

文章编号: 1005-8451 (2013) 10-0054-05

基于用户个性的微博网络传播模型研究

张文杰¹, 穆海冰¹, 张长伦²

(1.北京交通大学 通信与信息系统北京市重点实验室, 北京 100044;

2.北京建筑工程学院 理学院, 北京 100044)

摘要: 微博的转发过程易受各种因素影响, 构建合理、完善的传播模型准确描述信息的传播过程, 对网络舆论的预测和引导具有重要的支撑作用。本文从信息传播规则和网络拓扑结构2个角度建立信息传播模型, 并将用户个性作为影响因素纳入到模型当中, 在实际微博拓扑子网中模拟信息的转发过程。仿真结果表明用户个性对信息扩散的影响作用明显, 信息吸引力临界值在不同类型个体居多的网络中会发生变化, 并且不同个性用户对信息扩散的影响作用是不同的, 自主型个体对信息扩散具有很强的推动力, 中立型个体次之, 无主见型个体最差。

关键词: 微博网络; 传播模型; 用户个性; 信息吸引力; 信息扩散

中图分类号: U285.5 : TP39 **文献标识码:** A

Research on propagation model of microblog network based on personalities of users

ZHANG Wenjie¹, MU Haibing¹, ZHANG Changlun²

(1.Key Laboratory of Communication & Information Systems, Beijing Jiaotong University,
Beijing Municipal Commission of Education, Beijing 100044, China;

2.Science School, Beijing University of Civil Engineering and Architecture, Beijing 100044, China)

Abstract: The forwarding process of microblog was susceptible to a variety of factors. Constructing a reasonable, perfect propagation model to describe accurately the propagation process of information had an important supporting role of forecasting and guiding the public opinion. In this paper, the propagation model was built based on both information propagation mechanism and the network structure, and the personalities of user were incorporated into the model as influencing factors, and then analog the forwarding process of information in the actual microblog topology subnet. Simulation results showed that the personalities of user had obvious effect on information diffusion and the critical value of information attraction would change in the different network where some kind of personality was the majority. The user with different personalities had different effect on information diffusion. The autonomous individual had a strong impetus to information diffusion, the neutral individual follows, and non-assertive individual does worst.

Key words: microblog network; propagation model; personalities of user; information attraction; information diffusion

微博作为一种新兴的开放互联网社交服务工具, 具有独特的传播特征和强大的传播能量。它以发布便捷、时效性、交互性强等特点吸引了强大的用户群体。微博信息的转发是很复杂的演化过程, 综合考虑微博网络中的信息传播规则和网络拓扑结构2个角度, 构建基于微博网络的信息传播模型, 准确模拟信息在实际微博拓扑子网中的信息流向, 研究微博网络中信息的传播性质和

规律, 对网络舆论的预测、引导具有非常重要的理论价值和现实意义。

微博网络中用户的个性对微博的转发具有不可忽视的影响作用, 本文将用户个性作为影响因素纳入到传播模型中, 研究不同比例混合群体构成的网络中信息的传播性质并对仿真结果进行分析。

1 复杂网络传播模型研究现状

信息传播的过程是在以个体为节点、以连接

收稿日期: 2013-01-25

作者简介: 张文杰, 在读硕士研究生; 穆海冰, 教授。

为边的复杂网络中信息的复制、转移、变化和扩散的过程。实证的研究表明基于用户个体交互规则的概率影响模型很适合描述实际的传播过程,其基本假设是:用户个体可能在人际关系网络中其它邻居用户的影响下加入信息的传播,并且这种影响关系是概率方式,即每一个加入传播的个体都以一定的概率影响其邻居加入传播。实际建模过程中一般会根据需要在影响模型中考虑具体的影响因素,并设定一定的人际网络拓扑结构,对传播的范围、时间、持续性等特征进行研究。

在传染病动力学研究中,通过构建个体交互概率影响模型来描述传染病的扩散过程。现有的研究中,很多是借鉴传染病模型,结合具体网络的特性对模型改进,研究用户节点状态随时间的演化关系。刘俊^[1]在基于不同网络拓扑结构的E-mail病毒传播模型研究的基础上,针对网络拓扑的均匀性和缺少理论依据两方面的不足,提出了无标度网络环境下E-mail病毒的传播模型,通过对模型的求解,得到E-mail用户感染密度随传播率、恢复率和网络平均度变化的计算表达式,并通过实验证明了传播率与网络平均度是影响E-mail病毒蔓延的关键性因素。张彦超^[2]构造基于在线社交网络的信息传播模型,模型考虑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,建立动力学演化方程组,刻画不同类型节点随着时间的演化关系,分析其在网络的行为规律。许晓东^[3]根据微博社区网络呈现短平均路径、度分布指数为常值、信息传播有向的特性,结合文献[1]中病毒传播的相关算法,将网络中的节点按照其度值进行分类,建立微博社区谣言传播模型研究谣言在网络中的传播机理,研究表明谣言传播受有效传染率和网络度分布熵影响。杨春霞^[4]借鉴SIR模型,针对微博病毒传播的有向性在有向无标度网络上建立微博病毒的仿真模型,将传播过程分为自由传播阶段和警告抑制阶段,研究用户点击率、转发率以及对信息的反应速度等因素对病毒传播的影响。

在基于用户个体交互规则的概率影响模型中,个体间的作用规则 and 影响因素的定义至关重要,但现有的研究对于如何定义作用规则 and 影响因素还没有达成一定的共识。丁飞^[5]对舆论事件中参与个体的交互过程进行建模,采用Multi-Agent方

法对提出的模型进行仿真,研究舆论事件之间的竞争和影响关系。孙庆川^[6]对传播规则和传播网络的结构进行限定,建立人际网络传播模型的五元组,并在计算机上模拟,发现存在信息传播临界值,当信息吸引力大于临界值时信息才能传播开,否则几乎不能传播。刘丰^[7]参照文献[6]提出的信息传播模型五元组,在微博环境下,分别根据用户关系、用户关系的度量、用户的转发行为、用户及微博项的特征相应的生成网络底图、节点间信息亲和力、节点的信息接受能力、节点的信息传播能力,构建信息传播模型五元组,模拟信息在微博中的传播过程。郑蕾^[8]通过对参数设置决定选择转发信息的节点方式,构建一种新的基于微博网络的信息传播模型,并在新浪微博拓子网中模拟信息流向,得出明星用户在信息扩散中的意见领袖作用随着信息敏感度的增加逐渐弱化的结论。

这些实证研究表明,通过基于用户个体交互规则的概率影响模型来描述复杂网络特别是互联网的传播过程是可行的。到目前为止,研究者们对这种建模方法达成了一定的共识,但对互联网中各种复杂的传播现象,以及一些重要影响因素对传播过程的影响缺乏足够的认识。上述部分研究侧重于构造传播背景和环境,将用户抽象成无主观意识的节点,忽略了用户个性等主观意识对信息传播造成的影响,本文将用户个性作为影响因素纳入到模型当中,定量分析结果,来仿真研究实际的传播过程。

2 基于用户个性的信息传播模型

参照文献[6]提出的传播模型,针对微博时效性强、传播具有有向性等特点,考虑影响传播过程具体影响因素,对模型进行一定的改进。本文将微博中的用户抽象成网络节点 V ,用户之间的粉丝关系抽象成边 E ,基于一定人际网络拓扑结构的微博拓子网抽象成信息传播的空间 $G=(V,E)$,采用基于用户个体交互规则的概率影响模型来模拟微博中信息的传播过程。

2.1 模型传播规则描述

初始时刻 $T=0$,随机选择一个用户节点,作为发布微博信息的源节点, $T=t$ 时刻转发微博信

息者的集合为 $A^{(t)}$, $T=t+1$ 时刻, 一个不属于 $A^{(t)}$ 的个体会根据 t 时刻信息的吸引力 Φ_e 、受邻居用户的影响程度 $Nei_v^{(t)}$ 以及个体的接受信息临界值 $Acp_v^{(t)}$ 来决定自己在 $t+1$ 时刻是否转发该信息, 若满足转发条件, 则转发信息, 进入 $A^{(t+1)}$ 集合, 反之, 不转发信息。

2.2 模型假设

(1) 模型中只考虑用户之间的粉丝关系存在, 忽略这种联系的强度。

(2) 模型中不区分用户节点的影响力的大小。

(3) 模型中假设信息吸引力只取决于传播的话题, 而无个体间差异。

2.3 模型构建

模型是一个五元组 $(G, \Phi_e, Acp_v^{(t)}, w, A^{(t)})$ 。其中, G 表示信息传播空间, Φ_e 表示信息的吸引力, $Acp_v^{(t)}$ 表示个体的接受信息临界值, w 表示信息吸引力在影响因素中所占的权重, $A^{(t)}$ 表示转发微博信息者的集合。

假设用户转发信息受两种因素影响:

(1) 信息的吸引力 Φ_e , $\Phi_e \in [0,1]$ 。由于微博的时效性、交互性强等特点, 在不受外界因素影响下, 信息吸引力会逐渐衰减, 假设呈指数衰减, 衰减系数设为 $Decay$ 。

(2) 受邻居用户的影响程度 $Nei_v^{(t)}$, 具体量化为用户关注的人群中转发信息的人群数量在关注人群当中所占的比重, 即:

$$Nei_v^{(t)} = |M_v^{(t)}| / dou_v$$

其中, $M_v^{(t)} = \{v' | (v', v) \in E, v' \in A^{(t-1)}\}$, dou_v 表示该用户关注的人群数量。

用户转发信息受到的综合影响力, 记为 $e_v^{(t)}$, 其值为:

$$e_v^{(t)} = w \cdot \Phi_e \cdot Decay^t + (1-w) \cdot Nei_v^{(t)}$$

其中, $w \in [0,1]$ 。比较 $e_v^{(t)}$ 和 $Acp_v^{(t)}$, 假如 $e_v^{(t)} > Acp_v^{(t)}$, 则转发信息, 加入到 $B^{(t)}$ 中, $B^{(t)}$ 表示 t 时刻新增的信息转发者的集合, $B^{(t)} = \{v | v \in V, v \in A^{(t-1)}, e_v^{(t)} > Acp_v^{(t)}\}$, 否则不转发信息。从而, $A^{(t)} = A^{(t-1)} \cup B^{(t)}$ 。

网络群体中存在着不同个性的用户个体, 为了研究用户个性对信息转发的影响, 将群体中个体类型划分为: 自主型、无主见型和中立型 3 类, 将这 3 类个性纳入到传播模型中, 体现为信息吸引力在影响因素中所占的权重 w 的不同, 分别取

值为 0.9、0.1 和 0.5。

3 仿真分析

新浪微博属于在线社交网络。国外关于在线社交网络拓扑的实证研究表明它具有有向无标度特性, 例如 4 个受欢迎的在线社交网络 Orkut, YouTube, Flickr 和 LiveJournal 已被证实具有有向无标度特性, 其用户的入度与出度都具有幂律特性^[9]。国内同样证实新浪微博与人人网也具有相同的特性^[10]。笔者利用网络爬虫到新浪微博爬取数据, 获得了一个真实的关注微博网络边关系, 得到了具有一定拓扑结构的微博拓扑子网, 网络中的节点数 $N=5\ 906$, 平均度 $dout_average=4.5$, 最大度 $dout_max=732$, 度分布指数为 2.36, 参数 $Decay$ 设为 0.5。网络拓扑结构如图 1 所示, 网络度分布如图 2 所示。度分布函数 $P(k)$ 是指任意一个节点度为 k 的概率, 其值等于网络中度为 k 的节点个数与网络总节点个数的比值。从图 2 中可以看出, $P(k)$ 服从幂律分布, 即 $P(k) \sim k^{-\gamma}$, γ 代表度分布指数。微博拓扑子网的度分布指数符合无标度网络度分布指数介于 2 和 3 这一特性。

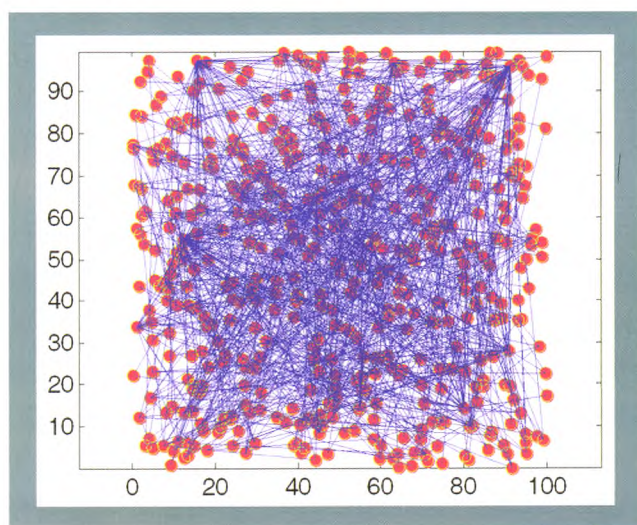


图1 网络拓扑结构

依据网络中 3 种类型的个体数量的不同比例混合, 得到 4 个不同的网络, 分别为 3 种类型个体数量均匀、自主型个体数量居多、无主见型个体数量居多、中立型个体数量居多。根据统计学中的“二八法则”, 任何一组东西中, 多数所占比例为 80%, 少数占 20%。基于此, 本文将数量居

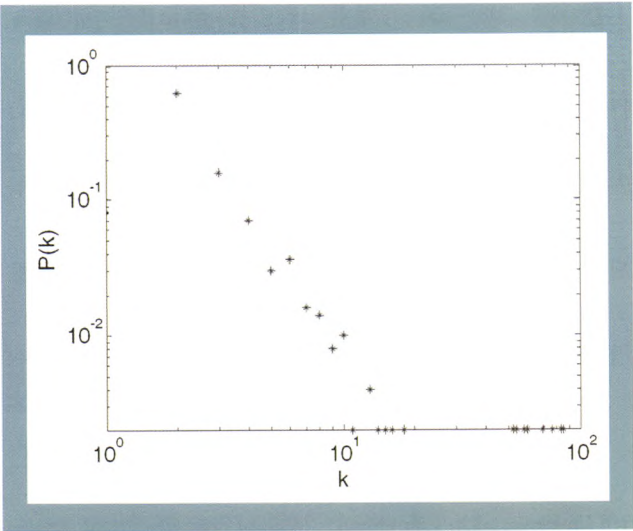


图2 网络度分布

多的个体在网络中所占比例设置为 80%，其余 2 种个体数量各占 10%，具体网络中 3 种类型个体所占比例如表 1 所示。

表1 3种类型个体在不同网络中的比例构成

不同比例个体网络 \ 个体类型	自主型	中立型	无主见型
3 种类型个体均匀	33%	34%	33%
自主型个体居多	80%	10%	10%
无主见型个体居多	10%	10%	80%
中立型个体居多	10%	80%	10%

当信息吸引力 Φ_e 取不同值时，信息在 4 个网络中传播结束时的扩散密度 ρ ($|A^{(t)}|$ 与传播空间节点集合的模的比值) 和扩散平均度 $AvgDegree$ ($A^{(t)}$ 中节点度的平均值) 与信息吸引力 Φ_e 的关系分别如图 3、图 4 所示。

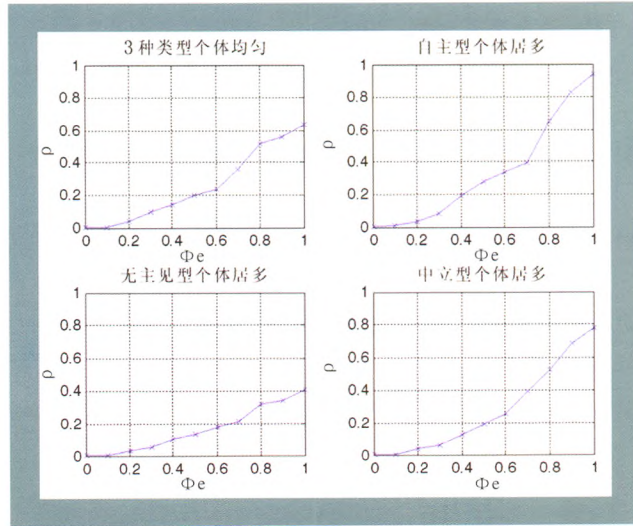


图3 4个网络中信息扩散密度与信息吸引力的关系

从 4 个网络信息扩散密度与信息吸引力的关

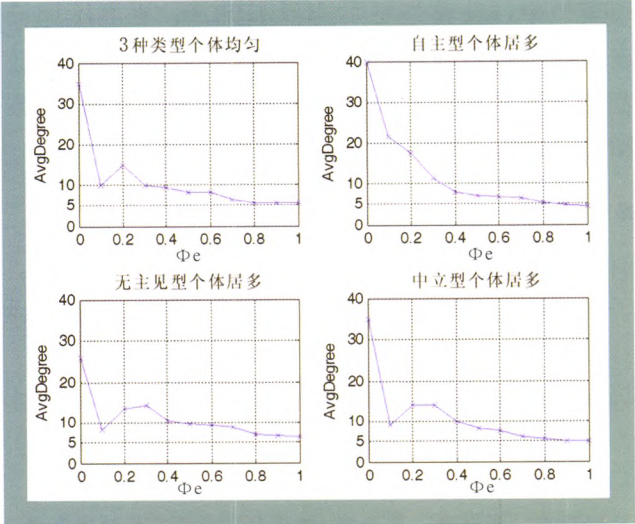


图4 4个网络中信息扩散平均度与信息吸引力的关系

系比较中，可以看到存在信息吸引力临界值（扩散密度 ≥ 0.6 的吸引力最小值），当信息吸引力大于临界值时，信息可以在网络中大规模扩散，并且由于用户个性影响因素的存在，临界值在不同类型个体居多的网络中会发生变化；信息在自主型个体居多的网络中扩散程度最高，很难在无主见型个体居多的网络中得到扩散，在中立型个体居多的网络和 3 种类型个体均匀分布的网络扩散程度相差不大。

从 4 个网络信息扩散平均度与信息吸引力的关系比较中，扩散平均度随着信息吸引力的增加逐渐趋于稳定并接近于整个网络的平均度。自主型个体居多的网络较之于其他 3 个网络，其扩散平均度随着信息吸引力的不同变化较为明显，对信息的吸引力较为敏感。

4 结束语

本文采用基于概率方式的传播影响模型来模拟微博信息的转发过程，通过在构建的传播模型中纳入用户个性这个影响因素，研究信息在不同比例混合群体构成的网络中的传播，得出结论：将用户个性作为影响因素纳入到传播模型中来模拟信息的传播过程是合理的，并且不同个性用户对信息扩散的影响作用不同，自主型个体对信息扩散具有很强的推动力，中立型个体次之，无主见型个体最差。由于实际当中微博信息的转发是一个很复杂的演化过程，其会受到各种不确定因素的影响，如何结合实际将更多的影响因素纳入

到模型中,使模型更加合理和完善,更准确地描述信息的传播过程,还需要进一步研究。相信随着更多的实证研究支持,可以为网络舆论的预测和引导提供更好的支撑。

参考文献:

- [1] 刘俊,金聪,邓清华.无标度网络环境下E-mail病毒的传播模型[J].计算机工程,2009,35(21):131-133,137.
- [2] 张彦超,刘云,张海峰,程辉,熊菲.基于在线社交网络的信息传播模型[J].物理学报,2011,60(5):1-7.
- [3] 许晓东,肖银涛,朱士瑞.微博社区的谣言传播仿真研究[J].计算机工程,2011,37(10):272-274.
- [4] 杨春霞,胡丹婷,胡森.微博病毒传播模型研究[J].计算机工程,2012,38(15):100-103.

- [5] 丁飞,刘云,司夏萌,张彦超.舆论话题的传播与竞争[J].系统仿真学报,2009,21(23):7660-7664.
- [6] 孙庆川.人际信息传播模型及其模拟[D].上海:上海大学,2009.
- [7] 刘丰.基于微博的突发事件检测和信息传播建模[D].哈尔滨:哈尔滨工业大学,2011.
- [8] 郑蕾,李生红.基于微博网络的信息传播模型[J].通信技术,2012,45(2):39-41.
- [9] Mislove A, Marcon M, Gummadi K P, et al. Measurement and Analysis of Online Social Networks[C]. Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement. [S. l.]: ACM Press, 2007.
- [10] Fu Feng, Liu Lianghuan, Wang Long. Empirical Analysis of Online Social Networks in the Age of Web 2.0[J]. Physical A, 2008, 387(2): 675-684.

责任编辑 方 圆

(上接 P53)

软件开发环境:

(1) 编译器要求: Visual C++ 6.0; (2) 操作系统: Windows XP Professional SP3; (3) 通信中间件: Wincap-4-0-1.exe, wdpack. zip; (4) 通用运行库: <ANSI stdio, stdlib, winpc-ap>。

3.2 系统的测试

在部署好环境后,运行协议分析软件,点击设置菜单下的“设置网卡”子菜单,进行网卡的设置,在分析菜单下面有“协议帧格式分析”,“协议工作流程分析”,“应用报文封装分析”和“应用报文工作流程分析”等子菜单,点击任何一个便开始不同功能的分析,在涉及到报文封装分析的部分,会生成一个名为“packet.txt”的文件,文件中包含报文封装相关的信息,前面已有所阐述,若要停止分析,可直接点击分析菜单下的“停止分析子菜单”。经过系统的测试,协议报文封装和工作流程以及应用报文封装和工作流程都符合规范的要求,从而检验了项目中所开发协议的正确性,达到了协议实现分析软件的功能

4 结束语

本文对运行在 CRH2 型列车控制网络上的 ARCNET 协议实现情况进行了分析,具体涵盖了协议本身 5 种帧格式的封装正确与否的分析以及

协议工作流程的分析,封装在协议 PAC 帧中的应用报文封装格式的分析以及不同应用报文工作流程的分析。这种分析对于验证协议的实现和协议的进一步开发以及对列车控制网络有条不紊的运行都有着重要的意义。

参考文献:

- [1] 路向阳.列车通信网络的发展与应用综述[J].机车电传动,2002(1):5-9.
- [2] 倪文波,王雪梅.高速列车网络与控制技术[M].成都:西南交通大学出版社,2011.
- [3] 管婷.TCN 实时协议和网络管理的研究与实现[D].成都:西南交通大学,2011.
- [4] 王兵兵.WTB 和 MVB 协议数据分析软件设计[D].武汉:华中科技大学,2011.
- [5] 王克举.TCN 协议一致性测试研究[D].上海:同济大学,2007.
- [6] 高渊.协议识别与分析技术[D].上海:同济大学,2007.
- [7] 中国南车株洲电力机车研究所.列车网络系统随车技师教材[Z].2008.
- [8] 张曙光.CRH2 型动车组[M].北京:中国铁道出版社,2008.

责任编辑 方 圆