doi：10. 12068/j. issn. 1005－3026. 2025. 20240057

基于多种特征综合识别时序网络中的影响力传播者

赵 海，杨树坤，缪九男，尉雪龙
（东北大学 计算机科学与工程学院，辽宁 沈阳 110169）

摘 要：在时序网络中精准识别有影响力的传播者对产品推广、谣言抑制等领域至关重要．针对现有方法多依赖单一特征（邻居数量、节点位置或传播能力）而忽略特征间相互作用导致准确率低的问题，提出基于时序引力（TG）模型和信息熵的识别方法（TGBISR），旨在从多特征融合角度提升识别准确性．该方法首先利用TG模型分析用户的度中心性、紧密中心性和介数中心性，分别刻画其局部、位置和全局特征；进而通过信息熵衡量各特征的信息含量并赋予不同权重，加权综合计算用户影响力．为评估效果，在4个真实数据集上使用易感–感染–恢复（SIR）模型模拟信息传播以获取用户真实影响力，并通过肯德尔相关系数和回归分析比较TGBISR计算结果与真实值的相关性．实验结果表明，TGBISR方法在识别有影响力传播者方面，其计算结果与SIR模型真实影响力展现出更高的统计相关性，准确性显著且稳定地优于其他5种基准算法．

关 键 词：时序网络；影响力传播者；时序引力模型；信息熵
中图分类号：TP 301.6 **文献标志码：**A **文章编号：**1005-3026（2025）10-0010-09

Comprehensive Identification of Influential Spreaders in Temporal Networks Considering Multiple Features

ZHAO Hai，*YANG Shu-kun*，*MIAO Jiu-nan*，*YU Xue-long*
（School of Computer Science & Engineering，Northeastern University，Shenyang 110169，China. Corresponding author：YU Xue-long，E-mail：primelongyu@gmail.com）

Abstract：Accurately identifying influential spreaders in temporal networks is crucial for product promotion，rumor suppression，and other aspects. Existing methods mostly rely on a single feature（the number of neighbors，node location，or propagation ability）and ignore interactions between features，resulting in low accuracy. Therefore，a temporal gravity（TG）model and an information entropy-based identification method（TGBISR）were proposed to improve identification accuracy by fusing multiple features. First，the TG model was used to analyze the degree centrality，closeness centrality，and betweenness centrality of the user，portraying their local，positional，and global features，respectively. Then，the information content of each feature was measured through information entropy，and different weights were assigned to them to comprehensively compute the user's influence. To verify the result，the susceptible–infected–recovered（SIR）model was used to simulate information dissemination on four real datasets to obtain the real influence of users. The correlation between the TGBISR calculation results and the real values was then compared using Kendall's correlation coefficient and regression analysis. The experimental results show that the TGBISR method's calculated results exhibit a higher statistical correlation with the true influence of the SIR model when identifying influential spreaders，and its accuracy significantly and consistently outperforms that of the other five benchmark algorithms.

Key words：temporal network；influential spreader；temporal gravity model；information entropy

收稿日期：2024-03-11
基金项目：中央高校基本科研业务费专项资金资助项目（2020GFZD014）.
作者简介：赵 海（1959—），男，辽宁沈阳人，东北大学教授，博士生导师.

识别有影响力的传播者对社会稳定、经济发展和信息安全具有重要作用[1].影响力传播者是在社交网络中负责大范围信息传播和接收的用户[2],而社交网络则是由大量用户以及每时每刻频繁的信息交互构成的[3].随着社交网络的普及,人们之间的交流变得更加方便快捷,社交网络影响着人类的行为[4],并成为传播思想、信息和广告的理想平台.由于影响力传播者承担了社交网络中绝大部分的交互,因此他们不仅成为了积极信息的理想传播介质,也可能被负面信息利用[5],对社会造成严重的负面影响.在此情况下,如何高效地识别有影响力的传播者已成为一个亟待解决的问题.

为了识别有影响力的传播者,绝大多数现有方法都采用了识别复杂网络中关键节点的技术.在关键节点识别的过程中,这些方法将用户及用户之间的互动视为复杂网络中的节点和边[6],而有影响力的传播者则用具有高中心性或重要性的节点表示[7].主流的复杂网络关键节点识别方法可分为两类:基于静态网络的方法和基于时序网络的方法.

静态网络的固定边能够正确描述社交网络中两个用户之间是否存在交互,但是这种恒定边却存在局限性,即静态网络认为社交网络是静态的而不是动态的,这就导致了静态网络忽略了真实社交网络中信息的间歇性和短暂性[6].与静态网络相比,时序网络能够更恰当地表示具有动态网络结构的社交网络[2].时序网络在每一时刻的网络结构都可被视为静态网络[8],因此用户在时序网络中的影响力可通过衡量其在每个时刻的影响力来计算[9].

人们提出了许多方法来识别时序网络中有影响力的传播者,在这些方法中,有影响力的传播者被视作时序网络中的关键节点.例如,Rocha等[10]提出了基于周期性边界条件下随机游走的TempoRank中心性来衡量节点的重要性;Wu等[11]将一段时间内节点邻居的变化作为节点的影响力指标,提出了时序邻域变化中心性.

但是,现有的方法存在一些不足.首先,这些方法仅从单一角度对用户的影响力进行评估.社交网络中的用户并不具备同等影响力,与普通用户相比,具有影响力的用户在网络的主要功能以及整体结构中扮演更为重要的角色,而且具有影响力的用户拥有更多显著特征[12].因此仅从一个角度或者特征评估用户的影响力是不合适的.其次,用户在时序网络中的局部、位置以及全局特征没有被综合考虑.在现实生活中,相对于普通用户,具有影响力的用户(传播者)往往在以下方面表现出色:网络中传播信息的位置,广泛接收并传播信息的能力,以及与该用户直接交互的用户数量.在时序网络的每一个时刻,用户的度中心性代表了节点的局部特征[13].因为度中心性[14]与该用户直接交互的邻居数量成正比,这代表了该用户的局部影响力.用户与网络中其他用户之间的距离表示紧密中心性[15],这被视为该用户的位置特征[16].紧密中心性越高,表示该用户与其他用户之间的距离越短,则该用户在传播或接收消息时能更迅速.从整个网络的全局视角看,介数中心性[17]表示整个网络中经过某一个节点的信息传播路径.当某一节点被多个信息传播路径所包含时,则该节点在整个网络中信息的传播和接收能力越突出,节点的全局特征[18]越显著.为综合评估以上特征,本文使用TG模型分析用户在每个时刻的局部、位置以及全局特征,并将平均值作为该用户在整个时序网络中的特征[9].

基于上述观点,本文使用时序网络中的节点、边以及中心性表示用户、交互以及影响力.为了综合识别影响力传播者,本文利用TG模型和信息熵提出基于时序引力模型的时序网络传播者识别方法(temporal gravity model based influential spreader recognition,TGBISR).

1　节点中心性相关工作

讨论在时序网络中识别影响力传播者技术的研究现状.所有方法都将有影响力的传播者和用户的影响力分别表示为一个关键节点和节点的中心性.现有方法可分为基于网络拓扑的方法[11]、基于动力学的方法[10]和基于机器学习的方法[19].

基于拓扑的方法大多通过分析时序网络中节点的拓扑特征来评估用户的影响力.Kim等[20]将静态网络中度中心性、紧密中心性和介数中心性进行拓展延伸,并应用至时序网络中.Wang等[21]对时序度的方法进行偏差计算,提出度偏差中心性.Elmezain等[22]将邻居节点作为一项评价指标,提出Temporal Degree-Degree(TDD)和Temporal Closeness-Closeness(TCC)中心性.两种方法都认为邻居是中心性的主要组成部分.Wu等[11]根据节点一段时间内邻居数量的变化提出

时序邻域变化中心性.这些方法在单一特征明显的网络中表现良好,但是它们缺乏对用户多种特征的综合评估.

关于基于动力学的方法,所有用户之间的交互都是通过模拟节点间的信息传播来表示.它们从用户之间交互以及信息传播的角度来衡量用户的影响力.Rocha等[10]将随机游走应用到时序网络,以此来衡量节点的中心性.Taylor等[23]对现有的以特征向量为基础的中心性进行总结,提出了边际中心性和条件中心性.Lv等[24]使用6阶张量表示时序网络并提出了适用于多层网络的特征向量中心性以及PageRank中心性.与基于网络拓扑结构的方法相比,基于动力学的方法能更真实地模拟信息传播过程,但由于过程的复杂性,这种方法会消耗大量计算资源.

在机器学习方面,很多框架都被用来识别时序网络中的影响力传播者,如表示学习[25]、图神经网络[26]、图卷积网络[6,19,27]、强化学习[28]等.为了更准确地识别影响力传播者,这些方法通过在大量数据集上训练模型,以此训练模型识别最具影响力的传播者.与基于动力学的方法相比,这些方法的优势在于对计算资源的要求较低[12].然而,能在质量和数量两个方面满足训练需求的数据集却极为罕见.

基于动力学的方法在模拟信息传播的过程中需要大量计算资源.在机器学习中,理想的模型需要通过在大量的高质量数据集上进行训练才能得到,但高质量的数据集在生活中难以收集.基于网络拓扑的方法所需的计算资源和数据集较少,但大部分只能从度中心性或紧密中心性等特征来衡量节点的中心性,无法全面评估整个网络中节点的中心性.以上3类方法都有各自的缺点,这导致识别时序网络中影响力传播者的准确率较低,但识别影响力传播者仍是一项需要高效方法的重要任务.

2 节点中心性基本方法

给出时序网络的定义、现有的时序网络中关键节点算法,以及TG模型,为后续TGBISR算法以及实验设计做准备.

2.1 时序网络

一个时序网络可被定义为$G=(V,E,T)$,其中V表示节点集,E表示时序边集,T表示网络G的持续时间为$[1,T]$.在时序网络G中,每一条边$e(v_i,v_j,t)\in E$表示一条连接节点v_i和v_j的时序边存在于t时刻.假设时间分辨率为Δt,时序网络G可以被分解为一系列的快照网络$G_1,G_2,\cdots,G_n$,其中$n=T/\Delta t$,每一个$G_m(1\leqslant m\leqslant n)$是一个静态网络.图1为一个时序网络和组成该时序网络的快照网络.时序网络G中所有的节点和边表示用户和用户之间的交互,节点的中心性表示用户的影响力.在本文的后续部分,节点的中心性将被用来表示用户的影响力,关键节点即为影响力传播者.

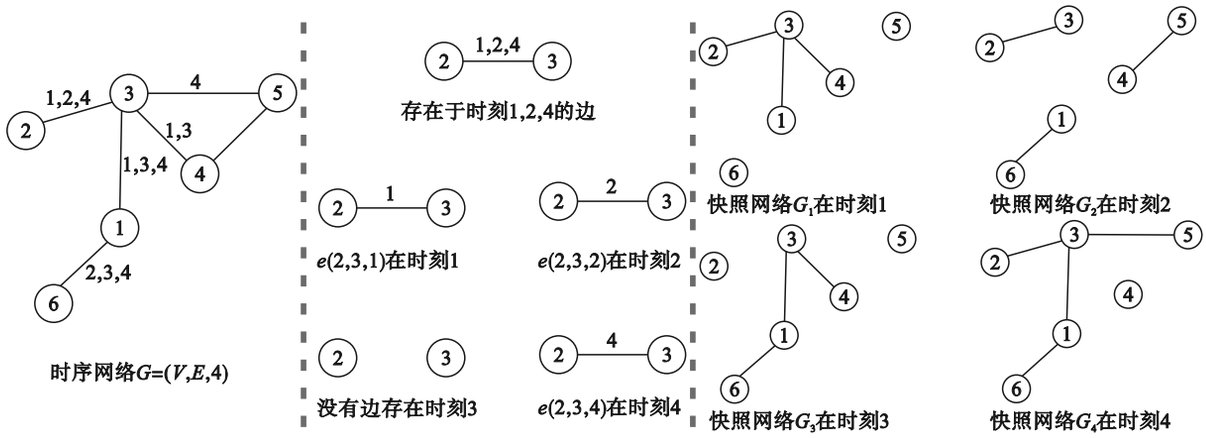

图1 具有6个节点和4个时间步的时序网络

Fig. 1 Temporal network with 6 nodes and 4 time steps

2.2 节点中心性

1)度中心性[14](DC).节点邻居越多,节点越重要.

$$\mathrm{DC}(v_i)=\frac{k(v_i)}{N-1}. \tag{1}$$

其中:$\mathrm{DC}(v_i)$表示节点v_i的度中心性;N为时序网络中节点的数量;$k(v_i)$表示节点的邻居数量.

$$\mathrm{DCM}(v_i)=\frac{1}{n}\sum_{m=1}^{n}\mathrm{DC}_m(v_i). \tag{2}$$

其中:$\mathrm{DCM}(v_i)$表示节点v_i的平均度中心性;

$\mathrm{DC}_m(v_i)$表示节点v_i在G_m中的度中心性.

2) 紧密中心性[15](CC). 节点与其他所有节点之间的平均最短距离越近,其紧密中心性越高.

$$\mathrm{CC}(v_i)=\frac{N-1}{\sum_j^N d(v_i,v_j)}. \tag{3}$$

其中:$\mathrm{CC}(v_i)$表示节点v_i的紧密中心性;$d(v_i,v_j)$表示节点v_i和v_j之间的最短路径长度.

$$\mathrm{CCM}(v_i)=\frac{1}{n}\sum_{m=1}^{n}\mathrm{CC}_m(v_i). \tag{4}$$

其中:$\mathrm{CCM}(v_i)$表示节点v_i的平均紧密中心性;$\mathrm{CC}_m(v_i)$表示节点v_i在G_m中的紧密中心性.

3) 介数中心性[17](BC). 表示不以节点v_i作为起点和终点,但是经过v_i的最短路径数量与图1中所有最短路径数量的比值.

$$\mathrm{BC}(v_i)=\sum_{s\neq i\neq t}\frac{\sigma^{v_i}(v_s,v_t)}{\sigma(v_s,v_t)}. \tag{5}$$

其中:$\mathrm{BC}(v_i)$表示节点v_i的介数中心性;$\sigma(v_s,v_t)$表示节点v_s和v_t之间的最短路径数量;$\sigma^{v_i}(v_s,v_t)$表示$\sigma(v_s,v_t)$中经过节点v_i的最短路径数量.

$$\mathrm{BCM}(v_i)=\frac{1}{n}\sum_{m=1}^{n}\mathrm{BC}_m(v_i). \tag{6}$$

其中:$\mathrm{BCM}(v_i)$表示节点v_i的平均介数中心性;$\mathrm{BC}_m(v_i)$表示节点v_i在G_m中的介数中心性.

2.3　TG模型

Bi等[9]提出TG模型,将节点的特征视为该节点的质量,最早到达路径(fastest arrival path, FAP)的长度作为距离,FAP如图2所示.v_i的时序中心性基础公式如下:

$$\mathrm{TG}(v_i)=\sum_{d(v_i,v_j)\leqslant R,i\neq j}\frac{M_{v_i}M_{v_j}}{d^2(v_i,v_j)}. \tag{7}$$

其中:$\mathrm{TG}(v_i)$表示基于不同特征M_{v_i}下节点v_i的中心性;M_{v_i}表示节点v_i的特征;$d(v_i,v_j)$表示节点v_i和v_j之间时序路径的长度;R是截断半径,表示算法只考虑长度小于R的FAP.

G中节点v_i的$\mathrm{DCM}(v_i)$,$\mathrm{CCM}(v_i)$和$\mathrm{BCM}(v_i)$作为节点v_i的节点特征M_{vi}.用G中节点v_i和v_j之间长度小于截断半径R的FAP作为$d(v_i,v_j)$,便可得到基于TG模型的TG-DC,TG-CC以及TG-BC[9,29],计算公式如下:

$$\mathrm{TG-DC}(v_i)=\sum_{d(v_i,v_j)\leqslant R,i\neq j}\frac{\mathrm{DCM}(v_i)\cdot\mathrm{DCM}(v_j)}{d^2(v_i,v_j)}, \tag{8}$$

$$\mathrm{TG-CC}(v_i)=\sum_{d(v_i,v_j)\leqslant R,i\neq j}\frac{\mathrm{CCM}(v_i)\cdot\mathrm{CCM}(v_j)}{d^2(v_i,v_j)}, \tag{9}$$

$$\mathrm{TG-BC}(v_i)=\sum_{d(v_i,v_j)\leqslant R,i\neq j}\frac{\mathrm{BCM}(v_i)\cdot\mathrm{BCM}(v_j)}{d^2(v_i,v_j)}. \tag{10}$$

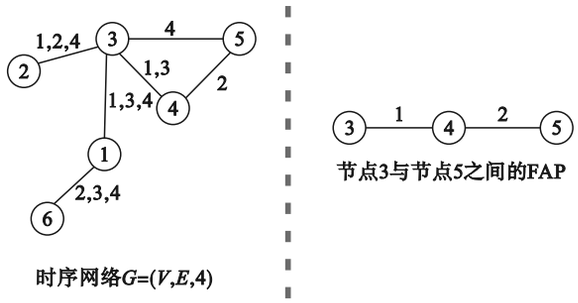

图2　最早到达路径

Fig. 2　FAP

3　TGBISR算法

一个时序网络G由一系列快照网络组成,且每个快照网络都是静态网络,因此节点在整个时序网络中的特征可由该节点在每个快照网络中的特征得出[9].在每一时刻的快照网络中,度中心性可以直观体现节点在网络中的局部特征;紧密中心性能够通过计算节点与其他节点之间的最短路径来体现节点在网络中的位置特征;而介数中心性以网络中经过某一节点的最短路径的数量来体现该节点在网络中的全局特征.TGBISR算法(表1)利用TG模型计算得到某一节点的TG-DC,TG-CC以及TG-BC,并将这3种中心性作为节点的局部、位置以及全局特征构建决策矩阵.

由于时序网络具有动态的网络结构,每种特征所含有的信息量并非恒定不变的,因此每种特征不应获得相同的权重.TGBISR使用信息熵根据每种特征所含有的信息量为其分配合理的权重,并将3种特征的加权和作为判断用户影响力的评价指标.算法具体步骤如下.

步骤1　(算法第1~4行):计算每个节点的TG-DC,TG-CC以及TG-BC,并生成决策矩阵$\boldsymbol{X}=(x_{ij})_{N\times 3}$,决策矩阵的每一列表示一种中心性.

决策矩阵$\boldsymbol{X}=(x_{ij})_{N\times 3}$:

$$\boldsymbol{X}=\begin{bmatrix} x_{11}^{\mathrm{TG-DC}} & x_{12}^{\mathrm{TG-CC}} & x_{13}^{\mathrm{TG-BC}} \\ x_{21}^{\mathrm{TG-DC}} & x_{22}^{\mathrm{TG-CC}} & x_{23}^{\mathrm{TG-BC}} \\ \vdots & \vdots & \vdots \\ x_{N1}^{\mathrm{TG-DC}} & x_{N2}^{\mathrm{TG-CC}} & x_{N3}^{\mathrm{TG-BC}} \end{bmatrix}. \tag{11}$$

表 1 TGBISR 算法
Table 1 TGBISR algorithm

输入: 时序网络 G, 节点数量 N
输出: 节点中心性序列 $\boldsymbol{C}$

1. 生成决策矩阵 $\boldsymbol{X}=(x_{ij})_{N\times 3}$;
2. **for** $i\leftarrow 1,2,\cdots,N$ do
3. $x_{i1}\leftarrow \mathrm{TG-DC}(v_i)$, $x_{i2}\leftarrow \mathrm{TG-CC}(v_i)$;
4. $x_{i3}\leftarrow \mathrm{TG-BC}(v_i)$;
5. $S_1\leftarrow 0,\ S_2\leftarrow 0,\ S_3\leftarrow 0$;
6. **for** $i\leftarrow 1,2,\cdots,N$ **do**
7. **for** $j\leftarrow 1,2,3$ **do**
8. $S_j\leftarrow S_j+x_{ij}^2$;
9. **for** $i\leftarrow 1,2,\cdots,N$ **do**
10. **for** $j\leftarrow 1,2,3$ **do**
11. $x'_{ij}\leftarrow \dfrac{x_{ij}}{\sqrt{S_j}}$;
12. $\boldsymbol{x}'_i\leftarrow(x'_{i1},x'_{i2},x'_{i3})$;
13. $\boldsymbol{X}'\leftarrow[\boldsymbol{x}'_1,\boldsymbol{x}'_2,\cdots,\boldsymbol{x}'_N]^{\mathrm{T}}$;
14. $F_1\leftarrow 0,F_2\leftarrow 0,F_3\leftarrow 0$ **for** $j\leftarrow 1,2,3$ **do**
15. **for** $i\leftarrow 1,2,\cdots,N$ **do**
16. $F_j\leftarrow F_j-\dfrac{1}{\ln N}\sum\limits_{i=1}^{N}x'_{ij}\times\ln x'_{ij}$;
17. $\omega_1\leftarrow\dfrac{F_1}{\sum\limits_{k=1}^{3}F_k},\omega_2\leftarrow\dfrac{F_2}{\sum\limits_{k=1}^{3}F_k},\omega_3\leftarrow\dfrac{F_3}{\sum\limits_{k=1}^{3}F_k}$;
18. $\boldsymbol{\omega}\leftarrow(\omega_1,\omega_2,\omega_3)^{\mathrm{T}}$ **for each** x'_i in $\boldsymbol{X}'$ **do**
19. $C(v_i)\leftarrow\boldsymbol{x}'_i\times\boldsymbol{\omega}$;
20. $\boldsymbol{C}\leftarrow(C(v_1),C(v_2),\cdots,C(v_N))^{\mathrm{T}}$;
21. 将 $\boldsymbol{C}$ 中所有的元素排序

步骤 2 (算法第 5~13 行): 将决策矩阵 $\boldsymbol{X}$ 中的每一个元素进行标准化.

$$x'_{ij}=\frac{x_{ij}}{\sqrt{\sum\limits_{k=1}^{N}x_{kj}^2}},\ i=1,2,\cdots,N,\ j=1,\ 2,\ 3. \tag{12}$$

得到节点 v_i 的信息向量 $\boldsymbol{x}'_i=(x'_{i1},x'_{i2},x'_{i3})$, 以及标准化后的决策矩阵 $\boldsymbol{X}'$.

$$\boldsymbol{X}'=\left[\boldsymbol{x}'_1,\boldsymbol{x}'_2,\cdots,\boldsymbol{x}'_N\right]^{\mathrm{T}}. \tag{13}$$

步骤 3 (算法第 14~17 行): 构建权重向量 $\boldsymbol{\omega}=\left(\omega_1,\omega_2,\omega_3\right)^{\mathrm{T}}$, 其定义为

$$\omega_j=\frac{F_j}{\sum\limits_{k=1}^{3}F_k},\ \ j=1,2,3, \tag{14}$$

$$F_j=-\frac{1}{\ln N}\sum_{i=1}^{N}x'_{ij}\ln x'_{ij}. \tag{15}$$

步骤 4 (算法第 18 行): 计算节点的最终中心性, 将标准化的决策矩阵 $\boldsymbol{X}'$ 与权重向量 $\boldsymbol{\omega}$ 相乘, 得到结果向量 $\boldsymbol{C}$:

$$\boldsymbol{C}=\left(C\left(v_1\right),C\left(v_2\right),\cdots,C\left(v_N\right)\right)^{\mathrm{T}}, \tag{16}$$

$$C\left(v_i\right)=\boldsymbol{x}'_i\boldsymbol{\omega},i=1,2,\cdots,N. \tag{17}$$

其中: $C\left(v_i\right)$ 为节点 v_i 由 TGBISR 计算得到的中心性指标; $\boldsymbol{\omega}$ 为权重向量; $\boldsymbol{x}'_i$ 为节点 v_i 的信息向量.

步骤 5 (算法第 19~20 行): 将 $\boldsymbol{C}$ 中所有元素从大到小进行排序, 排序结果即为节点重要性的顺序.

4 实 验

4.1 实验设计

本文使用 4 个从新浪微博热点大数据研究院收集的数据集对所提出的 TGBISR 的准确性进行评估. 这些数据集包含 4 个不同的时间段, 并且均为以时序网络形式存储的真实社交网络.

算法准确率的计算过程分为 3 步: 首先, 用 TGBISR 计算每个节点的中心性, 并用计算结果构建序列 X, 该序列包含所有节点的中心性; 其次, 使用 SIR 模型对信息的传播过程进行模拟, 模拟结束后得到包含每个节点感染数量的序列 Y, 感染数量作为节点的真实传播影响力; 最后, 计算序列 X 与序列 Y 之间的相关性, 并将相关性作为 TGBISR 的准确率.

本文以肯德尔系数以及回归分析作为评价指标, 作为对比, 本文用 TNCC[11], TDD[22], TD[20], TG-CC 以及 TG-BC 作为对比算法, 这些算法准确率的计算过程与 TGBISR 相同.

4.2 实验数据

如表 2 所示, 对于每个网络, $|V|$ 和 $|E|$ 分别表示时序网络中节点的数量和边的数量; $\langle d\rangle$ 表示平均最短距离, 代表网络中所有节点的平均位置特征; $\langle k\rangle$ 表示网络中每个节点的直接邻居; $k_{\max}$ 表示时序网络中度的最大值, 它表示该网络中的最大局部特征以及用户之间的交互分布; T 表示网络的持续时间; n 为该网络包含的快照网络数量; $\langle E\rangle$ 表示时序网络的每个快照网络所含边的平均数. $\langle E\rangle$ 和 $k_{\max}$ 可以反映出某些节点在整个网络中具有突出的信息传播和接收能力, 即全局特征. 如表 2 所示, 网络 1 中的交互是最频繁且集中的, 另外网络 1 拥有最少的快照网络以及最短的持续时间. 作为网络 1 的对比, 网络 4 拥有最长的持续时间、最多的快照网络、最分散以及最不频繁的交互. 网络 2 和网络 3 各项属性处于中间值, 作为过渡.

表2　真实数据的统计特性

Table 2　Statistical characteristics of real data

网络	$\|V\|$	$\|E\|$	$\langle d\rangle$	$\langle k\rangle$	$k_{\max}$	T	n	$\langle E\rangle$
社交网络1	3 472	3 484	59. 934 0	2. 006 912 4	1 877	120	10	374
社交网络2	3 183	3 351	52. 302 8	2. 105 560 7	1 330	444	37	124. 891 891
社交网络3	3 085	3 317	108. 698 2	2. 150 405 1	473	996	83	56. 481 927
社交网络4	3 620	3 828	46. 473 4	2. 114 917 1	142	3 996	333	19. 453 453

4.3　SIR模型

为了准确地衡量节点的中心性,本文使用被广泛采用的SIR模型来模拟信息传播的过程[9,11,30].SIR模型将所有节点划分为3类:易感状态(S)、感染状态(I)以及恢复状态(R).在初始阶段,除被检测节点v_i为感染状态以外,其余所有节点均为易感状态.在每一时刻,每个感染节点有β的概率感染其相邻的易感节点,并且每个感染节点有$2\times\beta$的概率变为恢复状态.所有恢复状态的节点均不会被再次感染.当整个网络中不存在感染节点时,模拟过程将会停止.模拟结束时,恢复状态的节点代表了节点v_i的中心性,即节点v_i的真实影响力.对于每一个感染概率β,模拟传播过程将会重复100次以避免实验中的随机情况.感染概率β的数值区间为[0.1, 0.2],增长步长为0.01.网络中的每个节点在每个时刻的感染机制服从式(18).

$$\left.\begin{aligned} S(v_i)+I(v_j)\xrightarrow{\beta} I(v_i)+I(v_j),\\ I(v_i)\xrightarrow{2\times\beta} R(v_j). \end{aligned}\right\} \tag{18}$$

其中:$S(v_i)$,$I(v_i)$,$I(v_j)$和$R(v_j)$分别表示易感、感染、感染的邻居和恢复节点.

4.4　肯德尔系数

肯德尔系数常用于表示2个长度相等的序列之间的相关性,用τ表示,取值区间为[−1,1],2个序列用X和Y表示.对于2个包含N个元素的序列X,Y,τ取值为1表示2个序列完全正相关,−1表示完全负相关,0表示两者独立.

$$\tau=\frac{2\times(n_{\mathrm{c}}-n_{\mathrm{d}})}{N(N-1)}. \tag{19}$$

其中:n_{c}表示一致的序列对;n_{d}表示不一致的序列对;N是序列中元素个数.假设包含所有节点中心性的序列$X=\{x_1,x_2,\cdots,x_N\}$和$Y=\{y_1,y_2,\cdots,y_N\}$,其中X的结果由一种算法计算得到,Y由SIR模型计算得到.对于每一对数据(x_i,x_j)和(y_i,y_j),如果$x_i<x_j$的同时$y_i<y_j$,或$x_i>x_j$的同时$y_i>y_j$,则该数据对为一致序列对.如果$x_i<x_j$的同时$y_i>y_j$,或$x_i>x_j$的同时$y_i<y_j$,则该数据对为不一致序列对.

显然,如果一个算法能够更准确识别有影响力的用户,则该算法得到的结果与SIR模型的结果具有更高的肯德尔系数.当肯德尔系数等于1时,表示该算法的结果与SIR模型得到的结果相同,可以没有任何错误地识别网络中的影响力传播者.

4.5　回归分析

回归分析常用于检测两个随机变量之间是否存在依赖关系.假设由一种算法计算得到的序列为X,由SIR模型得到的序列为Y,X和Y之间回归分析得到的回归线的斜率代表了2个序列之间的相关性.在最理想的情况下,回归线的斜率为1,表示该算法得到的结果和SIR模型得到的结果完全匹配,算法准确率为100%.本文通过回归线的斜率与理想回归线的斜率差(slope-difference)的绝对值来描述算法与SIR模型之间的相关程度,斜率差的绝对值越小,表明算法的准确率越高.

5　结果与讨论

5.1　不同感染率下肯德尔系数的对比

图3给出了TGBISR以及5种对比算法在4个社交网络中以不同的感染率β得到的节点中心性与SIR模型得到的中心性的肯德尔系数τ.肯德尔系数的值越大,表示算法识别影响力传播者的准确率越高.显然,TGBISR的准确率明显优于其他算法.TDD算法由于仅关注节点邻居的局部特征而忽略了节点本身的局部特征,从而导致了该算法的准确率较低.社交网络1和2具有较大的$k_{\max}$,这表示该网络中的最大局部特征突出.TD算法是从局部特征的角度来衡量节点的中心性,因此该算法可以在社交网络1和2中具有较好准确率,当局部特征的区别不明显时,该算法将无法作出准确的识别.TNCC算法通过平均邻域变化来衡量节点的中心性,在网络1和2中具有不错的表现,然而网络3和4中的交互频率降低,分布逐渐分散,节点的邻居变化逐渐不明显,导致

TNCC的准确率逐渐下降.快照网络中的边越少,表示交互数量越少,在这种快照网络中的节点将无法估计自己对于信息接收以及信息传播的能力,这意味着全局特征越不明显,因此TG-BC在前两个网络中具有不错的表现,而在后两个网络中表现不佳.

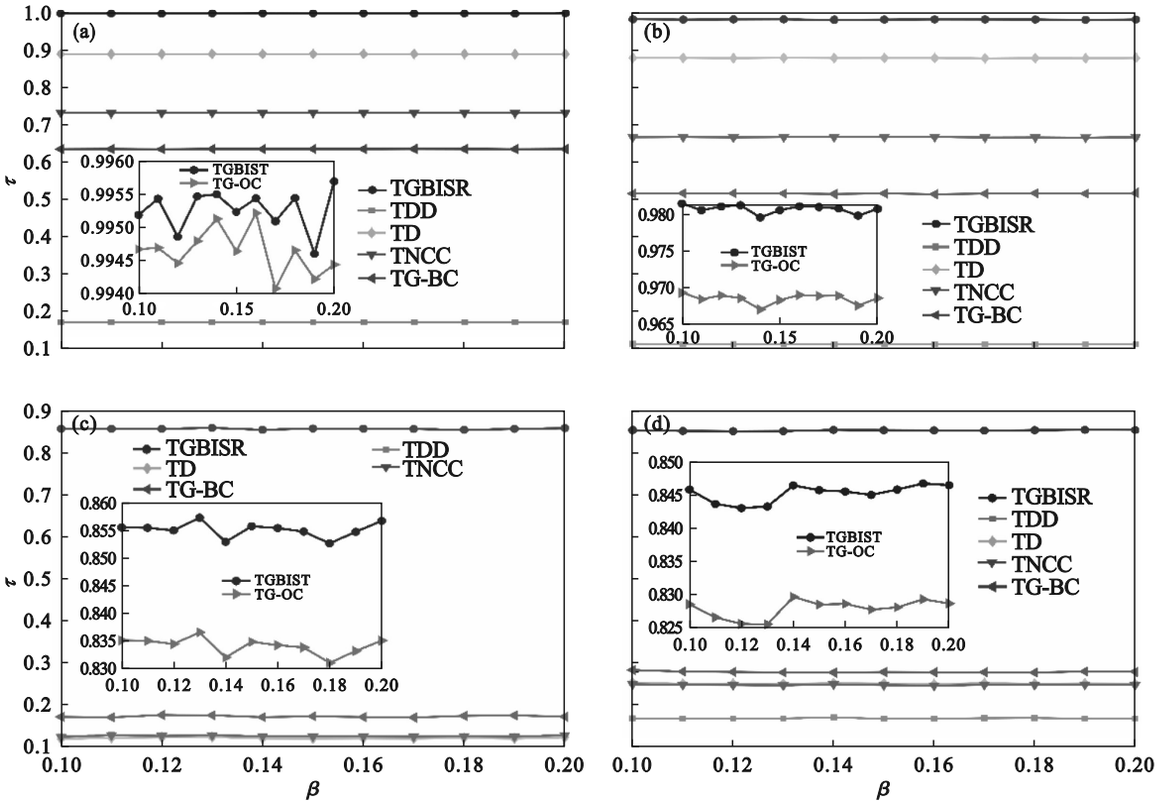

图3 6种算法在4个社交网络中的肯德尔系数

Fig. 3 Kendall coefficient of six algorithms in four social networks

(a)—网络1;(b)—网络2;(c)—网络3;(d)—网络4.

TGBISR是一种综合考虑局部特征、位置特征以及全局特征的方法,能够利用信息熵综合评估节点的中心性,因此,在4个网络中TGBISR能够始终保持最高的准确率.可见,TGBISR通过多种特征衡量节点中心性来识别影响力传播者的方法是更有效的.

5.2 不同算法节点中心性分布的回归分析

节点中心性分布的回归分析.每种算法在$\beta=0.1$时与SIR模型得到的结果的回归分析如表3所示.slope-difference越小,表示该算法的准确率越高.

表3 6种算法与SIR的节点中心性分布的回归分析

Table 3 Regression analysis of node centrality distributions between six algorithms and SIR

网络	算法	斜率差
网络1	TGBISR	0. 468 241
网络1	TDD	0. 526 443
网络1	TD	0. 817 011
网络1	TNCC	−0. 952 093
网络1	TG−CC	0. 637 825
网络1	TG−BC	1. 591 236
网络2	TGBISR	0. 427 741
网络2	TDD	0. 561 564
网络2	TD	0. 568 600
网络2	TNCC	0. 735 848
网络2	TG−CC	0. 543 336
网络2	TG−BC	1. 354 619
网络3	TGBISR	0. 059 830
网络3	TDD	0. 658 774
网络3	TD	0. 716 199
网络3	TNCC	1. 001 581
网络3	TG−CC	0. 269 897
网络3	TG−BC	1. 178 158
网络4	TGBISR	0. 024 723
网络4	TDD	0. 530 435
网络4	TD	0. 583 589
网络4	TNCC	−0. 990 811
网络4	TG−CC	0. 068 356
网络4	TG−BC	0. 576 472

续表3

TGBISR与TD,TG-CC以及TG-BC具有相似的分布.这是因为上述3种方法分别从局部、位置以及全局特征方面对节点的中心性进行评估,而TGBISR能够综合地考虑这3种特性,从而使得TGBISR的slope-difference是最小的.结果表明,在时序网络中识别影响力传播者时,综合地考虑局部、位置以及全局特征能够取得更高的准确率.

6　结　　语

1)本文基于TG模型以及信息熵提出TGBISR方法来识别时序网络中的影响力传播者.将TG模型以及信息熵相结合,综合考虑节点的局部特征、位置特征以及全局特征,TGBISR能够更准确地识别影响力传播者.在TGBISR计算节点的过程中,不同特征根据其所含有的信息量被分配不同的权重.

2)算法的评估实验在4个时序网络上进行,所有的算法都以肯德尔系数和回归分析作为统一评价指标.实验结果表明,在每个网络的每一个感染率β下,TGBISR总是能取得最高的肯德尔系数以及最低的斜率差.这表明本文提出的TGBISR算法相对于其余5种对比算法能够更有效且更精确地识别出时序网络中的影响力传播者.

在未来的研究中,不同快照网络的结构相似性与节点中心性之间的关系将作为提升识别准确率的研究目标.

参考文献:

[1] Kitsak M, Gallos L K, Havlin S, et al. Identification of influential spreaders in complex networks [J]. *Nature Physics*, 2010, 6(11): 888-893.

[2] Li A, Cornelius S P, Liu Y Y, et al. The fundamental advantages of temporal networks [J]. *Science*, 2017, 358(6366): 1042-1046.

[3] Kempe D, Kleinberg J, Tardos É. Maximizing the spread of influence through a social network[C]// Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Washington, DC,2003: 137-146.

[4] Zareie A, Sheikhahmadi A, Jalili M. Identification of influential users in social network using gray wolf optimization algorithm [J]. *Expert Systems with Applications*, 2020, 142: 112971.

[5] Malang K, Wang S, Phaphuangwittayakul A, et al. Identifying influential nodes of global terrorism network: a comparison for skeleton network extraction [J]. *Physica A: Statistical Mechanics and Its Applications*, 2020, 545: 123769.

[6] Gao C, Zhu J Y, Zhang F, et al. A novel representation learning for dynamic graphs based on graph convolutional networks [J]. *IEEE Transactions on Cybernetics*, 2023, 53(6): 3599-3612.

[7] Wang L, Ma L, Wang C, et al. Identifying influential spreaders in social networks through discrete moth-flame optimization [J]. *IEEE Transactions on Evolutionary Computation*, 2021, 25(6): 1091-1102.

[8] Liu J X, Xu C, Yin C, et al. K-core based temporal graph convolutional network for dynamic graphs [J]. *IEEE Transactions on Knowledge and Data Engineering*, 2022, 34(8): 3841-3853.

[9] Bi J L, Jin J Z, Qu C, et al. Temporal gravity model for important node identification in temporal networks [J]. *Chaos, Solitons & Fractals*, 2021, 147: 110934.

[10] Rocha L E C, Masuda N. Random walk centrality for temporal networks [J]. *New Journal of Physics*, 2014, 16(6): 063023.

[11] Wu Z Z, He L Z, Tao L, et al. Temporal neighborhood change centrality for important node identification in temporal networks [C]// 29th International Conference International Conference on Neural Information Processing (ICONIP 2022). Cham: Springer, 2023: 455-467.

[12] Tao L, Kong S Z, He L Z, et al. A sequential-path tree-based centrality for identifying influential spreaders in temporal networks [J]. *Chaos, Solitons & Fractals*, 2022, 165: 112766.

[13] Liu W Z, Lu P L, Zhang T. Identifying influential nodes in complex networks from semi-local and global perspective [J]. *IEEE Transactions on Computational Social Systems*, 2023,11: 1-16.

[14] Freeman L C. Centrality in social networks conceptual clarification[J]. *Social Networks*, 1978, 1(3): 215-239.

[15] Sabidussi G. The centrality index of a graph [J]. *Psychometrika*, 1966, 31(4): 581-603.

[16] Das K, Samanta S, Pal M. Study on centrality measures in social networks: a survey[J]. *Social Network Analysis and Mining*, 2018, 8(1): 13.

[17] Freeman L C. A set of measures of centrality based on betweenness[J]. *Sociometry*, 1977, 40(1): 35-41.

[18] Ullah A, Wang B, Sheng J, et al. Identifying vital nodes from local and global perspectives in complex networks[J]. *Expert Systems with Applications*, 2021, 186: 115778.

[19] Yu E Y, Fu Y, Zhou J L, et al. Predicting critical nodes in temporal networks by dynamic graph convolutional networks[J]. *Applied Sciences*, 2023, 13(12): 7272.

[20] Kim H, Anderson R. Temporal node centrality in complex networks[J]. *Physical Review E*, 2012, 85(2): 026107.

[21] Wang Z Q, Pei X B, Wang Y B, et al. Ranking the key nodes with temporal degree deviation centrality on complex networks [C]// 29th Chinese Control and Decision Conference (CCDC). Chongqing, 2017: 1484-1489.

[22] Elmezain M, Othman E A, Ibrahim H M. Temporal degree-degree and closeness-closeness: a new centrality metrics for social network analysis[J]. *Mathematics*, 2021, 9(22): 2850.

[23] 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.

[24] Lv L S, Zhang K, Zhang T, et al. Eigenvector-based centralities for multilayer temporal networks under the framework of tensor computation [J]. *Expert Systems with Applications*, 2021, 184: 115471.

[25] Kazemi S M, Goel R, Jain K, et al. Representation learning for dynamic graphs: a survey [J]. *Journal of Machine Learning Research*, 2020, 21(1): 70.

(下转第58页)