

基于社交网络的观点传播动力学研究*

熊熙 胡勇†

(四川大学信息安全研究所, 成都 610065)

(2012 年 2 月 7 日收到; 2012 年 4 月 5 日收到修改稿)

社交网络和微博是重要的 Web2.0 应用模式, 其观点传播模式与其他网络媒体以及传统媒体相比有很大差异. 本文提出一种基于在线社交网络的观点传播模型, 研究社交网络中舆论观点扩散的形式与特征. 仿真结果表明: 模型中信息传播的速度与六度分割理论的结论十分符合; 一个带强烈倾向性的观点在固有观点均匀分布的网络中传播的情况下, 稳定时网络中不会出现相反的观点; 稳定时的观点分布与源节点的度和回溯深度有关, 并不受信任界限的限制, 这与 Deffuant 模型和 Hegselmann-Krause 模型不同. 同时, 本文还分析了传播意愿、观点变更率和信任界限对弛豫时间的影响.

关键词: 复杂网络, 社交网络, 舆论传播动力学, 有界信任

PACS: 05.10.-a

1 引言

近年来, 复杂网络的拓扑结构、动力学行为引起了人们的极大关注. 在提出的众多网络模型中, 其中最受关注的无疑是 Watts 和 Strogatz 提出的小世界网络模型 (简称 WS 模型)^[1] 与 Barabasi 和 Albert 提出的无标度网络模型 (简称 BA 模型)^[2,3]. 无标度网络结构的泛在性彻底终结了 E-R 随机网络^[4] 统治的时代.

复杂网络的传播动力学理论同样是复杂网络的重要研究方向. 目前的大多研究工作主要使用传染病模型进行分析, 而目前研究最为彻底、应用最为广泛的传染病模型是 SIR 模型^[5] 和 SIS 模型^[6]. 文献 [7] 中得出了无标度网络中传染概率的阈值为零的重要结论. 舆论的传播过程与传染病的传播过程有诸多相似的地方, 比如状态在某些条件下可以相互转换, 因而舆论传播动力学也可以看作服从 SIR 模型^[8,9].

舆论观点动力学源于日常生活中的决策. Sznajd-Weron 和 Sznajd 提出的 Sznajd 模型^[10], 隐

含着 “United we stand, divided we fall” 的思想. 所有人都存在于网络的格点上, 需要在两个观点中做出选择. 任意两个相邻的人持有相同观点时, 就可以劝说他们的邻居采取相同的观点. 随后出现的 Deffuant 模型^[11] 和 Hegselmann-Krause 模型^[12] 将观点值建立在 0 和 1 之间的连续尺度上, 具有更精细的量化特点.

本文提出一种基于在线社交网络的观点传播模型, 研究社交网络中舆论观点扩散的形式与特征. 仿真结果表明: 模型中信息传播的速度与六度分割理论的结论十分符合; 一个带强烈倾向性的观点在固有观点均匀分布的网络中传播的情况下, 稳定时网络中不会出现相反的观点; 稳定时的观点分布与源节点的度和回溯深度有关, 并不受信任界限的限制, 这与 Deffuant 模型和 Hegselmann-Krause 模型不同. 同时, 本文还分析了传播意愿、观点变更率和信任界限对弛豫时间的影响.

2 社交网络

社会性网络服务 (social networking services,

* 国家自然科学基金 (批准号: 60873246) 和中国信息安全测评中心资助的课题.

† E-mail:

SNS)^[13], 作为 Web 2.0 时代的产物, 是继博客、wiki 等互联网应用服务形式之后, 对以“交互”为核心的互联网浪潮的扩展和深化. 一般地讲, 人们通常所提到的 SNS 即是指社交网站 (social network sites) 提供的特定群体的交互服务. SNS 网站近年来发展迅速, 从国外的 Facebook^[14] 到国内的人人网^[15], 均受到大量互联网用户的青睐. 社交网络正逐渐成为网络中各节点信息传播的工具, 建立了以“兴趣、话题、爱好、人际关系”为中心的社交网络架构. SNS 特别强调用户创造内容. 用户不但是信息接受者, 同时也是信息的制造者、发布者和传播者, 成为网络舆论形式中不可分割的一部分^[16].

社交网络“去中心化”的特征为消息高速传播创造了条件, 同时也容易使网络舆论的呈现一种混乱和非理性的状态. 目前对社交网络的研究, 大多集中在对网络结构和消息传播机理的研究^[17], 对社交网络中舆论观点的扩散与演化并无涉及. 如何有效地阻止谣言在社交网络中的扩散, 充分利用意见领袖的影响力引导公众的认知, 是基于社交网络的舆论传播研究工作面临的重大学术和技术问题, 具有重要的社会意义.

我们的研究从提取国内某著名社交网站的数据开始, 其中包括如下参数: 用户总数为 4221, 利用 SNA 方法^[18] 计算出平均度为 10.8, 簇系数为 0.13. 节点度分布如图 1 所示, 呈现幂律分布.

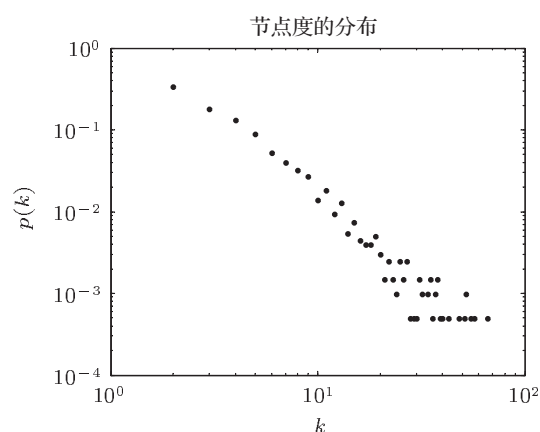


图 1 社交网络的度分布

在此基础上, 我们提取了该网站对某热门事件的所有评论数据 (包括从一个用户引入话题到整个社交网站形成广泛议论的全过程的数据). 在这过程中, 有 719 个用户对某一事件发表了 5770

条评论. 去掉冗余信息后, 从提取的数据中可以发现以下特征:

- 1) 网民通过多条消息在好友关系网中来回多次传递对该话题的评论, 以形成对该话题的讨论.
- 2) 每条消息都包含若干不同用户的评论, 这些用户构成了一条传播路径.
- 3) 网民因为拥有某一共同的特征 (比如对某个话题感兴趣) 而形成一个联系紧密的簇. 话题在某些簇内, 能更频繁地被讨论; 而在另外一些簇里, 用户对其缺乏兴趣, 不会对其过分或根本不关注.
- 4) 有界信任. 网民只与意见相近的网民交流意见.

- 5) 网民会因为各种原因 (如失去传播能力或兴趣), 退出传播过程, 从而引起话题热度减弱.

由此可见, 社交网络中的信息传播不同于传统在线网络传播方式, 如电子布告栏 (bulletin board system, BBS) 等. “去中心化”的特点使得舆论的传播更为便捷和迅速.

3 基于社交网络的观点传播模型

我们借用传染病模型中的几个有关术语来对社交网络中的节点分类, 易感染节点 (S 态)、传播节点 (I 态)、免疫节点 (R 态), 分别对应不知道消息的节点, 知道消息并有能力继续传播的节点, 以及知道消息但已失去传播能力或兴趣的节点. 易感节点在首次接受了来自其邻居节点的信息, 就变为了传播节点. 传播节点在一定情况下转化为免疫节点. 定义如下参数:

为便于建立模型, 我们作如下定义: 观点倾向表示节点所持有的对某一话题支持、反对和中立三种观点中的一种. 观点强度 T_i 表示节点 i 当前具有的观点倾向的强度, 用 -1 到 1 的连续变化的实数来表示. 其中, 区间 $[-1, -0.33]$ 表示观点持反对态度; 区间 $(-0.33, 0.33)$ 表示中立; 区间 $[0.33, 1]$ 则表示支持. 一次传播 (即一条有向线段) 中的源节点称为父节点, 目的节点称为子节点.

首先根据所提取的社交网络数据, 生成好友关系网. 好友关系网是无向网络, 消息只能沿好友关系路径传播. 每个节点都有对该话题的固有观点强度, 即在不受他人影响的情况下该节点持有的初始观点强度. 传播过程采用异步更新的方式, 每次随机选中一个节点进行操作, 直到网络中所有节点都

被选中. 这样一个过程被称为一个时步. 每个时步, 观点的传播过程可以描述如下:

在每个时刻, I 态节点以概率 u_i 向除了父节点以外的所有邻点传播消息, 邻点以概率 v_j 接收消息. 如果子节点是 S 态节点, 则该节点转变为 I 态节点, 并按步骤 2) 的方式重新计算观点强度.

$$I_i + S_j \xrightarrow{u_i v_j} I_i + I_j, \quad (1)$$

其中, 概率 u_i 表示节点 i 向邻点传播消息的倾向程度, 称为传播意愿. 概率 v_i 称为观点变更率, 该值越大, 表示节点 i 的观点越容易受邻点影响. s_i 称为退出概率, 表示节点 i 在每个时步, 由 I 态转化为 R 态的概率. 传播意愿 u_i 与观点变更率 v_i 都是表示观点在网络中顺利传播的概率. 用户对该话题越感兴趣, u_i 和 v_i 越大. 如果用户 i 和 j 之间是存在传播路径的邻点, 那么 $u_i v_j$ 就表示观点从 I 态用户 i 成功传播到 S 态用户 j 的概率. 概率 $u_i v_j$ 被称为用户 i 到用户 j 的传播概率.

2) 如果步骤 1) 中的子节点是 I 态, 则按以下方式重新计算观点强度.

$$T_j(t+1) = C_0 T_j(t) + \sum_{\substack{d(j,k) \leq D \\ |T_j - T_k| \leq B \\ 1 \leq n \leq D}} C_n T_k(t), \quad (2)$$

其中, D 称为回溯深度, $d(j, k) \leq D$ 表示节点 k 到节点 j 的距离小于等于 D , 即该条消息的传播路径上最后 D 个节点. $|T_j - T_k| \leq B$ 表示节点 j 和 k 的观点强度差不超过信任界限 B . 信任界限 B 与 Deffuant 模型和 Hegselmann-Krause 模型中的置信区间具有相同的作用, 都是将节点间观点的交流限制在一定的范围内.

系数 C_0, C_1, \dots, C_n 表示各节点在 t 时刻的观点强度对 $t+1$ 时刻节点 j 的观点强度的影响程度. 如果系数取典型值 $C_0 = 1/2, C_n = 1/2D (1 \leq n \leq D)$, 则 (1) 式就变为

$$T_j(t+1) = \frac{1}{2} \left[T_j(t) + \frac{1}{D} \sum_{\substack{d(j,k) \leq D \\ |T_j - T_k| \leq B}} T_k(t) \right]. \quad (3)$$

该式表示先对 D 个节点求算术平均, 结果再与 $T_i(t)$ 求算术平均. 此时, 节点的影响力只会随着

节点间距离的增长而衰减, 不会受其他因素的影响.

3) 节点 i 在每一个时步, 以概率 s_i 退出演化, 变为 R 态.

从上述对传播过程的描述中可以看出, 该模型引入了回溯的方法, 即节点所持有的观点不光会影响直接相连的邻点, 而且对下游非相邻节点的观点也会造成一定程度的影响.

4 仿真与分析

4.1 信息传播过程分析

利用本模型研究观点传播, 首先根据从网上采集的社交网络数据设置每个节点的参数: u_i, v_i, s_i . u_i 和 v_i 的近似计算方法为: $u_i = v_i = (i \text{ 收到然后马上转发消息的次数} / i \text{ 收到消息的次数})$. s_i 则可以取 i 退出的时步与总的演化时步的比值. 个别节点无法得到其传播参数, 则取中间值 $u_i v_i = 0.5, s_i = 0.005$. 通过 500 个时步 (达到稳定为止) 的循环过程, 来模拟观点的传播. 初始时选取一个节点作为传播源节点, 该节点的度为 42, 观点强度为 0.9, 持支持态度. 其他节点都未获知消息, 处于 S 态, 固有观点在 $[-1, 1]$ 范围内随机分布.

本模型中的消息传播过程可用下式表示:

$$I_i(t) + S_j(t) \xrightarrow{u_i v_j} I_i(t) + I_j(t), \quad (4)$$

$$I_i(t) \xrightarrow{s_i} R_i(t). \quad (5)$$

用 $s(t), i(t), r(t)$ 分别表示群体中处于 S 态、 I 态和 R 态的节点密度. 在各参数取上述值时, 通过仿真可以得到传播节点、易感节点和免疫节点的密度随时间变化的情况, 结果如图 2 所示.

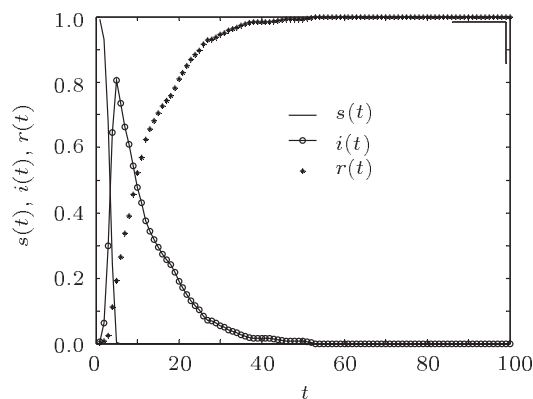


图2 三种传播状态的节点密度随时间的变化

从图 2 可以看出, 易感节点密度 $s(t)$ 迅速递减到 0, 表明消息在社交网络中传播极快, 所有人能在短时间内获得消息. 传播节点密度 $i(t)$ 先增大后减小, 最后趋近于 0. 消息传播初期, 因大量的节点得到消息, 曲线快速增长; 当达到顶点后, 不再有新的节点获得消息, 并且用户逐渐对传播消息失去兴趣, 引起曲线的衰减. 免疫节点密度 $r(t)$ 从 0 开始呈现由快到慢的增长态势, 最后达到 1.

社交网络源于 Stanley Milgram 在 1967 年创立的六度分割理论^[19], 即“最多通过六个人你就能认识任何一个陌生人”. 从图 2 中可以看出, 度为 42 的源节点在社交网络中传播观点, 只需要经过 5 个时步, 就可以达到每一个用户. 对度为 10 和 2 的另外两个源节点进行同样的仿真, 得出的结论是 5 个和 6 个时步. 可见上述仿真的结果与六度分割的结论十分符合. 因此, 六度分割理论在舆论传播方面同样适用. 即舆论可以在 6 次好友间传递中到达任何一个人.

4.2 观点传播过程与特征

为迅速传播观点, 源节点一般都是度比较大的节点. 在这种情况下, 图 3 描述了在均匀分布与一个源节点 (度为 42, 观点强度为 0.9) 的条件下, I 态的节点中, 分别持支持、中立和反对态度的节点密度随时间的变化情况. 取回溯深度 $D = 3$.

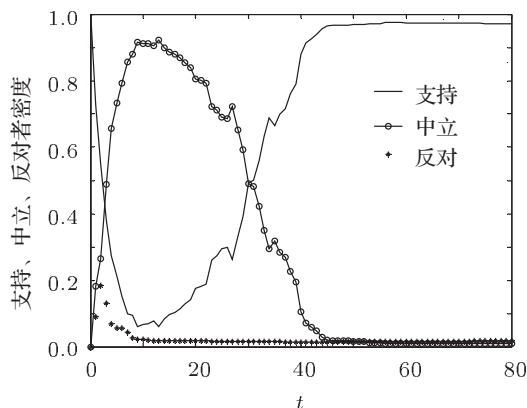


图 3 三种观点倾向的节点密度随时间的变化 (源节点的观点强度为 0.9; $k_0 = 42$; $D = 3$)

初期, 中立和反对的节点获知该消息引起这两种曲线的上升和支持曲线的下降. 随后正反双方激烈讨论, 引起众多态度不坚定的用户趋向于保留意见而转为中立. 当中立曲线达到顶点的同时, 其他

两支曲线达到最低点. 此时的正反方节点属于“铁杆的”支持者与反对者, 其观点强度的绝对值接近于 1. 随后中立曲线衰减为 0, 并且支持曲线递增, 表明认清形势后的中立观望者全部逐渐转化为支持. 个别反对态节点与其邻点的观点强度差大于信任界限, 因而无法进行沟通, 陷入孤立状态, 形成稳定时残留的反对节点.

稳定时, 从表面上看, 支持态节点占绝大多数. 这种情况表明, 固有观点均匀分布条件下, 当初始只有一种观点时, 网络中几乎不会听见相反的声音. 网络中的用户呈现较大程度的“盲从”. 通过计算此时的磁化率, 可以发现该值出现在 0.33 附近, 查看此时所有 I 态节点的观点强度, 可以发现绝大部分 I 态节点的观点强度在 0.33 这个支持与中立的交界线靠支持一侧. 虽然这些节点持支持态, 但是这种临界状态使他们随时会根据外界条件的变化而转变态度.

在个别情况下, 观点是由网络中的普通用户发布, 这些用户的度通常比较小. 图 4 表示源节点度 k_0 为 3 时, 三种观点倾向的节点密度随时间的变化情况.

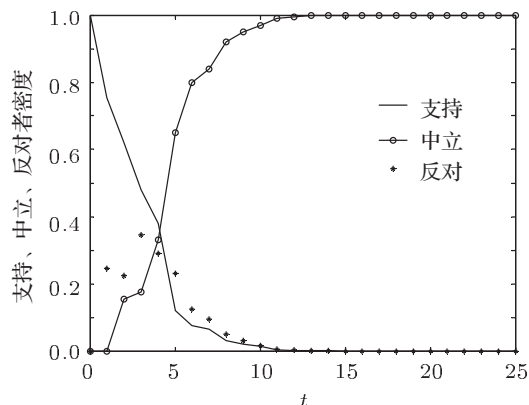


图 4 三种观点倾向的节点密度随时间的变化 (源节点的观点强度为 0.9; $k_0 = 3$; $D = 3$)

从图 4 中可以看出, 支持态节点衰减为 0, 反对态节点密度虽然先有一定增长, 最后还是慢慢变为 0. 所有节点都演化为中立说明度较小的节点 (通常称为“被边缘化”) 很难在网络中传播自己的观点. 演化过程中出现的正反观点, 都是相互辩论的中间状态. 网络最终还是会趋向于中立的目标状态.

当源节点的度 k_0 较大时, 图 3 的情况占多数, 而图 4 中的情况很少出现. 随着 k_0 减小, 图 3 的情

况所占比例逐渐减小, 而图 4 的情况所占比例逐渐增大. 接下来, 在每种条件下, 我们把重点放在两种情况中占多数的情况. 在不同传播参数条件下, 稳态磁化率 $m_e(t)$ 随回溯深度 D 与源节点度 k_0 的变化如图 5 所示.

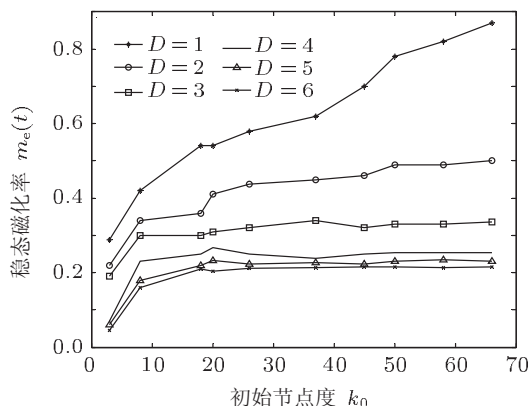


图 5 在回溯深度 D 取不同值的情况下, 稳态磁化率 $m_e(t)$ 随源节点度 k_0 的变化

当 k_0 取某个固定值时, $m_e(t)$ 随 D 的增大而减小. 当 D 取某个固定值, 在源节点度 k_0 很小时, 源节点只能与极少的人交流意见, 不能将自己的观点传播出去. 此时 $m_e(t)$ 主要受 k_0 的影响, 因此曲线呈快速上升趋势. 当 $D = 1$ 时, $m_e(t)$ 会随 k_0 的增长而持续增长. 这是因为每个节点的每次观点变更都只参考父节点, 源节点的观点特别容易影响它周围的节点, 并且这种影响会延伸到整个网络. k_0 越大, 源节点的影响越迅速和猛烈. 而当 $D > 1$ 时, 每个节点的每次观点变更, 都要参考前面不止一个节点的意见, 从而使源节点的影响大为削弱. 当 k_0 大到可将观点传播出去以后, $m_e(t)$ 就在稳定在某个值附近, 不再随 k_0 变化.

根据上面的分析可得, 社交网络的稳态观点分布主要取决于源节点度 k_0 和回溯深度 D , 并不受信任界限 B 的影响. 这与 Deffuant 模型和 Hegselmann-Krause 模型的稳态观点分布主要由置信区间决定不同.

在社交网络中, 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. 由此可见, 中心节点具有较大的“社会影响力”^[17]. 例如, 谣言通常就是从中心节点 (度数大) 发出, 利用社交网络结构上的高连通性, 使网络中的用户在短时间里接受其观点与主张. 同时, 从图 1

可以看出, 大部分节点都具有较小的度, 这些节点都处于边缘位置, 其影响力十分有限, 他们的观点很难得到网络中其他节点的认同.

4.3 其他参数对观点传播的影响

前面已经提到, 传播概率 $u_i v_j$ 被称为用户 i 到用户 j 的传播概率. 为定性研究传播概率 $u_i v_j$ 对观点传播的影响, 可以假设网络中所有节点的 u_i 和 v_i 相同, 即 $u_i = u, v_i = v, u$ 和 v 都是常数. 那么 uv 就表示观点在网络中任意一对用户间成功传递的概率. 当 $uv < 0.3$ 时, 用户对该话题几乎没有兴趣, 使得观点难以传播. 同样, 信任界限过小时, 用户间不能进行充分而有效地沟通, 不会出现一种态度占据压倒性优势, 反而会引起三种态度共存和分区域聚集的现象. 通常取信任界限 $B \geq 0.8$. 系统从一个初始态演化到稳定态所需要的时间称为弛豫时间 μ ^[20]. 图 6 表示在 B 取不同值的情况下, μ 随 uv 的变化. 其中, $uv \geq 0.3, B \geq 0.8$.

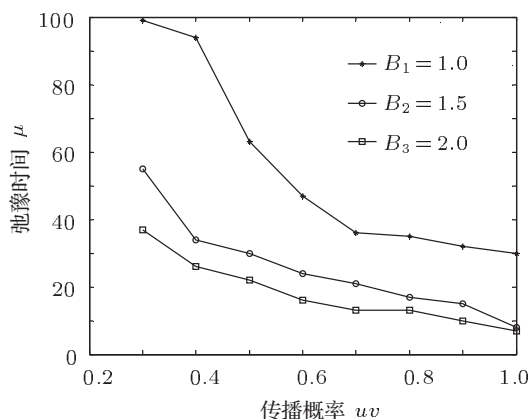


图 6 在信任界限 B 取不同值的情况下, 弛豫时间 μ 随传播概率 uv 的变化

可以明显地看到, 在同一个信任界限下, 弛豫时间 μ 随传播概率 uv 的增大而减小. 在以同一个传播概率 uv 传播消息时, 弛豫时间 μ 也是随传播界限 B 的增大而减小.

5 结论

随着对复杂网络 (包括社交网络在内) 结构的研究逐渐深入, 网络结构与传播的耦合动力学的研究正逐渐成为热点. 基于社交网络的观点传播和演化是一个复杂的过程, 各种因素相互影响, 呈现多

种特征. 研究的目标就是从网络信息的传播规律中找出正确监控、预测和引导的方法.

本文提出的基于社交网络结构的观点传播模型引入了回溯的方法, 即观点传播路径上的节点会影响其下游非相邻节点的观点. 数值仿真表明该网络模型与实际社交网络数据的统计结果基本符合,

揭示了一些重要的规律, 比如稳态网络中的观点分布取决于源节点的度、观点强度以及回溯深度, 与信任界限无关.

本文工作有助于进一步理解和研究社交网络中舆论的传播规律. 网络结构与传播交互反馈、动态变化的动力学将是未来的研究方向.

-
- [1] Watts D J, Strogatz S H 1998 *Nature* **393** 440
 - [2] Barabasi A L, Albert R 1999 *Science* **286** 509
 - [3] Albert R, Barabasi A L 2002 *Phys. Rev. Mod.* **74** 47
 - [4] Erdos P, Renyi A 1959 *Publ. Math.* **6** 290
 - [5] Kermack W O, Mckendrick A G 1927 *Proc. Roy. Soc.* **A115** 700
 - [6] Kermack W O, Mckendrick A G 1932 *Proc. Roy. oc.* **A138** 55
 - [7] Pastor-Satorras R, Vespignani A 2001 *Phys. Rev. Lett.* **86** 3200
 - [8] Sudbury A J 1985 *Appl. Prob.* **22** 443
 - [9] Zhou J, Liu Z, Li B 2007 *Phys. Lett. A* **368** 458
 - [10] Sznajd-Weron K, Sznajd J 2000 *Int. J. Mod. Phys. C* **11** 1157
 - [11] Deffuant G, Amblard F, Weisbuch G, Faure T 2002 *JASSS-The Journal of Artificia* **5** 1
 - [12] Hegselmann R, Krause U 2002 *JASSS-The Journal of Artificia* **5** 2
 - [13] Hu H B, Wang X F 2009 *Phys. Lett. A* **37** 1105
 - [14] www.facebook.com
 - [15] www.renren.com
 - [16] http://blog.sina.com.cn/s/blog_67caa13b0100ip5x.html
 - [17] Zhang Y C, Liu Y, Zhang H F, Cheng H, Xiong F 2011 *Acta Phys. Sin.* **60** 50501 (in Chinese) [张彦超, 刘云, 张海峰, 程辉, 熊菲 2011 物理学报 **60** 50501]
 - [18] Liu J 2004 *An Introduction to Social Network Analysis* (Beijing: Social Sciences Academic Press) p116 (in Chinese) [刘军 2004 社会网络分析导论 (北京: 社会科学文献出版社) 第 116 页]
 - [19] <http://baike.baidu.com/view/200573.htm>
 - [20] Sznajd W K 2005 *Acta Phys. Polonica B* **36** 1001

Research on the dynamics of opinion spread based on social network services*

Xiong Xi Hu Yong[†]

(*Institute of Information Security, Sichuan University, Chengdu 610065, China*)

(Received 7 February 2012; revised manuscript received 5 April 2012)

Abstract

Social network services and microblog are important application modes of Web2.0, in which opinion spread is quite different from that in other cyber-media and in traditional media. In this paper, we present an opinion-spreading model based on online social network services, to study the forms and features of the spread of public opinion in social network services. The simulation results show that the model can fit the actual data from a social network site. The speed of information spread is consistent with the conclusion of six degrees of separation theory. When an opinion with strong tendency spreads in a network in which intrinsic views obey uniform distribution, the opposite view cannot exist in the stable network. In a stable network, view distribution is related to the degree of source node and the depth of backtrack, but not related to confidence limit, which is different from Deffuant model and Hegselmann-Krause model. Meanwhile, in this paper the also analyze the influence of the probability of spreading will, the probability of opinion change and confidence limit on relaxation time. Finally, in the paper are shown two applications of the model: the spread of rumors and the role of opinion leaders.

Keywords: complex networks, social networking services, dynamics of opinion spread, limitary trust

PACS: 05.10.-a

* Project supported by the National Natural Science Foundation of China (Grant No. 60873246) and the China Information Technology Security Evaluation Center.

[†] E-mail: huyong@scu.edu.cn