

基于在线社交网络的信息传播模型^{*}

张彦超[†] 刘 云 张海峰 程 辉 熊 菲

(北京交通大学通信与信息系统北京市重点实验室, 北京 100044)

(2010 年 7 月 23 日收到; 2010 年 9 月 8 日收到修改稿)

本文构造了一个基于在线社交网络的信息传播模型. 该模型考虑了节点度和传播机理的影响, 结合复杂网络和传染病动力学理论, 进一步建立了动力学演化方程组. 该方程组刻画了不同类型节点随着时间的演化关系, 反映了传播动力学过程受到网络拓扑结构和传播机理的影响. 本文模拟了在线社交网络中的信息传播过程, 并分析了不同类型节点在网络中的行为规律. 仿真结果表明: 由于在线社交网络的高度连通性, 信息在网络中传播的门槛几乎为零; 初始传播节点的度越大, 信息越容易在网络中迅速传播; 中心节点具有较大的社会影响力; 具有不同度数的节点在网络中的变化趋势大体相同. 研究表明该模型符合在线社交网络的特性, 有助于更深刻地理解发生在在线社交网络中的传播行为.

关键词: 在线社交网络, 信息传播, 微分方程, 传染病动力学

PACS: 05.10.-a

1. 引言

社会性网络服务 (social network service, SNS)^[1]是 Web2.0 体系下的一个技术应用架构. SNS 通过网络聊天、博客、播客和社区共享等途径, 实现个体社交圈的逐步扩大, 最终形成一个联结“熟人的熟人”的大型网络社交圈, 充分反映出人类社会的六度分离特征. 在国外, 首屈一指的 SNS 网络——Facebook^[2]拥有超过千万的用户, 成为全世界最大的社交网站. 在国内, 人人网^[3]、开心网^[4]等提供 SNS 服务的网站的发展也势如破竹. 在线社交网络作为一个虚拟社会, 人在 SNS 网络上的行为与现实生活的行为具有共性. 分析网络用户的行为规律^[5], 例如人在 SNS 网络上的个体流行程度和活跃程度^[6]等已经成为了研究热点. 同时, 在线社交网络作为复杂网络的一种, 研究其网络结构性质同样具有重要的意义. 文献 [1, 7, 8] 研究了在线社交网络的网络拓扑结构的演化特性, 包括度分布、聚类系数、群聚系数、顶点度相关系数等网络属性; 文献 [9, 10] 将几个大型的 SNS 网络进行对比. 但是, 目前的研究还局限于定性或者半定量的统计性质分

析, 缺少理论以及背后机理分析.

在线社交网络与传统的 Web 网络在信息传播方式上存在着根本的不同: 传统的 Web 网络是以信息内容为主体进行传播; 在线社交网络是以人为中心, 依靠人与人之间的好友关系进行信息的传播. 如今, 人们在获取信息的时候更加在乎信息的来源. 这种获取信息的方式将关注的重心放在人们会从谁那里获取信息, 又会与谁分享信息. 传统的传播行为, 如计算机病毒在计算机网络上的蔓延^[11]、传染病在人群中的流行^[12]、谣言在社会中的扩散^[13, 14]、信息在虚拟社区和博客中的传播^[15, 16], 都可以看作是服从某种规律的传播行为. SNS 网络打破了传统传播方法, 它利用人与人的关系改变人与信息的关系, 反过来又用人与信息的关系影响人与人的关系. 如何去描述在线社交网络中的传播行为, 揭示它的特性, 具有重要的理论和应用价值, 是我们关注的焦点.

以前的传播模型大都是基于规则网络的, 但是在在线社交网络上的传播行为与规则网络相比确实存在根本上的不同. 本文针对该问题, 利用传染病动力学和复杂网络理论, 对在线社交网络中的信息传播行为进行详细的理论建模和数值仿真研究, 有

^{*} 国家自然科学基金 (批准号: 60972012), 北京市自然科学基金 (批准号: 4102047), 科技人员服务企业项目 (批准号: 2009GJA00048), 教育部哲学人文社会科学研究重大课题 (批准号: 08WL1101) 和北京市教育委员会学科建设与研究生建设项目资助的课题.

[†] E-mail: 07111026@bjtu.edu.cn

助于更深刻地理解发生在社交网络中的传播行为,是进一步研究网络舆论传播的基础.下面首先介绍本文提出的模型,然后介绍数值模拟结果,最后给出结论.

2. 模型

2.1. 传播机理

在 SNS 网络中,一个人发布的消息会被其好友看到,并以一定的概率分享、传播.同时,若其好友对其内容不感兴趣则成为“免疫者”且不会传播.本文把 SNS 网络上的用户定义为节点,个体之间的好友关系则可以抽象地用节点之间的边来表示,信息只沿着边传播.

根据信息在 SNS 网络中的传播规律,我们把网络中的节点分为三类:传播节点、未感染节点、免疫节点.传播节点表示该节点接受了来自其邻居节点的信息,并具有传播该信息的能力.未感染节点表示该节点没有接受过来自其邻居节点的信息,并有机会接受信息,即有概率被感染.免疫节点表示该节点已经接受了其邻居节点的信息,但是不具有传播能力.节点在传播状态、未感染状态和免疫状态之间的转移不仅依赖于节点自身的状态,还与它的邻居节点的状态相关.定义以下传播规则:

- 1) 如果一个传播节点与一个未感染节点接触,则未感染节点会以概率 p_1 成为传播节点.
- 2) 如果一个传播节点与一个免疫节点接触,则传播节点会以概率 p_2 成为免疫节点.
- 3) 传播节点不会无休止地传播下去,会以一定的速度 v 停止传播而变为免疫节点,且无需与其他节点接触.

2.2. 基于 SNS 网络的信息传播模型

假设一个节点 j 在 t 时刻处于未感染状态 p_{ii}^j 表示节点 j 在 $[t, t + \Delta t]$ 时段内处于未感染状态的概率 p_{is}^j 表示节点 j 从未感染状态转移到传播状态的概率且 $p_{is}^j = 1 - p_{ii}^j$. 其中

$$p_{ii}^j = (1 - \Delta t p_1)^g, \quad (1)$$

式中 $g = g(t)$ 表示在 t 时刻节点 j 的邻居中传播节点的数量.

假设节点 j 含有 k 条边 g 是具有如下二项分布的随机变量:

$$\Pi(g, t) = \binom{k}{g} \omega(k, t)^g (1 - \omega(k, t))^{k-g}, \quad (2)$$

式中 $\omega(k, t)$ 表示在 t 时刻从具有 k 条边的未感染节点连接到一个传播节点的概率. $\omega(k, t)$ 可写为如下形式:

$$\begin{aligned} \omega(k, t) &= \sum_{k'} p(k' | k) p(s_{k'} | i_k) \\ &\approx \sum_{k'} p(k' | k) \rho^s(k', t), \end{aligned} \quad (3)$$

式中 $p(k' | k)$ 为度相关函数,表示度为 k 的节点与度为 k' 的节点邻接的条件概率; $p(s_{k'} | i_k)$ 表示一个拥有 k' 条边的节点在它连接到一个度为 k 的未感染节点条件下,处于传播状态的概率; $\rho^s(k', t)$ 表示在 t 时刻度为 k' 的传播节点的密度.

所以,得到度为 k 的节点在 $[t, t + \Delta t]$ 时段内处于未感染状态的平均转移概率 $\bar{p}_{ii}(k, t)$ 如下:

$$\begin{aligned} \bar{p}_{ii}(k, t) &= \sum_{g=0}^k \binom{k}{g} (1 - p_1 \Delta t)^g \omega(k, t)^g \\ &\quad \times (1 - \omega(k, t))^{k-g} \\ &= (1 - p_1 \Delta t \omega(k, t))^k. \end{aligned}$$

将(3)式代入上式得到

$$\bar{p}_{ii}(k, t) = (1 - p_1 \Delta t \sum_{k'} p(k' | k) \rho^s(k', t))^k. \quad (4)$$

同理,假设 p_{sr}^j 表示节点 j 在 $[t, t + \Delta t]$ 时段内从传播状态转移到免疫状态的概率, p_{ss}^j 表示节点 j 保持传播状态的概率且 $p_{ss}^j = (1 - \Delta t p_2)^g (1 - v \Delta t)$.

于是,得到度为 k 的节点在 $[t, t + \Delta t]$ 时段内处于感染状态的平均转移概率 $\bar{p}_{ss}(k, t)$ 如下:

$$\begin{aligned} \bar{p}_{ss}(k, t) &= (1 - \Delta t p_2 \sum_{k'} p(k' | k) \rho^r(k', t))^k \\ &\quad \times (1 - v \Delta t). \end{aligned} \quad (5)$$

则节点从传播状态转移到免疫状态的转移概率为 $\bar{p}_{sr}(k, t) = 1 - \bar{p}_{ss}(k, t)$.

假设 $N(k, t)$ 为在 t 时刻网络中度为 k 的节点总数量 $I(k, t)$, $S(k, t)$, $R(k, t)$ 分别为在 t 时刻网络中度为 k 的未感染节点、传播节点和免疫节点的数量,则

$$I(k, t) + S(k, t) + R(k, t) = N(k, t). \quad (6)$$

于是,网络中度为 k 的未感染节点的数量在 $[t, t + \Delta t]$ 时段内的变化情况如下:

$$\begin{aligned} I(k, t + \Delta t) &= I(k, t) - I(k, t) (1 - \bar{p}_{ii}(k, t)) \\ &= I(k, t) - I(k, t) \\ &\quad \times [1 - (1 - p_1 \Delta t \sum_{k'} \rho^s(k', t) p(k' | k))]^k. \end{aligned} \quad (7)$$

同理,可相应得到度为 k 的传播节点和免疫节点的数量在 $[t, t + \Delta t]$ 时段内的变化情况,如下

$$\begin{aligned} S(k, t + \Delta t) &= S(k, t) + I(k, t)(1 - \bar{p}_{ii}(k, t)) \\ &\quad - S(k, t)(1 - \bar{p}_{ss}(k, t)) \\ &= S(k, t) + I(k, t) \\ &\quad \times [1 - (1 - p_1 \Delta t \\ &\quad \times \sum_{k'} \rho^s(k', t) p(k' | k))]^k \\ &\quad - S(k, t) [1 - (1 - p_2 \Delta t \\ &\quad \times \sum_{k'} \rho^r(k', t) p(k' | k))]^k (1 - v \Delta t), \end{aligned} \quad (8)$$

$$\begin{aligned} R(k, t + \Delta t) &= R(k, t) + S(k, t)(1 - \bar{p}_{ss}(k, t)) \\ &= S(k, t) [1 - (1 - p_2 \Delta t \\ &\quad \times \sum_{k'} \rho^r(k', t) p(k' | k))]^k \\ &\quad \times (1 - v \Delta t) + R(k, t). \end{aligned} \quad (9)$$

由(6),(7)式可以得到

$$\begin{aligned} \frac{I(k, t + \Delta t) - I(k, t)}{N(k, t) \Delta t} &= - \frac{I(k, t)}{N(k, t) \Delta t} [1 - (1 - p_1 \Delta t \\ &\quad \times \sum_{k'} \rho^s(k', t) p(k' | k))]^k. \end{aligned} \quad (10)$$

当 $\Delta t \rightarrow 0$ 时,对(10)式右侧进行泰勒展开得到

$$\frac{\partial \rho^i(k, t)}{\partial t} = -k p_1 \rho^i(k, t) \sum_{k'} \rho^s(k', t) p(k' | k). \quad (11)$$

同理,由(8)式可以得到

$$\begin{aligned} \frac{\partial \rho^s(k, t)}{\partial t} &= k p_1 \rho^i(k, t) \sum_{k'} \rho^s(k', t) p(k' | k) \\ &\quad - k p_2 \rho^s(k, t) \sum_{k'} \rho^r(k', t) p(k' | k) \\ &\quad - v \rho^s(k, t). \end{aligned} \quad (12)$$

由(9)式得到

$$\begin{aligned} \frac{\partial \rho^r(k, t)}{\partial t} &= k p_2 \rho^s(k, t) \sum_{k'} \rho^r(k', t) p(k' | k) \\ &\quad + v \rho^s(k, t). \end{aligned} \quad (13)$$

由(11),(12),(13)式联立得到信息传播的动力学演化方程组,用于刻画传播节点、未感染节点和免疫节点的密度随着时间的变化关系,传播动力学过程同时受到网络拓扑结构和传播机理的影响.

网络中节点的度相关性^[17]被用来描述不同网络结构之间的差异.节点的度相关性,也称为网络选型连接性,指的是网络中与高度数(或低度数)节

点相连接的节点的度数偏向于高还是低.若连接度大的节点趋向于和其他连接度大的节点连接,则认为网络呈现协调混合;若连接度大的节点趋向于和其他连接度小的节点连接,则认为网络非呈现协调混合.

实际的网络的选型连接性有一些呈现协调混合,一些呈现非协调混合.如社会网络(电影演员合作网络^[18]、科学家合作网络^[19])中节点具有正的相关性,其他类型的网络(信息网络、技术网络、生物网络^[20])则相反.

根据文献[21],SNS网络的度相关性一般小于零,例如人人网^[3]的度相关性为 -0.0036 .所以,根据文献[22],本文(3)式中的度相关函数可以写为

$$P(k' | k) = \frac{k' P(k')}{\bar{k}}, \quad (14)$$

式中 $P(k')$ 为度分布函数, \bar{k} 为网络节点平均度.

3. 数值模拟及结果

1) 为验证提出的模型,本文首先生成一个在线社交网络,该网络的基本数据如下:节点总数量为 10000,平均度为 14.85,最大度为 334,聚类系数为 0.0838189,同配系数为 -0.0045869 .网络初始度分布如图 1 所示.

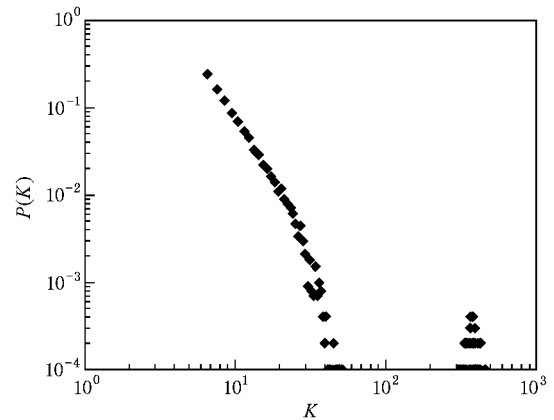


图 1 网络初始度分布图

初始状态设置网络中只有一个传播节点,其余节点全部为未感染节点.其次,设置模型参数如下: $p_1 = 0.3$, $p_2 = 0.1$, $v = 0.05$,迭代次数 $T = 300$.现在研究网络中传播节点、未感染节点和免疫节点的密度随着时间的演化情况,如图 2 所示.

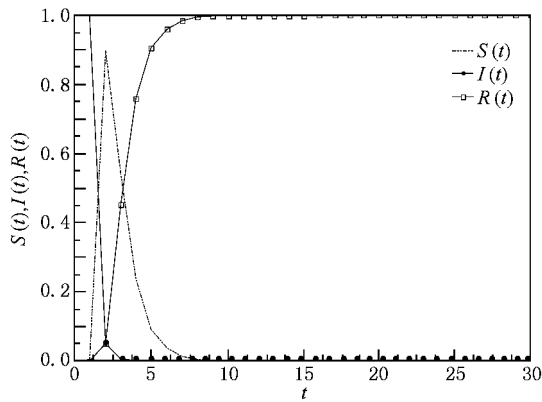
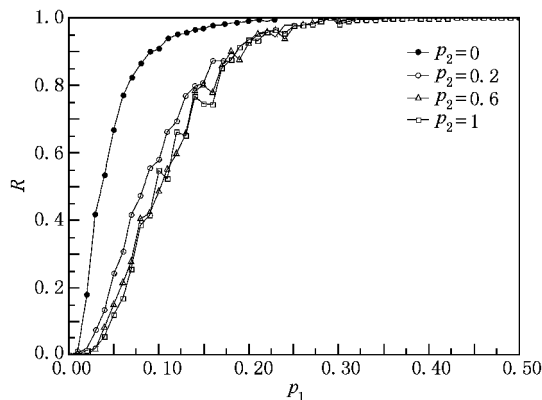


图2 不同节点密度随着时间的变化关系

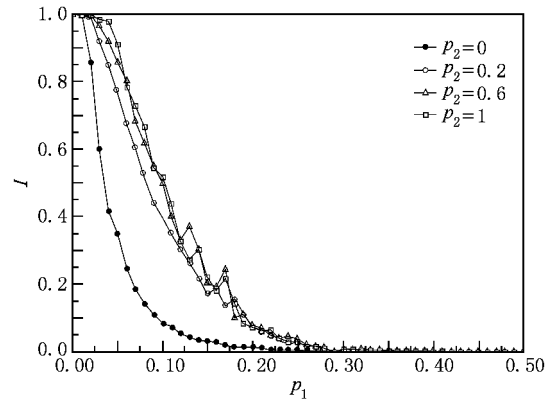
从图2可以看出:免疫节点密度 $R(t)$ 初期呈现出较快的上升趋势,在一定阶段后趋于平稳,直到趋向于1;传播节点密度 $S(t)$ 在初始阶段快速增长,当达到最大值后迅速减小,直到趋向于零;未感染节点密度 $I(t)$ 不断衰减直到趋向于零.

2) 定义 $R = \sum_k \rho^r(k, t_\infty)$, R 表示免疫节点密度的最终大小. 当 t 趋向于无穷大时,传播过程已经达到平衡状态,各类节点的数量、密度不会再发生变化. 我们重点研究模型中各参数对 R 的影响. 图3给出在 p_2 取不同的值的情况下, R 随着 p_1 的变化关系. 可以看出,在 p_1 很小的情况下,信息仍能在网络中传播,即信息在网络中传播的门槛几乎为零;而基于规则网络的传统的传播行为是在传播概率大于某个阈值的情况下才能传播,这是两者较大的区别,是由网络本身的拓扑结构性决定的. 从图3可以进一步了解 p_1 和 p_2 分别对 R 的影响:在 p_1 相同的情况下, R 随着 p_2 的增大而减小;在 p_2 相同的情

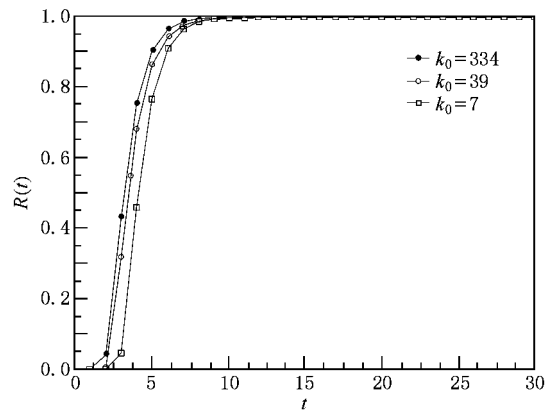
图3 免疫节点最终密度 R 在 p_2 取不同值的情况下随着 p_1 的变化关系

况下 R 随着 p_1 的增大而增大.

同理,定义 $I = \sum_k \rho^i(k, t_\infty)$, I 表示未感染节点密度的最终大小. 图4显示的是 I 在 p_2 取不同值的情况下随着 p_1 的变化关系. 在 p_2 相同的情况下, I 随着 p_1 的增大而减小;在 p_1 相同的情况下, I 随着 p_2 的增大而增大.

图4 未感染节点最终密度 I 在 p_2 取不同值的情况下随着 p_1 的变化关系

3) 研究初始传播节点的度为不同值时对传播行为的影响. 假设初始传播节点的度为 k_0 , 在 $k_0 = 334$, $k_0 = 39$, $k_0 = 7$ 这三种情况下,免疫节点密度和传播节点密度随着时间的变化情况分别如图5和图6所示.

图5 初始传播节点的度分别为 $k_0 = 334$, $k_0 = 39$, $k_0 = 7$ 的情况下免疫节点密度 $R(t)$ 随着时间的变化关系

从图5和图6可以看出,当初始传播节点的度较大时,信息在网络中的传播速度较快. 相反,当初始传播节点的度较小时,信息的传播具有滞后性,即信息从没有传播到开始传播需要一定的时间. 但

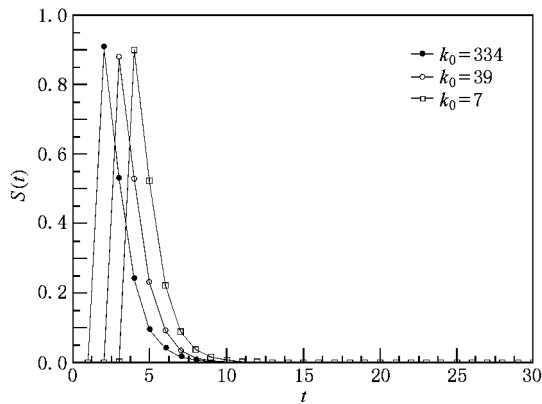


图6 初始传播节点的度分别为 $k_0 = 334$ $k_0 = 39$ $k_0 = 7$ 的情况下传播节点密度 $S(t)$ 随着时间的变化关系

是当时间 t 趋向无穷大时,不论初始传播节点的度为多少,最终网络中免疫节点密度都趋向于 1,这是由于在线社交网络具有较高连通性的缘故。

在 SNS 网络中,高度数的节点,或称为中心节点,它们的存在大大加快了信息在网络中的传播速度。虽然中心节点能够加速信息的传播,但是一旦它们成为免疫节点,它们同样能够有效地阻碍信息的传播。由此可见,中心节点具有较大的“社会影响力”,这与实际生活中的情况基本符合,也是社会网络所特有的一个重要特征。在现实生活中,往往名人发表的观点能够在网络中迅速传播,造成较大的社会影响。相反,普通人发表的观点得到的关注较小。这与本文的仿真结果相符。

4) 研究具有不同度数的节点在网络中所占比

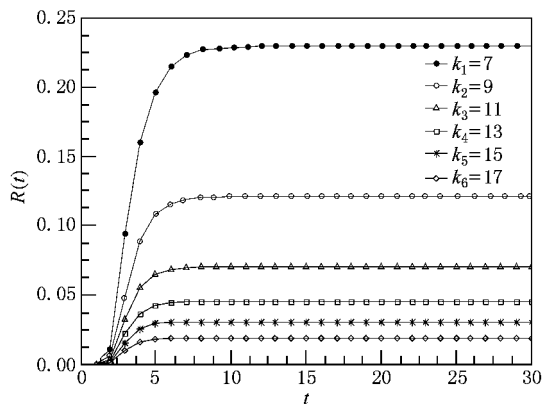


图7 度为 k_1 k_2 k_3 k_4 k_5 k_6 的免疫节点密度 $R(t)$ 随着时间的变化关系

例随着时间的演化关系,即不同度数的节点对于信息传播的影响。图 7 和图 8 所示的是度分别为 $k_1 = 7$ $k_2 = 9$ $k_3 = 11$ $k_4 = 13$ $k_5 = 15$ $k_6 = 17$ 的免疫节点密度和传播节点密度随着时间的变化关系。

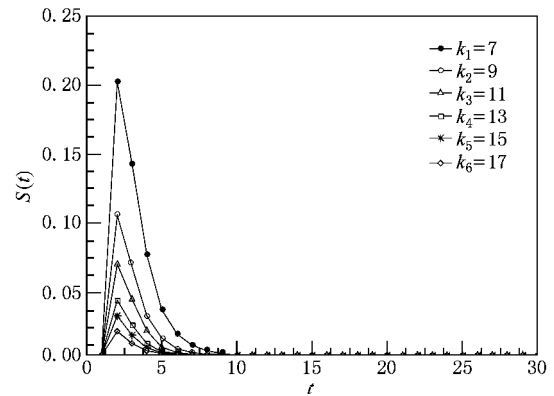


图8 度为 k_1 k_2 k_3 k_4 k_5 k_6 的传播节点密度 $S(t)$ 随着时间的变化关系

从图 7 和图 8 可以看出,虽然具有不同度数的节点在网络中所占的比例不同,但是它们的变化趋势是大体相同的。即由于受到传播机理的作用,节点表现出相似的行为特征。

4. 结 论

本文研究了在线社交网络中的信息传播,提出了一个基于 SNS 网络的信息传播模型。该模型定义了三个传播规则,考虑了节点度的影响,结合复杂网络和传染病动力学理论,进一步建立了微分方程组,使建立的模型更符合 SNS 网络的特点。本文模拟了一个在线社交网络,并分析了传播节点、未感染节点、免疫节点在网络中的传播行为。仿真结果表明:由于在线社交网络的高度连通性,信息在网络中传播的门槛几乎为零;初始传播节点的度越大,信息越容易在网络中迅速传播;中心节点具有较大的社会影响力;具有不同度数的节点在网络中所占的比例不同,且它们呈现出相同的变化趋势。本文工作有助于更深刻地理解发生在 SNS 网络中的传播行为,是进一步研究网络舆论传播的基础。如何在动态网络中对传播行为进行建模是今后的研究方向。

- [1] Hu H B , Wang X F 2009 *Phys. Lett. A* **37** 1105
- [2] <http://www.facebook.com>
- [3] <http://www.renren.com>
- [4] <http://www.kaixin001.com>
- [5] Song X D , Lin C Y , Tseng B L , Sun M T 2005 *International Conference on Knowledge Discovery and Data Mining , Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining , Chicago , Illinois , USA , August 21—24 , 2005* , 2005 p479
- [6] Hu H B , Han D Y , Wang X F 2010 *Physica A* **389** 1065
- [7] Kumar R , Novak J , Tomkins A 2006 *International Conference on Knowledge Discovery and Data Mining , Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining , Philadelphia , PA , USA , August 20—23 , 2006* p611
- [8] Mislove A , Marcon M , Gummadi K P 2007 *Internet Measurement Conference , Proceedings of the 7th ACM SIGCOMM conference on Internet measurement , San Diego , California , USA , October 24—26 , 2007* p29
- [9] Chun H , Kwak H , Eom Y H , Ahn Y Y , Moon S , Jeong H 2008 *Internet Measurement Conference , Proceedings of the 8th ACM SIGCOMM conference on Internet measurement , Vouliagmeni , Greece , October 20—22 , 2008* p57
- [10] Ahn Y Y , Han S , Kwak H , Moon S , Jeong H 2007 *International World Wide Web Conference , Proceedings of the 16th international conference on World Wide Web , Banff , Alberta , Canada , May 8—12 , 2007* p835
- [11] Newman M E J , Forest S , Balthrop J 2002 *Phys. Rev. E* **66** 035101
- [12] Ni S J , Weng W G , Fan W C 2009 *Acta Phys. Sin.* **58** 3707 (in Chinese) [倪顺江、翁文国、范维澄 2009 物理学报 **58** 3707]
- [13] Moreno Y , Nekovee M , Pacheco A F 2004 *Phys. Rev. E* **69** 066130
- [14] Zanette D H 2002 *Phys. Rev. E* **65** 041908
- [15] Zhang L , Liu Y 2008 *Acta Phys. Sin.* **57** 5419 (in Chinese) [张立、刘云 2008 物理学报 **57** 5419]
- [16] Java A , Kolar P , Finin T , Oates T 2006 *The 15th International World Wide Web Conference , Edinburgh , UK , May 22—26 , 2006*
- [17] Newman M E J 2002 *Phys. Rev. Lett.* **89** 208701
- [18] Albert R , Barabási A L 2000 *Phys. Rev. Lett.* **85** 5234
- [19] Newman M E J 2001 *Phys. Rev. E* **64** 016132
- [20] Jeong H , Mason S P , Barabási A L , Oltvai Z N 2001 *Nature* **411** 41
- [21] Fu F , Chen X J , Liu L H , Wang L 2007 arXiv: 0701323 [physics]
- [22] Vázquez A , Weigt M 2003 *Phys. Rev. E* **67** 027101

The research of information dissemination model on online social network^{*}

Zhang Yan-Chao[†] Liu Yun Zhang Hai-Feng Cheng Hui Xiong Fei

(Key Laboratory of Communication and Information Systems-Beijing Jiaotong University ,

Beijing Municipal Commission of Education , Beijing 100044 , China)

(Received 23 July 2010; revised manuscript received 8 September 2010)

Abstract

In this paper , we propose a general stochastic model for the information dissemination on the online social network. The model considers the node of degree and propagation mechanism , utilizes complex network theory and dynamics of infectious diseases , and finally establishes the dynamic evolution equations. The dynamic evolution equations describe the evolution process of different types of nodes , and show that the propagation process is influenced by network topology and propagation mechanism. We simulate the information spreading process , and analyze the behavior of different types of nodes on online social network. Simulation results show that information can spread easily on the online social network because of the good connectivity. The greater the degree of the initial spread node , the faster the information spreads on online social network. Center nodes have great social influence , and the nodes with different degrees have the similar trend on online social network. Research shows that the model , having the same characteristics with online social network , contributes to a more profound understanding of information dissemination behavior on online social network.

Keywords: online social network , information dissemination , differential equation , dynamics of infectious disease

PACS: 05.10.-a

^{*} Project supported by the National Natural Science Foundation of China (Grant No. 60972012) , the Natural Science Foundation of Beijing , China (Grant No. 4102047) , the Service Business of Scientists and Engineers Project (Grant No. 2009GJA00048) , the Major Program for Research on Philosophy & Humanity Social Sciences of the Ministry of Education of China (Grant No. 08WL1101) and the Academic Discipline and Postgraduate Education Project of Beijing Municipal Commission of Education , China.

[†] E-mail: 07111026@bjtu.edu.cn