

基于在线社交网络的动态消息传播模型

蒙在桥*, 傅秀芬

(广东工业大学 计算机学院, 广州 510006)

(*通信作者电子邮箱 joeman1003@gmail.com)

摘要:传统传播模型较难描述在线社交网络中的复杂活跃模式以及节点间的拓扑差异,并且其接触式的传播者退化方式也与现实不符。针对理论模型模拟与现实消息传播的不符,提出一个基于在线社交网络的动态消息传播模型 D-SIR。该模型考虑了在线社交网络中影响消息传播的一些实际因素,引入基于传播延迟的退化方式使传播者自发地退化成免疫者,动态指定节点的权威度和免疫力以适应非均质网络,并考虑接收增强信号效应以及外部社会加强效果。在采集的新浪微博真实传播网络数据中,通过参数变化的传播仿真实验验证了 D-SIR 模型可以有效反映在线社交网络的现实传播情形,并且较传统模型更具灵活性及可扩展性。

关键词:在线社交网络;消息传播模型;建模仿真;传播模型;复杂网络

中图分类号: TP391.9 **文献标志码:** A

Dynamic information spreading model based on online social network

MENG Zaiqiao*, FU Xiufen

(School of Computer, Guangdong University of Technology, Guangzhou Guangdong 510006, China)

Abstract: Traditional spreading models have difficulties in describing the complex activity patterns and the topological differences between nodes in online social networks, and the contact-based spreader annihilation mechanisms in these models do not fit with the reality. To filling the gap between spreading simulations of theoretical model and realities of information spreading, a new dynamic information spreading model (D-SIR) based on online social network was proposed. With consideration of some practical factors in information dissemination process, this model introduced the time delay annihilation mechanism that spreaders changed to stiflers spontaneously and the dynamic authority and resistance of nodes mechanism to apply to inhomogeneous networks, and considered the receiving reinforced signal effect and the social reinforcement. With the variances of parameters, the simulations on the real-world online social network which is constructed by crawled Sina microblog data verify that D-SIR model can reflect the real spreading situation in online social network. And compared to the traditional spreading model, the new model is more flexible and extensible.

Key words: online social network; information spreading model; modeling and simulation; spreading model; complex network

0 引言

近年来大规模在线社交网络的盛行给日常生活中消息和舆论的传播带来了极大的便利,也为复杂网络中的传播动力学研究提供了丰富的数据来源。由于现实的人类信息交互系统是极其复杂的,消息的传播研究一般都是先通过构建合理的消息传播模型或谣言传播模型^[1-10],然后运用这些模型仿真现实消息的传播以求获得有价值的传播机理。传播模型的研究大多数源于经典的 SIS (Susceptible-Infected-Susceptible) 和 SIR (Susceptible-Infective-Removal) 传染病传播模型,比如较早前的 DK (Daley Kendal) 和 MK (Maki Thomson) 谣言传播模型^[1]。后来 Moreno 等^[2]、Nekovee 等^[3]发展了 Daley 等提出的 DK 谣言传播模型,并引入了该模型在均匀网络及非均匀网络中的均场方程。然而这些传统的传播模型都过于简单,其理论上的一些限制条件往往很难应用于实际,尤其是很难描述消息在大规模在线社交网络中的传播^[1,4]。因此,许

多学者在传统模型的基础上纳入了一些传播中的实际因素,以减小理论模型仿真与实际传播间的差距。比如, Zhao 等^[1,5]在 SIR 的基础上拓展出了一类节点 Hibernators,提出基于回忆机制和遗忘机制的 SIHR (Susceptible-Infected-Hibernator-Removed) 模型;Lyu 等^[6]在其消息传播模型中考虑到多次感染以及外部社会因素的加强效果;文献[7]在传播模型中引入传播者节点以一定速度自发变成免疫者的机制;王辉等^[8]在其提出的 CSR (Credulous-Spreader-Rationals) 谣言传播模型中加入记忆效应和接受阈值,使其更适用于移动社交网络中谣言的传播仿真;Zhao 等^[9]提出基于指数遗忘函数的可变遗忘率谣言传播模型。但是,这些模型大都没有考虑社交网络的在线特性以及不同节点间的拓扑特征差异。

为使传播模型中的仿真与真实社交网络中的传播情形更加吻合,本文首先分析了传统传播模型在社交网络中仿真消息传播的不足,然后提出一种基于在线社交网络的动态消息传播模型。该模型采用基于传播延迟的传播者自行退化机制,可根

收稿日期:2014-01-28;修回日期:2014-02-27。 基金项目:广东省自然科学基金资助项目(10451009001004804)。

作者简介:蒙在桥(1988-),男,江西赣州人,硕士研究生,主要研究方向:复杂网络、数据挖掘;傅秀芬(1957-),女,福建漳州人,教授,主要研究方向:网络多媒体、协同软件、网络安全。

据不同用户的拓扑差异动态指定权威度和免疫力,并且还考虑了接收增强信号机制以及社会加强效应。最后运用该模型在真实的微博传播网络中仿真传播以验证模型的有效性。

1 传统模型的分析

传统的传播模型通常会应用动力学中常用的“平均场”假设,其思想认为网络中相互作用的整体效果等价于一个“平均场”^[8],它抛开个体之间具体的细节,假设网络中的个体都有一致的平均传染可能性和免疫力(或称治愈率)。这种假设在均匀网络中是合理的,因为均匀网络中可以忽略节点间的拓扑差异,令每个节点度为网络的平均度,传播者也会在相同的接触数量中慢慢退化。但社交网络是一个不规则的非均匀无标度网络,节点的度分布一般都呈现出幂律分布^[11],优先连接的特性使节点的度呈现出“富者愈富”的效应,节点之间存在着较大的拓扑差异。在实际的社交网络中,普通用户与明星用户的影响力或权威度显然都是不同的,因此在构建传播模型时需考虑网络非均质的影响^[12]。

现有传播模型中大多数都是以接触的方式传播消息或者谣言,传播者则会在接触的过程中退化成无知者或免疫者。在较为经典的谣言传播模型^[2-3]中,传播者是在接触其他传播者或免疫者时以某个固定概率 α 变为免疫者。然而在实际的传播演化过程中,这种接触的传播者退化方式导致了一些网络的中心节点较快退化成免疫者而中止传播^[13]。以经典谣言传播模型为例,社交网络的无标度特性使模型不得不脱离均匀假设,传播者 i 的每次传播都将接触其所有的出度邻居,那么 i 在第 t 时刻成为免疫者的概率为:

$$p_{sr}^i(t) = 1 - (1 - \alpha)^{g(i,t)} \quad (1)$$

其中: $\alpha \in (0, 1]$ 为模型中一固定概率值, $g(i, t)$ 代表在 t 时刻 i 的出度邻居中传播者和免疫者的数量。在同一时刻出度越大的节点接触网络中传播者或免疫者的可能性越大,那么该节点退化成免疫者的概率 $p_{sr}^i(t)$ 也越大。这意味着度高的节点比度小的节点更快退化成免疫者而停止传播,这种现象显然有悖于现实的情况。因此,这种退化方式在仿真节点传播状态退化时缺乏灵活性,许多模型已经开始考虑应用传播者自发的退化方式^[1,7,9]。

2 在线社交网络消息传播的分析与建模

2.1 在线社交网络的活跃模式

社交网络中的个体都呈现出一种复杂的活跃模式^[12]。在社交网络中,用户并不是一直都能保持活跃在线,只有处于在线状态的用户才能接收到消息。为了模拟现实社交网络中的这种在线的活跃行为,本文引入在线概率 θ ,即用户在线并且能接收消息的概率。尽管在社交网络中每个人在线的概率完全不同,为简化模型,本文设 θ 为常数,用于反映网络的整体在线活跃度。

在线社交网络中不同节点的拓扑差异导致节点之间的用户活跃性也不同,很多用户接收到消息,但愿意转发使其传播的可能性很小^[4],通常粉丝量越多的用户越不轻易转发别人的消息。因此,本文在传播模型中赋予节点对消息不同的免疫力。直观上来说,具有更高权威度的文档消息往往会更有可能会引入更多的原创内容^[14],较高权威度的传播者发布的消息往往具有更高的可信度,而该消息也更可能被传播。那么

未知者接收消息后,不仅要考虑自身的免疫力还应考虑消息传播者的权威度。在社交网络中,接收更多来自邻居传来的增强信号可以使用户更愿意接受某个行为^[15],接收到不同传播者消息的信号次数和传播者的权重是未知者做判断的重要依据^[8]。因此,本文在构建传播模型时动态指定不同节点的权威度和免疫力,并考虑接收信号的增强效应。

2.2 基于在线社交网络的动态消息传播模型

结合上述的分析,本文提出基于在线社交网络的动态传播模型(D-SIR)。该模型考虑单条消息在网络中的传播仿真,将网络中节点分为三类角色:未知者、传播者和免疫者。未知者是指未接收到消息的人,传播者是指消息的传播者,免疫者是指收到消息后对消息不感兴趣或者传播者经过一段时间而失去传播能力的人。D-SIR的传播仿真过程是基于离散时间的模型,将一次传播迭代作为一个离散时间片,从而模拟消息在社交网络中随时间变化的传播演变过程。

初始时,先将所有节点的状态都设为未知者,并根据出度动态指定每个节点的权威度和免疫力。然后根据相应的选择策略选择最初的传播源,将所选节点的状态置为传播者。最后根据传播规则迭代传播直至网络中的传播者都退化成免疫者。将每一次传播迭代分为2步:第一是广播步,网络中的所有传播者向邻居广播消息,其出度的在线未知者邻居以在线概率 θ 接收该消息,并记录传播者存活时间 t_s ,传播者再根据退化机制退化成免疫者;第二是响应步,接收到消息的在线未知者根据自身的免疫力、接收消息的次数、传播者的权威度以及外部社会加强综合计算被感染的概率,如被感染则成为传播者,否则成为免疫者。

下面详细介绍模型中传播者的退化机制、动态指定节点权威度和免疫力的方法以及模型的迭代仿真过程。

2.2.1 传播者的退化机制

研究发现,在疾病传播网络中传播延迟的存在对传播会产生影响,在传播模型中加入传播的延迟是合理的^[10]。传播延迟是指传播者接触其他个体后不立即转换状态而保持传播者状态一段时间。因为传播者在发出消息时未知者可能不在线,但它可以在线经过一段时间上线后继续接收消息,因此应用传播延迟更符合现实在线社交网络的消息传播。我们将传统接触退化方式修改为基于传播延迟的自发退化方式,并引用Parato分布函数^[16]作为传播者的退化概率函数。

$$\eta(t_s) = 1 - (t_s/t_{\min})^{-\alpha+1} \quad (2)$$

其中: $t_s \geq 1$ 是指传播者的存活时间, $\alpha \geq 1, t_{\min} \geq 1$ 作为参数决定了函数的形状。从图1中可以看出:退化概率 $\eta(t_s)$ 随着 t_s 的增加而增大;在固定 t_{\min} 时,参数 α 的值越大, $\eta(t_s)$ 上升的速率越快,该参数可以视为是传播者的退化速率,反映消息被其他消息淹没的速度;固定 α 增大 t_{\min} 时函数图像向右平移,因此 t_{\min} 定义为传播者存活时间的最小值,即传播者在经历了 t_{\min} 个时间片后才开始退化。运用该函数可以较灵活地调节传播者的最小存活时间及退化速率。不失一般性,在本文实验取 t_{\min} 的最小值1。

2.2.2 动态的权威度和免疫力

结合2.1节中对用户活跃模式的分析并考虑网络非均质的影响^[12],分别定义 $d_a(i)$ 、 $d_r(i)$ 分别为节点 i 的权威度和免疫力,两者值相同且都服从出度相关的对数函数:

$$d_a(i) = d_r(i) = \log_{\mu}(k_i + 1) \quad (3)$$

其中 k_i 为 i 的出度,在微博中即为粉丝数量,该函数的作用是将出度的较大数量级映射在一个较小的数量级内以衡量用户权威度和免疫力的不同等级。本文中令 $\mu = 10$,因在该对数下,幂律分布图像呈现出一条近似直线,这样可将用户出度按递增顺序划分成数量呈线性递减的不同等级。在消息的响应步,未知者会接收到多个传播者的信号,为体现接收多信号的增强效应,定义接收消息的总剂量 $\delta(i)$ 为消息传播者的权威度之和:

$$\delta(i) = \sum_{j \in ajs(i)} d_a(j) \quad (4)$$

其中 $ajs(i)$ 为用户 i 的入度邻居传播者集合。在消息的响应步,定义未知者接收消息后成为传播者的概率为:

$$p_{is}(i) = \theta(1 - (1 - \lambda)^{\frac{\delta(i)}{B}}) \quad (5)$$

其中, λ 为消息的传染率,该参数反映消息本身的影响力而与网络无关。研究表明外部社会加强对传播有重要影响^[6, 15],因此令 B 作为外部社会加强。根据式(5)可以看出:当 $B \in (0, 1)$ 时,对 $p_{is}(i)$ 有负加强的效果; $B \in (1, +\infty)$ 时有正加强的效果,本文实验暂不考虑社会加强效果,因此令 $B = 1$ 。

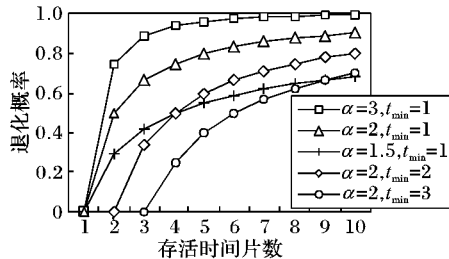


图1 退化函数随参数的变化

2.2.3 迭代仿真过程

在每一次的迭代传播过程中,网络中的所有传播者都会向其出度邻居广播消息并记录自己的存活时间 t_s 。当这些邻居中的未知者以概率 θ 在线时,表示该用户查阅了消息,将以 $p_{is}(i)$ 的概率成为传播者,否则成为免疫者;不在线的未知者将保持未知者状态,并在本次迭代结束前将消息的接收剂量 $\delta(i)$ 置0,这样是为了防止消息的接收冗余^[6, 15]。而传播者则根据存活时间 t_s 以及退化机制逐渐退化成免疫者。迭代过程中当传播者都退化成免疫者时传播结束。最终的免疫者数量可反映消息的传播范围,传播者数量的最大值可反映参与传播的人数,迭代的次数可反映传播结束时所用的时间。D-SIR 仿真传播中节点的状态转移过程如图2所示。

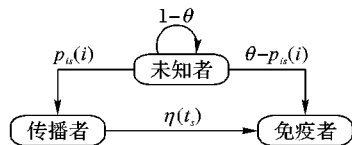


图2 D-SIR 中节点的状态转移

3 实验与评估

3.1 数据集描述

新浪微博,是一个消息有向传播的在线社交网络,并且具有无标度和小世界的网络特性,其节点的度服从幂律分布。本文使用网络爬虫从用户姚晨出发爬取三层以内的新浪微博粉丝和关注关系数据,然后将数据中的用户作为网络的节点,粉丝及关注关系作为网络的边,最终构成了实验的有向传播网络,其网络的基本特征如表1所示。

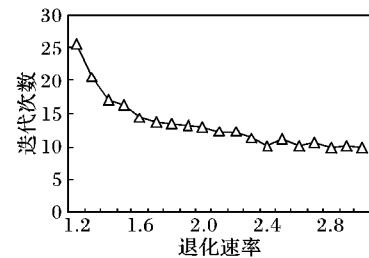
表1 采集微博传播网络的基本拓扑特征

节点数	边数	平均度	平均路径长度	网络直径	平均聚类系数
63 985	3 301 712	103.20	5.92	12	0.047

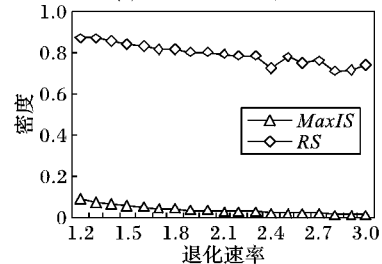
根据该网络的平均路径长度、网络直径和平均聚类系数可以看出,该网络符合小世界网络的特征。

3.2 仿真结果

为了验证模型各个参数对仿真传播的影响,本文根据 D-SIR 模型用 Java 编写了仿真程序,在真实微博网络中仿真消息的传播并记录传播过程中传播者密度的最大值 $MaxIS$ 、传播结束时免疫者的密度 RS 以及迭代次数 $ItNum$ 作为仿真传播的评估量。选择最大出度的 10 个用户作为消息的初始传播源,每组实验仿真 20 次并求取平均结果,最终实验结果如图3~6所示。



(a) 对 $ItNum$ 的影响



(b) 对 RS 及 $MaxIS$ 的影响

图3 退化速率 α 对传播的影响

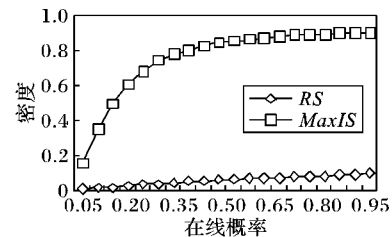


图4 在线概率 θ 对传播的影响

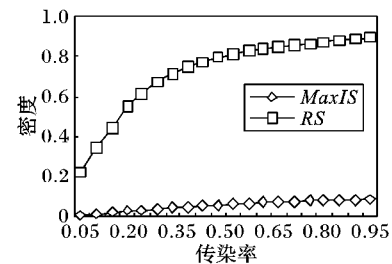


图5 传染率 λ 对传播的影响

图3为不同退化速率 α 下的传播结果,可以看出,其他参数不变时($\theta = 0.5$, $\lambda = 0.1$)随着 α 增大,传播的迭代次数减小(如图3(a)),并且 RS 、 $MaxIS$ 也随之略微下降(如图3(b)),当 α 值大于 2.7 时传播效果趋于稳定。说明传播者退化得越慢,消息可以传播得越久越广。这个现象说明传

播的延迟会对传播造成影响^[10],事实上,谣言被其他消息淹没的速度越快,其能够影响的人群就越小,但由于在线社交网络中消息的响应速度较快,这个影响能力并不大。

从图4中可以看出,在线概率 θ 越高(固定 $\lambda = 0.1$, $\alpha = 1.5$), RS 和 $MaxIS$ 都越大,意味着在线活跃度越高的社交网络越容易使消息扩散开来。从图5可以看出(固定 $\theta = 0.5$, $\alpha = 1.5$),消息的传染率 λ 也与 RS 和 $MaxIS$ 正向同步,意味着影响力越大的消息越容易在网络中普遍传播。图6显示了运用D-SIR仿真传播时未知者密度 SS 、传播者密度 IS 以及免疫者密度 RS 随着时间片 t 改变时的变化($\theta = 0.5$, $\alpha = 1.5$, $\lambda = 0.1$)。可以看出:该传播过程符合真实传播中急剧上升再缓慢结束的特征;并且随着代表不同物理意义的参数变化,传播结果的变化皆符合实际传播中的现象。因此,本文的D-SIR模型可以较好地描述在线社交网络中的消息传播。

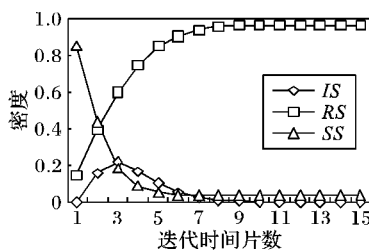


图6 D-SIR模型仿真传播过程

4 结语

传统传播模型过于简单,难于描述在线社交网络中的复杂活跃模式以及节点间的差异,因此提出一种基于在线社交网络的动态消息传播模型D-SIR。该模型针对接触退化方式的不足,引入一个退化函数使传播者自发地退化成免疫者,并且函数的参数可以灵活调节传播者的退化速率及最小存活时间。考虑到社交网络的无标度特性,模型根据节点出度动态指定节点的权威度和免疫力,并且该模型还可以扩展外部社会加强的影响因素。运用D-SIR模型在真实微博传播网络下进行仿真实验,仿真结果验证了模型可以有效反映在线社交网络中消息的传播情形。该模型也为进一步研究在线社交网络中消息、谣言等的传播机理打下了基础。

参考文献:

- [1] ZHAO L, WANG J, CHEN Y, *et al.* SIHR rumor spreading model in social networks[J]. *Physica A*, 2012, 391(7): 2444–2453.
- [2] MORENO Y, NEKOVEE M, PACHECO A F. Dynamics of rumor spreading in complex networks[J]. *Physical Review E*, 2004, 69(6): 066130.
- [3] NEKOVEE M, MORENO Y, BIANCONI G, *et al.* Theory of rumour spreading in complex social networks[J]. *Physica A*, 2007, 374(1): 457–70.
- [4] BORGE-HOLTHOEFER J, MELONI S, GONALVES B, *et al.* Emergence of influential spreaders in modified rumor models[J]. *Journal of Statistical Physics*, 2013, 151(1/2): 383–393.
- [5] ZHAO L, QIU X, WANG X, *et al.* Rumor spreading model considering forgetting and remembering mechanisms in inhomogeneous networks[J]. *Physica A*, 2012, 392(4): 987–994.
- [6] LYU L, CHEN D B, ZHOU T. The small world yields the most effective information spreading[J]. *New Journal of Physics*, 2011, 13(12): 123005.
- [7] ZHANG Y, LIU Y, ZHANG H, *et al.* The research of information dissemination model on online social network[J]. *Acta Physica Sinica*, 2011, 60(5): 50501. (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): 50501.)
- [8] WANG H, HAN J, DENG L, *et al.* Dynamics of rumor spreading in mobile social networks[J]. *Acta Physica Sinica*, 2013, 62(11): 110505. (王辉, 韩江洪, 邓林, 等. 基于移动社交网络的谣言传播动力学研究[J]. *物理学报*, 2013, 62(11): 110505.)
- [9] ZHAO L, XIE W, GAO H O, *et al.* A rumor spreading model with variable forgetting rate[J]. *Physica A*, 2013, 392(23): 6146–6154.
- [10] XU X J, CHEN G. The SIS model with time delay on complex networks[J]. *International Journal of Bifurcation and Chaos*, 2009, 19(2): 623–628.
- [11] 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*. New York: ACM, 2007: 29–42.
- [12] PERRA N, GONALVES B, PASTOR-SATORRAS R, *et al.* Activity driven modeling of time varying networks[J]. *Scientific Reports*, 2012, 2: Article no. 469.
- [13] BORGE-HOLTHOEFER J, MORENO Y. Absence of influential spreaders in rumor dynamics[J]. *Physical Review E*, 2012, 85(2): 026116.
- [14] LIN C X, MEI Q, HAN J, *et al.* The joint inference of topic diffusion and evolution in social communities[C]// *Proceedings of the 11th IEEE International Conference on Data Mining*. Piscataway: IEEE, 2011: 378–387.
- [15] CENTOLA D. The spread of behavior in an online social network experiment[J]. *Science*, 2010, 329(5996): 1194–1197.
- [16] NEWMAN M E. Power laws, Pareto distributions and Zipf's law[J]. *Contemporary Physics*, 2005, 46(5): 323–351.
- [17] LUO H, YE F, CHENG J, *et al.* TTDD: two-tier data dissemination in large-scale wireless sensor networks[J]. *Elsevier/ACM Wireless Networks*, 2005, 11(1/2): 161–175.
- [18] KIM H S, ABDELZAHER T F, KWON W H. Minimum energy asynchronous dissemination to mobile sinks in wireless sensor networks[C]// *Proceedings of the 1st International Conference on Embedded Networked Sensor Systems*. New York: ACM, 2003: 193–204.
- [19] YE Y. An $O(n^3L)$ potential reduction algorithm for linear programming[J]. *Mathematical Programming*, 1991, 50(1/2/3): 239–258.

(上接第1946页)

- [15] ZHANG X, DAI H, XU L, *et al.* Mobility-assisted data gathering strategies in WSNs[J]. *Journal of Software*, 2013, 24(2): 198–214. (张希伟, 戴海鹏, 徐力杰, 等. 无线传感器网络中移动协助的数据收集策略[J]. *软件学报*, 2013, 24(2): 198–214.)
- [16] LI B, LIN Y, ZHOU S, *et al.* Efficient data collection algorithm in sensor networks with optimal-path mobile sink[J]. *Journal of Computer Applications*, 2011, 31(10): 2625–2629. (李彬, 林亚平, 周四望, 等. 传感器网络中基于移动Sink最优穿越路径的高效数据收集算法[J]. *计算机应用*, 2011, 31(10): 2625–2629.)