

时序超网络上重要节点挖掘方法研究

詹秀秀¹, 余小燕¹, 刘闯¹, 张子柯²

(1. 杭州师范大学 阿里巴巴复杂性科学研究中心, 杭州 311121; 2. 浙江大学 传媒与国际文化学院, 杭州 310058)

摘要: 在如何识别时序超网络上的重要节点方面取得了一定的进展。定义了该类网络上度量节点重要性程度的8个中心性方法, 分别侧重于网络不同的拓扑结构性质和时间特征, 从多个角度综合考虑了该类网络上节点的重要性。同时, 构建了时序超网络上的SI传播模型, 基于该模型提出了随机移除节点的评估方法来衡量所定义的中心性方法的有效性。研究表明, 在时序超网络上, 基于最快到达路径的介数中心性方法是评价该类网络上节点重要性的良好指标。此外, 基于时间分辨率的度和超度中心性方法通过寻找网络的最佳时间分辨率, 可以进一步优化普通的度和超度中心性方法, 弥补了普通方法不能有效考虑网络时间信息的缺点, 且在多个真实网络上表现出与介数中心性方法相当的性能。

关键词: 时序超网络; SI传播模型; 重要节点; 中心性方法

中图分类号: N 94 **文献标志码:** A

Mining methods of important nodes in temporal hyper-networks

ZHAN Xiuxiu¹, YU Xiaoyan¹, LIU Chuang¹, ZHANG Zike²

(1. Alibaba Research Center for Complexity Sciences, Hangzhou Normal University, Hangzhou 311121, China; 2. College of Media and International Culture, Zhejiang University, Hangzhou 310058, China)

Abstract: Some progress has been made on how to identify important nodes in temporal hyper-networks. Eight centrality methods for measuring the importance of nodes in this type of network were defined, focusing on the different topological properties and time characteristics of the network, and comprehensively considering the importance of nodes in this type of network from multiple perspectives. At the same time, a SI spreading model in temporal hyper-networks was constructed. Based on this model, an evaluation method for randomly removing nodes was put forward to measure the effectiveness of the defined centrality methods. The results show that the betweenness centrality method which is based on the fastest arrival path is a good indicator in identifying important nodes in temporal hyper-networks. Furthermore, degree and hyper-degree centrality methods which consider time resolution can optimize the conventional degree and hyper-degree centrality methods by finding the optimal time resolution of the network. This addresses the limitation of traditional methods in

收稿日期: 2023-01-19

基金项目: 国家自然科学基金重大项目(92146001); 浙江省自然科学基金资助项目(LQ22F030008); 国家社会科学基金重大项目(19ZDA324); 杭州师范大学科研启动项目(2021QDL030); 杭州师范大学研究生科研创新推进项目(1115B20500241); 中央高校基本科研项目

第一作者: 詹秀秀(1989-), 女, 副教授. 研究方向: 复杂网络、智能计算. E-mail: zhanxiuxiu@hznu.edu.cn

通信作者: 张子柯(1982-), 男, 教授. 研究方向: 计算传播学. E-mail: zhangzike@gmail.com

effectively considering network temporal information and performs comparably to the betweenness centrality method in multiple real networks.

Keywords: temporal hyper-network; SI spreading model; important nodes; centrality methods

网络结构的异质性和复杂性使得在特定观察角度下,某些节点相比于其他节点更具重要性,这种重要性可通过中心性来量化。如何衡量网络中节点的中心性是网络分析的基础问题之一,该问题研究目标的实质是确定网络中节点的重要性程度,并对网络中的节点进行重要性排序^[1-2]。具有较高中心性值的节点往往是高影响力节点,拥有较强的信息扩散或者疾病传播的能力。因此,找到网络中的高影响力节点具有重要的实际意义。以往的中心性研究成果在现实生活中也获得了广泛应用,如疾病传播控制^[3-4]、广告精准投放^[5]、交通堵塞疏导^[6-8]等。

一般情况下,每个网络都有特定的节点重要性排名,不同的识别方法考虑网络的不同结构属性,会给出不同的排名列表^[9]。截至目前,研究者们已经提出了许多用于识别网络中重要节点的中心性方法。例如:Zareie等^[10]使用熵对有影响力的节点进行排序;Pu等^[11]提出了局部维度来识别重要节点;Teng等^[12]追踪了社交网络中的真实信息流,找到了有影响力的传播者;张喜平等^[13]引入 m 阶邻居节点的概念,提出了一种基于 m 阶邻居节点重要度贡献的复杂网络节点重要度方法;顾亦然等^[14]引入影响度概念,用节点自身的 K -Shell值与对其邻居节点的影响度来表征节点对邻居节点的贡献重要度。

然而,这些对网络节点中心性的研究仅停留在表示成对节点交互的普通静态网络上,当网络的结构发生动态变化时,静态网络并不能完全准确地描述各式各样的现实世界系统,即现实中的交互均是非静态的、随时间不断发展和变化的,而时序网络可以很好地表达这一概念^[15-18]。另外,相比于普通网络所表达的二元交互,现实世界更普遍存在的是表达多个个体间联系的高阶交互,如电子邮件的发送者和接收者、出版物的共同作者、购物网站共同购买一样物品的用户等。放松节点数量限制的超网络是这一类型交互的理想载体^[19-20]。因此,为更准确地表达现实世界的复杂关系,本研究将这两类网络组合,构造了一类全新

的网络——时序超网络。该类网络随着时间的推移,网络的拓扑结构会随时建立和瓦解,且可形成节点数大于2的超边,其为研究真实复杂系统的动力学提供了良好的表示工具^[21]。该框架尤其适用于描述社交系统,典型的如物理接触网络和在线社交网络。在这两种网络所对应的实际场景中,人群间的真实接触和在线会话等交互会随着时间的推移发生变化,此类交互往往是高阶的,并在时间上具有突变性和持续性。已有研究表明,在研究人类时间动力学时,考虑社会互动的高阶结构是非常重要的^[21]。时序超网络能以超高的时间分辨率描述现实世界动态演化过程,并从中挖掘到具有高影响力的重要节点,该过程对于上述问题的解决具有非常重要的参考意义。然而,目前关于时序超网络上节点中心性评估方法的研究还寥寥无几,且还未形成系统化的定义。

本文的研究目标是在时序超网络上寻找重要的节点,通过构建基于时序超网络的易感-感染(susceptible-infected, SI)传播模型来模拟网络的动态演化过程,定义时序超网络上8个节点中心性评估方法。此外,通过攻击不同中心性方法所寻找到的重要性排名较高的节点来评估所提出的中心性指标的有效性,为挖掘时序超网络上的重要节点提供了理论支撑。

1 方 法

1.1 时序超网络的定义与表示

为方便讨论,假设记录网络的时间有限,且开始记录的时间 $t_{\text{start}} = 0$,结束记录的时间 $t_{\text{end}} = T$ 。在时间窗口 $[0, T]$ 上,根据不同时间戳的节点与超边的隶属关系,构建时序超网络 $H_0^T = \{H_0, H_1, \dots, H_T\}$,即时序超网络可以由每个时刻的静态超网络 $H_t = (V_t, E_t)$ 来表示,且 $t \in [0, T]$ 。其中, $H_t = (V_t, E_t)$ 表示属于 t 时刻的一组节点 $V_t = \{v_1, v_2, \dots, v_N\}$ 和一组超边 $E_t = \{e_1, e_2, \dots, e_m\}$ 构成的超网络, N 表示时序超网络中的总节点数, m 表示 H_t 中的超边数。值得注意的是,任意一条超边 e_j 可以包含多个节点,任意一个节点 v_i 也可以隶属于多条超边。在时序超网络

中, 假设每个时间片中的节点个数相同, 而超边的数量会随着时间的推移而发生改变。时序超网络若忽略不同交互所发生的时间, 即为 H_0^T 所对应的静态超网络 $H = (V, E)$, 其中, $V = \{v_1, v_2, \dots, v_N\}$, 超边 $e_j \in E$, 当且仅当 e_j 在 $E_i(t \in [0, T])$ 中至少出现过一次。

图 1(a) 为一个由 3 个时间戳和 6 个节点组成的时序超网络示例, 根据不同时间戳发生的交互所形成的不同连接进行着色。中间的纵轴线代表时间流动的路径, 曲线连接着一个交互所涉及到的多个节点, 不同颜色代表该时刻的高阶交互所

形成的不同超边: 如在 t_1 时刻, 节点 v_1, v_2, v_3 之间的两两交互形成一个拥有 3 个节点的超边 (3-超边); 节点 v_2 和 v_4 之间的交互形成拥有两个节点的超边 (2-超边)。若忽略图 1(a) 中时序超网络每条超边出现的时间信息, 就可以得到其对应的静态超网络, 如图 1(b) 所示。

1.2 时序超网络上节点中心性方法

为了挖掘时序超网络上的重要节点, 定义度量该类网络上节点重要性的一系列中心性指标, 包括基于节点的度、超度、接触数、节点之间的距离以及网络的时序信息等。

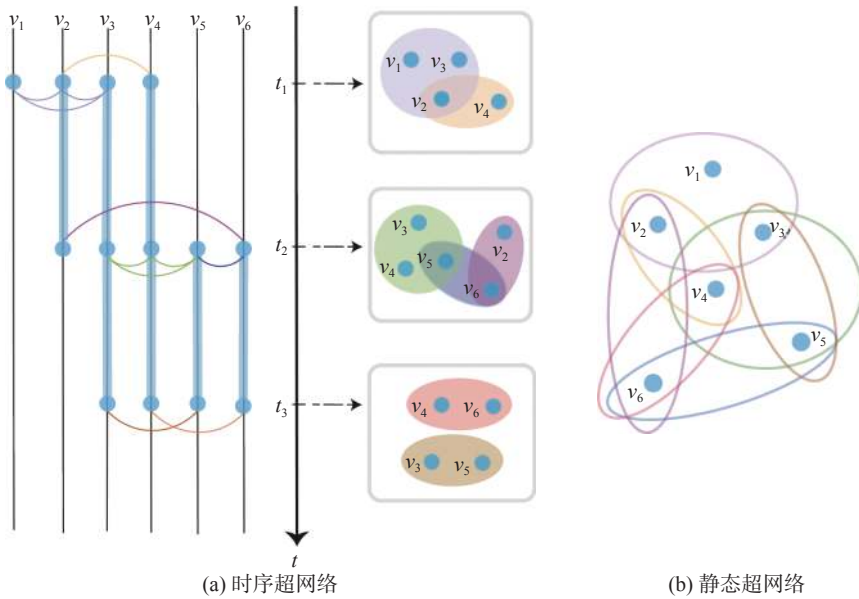


图 1 时序超网络示例

Fig.1 Example of a temporal hyper-network

1.2.1 度中心性

度中心性 (D) 认为节点的重要性与其邻居数量呈正相关。若两节点共同属于同一条超边, 则它们互为邻居。节点 v_i 的度中心性定义为

$$D(i) = f_i \quad (1)$$

式中, f_i 表示时序超网络所对应的静态超网络中节点 v_i 的邻居数。

1.2.2 基于时间分辨率的度中心性

对于一个时序超网络 H_0^T , 可按照时间分辨率 Δt 的大小, 将其划分为若干个时间跨度相同、但网络大小不同的子网络, 即 $H_0^T = \{H_0^{\Delta t}, H_1^{\Delta t}, \dots, H_n^{\Delta t}\}$ 。其中, n 表示子网络的个数。值得注意的是, 每一个子网络仍然是一个时序超网络。

对于给定的时间分辨率 Δt , 节点 v_i 的基于 Δt 的

度中心性 (D_{-t}) 定义为

$$D_{-t}(i, \Delta t) = \frac{1}{n} \sum_{s=1}^n f_i^s \quad (2)$$

式中, f_i^s 表示在时间分辨率 Δt 下划分的第 s 个子网络所对应的静态超网络中节点 v_i 的邻居数。

1.2.3 超度中心性

超度中心性 (W) 认为节点的重要性与其所处的超边数量呈正相关。节点 v_i 的超度中心性定义为

$$W(i) = g_i \quad (3)$$

式中, g_i 表示时序超网络所对应的静态超网络中节点 v_i 处的超边数。

1.2.4 基于时间分辨率的超度中心性

类似于基于时间分辨率的度中心性, 对于给

定的分辨率 Δt ，节点 v_i 的基于 Δt 的超度中心性(W_t)定义为

$$W_t(i, \Delta t) = \frac{1}{n} \sum_{s=1}^n g_i^s \quad (4)$$

式中， g_i^s 表示在时间分辨率 Δt 下划分的第 s 个子网络所对应的静态超网络中节点 v_i 处的超边数。

1.2.5 介数中心性

时序超网络上节点的介数中心性(B)是通过该节点的最快到达路径的比例。节点之间的最快到达路径(p)是从 $t=0$ 开始最早到达下一个节点的路径。令 $S(j, z)$ 表示在时间间隔 $[0, T]$ 上从起始节点 v_j 到目的节点 v_z 的 p 集合，而 $S(j, i, z)$ 则表示 $S(j, z)$ 路径中经过节点 v_i 的 p 集合。那么，时序超网络上节点 v_i 的介数中心性定义为

$$B(i) = \frac{1}{N^2} \sum_{j \neq i \neq z} \frac{\sigma(j, i, z)}{\sigma(j, z)} \quad (5)$$

式中， $\sigma(j, z) \equiv |S(j, z)|$ ， $\sigma(j, i, z) \equiv |S(j, i, z)|$ 。B方法值越高的节点，说明其在网路中充当桥梁的效果越明显，即连接着网络中大部分处于不同时刻超边的节点。

1.2.6 接近中心性

时序超网络上的接近中心性(Q)是衡量节点在其连通分量中到其他各节点的平均最快到达距离。节点之间最快到达距离(d)是其最快到达路径的长度，其值为该路径上包含的节点个数。值得注意的是，大型真实网络往往是不连通的，时序超网络由于时间性质和高阶性质的加持，这种不连通性可能会增加，这意味着网络中的两个节点之间可能并不存在连通的路径。此时，视两个节点间的最快到达距离为无穷大^[22]。

为了避免这种发散问题的出现，受到谐波中心性^[23]启发，将时序超网络上节点 v_i 的接近中心性定义为

$$Q(i) = \frac{1}{N-1} \sum_{i \neq j} \frac{1}{d_{ij}} \quad (6)$$

式中， d_{ij} 表示从节点 v_i 到节点 v_j 的最快到达距离，当不存在从节点 v_i 到节点 v_j 的最快到达路径时，则定义 $\frac{1}{d_{ij}} = 0$ 。Q方法值越高的节点，说明其到达其他节点的 d 值越小，该节点越重要。

1.2.7 接触数中心性

接触数中心性(U)认为节点重要性与节点的接

触数呈正相关。将节点出现在某一时刻的某一超边的情况视作一个接触。节点 v_i 的接触数中心性定义为

$$U(i) = \frac{1}{T} \sum_{t=0}^T c_i^t \quad (7)$$

式中， c_i^t 表示节点 v_i 在 $t \in [0, T]$ 时刻的接触数。

1.2.8 接触邻居数中心性

接触邻居数中心性(V)认为节点的重要性与节点的接触邻居数目呈正相关。接触邻居数即节点每出现在某一时刻时，则统计该节点的邻居数。节点 v_i 的接触邻居数中心性定义为

$$V(i) = \frac{1}{T} \sum_{t=0}^T y_i^t \quad (8)$$

式中， y_i^t 表示节点 v_i 在 $t \in [0, T]$ 时刻的接触邻居节点数。

1.3 评估方法

针对上述8种时序超网络上重要节点挖掘的中心性指标，提出合理可靠的方法来评估这些度量指标的有效性。

1.3.1 SI传播模型

构建时序超网络的SI传播模型，在该模型中，个体被分为两组：易感者(susceptible)和感染者(infected)。为简单起见，让所有传播过程都从 $t=0$ 时刻的一个初始被感染的种子节点开始，其余所有节点的初始状态均被设置为易感状态。按照时间 t 推移，依次遍历每一个时刻，若感染节点出现在 t 时刻的超边中，就会以传播概率 β 随机去感染其邻居易感节点，该时刻的节点感染完成之后，推移到下一时刻进行感染，传播过程在 $t=T$ 时刻结束。最后，依次迭代序列中的每一个节点作为初始被感染的种子节点，进行上述传播过程^[24]。用 $\varphi(t)$ 表示 t 时刻以每个节点作为传播源的平均被感染节点总数。

1.3.2 节点中心性评估方法

采用不同的中心性方法作为节点攻击策略来评估节点重要性方法的有效性。具体步骤如下：a. 对于给定的节点中心性方法，首先用该方法对网络中的节点进行重要性排序；b. 删除重要性排名前 k 的节点序列，构造攻击节点序列后的时序超网络；c. 将原始时序超网络和攻击节点序列后的时序超网络分别执行传播概率 β 下的SI传播模拟，分别得到两条传播曲线 $\varphi_0(t)$ 和 $\varphi_k(t)$ ；d. 比较两

条传播曲线的差异, 即计算两条传播曲线间的面积差。如图 2 所示, 用 φ_0 和 φ_k 之间面积差(蓝色标出)评估所提出的中心性指标的有效性, 评价指标的数学定义如下:

$$E(k) = \frac{\sum_{t=0}^T (\varphi_0(t) - \varphi_k(t))}{\sum_{t=0}^T \varphi_0(t)}$$

(9)

$E(k)$ 越大, 说明攻击该中心性方法挖掘到的重要节点能有效抑制传播, 同时也验证了该方法的有效性。除了上述 8 种中心性指标, 还将随机删除包含 k 个节点的序列作为对比方法(用 Random 表示), 随机实验的平均次数为 50 次。

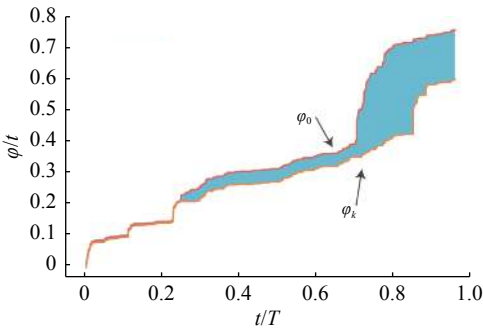


图 2 原始网络和受攻击网络传播曲线之间的面积差示例
Fig.2 Example of the area difference of the spreading curve between the original network and the attacked network

2 实验

2.1 数据集

选择 9 个真实的时序超网络数据集, 个体的

交互覆盖医院、会议、艺术馆、工作场所、研究机构及学校等。通过这些网络来探究不同的中心性方法在具有不同社会背景和拓扑结构性质的真实场景中的适用性。EEU1 和 EEU2 网络描述了欧洲大型研究机构两个不同部门所有成员之间电子邮件的时间网络^[25]。HT2009 网络是 ACM Hypertext 2009 会议期间的的时间网络^[26]。Gallery 网络记录的是一家艺术馆的参观人员的时间网络^[27]。Hospital 网络描述的是法国里昂医院病房中患者、患者和卫生保健工作者以及医护人员之间接触的时间网络^[21]。Workplace 网络描述的是 2015 年在法国办公楼测量的个人之间接触的时间网络^[28]。HS2011 和 HS2012 网络描述的是 2011 年和 2012 年法国马赛一所高中学生之间的联系时间网络^[29]。PS 网络是儿童和教师之间的联系时间网络^[29]。这些网络的基本拓扑性质如表 1 所示。其中: N 为节点数; C 为接触数; T 为网络的最大时间戳; $\langle f \rangle$ 为节点平均度; $\langle g \rangle$ 为节点平均超度; $\langle d \rangle$ 为网络的平均最快到达距离。表 2 展示的是时序超网络中拥有不同度数的超边规模。表中, 超边的度表示一条超边中包含的节点数, 这里只统计至网络中超边的度为 10 的超边数量, “—” 表示网络中不存在该度数的超边。从表 2 可以发现, 随着超边的度的增加, 其超边数量是逐渐减少的, 大多数的真实时序超网络中的交互局限于度为 2, 3, 4 的超边之间, 而邮件网络的超边规模偏大。

2.2 节点重要性指标的性能比较

上文定义了不同时间分辨率下的度和超度, 为了探索不同时间分辨率的中心性指标的性能, 将基于时间分辨率的中心性方法的性能定义为从 $k = 5$ 到 $k = 30$ 范围内有效性 $E(k)$ 曲线下的面积。以

表 1 时序超网络的基本拓扑性质
Tab.1 Basic topological properties of temporal hyper-networks

网络	N	C	T	$\langle f \rangle$	$\langle g \rangle$	$\langle d \rangle$	场景
Hospital	75	27835	9453	30.37	2.43	3.41	医院
HT2009	113	19037	5246	38.87	2.15	2.87	会议
Gallery	186	4647	1240	6.61	2.36	11.86	艺术馆
Workplace	217	73823	18488	39.39	2.16	4.83	工作场所
EEU1	89	8912	8911	37.78	3.04	2.89	研究机构
EEU2	142	26497	26496	19.49	4.74	3.96	研究机构
HS2011	126	26385	5609	27.13	2.27	4.11	学校
HS2012	180	42105	11273	24.67	2.22	4.76	学校
PS	242	106879	3100	68.74	2.42	5.07	学校

EEU2网络为例，在图3(a)中显示了D和W方法的有效性结果，并将这两种方法的性能作为对照组(图3(b)中的static组)。改变网络的时间分辨率 Δt ，从100增至1000，计算基于不同时间分辨率的度(D_t)和超度中心性(W_t)方法的性能。如在图3(b)中，D_t方法的最佳时间分辨率为 $\Delta t = 100$ ，W_t方法的最佳时间分辨率为 $\Delta t = 300$ 。按照这种方式，分别在9个真实网络中，寻找不同方法的最佳时间分辨率，结果如表3所示。如果在某一 Δt 情况下，D_t和W_t方法的性能值大于static组中D和W方法的性能值，则该网络的最佳时间分辨率在表3中用粗体进行标注，可以发现所有网

络均可以通过该方法寻找到最佳的时间分辨率。同时，在绝大多数网络上，D_t和W_t方法的性能比D和W方法的更好。因为这两种方法侧重于考虑不同时期网络中节点的重要性，同时结合了网络的时间性质和高阶性质。

实验选取 $\beta = 1$ 来评估不同中心性方法的效果，结果如图4和图5(a)所示。值得说明的是，D_t和W_t方法选取的均是表3中所列出的最佳时间分辨率。在图5(a)中，以Hospital网络为例，将9个中心性方法在该网络中所得到的性能值组成一个序列X，并将其按照最大最小值方法进行归一化，即 $X_i^* = \frac{X_i - X_{\min}}{X_{\max} - X_{\min}}$ 。式中： X_i 表示序列X的第

表 2 不同时序超网络的超边数量

Tab.2 Number of hyper-edges in different temporal hyper-networks

超边的度	Hospital	HT2009	Gallery	Workplace	EEU1	EEU2	HS2011	HS2012	PS
2	1108	2103	518	4142	806	632	1608	2125	7748
3	657	302	211	755	203	324	470	466	4600
4	58	18	26	12	97	180	50	50	347
5	2	7	3	—	59	105	1	4	9
6	—	4	—	—	40	78	—	—	—
7	—	—	—	—	21	42	—	—	—
8	—	—	—	—	11	25	—	—	—
9	—	—	—	—	5	23	—	—	—
10	—	—	—	—	5	10	—	—	—

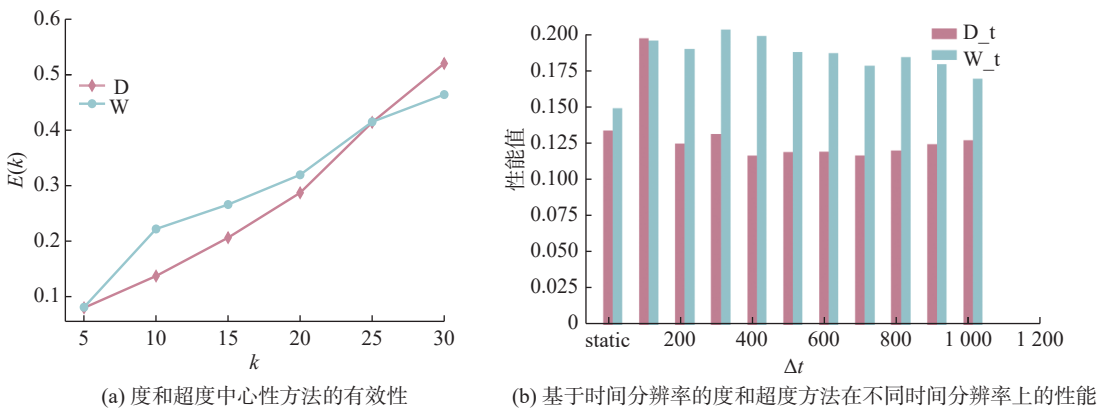


图 3 EEU2 网络中不同中心性方法的有效性和性能

Fig.3 Effectiveness and performance of different centrality methods for EEU2 network

表 3 不同时序超网络的最佳时间分辨率

Tab.3 Optimal time resolutions for different temporal hyper-networks

方法	Hospital	HT2009	Gallery	Workplace	EEU1	EEU2	HS2011	HS2012	PS
D _t	1000	500	400	900	900	100	700	500	300
W _t	700	800	900	900	1000	300	600	900	900

i 个数据; X_i^* 表示序列 $\mathbf{X}$ 的第 i 个数据归一化后得到的结果; $X_{\max}$ 和 $X_{\min}$ 分别表示序列 $\mathbf{X}$ 中的最大值和最小值。归一化后, 得到对应的序列 $\mathbf{X}^*$, 即为 9 个中心性方法的平均性能序列, 在图 5(a)中以不同颜色的柱子进行显示。图 4 和图 5(a)的结果表明, 在时序超网络上定义的中心性方法中, B 方法的有效性和平均性能均明显优于其他中心性方法。可能原因有 2 个: a. B 方法的节点介数中心性值越高, 说明节点更大范围地连接着不同时刻的超边, 在网络中充当桥梁的作用越明显; b. B 方法在计算距离时是基于节点的最快到达路径的思想, 因此在考虑网络全局特征的同时也考虑了时间的全域性, 更加适用于时序超网络。另外, 图 4 出现了两个有趣的现象。首先, 在 Gallery 网络中, Q 方法表现优异, 这可能是因为 Gallery 网络的节点平均度较小, 平均最快到达距离较长, 如表 1 所示。但在其余真实的时序超网络中, 该方法的有效性并不是很高, 尤其在 Workplace

网络中, Q 方法的有效性甚至和对比方法的接近, 这说明 Q 方法在度量时序超网络上节点的重要性时并不具备普适性。其次, 拟合的曲线会出现下降的趋势, 如 Gallery 网络中的 V 方法和 EEU2 网络中的 Q 方法, 此现象意味着这些方法挖掘到的节点在网络中的重要性程度并不高。图 5(a)中, Q, U, V 等方法的效果不如其他方法。可以清晰地观察到, 在大多数网络中, D_t和 W_t方法的平均性能比 D 和 W 方法的更好。这说明通过寻找最佳时间分辨率, 可以进一步有效优化普通的 D 和 W 方法。同时, 实验验证了在 $\beta = 0.5$ 的情况下, 不同中心性方法及对比方法之间的平均性能与上述 $\beta = 1$ 情况下的实验结果类似, 如图 5(b)所示。

2.3 肯德尔相关性

肯德尔相关性系数 τ 可以衡量两个不同变量之间的相关性^[30]。假设对于两个随机变量 Y 和 Z , 它们的第 i 个组合用 (Y_i, Z_i) 表示。当 $Y_i > Y_j$ 和 $Z_i > Z_j$ 或 $Y_i < Y_j$ 和 $Z_i < Z_j$ 同时出现时, (Y_i, Z_i) 和 (Y_j, Z_j) 被认

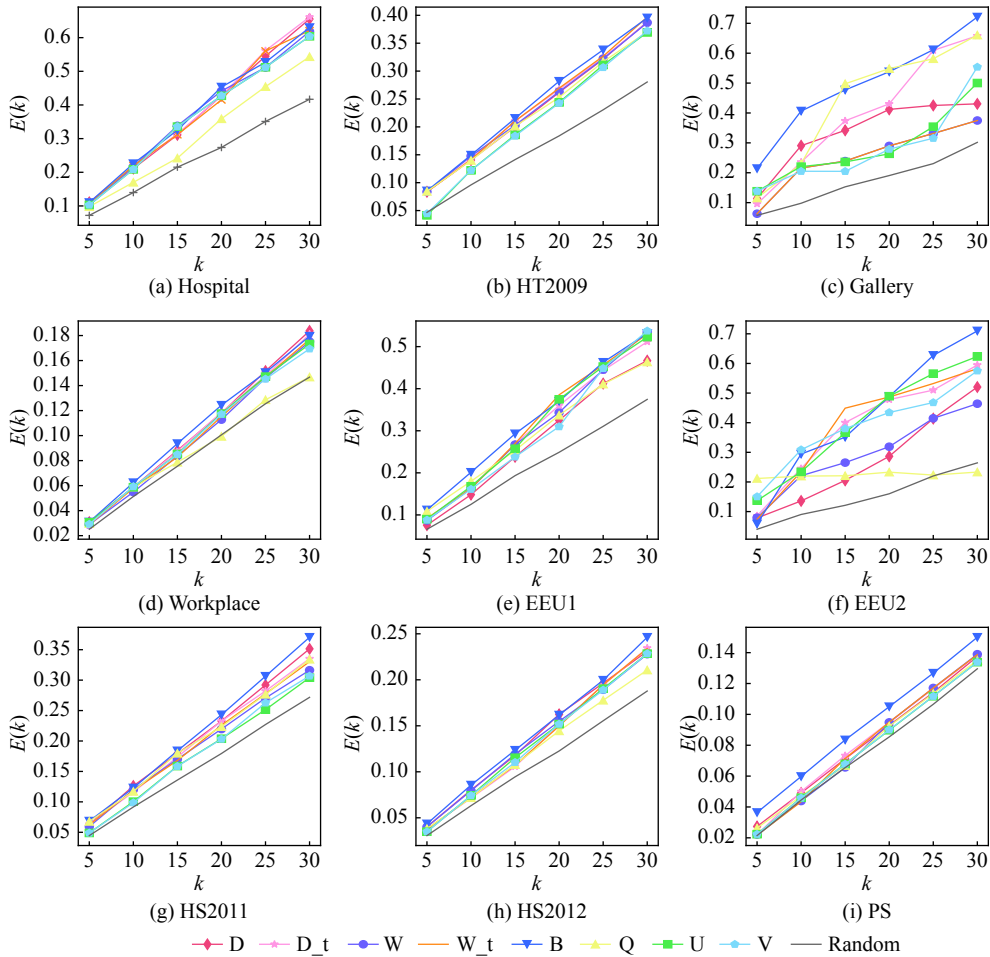


图 4 节点中心性方法的性能($\beta = 1$)

Fig.4 Performance of node centrality methods ($\beta = 1$)

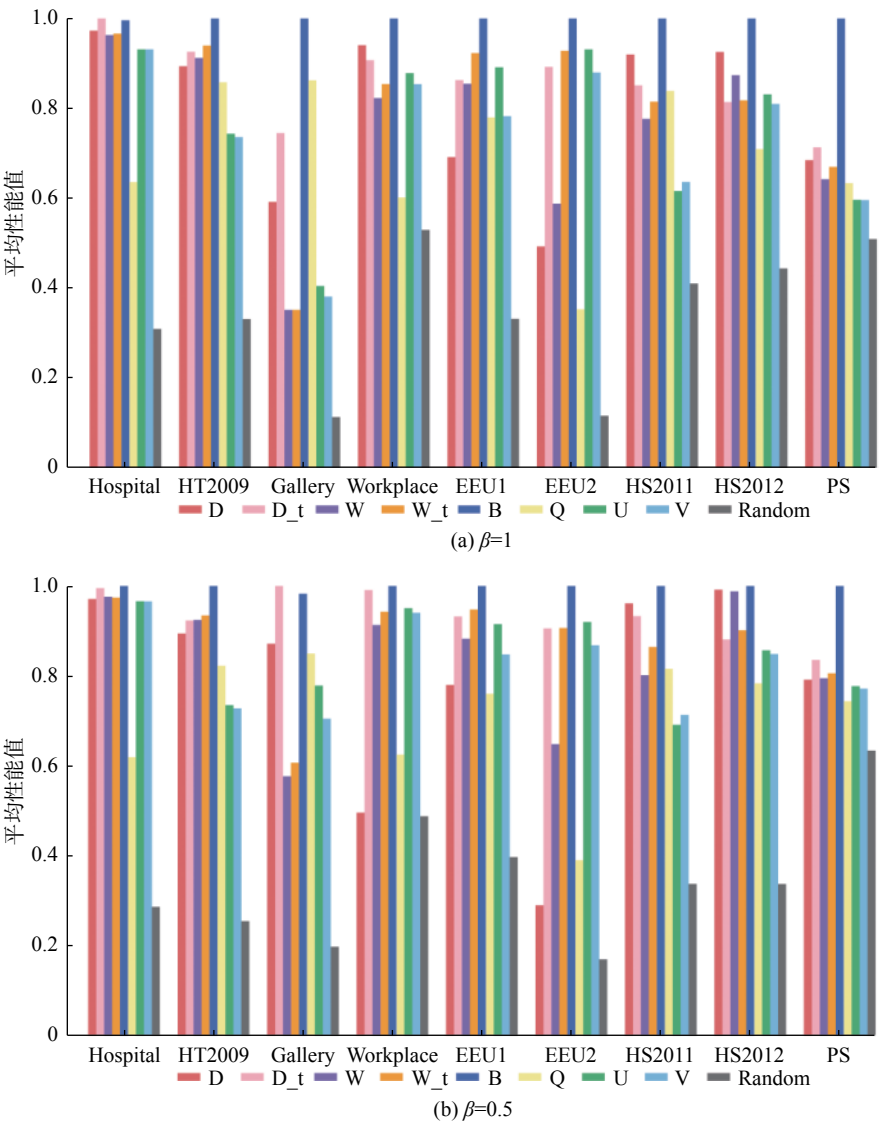


图 5 中心性方法在不同真实时序超网络中的平均性能比较

Fig.5 Comparison of average performance of centrality methods in different real temporal hyper-networks

为具有一致性，反之则不具有。因此，肯德尔相关性系数 τ 的定义如下：

$$\tau = \frac{2(n^+ - n^-)}{N(N - 1)} \tag{10}$$

式中： n^+ 表示具有一致性的序列个数； n^- 表示不一致的序列个数。 $\tau = 1$ 表示不同方法得到的节点重要性排序列表相同； $\tau = 0$ 则表示相反的情况。 τ 的值越大，说明两种中心性方法的关系越相关。本文采用该方法来检测定义的不同中心性方法之间的相关性。

图 6 显示了 9 个真实时序超网络不同中心性方法之间的肯德尔相关性结果。首先，可以观察到 D, D_t, W, W_t这 4 个方法之间的颜色最深，即

它们之间的相关性最高。虽然这 4 个方法都是侧重于度量网络中节点关于度的信息来判断节点的重要性，但是从图 5(a)可以看出，D_t, W_t方法会明显比 D, W 方法的平均性能更高，这也进一步说明度量节点中心性时考虑网络时间信息的重要性。其次，U 和 V 方法都是从节点的接触次数出发，判断节点的重要性时考虑的因素是相似的，所以这两种方法之间的相关性程度很高。且在图 4 中，这两种方法的曲线重合度也很高，这种情况再次说明了这两种方法具有很高的相似性。但是，节点的接触数和接触邻居数同节点的度和超度是有差别的，因为它们考虑了节点出现的时间信息，统计了节点在不同时刻出现的次数，所以

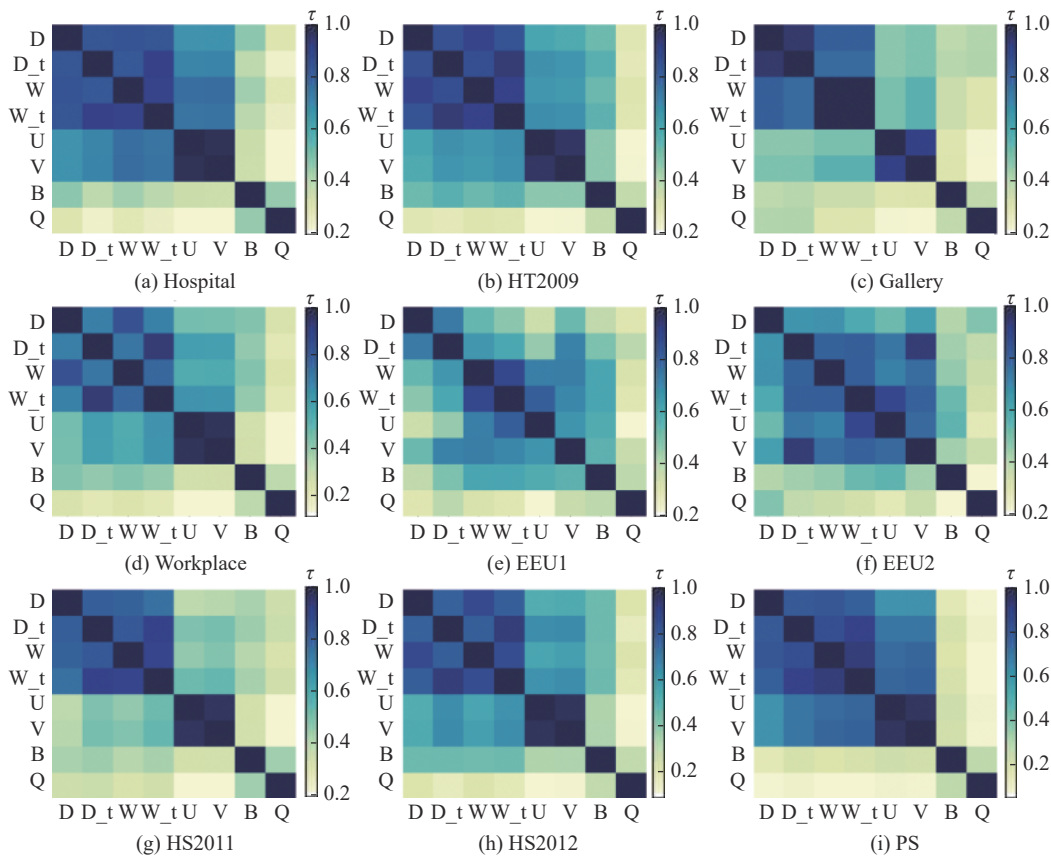


图 6 不同真实时序超网络上中心性方法之间的相关性

Fig.6 Correlation between centrality methods in different real temporal hyper-networks

U和V方法同D, D_t, W, W_t方法之间的相关性程度减弱。最后, B和Q方法分别从节点位于网络中的位置和节点到其他节点之间的最快到达距离等方面来衡量节点的重要性, 因此, 它们之间的相关性程度略低, 同其他方法之间的相关性也类似。

3 总结与展望

构建了一种全新类型的网络——时序超网络, 其同时兼具网络的高阶性质和时间性质, 能更加准确、有效地描绘真实世界的社会系统。此外, 基于节点的最快到达路径计算了时序超网络上节点之间的最快到达距离。接下来, 针对挖掘该类网络上重要节点的中心性方法匮乏且不成体系等问题, 提出了 8 个中心性方法和评估指标, 并在 9 个不同领域的真实网络上进行了实验研究。结果表明: 提出的所有中心性方法中, 介数中心性方法总是具有优越性的, 其能有效地挖掘出时序超网络上具有高影响力的重要节点; 在基于时间分辨率的中心性方法中, 通过寻找网络的最佳时间分辨率, 可以进一步优化普通的度和超

度中心性方法, 解决了在时序超网络上计算度或者超度时不能有效地考虑网络时间信息的问题, 该想法的合理性经过实验验证得到了充分的证实。同时, 基于时间分辨率的度和超度中心性方法也表现得极具竞争力。综上所述, 在真实世界中, 为了快速且有效地控制疾病和信息的传播, 一方面可以通过寻找具有“枢纽”作用的节点并中断其传播能力, 另一方面要关注节点出现的时间信息, 以及节点在不同时间窗口下影响力的差异。

本文基于时序超网络定义的一系列中心性方法, 重点考虑了节点的度、超度和接触数、节点之间的路径和距离等信息, 并结合网络的时间性质和高阶性质, 提出了一种能够囊括这两个特性且效果优异的中心性方法。因此, 本文工作弥补了当前研究的不足, 为后续深入分析时序超网络奠定了较为坚实的基础。

然而, 在时序超网络上挖掘重要节点的中心性方法仍存在许多潜在问题: 首先, 本文所提出的中心性方法还未结合节点的更多方面特征来综合度量节点的重要性; 其次, 在寻找网络的最佳时间分辨率时, 本文将网络按照固定时间分辨率

进行划分,可能会遗漏一些重要信息。因此,可以根据网络时间讯息的不同,按小时、天数、月份、年份等信息来划分网络,当然,这不可避免地会增加网络信息收集工作的复杂度和任务量。未来研究中,可以根据此类网络的特点改变或补充一些因素,使方法更具普适性和有效性。

参考文献:

- [1] 郭晓成, 马润年, 王刚. 复杂网络中节点重要性综合评价方法研究[J]. *计算机仿真*, 2017, 34(7): 264–268.
- [2] WEN T, DENG Y. Identification of influencers in complex networks by local information dimensionality[J]. *Information Sciences*, 2020, 512: 549–562.
- [3] BELL D C, ATKINSON J S, CARLSON J W. Centrality measures for disease transmission networks[J]. *Social Networks*, 1999, 21(1): 1–21.
- [4] ZENG Q, LIU Y, TANG M, et al. Identifying super-spreaders in information-epidemic coevolving dynamics on multiplex networks[J]. *Knowledge-Based Systems*, 2021, 229: 107365.
- [5] 陈峥, 高翔, 梁恒, 等. 基于复杂网络的在线广告投放方法: 201610741623.3[P]. 2017-01-04.
- [6] 李树彬, 吴建军, 高自友, 等. 基于复杂网络的交通拥堵与传播动力学分析[J]. *物理学报*, 2011, 60(5): 050701.
- [7] 周艳, 李妍羲, 江荣贵, 等. 交通拥堵与预警信息交互传播动力学分析[J]. *地球信息科学学报*, 2017, 19(10): 1279–1286.
- [8] 邢国正, 杜萍萍, 方晗琦, 等. 基于复合复杂网络交通运输系统研究[J]. *南阳理工学院学报*, 2017, 9(4): 54–57.
- [9] LI Z, HUANG X Y. Identifying influential spreaders by gravity model considering multi-characteristics of nodes[J]. *Scientific Reports*, 2022, 12(1): 9879.
- [10] ZAREIE A, SHEIKHAHMADI A, FATEM A. Influential nodes ranking in complex networks: an entropy-based approach[J]. *Chaos, Solitons & Fractals*, 2017, 104: 485–494.
- [11] PU J, CHEN X W, WEI D J, et al. Identifying influential nodes based on local dimension[J]. *Europhysics Letters*, 2014, 107(1): 10010.
- [12] TENG X, PEI S, MORONE F, et al. Collective influence of multiple spreaders evaluated by tracing real information flow in large-scale social networks[J]. *Scientific Reports*, 2016, 6(1): 36043.
- [13] 张喜平, 李永树, 刘刚, 等. 节点重要度贡献的复杂网络节点重要度评估方法[J]. *复杂系统与复杂性科学*, 2014, 11(3): 26–32, 49.
- [14] 顾亦然, 王兵, 孟繁荣. 一种基于 K -Shell 的复杂网络重要节点发现算法[J]. *计算机技术与发展*, 2015, 25(9): 70–74.
- [15] 李佳佳. 动态网络中心性方法分析[D]. 成都: 西安电子科技大学, 2012: 12–15.
- [16] 董永强, 祖倩倩, 陶桦. 基于真实数据集的动态网络时间中心性分析[J]. *华中科技大学学报(自然科学版)*, 2015, 43(5): 109–113.
- [17] 付凯, 夏靖波, 赵小欢. 动态融合复杂网络节点重要度评估方法[J]. *哈尔滨工业大学学报*, 2017, 49(10): 112–119.
- [18] 田亮. 复杂网络的统计性质及其上动力学分析[D]. 南京: 南京航空航天大学, 2012.
- [19] 胡枫, 赵海兴, 马秀娟. 一种超网络演化模型构建及特性分析[J]. *中国科学: 物理学 力学 天文学*, 2013, 43(1): 16–22.
- [20] 王众托. 关于超网络的一点思考[J]. *上海理工大学学报*, 2011, 33(3): 229–237.
- [21] CENCETTI G, BATTISTON F, LEPRI B, et al. Temporal properties of higher-order interactions in social networks[J]. *Scientific Reports*, 2021, 11(1): 7028.
- [22] 汪小帆, 李翔, 陈关荣. 复杂网络理论及其应用[M]. 北京: 清华大学出版社, 2006.
- [23] AKSOY S G, JOSLYN C, MARRERO C O, et al. Hypernetwork science via high-order hypergraph walks[J]. *EPJ Data Science*, 2020, 9(1): 16.
- [24] ZHAN X X, HANJALIC A, WANG H J. Suppressing information diffusion via link blocking in temporal networks[C]//Proceedings of the Eighth International Conference on Complex Networks and Their Applications. Cham: Springer, 2019: 448–458.
- [25] PARANJAPPE A, BENSON A R, LESKOVEC J. Motifs in temporal networks[C]//Proceedings of the Tenth ACM International Conference on Web Search and Data Mining. Cambridge: ACM, 2017: 601–610.
- [26] ISELLA L, STEHLÉ J, BARRAT A, et al. What's in a crowd? Analysis of face-to-face behavioral networks[J]. *Journal of Theoretical Biology*, 2011, 271(1): 166–180.
- [27] VAN DEN BROECK W, QUAGGIOTTO M, ISELLA L, et al. The making of sixty-nine days of close encounters at the science gallery[J]. *Leonardo*, 2012, 45(3): 285–285.
- [28] ZHANG S L, ZHAO X Y, WANG H J. Mitigate SIR epidemic spreading via contact blocking in temporal networks[J]. *Applied Network Science*, 2022, 7(1): 2.
- [29] ZHAN X X, LIU C, WANG Z P, et al. Measuring and utilizing temporal network dissimilarity[DB/OL]. (2021–11–02)[2023-01-02]. <https://arxiv.org/abs/2111.01334>.
- [30] LV L Y, CHEN D B, REN X L, et al. Vital nodes identification in complex networks[J]. *Physics Reports*, 2016, 650: 1–63.