

基于微博网络的信息传播模型

郑 蕾, 李生红

(上海交通大学 电子信息与电气工程学院, 上海 200240)

【摘 要】微博(Microblog)中信息传播的面积、速度、效率都得到了极大提高,研究信息在微博中的传播规律,对控制和引导舆论具有非常重要的理论价值和现实意义。结合信息在微博网络中的传播规则和微博网络的拓扑结构,通过构建一种新的基于微博网络的信息传播模型,并在实际新浪微博拓扑子网中模拟信息流向,得到如下结论:当信息敏感度大于某一临界值后,明星用户在信息扩散中的意见领袖角色随着信息敏感度的增加而逐渐弱化,当信息敏感度趋于1时,明星用户在信息传播中的优势趋于0。

【关键词】微博网络;信息传播模型;明星用户;信息敏感度

【中图分类号】TP309

【文献标识码】A

【文章编号】1002-0802(2012)02-0039-03

A Novel Information Diffusion Model based on Microblog Network

ZHENG Lei, LI Sheng-hong

(School of Electronic Information and Electrical Engineering, Shanghai Jiaotong University, Shanghai 200240, China)

【Abstract】The emergence of Microblog provides a powerful means for sharing information among between the ordinary people. In this paper, a novel information diffusion model based on Microblog network is proposed, the network structure, coupled with information diffusion mechanism, is considered. And based on the proposed model, the role of super-users and ordinary users in the information diffusion process is investigated. The simulation on information flow over a large static social network derived from Sina Microblog gives some interesting results.

【Key words】Microblog; information diffusion model; super-users; information sensitivity

0 引言

微博,是一种基于用户关系的信息分享,传播以及获取的平台,用户通过WEB, WAP以及各种客户端组件便能随时随地地实现信息的更新和即时分享。与BBS和博客不同的是,微博用户可以不受时间空间的限制,持续不断地对某条信息作出评论和转发,使得该信息可能在极短时间内以“核裂变”的方式被成百上千的用户转发,从而演化成引起舆论关注的社会新闻事件,例如,直播玉树大地震,“我爸是李刚”,唐俊“学历门”事件等。在微

博网络上,信息的传播速度,广度和效率都得到了极大提高,但这无疑也为一些带有敏感字眼的失实消息(如“无偿献血被红十字以一袋200元卖给医院”)在微博上的传播创造了更有利的条件。因此,研究信息(舆论)在微博网络上的传播及其传播的性质、规律等,进而研究如何控制、引导舆论,具有重要的理论价值和现实意义。

现有的文献大多从单一的角度来研究人际网络上的信息传播,如,信息传播所在网络的拓扑结构^[1-3],个体相互作用规则^[4-7]。但实际中信息(舆论)的传播与形成是一个典型的复杂系统的演化过程,需要将个体的相互作用和网络结构进行综合考虑才能更准确地描述信息在网络中的传播。文中在基于信息在微博网络上的传播规则,并结合微博网络拓扑结构对传播行为的影响,初步建立了微博网络中信息传播的随机演化仿真模型。

在微博网络中,入度越大,说明关注的用户越

收稿日期:2011-10-31。

基金项目:国家自然科学基金(批准号:61071152,61071081,61171173);973项目(No.2010CB731403, No.2010CB731406);北京交通大学通信与信息系统重点实验室项目。

作者简介:郑蕾(1989-),女,硕士,主要研究方向为信息安全;李生红(1971-),男,教授,博士生导师,主要研究方向为信息安全。

多,信息来源越丰富,而出度越大,说明关注自己的人越多,影响力越大。早在 20 世纪 40 年代,拉扎斯菲尔德等人在《人民的选择》中提出了“意见领袖”的概念,所谓意见领袖,维基百科是这样描述的,在人际传播网中经常为他人提供信息,同时对他人施加影响的“活跃分子”,他们在大众传播效果的形成过程中起着重要的中介或过滤的作用。通常而言,大部分微博用户的粉丝一般是几十至几百的数量级,而少数用户的粉丝超过上千乃至上万,称其为明星用户(意见领袖)。文中基于提出的信息传播模型,探讨了明星用户和小级别用户在微博网络信息扩散中扮演的角色,通过仿真实验,得到了一些有趣的发现。

1 微博网络信息传播规则

微博的每个用户通常都拥有一定数量的“粉丝”(跟随者),其发布的信息会以广播的形式通知每个“粉丝”,从而受到关注甚至转发。以新浪微博为例,它具有 4 大功能模块:发布功能,转发功能,关注功能及评论功能。发布功能,用户可以随时随地发布消息;转发功能,用户可以把自己喜欢的消息一键转发到自己的微博,在转发过程中原帖内容不会被篡改,且转发记录会以“//@用户名”的方式保留转发路径;关注功能,用户 u 可以对自己喜欢的用户 v 进行关注,成为用户 v 的“粉丝”,这一行为不需要得到受关注用户的同意,之后用户 v 更新的所有消息便会同步出现在用户 u 的微博首页上;评论功能,用户可以对任何一条微博进行评论。由此,消息 I 在微博网络上的传播过程可以简述如下:

- 1) 用户 u 发布消息 I , 消息的形式可能是一句话,一张照片等。
- 2) u 的粉丝可通过个人首页上系统的新鲜消息提醒和主动访问 u 的微博页面这两种途径来获知消息 I 。
- 3) 对消息 I 感兴趣的粉丝 v 可能会对该消息进行评论,收藏和转发,从而将信息 I 进行再次传播。
- 4) 如此重复过程 2) 和 3)。

2 信息传播模型描述

将微博网络中的信息传播机制和网络拓扑结构相结合,提出一种基于微博网络的信息传播模型 $IDM(\alpha, \beta)$, 其中 $\alpha \in [0, 1]$ 表示消息的敏感度,其值是由消息的来源,内容及表现形式决定的; $\beta \in \{-1, 0, 1\}$ 是模型的另一个参数,用来决定选择转发信息的节点的方式。在这个模型中,网络中的节点只可能处于 3 种状态,一是未激活状态,即不知道消息 I , 记为 S_0 , 二是被消息 I 激活但选择不作为

状态,记为 S_1 , 三是被消息 I 激活后对消息进行转发的状态,记为 S_2 。时间步序列是一个离散的等间隔(如, 5 min)时间序列。模型定义如下:

1) 初始时刻,将网络中所有节点的状态初始化为 S_0 。在时间步 $T=0$, 随机选择一个节点 v 作为发布消息 I 的源节点。节点 v 的出度是 $dout_v$ 。

2) 在时间步 $T=T+1$, 关注节点 v 的 $dout_v$ 个节点被激活, 其中转发消息 I 的节点数量为 $M(M \leqslant dout_v)$, M 计算如下:

$$M = dout_v \cdot a, \quad (1)$$

设 P^T 是时间步 T 内选择转发的节点的集合, 即 $P^T = \{u_j\}$, 其中 u_j 的状态为 S_2 。

3) 对关注节点 v 且不属于集合 P^{T-1} 的 $n(n \leqslant dout_v)$ 个节点的出度进行排序, 按照从小到大的顺序记为 $Q = \{u_{(1)}, u_{(2)}, \dots, u_{(dout_v)}\}$ 。若 $\beta=1$, 顺序选择序列 Q 中前 M 个节点, 将它们状态设为 S_2 , 放入集合 P^{T+1} ; 若 $\beta=-1$, 顺序选择序列 Q 最后 M 个节点, 将它们状态设为 S_2 , 放入集合 P^{T+1} ; 若 $\beta=0$, 从序列 Q 中随机选出 M 个节点, 将它们状态设为 S_2 , 放入集合 P^{T+1} 。将节点 v 从集合 P^T 中删去。

4) 对每一个节点 $v \in P^T$ 递归执行 2) 和 3), 直到集合 P^T 为空。

由式(1)可知, 微博网络中节点 v 对消息传播的贡献是与它的出度及消息的敏感度成正比的, 这与实际情况中用户的粉丝越多, 其发布(转发)的消息被进一步转发的可能性就越大及消息的敏感度越高, 吸引力越大, 其转发率越大是吻合的。在该模型中, 利用参数 β 来选择节点 v 的“粉丝”节点中会进一步转发消息 I 的节点, 当 $\beta=-1$ 时, 选择出度小的节点对信息进行转发, 当 $\beta=1$ 时, 选择出度大的节点对信息进行转发, 当 $\beta=0$ 时, 随机选择出转发节点。

3 数据库

现利用自动搜索程序, 获得了新浪微博网络的一个子网络, 该网络是个有向图, 记为 $G=(V, E)$, 其中节点 $v \in V$ 表示一个微博 ID, 有向边 $(u, v) \in E$ 表示用户 v 是用户 u 的粉丝, 信息从 u 流向 v 。所获取的子图的节点数 $N=62\,668$, 平均入度 $din_mean=248.12$, 平均出度 $dout_mean=329.49$ 。累积出度的分布如图 1 所示, p 表示出度值。

4 仿真结果分析

图 2 显示的是信息敏感度取不同值时, 采取不同的转发节点选择策略, 信息在网络上的扩散面积

和扩散耗时的比较,图2中空点,倒三角,叉分别代表 β 取1,0,-1时信息在网络中的扩散比例,每一条曲线是通过每对参数50次仿真结果的统计平均得到的。从图2可以看出当信息敏感度较小时,明星用户在信息传播过程中扮演着意见领袖的角色,对舆论的传播起着绝对的推动作用;但当信息敏感度大于某一临界值后,明星用户对提高信息传播速度、扩大信息传播面积的作用逐渐减小;当信息敏感度趋于1时,明星用户在信息传播中的优势趋于0。

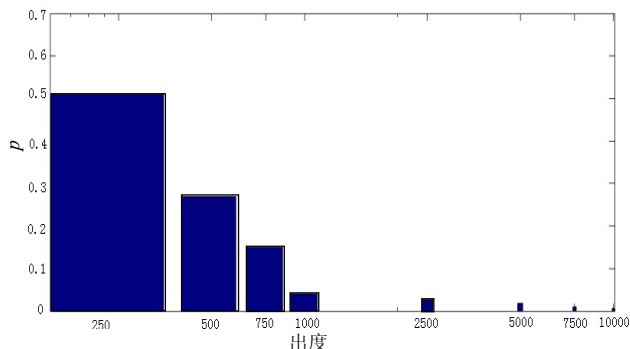
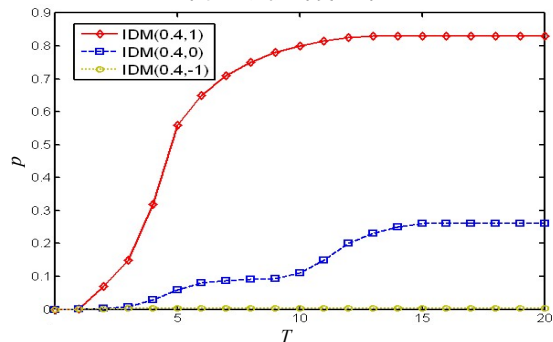
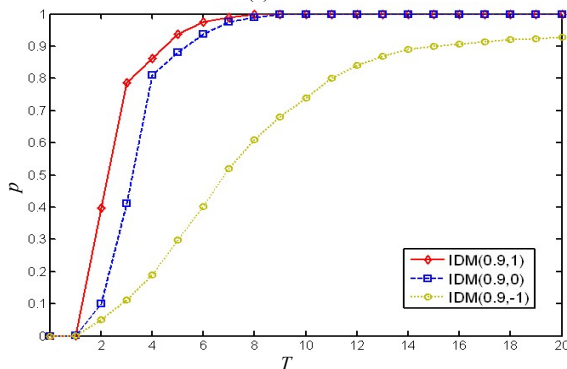


图1 累积出度分布



(a) $\alpha=0.4$



(b) $\alpha=0.9$

图2 信息扩散密度随时间的演化

5 结语

微博网络是复杂网络在实际人际网络的一个实例,文中初步构建了一个基于微博网络的信息传播

模型,该模型准确描述了微博网络中信息传播的基本特点,以及舆论演化过程中每次节点的选取、节点策略的转移。通过在新浪微博的拓扑子网上进行仿真,得到如下结论:明星用户在信息扩散的过程中起着意见领袖的作用,但当信息敏感度大于某一临界值后,明星用户对提高信息传播速度、扩大信息传播面积的作用逐渐减小,当信息敏感度趋于1时,明星用户在信息传播中的优势趋于0。由于舆论传播过程中存在着许多会影响传播过程的不确定因素,因此该模型还有很大的改善空间,例如,信息敏感度的划分需要更多的实证研究来支持。相信随着这方面研究的深入,复杂网络中的信息传播机制会越来越得到重视,进而更好的提升网络空间的文明度和保障人们对网络安全的需求。

参考文献

- [1] JAEWON Y, LESKOVEC J. Modeling Information Diffusion in Implicit Networks[C]//IEEE. 2010 IEEE 10th International Conference on Data Mining. Berlin, Germany: IEEE, 2010:599-608.
- [2] GOLDENBERG J, LIBAI B, MOLDOVAN S, et al. The NPV of Bad News[J]. International Journal of Research in Marketing, 2007, 24 (03):186-200.
- [3] LAHIRI M, CEBRIN M. The Genetic Algorithm as a General Diffusion Model for Social Networks[C]//The 24th AAAI Conference on Artificial Intelligence. Atlanta, Georgia, USA:[s. n.], 2010: 76-81.
- [4] SAITO K, KIMURA M, OHARA K, et al. Learning Continuous Time Information Diffusion Model for Social Behavioral Data Analysis[J]. Advances in Machine Learning, Springer, 2009 (5828): 322-337.
- [5] 惠彦涛, 吴静. 互补结构网络建模与统计特性分析[J]. 通信技术, 2011, 44 (05): 94-96.
- [6] HADDADI H, BENEVENUTO F, GUMMADI K P. Measuring User Influence in Twitter: The Million Follower Fallacy[C]// Proceedings of International AAAI Conference on Weblogs and Social. Washington, DC: ICWSM, 2010:97-105.
- [7] BHARATHI S, DAVID K, MAYHAR S. Competitive Influence Maximization in Social Networks Internet and Network Economics[J]. Springer, 2007 (2315): 306-311.

欢迎广大读者踊跃投稿!