

## 面向节点演化波动的社会网络事件检测方法<sup>\*</sup>

胡文斌, 王欢, 严丽平, 邱振宇, 聂聪, 杜博



(武汉大学 计算机学院, 湖北 武汉 430072)

通讯作者: 胡文斌, E-mail: hwb@whu.edu.cn; 杜博, E-mail: whuhero@whu.edu.cn

**摘要:** 社会网络特征千差万别, 演化规律错综复杂. 合理地分析网络演化规律, 及时地检测网络事件具有重大意义. 基于链路预测的社会网络事件检测方法利用有限的网络拓扑信息, 能够有效地发现网络演化的异常波动, 准确地检测网络事件. 然而, 现有方法大多受到链路预测的宏观评价指标的限制, 忽略了不同节点演化波动的差异, 用相同的相似性计算指标去描述所有节点的演化波动, 不利于提升事件检测的表现. 为了进一步提升事件检测的精确性和敏感性, 提出一种面向节点演化波动的社会网络事件检测方法 NodeED, 由节点相似性计算指标判定算法 SimJudge 和网络微观演化波动检测算法 MicroFluc 组成. 主要工作如下: (1) 结合粒子群优化算法, 提出 SimJudge 定量地比较不同的相似性计算指标对节点演化波动的描述程度, 确定每个节点在不同时段的最佳相似性计算指标; (2) 为了量化事件对网络演化的影响, 提出了 MicroFluc, 充分考虑节点演化波动的差异, 从节点演化波动的角度对不同时段的网络整体演化波动进行定量评估; (3) 在真实社会网络 VAST 和 ENRON 中进行对比实验, 其结果表明, NodeED 在 VAST 中的事件敏感性提升了 100%, 在 ENRON 中的事件敏感性提升了 50%, 更有利于精确地检测社会网络中发生的事件.

**关键词:** 事件检测; 链路预测; 节点演化波动; 社会网络; 网络演化规律

**中图法分类号:** TP181

中文引用格式: 胡文斌, 王欢, 严丽平, 邱振宇, 聂聪, 杜博. 面向节点演化波动的社会网络事件检测方法. 软件学报, 2017, 28(10): 2693-2703. <http://www.jos.org.cn/1000-9825/5153.htm>

英文引用格式: Hu WB, Wang H, Yan LP, Qiu ZY, Nie C, Du B. Event detection method for social networks based on node evolution fluctuations. Ruan Jian Xue Bao/Journal of Software, 2017, 28(10): 2693-2703 (in Chinese). <http://www.jos.org.cn/1000-9825/5153.htm>

### Event Detection Method for Social Networks Based on Node Evolution Fluctuations

HU Wen-Bin, WANG Huan, YAN Li-Ping, QIU Zhen-Yu, NIE Cong, DU Bo

(Computer School, Wuhan University, Wuhan 430072, China)

**Abstract:** The social network is complicated with different evolution mechanisms. It is of great significance to reasonably analyze social network evolutions and effectively detect social events. The event detection methods based on link prediction make most of the limited network topological information, discover the network evolution fluctuation, and detect events. However, most of existing methods are limited by the assessment measures of link prediction, and neglect the otherness of micro node evolution mechanisms. They use the same similarity index to describe evolution fluctuations of different nodes, which is adverse to the performance of event detection. To improve the accuracy and sensitivity of event detection, this paper proposes an event detection method based on node evolution fluctuation for social networks (NodeED). The method consists of a node similarity index judgement algorithm (SimJudge) and a micro

<sup>\*</sup> 基金项目: 国家重点基础研究发展计划(973)(2012CB719905); 国家自然科学基金(61572369, 61471274); 湖北省自然科学基金(2015CFB423); 武汉市重大科技计划(2015010101010023)

Foundation item: National Program on Key Basic Research Project of China (973) (2012CB719905); National Natural Science Foundation of China (61572369, 61471274); National Natural Science Foundation of Hubei Province (2015CFB423); Wuhan Major Science and Technology Program (2015010101010023)

收稿时间: 2016-01-23; 修改时间: 2016-07-15; 采用时间: 2016-09-29

evolution fluctuation detection algorithm (MicroFluc). The main work is as follow: (1) Based on the particle swarm algorithm, SimJudge is proposed to compare the description performances of different similarity indexes for a node evolution fluctuation. Different nodes can find their optimal similarity indexes at different periods by SimJudge; (2) To quantify the effect of events, MicroFluc is proposed to consider the diversity of node evolution fluctuations and evaluate the entire network evolution fluctuation; (3) In real social networks VAST and ENRON, NodeED results in the event detection sensibility increase by 100% in VAST and 50% in ENRON, which shows NodeED has more advantages to detect events in social networks than other methods.

**Key words:** event detection; link prediction; node evolution fluctuation; social network; network evolution

## 1 引言

如今社会网络高速发展, Facebook、Twitter 和 QQ 等众多在线社交网络的规模不断扩大, 社会网络受到越来越多学者的关注. 其中, 网络演化规律分析和事件检测是当前社会网络的两个重要研究方向<sup>[1,2]</sup>. 网络演化规律分析为进行社会网络事件检测提供了重要的方法基础. 网络演化规律分析是指通过跟踪社会网络不同阶段的特征变化来描述其演化规律, 分析网络的增长、传播等行为, 预测网络未来结构. 事件是由个体或者群体的异常行为所引起, 具有一定的时间持续性和扩散性, 通常会影响网络的正常演化. 事件检测是借助网络演化规律分析的方法, 识别事件造成的网络演化异常, 判定当前社会网络是否有事件发生.

现有研究已经表明了通过链路预测分析网络演化规律、进行社会网络事件检测的方法的有效性<sup>[3,4]</sup>. 链路预测是指利用网络的结构或节点的属性等信息, 计算节点间的相似性, 预测节点间未来产生连边的可能性<sup>[5]</sup>. 链路预测和网络演化规律具有内在的一致性<sup>[4]</sup>. 借助于链路预测的理论框架和评价方法, 通过比较不同的网络演化规律对应的链路预测算法的准确度, 可以间接地评价对不同网络演化规律的表现. 现有的基于链路预测的社会网络事件检测方法的常见步骤是: (1) 通过链路预测的主要评价指标 AUC<sup>[6]</sup>、Precision<sup>[7]</sup>和 Ranking Score<sup>[8]</sup>, 比较基于不同相似性计算指标的链路预测算法的链路预测准确度, 选取准确度最高的相似性计算指标; (2) 基于选取的相似性计算指标设计策略去量化不同时段的网络演化波动, 演化波动显著的时段作为潜在的事件发生时段. 然而, AUC、Precision 和 Ranking Score 都需要假设所有的节点都遵循相同的演化规律, 进而用同一种相似性计算指标计算所有节点间的相似性. 该假设导致现有方法都是宏观上描述整体网络的演化波动, 忽略了个体节点演化波动的差异. 在真实的社会网络中, 不同的节点常常遵循不同的演化规律, 事件的发生对不同的节点演化规律产生的影响也不一样, 部分节点演化规律发生变动的同时, 依旧会有部分节点遵循原来的演化规律, 导致很难用基于单一的相似性计算指标去反映不同节点的演化波动.

基于上述问题, 本文希望从节点的微观角度分析网络的演化波动, 提升事件检测的表现. 本文主要面临如下两个问题: (1) 一个节点可基于多种相似性计算指标来反映其演化波动, 如何确定哪个相似性计算指标最能反映节点的演化波动? (2) 基于不同相似性计算指标的节点演化波动具有不同的度量单位, 如何基于个体节点的演化波动去量化整体网络的演化波动? 为了解决上述两个问题, 在先前工作中提出的社会网络事件检测方法 LinkEvent<sup>[9]</sup>的基础之上, 进一步考虑不同节点演化波动的差异, 本文提出了一种面向节点演化波动的社会网络事件检测方法 NodeED. NodeED 由节点相似性计算指标判定算法 SimJudge 和网络微观演化波动检测算法 MicroFluc 组成. 本文主要工作可总结如下.

(1) 结合粒子群优化算法(PSO), 提出了 SimJudge, 定量地比较不同的相似性计算指标对节点演化波动的描述程度, 确定不同时段最能反映每个节点演化波动的相似性计算指标.

(2) 提出了 MicroFluc, 计算不同节点的加权相似性, 基于每个节点的加权相似性统一量化整体网络的演化波动, 确定网络演化序列.

(3) 在 SimJudge 和 MicroFluc 的基础上, 提出 NodeED, 从节点的角度有效地分析真实网络的演化规律, 提升事件检测的精确性和敏感性.

本文第 2 节介绍相关的重要研究工作. 第 3 节详细介绍 NodeED 方法. 第 4 节在真实的通信网络和邮件网络上进行对比实验以验证 NodeED 方法的表现. 第 5 节是结论及未来研究展望.

## 2 相关工作

目前网络演化规律分析的工作主要集中在网络演化特征的统计研究上,包括密度、度分布、聚集系数、社团等<sup>[10,11]</sup>.图被用来描述特定时间的网络快照,网络的波动性通过图序列来体现<sup>[12]</sup>.常见的网络演化规律分析方法有直接建立网络演化模型的分析方法和基于链路预测的分析方法.直接建立网络演化模型有两个广泛接受的演化机制:三元闭包<sup>[13]</sup>与优先链接<sup>[14]</sup>.常见的网络演化模型正是基于这些机制而展开的,包括 E-R 模型<sup>[15]</sup>、W-S 模型<sup>[16]</sup>、B-A 无标度模型<sup>[14]</sup>、Marsili-Vega-Slanina 模型<sup>[17]</sup>.然而,这些模型都是基于一种或多种演化机制进行构建的,调整模型参数使模型仿真的各时段网络更接近真实网络,通过度分布、聚集系数等统计特征评价模型优劣.由于可供选取比较的网络统计特征太多,造成不同的网络演化模型之间难以进行客观、公正的比较<sup>[4]</sup>.同时,对真实网络而言,很难在预知其演化规律的前提下建立合适的演化模型以进行相关分析.Liu 等人<sup>[3]</sup>率先提出利用链路预测分析网络演化规律,并在推测中国城市航空网络演化机制的例子中,得到了与直接建立网络演化模型分析演化的主导因素一致的结论.Zhang 等人<sup>[18]</sup>引入了似然分析的方法量化不同演化规律的贡献,观察真实网络中流行性和聚类性这两种机制的影响,发现大多数网络的演化都同时受到这两种规律的影响,两种机制的比重会随着时间的推移而发生变化.然而,Zhang 等人<sup>[18]</sup>并没有注意不同相似性计算值度量单位的差异,影响到对最终演化规律的定量分析.

在事件检测方面,Noble 等人<sup>[19]</sup>提出了通过迭代比较发现异常网络结构的方法,利用子结构条件熵来量化图形结构的异常程度.但是该方法侧重于理论研究,实际计算复杂.Papadimitriou<sup>[20]</sup>检测网络的异常,判断面向较大规模的服务器、爬虫程序等是否有异常发生.McCulloh 等人<sup>[21]</sup>尝试在社会网络中屏蔽正常网络波动的干扰,对网络实质变化进行分离,提出了社会网络变化检测方法.但是,这种检测方法要求网络参数满足正态分布,使其应用范围受到了极大的限制.Wan 等人<sup>[22]</sup>通过分别检测网络特征参数的变化和社团通联结构的变化来发现邮件网络的异常事件.Priebe 等人<sup>[23]</sup>采用检测统计量方法对邮件网络进行检测,发现网络中的“震荡”区域.Wan<sup>[22]</sup>和 Priebe<sup>[23]</sup>的工作中要计算的统计量过多,导致计算量巨大.Wu 等人<sup>[14]</sup>提出一种基于事件的社会网络演化分析框架,发现网络演化过程中的事件,分析事件对网络演化产生的影响.Hu 等人<sup>[9]</sup>通过链路预测推断网络的内部演化机制,然后考虑每个点与其邻居在何种程度上符合这种演化机制,提出网络相似性计算算法,进而构建事件检测算法.Wu<sup>[14]</sup>和 Hu<sup>[9]</sup>的工作侧重于分析整体网络的宏观演化规律,忽略了节点微观演化的差异性.

综上所述,现有网络演化模型常常假设网络中所有节点统一遵循一种或者多种演化机制,只能宏观地分析整体网络演化,忽略了节点微观演化机制的特异性,不利于深入分析网络演化.同时,当社会网络中事件发生时,不同的节点受到事件的影响程度不同.基于宏观网络演化规律的事件检测方法无法区分事件对节点演化造成的扰动差异,从微观节点的角度分析网络演化有助于更准确地量化事件引起的网络异常波动.本文在先前工作的基础上<sup>[9]</sup>,充分考虑了不同节点演化波动的差异,提出了一种面向节点演化波动的社会网络事件检测方法 NodeED.

## 3 NodeED 方法

在真实的社会网络中,不同的节点遵循的演化规律不一定相同,事件对不同节点造成的影响也不一样.不同于现有的事件检测方法忽略了节点演化存在差异这一点,NodeED 方法从节点微观演化波动的角度,分析事件对不同节点产生的影响,进行事件检测.第 3.1 节将对本文涉及到的先前工作 LinkEvent<sup>[6]</sup>中的相关概念进行说明.第 3.2 节介绍 NodeED 方法的框架和实现步骤.第 3.3 节和第 3.4 节分别详细介绍 NodeED 方法中的 SimJudge 算法和 MicroFluc 算法.

### 3.1 相关概念定义

在社会网络的研究过程中,经常借助图来描述某一个特定时间的网络快照.图中的节点代表网络中的个体,图中的边代表网络中个体之间的联系(如通话、邮件、论文合作).网络演化的动态性常常通过图序列来刻画,

对于由  $n$  个网络快照组成的图序列表示如下:

$$G = \{g^1, g^2, \dots, g^t, \dots, g^n\} \tag{1}$$

(1) 节点相似性计算指标:链路预测中的节点相似性计算指标用来衡量图中两个不同节点间的相似程度,两个节点的相似性大小取决于它们的共同邻居、节点的度等拓扑信息. $G$  中节点  $i$  在  $g^t$  和  $g^{t+1}$  中可以被看成是两个不同的节点  $v_i^t$  和  $v_i^{t+1}$ ,  $v_i^t$  和  $v_i^{t+1}$  之间的相似性也可以通过它们的拓扑信息来描述.为了避免新增和消失的节点对网络演化影响的丢失,引入虚拟点  $V_{virtual}$  来完整地描述网络演化波动( $V_{virtual}$  与网络中所有的节点都存在一条虚拟边).引入虚拟节点后,8 种常见的节点相似性计算可见表 1.

**Table 1** The adjusted similarity indexes by introducing  $v_{virtual}$

**表 1** 引入基准点  $v_{virtual}$  后的节点的相似性计算指标

名称	节点的相似性	名称	节点的相似性
共同邻居指标 (CNS) <sup>[13]</sup>	$s(v_i^t, v_i^{t+1}) =  \Gamma(v_i^t) \cap \Gamma(v_i^{t+1})  + 1$	Jaccard 指标 (JAS) <sup>[24]</sup>	$s(v_i^t, v_i^{t+1}) = \frac{ \Gamma(v_i^t) \cap \Gamma(v_i^{t+1})  + 1}{ \Gamma(v_i^t) \cup \Gamma(v_i^{t+1})  + 1}$
优先链接指标 (PAS) <sup>[25]</sup>	$s(v_i^t, v_i^{t+1}) = (k(v_i^t) + 1) \times (k(v_i^{t+1}) + 1)$	Sorenson 指标 (SOS) <sup>[26]</sup>	$s(v_i^t, v_i^{t+1}) = \frac{2( \Gamma(v_i^t) \cap \Gamma(v_i^{t+1})  + 1)}{k(v_i^t) + k(v_i^{t+1}) + 2}$
Adamic-Adar 指标 (AAS) <sup>[27]</sup>	$s(v_i^t, v_i^{t+1}) = \sum_{z \in \Gamma(v_i^t) \cap \Gamma(v_i^{t+1})} \frac{1}{\frac{k(v_i^t) + k(v_i^{t+1})}{2}}$	大度节点有利指标 (HPIS) <sup>[28]</sup>	$s(v_i^t, v_i^{t+1}) = \frac{ \Gamma(v_i^t) \cap \Gamma(v_i^{t+1})  + 1}{\min\{k(v_i^t) + 1, k(v_i^{t+1}) + 1\}}$
Salton 指标 (SAS) <sup>[29]</sup>	$s(v_i^t, v_i^{t+1}) = \frac{ \Gamma(v_i^t) \cap \Gamma(v_i^{t+1})  + 1}{\sqrt{(k(v_i^t) + 1) \times (k(v_i^{t+1}) + 1)}}$	LNH-I 指标 (LNHS) <sup>[30]</sup>	$s(v_i^t, v_i^{t+1}) = \frac{ \Gamma(v_i^t) \cap \Gamma(v_i^{t+1})  + 1}{(k(v_i^t) + 1) \times (k(v_i^{t+1}) + 1)}$

(2) 图的基本相似性:图  $g^t$  和  $g^{t+1}$  的基本相似性是图中各个节点相似性叠加后的宏观表现,用  $S(g^t, g^{t+1})$  表示,定义为式(2).

$$S(g^t, g^{t+1}) = \sum_{i \in U_{t,t+1}} s(v_i^t, v_i^{t+1}) \times \frac{1}{|U_{t,t+1}|}, \text{其中, } U_{t,t+1} = g^t \cup g^{t+1} \tag{2}$$

$S(g^t, g^{t+1})$  反映两个图之间的近似程度,其值越大表示网络在  $[t, t+1]$  时段内变化越小,网络的波动程度越小. $[t, t+1]$  时段内网络的波动性  $\hat{D}(g^{t+1} \| g^t)$  用式(3)表示如下:

$$\hat{D}(g^{t+1} \| g^t) = \frac{1}{S(g^t, g^{t+1})} \tag{3}$$

网络演化序列 *GraphS* 定义为各个时段波动性的集合,如式(4)所示.

$$GraphS = \{\hat{D}(g^2 \| g^1), \hat{D}(g^3 \| g^2), \dots, \hat{D}(g^n \| g^{n-1})\} \tag{4}$$

(3) 事件检测敏感表现:假定  $k_1, k_2$  为网络  $G$  中紧邻的两个事件点( $k_1 + 1 < k_2$ ),  $G$  在  $[k_1 + 1, k_2]$  时段正常演化,称为网络平稳时段.  $G$  在  $[k_2, k_2 + 1]$  时段受到了事件影响,称为事件波动时段.事件敏感性表现指标 *Per* 反映了网络在事件波动时段与网络平稳时段的平均波动性的比值. *Per* 用来衡量事件检测方法的好坏,值越大则表明事件检测方法表现越好,越容易检测出事件. *Per* 的定义如式(5)所示.

$$Per = \frac{\sum_{h=k_1+1}^{k_2-1} \hat{D}(g^{h+1} \| g^h)}{\hat{D}(g^{k_2+1} \| g^{k_2}) - \frac{\sum_{h=k_1+1}^{k_2-1} \hat{D}(g^{h+1} \| g^h)}{k_2 - k_1 - 2}} \tag{5}$$

**3.2 整体框架**

NodeED 方法框架如图 1 所示. NodeED 方法输入信息是目标社会网络的图序列,仅仅依赖社会网络的边和节点的信息,这也增强了 NodeED 方法的适用范围. NodeED 方法关注的目标事件是对网络演化产生了重要影响

的事件,这些事件网络对网络演化的描述具有重大意义.如果一个发生的事件并没有对网络演化产生任何扰动,则 NodeED 方法无法对该事件进行有效检测.NodeED 方法主要是由 SimJudge 算法和 MicroFluc 算法组成.SimJudge 算法确定不同时段能反映节点演化波动的最佳相似性计算指标.MicroFluc 算法在 SimJudge 算法的基础上量化不同时段网络演化波动,得到网络演化序列.NodeED 方法通过与事件检测阈值进行比较,判定哪些时段发生了事件.假设  $EventS$  表示事件发生时段集合,NodeED 方法的实现步骤见表 2.

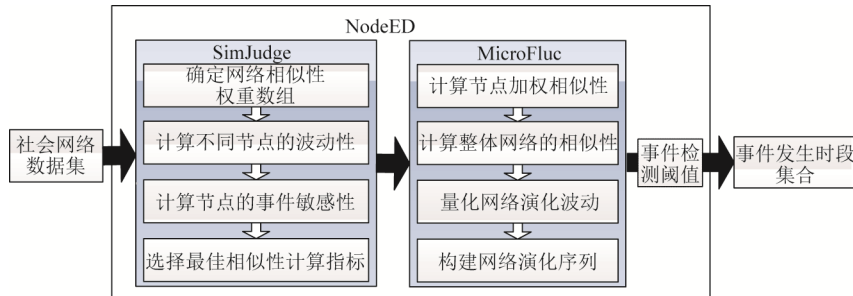


Fig.1 The framework of NodeED

图 1 NodeED 方法框架

Table 2 The implementation steps of NodeED

表 2 NodeED 方法实现步骤

输入:社会网络的数据集.
1. 执行 SimJudge 算法确定节点在不同时段的最佳相似性计算指标.
2. 执行 MicroFluc 算法构建网络演化序列 $Graphs$ .
3. For $t=1$ to $n-1$ :
$H = \max \{ \hat{D}(g^{t+1} \  g^t) \}$
$L = \min \{ \hat{D}(g^{t+1} \  g^t) \}$
End
4. 人工选取事件检测阈值 $T_i \in [L, H]$
5. For $t=1$ to $n-1$ :
If $\hat{D}(g^{t+1} \  g^t) > T_i$ :
$[t, t+1] \in EventS$
End
End
7. return $EventS$

### 3.3 SimJudge 算法

在 PSO 的基础上,本节提出 SimJudge 算法,定量分析不同相似性计算指标对节点演化波动的描述程度,确定每个节点的最佳相似性计算指标.不同于现有方法通过统计网络特征变化来分析不同时段网络的相似性,SimJudge 算法首次引入链路预测中的相似性计算指标从节点的角度分析整体网络的相似性变化.

节点的演化规律实质上可以看成是节点的连边规律,表现为节点及其周边拓扑结构的变化.节点的演化波动指的是节点演化规律的变动.借助于表 2 中的节点相似性计算指标描述节点及其周边拓扑结构变化,可以间接反映节点的演化波动.然而,不同的相似性计算指标得到的相似性计算值具有不同的度量单.以表 1 中列出的共同邻居指标(CNS)和优先链接指标(PAS)为例,两者分别是从小世界<sup>[13]</sup>和富者愈富<sup>[14]</sup>两个基本规律出发得到的.我们可以证明始终有  $s(v_i^t, v_i^{t+1})_{PAS} > s(v_i^t, v_i^{t+1})_{CNS}$ ,证明如下.

$$\begin{aligned}
 s(v_i^t, v_i^{t+1})_{PAS} &= (k(v_i^t) + 1) \times (k(v_i^{t+1}) + 1) \\
 &= k(v_i^t) \times k(v_i^{t+1}) + k(v_i^t) + k(v_i^{t+1}) + 1 \\
 &\geq k(v_i^t) \times k(v_i^{t+1}) + \max(k(v_i^t), k(v_i^{t+1})) + 1
 \end{aligned}$$

$$\begin{aligned}
&\geq k(v_i^t) \times k(v_i^{t+1}) + |\Gamma(v_i^t) \cap \Gamma(v_i^{t+1})| + 1 \\
&> |\Gamma(v_i^t) \cap \Gamma(v_i^{t+1})| + 1 \\
&= s(v_i^t, v_i^{t+1})_{CNS}.
\end{aligned}$$

为了定量分析不同相似性计算指标的计算值,我们构建相似性权重数组  $params$  对不同的相似性计算值进行加权调整.本文考虑表 1 中的 8 种相似性计算指标组成相似性计算指标集合  $EList=[CNS,JAS,PAS,SOS,AAS,HPIS,SAS,LNHS]$ ,对应的相似性权重数组为  $params=(w_{CNS},w_{JAS},w_{PAS},w_{SOS},w_{AAS},w_{HPIS},w_{SAS},w_{LNHS})$ .同时,假设网络演化的过程中,同一时段网络的相似性权重数组固定,在不同的时段可以对应不同的相似性权重数组.

在相似性权重数组的基础上,我们提出节点的波动性,反映节点在网络整体演化过程中的个体演化波动.对于节点  $i$  而言,  $v_i^t$  和  $v_i^{t+1}$  分别表示在  $g^t$  和  $g^{t+1}$  的节点  $i$ .  $params_{(t,t+1)}[index]$  表示  $[t,t+1]$  时段相似性计算指标  $index$  对应的相似性权重,  $s_{index}(v_i^t, v_i^{t+1})$  是从表 1 中得到的  $index$  对应的相似性计算指标公式.  $[t,t+1]$  时段相似性计算指标  $index$  所反映的节点  $i$  的波动性  $v\hat{D}_{index}(v_i^{t+1} \| v_i^t)$  如式(6)所定义.

$$v\hat{D}(v_i^{t+1} \| v_i^t) = \frac{1}{params_{(t,t+1)}[index] \times s_{index}(v_i^t, v_i^{t+1})} \quad (6)$$

为了评价不同相似性计算指标对节点波动性的描述程度,我们针对事件检测方法的事件敏感性表现  $Per$  进行扩展,提出了节点的敏感性表现  $vPer$ .在相似性计算指标  $index$  下,节点  $i$  在  $[t,t+1]$  时段的敏感性表现  $vPer_{[t,t+1]}^{index}(i)$  如式(7)所定义.

$$vPer_{[t,t+1]}^{index}(i) = \frac{v\hat{D}(v_i^{t+1} \| v_i^t) - \frac{\sum_{h=1}^{t-1} v\hat{D}(v_i^{h+1} \| v_i^h)}{t-1}}{\frac{\sum_{h=1}^t v\hat{D}(v_i^{h+1} \| v_i^h)}{t}} \quad (7)$$

$vPer_{[t,t+1]}^{index}(i)$  越大,则表示  $index$  越能更好地描述节点  $i$  在当前时段  $[t,t+1]$  的网络演化异常.使  $vPer_{[t,t+1]}^{index}(i)$  值最大的  $index$  称为节点  $i$  在  $[t,t+1]$  时段的最佳相似性计算指标,表示为  $Bindex_{[t,t+1]}(i)$ .  $vPer_{[t,t+1]}^{index}(i)$  对应的  $index$  用来计算节点  $i$  在  $[t,t+1]$  时段的节点波动性  $v\hat{D}(v_i^t \| v_i^{t-1})$ ,在对  $[t,t+1]$  时段之前各时段节点波动性进行计算时分别使用它们各自的最佳相似性计算指标.

最后,如何确定网络在不同时段的相似性权重数组,进而通过式(6)和式(7)判定不同节点在不同时段的最佳相似性计算指标,将是本节提出的  $SimJudge$  算法要解决的问题.在式(2)的基础上进一步提出了扩展相似性  $S'(g^t, g^{t+1})$  用来量化  $g^t$  和  $g^{t+1}$  相似性,定义如式(8)所示.

$$S'(g^t, g^{t+1}) = \sum_{i \in U_{t,t+1}} \sum_{index \in EList} params_{(t,t+1)}[index] \times s_{index}(v_i^t, v_i^{t+1}) \times \frac{1}{|U_{t,t+1}|} \quad (8)$$

$S'(g^t, g^{t+1})$  对所有相似性计算指标下的相似性计算值进行加权求和,综合地考虑不同的相似性计算指标对同一个节点的相似性评价.对于  $[t,t+1]$  时段而言,在式(8)所示的  $S'(g^t, g^{t+1})$  的基础上,进一步通过式(3)~式(5)计算求得  $[t,t+1]$  时段对应的事件拟合值  $Per$ ,并将其作为  $SimJudge$  算法在  $[t,t+1]$  时段的适应度函数.  $SimJudge$  算法通过调整  $[t,t+1]$  时段的  $params_{(t,t+1)}$ ,使  $[t,t+1]$  时段对应的  $Per$  值增大,更好地协调  $[t,t+1]$  时段不同的相似性计算指标之间的度量单位差异,提升对  $[t,t+1]$  时段网络波动的敏感性.

假设  $SimJudge$  算法中共有  $m$  个粒子,每个粒子初始状态都携带一个随机生成的  $params$ ,共经历了  $g_{max}$  次迭代.  $Per(params_k^g)$  表示粒子  $k$  在当前迭代  $g$  结束时的  $params_k^g$  对应的事件拟合值.  $params_{kb}$  表示粒子  $k$  当前发现的最优  $params$ .整个粒子群的最优  $params$  即  $params_G$ .  $pv_k^g$  表示粒子  $k$  在第  $g$  次迭代时  $params_k^{g-1}$  对应的调整速度.  $c_1$  和  $c_2$  是学习因子.  $r_2$  和  $r_1$  是 0 和 1 之间的随机数.对每个时段,在迭代优化的过程中,每个粒子的  $params$  不断调整以趋近  $params_G$ .  $SimJudge$  算法依次确定每个时段对应的  $params_G$ ,进而确定每个时段不同节点对应的最佳相似性计算指标.二维数组  $IndexList$  用来保存节点和网络演化时段、节点最佳相

似性计算指标之间的对应关系.  $IndexList[i][t]$  表示节点  $i$  在时段  $[t, t+1]$  的最佳相似性计算指标. 算法 SimJudge 的具体表示如下.

**算法. SimJudge.**

输入: 社会网络数据集;

输出:  $IndexList$ , 不同时间段的  $params_G$ .

1. 构建网络演化规律指标集合  $EList$ .

2. For  $t=1$  to  $n-1$ :

    For  $g=1$  to  $g_{max}$ :

        For  $k=1$  to  $m$ :

            If  $Per(params_k^g) > Per(params_{kb})$  then  $params_{kb} = params_k^g$ ;

            If  $Per(params_{kb}) > Per(params_G)$  then  $params_G = params_{kb}$ ;

$sv_k^g = sv_k^{g-1} + c_1 r_1 (params_{kb} - params_k^{g-1}) + c_2 r_2 (params_G - params_k^{g-1})$ ;

$params_k^g = params_k^{g-1} + p v_k^g$ ;

        End

$IndexList[i][t] = Bindex_i, Bindex_i$  使  $vPer_{[t,t+1]}^{Bindex_i}(i) = \max\{vPer_{[t,t+1]}^{index}(i), index \in EList\}$ ;

    Return  $params_{(t,t+1)}$ ;

    End

End

Return  $IndexList$ .

### 3.4 MicroFluc 算法

基于网络快照序列  $G = \{g_1, g_2, \dots, g^t, \dots, g^n\}$ , 事件检测的目的就是要判定每个网络演化时段是不是有重大事件发生, 即该时段是属于网络平稳时段还是事件波动时段. 先前的工作中事件检测方法<sup>[9]</sup>采取的步骤是: (1) 把  $g_1$  中的边划分为训练集和测试集, 基于各种指标进行链路预测, 选取在  $g_1$  上评价指标 AUC<sup>[6]</sup> 值最大的相似性计算指标作为网络的最佳相似性计算指标; (2) 从表 1 中选取网络最佳相似性计算指标对应的公式, 计算在相邻网络快照中每个节点的相似性, 通过公式(2)~公式(4)求得网络演化序列; (3) 设置事件检测阈值  $T_i$ , 当网络演化序列中网络波动值大于  $T_i$  时, 则该时段为事件波动时段, 反之为网络平稳时段.

然而, 事件发生时, 不同节点可能遵循不同的演化规律, 能够描述每个节点受事件影响的最佳相似性计算指标有所不同. 不同于现有的 LinkEvent 方法, NodeED 方法中 SimJudge 算法充分考虑了事件的发生对不同节点影响的差异, 确定每个节点对应的最佳相似性计算指标, 充分反映了每个节点的演化波动. 在 SimJudge 算法的基础上, 本节进一步提出了 MicroEvent 算法, 基于节点的演化波动去量化整体网络的演化波动, 构建网络演化序列.

本节在表 1 中相似性计算指标的基础上提出了节点的加权相似性, 定义如式(9)所示.

$$s^n(v_i^{t+1} \| v_i^t) = params_{(t,t+1)}[Bindex_{(t,t+1)}(i)] \times s_{[Bindex_{(t,t+1)}(i)]}(v_i^t, v_i^{t+1}) \quad (9)$$

在计算加权相似性  $s^n(v_i^{t+1} \| v_i^t)$  时, 每个节点  $i$  根据  $IndexList$  选取  $[t, t+1]$  时段对应的  $Bindex_{(t,t+1)}(i)$  ( $Bindex_{(t,t+1)}(i) = IndexList[i][t]$ ). 通过  $params_{(t,t+1)}[Bindex_{(t,t+1)}(i)]$  对相似性计算值  $s_{[Bindex_{(t,t+1)}(i)]}(v_i^t, v_i^{t+1})$  进行加权调整, 得到加权相似性值  $s^n(v_i^{t+1} \| v_i^t)$ . 通过加权相似性, 每个节点在每个时段对应的演化波动基于不同的相似性计算指标来描述, 反映事件造成的节点在不同时间段的演化波动. 同一时段, 不同节点的加权相似性可以定量比较. 在加权相似性的基础上, 利用式(2)对不同时段网络中节点的加权相似性进行累加, 通过式(3)和式(4)得到的最终的网络演化序列. 假设  $g^t$  中共有  $N^t$  个粒子, MicroFluc 算法的具体表示如下.

**算法. MicroFluc.**

输入: 社会网络数据集;

输出: *GraphS*.

1. For  $t=1$  to  $n-1$ :

For  $i=1$  to  $N^t$ :

$Bindex_{(t,t+1)}(i)=IndexList[i][t]$ ;

通过式(9)基于  $Bindex_{(t,t+1)}(i)$  计算  $s^r(v_i^{t+1} \| v_i^t)$ ;

End

通过式(2)累加所有节点的加权相似性得到  $S(g^t, g^{t+1})$ ;

通过式(3)计算  $\hat{D}(g^{t+1} \| g^t)$ ;

$GraphS[t]=\hat{D}(g^{t+1} \| g^t)$ ;

End

2. Return *GraphS*

## 4 实验分析

本节通过在真实社会网络数据集上的实验来验证 NodeED 方法的表现.第 4.1 节介绍实验中使用的真实社会网络数据集.第 4.2 节利用 SimJudge 算法确定真实网络中不同节点的最佳相似性计算指标,并通过与先前工作中宏观最佳相似性计算指标的分析结果进行对比,证明 SimJudge 算法从节点的角度分析网络演化波动的必要性.第 4.3 节基于 MicroFluc 算法得到网络演化序列,通过对比实验验证 NodeED 方法进行事件检测的精确性和敏感性.

### 4.1 数据集描述

我们选取真实社会网络数据集 ENRON 和 VAST 进行实验.VAST 数据集源自 IEEE VAST 2008,它涉及 400 人组成的社会网络在 10 天内的通话数据,并且已知 10 天里发生了一次高层变动的事件.ENRON 数据集源自 ENRON 公司的内部邮件联系网络,它共涉及 150 人在 111 周里的通信数据,本文选择其中有代表性的 20 周,期间发生公司发布财务报表,提交交易等多次事件.VAST 数据集的特点是只有独立的事件发生,并且事件发生的时间和原因确定,有利于分析独立事件发生前后节点和网络的演化波动.ENRON 数据集的特点是多重事件连续发生,有助于分析多重连续事件之间的相互作用对网络演化的影响.

### 4.2 节点演化波动分析的必要性验证

实验考虑表 1 中 8 种相似性计算指标,  $Elist=[CNS, JAS, PAS, SOS, AAS, HPIS, SAS, LNHS]$ .参照 PSO 中常见的参数设置<sup>[31]</sup>,设置 NodeEvo 算法中粒子数为 100,迭代次数设置为 1 000,  $w=0.8, c_1=c_2=2, r_1=0.6, r_2=0.3$ .先前工作已经表明,VAST 和 ENRON 的宏观上最佳的相似性计算指标分别是 PAS 和 JAS<sup>[9]</sup>.利用 NodeEvo 算法确定 VAST 和 ENRON 网络演化过程中不同时段不同节点的最佳相似性计算指标,进而统计整个网络演化过程中每个最佳相似性计算指标对应的节点比例,见表 3.通过与宏观上最佳的相似性计算指标进行对比,从节点的角度分析网络演化波动的必要性主要体现在如下两点.

(1) 在真实社会网络 VAST 和 ENRON 的演化过程中,网络中的节点会分别对应多种最佳相似性计算指标,用单一的宏观最佳相似性计算指标不能充分反映所有节点的演化波动.在 VAST 中,节点的最佳相似性计算指标的比例比较分散,节点比例较高的最佳相似性计算指标为 PAS, AAS, JAS 和 LNHS,它们的比例都高于 10%,其中, LNHS 最高为 37.25%.在 ENRON 中,最佳相似性计算指标集中为 PAS 和 LNHS,它们对应的比例分别为 48.67%和 40.67%,远远高于其他最佳相似性计算指标.

(2) 宏观上网络的最佳相似性计算指标并不总能充分反映节点对应的最佳相似性指标的比例.对于 VAST 网络,最佳相似性计算指标为 PAS 的节点比例为 22%,与宏观上满足 PAS 相一致.但是,比例最大的 LNHS 却无法在宏观分析中体现出来, LNHS 的节点比例高达 37.25%.对于 ENRON 网络,宏观上最佳相似性计算指标 JAS 的比例仅为 1.33%,而 PAS 和 LNHS 的比例之和高达 89.34%.



**Table 3** The ratios of nodes with different optimal similarity indexes in VAST and ENRON

**表 3** VAST 和 ENRON 中不同最佳相似性计算值对应的节点比例(%)

	PAS	AAS	SAS	JAS	SOS	HPIS	LNHS	CNS
VAST	22.00	19.00	0.00	14.00	0.50	5.25	37.25	2.00
ENRON	48.67	4.67	0.00	1.33	2.00	2.00	40.67	0.67

**4.3 事件检测的精确性和敏感性验证**

本节对采用 NodeED 方法在基于 MicroFluc 算法得到的网络演化序列的基础上进行事件检测加以介绍.直接基于宏观最佳的相似性计算指标的事件检测方法<sup>[9]</sup>和 LinkEvent 中表现最好的基于策略 1 的事件检测方法<sup>[9]</sup>被选取作为对比实验.基于 3 种事件检测方法得到的 VAST 和 ENRON 的网络演化序列分别如图 2 和图 3 所示.

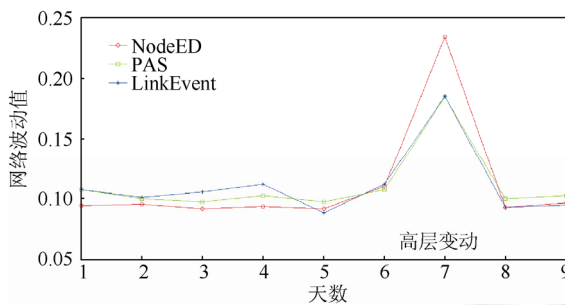


Fig.2 The network evolution sequence for VAST

图 2 VAST 中 3 种检测方法对应的网络演化序列

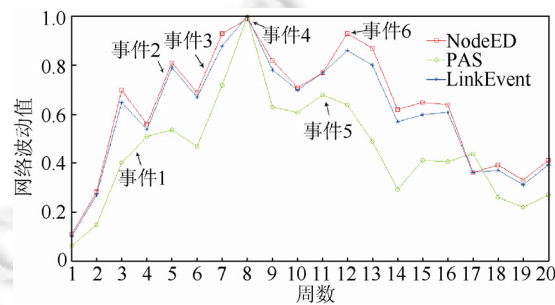


Fig.3 The network evolution sequence for ENRON

图 3 ENRON 中 3 种检测方法对应的网络演化序列

3 种事件检测方法在 VAST 和 ENRON 中所有演化时段的平均事件敏感性值 *Per* 见表 4.表 5 显示了 ENRON 的网络演化序列反映的事件与真实事件的对应关系.我们可以得到如下结论.

(1) 在网络 VAST 和 ENRON 中,NodeED 方法能够精确地反映事件引起的网络演化波动.3 种事件检测方法得到的网络波动序列都与真实事件具有较好的匹配关系.在 VAST 网络中,第 7 周公司高层发生变动时,NodeED 方法能够更显著地反映高层变动事件引起的网络演化波动.在 ENRON 中,不同的网络事件之间相互干涉影响.在有连续事件发生的时段,3 种事件检测方法都具有相对较高的网络波动值,NodeED 方法反映出的整体网络波动值最高.事件 3 的美国证券交易委员会的正式调查对 ENRON 网络的影响最大,是整个安然事件的爆发点,使 3 种方法的网络波动值同时达到最大值.

(2) NodeED 方法显著地提升了事件检测敏感性,更有利于检测社会网络中发生的事件.该方法在 VAST 和 ENRON 中事件敏感性值分别为 31.52 和 6.5,在 3 种方法中均为最高.基于宏观最佳相似性计算指标的事件检测方法在 VAST 和 ENRON 中事件敏感性值分别为 4.46 和 3.76,远低于 NodeED 方法的表现.相对于次优 LinkEvent 方法,NodeED 方法在 VAST 中的事件敏感性表现提升了 100%,在 ENRON 中的事件敏感性表现提升了 50%.

**Table 4** The *Per* values by three detection methods

**表 4** 3 种方法事件敏感性表现 *Per* 值

检测方法	VAST			ENRON		
	NodeED	PAS	LinkEvent	NodeED	LAS	LinkEven
<i>Per</i> 值	31.52	4.46	15.73	6.50	3.76	4.21

**Table 5** The network evolution sequence between detected events and real events

**表 5** ENRON 网络演化序列反映的事件和真实事件对应关系

事件编号	真实事件描述
事件 1	安然发布 2001 年第二季财务报表
事件 2	美国证券交易委员会要求提交交易内容
事件 3	美国证券交易委员会开始对安然进行正式调查,次日安然抵押了部分资产
事件 4	安然承认做了假账,次日迪诺基公司宣布准备收购安然
事件 5	标准普尔将安然调低至“垃圾债券”
事件 6	安然申请破产保护

## 5 结论及展望

为了提高社会网络事件检测的表现,本文提出了一种面向节点演化波动的社会网络事件检测方法 NodeED. NodeED 方法首先通过 SimJudge 算法确定节点在不同时段的最佳相似性指标,然后利用 MicroFluc 算法得到网络演化序列,进而与事件检测阈值比较判定事件.为验证 NodeED 方法的性能,本文选取真实社会网络数据集 VAST 和 ENRON 进行对比实验,并得出以下结论:(1) 在真实社会网络中,SimJudge 算法能够有效地确定不同节点在不同时段的最佳相似性计算指标,充分反映每个节点的演化波动;(2) MicroFluc 算法得到的网络演化序列与真实事件具有很好的匹配关系,考虑节点演化波动的差异有助于量化网络演化波动;(3) 在发生独立事件和连续多事件的社会网络的演化过程中,NodeED 方法事件检测的精确性和敏感性都优于对比方法,表明 NodeED 方法从节点演化波动的角度能够更好地量化网络演化波动,进行事件检测.

进一步的研究需要在以下两个方面展开:(1) 探究相似性计算指标和网络演化规律之间具体的对应关系;(2) 从节点的角度分析网络演化规律,对未来网络拓扑结构进行预测.

### References:

- [1] Li R, Lei KH, Khadiwala R, Chang CC. TEDAS: A Twitter-based event detection and analysis system. In: Proc. of the 28th IEEE Int'l Conf. on Data Engineering. 2012,41:1273–1276. [doi: 10.1109/ICDE.2012.125]
- [2] Eyal R, Rosenfeld A, Sina S, Kraus S. Predicting and identifying missing node information in social networks. ACM Trans. on Knowledge Discovery from Data, 2014,8(3):1–35. [doi: 10.1145/2536775]
- [3] Liu HK, Lü LY, Zhou T. Uncovering the network evolution mechanism by linkprediction. Scientia Sinica: Physica-Mechanica-Astronomica, 2011,41(7):816–823 (in Chinese with English abstract). [doi: 10.1360/132010-922]
- [4] Wang H, Hu W, Qiu Z. Nodes' evolution diversity and link prediction in social networks. IEEE Trans. on Knowledge and Data Engineering, 2017,29(10):1.
- [5] Ibrahim NMA, Chen L. Link prediction in dynamic social networks by integrating different types of information. Applied Intelligence, 2015,42(4):738–750. [doi: 10.1007/s10489-014-0631-0]
- [6] Hanley JA, Meneil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology, 1982,143(1):29–36. [doi: 10.1148/radiology.143.1.7063747]
- [7] Herlocker JL, Konstan JA, Terveen LG, Riedl JT. Evaluating collaborative filtering recommender systems. ACM Trans. on Information Systems, 2004,22(1):5–53. [doi: 10.1145/963770.963772]
- [8] Zhou T, Ren J, Medo M, Zhang YC. Bipartite network projection and personal recommendation. Physical Review E (Statistical, Nonlinear, and Soft Matter Physics), 2007,76(4):70–80. [doi: 10.1103/PhysRevE.76.046115]
- [9] Hu WB, Peng C, Liang HL, Du B. Event detection method based on link prediction for social network evolution. Ruan Jian Xue Bao/Journal of Software, 2015,26(9):2339–2355 (in Chinese with English abstract). . <http://www.jos.org.cn/1000-9825/4703.htm> [doi: 10.13328/j.cnki.jos.004703]
- [10] Fonseca JRS, Xerez R. How latent class models matter to social network analysis and mining: Exploring the emergence of community. In: The Influence of Technology on Social Network Analysis and Mining. 2013. 563–589. [doi: 10.1007/978-3-7091-1346-2\_25]
- [11] Erdos ED, Gemulla R, Terzi E. Reconstructing graphs from neighborhood data. ACM Trans. on Knowledge Discovery from Data, 2014,8(4):231–240. [doi: 10.1145/2641761]
- [12] Palla G, Barabási AL, Vicsek T. Quantifying social group evolution. Nature, 2007,446(7136):664–667. [doi: 10.1038/nature05670]
- [13] Rapoport A. Spread of information through a population with socio-structural bias: I. Assumption of transitivity. The Bulletin of Mathematical Biophysics, 1953,15(4):523–533. [doi: 10.1007/BF02476440]
- [14] Barabasi AL, Albert R. Emergence of scaling in random networks. Science, 1999, 509–512. [doi: 10.1126/science.286.5439.509]
- [15] Erdős P, Renyi A. On the strength of connectedness of a random graph. Acta Mathematica Hungarica, 1961,12(1):261–267. [doi: 10.1007/BF02066689]
- [16] Watts DJ, Strogatz SH. Collective dynamics of 'small-world' networks. Nature, 1998,393(6684):440–442. [doi: 10.1038/30918]
- [17] Marsili M, Vega-Redondo F, Slanina F. The rise and fall of a networked society: A formal model. Proc. of the National Academy of Sciences of the United States of America, 2004,101(6):1439–1442. [doi: 10.1073/pnas.0305684101]
- [18] Zhang QM, Xu XK, Zhu YX, Zhou T. Measuring multiple evolution mechanisms of complex networks. Scientific Reports, 2016,5: 10350.

- [19] Noble CC, Cook DJ. Graph-Based anomaly detection. Knowledge Discovery and Data Mining, 2003, 631–636. [doi: 10.1145/956750.956831]
- [20] Papadimitriou P, Dasdan A, Garcia-Molina H. Web graph similarity for anomaly detection. Journal of Internet Services and Applications, 2010,1(1):19–30. [doi: 10.1007/s13174-010-0003-x]
- [21] Mcculloh IA, Carley KM, Mcculloh IA, Carley KM. Social network change detection [Ph.D. Thesis]. Carnegie Mellon University School of Computer Science Institute for Software Research, 2008.
- [22] Wan X, Milios E, Kalyaniwalla N, Janssen J. Link-Based event detection in email communication networks. In: SAC Proc. of the ACM Symp. on Applied Computing. 2009. [doi: 10.1145/1529282.1529618]
- [23] Priebe CE, Conroy JM, Marchette DJ. Scan statistics on enron graphs. Computational and Mathematical Organization Theory, 2005,11(3):229–247. [doi: 10.1007/s10588-005-5378-z]
- [24] Wu B, Wang B, Yang SQ. Framework for tracking the event-based evolution in social networks. Ruan Jian Xue Bao/Journal of Software, 2011,22(7):1488–1502 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/3841.htm> [doi: 10.3724/SP.J.1001.2011.03841]
- [25] Jaccard P. Etude comparative de la distribution florale dans une portion des alpes et du jura. Bulletin De La Societe Vaudoise Des Sciences Naturelles, 1901,37(142):547–579.
- [26] Barabasi AAR. Emergence of scaling in random networks. Science, 1999,286(5439):509–512. [doi: 10.1126/science.286.5439.509]
- [27] Sørensen T. A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on Danish commons. Biol. Skr., 1957,5:1–34.
- [28] Adamic LA, Adar E. Friends and neighbors on the Web. Social Networks, 2003,25(3):211–230. [doi: 10.1016/S0378-8733(03)00009-1]
- [29] Ravasz E, Somera AL, Mongru DA. Hierarchical organization of modularity in metabolic networks. Science, 2002,297(5586):1551–1555. [doi: 10.1126/science.1073374]
- [30] Salton G, McGill MH. Introduction to Modern Information Retrieval. New York: McGraw-Hill, 1983.
- [31] Cai Q, Gong M, Ma L, Ruan S, Yuan F, Jiao L. Greedy discrete particle swarm optimization for large-scale social network clustering. Information Sciences, 2015,316(41):503–516. [doi: 10.1016/j.ins.2014.09.041]

#### 附中文参考文献:

- [3] 刘宏鲲,吕琳媛,周涛.利用链路预测推断网络演化机制.中国科学:物理学·力学·天文学,2011,41(7):816–823. [doi: 10.1360/132010-922]
- [9] 胡文斌,彭超,梁欢乐,杜博.基于链路预测的社会网络事件检测方法.软件学报,2015,26(9):2339–2355. <http://www.jos.org.cn/1000-9825/4703.htm> [doi: 10.13328/j.cnki.jos.004703]
- [24] 吴斌,王柏,杨胜琦.基于事件的社会网络演化分析框架.软件学报,2011,22(7):1488–1502. <http://www.jos.org.cn/1000-9825/3841.htm> [doi: 10.3724/SP.J.1001.2011.03841]



胡文斌(1976—),男,湖北武汉人,博士,副教授,博士生导师,主要研究领域为人工智能与智能仿真优化,社会网络仿真与优化,大数据分析 with 优化理论,智能计算与智能决策,智能交通系统.



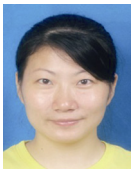
邱振宇(1992—),男,硕士,主要研究领域为社会网络仿真与优化,大数据分析 with 优化理论,智能交通系统.



王欢(1989—),男,硕士,主要研究领域为社会网络仿真与优化,大数据分析 with 优化理论,智能交通系统.



聂聪(1993—),男,硕士,主要研究领域为社会网络仿真与优化,大数据分析 with 优化理论.



严丽平(1980—),女,硕士,主要研究领域为社会网络仿真与优化,大数据分析 with 优化理论,智能交通系统.



杜博(1983—),男,博士,副教授,博士生导师,主要研究领域为人工智能,图像分析.