

## 在线社交网络中信息传播模式的特征分析

韩佳<sup>1</sup>, 肖如良<sup>1\*</sup>, 胡耀<sup>1</sup>, 唐涛<sup>1</sup>, 房丽娜<sup>2</sup>

(1. 福建师范大学 软件学院, 福州 350108; 2. 深圳信息职业技术学院 软件学院, 广东 深圳 518172)

(\*通信作者电子邮箱 xiaoruliang@163.com)

**摘要:**在线社交网络以其独特的传播优势,已成为一种流行的社交媒体平台。针对在线社交网络中信息传播模式的形式特点,结合传染病动力学原理,提出了在线社交网络中的信息传播模型。模型考虑了不同用户行为对传播机理的影响,并建立了不同用户节点的演化方程组,模拟了信息传播的过程,分析了不同类型的用户在网络中的行为特征以及影响信息传播的主要因素。实验结果表明:不同类型的用户在信息传播过程中有着特定的行为规律,信息不会无限制地传播,并在最终达到平稳状态,并且传播系数和免疫系数越大,信息传播达到稳态的速度就越快。

**关键词:**在线社交网络;信息传播;传播模式;特征分析

**中图分类号:** TP311.1 **文献标志码:** A

### Characteristic analysis of information propagation pattern in online social network

HAN Jia<sup>1</sup>, XIAO Ruliang<sup>1\*</sup>, HU Yao<sup>1</sup>, TANG Tao<sup>1</sup>, FANG Lina<sup>2</sup>

(1. Faculty of Software, Fujian Normal University, Fuzhou Fujian 350108, China;

2. Department of Software, Shenzhen Institute of Information Technology, Shenzhen Guangdong 518172, China)

**Abstract:** Because of its unique advantage of information propagation, the online social network has been a popular social communication platform. In view of the characteristics of the form of information propagation and the dynamics theory of infectious diseases, this paper put forward the model of information propagation through online social network. The model considered the influence of different users' behaviors on the transmission mechanism, set up the evolution equations of different user nodes, simulated the process of information propagation, and analyzed the behavior characteristics of the different types of users and main factors that influenced the information propagation. The experimental results show that different types of users have special behavior rules in the process of information propagation, i.e., information cannot be transported endlessly, and be reached at a stationary state, and the larger the spread coefficient or immune coefficient is, the faster it reached the stationary state.

**Key words:** online social network; information propagation; transmission pattern; characteristic analysis

## 0 引言

近年来社交网络得到了迅猛的发展,人们对信息的获取和消息的传递开始变得越来越依赖于网络,如国外的 Facebook<sup>[1]</sup>、MySpace<sup>[2]</sup>等,国内的人人网<sup>[3]</sup>、开心网<sup>[4]</sup>等。用户可以通过分享新闻、日志、视频、音乐及相片等方法,维持和开拓人际关系。作为一种流行的社交媒体平台,社交网络已经成为在线社交网络,正以前所未有的传播优势和巨大影响力,成为拥有强大社会影响力和舆论动员力的重要传播工具。

在线社交网络是一个复杂的虚拟社会,人们在现实生活中的行为和他们在社交网络中的行为有很大程度的相似性,分析社交网络中的用户行为和社交网络的结构性质等已经成为学者们研究的热点问题。例如,文献[5-6]研究了在线社交网络的网络拓补结构的演化特性以及度分布、聚类系数、簇系数、顶点的度相关性等社交网络的相关属性;文献[7-9]从用户点击网站、用户群以及影响最大化等角度,对信息产生和扩散中的用户行为进行了研究;文献[10]从网络安全和隐私保护的角度,对几个大型的社交网络进行了分析;Mislove等<sup>[11]</sup>通过收集 Flickr 网络的数据生长来研究在线社交网络中用户的行为模式,结果发现社交网络中信息传播的新路径

是由度数大的节点创建的。总之,尽管在线社交网络问题已经有了大量的研究成果,但仍需要进一步的研究来揭示以人为主体的在线社交网络中信息的传播机制。

在线社交网络与传统的 Web 网络最大的不同之处在于:传统的 Web 网络的主体是内容信息,依靠内容信息组织在一起,呈现给用户;而社交网络的主体是人,依靠人与人之间的朋友关系组织在一起。在线社交网络必须具备 3 项基本功能,即允许用户:1) 创建和维护朋友关系;2) 上传自己预分享的内容信息;3) 浏览其他用户分享的内容信息并能进行评论、分享、转发等操作。基于这些用户所能拥有的功能,人与人之间的关系开始向“虚拟化”转变,而这种“虚拟化”的关系又对人与人之间的关系产生着一定的影响。如何去描述在线社交网络中的信息传播模式,揭示其传播特性,对以网络为主体的社会来说,具有重要的理论和应用价值。以前人们大都是在规则网络的基础上研究信息的传播方式,但是鉴于在线社交网络所具有的网状式的结构特征,在线社交网络中信息的传播行为与规则网络相比有着本质上的区别。本文针对该问题,利用传染病动力学模型,对在线社交网络中信息传播模式进行了理论建模和仿真分析,有助于更深刻地理解在线社交网络中的信息传播模式及其特征,也对研究网络舆论传播

收稿日期:2012-07-04;修回日期:2012-08-09。

基金项目:教育部规划基金资助项目(11YJA860028);福建省科技计划重大项目(2011H6006)。

作者简介:韩佳(1987-),女,湖北天门人,硕士研究生,主要研究方向:社交网络、软件工程;肖如良(1966-),男,湖南娄底人,教授,博士,CCF 高级会员,主要研究方向:社交网络、虚拟化、软件工程;胡耀(1987-),男,湖南永州人,硕士研究生,主要研究方向:社交网络、虚拟化。

等有一定的价值。

## 1 在线社交网络的信息传播特征

相比于传统的 Web 网,在线社交网络不仅提供信息供用户使用,而且允许用户彼此之间进行交流,是提供信息共享的平台,并且信息更新速度快,数据量大。在线社交网络中,一个用户发出的信息会在其朋友的个人界面展示,其好友可能会以一定的概率传播该信息,也可能只是纯粹地浏览信息。也就是说<sup>[12]</sup>,若该用户的好友对该信息感兴趣则可能传播信息;若其好友对该信息内容不感兴趣,则不会传播该信息而成为免疫者。

举个例子来说,在某个时刻用户 A 在自己好友动态中看到了用户 B 的动态更新,那么用户 A 可能对其进行分享或转发,这一行为就会出现在用户 A 的朋友的好友动态中,用户 A 的朋友就能见到这些消息;若是用户 A 对用户 B 发出的信息不感兴趣而不作出任何回应,那么用户 A 的朋友就见不到这条消息,那么通过用户 A 的消息链就可能被迫中断。接下来,用户 A 的朋友,用户 A 的朋友的朋友都可能具有类似于用户 A 的操作。此外,用户 B 可能不止有用户 A 一个朋友,其他的朋友也会有类似 A 的行为,依此类推,信息就以网状的形式像四周扩散开来。信息传播模式如图 1 所示。

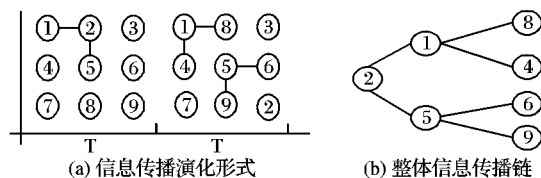


图1 信息传播模式

当一个用户发出消息后,与它相连的所有朋友用户都有可能以一定的概率分享转发消息,或浏览信息而不进行操作。图1中,若用户2与用户1和5之间有链接,则2发出的信息I则会以一定的概率传递到1和5;而用户1又与用户4和用户8之间有链接,用户5又与用户6和用户9之间有链接,则信息I将有可能传递到用户4,6,8,9,而这些用户又会有他们自己的朋友关系网,信息会通过这些朋友关系传递下去。

图1是在线社交网络中信息传播模式的简单描述。就某一特定的消息来说,它是以信息源为中心向四周散射传播开去的,信息的走向具有不规则性,只有对该信息感兴趣的用户对信息进行分享、转发等,信息才会沿着这条信息链继续传播;否则,通过这个用户的信息链将中断。

## 2 传染病传播的 SIR 模型分析

在流行病传播的研究中,大多是基于 SIR (Susceptible-Infectious-Removed) 模型<sup>[13]</sup>。该模型将总人口分为以下三类:易感者 (susceptibles), 其数量记为  $S(t)$ , 表示  $t$  时刻未染病但有可能被该类疾病传染的人数;染病者 (infectives), 其数量记为  $I(t)$ , 表示  $t$  时刻已被感染成为病人而且具有传染力的人数;恢复者 (recovered), 其数量记为  $R(t)$ , 表示  $t$  时刻已从染病者中移出的人数。设总人口为  $N(t)$ , 则有  $N(t) = S(t) + I(t) + R(t)$ 。

SIR 模型的建立基于以下 3 个假设。

1) 不考虑人口的出生、死亡、流动等种群动力因素。人口始终保持一个常数,即  $N(t)$ 。

2) 一个病人一旦与易感者接触就必然具有一定的传染力。假设  $t$  时刻单位时间内,一个病人能传染的易感者数目与此环境中易感者总数  $S(t)$  成正比,比例系数为  $\beta$ ,从而在  $t$  时刻单位时间内被所有病人传染的人数为  $\beta S(t)I(t)$ 。

3)  $t$  时刻,单位时间内从染病者中移出的人数与病人数量成正比,比例系数为  $\gamma$ ,单位时间内移出者的数量为  $\gamma I(t)$ 。因此有式(1):

$$\begin{cases} \frac{dI}{dt} = \beta SI - \gamma I \\ \frac{dS}{dt} = -\beta SI \\ \frac{dR}{dt} = \gamma I \end{cases} \quad (1)$$

通过第1章的分析可知,社交网络中的信息传播是非线性的动力学系统的一个例子,信息的这种传播行为和人群中的流行病十分相似,因此,可以借助生物流行病领域已有的数学方法来对在线社交网络中的信息传播进行研究。

但用来描述在线社交网络中的信息传播,SIR 模型则存在着不足之处。SIR 模型认为某封闭群体中的 3 种状态 SIR 随时间  $t$  的变化之和其自身及两个常量  $\beta, \gamma$  有关,而没有考虑到其他因素的影响。在现实生活中,当流行病出现的时候人们往往会采取一些措施如对居住的环境进行消毒或进行免疫接种等,使自身从易感者向免疫者转变;在社交网络中,用户对信息的兴趣度也可能随着时间的推移或周围环境的影响等发生改变,易感染的用户也有可能直接转变为免疫者。因此,若单纯地使用 SIR 模型,则就成了一种很被动的信息传播方式。所以要研究在线社交网络中的信息传播,必须在 SIR 模型的基础上进行改进。

## 3 基于 SIR 的在线社交网络信息传播模型

本文将在线社交网络中的用户定义为节点,节点之间的边表示用户的朋友关系,信息沿着边进行传播。根据在线社交网络中信息的传播模式,将网络中的节点分为 3 类<sup>[12]</sup>:传播节点(I),未感染节点(S)和免疫节点(R)。传播节点指该节点能接受其邻居节点的信息,并具有传播该信息的能力;未感染节点指该节点没有接受过来自其邻居节点的信息,但有概率被感染;免疫节点指该节点能接受其邻居节点的信息,但不具有传播能力。

在线社交网络中三类节点的态机制,如图2所示。

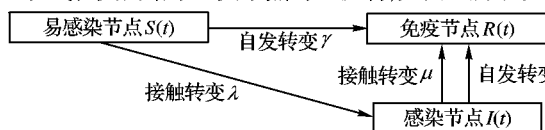


图2 信息传播的状态机制

在线社交网络中信息传播的 3 种状态转变规则。

1) 一个处于易感染状态的节点接触到感染节点时,易感染节点可能以概率  $\lambda$  转变为感染节点,  $\lambda$  为传播系数。

2) 一个处于感染状态的节点与一个免疫节点接触时,传播节点可能以概率  $\mu$  转变为免疫节点,  $\mu$  为免疫系数。

3) 随着时间的推移和周围环境的影响,处于易感染状态的节点将会以概率  $\gamma$  向免疫节点转变,  $\gamma$  为自免疫系数;处于感染状态的节点将会以概率  $\delta$  自动向免疫节点转变,  $\delta$  为自免疫系数。

4) 感染节点的信息传播链会在一定时刻停止,即信息不会无限地传播。

显然,  $t$  时刻,易感染节点的变化率除了与感染节点的传播系数  $\lambda$  有关,还与其自身的免疫系数  $\gamma$  有关;免疫节点的变化率除了与感染节点的免疫系数  $\mu$  有关,还与易感染节点的免疫系数  $\gamma$  和感染节点的自发免疫系数  $\delta$  有关。用  $R_i(t)$  和  $R_e(t)$  表示易感染节点和感染节点能够自发转变为免疫节点的数量,则有式(2):

$$\begin{cases} \frac{dS(t)}{dt} = -\lambda S(t)I(t) - \frac{dR_s(t)}{dt} \\ \frac{dR(t)}{dt} = \frac{dR_s(t)}{dt} + \frac{dR_i(t)}{dt} \end{cases} \quad (2)$$

由以上分析可知,  $R_s(t)$  的变化率与易感染节点自发转变的免疫系数  $\gamma$  成正比;  $R_i(t)$  的变化率与感染节点自发转变的免疫系数  $\delta$  和接触的免疫系数  $\mu$  之和成正比, 因此有式(3):

$$\begin{cases} dR_s(t) = \gamma S(t) \\ dR_i(t) = (\mu + \delta) I(t) \end{cases} \quad (3)$$

又由于  $S(t) + I(t) + R(t) = N$ , 即有式(4):

$$\begin{cases} \frac{dS(t)}{dt} = -\lambda S(t)I(t) - \gamma S(t) \\ \frac{dI(t)}{dt} = \lambda S(t)I(t) - (\mu + \delta) I(t) \\ \frac{dR(t)}{dt} = \gamma S(t) + (\mu + \delta) I(t) \end{cases} \quad (4)$$

其中  $\lambda, \gamma, \mu, \delta$  均为正常数。 $\lambda, \mu$  为一般情况下的传播系数和免疫系数;  $\gamma, \delta$  为在环境影响下易感染节点和感染节点自发转变为免疫节点的免疫系数。

#### 4 实验结果及分析

实验采用的数据是 Epinions.com 的 who-trust-whom 在线社交网络的真实数据集。该数据集的基本数据如下: 节点总数量为  $N = 75\,879$ , 节点之间边的数量为  $508\,837$ , 平均度为  $6.7059$ , 平均聚类系数为  $0.2283$ , 最长的最短路径为  $13$ 。该网络初始度分布如图3所示。

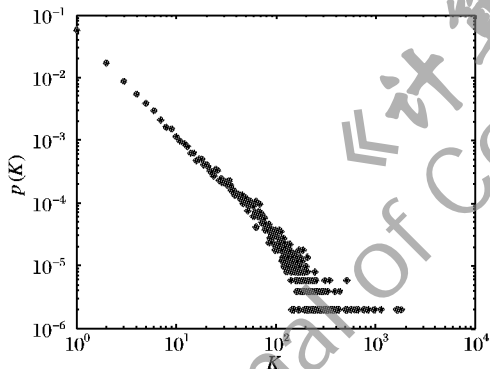


图3 网络初始度分布

设置初始状态网络中只有一个传播节点, 其余节点全部为未感染节点, 即  $I(0) = 1 \ll N, R(0) = 0, S(0) = N - 1$ , 设置模型参数:  $\lambda = 0.3, \mu = 0.1$ , 迭代次数  $T = 300$ 。

随着时间的推移和周围环境的影响, 用户中的易感染节点和感染节点用户能够自动转变为免疫节点用户, 取  $\gamma = \delta = 0.0125$ 。这样即可求得, 网络中传播节点、未感染节点和免疫节点的数量所占节点总数的比例(分别用  $s(t), i(t), r(t)$  表示)随着时间的演化情况, 如图4所示。

从图4中可以看出: 当  $t$  趋向于无穷大时, 信息的传播过程将达到平稳状态, 各类节点的数量所占的比例将不再发生变化。传播节点所占的比例  $i(t)$  在初始阶段快速增长, 当达到最大值后迅速减小, 直到趋向于零; 未感染节点所占的比例  $s(t)$  不断衰减直到趋向于零; 免疫节点所占的比例  $r(t)$  在初期呈现较快的增长趋势, 在一定阶段后趋于平稳直到达到最大值。

在信息传播趋于稳态时, 网络中用户节点最终都将转变为免疫状态 R 态。尽管易感染节点和感染节点用户的自免疫

系数不可忽略, 但它们的变化范围很小, 故在线社交网络中对信息的传播起主要作用的是传播系数  $\lambda$  和免疫系数  $\mu$ 。

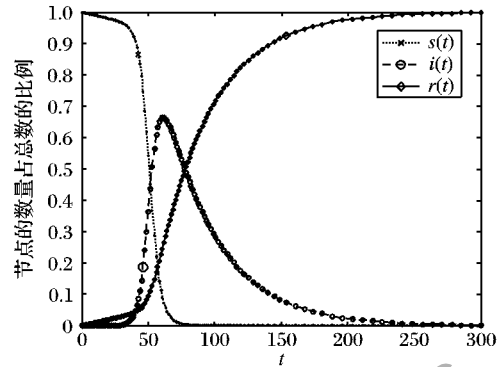


图4 三类节点所占比例随时间的变化关系

改变传播系数  $\lambda$  值, 而其他参数保持不变时, R 态节点所占的比例随时间的变化情况如图5所示。

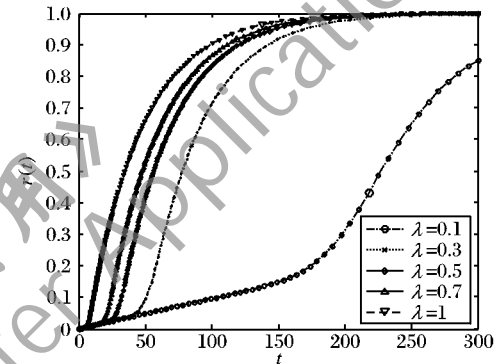


图5  $r(t)$  在  $\lambda$  取不同值的情况下随时间的变化关系

改变免疫系数  $\mu$  的值, 而其他参数保持不变时, R 态节点所占的比例随时间的变化情况如图6所示。

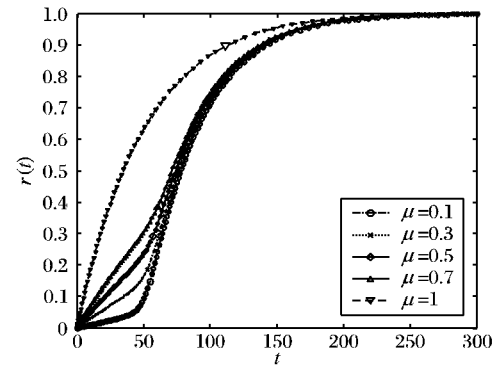


图6  $r(t)$  在  $\mu$  取不同值的情况下随时间的变化关系

从图5~6中可以看出, 当传播系数  $\lambda$  越大,  $r(t)$  增大得越快, 信息传播达到平稳状态的时间就越短; 当传播系数  $\lambda$  很小时, 信息仍能在网络中传播, 但该信息传播达到趋于平稳的时间将会较长, 即信息的传播具有一定的时延性。当免疫系数  $\mu$  逐渐增大时,  $r(t)$  也逐渐增大; 当免疫系数  $\mu$  较小时, 在信息传播的初期, R 态所占的比例增加得较为缓慢, 经过一定阶段将快速增长直至平稳状态。

#### 5 结语

通过分析在线社交网络中信息传播模式的形式特点, 结合传染病动力学原理, 在对 SIR 模型改进的基础上, 提出了在线社交网络中的信息传播模型, 并建立了微分方程组。仿真分析了社交网络中的信息传播过程以及三种不同类型的用户节点在信息传播中的行为规律, 结果表明: 信息不可能无限制

(下转第111页)

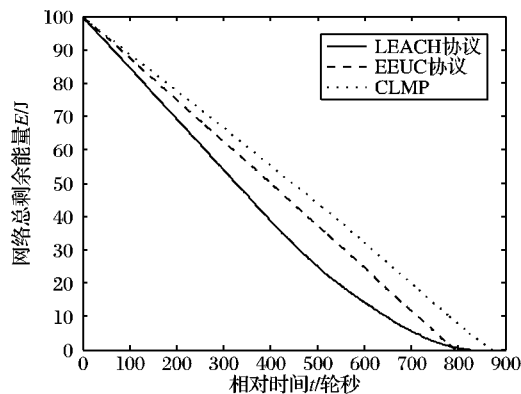


图4 网络总剩余能量图

#### 4 结语

本文针对多跳分簇无线传感网中普遍问题与 EEUC 中的不足进行了深入分析,通过控制成簇大小与数量,建立能量高效的多跳链路,利用将整个网络生命期划分为两部分实现减少成簇和建立多跳链路的次数,弥补了 EEUC 中的不足,有效地降低了网络总能耗。仿真结果表明,与 LEACH 和 EEUC 算法比较,CLMP 算法减少了网络总能耗,使负载更均衡,延长了网络生命期。

#### 参考文献:

- [1] 李利民,李建中,陈渝. 无线传感器网络[M]. 北京:清华大学出版社,2005.
- [2] 沈波,张世永,钟亦平. 无线传感器网络分簇路由协议[J]. 软件学报,2006,17(7):1588-1600.
- [3] 余勇昌,韦岗. 无线传感器网络中基于 PEGASIS 协议的改进算法[J]. 电子学报,2008,36(7):1309-1313.
- [4] MHATRE V, ROSENBERG C. Design guidelines for wireless sensor network: communication, clustering and aggregation [J]. Ad Hoc Networks, 2004, 2(1): 45-63.
- [5] HEINZELMAN W R, CHANDRAKASAN A, BALALCRISHNAN H. Energy efficient communication protocol for wireless microsensor networks [C]// Proceedings of the 33rd Hawaii International Conference on System Science. Washington, DC: IEEE Computer Society, 2000: 300-304.

- [6] MANJESHWAR A, AGRAWAL D. TEEN: a routing protocol for enhanced efficiency in wireless sensor networks [C]// Proceedings of the 15th International Symposium on Parallel and Distributed Processing. Washington, DC: IEEE Computer Society, 2001: 2001-2009.
- [7] MANJESHWAR A, AGRAWAL D. APTEEN: a hybrid protocol for efficient routing and comprehensive information retrieval in wireless sensor networks [C]// Proceedings of the 16th International Parallel and Distributed Processing Symposium. Washington, DC: IEEE Computer Society, 2002: 195-202.
- [8] HE T, STANKOVIC J A, LU C Y, *et al.* SPEED: a stateless protocol for real time communication in sensor networks [C]// Proceedings of the 23rd International Conference on Distributed Computing System. Washington, DC: IEEE Computer Society, 2003: 46-55.
- [9] LINDSEY S, RAGHAVENDRA C S. PEGASIS: power efficient gathering in sensor information systems [C]// Proceedings of the 2002 Aerospace Conference. Washington, DC: IEEE Computer Society, 2002: 3-1125-3-1130.
- [10] YOUNIS O, FAHMY S. HEED: a hybrid, energy-efficient, distributed clustering approach for Ad Hoc sensor networks [J]. IEEE Transactions on Mobile Computing, 2004, 3(4): 366-379.
- [11] LOSCRI V, MORABITO G, MARANO S. A two level hierarchy for low energy adaptive clustering hierarchy (TL-LEACH) [C]// Proceedings of the 62nd IEEE Vehicular Technology Conference. Washington, DC: IEEE Computer Society, 2005: 1809-1813.
- [12] SORO S, HEINZELMAN W B. Prolonging the lifetime of wireless sensor networks via unequal clustering [C]// Proceedings of the 19th International Parallel and Distributed Processing Symposium. Washington, DC: IEEE Computer Society, 2005: 8-15.
- [13] YE M, LI C, CHEN G, *et al.* EECS: an energy efficient clustering scheme in wireless sensor networks [C]// Proceedings of the 24th IEEE International Conference on Performance, Computing and Communications. Piscataway: IEEE Press, 2005: 535-540.
- [14] 李法成,陈贵海,叶懋,等. 一种基于非均匀分簇的无线传感器网络路由协议[J]. 计算机学报,2007,30(1):27-36.
- [15] 施叶玲,陈彬兵. 无线传感器网络改进的 LEACH-ID 算法[J]. 计算机应用,2011,31(2):324-327.

(上接第107页)

地传播,当  $t$  趋向于无穷大时,信息的传播过程将达到平稳状态,各类节点的数量将不再变化;当传播系数较小时,信息仍能在网络中传播,并且传播系数和免疫系数的大小发生变化时,信息传播的快慢程度将发生相应的改变。本文有助于更深刻地理解在线社交网络中的信息传播行为和合理地引导网络舆论。下一阶段的任务将是研究动态网络中,具有不同度的用户节点对信息传播产生的影响。

#### 参考文献:

- [1] Facebook[EB/OL]. [2012-03-12]. <http://www.facebook.com/>.
- [2] MySpace[EB/OL]. [2012-03-12]. <http://www.myspace.com/>.
- [3] 人人网[EB/OL]. [2012-03-25]. <http://www.renren.com/>.
- [4] 开心网[EB/OL]. [2012-03-25]. <http://www.kaixin001.com/>.
- [5] KUMAR R, NOVAK J, TOMKINS A. Structure and evolution of online social networks [C]// Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM Press, 2006: 611-617.
- [6] MISLOVE A, MARCON M, GUMMAD K P, *et al.* Measurement and analysis of online social networks [C]// IMC'07: Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement. New

- York: ACM Press, 2007: 29-42.
- [7] BENEVENUTO F, RODRIGUES T, CHA M, *et al.* Characterizing user behavior in online social networks [C]// IMC'09: Proceedings of the 9th ACM SIGCOMM Conference on Internet Measurement Conference. New York: ACM Press, 2009: 49-62.
- [8] CHEN W, WANG Y J, YANG S Y. Efficient influence maximization in social networks [C]// KDD'09: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM Press, 2009: 199-208.
- [9] SHI X L, ZHU J, CAI R, *et al.* User grouping behavior in online forums [C]// KDD'09: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM Press, 2009: 777-786.
- [10] 刘建伟,李为宇,孙钰. 社交网络安全问题及其解决方案[J]. 中国科学技术大学学报,2011,41(7):565-574.
- [11] MISLOVE A, KOPPULA H S, GUMMADI K P, *et al.* Growth of the flickr social network [C]// Proceedings of the 1st ACM SIGCOMM Workshop on Social Networks. New York: ACM Press, 2008: 25-30.
- [12] 张彦超,刘云,张海峰,等. 基于在线社交网络的信息传播模型[J]. 物理学报,2011,60(5):60-66.
- [13] FRAUENTHAL J C. Mathematical modeling in epidemiology [M]. Berlin: Springer, 1980.