

基于用户相对影响权重的热点事件传播阈值模型

张德阳¹, 韩益亮¹, 李晓龙¹, 潘峰¹, 周宜兴^{1,2}

(1. 武警工程大学电子技术系, 陕西 西安 710086;

2. 中国人民武装警察部队河北总队, 河北 石家庄 050000)

摘要: 结合现有的线性阈值模型, 提出一种新的用户相对影响权重函数; 结合信息自身敏感性以及节点自身的接受阈值特性, 提出一种能衡量信息传播能力的阈值模型, 即 URLT 模型。通过仿真实验模拟在不同网络、不同敏感度信息以及不同节点阈值的传播, 并对最终传播情况进行对比, 实验结果证明, 其最终影响范围与真实传播情况符合。因此, 该模型对信息传播规律的发现与抑制有一定的参考价值。

关键词: 热点事件; 阈值; 相对权重; 网络传播

中图分类号: TN929

文献标识码: A

doi: 10.11959/j.issn.2096-109x.2017.00195

Threshold model for hot events spreading based on relative impact weight of the user

ZHANG De-yang¹, HAN Yi-liang¹, LI Xiao-long¹, PAN Feng¹, ZHOU Yi-xing^{1,2}

(1. Department of Electronic Technology, Engineering University of the PAP, Xian 710086, China;

2. Hebei Corps of the PAP, Shijiazhuang 050000, China)

Abstract: A new relative influence weight function was proposed in combination with the existing linear threshold model. Combining with the information itself sensitivity and the threshold characteristic of the node itself, a threshold model(URLT model) was proposed, which can measure the information communication ability. By simulating the spread of different networks, different sensitivity information and different node thresholds, and comparing the final propagation situation, the experimental results show that the final influence range is consistent with the real spread situation. Therefore, the model has some reference value for the discovery and suppression of the law of information dissemination.

Key words: hot event, threshold, relative weight, network spread

1 引言

随着因特网的飞速发展, 各种基于网络的社交媒体已经形成了全新的信息传播方式。尤其是对当今社会中一些热点问题, 社会舆论的发展更是达到了一个前所未有的态势。国际、国内的重

大事件, 人们都能通过网络以及相关媒体在第一时间获取信息。也正是由于网络的便捷性, 接收信息和发布信息的速度都有了极大提高。这种通过网络表达观点, 诉说想法的行为, 无疑对社会稳定、舆论引导造成了很大的压力。

通常意义下的社会热点事件包括 2 个方面的

收稿日期: 2017-08-17; 修回日期: 2017-09-02。通信作者: 张德阳, 593251893@qq.com

基金项目: 国家自然科学基金资助项目 (No.61572521); 军事科学研究计划课题基金资助项目 (No.16QJ003-097)

Foundation Items: The National Natural Science Foundation of China (No.61572521), Research Project of Military Science (No.16QJ003-097)

内容: 一是由于社会环境长期形成的、较短时间内无法得到妥善解决的, 具有一定的周期性, 如大学生就业问题、开放“二胎”问题、全球气候变暖问题等, 该类事件在一段时间内具有较高的关注度, 之后重新回到正常水平; 二是由于国际国内热点事件的发生引起的, 具有突发性, 诸如此类的事件中的部分带有故意炒作成分, 目的是为了提提高事件关注度。

因此, 对社会中热点事件的研究是一个亟待解决的问题, 目前已经存在的对热点事件的研究仅停留在理论阶段, 没有将具体传播中的数据进行量化、建立数学模型。目前, 国内对热点事件的影响及对策方面的研究占了大多数。本文重点对热点事件中的舆论传播问题进行研究, 利用已有的传播模型进行综合分析。本文综合了用户相对权重与线性阈值模型的优点, 利用用户的相对权重函数 $R(x, y)$ 与信息敏感度 P 和阈值 θ 相结合, 提出了一种新的 URLT 模型。

2 相关概念及定义

2.1 相关概念

2.1.1 社会热点事件

社会热点事件定义为比较受广大群众关注或者欢迎的新闻或信息, 或指某时期引人注目的地方或问题^[1]。

社会热点事件一般具有鲜明的时代特征, 它是伴随着生产力的发展而来的产物, 反映的是社会中存在的矛盾性问题。具体体现在 3 个方面: 一是突发性, 热点事件的发生与发展有可能只是在很短的时间内达到较大的传播范围; 二是敏感性, 热点事件所反映的问题大部分是在普通民众之间产生的, 代表了广大民众的诉求, 不可避免地要涉及人民群众的切身利益; 三是流动性, 可能今天所谓的热点问题, 到了明天就不是, 或者民众对这个问题的兴趣程度大大减弱。

2.1.2 传播方式

社会网络由无数的个体组成, 每个个体都有不同的家庭背景、性格习惯, 因此每个人在得到信息后的处理方式也不同。当接收到自己比较感兴趣的信息时, 个体通常会选择接受并且进行下一步的传播; 当接收到自己不擅长或者不感兴趣

的信息时, 则会选择不接受、不传播; 如果认为信息只对自己有利, 对他人作用不大的时候, 则会选择只接受不传播。

如今社会的传播方式^[2]比历史上的任何时期都要复杂, 较之以往线下口口相传的方式, 现在的传播方式多趋于网络化, 人们足不出户就可以获取自己想要的信息, 这一切都得益于互联网的发展, 人们可以在互联网上应聘求职、聊天交友、买卖交易等。然而, 也正是互联网的便利性, 使社会热点问题中舆论的传播形式越来越复杂, 对舆情的管理难度也越来越高。

2.1.3 影响力

社交影响力只能通过人们之间的相互活动体现出来, 其定义有如下 2 种。

1) Rashotte^[3]根据用户之间的行为及其产生的效果, 将社交影响力定义成由于交互行为的存在改变人的思想、感情、行为的现象。

2) 一些学者根据统计学的特点, 将其定义为在影响力度分布中度值较大的一部分人。

2.1.4 敏感度

敏感度描述的是个体受他人影响的趋势, 实际意义表示的是用户本身对事件被感染的程度, 近来研究敏感度与影响力在社会传播中哪个的作用更大。文献[4]发现在部分社交软件中影响力的作用和敏感性的作用呈现此消彼长之势, 同一个用户不可能同时具有高影响力和高敏感性。那么影响力和敏感度在传播过程中哪个因素的作用更大, 这也是一个新的问题。

2.2 影响力的度量方法

影响力度量的传统方法^[5]是基于网络拓扑结构的度量, 包括基于节点的度量和基于路径的度量。一般将社交网络定义为 $G=(V, E)$, 其中, V 定义为节点的集合, 每个节点表示社交网络中的用户; E 定义为边的集合, 用边来表示用户之间的关系(亲人, 朋友, 敌人), 节点的个数为 $n=|V|$ 。节点 u 对节点 v 的影响为 b_{uv} 。

2.3 影响力最大化模型

社交信息在社交网络中的传播过程简单来说就是从最初的已经被激活的群体扩散到最终没有可以被激活节点。节点有激活和待激活 2 种状态, 处于激活态的用户以一定概率 λ 通过激活行为为

表1 节点影响力的度量方法

类型	度量方法	公式	方法描述
基于节点度	入度	$Deg^{in}(v_i) = \sum_j a_{j,i}$	反映该用户的受关注程度
	出度	$Deg^{out}(v_i) = \sum_j a_{i,j}$	反映该节点对其他节点的关注程度
	度中心度	$C^{DEG}(v_i) = \frac{Deg(v_i)}{n-1}$	反映一个节点对邻居节点的平均影响力
基于最短路径	紧密中心度	$C^{CLO}(v_i) = \frac{1}{\sum_{v_j \in V \setminus \{v_i\}} g'_{ij}}$	一个节点对其他节点的影响, 反映用户的社会关系强弱
	介数中心度	$C^{BET}(v_i) = \frac{\sum_{s \neq i} \{g'_{st}\} \{g_{st}\} }{n(n-1)/2}$	反映社会网络中节点的位置重要程度

活待激活的用户, 且这一激活过程是不可逆的。用数学语言就是选定不同的 k 个节点, 同时在网络中进行扩散传播, 对比最后影响范围 $Inf(k)$, 选定影响范围最大的一种情况为达到了影响力最大化。已有的对影响力最大化问题的研究包括贪心算法^[6]、独立级联模型^[6,7]和线性阈值模型^[8,9]以及基于贪心算法和线性阈值模型的“积累”特性结合的一种新的影响力最大化算法 HPG 算法^[10]。本文综合反映出真实世界中的传播规律, 当存在多方信息的来源时, 接收者总会选择对自己影响最大的一个进行接受, 体现了信息传播的竞争机制, 利用用户的相对权重 $R(x, y)$ 与信息的敏感度 P 和阈值 θ 相结合, 提出了一种新的 URLT 模型。

2.3.1 线性阈值模型

线性阈值模型简称 LT 模型, 是一种基于节点特异性的模型。第 2.2 节已对 $G = (V, E)$, b_{uv} 等有关参数进行了定义, 在此不再说明。定义新的参数 θ 为特异性阈值, 表示一个节点被激活时的临界权重值, 这是由节点本身所决定的。定义 $A(v)$ 为节点 v 已经激活的邻居节点集合, 并且所有邻居节点对节点 v 的权重和不超过 1, 即

$$\sum_{u \in A(v)} b_{uv} \leq 1$$

每一个节点都有自己的特异性阈值 $\theta (0 < \theta \leq 1)$, 只有当

$$\sum_{u \in A(v)} b_{uv} \geq \theta$$

时, 节点 v 的下一时刻变为激活状态, 且保持激活状态不变。需要指出的是, 在 LT 模型中, 当一个节点 u 尝试激活节点 v 而没有成功时, 所产

生的影响力 b_{uv} 会被积累下来, 直到节点 v 被激活或传播过程结束。

2.3.2 HPG 算法中的改进权重

在已有的线性阈值模型中用 b_{uv} 表示节点 u 对节点 v 的影响, 通常用式(1)表示。其中, $d(v)$ 表示节点 v 的度数。

$$b_{uv} = \frac{1}{d(v)} \quad (1)$$

在 HPG 算法中, 认为节点 u 与节点 v 之间存在相互影响, 同时认为在同一网络中节点 u 对 v 的影响 b_{uv} 与节点 v 对 u 的影响 b_{vu} 不同。下面给出它的定义。

以往模型往往忽视了节点本身的作用, 将所有的邻居都作为同等的地位, 忽视了用户的差异性, 与现实情况不符。Tian 等^[10]指出 b_{uv} 的大小是节点 v 自身的感受, 反映的是对于指向它的节点的一个权威性的体现, 与其他节点没有关系, 因此只需要考虑与之相连的节点即可, 也就是邻接图的思想, 即只由节点 v 与指向它的邻居节点组成, 并且改进了 b_{uv} 的估计公式。

$$\begin{cases} NG(v) = G'(V', E') \\ V' = \{v\} \cup N(v) \\ E' = \{(u, v) \mid x, y \in V', (x, y) \in E\} \\ b_{uv} = \frac{Deg^{out}(u)}{\sum_{w \in N(v)} Deg^{out}(w)} \end{cases} \quad (2)$$

式(2)表示的是节点 u 与节点 v 之间的影响力主要由 v 的邻接结构决定。 $N(v)$ 为节点 v 的邻居节点集合, 这里的邻居指的是节点 v 的入边邻居。

3 基于用户相对影响权重的阈值模型

3.1 信息的敏感度 P

信息的传播过程大概分为3个阶段：信息的启发阶段、信息的渲染阶段、信息的传播阶段^[11]，在每一个阶段中信息的敏感性是不同的。本文将信息在传播过程中最为敏感时的敏感度 P 记为1，但一个信息的敏感度最高的时间是短的，因此通过信息敏感度判断一个信息是否敏感。通常将敏感度在0.5~1这一区间内的信息定义为敏感信息。从图1可以看出，一个信息的传播是有一定的生存期的，在其敏感度达到最大后，其敏感度又会逐渐下降到0。

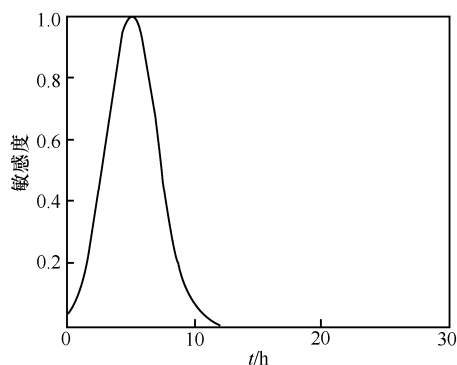


图1 信息敏感度变化曲线

3.2 节点的特异性阈值 θ

节点的特异性阈值表明一个用户对一个信息的接受程度，该值与节点自身有密切联系，不同类型人群的阈值不同，但为了方便实验论证，同时也考虑大多数人的选择情况，本文将每个节点的特异性阈值 θ 设定为0.5~1，阈值越小，该节点接受信息的可能越大；反之，接受信息的可能越小。在第4节的仿真实验中，本文对节点的不同阈值进行了对比实验，并验证了理论的真实性的。

3.3 算法定义

3.3.1 算法描述

现有模型中主要考虑的是在社交网络中已有的网络拓扑结构，很少考虑用户自身的影响力。现有工作基本上都是将每个用户平等对待，在模型中赋予相同的权重，但在实际生活中由于每个个体的自身经历、学历、经验水平各不相同，每个人的想法又被其他人，即邻居所影响，因此本文提出基于用户之间影响函数^[12] $R(x,y)$ 来表示

节点 x 对节点 y 的相对影响权重，由于在社交网络中的每个节点的邻居都不同，因此在真正判断2个节点间的影响力哪个更大时，如果哪个节点的相对影响权重更大，那么它的影响力就更强。

结合式(2)，本文将影响函数定义为

$$R(u,v) = 2 \frac{b_{uv}}{b_{uv} + b_{vu}} \quad (3)$$

需要说明的是， $R(u,v)$ 一般不等于 $R(v,u)$ ，这是由节点 u 与节点 v 之间的相对权威性决定的。下面进行实例分析。

3.3.2 实例分析

如图2所示的有向网络拓扑图中，节点1与节点7的度数较大，可以将其视为在网络中处于关键位置。对节点1而言，它的入边节点有 $N(v) = \{13, 18, 11, 7\}$ ，因此可以将其邻接图简化为图3。

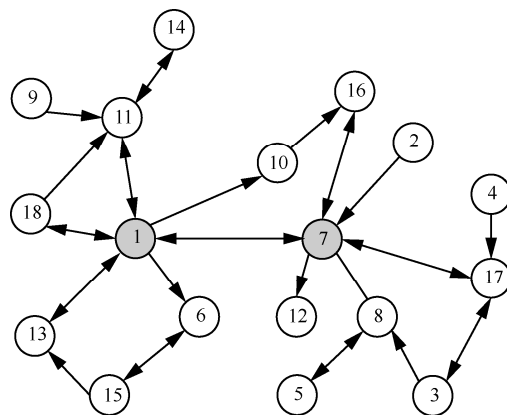


图2 有向网络拓扑图

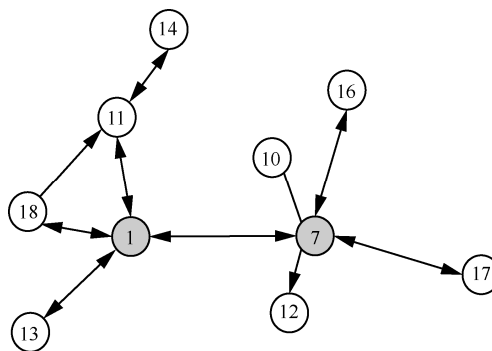


图3 节点1的邻接图

节点1各个入边节点的出度 Deg^{out} 如表1所示。

表1 节点1各个入边节点的出度 Deg^{out}

u_i	13	18	11	7
$Deg^{out}(u_i)$	1	2	2	4

那么,可以得到 b_{u1} 分别是 $b_{13,1}=0.111$, $b_{8,1}=0.222$, $b_{11,1}=0.222$, $b_{7,1}=0.444$ 。同样地,可以得到节点7的邻接图如图4所示。

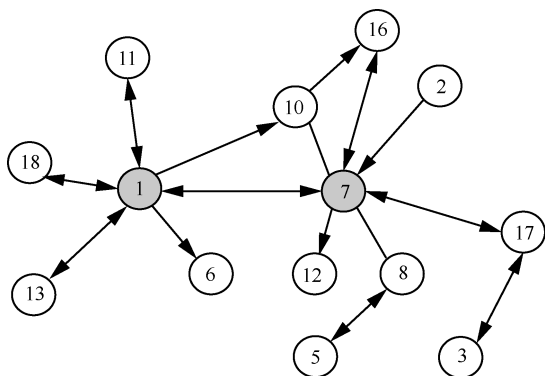


图4 节点7的邻接图

可以计算得到 $b_{1,7}=0.428$, $b_{10,7}=0.143$, $b_{16,7}=0.071$, $b_{2,7}=0.071$, $b_{17,7}=0.143$, $b_{8,7}=0.143$ 。

由式(3)可以得到 $R(7,1)=1.018$, $R(1,7)=0.981$ 。显然, $R(u,v)$ 的作用是将连接节点 u 与节点 v 的权重根据双方的地位重新做了调整,能够较好地反映二者之间的相互影响权重大小。

3.4 URLT 模型

根据信息在网络传播中的特点:信息的传播不仅与节点之间的影响力有关,也与信息本身有关,本文提出了一种基于用户相对影响函数与线性阈值模型相结合的新的模型。基于网络拓扑图中的节点特性,本文将用户相对权重函数 $R(x,y)$ 与信息敏感度 P 相乘,与节点的特异性阈值相比,得到实际的传播概率 λ 。

具体规定如下。

1) 如果与该节点相连的其他所有节点中 $R(x,y)P$ 的最大值仍然没有超过节点的阈值 θ ,则节点不被激活。

2) 如果该节点相连的其他所有节点中的 $R(x,y)P$ 的最大值大于等于节点的阈值,则认为该节点被成功激活。

3) 由于 θ 不超过1,该节点相连的其他所有节点中的 $R(x,y)P$ 如果超出1,则认为该节点必然被激活。

可表示为

$$\lambda = \begin{cases} 1, \text{Max}(R(x,y)P) > 1 \\ 1, \text{Max}(R(x,y)P) \geq \theta \\ 0, \text{Max}(R(x,y)P) < \theta \end{cases}$$

以上规定可以反映出真实世界中的传播规律,当存在多方信息的来源时,接收者总会选择对自己影响最大的一个接受,体现了信息传播的竞争机制。

4 仿真实验分析

本文实验环境如下:使用 Matlab R2015b 在 Windows10 系统下,通过计算机模拟在真实世界中存在的 BA 无标度网络和 WS 小世界网络,观察该模型在仿真网络中的传播情况,并对不同网络模型中的信息传播情况进行对比,对比不同敏感度的信息在相同网络情况下的最终影响力范围。

4.1 不同网络中的传播对比

图5是在 BA 无标度网络中,当信息敏感度 P 为0.5、节点阈值 θ 为0.5时的传播情况,图5反映在不同个数的初始节点的情况下,随着初始节点个数的增加,最终的影响范围呈现增长速度变慢的趋势,并且从图5的局部坐标来看,虽然初始节点的个数增加,但其最终的影响范围还有可能减小,可以比较客观地反映真实社会中的信息传播规律,即初始节点的个数越多,最终的影响范围不一定就越大。

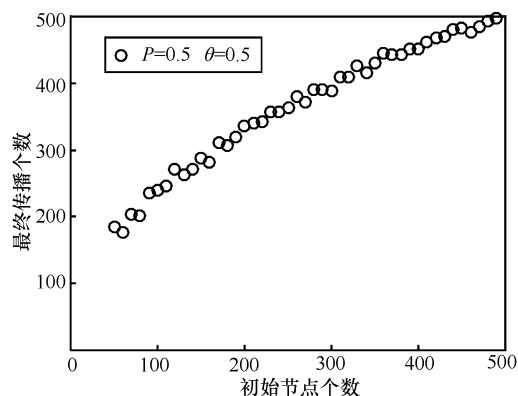


图5 BA 无标度网络中 URLT 模型的传播图

图6模拟的是在小世界网络中的传播情况,小世界网络节点与节点之间的耦合性较强,每个节点都有固定个数的邻居节点,因此信息的传播过程较 BA 无标度网络速度快。与图5相比,在网络规模为500的小世界网络中,仅100个初始节点就可以达到最大的影响范围,随着初始节点个数的增加,所影响的范围也呈指数级快速增长,与小范围内的信息呈现病毒性传播符合,也反映

了在人群密集的社会网络中，敏感信息的传播速度是极其迅速的。

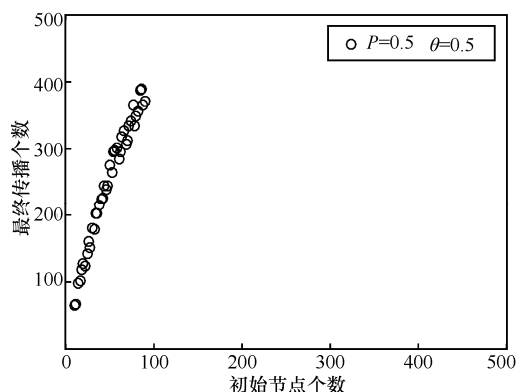


图6 WS网络中URLT模型的传播情况

4.2 不同敏感度信息的对比

如图7所示，在同一网络条件以及相同节点阈值的情况下，分别用不同敏感度的信息传播，可以看出在初始节点个数相同的情况下，其最终影响范围不同，且高敏感度信息的传播速度明显快于低敏感度信息。结合真实世界的传播规律，高敏感度的信息比低敏感度的传播速度快，但其传播速度的趋势逐渐变慢。反之，低敏感度信息的最终影响范围呈现斜率基本不变的直线，对于增加的初始节点数目，其最终影响范围的增长速度保持不变，没有出现增长速度变快并最终趋于平衡。

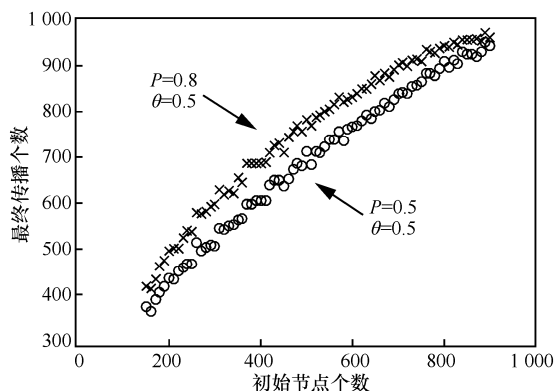


图7 不同敏感度的传播对比

4.3 在不同阈值中的对比

与4.2节的对比实验类似，分别选取信息敏感度 P 为0.5，阈值 θ 为0.5和1的两组传播曲线，这里用传播曲线的斜率变化作为信息传播速度的快慢，实验表明，在阈值为0.5时，即当每个节点对信息的接受程度为半信半疑的时候，信

息的传播速度是快速的；而当阈值为1时，此时的传播速度随着初始节点的增加并没有明显的提升，即节点自身对信息的判断能力较强时，信息传播是缓慢的，这个结论与真实社会网络中的传播一致。

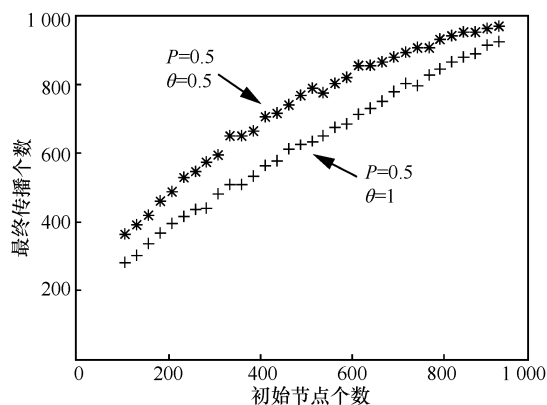


图8 不同阈值的传播对比

通过以上的几组对比实验，可以发现本文提出的URLT模型基本符合真实社会网络中的信息传播规律，其中，信息的敏感程度与节点的阈值等重要参数，也代表了信息和节点本身的属性，是一个复杂的变量函数。因此，如何将这些参数从理论层次上升到科学计算，是下一步要关注的问题。

5 结束语

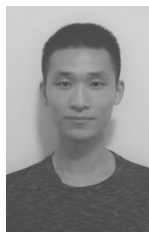
本文以社会热点事件作为研究背景，分析了热点事件与其他信息的异同，对相关概念进行了界定，并对已有的研究进行了说明。针对热点事件的传播规律，在已有研究的基础上，提出了本文的创新点，即不仅要考虑网络拓扑结构，也要考虑节点之间的相互作用，即本文所说的相对影响权重。根据信息的传播规律，将信息在不同时期的敏感度定义在一定的区间内，方便进行数值计算。结合已有的线性阈值模型，最终本文提出了用户相对影响权重阈值模型，经过仿真实验分析，该模型可以较好地反映信息的传播规律。本文的研究仅是对静态网络进行了分析，并没有对动态网络进行分析，如何在动态网络中进行建模分析也是今后研究的重要内容。

参考文献：

- [1] LIU T, ZHONG Y, CHEN K. Interdisciplinary study on popularity

- prediction of social classified hot online events in China[J]. Tele-
matics and Informatics, 2017,34(3):755-764.
- [2] JUNG S H, KIM J. A new way of extending network coverage:
relay-assisted D2D communications in 3GPP[J]. ICT Express, 2016,
2(3): 117-121.
- [3] RASSHOTTE. Social influence: the blackwell encyclopedia of
social psychology[M]. Malden: Blackwell Publishing. 2007.
- [4] ARAL S, WALKER D. Identifying influential and susceptible
members of social networks[J]. Science. 2012, 337(6092): 337-341.
- [5] 吴信东.在线社交网络影响力分析[J]. 计算机学报, 2015, 37(4):
735-751.
WU X D. Analysis of the influence of online social networks[J].
Journal of computer science, 2015,37 (4): 735-751.
- [6] KEMPE D, KLEINBERG J, TARDOS E. Maximizing the spread of
influence in a social network[C]//The 9th ACM SIGKDD Interna-
tional Conference on Knowledge Discovery and Data Mining.
2003:137-146.
- [7] GRUHL D, GUHA R, LIBEN-NOWELL D, et al. Information
diffusion though blogspace[C]//The 13th International Conference
on World Wide Web. 2004:491-501.
- [8] CHEN W, YUAN Y, ZHANG L. Scalable influence maximization
in social networks under the liner threshold model[C]//The 2010
IEEE International Conference on Data Mining. 2010:88-97.
- [9] CHEN W, WANG C, WANG Y. Scalable influence maximization
for prevalent viral marketing in large-scale social networks[C]//The
16th ACM SIGKDD International Conference on Knowledge Dis-
covery and Data Mining. 2010:1029-1038.
- [10] TIAN J T, WANG Y T, FENG X J. A new hybrid algorithm for
influence maximization in social networks[J]. Chinese Journal of
Computers, 2011,34(10):1956-1964.
- [11] 李弼程,戴锋. 网络舆情分析——理论、技术与应用策略[M]. 北
京:国防工业出版社. 2015.
LI B C, DAI F. Analysis of network public opinion: theory tech-
nology and application strategy[M]. Beijing: National Defense In-
dustry Press. 2015.
- [12] WANG J L. An online social network information propagation
model based on relative weight of users[J]. Journal of Physics 2015,
64(5).

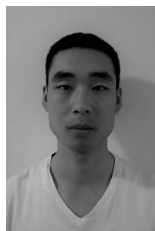
作者简介:



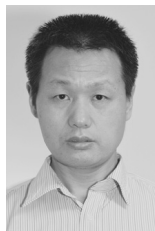
张德阳 (1993-), 男, 河南新乡人,
武警工程大学硕士生, 主要研究方向为
舆情分析、社交网络。



韩益亮 (1977-), 男, 甘肃会宁人,
武警工程大学教授、博士生导师, 主要
研究方向为密码学、复杂网络分析。



李晓龙 (1991-), 男, 河南安阳人,
武警工程大学硕士生, 主要研究方向为
网络安全、舆情分析。



潘峰 (1967-), 男, 北京人, 武警
工程大学副教授, 主要研究方向为密码
学、复杂网络分析。



周宜兴 (1976-), 男, 河北沧州人,
武警工程大学硕士生, 主要研究方向为
复杂网络分析。