

Factors Analysis for Improved Information Dissemination Model on Microblog

Liang Li, Man Yu, Ping Sun, Gaoxia Wang*

College of Science, China Three Gorges University, Yichang Hubei
Email: gaoxiawang@163.com

Received: Dec. 29th, 2014; accepted Jan. 18th, 2015; published Jan. 26th, 2015

Copyright © 2015 by authors and Hans Publishers Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY).
<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Microblog is a popular platform for the information dissemination of social networks, which plays an important role for communication and supervision of public voice. Based on the classic SIR model and the process of microblog information dissemination on the state transformation between user nodes, including easy disseminators, disseminators and silencers, the improved microblog dissemination model with certified users and absolute silencers was built. The factors affecting the spread of microblog information were analyzed and simulated. Researches show that the higher the transmission rate of authenticated users and the normal nodes, the lower the stop transmission rate, and the higher the rate of absolute silence, the sooner time to reach steady state of the microblog information dissemination. It has some theoretical basis and practical significance for actually controlling information dissemination of the popular social platform.

Keywords

Social Network, SIR Model, Certified User, Absolute Silence Node

改进的微博传播模型的影响因素分析

李亮, 余曼, 孙平, 王高峡*

三峡大学理学院, 湖北 宜昌
Email: gaoxiawang@163.com

收稿日期: 2014年12月29日; 录用日期: 2015年1月18日; 发布日期: 2015年1月26日

*通讯作者。

摘要

微博作为社会网络信息传播的热门平台,对舆论传播与监督具有重要作用。本文基于经典SIR传播模型,以微博信息在易传播者、传播者和沉默者之间的状态转换为基础,建立了基于认证用户和绝对沉默者的微博网络传播模型,并通过仿真分析了微博传播的影响因素。研究表明认证用户节点和普通节点的传播率越高,停止传播率越低或者绝对沉默率越高,则微博信息传播达到稳定状态越快,对实际控制社交平台的传播具有一定的理论分析依据和实践指导意义。

关键词

社会网络, SIR模型, 认证用户, 绝对沉默节点

1. 引言

自然界和人类社会中的大量复杂系统都可以抽象为网络结构,反之复杂网络也可视为由若干节点通过一定方式连接而成的复杂系统[1]。在 *Strogatz* 和 *Watt* 的小世界网络模型[2], *Albert* 和 *Barabasi* 的无标度网络模型[3]的基础上,众多学者致力于现实系统的网络特性实证研究,与此同时各个领域掀起了研究复杂网络的热潮[4]。复杂网络理论不仅提供了探究现实生活中复杂系统的全新科学视角,而且成为改造客观世界的新方法和有力武器。

社交网络作为以信息交流为核心的服务平台,成为互联网发展的热点领域,包括 Facebook、Bebo 和 Mini 等。目前国内影响力最深的新浪微博,其日益攀升的用户规模和迅猛提升的影响力在国内掀起了“微革命”[5]。基于微博网络平台,用户依据教育背景、兴趣爱好、职业背景等方面的差异,根据各自的喜好,可选择性地对其他用户关注,从而形成虚拟的“关系圈”。在此微博信息传播的“关系圈”中,“关系”用于表示信息传播用户之间某种特定的联系。基于复杂网络基本理论,微博用户可抽象为网络节点,用户关系抽象为网络连边,而连边方向表示信息传播的路径[6]。微博信息传播网络本质上是社交网络,研究表明其具有无标度和小世界特性[7]。

社会信息传播网络的研究分别基于不同机制的传播模型[8] [9]。*Watts* 和 *Strogatz* 研究了 *SW* 网络模型,提出 *SW* 效应可加快传染病的传播过程。王林等人研究了 *SIS* (*Susceptible Infected Susceptible*)和 *SIR* (*Susceptible Infected Recovered*)模型与网络模型的联系[10]。对于微博信息传播网络,若干学者以 *SIR* 模型为基础进行了相关研究[11]。李可嘉,王义康以 *SIR* 模型为基础,研究了基于“热传播节点”的社交网络信息传播模型与仿真分析[12]。刘丹等模拟了基于 *SIR* 模型的微博信息传播过程[13]。本文根据微博传播的实际情况,在微博传播过程中提出“认证用户”和“绝对沉默节点”,建立了基于 *SIR* 的微博信息传播改进模型,通过模型仿真分析了实际微博传播网络参量对信息传播的影响因素。

2. 改进的微博信息传播模型

用 $G=(V,E)$ 表示复杂网络的有向图, $A=(a_{ij})$ 表示图 G 的邻接矩阵,其中 V 表示节点集, $v \in V$ 表示 V 中的节点, E 表示边集, $e_{(i,j)} \in E$ 表示从节点 i 到节点 j 的有向边。出边的数量称为出度,表示为 $k_i^{out} = \sum_j a_{ij}$, 入边的数量称为入度,表示为 $k_i^{in} = \sum_j a_{ji}$ 。在微博信息传播网络中,节点代表微博用户,如果用户 i 关注了其它用户 j , 则有出边 $e_{(i,j)}$, 如果受到其他用户 j 的关注, 则有入边 $e_{(j,i)}$ 。关注数对应节点的出度 k_i^{out} , 粉丝数即被关注数, 对应节点的入度 k_i^{in} 。

SIR 模型考虑网络结构对病毒传播行为的影响时,将人群中所有个体分成三类,即易感染者、感染者和免疫者。微博信息传播中,将微博用户也分为三类:易传播者、传播者和沉默者,在微博网络中分别对

应易传播节点、传播节点和沉默节点。下面通过三类节点之间的状态转换来建立微博信息传播模型。

假设 t 时刻微博用户总数不变, 记 $s(t)$ 为 t 时刻还没有看到并且有可能看到该微博信息的用户所占的比例(易传播者), 记 $i(t)$ 为在 t 时刻看到该信息并具有转发能力的用户所占的比例(传播者), 记 $r(t)$ 为在 t 时刻看到该信息但并不具有转发能力的用户所占的比例(沉默者), 且有 $s(t)+i(t)+r(t)=1$ 。假设易传播者 s 接受信息后以传播率 λ 向其它用户传递信息, 称 λ 为 s 的传播率。 μ 为信息传播者 i 将信息传递给沉默者 r 的概率(停止传播率)。则基于经典 *SIR* 的微博信息传播模型微分方程如下[10]:

$$\begin{cases} \frac{ds(t)}{dt} = -\lambda s(t)i(t) \\ \frac{di(t)}{dt} = \lambda s(t)i(t) - \mu i(t) \\ \frac{dr(t)}{dt} = \mu i(t) \end{cases} \quad (1)$$

在实际微博网络中, 存在普通用户和认证用户之分, 认证用户拥有庞大的粉丝数, 粉丝数可以直接影响微博信息的传播情况。因此在微博信息传播中, 认证用户对信息的传播起着至关重要的作用。另外, 当易传播者始终没有看到传递的信息或者对此类信息不感兴趣而不去理会, 则易传播者会直接变为沉默者, 我们称这类沉默者用户为绝对沉默节点。本文考虑微博信息传播网络中同时含有认证用户和绝对沉默节点的模型。

假设 λ_1 为认证用户的传播率, λ_2 为普通用户的传播率, ρ 为易传播节点变为沉默节点的概率, 即绝对沉默率。 i_0 表示网络中认证用户节点的初始比例, $i(t)$ 表示 t 时刻普通传播节点所占的比例, 由 $i_1(t)$ 和 $i_2(t)$ 两部分组成, 其中 $i_1(t)$ 表示由认证用户节点传播产生的普通传播节点, $i_2(t)$ 表示由普通传播节点产生的新的普通传播节点。则由认证用户节点产生的新的普通传播节点数为 $\lambda_1 i_0 s(t)$, 由普通传播节点产生的新的普通传播节点数为 $\lambda_2 s(t)i(t)$, 由传播节点转化成沉默节点的数为 $(1-\mu)[\lambda_1 i_0 s(t) + \lambda_2 s(t)i(t)]$, 由传播节点转化为绝对沉默节点的数为 $\rho s(t)r(t)$ 。随着信息的传播, $\rho s(t)r(t)$ 会使得易传播节点的数量减少, 同时使得沉默节点的数量增加。因此可以得到 t 时刻关于 $s(t)$ 和 $r(t)$ 的微分方程:

$$\frac{ds(t)}{dt} = -\lambda_1 i_0 s(t) - \lambda_2 s(t)i(t) - \rho s(t)r(t) \quad (2)$$

$$\frac{dr(t)}{dt} = (1-\mu)[\lambda_1 i_0 s(t) + \lambda_2 s(t)i(t)] + \rho s(t)r(t) \quad (3)$$

又此模型满足 $s(t)+i(t)+i_0+r(t)=1$, 可得到关于 $i(t)$ 的微分方程:

$$\frac{di(t)}{dt} = \mu[\lambda_1 i_0 s(t) + \lambda_2 s(t)i(t)] \quad (4)$$

分别设初始状态下的沉默节点所占比例和易传播节点所占比例为 $r(0)=r_0$, $s(0)=s_0$ 。则建立含有认证用户节点和绝对沉默节点的微博传播模型如下:

$$\begin{cases} s(t)+i(t)+i_0+r(t)=1 \\ \frac{ds(t)}{dt} = -\lambda_1 i_0 s(t) - \lambda_2 s(t)i(t) - \rho s(t)r(t) \\ \frac{di(t)}{dt} = \mu[\lambda_1 i_0 s(t) + \lambda_2 s(t)i(t)] \\ \frac{dr(t)}{dt} = (1-\mu)[\lambda_1 i_0 s(t) + \lambda_2 s(t)i(t)] + \rho s(t)r(t) \\ s(0)=s_0, r(0)=r_0 \end{cases} \quad (5)$$

3. 微博传播影响因素的仿真分析

3.1. 微博传播模型仿真

本文选新浪微博上某博主所发的微博信息进行分析,信息内容是有关该博主和粉丝之间的互动和交流。该条信息发出后,受关注的情况和主要关系如下表 1 和图 1 所示(一找微分析网站)。

在表 1 中,点赞指通过点击相应图标的方式来表达用户对于该信息的赞同和喜爱,点赞数即点赞的用户数;曝光量指在一段时间内看到该博文的用户数;真实点赞数是指去除水军后的点赞数;参与用户数是指参与评论和转发的用户总数,是除去了重复、水军以及被新浪屏蔽的用户。(在微博上,水军指为他人转发评论造势的批量微博用户,这批用户由机器算法操纵去造势,使用大量虚假用户去共同完成,而这些临时征集来为特定微博服务的用户,即称为水军。)

根据表 1 的统计显示,作为一个拥有 3730 万粉丝的微博认证用户发了一条信息后,一天内约有 14,000 人对该信息进行了转发,约 34,000 人进行了评论。又由图 1 看出,转发和评论的高峰基本集中在前十小时以内,经估计该信息的传播率约为 0.0016%,则停止传播率 μ 为 99.9984%。统计得到认证用户传播节点的每小时传播率 λ_1 约为 100,普通传播节点的每小时传播率 λ_2 约为 5。在注册用户数约为 5 亿的新浪微博网络中,假定初始传播的认证用户数约为 10,即 $i_0 = 2 \times 10^{-8}$ 。在信息发布最初,除了认证用户传播节点,其他所有节点应都为易传播节点,即 $s_0 \approx 1$ 。综合以上有如下模型参数初始值:

$$\lambda_1 = 100, \lambda_2 = 5, \mu = 99.9984\%, i_0 = 2 \times 10^{-8}, i(0) = 0, s_0 \approx 1, r_0 = 0$$

因为 $i_0 = 2 \times 10^{-8}$ 极小,为计算方便可假设 $s_0 = 1$ 。另外,假设微博网络中模型参数在一定时间内保持稳定,对模型进行求解可得到微博信息传播网络中各类节点所占比例随时间的变化情况如图 2。

图 2 是在绝对沉默率 $\rho = 0$ 时的仿真。由图可得:(1) 该信息传播约 10 小时后,网络中约 50% 的用户已经获得信息,约 14 小时后,约有 97% 的用户已经获得信息,最终 3 类节点在网络中所占的比例趋于稳定,信息传播过程基本结束。在信息传播过程中,所有节点所占比例在约 14 小时达到稳定,即网络处于稳定状态时所需时间约为 14 小时。(2) 该信息在最初的 8 小时内各节点所占比例的变化速度较慢,在 8-11 小时内快速变化,12 小时后比例变化又减缓。(3) 传播节点所占比例 $i(t)$ 不断增大,沉默节点所占比例 $r(t)$ 一直比较小,且这两类节点所占比例的增长趋势均为先缓慢增长,后快速增长,最后减缓增长(S 型增长),只至稳定状态。

3.2. 微博传播影响因素分析

模型中的参数与微博信息的传播密切相关,本节将分别讨论当 $\rho = 0$ 和 $\rho > 0$ 时,模型的参数对信息传播的影响,以下仿真均以 3.1 中的模型参数初始值为参照。

当 $\rho = 0$ 时,考虑模型中三个参数对信息传播的影响。

只改变认证用户的每小时传播率 λ_1 ,其他参数保持不变,即对以下两组数据进行仿真:

(1) $\lambda_1 = 10, \lambda_2 = 5, \mu = 99.9984\%$; (2) $\lambda_1 = 1000, \lambda_2 = 5, \mu = 99.9984\%$,结果如图 3。

结合图 2 和图 3 分析发现:当认证用户传播节点的传播率分别为 10、100、1000 时,微博信息传播的稳定时间约为 15、14、12 小时。说明认证用户的传播率越高,微博信息传播达到稳定状态的时间越短。

只改变普通用户(普通传播节点)的每小时传播率 λ_2 ,其他参数保持不变,即对以下两组数据进行仿真:(1) $\lambda_1 = 100, \lambda_2 = 4, \mu = 99.9984\%$; (2) $\lambda_1 = 100, \lambda_2 = 10, \mu = 99.9984\%$,结果如图 4。

结合图 2 和图 4 分析发现:当普通传播节点的传播率分别为 4、5、10 时,微博信息传播的稳定时间约为 17、14、7。说明普通传播节点的传播率越高,微博信息传播达到稳定状态的时间越短。

只改变停止传播率 μ ,其他参数保持不变,若信息的转发率分别为 0.0001% 和 1.7500%,则对应的

Table 1. Microblog information statistics
表 1. 微博信息统计

被关注数	转发数	评论数	点赞数	总曝光量	真实点赞数	参与用户数
37,270,023	14,164	34,254	22,450	47,753,404	10,620	20,804

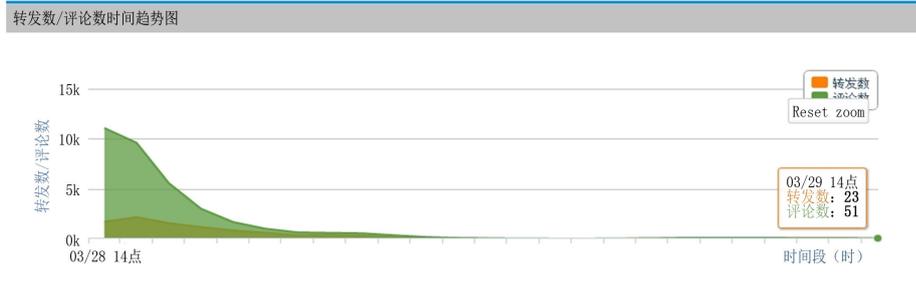


Figure 1. Diagram about forwarding number and comments number
图 1. 转发数和评论数随时间变化关系图

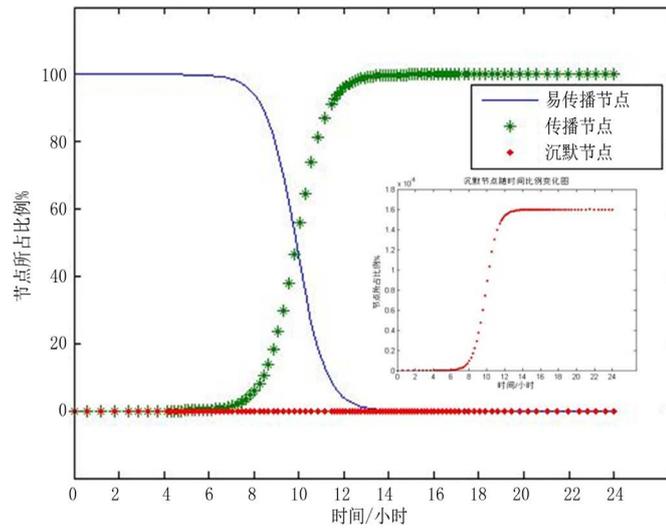


Figure 2. Proportion of three types of nodes on microblog
图 2. 微博中三类节点所占比例变化图

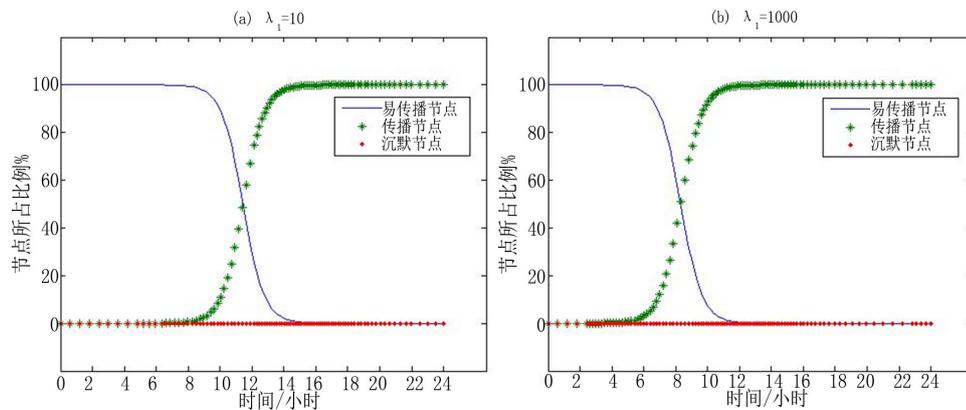


Figure 3. Impact of transmission rates of authenticated users on nodes proportion
图 3. 认证用户的传播率对节点比例的影响

停止传播率 μ 分别为 99.9999% 和 98.2500%。即对以下两组数据进行仿真：

- (1) $\lambda_1 = 100, \lambda_2 = 5, \mu = 99.9999\%$; (2) $\lambda_1 = 100, \lambda_2 = 5, \mu = 98.2500\%$, 结果如图 5。

结合图 2 和图 5 分析发现：当停止传播率 μ 分别为 99.9999%、99.9984% 和 98.2500% 时，微博信息传播的稳定时间约为 15、14、13。说明停止传播率越小，微博信息传播达到稳定状态的时间越短。

当 $\rho > 0$ 时，随着 ρ 的变化，网络中三类节点比例也随之发生变化，如图 6 所示。

结合以上仿真，当绝对沉默率 $\rho > 0$ 时，微博传播除了具有上面讨论的性质外，又有自己的特点：(1) 随着 ρ 的增大，传播节点 i 所占比例逐渐减少，而沉默节点 r 所占比例逐渐增加，且减少(增加)是依曲线 S 型减少(增加)的；(2) 在此数据模型中，当 $n \approx 9.1$ (约为认证用户传播节点的 9.1%)时，传播节点和沉默节点处于平衡状态；(3) 随着 ρ 的增大，信息传播的稳定时间逐渐缩短，因此提高绝对沉默率，信息传

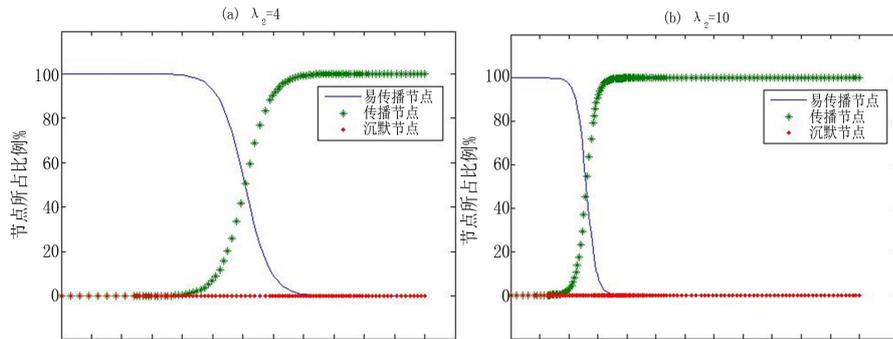


Figure 4. Impact of transmission rates of common nodes on nodes proportion
图 4. 普通节点的传播率对节点比例的影响

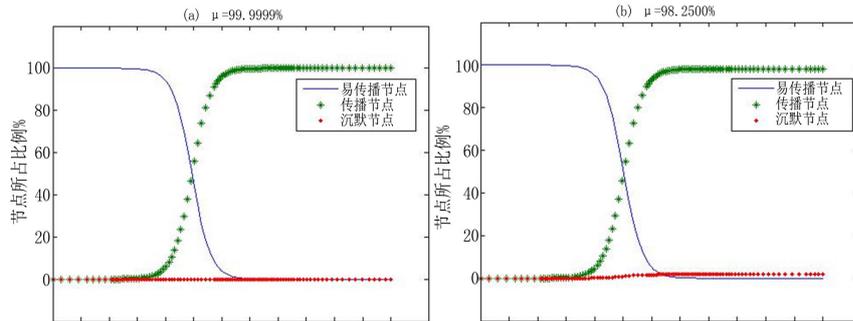


Figure 5. Impact of stop transmission rates on nodes proportion
图 5. 停止传播率对节点比例的影响

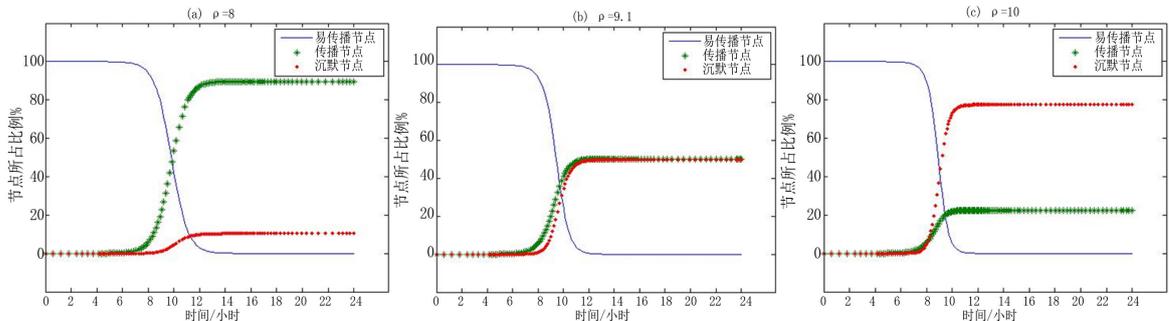


Figure 6. Impact of absolute silence rates on nodes proportion
图 6. 绝对沉默率对节点比例的影响

播达到稳定状态越快；(4) 当网络处于稳定状态时，主要存在的节点为沉默节点，易传播节点和传播节点基本趋于零。

3.3. 结果分析

通过以上分析，在微博信息传播过程中，提高认证用户和普通用户的传播率，降低停止传播率或者提高绝对沉默率，都能够使信息传播达到稳定状态的时间越短。如果考虑认证用户传播节点而忽略绝对沉默节点的影响，网络中主要变化的节点是易传播节点和传播节点，随着时间的推移，易传播节点会减少而传播节点会增加，且均成 S 型变化；当两者都考虑时，网络中易传播节点所占比例基本不变，且随着绝对沉默节点所占比例的增加，沉默节点总数会逐渐增加，而传播节点逐渐减少；当达到平衡状态时，网络中基本上只剩下沉默节点，这也符合现实中一条微博信息从发出到传播，再到消逝的传播过程。

4. 总结

本文建立了含认证用户和绝对沉默节点的微博信息传播网络模型，并仿真分析了模型各参数对信息传播的影响，以分析实际微博传播的影响因素，对现实中控制微博信息的传播有一定的指导意义。同时考虑认证用户和绝对沉默节点时，随着信息传播网络中易传播节点和传播节点趋于零，这更符合实际中的信息传播过程。本文的结果是在一定初始条件和参数下得到的，但实际微博网络会受到主客观条件等因素的影响，模型参数可能会发生变化。微博信息的传播存在一定的时效性，以至在通常情况下，信息很难传播到网络中的所有用户，但这并不影响模型仿真结果的变化趋势和规律。

基金项目

湖北省科技厅自然科学面上项目(2013CFB216)，三峡大学研究生科研创新基金(2014)。

参考文献 (References)

- [1] 汪小凡, 李翔, 陈关荣 (2006) 复杂网络理论及其应用. 清华大学出版社, 北京.
- [2] Watts, D.J. and Strogatz, S.H. (1998) Collective dynamics of “small world” networks. *Nature*, **393**, 440-442.
- [3] Barabasi, A.L. and Albert, R. (1999) Emergence of scaling in random networks. *Science*, **286**, 509-512.
- [4] Watts, D.J. (2004) The “new” science of networks. *Annual Review of Sociology*, **30**, 243-270.
- [5] 殷俊, 何芳 (2011) 微博在我国的传播现状及传播特征分析. *河南大学学报(社会科学版)*, **3**, 124-129.
- [6] 平亮, 宗利永 (2010) 基于社会网络中心性分析的微博信息传播研究——以 Sina 微博为例. *图书情报工作*, **6**, 92-97.
- [7] 何静, 郭进利, 徐雪娟 (2013) 微博关系网络模型研究. *计算机工程*, **11**, 105-108.
- [8] Moreno, Y., Pastor-Satorras, R., et al. (2002) Epidemic outbreaks in complex heterogeneous networks. *European Physical Journal B*, **26**, 521-529.
- [9] Zhou, J., Liu, Z.H. and Li, B.E. (2007) Influence of network structure on rumor propagation. *Physics Letters A*, **368**, 458-463.
- [10] 王林, 戴冠中 (2009) 复杂网络的 Scale-free 性、Scale-free 现象及其控制. 科学出版社, 北京, 253-267.
- [11] 钱颖, 张楠, 赵来军, 等 (2012) 微博舆情传播规律研究. *情报学报*, **12**, 1299-1304.
- [12] 李可嘉, 王义康 (2013) 改进 SIR 模型在社交网络信息传播中的应用. *电子科技*, **8**, 168-171.
- [13] 刘丹, 殷亚文, 宋明 (2014) 基于 SIR 模型的微博信息扩散规律仿真分析. *北京邮电大学学报*, **3**, 28-33.