关系, 但在第二层内节点 1 无任何交互关系变成孤立节点. 节点 1 在后续时间层内的交互关系处于连接与断开的动态变化中, 这种演化现象符合现实网络节点的动态活跃现象.

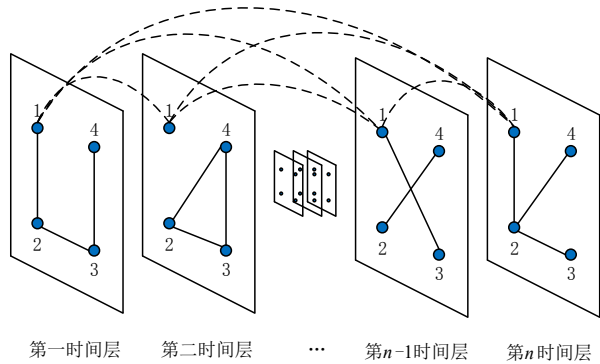


图 1 时序网络节点跨时间层耦合关系演化示意图

3.1 时序网络节点超链接信息熵定义

节点邻接信息熵描述节点与其直接相邻和间接相邻节点之间的交互关系. 但在时序网络中时间维度被分割并且压缩, 进而节点间的交互关系不完全等同于静态网络, 基于以上的信息, 本文提出动态时序网络节点超链接信息熵定义, 具体表示如下:

定义 1 时序网络节点超链接信息熵. 时序网络节点信息熵考虑网络时间层内多级节点间交互关系是否存在. 节点邻接度和节点概率函数可参考静态网络, 具体形式如下:

$$H_i^t = - \sum_{j \in \Gamma_i} (p_i \log_2 p_i) p_j \quad (8)$$

其中, p_i 为节点 v_i 的概率函数, p_j 为节点 v_j 的概率函数, Γ_i 为节点 v_i 邻居节点的集合, Γ_j 为节点 v_j 邻居节点的集合. 表示节点 i 在 t 时间层内超链接信息熵.

3.2 时序网络节点重要性排序定义

分层时序网络中, 层内和层间节点在其对应的时间间隔内都可能与其他节点存在交互, 如果只考虑本层间节点交互信息, 不考虑跨层节点耦合信息, 会降低节点在每个时间层内的重要性排序. 基于以上的信息, 本文提出时序网络节点超链接信息熵在动态网络相邻层内排序定义, 具体描述如下:

定义 2 时序网络相邻层间节点超链接排序系数. 时序网络相邻层间节点排序系数需考虑节点自身和其邻居节点的影响, 即在相邻层间节点交互信息越频繁, 时序节点超链接信息熵值越大, 节点在层内的排序越高, 将超链接信息熵作为节点重要性排序评判指标, 具体形式如下:

$$o_i^{(t,t+1)} = \frac{\sqrt{\left[\frac{1}{L_i(t)} \times \frac{1}{t} + \frac{1}{L_i(t+1)} \right] \times \frac{1}{(t+1)}}}{1 + \frac{|L_i(t) - L_i(t+1)|}{\text{Max}(L_i(t)) - \text{Min}(L_i(t))}} \quad (9)$$

时序网络相邻层间节点排序系数描述序结构在相邻层间持续迭代的递延关系. 其中, $L_i(t), L_i(t+1)$ 对应相邻时间网络层 G_t, G_{t+1} 的节点熵值层间排序结果, $L_i(t)$ 的值域为 $[1, N]$, N 表示网络中节点数目. 不妨对节点超链接信息熵值取倒数表示该节点在层内的节点重要性程度, $1/L_i(t)$ 值域在 $[0, 1]$ 之间; 相邻层内节点在第一层的重要性程度为 $1/L_i(t)$, 第二层内该节点的重要性程度相比于第一层该节点在本层的影响力由于时间的推进有所下降, 因此, 节点在每层的重要性程度乘以比较层的间隔层数. $|L_i(t) - L_i(t+1)|$ 表示相邻层间的节点排序变化系数. $\text{Max}(L_i(t)) - \text{Min}(L_i(t))$ 是对节点排序变化的归一化处理, 其中 $\text{Max}(L_i(t))$ 表示该层节点熵值排序的最大值. 观测节点在整个时间段或是所有时间层内变化, 相邻层的递阶关系无法表征节点多层递阶关系, 由此节点在整个时间段内重要性排序有较大影响. 基于以上分析, 本文提出时序网络节点跨层超链接信息熵排序系数定义.

定义 3 时序网络跨层节点超链接排序系数. 时序网络中每个活跃节点交互关系随着时间不断改变, 为更能准确地表述节点在整个时间段内节点重要性序结构变化, 本文给出时序网络跨层节点超链接排序系数, 具体形式如下:

$$o_i^{(t,t+n)} = \frac{\sqrt{\left[\frac{1}{L_i(t)} \times \frac{1}{t} + \frac{1}{L_i(t+n)} \right] \times \frac{1}{t+n}}}{1 + \sum_{\tau=1}^n \frac{|L_i(t) - L_i(t+\tau)|}{\text{Max}(L_i(t)) - \text{Min}(L_i(t))}} \quad (10)$$

如果 $n=1$ 表示相邻时间层, $o_i^{(t,t+n)}$ 即转变为定义 2 中的式(9).

如果 $n=2$ 表示跨一层, 即第 1 层至第 3 层节点排序关系:

$$o_i^{(1,3)} = \frac{\sqrt[3]{\left(\left(\frac{1}{L_i(1)} \times \frac{1}{1} + \frac{1}{L_i(2)}\right) \times \frac{1}{2} + \frac{1}{L_i(3)}\right) \times \frac{1}{3}}}{1 + \frac{|L_i(1) - L_i(2)|}{\text{Max}(L_i(1)) - \text{Min}(L_i(1))} + \frac{|L_i(1) - L_i(3)|}{\text{Max}(L_i(1)) - \text{Min}(L_i(1))}} \quad (11)$$

如果 $n=3$ 表示跨两层, 即第 1 层至第 4 层节点排序关系:

$$o_i^{(1,4)} = \frac{\sqrt[4]{\left(\left(\frac{1}{L_i(1)} \times \frac{1}{1} + \frac{1}{L_i(2)}\right) \times \frac{1}{2} + \frac{1}{L_i(4)}\right) \times \frac{1}{4}}}{1 + \frac{|L_i(1) - L_i(2)|}{\text{Max}(L_i(1)) - \text{Min}(L_i(1))} + \frac{|L_i(1) - L_i(3)|}{\text{Max}(L_i(1)) - \text{Min}(L_i(1))} + \frac{|L_i(1) - L_i(4)|}{\text{Max}(L_i(1)) - \text{Min}(L_i(1))}} \quad (12)$$

如果 $n=n$ 表示跨 $n-1$ 层, 即第 1 层至第 n 层节点排序关系:

$$o_i^{(t,t+n)} = \frac{\sqrt[n+1]{\left(\left(\frac{1}{L_i(t)} \times \frac{1}{1} + \frac{1}{L_i(t+1)}\right) \times \frac{1}{2} + \dots + \frac{1}{L_i(t+n)}\right) \times \frac{1}{n+1}}}{1 + \frac{|L_i(t) - L_i(t+1)|}{\text{Max}(L_i(t)) - \text{Min}(L_i(t))} + \frac{|L_i(t) - L_i(t+2)|}{\text{Max}(L_i(t)) - \text{Min}(L_i(t))} + \dots + \frac{|L_i(t) - L_i(t+n)|}{\text{Max}(L_i(t)) - \text{Min}(L_i(t))}} \quad (13)$$

网络节点超链接信息熵排序总体反映网络节点在一个时间层内或多个相邻时间层内网络节点重要性排序演化规律. 进而避免单一相邻层间节点排序波动变化较大时对整体排序产生影响, 降低时序网络节点重要性在全周期排序演化产生逆序的概率.

3.3 时序网络超链接信息熵矩阵模型构建

基于节点在相邻层间排序系数, 经典 SAM 模型无法描述节点在跨层间的相似性系数, 本文改进后的层间相似 OSAM 时序网络模型将经典的 SAM 模型中相邻层的单位矩阵 I 和可调参数 ω , 调整为相似性系数 O , 可更加准确描述节点相邻层和跨层在超邻接矩阵内的耦合关系, 降低可调参数对结果的主观影响. 在此基础上加入超链接节点跨层相似性系数矩阵, 保证时序网络节点跨层排序结果准确性. 具体形式如下:

$$A = \begin{bmatrix} A^{(1)} & O^{(1,2)} & O^{(1,3)} & \dots \\ O^{(1,2)} & A^{(2)} & O^{(2,3)} & \ddots \\ O^{(1,3)} & O^{(2,3)} & A^{(3)} & \ddots \\ \vdots & \ddots & \ddots & \ddots \end{bmatrix} \quad (14)$$

其中, $O^{(1,2)}, O^{(2,3)}, \dots$, 表示相邻层之间的相似性系数, $O^{(1,3)}, \dots$, 表示非相邻层之间的相似性系数; $O^{(1,2)}$ 为 $N \times N$ 的对角矩阵, 即 $O^{(1,2)} = \text{diag}(o_1^{(1,2)}, o_2^{(1,2)}, \dots, o_N^{(1,2)})$, $o_i^{(1,2)}$ 为时序网络排序层间相似性系数, 表示节点 i 在第一层和第二层的排序相似性. OSAM 时序网络超邻接矩阵将网络中每个节点在不同层内偏好排序动态集结, 用时序网络节点超链接信息熵解决了时序网络层内节点重

要性序结构递阶演化问题. 该模型的算法复杂度 $T_{OSAM}(n) = o(kn^2)$, 其中 k 是关于时间层数 T 的函数, 当 $T \ll n$ 时, $T_{OSAM}(n) = o(n^2)$. OSAM 算法流程图如下所示:

算法 1 OSAM 算法

输入: 复杂网络邻接矩阵 A

输出: 超邻接矩阵 A'

确定节点个数 N , 时间层数 T , 时序排序系数 o

建立空矩阵 $A, A^{(1)}, A^{(2)}, \dots, A^{(T)}$ 插入 A 的主对角线上

时间层 $i=1: T$, 时间层 $j=1: T, i \neq j$

IF $j=i+1$

 根据式(4)计算节点层内熵值, 得到排序结果 L

 根据式(9)计算时序网络相邻层间节点排序系数 $o_i^{(t,t+1)}, i=1, 2, \dots, N$

ELSE

 根据式(4)计算节点层内熵值, 得到排序结果 L

 根据式(10)计算时序网络相邻层间节点排序系数 $o_i^{(t,t+1)}, i=1, 2, \dots, N$

END IF

 构建节点跨层相似性系数 $O^{(t,t+n)}$, 并将 $O^{(t,t+n)}$ 作为对角线元素的对角矩阵插入 A_j 的相应位置 $A' = A$, 得到超邻接矩阵 A'

3.4 基于特征向量中心性节点重要性排序

特征向量中心性是评价复杂网络中节点重要性的指标, 其不仅考虑节点自身在网络中重要性的影响, 同时考虑邻居节点重要性对自身节点的影响. 本文采用

特征向量中心性对时序网络节点重要性进行度量,对上文构建的超邻接矩阵 A' 求主特征向量(最大特征值对应的特征向量) $\mathbf{v} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{NT}]^T$. 因此, 向量 \mathbf{v} 的第 $N(t-1) + i$ 项即表示第 t 个时间层上的节点 i 的特征向量中心性, 记为 $N \times T$ 矩阵:

$$\mathbf{W} = [w_{it}]_{N \times T} \quad (15)$$

则 $w_{it} = v_{N(t-1)+i}$, 其中, w_{it} 为矩阵 \mathbf{W} 的第 i 行第 t 列元素, 即为第 t 个时间层上节点 i 的特征向量中心性. 该指标不仅可以获得各时间层网络节点超链接重要性排序, 同时能够反映节点在每个时间层网络的重要性随时间变化的轨迹.

4 时序网络节点序结构实证数据分析

4.1 实证数据描述

本实验采用 Workspace 数据和 Email-Eu-core 数据检验 OSAM 方法对动态网络节点超链接特征演化序结构仿真分析. Workspace 数据集^[20]是 2013 年 6 月 24 日至 7 月 3 日在法国一栋办公楼测量到的个人之间面对面交互联系数据集. 本文将 Workspace 数据以 7 天为一个时间尺度进行划分. Email-Eu-core 数据集^[21]利用欧洲一家大型研究机构的电子邮件数据生成的网络数据集. 电子邮件时序网络数据包含 986 个匿名 ID 在历时 803 天中产生的交互信息, 本文以 30 天为一个时间尺度进行划分, 截取其中 360 天的交互信息进行仿真实验. 网络基本特征如下表 1 所示, 其中, N 代表网络节点数目, C 表示节点之间的交互次数, E 表示整个网络的连边数, During 是该数据的记录时间段, T 代表切分的网络层数.

表 1 实证网络基本统计特征

网络	N	C	E	During	T
Workspace	92	9827	755	2013. 6. 24-2013. 7. 3	10
Email-Eu-core	986	332334	24929	360d	12

4.2 基于传播模型的网络节点重要性评价

本文采用 SIR 传播模型度量节点重要性排序, 统计网络层中稳定状态下免疫节点的数量来计算此节点在网络中的影响力. SIR 模型将总人口分为以下三类: 易感者, 其数量记为 $S(t)$, 表示 t 时刻未染病但有可能被该类疾病传染的人数; 染病者, 其数量记为 $I(t)$, 表示 t 时刻已被感染成为病人而且具有传染力的人数; 免疫者, 其数量记为 $R(t)$, 表示 t 时刻已从染病者中移出且不会再被感染此疾病的人数. 设总人口为 $N(t)$, 则有 $N(t) = S(t) + I(t) + R(t)$. 传染率 β : 两人接触被传染的概率, 无论他是否为易感者. 人数 n : 针对于病人而言, 表示一个病人接触多少个人, 可接触的人包括除自己

以外种群中的所有. 治愈率 γ : 一位患者被治愈成为免疫者的概率. SIR 模型微分方程如下:

$$\begin{cases} \frac{ds(t)}{dt} = -\lambda I(t)S(t) \\ \frac{dI(t)}{dt} = \lambda I(t)S(t) - \eta I(t) \\ \frac{dR(t)}{dt} = \eta I(t) \end{cases} \quad (16)$$

本文进行仿真实验时, 假设传染者为节点 i , 开始时刻只有一个传染者, 剩余节点均为易感者. 在设定传播率 β 时过大或过小都无法衡量每个节点的传播能力. 因此, 本文设定的传播阈值^[22]为 $\beta_{th} \approx \langle k \rangle / \langle k^2 \rangle$, 其中 $\langle k \rangle$ 为网络平均度, $\langle k^2 \rangle$ 为网络二阶邻居平均度. 为保证传播过程正常进行, 本文传播率 β 大于传播阈值 β_{th} . 对每个节点进行 1000 次模拟仿真实验, 传染者节点 i 在当前层内传染后最终的免疫者数量取 1000 次实验结果的期望值. 不同网络时间层内节点 i 的邻居节点数和邻居节点信息熵可能不同, 故节点 i 在每一层实验仿真的排序结果都会有所差异.

4.3 时序网络节点重要性序结构肯德尔系数分析

为检验本文方法的有效性, 采用肯德尔相关系数 (Kendall' s τ)^[23]对特征向量矩阵 \mathbf{W} 和单位时间层内 SIR 模型仿真结果进行相关性分析. Kendall' s τ 用来测量两变量序列之间排序的相关性程度, 其取值范围为 $[-1, 1]$, 该值越大, 两序列相关性越强; 反之, 则两序列相关性越弱. Kendall' s τ -b 具体定义如下:

$$\text{Kendall' s } \tau = \frac{\sum_{i < j} \text{sgn}[(x_i - x_j)(y_i - y_j)]}{\sqrt{\left[\frac{n(n-1)}{2} - n_1\right] \left[\frac{n(n-1)}{2} - n_2\right]}} \quad (17)$$

其中, $\mathbf{X} = (x_1, x_2, \dots, x_n)^T$, $\mathbf{Y} = (y_1, y_2, \dots, y_n)^T$, \mathbf{X} 表示特征向量中心性矩阵 \mathbf{W} 中第 t 列向量, \mathbf{Y} 表示单位时间内 SIR 模型仿真结果排序值 ($t = 1, 2, \dots, T$); $\text{sgn}(z)$ 为一个分段函数, 当 $z > 0$ 时, $\text{sgn}(z) = +1$, 当 $z < 0$ 时, $\text{sgn}(z) = -1$, 当 $z = 0$ 时, $\text{sgn}(z) = 0$, 这里 n 为序列长度, 即节点总数; $n_1 = \sum_i t_i(t_i - 1)/2$, $n_2 = \sum_j u_j(u_j - 1)/2$, 其中, t_i 为 X 序列中第 i 个使得 $\text{sgn}(z) = 0$ 的 x_i 值的个数, u_j 为 Y 序列中第 j 个使得 $\text{sgn}(z) = 0$ 的 y_j 值的个数.

4.4 时序网络节点序结构演化分析

本文基于上述实证网络数据, 依据 OSAM 方法和 SAM 方法取值不同参数计算时序网络中各时间层网络节点的特征向量中心性, 得到 OSAM 和 SAM 方法对实证网络数据的节点重要性排序结果, 其中 SAM 方法参数取 $\omega \in \{0.1, 0.2, \dots, 1.0\}$. 为了直接检验 OSAM 方法

的效果,使用 SIR 模型得到节点免疫数量排序结果后,利用 Kendall's τ (肯德尔系数)评估排序相关性. 在图 2~4 中横坐标表示网络切分的各个时间层,纵坐标表示相应时间层对应的 Kendall's τ 值,图 2、图 4 中黑色为 OSAM 方法,其它颜色为 SAM 方法取不同参数仿真结果汇总图.

根据图 2、图 4 分析:(1)OSAM 方法得到 Kendall's τ 值总体高于 SAM 方法,说明基于时序网络节点层间序结构递阶的 OSAM 方法分析节点在不同层间的差异性可更准确地描述时序网络动态演化过程;(2)从实证网络 Workspace 数据结果来看,OSAM 方法比 SAM 方法的 Kendall's τ 值在部分时间层均有提高,如图 2 所示,最高为 8.722%;存在个别层,如 Workspace 数据的 $t=6$ 和 $t=8$ 上,OSAM 方法的计算结果劣于 SAM 方法,此结果是由于实际数据本身冗余度影响造成;(3)OSAM 方法在实证网络 Email-Eu-core 数据集的实验结果与 SAM 方法的 Kendall's τ 值分别如图 3 所示. 图 3 是 $\omega \in \{0.1, 0.2, \dots, 1.0\}$ 的 Kendall's τ 值汇总图. 因为实证网络中全时间段内孤立节点的影响 OSAM 方法效果不明显;实验中在不改变网络连边性质,舍去部分全时间段内孤立节点后,此时 SAM 方法的 Kendall's τ 值和全时间段内全部节点的 OSAM 的 Kendall's τ 值对比如图 4 所示. OSAM 方法比 SAM 方法的 Kendall's τ 值在各个时间层全面提高,最高为 11.310%. 实证分析证明动态网络全时间段内的孤立节点对时序网络超链接信息熵序结构演化有显著影响.

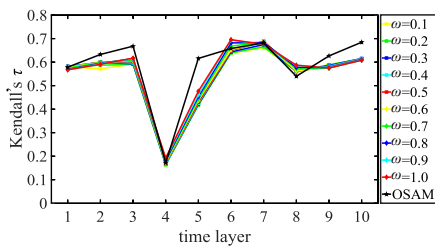


图 2 Workspace 数据特征向量中心性与 SIR 排序结果的 Kendall's τ 值

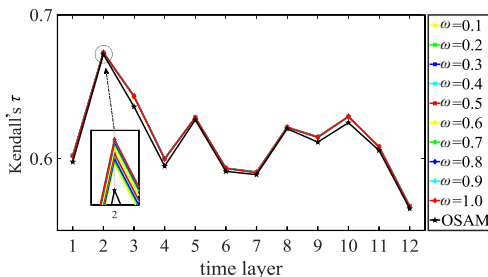


图 3 Email-Eu-core 数据特征向量中心性与 SIR 排序结果的 Kendall's τ 值

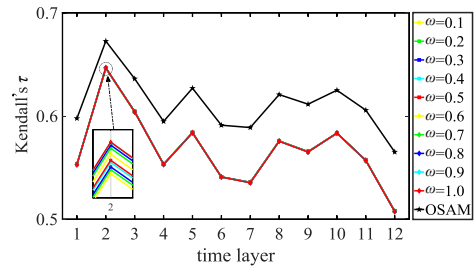


图 4 Email-Eu-core 数据删除部分孤立节点后特征向量中心性与 SIR 排序结果的 Kendall's τ 值

5 结论

本文研究动态时序网络超链接信息熵的节点重要性序结构演化模型,通过时序网络节点超链接信息熵构建时序网络相邻时间层与跨层节点重要性排序模型,表征时序网络节点层内交互关系和层间耦合关系序结构演化规律;并且进行模拟仿真与实践应用,系统建模分析结论科学有效. 同时,在现实网络应用过程中,发现部分动态网络层间悬挂孤立节点统计结果对网络整体序结构峰值影响偏大,导致网络节点超链接信息熵值在全周期内波动异常. 如何有效关联分析时序网络孤立节点超链接对层内层间的系统演化效应,对时序网络全周期序结构依托网络动力学进行更加精确的预测决策,需进一步研究.

参考文献

- [1] 纪秋磊, 梁伟, 傅伯杰, 等. 基于 Google Earth Engine 与复杂网络的黄河流域土地利用/覆被变化分析[J]. 生态学报, 2021, 42(06): 1-14.
JI Qiu-lei, LIANG Wei, FU Bo-jie, et al. Land use/cover change in the Yellow River Basin based on Google Earth Engine and complex network[J]. Acta Ecologica Sinica, 2021, 42(06): 1-14. (in Chinese)
- [2] 刘慧, 王炳琚, 陆君安, 等. 复杂网络牵制控制优化选点算法及节点组重要性排序[J]. 物理学报, 2021, 70(05): 284-294.
LIU Hui, WANG Bing-jun, LU Jun-an, et al. Node-set importance and optimization algorithm of nodes selection in complex networks based on pinning control[J]. Chinese Journal of Physics, 2021, 70(05): 284-294. (in Chinese)
- [3] CHENFENG X, SONGHUA H, MOFENG Y, et al. Mobile device data reveal the dynamics in a positive relationship between human mobility and COVID-19 infections[J]. Proceedings of the National Academy of Sciences, 2020, 117(44): 27087-27089.
- [4] YABE T, TSUBOUCHI K, FUJIWARA N, et al. Non-compulsory measures sufficiently reduced human mo-

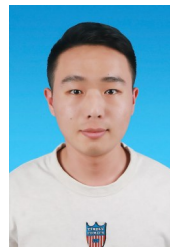
- bility in Tokyo during the COVID-19 epidemic[J]. Scientific Reports, 2020, 10(1): 1-9.
- [5] 刘金平, 何捷舟, 马天雨, 等. 基于 KELM 选择性集成的复杂网络环境入侵检测[J]. 电子学报, 2019, 47(5): 1070-1078.
LIU Jin-ping, HE Jie-zhou, MA Tian-yu, et al. Selective ensemble of KELM-Based complex network intrusion detection[J]. Acta Electronica Sinica, 2019, 47(5): 1070-1078. (in Chinese)
- [6] BONACICH P. Factoring and weighting approaches to status scores and clique identification[J]. Journal of Mathematical Sociology, 2010, 2(1): 113.
- [7] Freeman L C. A set of measures of centrality based on betweenness[J]. Sociometry, 1977, 40(1): 35-41.
- [8] SABIDUSSI G. The centrality index of a graph[J]. Psychometrika, 1966, 31(4): 581-603.
- [9] BORGATTI S P. Centrality and network flow[J]. Social Networks, 2005, 27(1): 55-71.
- [10] ZELLEN S M. Rethinking centrality: methods and examples[J]. Social Networks, 1989, 11(1): 1-37.
- [11] 曹玖新, 董丹, 徐顺, 等. 一种基于 k -核的社会网络影响最大化算法[J]. 计算机学报, 2015, 38(02): 238-248.
CAO Jiu-xin, DONG Dan, XU Shun, et al. A k -core based algorithm for influence maximization in social networks[J]. Chinese Journal of Computers, 2015, 38(02): 238-248. (in Chinese)
- [12] LIU Y, TANG M, ZHOU T, et al. Identify influential spreaders in complex networks, the role of neighborhood [J]. Physica A Statistical Mechanics & Its Applications, 2016, 452(12): 289-298.
- [13] LORRAIN F, WHITE H C. Structural equivalence of individuals in social networks[J]. Social Networks, 1977, 1(1): 67-98.
- [14] ADAMIC L A, ADAR E. Friends and neighbors on the Web[J]. Social Networks, 2003, 25(3): 211-230.
- [15] ZHOU T, LÜ L, ZHANG Y C. Predicting missing links via local information[J]. European Physical Journal B, 2009, 71(4): 623-630.
- [16] TAYLOR D, MYERS S A, CLAUSET A, et al. Eigenvector-based centrality measures for temporal networks[J]. Multiscale Modeling & Simulation, 2017, 15(1): 537-574.
- [17] 杨剑楠, 刘建国, 郭强. 基于层间相似性的时序网络节点重要性研究[J]. 物理学报, 2018, 67(04): 279-286.
YANG Jian-nan, LIU Jian-guo, GUO Qiang. Node importance identification for temporal network based on inter-layer similarity[J]. Chinese Journal of Physics, 2018, 67(04): 279-286. (in Chinese)
- [18] 胡钢, 许丽鹏, 徐翔. 基于时序网络层间同构率动态演化的重要节点辨识[J]. 物理学报, 2021, 70(10): 355-366.
HU Gang, XU Li-peng, XU Xiang. Identification of important nodes based on dynamic evolution of inter-layer isomorphism rate in temporal networks[J]. Chinese Journal of Physics, 2021, 70(10): 355-366. (in Chinese)
- [19] 胡钢, 高浩, 徐翔. 基于多属性偏好信息集结的复杂网络重要节点辨识[J]. 浙江理工大学学报, 2019, 041(004): 482-488.
HU Gang, GAO Hao, XU Xiang. Identify important nodes in complex network based on aggregation of multi-attribute preference information[J]. Journal of Zhejiang Sci-Tech University, 2019, 041(004): 482-488. (in Chinese)
- [20] GÉNOIS M, BARRAT A. Can co-location be used as a proxy for face-to-face contacts?[J]. EPJ Data Science, 2018, 7(1): 11-29.
- [21] PARANJAPE A, BENSON A R, LESKOVEC J. Motifs in Temporal Networks[M]. New York: Association Computing Machinery, 2017: 601-610.
- [22] CASTELLANO C, PASTOR-SATORRAS R. Thresholds for epidemic spreading in networks[J]. Physical Review Letters, 2010, 105(21): 218701-218705.
- [23] BORRONI C G. A new rank correlation measure[J]. Statistical Papers, 2013, 54(2): 255-270.

作者简介



胡 钢 男, 1970 年出生于甘肃省天水市; 安徽工业大学管理科学与工程学院副教授, 硕士生导师; 研究方向: 多属性决策、复杂网络系统建模仿真与均衡分析.

E-mail: hug_2004@126.com



牛 琼 男, 1997 年出生于安徽省亳州市; 安徽工业大学在读硕士研究生, 主要研究方向为复杂网络系统建模仿真与均衡分析、多属性决策.

E-mail: NiuQiongNQ@163.com