

基于层间相似性的时序网络节点重要性研究

杨剑楠 刘建国 郭强

Node importance identification for temporal network based on inter-layer similarity

Yang Jian-Nan Liu Jian-Guo Guo Qiang

引用信息 Citation: *Acta Physica Sinica*, 67, 048901 (2018) DOI: 10.7498/aps.20172255

在线阅读 View online: <http://dx.doi.org/10.7498/aps.20172255>

当期内容 View table of contents: <http://wulixb.iphy.ac.cn/CN/Y2018/V67/I4>

您可能感兴趣的其他文章

Articles you may be interested in

一种改进的基于信息传播率的复杂网络影响力评估算法

An improved evaluating method of node spreading influence in complex network based on information spreading probability

物理学报.2017, 66(20): 208901 <http://dx.doi.org/10.7498/aps.66.208901>

多层单向耦合星形网络的特征值谱及同步能力分析

Synchronizability and eigenvalues of multilayer star networks through unidirectionally coupling

物理学报.2017, 66(18): 188901 <http://dx.doi.org/10.7498/aps.66.188901>

基于领域相似度的复杂网络节点重要度评估算法

Node importance measurement based on neighborhood similarity in complex network

物理学报.2017, 66(3): 038902 <http://dx.doi.org/10.7498/aps.66.038902>

基于超图的超网络相继故障分析

Cascading failure analysis in hyper-network based on the hypergraph

物理学报.2016, 65(8): 088901 <http://dx.doi.org/10.7498/aps.65.088901>

负荷作用下相依网络中的级联故障

Load-induced cascading failure in interdependent network

物理学报.2015, 64(4): 048901 <http://dx.doi.org/10.7498/aps.64.048901>

基于层间相似性的时序网络节点重要性研究*

杨剑楠¹⁾ 刘建国^{2)†} 郭强¹⁾

1) (上海理工大学复杂系统科学研究中心, 上海 200093)

2) (上海财经大学金融科技研究院, 上海 200433)

(2017年10月18日收到; 2017年11月29日收到修改稿)

时序网络可以更加准确地描述节点之间的交互顺序和交互关系. 结合多层耦合网络分析法, 本文提出了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, 与经典的认为所有层间关系为常数不同, 层间关系用节点的邻居拓扑重叠系数进行度量. Workspace 和 Enrons 数据集上的结果显示: 相比经典的方法, 使用该方法得到的 Kendall's τ 值在各时间层上的平均提高, 最高为 17.72% 和 12.44%, 结果表明层间相似性的度量对于时序网络的节点重要性度量具有十分重要的意义.

关键词: 时序网络, 层间相似性, 特征向量中心性, 时序全局效率

PACS: 89.75.Da, 05.10.-a

DOI: 10.7498/aps.67.20172255

1 引言

时序网络考虑节点之间的交互关系和次序, 可以更加准确地刻画手机通讯、社交等复杂系统的交互关系^[1-5]. 节点重要性的评价方法有很多种, 如度中心性、介数中心性、紧密度中心性、特征向量中心性、K-核中心等, 不同的评价方法其侧重点也各有不同^[6-8]. 刘建国等^[6]分别从网络结构和传播动力学的角度, 对现有的复杂网络中节点重要性排序方法进行了系统的比较、归纳与总结. Liu 等^[9]综合网络结构和传播动力学特征进行耦合分析, 利用差分方程考察传播过程, 对节点重要性进行排序. 尽管对于节点重要性排序的研究在静态网络上已经取得了一定进展, 但时序网络不同于静态拓扑结构的网络, 由于时间维度的引入, 网络的连边随时间会间断性地出现和消失^[5,10]. 所以, 近些年人们开始对时序网络节点重要性的研究方法进行探究. Tang 等^[11,12]基于时间切片的网络, 通过时序最短路径定义了时序介数中心性和时序紧密度中心性等统计特性, 并提出了节点重要性预测及网络切片方法等. 邓冬梅等^[13,14]基于节点边贡献

值来评价节点重要性, 考虑网络时间属性并定义事件相关节点感染方式, 提出了时序社交网络中节点重要性评价指标与排序算法. Kim 和 Anderson^[15]在构建时序网络时规定每个时间切片上的事件仅有一次, 并将切片间用有向边连接, 转化为沿时间单向的有向静态图, 从而定义了有向时序图的度中心性、介数中心性和紧密度中心性. Huang 和 Yu^[16]将文献^[9]中考虑动力学指标的重要节点识别方法引入时序网络中, 对节点重要性进行度量.

但上述方法都只考虑时间切片上的连接关系, 为了完整地表示时序网络的结构演变及其动力学过程, 还需要考虑不同时间切片之间的连接关系. Taylor 等^[17]考虑用多层耦合网络分析的方法, 将时序网络按层间关系和层内关系建立超邻接矩阵 (supra-adjacency matrix, SAM), 并定义了基于特征向量的中心性指标和节点重要性随时间波动的评判指标. 然而, 经典的 SAM 方法中时序网络的构建, 其层间关系为参数表示, 忽略了不同节点层间连接关系的差异性. 基于此, 考虑到不同节点的层间连接关系应不同, 本文将节点的层间连接关系用邻居拓扑重叠系数表示, 提出了基于节点层间相似

* 国家自然科学基金 (批准号: 61773248, 71771152) 资助的课题.

† 通信作者. E-mail: liujg004@ustc.edu.cn

性的超邻接矩阵 (similarity-based supra-adjacency matrix, SSAM) 时序网络构建方法, 利用特征向量中心性指标, 获取节点在各个时间层上的重要性排序, 并得到节点重要性随时间变化的轨迹. 通过节点删除法判断节点在网络中的影响力, 利用删除节点后网络效率的变化情况来评价节点的重要性排序, 并与 SAM 方法的不同参数结果做比较. Workspace 和 Enrons 数据集上的实验结果显示, 相比于 SAM 方法, SSAM 方法最终的 Kendall's τ 值在各时间层上的平均提高, 最高为 17.72% 和 12.44%, 说明 SSAM 方法由于避免了节点共用参数可能造成的高估或者低估节点的层间连接关系, 所以可以更为恰当地描述时序网络, 从而更准确地识别时序网络中的重要节点.

2 理论基础

2.1 时序网络描述

时序网络是一个包含了个体及个体间按时间顺序相互作用的系统, 将个体视为节点, 个体间的相互作用视为节点间的连边, 节点间连边随时间变化且具有时间先后顺序. 通常定义一个网络 $G = (V, E)$, 其中所有节点构成节点集为 $V = \{v_1, v_2, \dots, v_N\}$, 如果不考虑时序网络中交互事件发生的时长, 则时序网络整个观察期 $[0, m]$ 内的边可以用三元组 $e_t = (i, j, t)$ 来描述, 表示节点 i 与节点 j 在 t 时刻发生了一次交互, 故所有这样的边构成边集 $E = \{e_1, e_2, \dots, e_t\}$. 可以将这样的时序网络整个观察期 $[0, m]$ 按一定时间间隔切分为 T 个时间窗口 (其中时间间隔为 m/T), 则网络被分为 T 个离散有序的时间层网络 $G_1, G_2, \dots, G_T$.

2.2 邻居拓扑重叠系数

邻居拓扑重叠系数源自文献 [12] 提出的时序相关系数, 时序相关系数具体表示为

$$c_i = \frac{1}{T-1} \sum_{t=1}^{T-1} c_i^{(t,t+1)}, \quad c = \frac{1}{N} \sum_{i=1}^N c_i, \quad (1)$$

其中 $c_i^{(t,t+1)}$ 为节点 i 在两个相邻时间层网络 G_t, G_{t+1} 中的共同邻居数量所占的比例, 称为节点 i 在时间层网络 G_t, G_{t+1} 间的邻居拓扑重叠系数 [18]. 其具体形式如下:

$$c_i^{(t,t+1)} = \frac{\sum_j a_{ij}(t) a_{ij}(t+1)}{\sqrt{\left[\sum_j a_{ij}(t) \right] \left[\sum_j a_{ij}(t+1) \right]}}, \quad (2)$$

其中 $a_{ij}(t), a_{ij}(t+1)$ 分别为相邻时间层网络 G_t, G_{t+1} 对应的邻接矩阵中的元素, 如果在网络 G_t 中节点 i 与节点 j 之间有连边, 那么 $a_{ij}(t) = 1$; 否则 $a_{ij}(t) = 0$. 且只有节点 j 既是节点 i 在时间层 G_t 上的邻居节点同时又是在时间层 G_{t+1} 上的邻居节点时, 才有 $a_{ij}(t) a_{ij}(t+1) = 1$; 其他情况时 $a_{ij}(t) a_{ij}(t+1) = 0$.

邻居拓扑重叠系数描述了节点持续出现度及节点邻居关系的层间相似性. 该系数越大表示节点持续出现在两个相继时间层, 且邻接关系保持稳定; 反之, 系数越小表示节点未持续出现在两个相继时间层或相邻时间层上邻接关系较不稳定.

2.3 经典的 SAM 时序网络模型

为了更完整地构建时序网络, 文献 [17] 将时序网络通过层内连接关系和层间连接关系来表示, 提出了经典的超邻接矩阵 SAM 模型, SAM 为 $NT \times NT$ 的分块矩阵. 其中层内连接关系用相应时间层网络的邻接关系 $A^{(1)}, A^{(2)}, \dots, A^{(T)}$ 表示, 均为 $N \times N$ 的矩阵; 层间连接关系用一个参数 ω 来表示, 参数大小调节网络层间连接关系的紧密程度. 对于一个给定节点数为 N 的时序网络 Γ , 其切分的有序时间层网络集合为 $\Gamma = \{G_t\}_{t=1,2,\dots,T}$, T 为切分的时间层总数, 则其 SAM 具体表示如下:

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

其中, $\mathbf{A}$ 表示经典的时序网络模型; $A^{(1)}, A^{(2)}, \dots, A^{(T)}$ 表示层内连接关系, 这里用切分的 T 个时间层网络对应的邻接矩阵表示, 依次位于 SAM 的对角线上, 表示有序的时间层网络: 定义 $a_{ij}(t)$ 为邻接矩阵 $A^{(t)}$ 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 间无连边, $a_{ij}(t) = 1$ 表示有连边关系; $\omega \mathbf{I}$ 表示层间连接关系, 其中 ω 为可调参数, $\mathbf{I}$ 为 $N \times N$ 单位矩阵. 由于这里仅考虑同一节点在相邻时间层间的连接关系, 所以矩阵其他部分均为 0.

3 改进方法

3.1 基于层间相似性SSAM时序网络模型

由于上述的超邻接矩阵SAM中, 层间关系的参数讨论使时序网络建模复杂化, 且节点在相邻时间层的连接关系用同一参数来表示, 忽略了不同节点间的差异性. 实际上, 不同节点对应的层间连接关系应做不同考虑, 才能更真实地反映网络连接的实际情况. 基于以上考虑, 本文对SAM的层间连接关系做了改进.

考虑到同一节点在相邻时间层之间的连接关系与其在两个时间层上的持续出现度及节点的邻居关系层间相似程度有关, 所以我们用邻居拓扑重叠系数((2)式)来确定层间连接关系, 则改进后基于层间相似性SSAM时序网络模型具体表示形式如下:

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

其中, $\mathbf{A}'$ 表示基于节点层间相似性的超邻接矩阵SSAM; $A^{(1)}, A^{(2)}, \dots, A^{(T)}$ 同样表示层内连接关系, 这里用切分的 T 个切片网络对应的邻接矩阵表示; $C^{(1,2)}, C^{(2,3)}, \dots, C^{(t-1,t)}$ 分别表示相邻时间层之间的连接关系, 如 $C^{(1,2)}$ 表示时间层网络 G_1 与时间层网络 G_2 之间的连接关系. $C^{(t-1,t)}$ 为 $N \times N$ 的对角矩阵, 即 $C^{(t-1,t)} =$

$\text{diag}(c_1^{(t-1,t)}, c_2^{(t-1,t)}, \dots, c_N^{(t-1,t)})$, 而 $c_i^{(t-1,t)}$ 即为(2)式节点的邻居拓扑重叠系数, 用来描述节点 i 的层间相似性, 即节点 i 在时间层网络 G_{t-1} 上与时间层网络 G_t 上的共同邻居数所占的比例. 本文也只考虑同一节点在相邻时间层之间的连接关系, 所以SSAM矩阵的其他部分均为0.

图1给出了一个包含4个节点和3个时间层的时序网络及SSAM方法的具体表示, 其中黑色实线表示层内连接关系, 黑色虚线表示层间连接关系.

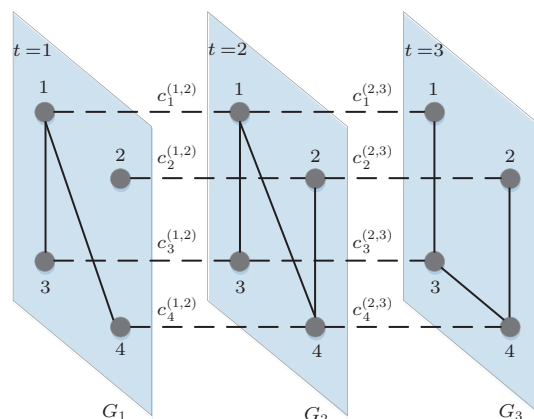


图1 基于层间相似性方法的时序网络建模实例

Fig. 1. An example of SSAM model for temporal network.

则图1对应的基于层间相似性的矩阵表示如(5)式. 具体地, 层内连接关系由各个时间层网络的邻接矩阵确定, 即(5)式的对角线矩阵块部分; 相邻时间层的层间连接关系则由各个节点的层间相似性, 即(2)式邻居拓扑重叠系数计算得到; 最后, 非相继时间层的矩阵块部分不做考虑, 其元素均为0.

$$\mathbf{A}' = \begin{bmatrix} 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & & & & & \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & & & & & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & & & & & \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/\sqrt{2} & & & & \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1/\sqrt{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1/\sqrt{2} & 0 & 0 \\ 0 & 0 & 0 & 1/\sqrt{2} & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1/\sqrt{4} \\ & & & 1/\sqrt{2} & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & & & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ & & & 0 & 0 & 1/\sqrt{2} & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ & & & 0 & 0 & 0 & 1/\sqrt{4} & 0 & 1 & 1 & 0 & 0 & 0 \end{bmatrix}, \quad (5)$$

3.2 特征向量中心性

特征向量中心性是评估网络节点重要性的一个重要指标^[6], 其不仅考虑节点的重要性程度, 同时综合考虑邻居节点的重要性程度. 所以本文利用特征向量中心性, 对时序网络的节点重要性进行度量, 即对上文构建的超邻接矩阵 $\mathbf{A}'$ 求主特征向量(最大特征值对应的特征向量) $\boldsymbol{\nu} = \{v_1, v_2, \dots, v_{NT}\}^T$. 这样, 向量 $\boldsymbol{\nu}$ 的第 $N(t-1)+i$ 个项即表示第 t 个时间层上的节点 i 的特征向量中心性, 记为 $N \times T$ 的矩阵 $\mathbf{W} = \{w_{it}\}_{N \times T}$, 则

$$w_{it} = v_{N(t-1)+i}, \quad (6)$$

表1 实例网络中节点的特征向量中心性

Table 1. Eigenvector centrality of nodes in temporal network of Fig. 1.

$W_{\text{SAM}}(\omega = 0.5)$				W_{SSAM}			
节点	G_1	G_2	G_3	节点	G_1	G_2	G_3
1	0.2809	0.4413	0.2392	1	0.3739	0.4742	0.2287
2	0.0542	0.2444	0.1978	2	0.0	0.1986	0.1629
3	0.1934	0.3094	0.3184	3	0.2760	0.3558	0.2695
4	0.2189	0.4247	0.3233	4	0.2383	0.3621	0.2320

结合图1的时序网络可以看到, 用共同的参数0.5来表示不同节点的层间连接关系, 弱化了持续出现且节点层间邻居相似度高的节点的重要程度, 而强化了孤立节点的重要程度. 例如第一时间层网络 G_1 中的1号节点, 持续出现在相继时间层且层内的邻接关系稳定, 虽然其在两方法中均为该时间层网络中最重要节点, 但是在SAM方法里1号节点的重要性值较小为0.2809, 而SSAM方法里1号节点的重要性值较大为0.3739, 说明SAM方法弱化了1号节点的重要性值; 对于2号节点, 其在 G_1 中为孤立节点, 其重要性值应接近于0, 而SAM方法高估了 G_1 中2号节点的重要性值. 所以不难推出, SAM方法中, 当参数较大时, 更加强了孤立节点和层间邻居相似度低的节点的重要性; 而参数较小时, 则更加弱化了持续出现且层间邻居相似度高的节点的重要性.

4 数据实验与结果分析

4.1 数据描述

为了检验SSAM方法对时序网络节点重要性排序的效果, 本文用了两个实证网络来进

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. 该指标可以获得各时间层上节点重要性的排序, 同时能够得到节点在各个时间层网络重要性随时间变化的轨迹.

表1为图1中实例网络的特征向量中心性指标的结果, 并给出了SAM方法参数 ω 取0.5的特征向量中心性结果做对比. 由表1结果可以得到各时间层节点的重要性排序及各个节点重要性随时间层的变化轨迹, 比如对于SSAM方法, 第一时间层的 G_1 中节点重要性排序为1-3-4-2, 且可以看出1号节点重要性排序随时间层变化轨迹为1-1-3.

行实验, 分别为Workspace数据和Enrons数据. Workspace^[19]为法国某公司通过移动射频设备获取的公司员工之间面对面交互产生的交互数据, 时间从2013年6月24日到2013年7月3日, 数据按天切分; Enrons^[20]为美国某公司员工的邮件往来数据, 数据从1999年到2002年, 为缩减数据, 取其中2001年的数据子集, 并按月切分. 网络基本统计特性具体描述如表2, 其中 N 表示节点总数, T 表示切分的时间层网络数, C 表示节点之间的总交互次数, E 表示整个聚合网络的连边数, During为该数据记录的时段.

表2 实证网络基本统计特性

Table 2. Basic statistical features of Workspace and Enrons networks.

网络	N	C	E	During	T
Workspace	92	9827	755	2013.6.24—2013.7.3	10
Enrons	151	33124	1270	2001	12

4.2 评价方法

通常认为, 节点重要性不仅体现在节点在网络中对信息的传播能力^[21–23], 也可体现在节点被

移除以后对网络连通的破坏性^[8]. 为进一步检验 SSAM 方法对节点重要性的排序效果, 并与 SAM 方法做对比, 文中通过删除节点后网络连通性的变化程度来评价节点重要性排序结果. 通常认为, 删除节点后, 网络连通性变化越大, 则被删除的节点在网络中越重要; 反之, 则节点重要性相对较低.

时序网络效率是评判时序网络连通性的一个重要方法, 其时序全局效率^[17]的具体形式如下:

$$e = \frac{1}{N(N-1)} \sum_{ij} \frac{1}{d_{ij}}, \quad (7)$$

其中, d_{ij} 为时序网络中各节点之间的时序距离^[12]. 时序距离通过时序最短路径计算, 时序最短路径不同于静态网络的最短路径, 时序路径的有效连接都是临时的, 会在某个特定的时间点建立或者断开, 所以时序网络中的时序最短路径需要遵从于不同连边的时间先后顺序^[4]. 信息从节点 i 经过节点 k 最终传到节点 j , 需要在时间维度上要求节点 i 到 k 之间的有效连接发生在节点 k 到 j 之前, 否则信息不能从节点 i 传到 j .

具体地, 图 1 所示的时序网络, 假设有信息从 $t = 1$ 时的 1 号节点开始传递, 并传递到最终 $t = 3$ 时刻结束, 且每个时间层上只传递一步, 则表 3 给出了最终时序最短路径的时序距离 d_{ij} 的结果.

表 3 图 1 时序网络中各节点之间的时序距离

Table 3. Temporal distance of nodes in temporal network of Fig. 1.

节点编号	1	2	3	4
1	0	2	2	1
2	∞	0	3	2
3	1	3	0	2
4	1	2	2	0

最后, 用删除节点前后时序全局效率的差值作为节点重要性的验证方法. 为了与特征向量中心性矩阵 $\mathbf{W}$ 对应, 依次删除各个时间层的节点后重新计算网络的时序全局效率, 得到一个 $N \times T$ 的矩阵 $\mathbf{E} = \{e_{it}\}_{N \times T}$, 再与原时序全局效率 e 做差值, 最终得到删除节点的时序全局效率差值矩阵 $\mathbf{E}'$, 其元素

$$E'_{i,t} = |e_{it} - e|, \quad (8)$$

其中, $E'_{i,t}$ 为删除第 t 个时间层网络上的第 i 个节点后网络时序全局效率差值, 对应的值越大, 说明该被删除节点越重要.

4.3 结果分析

本文基于上述两个实证网络数据, 依据 SSAM 方法和 SAM 方法的不同参数取值计算时序网络中的各时间层网络上节点的特征向量中心性, 分别得到了 SSAM 和 SAM 方法对实证网络数据的节点重要性排序 (其中 SAM 方法的参数 ω 取 $[0.1, 0.2, \dots, 1.0]$). 另外, 为了更直观地检验 SSAM 方法的效果, 用节点删除法得到节点时序全局效率差值排序后, 利用 Kendall's τ (肯德尔系数) 来评估排序相关性. 具体而言, 我们对特征向量中心性矩阵 $\mathbf{W}$ 和时序全局效率差值矩阵 $\mathbf{E}'$ 对应的列分别计算 Kendall's tau, 即得到相应时刻上节点重要性排序同该时刻删除节点后网络效率差值排序的一致性程度.

Kendall's τ ^[24] 被用来测量两序列之间排序的相关性程度, 其取值范围为 $[-1, 1]$, 该值越大, 则两序列相关性越强, 即说明节点重要性的排序方法更加精确. 其中对于两列数 $X = \{x_1, x_2, \dots, x_n\}$ 和 $Y = \{y_1, y_2, \dots, y_n\}$ 的 Kendall's τ -b^[25] 定义为

$$\tau = \frac{\sum_{i < j} \text{sgn}[(x_i - x_j)(y_i - y_j)]}{\sqrt{(n(n-1)/2 - n_1)(n(n-1)/2 - n_2)}}, \quad (9)$$

这里 X 表示特征向量中心性矩阵 $\mathbf{W}$ 中第 t 列的排序列表; Y 表示通过删除节点法得到的网络全局效率差值矩阵 $\mathbf{E}'$ 中对应第 t 列的排序列表 ($t = 1, 2, \dots, T$); $\text{sgn}(z)$ 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.

对 Workspace 数据和 Enrons 数据分别用 SSAM 方法和 SAM 方法的特征向量中心性矩阵与删除节点法的网络效率差值矩阵得到相应时间层的 Kendall's τ 值如图 2 所示: 图 2(a) 为 Workspace 数据的结果, 图 2(b) 为 Enrons 数据的结果; 图 2 中横坐标表示时序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 tau 值; 其中黑色小正方形为 SSAM 方法, 其他为 SAM 方法取不同参数的结果.

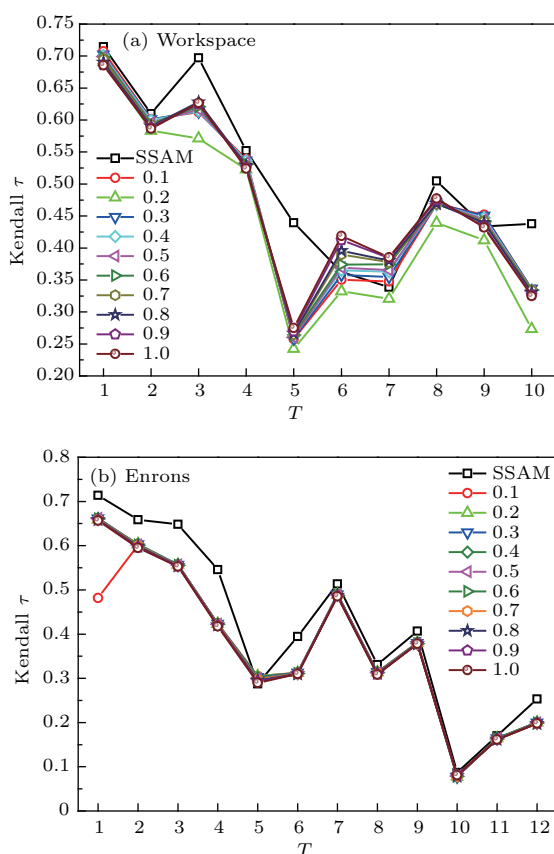


图2 特征向量中心性与时序网络效率差值的 Kendall's tau 结果 (a) Workspace 数据基于层间相似性的超邻接矩阵方法和经典超邻接矩阵方法不同参数的 Kendall's τ 结果; (b) Enrons 数据相应的结果

Fig. 2. Results of Kendall's τ for eigenvector centrality and difference of temporal global efficiency: (a) result for Workspace by SSAM method and SAM method; (b) result for Enrons by SSAM method and SAM method.

由图2中结果可以看到: 1) 对两组数据的时序网络构建中, SAM方法不同参数下得到的 Kendall's τ 结果大多相近, Enrons 数据集的结果在这点尤为显著, 也就是不同的参数选择对于节点的特征向量中心性在各个时间层的排序结果影响并不显著, 所以SAM方法中, 层间连接关系以参数来讨论使得时序网络构建复杂化, 可以考虑用确切的数值来描述层间连接关系; 2) SSAM方法得到的 Kendall's τ 系数结果大部分比SAM方法的高, 也就是基于层间相似性的SSAM方法考虑了不同节点的差异性, 能更准确地对时序网络进行描述, 得到的节点重要性排序也更为准确, 且对于Enrons数据的结果尤为显著, 其中两组实证数据SSAM方法的结果相比SAM方法和Kendall's τ 值在各时间层上的平均提高, 最高为17.72%和12.44%; 3) 对于Workspace数据的个别时间层, 如 $t = 6$, $t = 7$ 和

$t = 9$ 上, SSAM方法略低于SAM方法, 同时该时间层上SAM方法的不同参数结果差别也略大, 我们认为此结果是由于实际数据的影响所造成的。

5 结 论

识别网络中的重要节点一直是复杂网络的热点问题, 最近有不少工作开始考虑加入时间属性来研究节点重要性的度量方法, 而对于时序网络的建模也各有不同. 本文将时序网络用层内邻接关系和层间连接关系共同来描述, 并考虑不同节点的差异性及其相继时间层的时序相关性, 用节点的邻居拓扑重叠系数作为时序网络的层间连接关系, 提出了一种基于节点层间相似度的超邻接矩阵建模方法(SSAM方法), 同时利用特征向量中心性来度量节点的重要性. 另外, 本文利用节点删除法, 通过计算删除节点前后网络时序全局效率的变化情况, 来评估SSAM方法对节点重要性排序的效果.

两组实证数据的结果表明, SSAM方法得到的特征向量中心性排序结果大部分比SAM方法的好, 且SSAM方法两组实验的Kendall's τ 值较SAM方法在各时间层上的平均提高, 最高为17.72%和12.44%. SSAM方法一方面回避了参数讨论, 另一方面由于考虑了不同节点在相邻时间层间连接关系的差异性, 用基于节点的层间相似度构建时序网络, 避免了节点共用参数可能造成的高估或者低估节点的层间连接关系. 综上所述, 本文提出了一种基于层间相似性的时序网络构建方法, 该方法可以更为恰当地描述时序网络, 从而更准确地识别时序网络中的重要节点. 该工作对于时序网络建模和对时序网络节点重要性度量方法的研究具有重要意义.

然而, 本文基于层间相似性的SSAM方法是考虑了最直观的邻居拓扑重叠系数作为相似性的度量指标, 而其他相似性指标对于该方法的应用及效果是我们未来可以做进一步讨论和研究的问题. 另外, 本文只考虑同一个节点在相邻时间层的关系, 而实际中, 节点之间的联系不仅依赖于上一时间层内的信息, 还可依赖于之前多个时间片段内的信息, 并可影响之后多个时间片段, 因此具有非马尔可夫特性; 且本文的方法只适用于小规模的数据, 这就需要新的更为准确并能用于大规模数据的方法来描述时序网络. 最后, 如何进行时序网络的切分及时间窗大小的选择一直是个困难的问题^[5], 不同的切分情况会导致不同的实验结果, 因此如何结

合时序网络的全局结构和网络动力学来对时序网络进行构建, 并识别出重要节点是未来需要深入研究的问题.

感谢上海理工大学复杂系统科学研究中心林坚洪、杨凯、胡小军和郭昕字的交流与讨论.

参考文献

- [1] Holme P, Saramäki J 2013 *Temporal Networks* (Heidelberg: Springer) pp1–2
- [2] Holme P, Saramäki J 2012 *Phys. Rep.* **519** 97
- [3] Holme P 2015 *Eur. Phys. J. B* **88** 234
- [4] Liu J G, Ren Z M, Guo Q 2013 *Physica A* **392** 4154
- [5] Ren Z M, Zeng A, Chen D B, Liao H, Liu J G 2014 *EPL* **106** 48005
- [6] Liu J G, Ren Z M, Guo Q, Wang B H 2013 *Acta Phys. Sin.* **62** 178901 (in Chinese) [刘建国, 任卓明, 郭强, 汪秉宏 2013 物理学报 **62** 178901]
- [7] Lü L Y, Chen D B, Ren X L, Zhang Q M, Zhang Y C, Zhou T 2016 *Phys. Rep.* **650** 1
- [8] Ren Z M, Shao F, Liu J G, Guo Q, Wang B H 2013 *Acta Phys. Sin.* **62** 128901 (in Chinese) [任卓明, 邵凤, 刘建国, 郭强, 汪秉宏 2013 物理学报 **62** 128901]
- [9] Liu J G, Lin J H, Guo Q, Zhou T 2016 *Sci. Rep.* **6** 21380
- [10] Zhang Y Q, Cui J, Zhang S M, Zhang Q, Li X 2016 *Eur. Phys. J. B* **89** 26
- [11] Tang J, Musolesi M, Mascolo C, Latora V 2009 *Proceedings of the 2nd ACM Workshop on Online Social Networks* Barcelona, Spain, August 17–17, 2009 p31
- [12] Tang J, Scellato S, Musolesi M, Mascolo C, Latora V 2010 *Phys. Rev. E* **81** 055101
- [13] Deng D M, Zhu J, Chen D B, Gao H 2013 *Comput. Sci.* **40** 26 (in Chinese) [邓冬梅, 朱建, 陈端兵, 高辉 2013 计算机科学 **40** 26]
- [14] Deng D M 2014 *M. S. Dissertation* (Chengdu: University of Electronic Science and Technology of China) (in Chinese) [邓冬梅 2014 硕士学位论文 (成都: 电子科技大学)]
- [15] Kim H, Anderson R 2012 *Phys. Rev. E* **85** 026107
- [16] Huang D W, Yu Z G 2017 *Sci. Rep.* **7** 41454
- [17] Taylor D, Myers S A, Clauset A, Porter M A 2017 *Multiscale Model. Simul.* **15** 537
- [18] Zhu Y X, Zhang F L, Qin Z G 2014 *J. Comput. Appl.* **34** 3184 (in Chinese) [朱义鑫, 张凤荔, 秦志光 2014 计算机应用 **34** 3184]
- [19] Génois M, Vestergaard C L, Fournet J, Panisson A 2015 *Network Sci.* **3** 326
- [20] Klimt B, Yang Y 2004 *Machine Learning: ECML* **2004** 217
- [21] Zhang Z K, Liu C, Zhan X X, Lu X, Zhang C X, Zhang Y C 2016 *Phys. Rep.* **651** 1–34
- [22] Liu C, Zhan X X, Zhang Z K, Sun G Q, Hui P M 2015 *New J. Phys.* **17** 113045
- [23] Liu C, Zhang Z K 2014 *Commun. Nonlinear Sci. Numerical Simulat.* **19** 896
- [24] Kendall M G 1938 *Biometrika* **30** 81
- [25] Agresti A 2010 *Analysis of Ordinal Categorical Data* (2nd Ed.) (New York: John Wiley & Sons John Wiley & Sons) pp188–191

Node importance identification for temporal network based on inter-layer similarity*

Yang Jian-Nan¹⁾ Liu Jian-Guo^{2)†} Guo Qiang¹⁾

1) (Complex Systems Science Research Center, University of Shanghai for Science and Technology, Shanghai 200093, China)

2) (Institute of Financial Technology Laboratory, Shanghai University of Finance and Economics, Shanghai 200433, China)

(Received 18 October 2017; revised manuscript received 29 November 2017)

Abstract

Measuring node centrality is important for a wealth of applications, such as influential people identification, information promotion and traffic congestion prevention. Although there are many researches of node centrality proved, most of them have assumed that networks are static. However, many networks in our real life are dynamic, and the edges will appear or disappear over time. Temporal network could describe the interaction order and relationship among network nodes more accurately. It is of more important theoretical and more practical significance to construct proper temporal network model and identify vital nodes. In this paper, by taking into account the coupling strength between different network layers, we present a method, namely similarity-based supra-adjacency matrix (SSAM) method, to represent temporal network and further measure node importance. For a temporal network with N nodes and T layers, the SSAM is a matrix of size $NT \times NT$ with a collection of both intra-layer relationship and inter-layer relationship. We restrict our attention to inter-layer coupling. Regarding the traditional method of measuring the node similarity of nearest-neighbor layers as one constant value, the neighbor topological overlap information is used to measure the node similarity for the nearest-neighbor layers, which ensures that the couplings of different nodes of inter-layer relationship are different. We then compute the node importance for temporal network based on eigenvector centrality, the dominant eigenvector of similarity-based supra-adjacency matrix, which indicates not only the node i 's importance in layer t but also the changing trajectory of the node i 's importance across the time. To evaluate the ranking effect of node importance obtained by eigenvector-based centrality, we also study the network robustness and calculate the difference of temporal global efficiency with node deletion approach in this work. In order to compare with the traditional method, we measure the node ranking effect of different time layers by the Kendall rank correlation coefficient of eigenvector centrality and the difference of temporal global efficiency. According to the empirical results on the workspace and Enrons datasets for both SSAM method and tradition method, the SSAM method with neighbor topological overlap information, which takes into account the inter-layer similarity, can effectively avoid overestimating or underestimating the importance of nodes compared with traditional method with one constant value. Furthermore, the experiments for the two datasets show that the average Kendall's τ could be improved by 17.72% and 12.44% for each layer network, which indicates that the node similarity for different layers is significant to construct temporal network and measure the node importance in temporal network.

Keywords: temporal network, inter-layer similarity, eigenvector-based centrality, temporal global efficiency

PACS: 89.75.Da, 05.10.-a

DOI: 10.7498/aps.67.20172255

* Project supported by the National Natural Science Foundation of China (Grant Nos. 61773248, 71771152).

† Corresponding author. E-mail: liujg004@ustc.edu.cn