

文章编号: 2095-2163(2020)02-0278-06

中图分类号: TP393.0

文献标志码: A

基于记忆效应的社交网络信息传播模型

王晓萌, 方滨兴, 张宏莉, 王星

(哈尔滨工业大学 计算机科学与技术学院, 哈尔滨 150001)

摘要: 在线社交网络已成为当代社会信息交互的重要集散地,其信息传播日益活跃。为了探究社交网络中的信息传播规律,本文通过引入信息曝光度曲线与社交网络个体影响力,提出了一种基于信息转发传播过程中记忆效应的信息传播模型。在规则网络、随机网络与人人网数据集上进行仿真,研究发现谣言、爱好等具有兴趣累积型记忆效应的信息更容易在规模较大的规则网络上传播,且这种趋势会随着网络规模增大而更加显著,符合社交网络特性。而具有时效衰减性质的信息,在随机网络上传播的效果更好,这对 Centola 的行为传播实验结论提供了补充。

关键词: 社交网络; 信息传播; 记忆效应; 曝光度曲线

Memory effect based information diffusion model through online social networks

WANG Xiaomeng, FANG Binxing, ZHANG Hongli, WANG Xing

(School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China)

【Abstract】 Information diffusion is one of the major research focuses in Social Network Analysis, the information diffusion is the core content in this field. This paper presents a memory effect based information diffusion model, where the exposure curve and individual influence are both taken into consideration to improve the epidemic model. Through simulation experiments based on regular network, random network and Renren data, it is found that rumors, hobbies, and other types of information which has cumulative memory effect is more easily to spread in regular network with large size, and this trend becomes more significant as the size of the network increases. However, the random network performs better for the attenuation memory effect, and this result provides a supplement to the Centola's experiment.

【Key words】 social network; information diffusion; memory effect; exposure curve

0 引言

在线社交网络与生俱来的自由性和开放性,使其逐渐成为当代社会信息传播的重要集散地,社交网络中的信息活跃性达到了前所未有的程度。随着国内外大量在线社交网络服务的涌现以及用户的参与,针对社交网络信息传播分析的相关研究引起了国内外学者的高度关注。研究社交网络中的信息传播规律,建立传播模型,即能从信息传播的角度对网络结构、属性以及突发事件遵循的规律有进一步的认识。

信息传播模型是信息传播研究内容的核心,主要任务是分析信息传播过程中的影响因素,理解、模拟并验证扩散过程。信息传播与疾病、文化和行为的传播是人类社会中常见的现象,很多研究者多借鉴病毒传播模型^[1-2]来研究在线社交网络中的谣言传播^[3-4]。传统的传染病传播模型中的状态表征存在不足,个体只有感染、易感和免疫三种状态,许多扩

展模型并没有充分考虑信息传播的内在机理与用户的影响力,传播模拟实验结论也没有给出拓扑结构对传播机理影响的合理解释。信息传播具有记忆效应(memory effect)^[5],记忆效应是指人对于同类信息接触所具有的记忆上的累积特性,即同一类信息的多次冗余接触会改变人们对其初始看法,这种累积特性会对社交网络中用户的转发行为产生影响。部分在线社交网络上的实证研究^[6]已经验证的记忆效应在社交网络信息传播过程中存在且具有规律性。

以往的传播模型大都没有考虑记忆效应这种行为对传播的影响或考虑不全面。本文中,将同时考虑传播过程中的记忆效应的兴趣累积与时效衰减两方面,还将兼顾社交网络中用户影响力,对在线社交网络中的信息传播行为进行详细的理论建模与数值仿真,有助于深入理解社交网络中的传播行为,为舆情研究提供理论基础。

作者简介: 王晓萌(1987-),男,博士研究生,主要研究方向:在线社交网络、信息传播预测、舆情安全等;方滨兴(1960-),男,博士,教授,博士生导师,主要研究方向:并行计算、网络安全、拓扑发现等;张宏莉(1973-),女,博士,教授,博士生导师,主要研究方向:网络与信息安全、网络测量与建模、网络计算等;王星(1981-),男,博士,助理研究员,主要研究方向:网络与信息安全、网络舆情监控、知识迁移。

收稿日期: 2019-03-26

1 相关工作

社交网络中记忆效应的研究最早来自于 Centola 的邮件推荐实验^[5], Centola 将 1 528 名的注册用户以匿名的方式分配到度相同的规则网络(最近邻耦合)与随机网络^[7]中, 每个用户通过邮件的方式对邻居用户重复推荐一个健康社区网站, 当个体接收到多次邮件推荐后, 则更有可能接收该邮件的推荐内容并进行转发。而在此之前, 传统结论一般认为具有短平均路径的随机网络更有利于信息的传播, 但在线社交网络存在记忆效应, 与此同时还发现相比于平均路径较短的随机网络, 具有高聚集系数的规则网络更有利于行为的传播与扩散。

随后, Romero 等人^[6]对大量 Twitter 数据进行统计分类, 发现不同类型话题传播过程中记忆效应的的巅峰概率与接受信息次数的关联关系。文献^[8]对海量新浪微博数据进行统计, 发现不同类型的微博在转发过程中所表征出来的记忆效应曲线存在差异, 冗余的信息刺激可能会降低微博的转发概率。Myers 等人^[9]认为社交网络用户影响力可以近似拟合为一个与信息累积接受次数相关的曝光曲线(exposure curve)。文献^[10]建立了一个兴趣积累的记忆效应传播模型, 并讨论该模型在小世界网络上的传播能力, 但该模型只考虑了记忆效应带来的概率累积一种情况。

相对于之前的研究, 本文将整合上述研究者的工作, 探究符合记忆效应规律的信息传播特征, 借鉴传染病模型, 重新划分社交网络上节点的状态集, 给出传播动力学方程并在模拟网络与真实网络上进行验证分析。

2 传播机理与模型

在线社交网络中, 用户发布或分享的信息会传递给该用户的好友, 其好友会依据兴趣程度、是否可信、新鲜程度以及传播源的影响力等因素以一定概率对信息进行转发扩散。然而信息多次累积会产生记忆效应, 对用户的转发行为产生影响, 研究中给出了 2 个过程的表征含义可分述如下。

(1) 兴趣累积。在社交网络中, 爱好、谣言以及一些信息真实性辨识度不高的信息在传播过程中一般表现出兴趣累积的记忆效应, 即接收到同类信息刺激越多, 用户越易产生转发行为。如图 1 中所示“事件一”^[8]的传播概率随着被推荐次数的累积而增加, 近似呈现指数式递增。

(2) 时效衰减。即时性强的突发新闻, 社交网络用户越少接触, 则转发的意愿越强烈。若社交网络用户在某一时刻收到多次转发, 则用户的兴趣度就会随时间衰减, 不易再产生转发意图。如图 1 中即时性很强的新闻“事件二”^[8], 转发该条微博的用户一般初次收到就转发, 随接触次数增多转发概率便逐步递减。

本节将首先提出一种涵盖以上两种记忆效应表征的统一模型, 兼具兴趣累加与时效衰减两种特点。该模型中, 信息要经过兴趣累积过程, 传播概率在多次接触后会达到峰值, 而当经历了时效衰减过程, 传播概率逐渐降低, 概率函数性质符合图 1 中曝光度曲线。在此基础上将在线社交网络节点划分为 4 种状态, 其状态间的转移概率以及传播过程如图 2 所示。本节中涉及到的缩写和符号的含义见表 1。

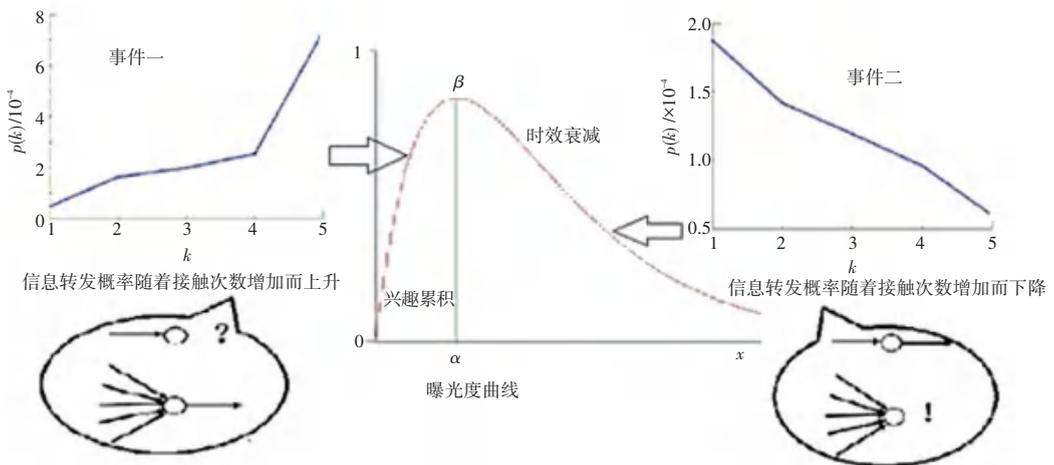


图 1 新浪微博中信息传播的记忆效应

Fig. 1 The memory effect of information transmission in Sina Weibo

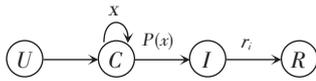


图2 状态转移示意图

Fig. 2 State transition diagram

表1 相关符号含义

Tab. 1 Related symbol meaning

符号(缩写)	含义
G	社交网络拓扑
V	整体拓扑点集
E	整体拓扑边集
$P(x)$	收到信息第 x 次后转发的概率
α	累积效应峰值
β	传播转发过程中所能达到的巅峰概率
r_i	节点 i 由传播态转为疲劳态的概率
C_i	节点 i 的个体影响力强度
$\text{deg}(i)$	节点 i 的度
R	终态时网络中疲劳态的数量
b	影响力差异系数

2.1 节点状态

社交网络形式化描述为 $G = (V, E)$, 其中 G 代表社交网络, V 是点的集合, 表示社交网络上的用户, E 为边的集合(无向边), 表示用户间的好友关系。借鉴 SIR 模型的思想, 将社交网络中节点的类型拓展为 4 种状态, 每一时间步, 每一个体处于如下 4 种状态之一:

(1) 未知态 U (Unknown)。不知道消息的人群, 类似于 SIR 模型中的易感人群。

(2) 积蓄态 C (Cumulative)。用户听到了这个信息, 但是由于不确定信息的准确性或者已经变得不感兴趣而不愿意传播。

(3) 传播态 I (Infected)。个体确认了该消息并将该信息扩散给其邻居。

(4) 疲劳态 R (Resistant)。个体传播了信息后失去继续传播该类话题兴趣的人群, 相当于 SIR 模型中的免疫态。

2.2 状态间迁移概率

(1) 传播概率。假设积蓄态节点转变为传播态节点的概率服从如图 1 所示的曝光度曲线, 曲线方程定义如下:

$$P(x) = \frac{\beta}{\alpha} x \cdot \exp\left(1 - \frac{x}{\alpha}\right), \quad (1)$$

其中, x 为截止到 t 时刻, 个体累计接收到的信息次数, 反映了信息传播的记忆性特征; $P(x)$ 为用

户第 x 次接收到其邻居转发的信息后选择转发的概率; $\beta \in (0, 1]$ 为该类型信息最流行时的传播概率峰值, 称之为巅峰概率; $\alpha \in N^*$ 为 $P(x) = \beta$ 时的转发次数, 称 α 为累积效应峰值。

当累积峰值 $\alpha = 1$ 时, 所对应的信息类型为具有时效衰减性质的即时新闻, 接触次数越多用户越易失去转发兴趣, 传播概率方程退化为式(2):

$$P(x) = \beta x \cdot \exp(1 - x), \quad (2)$$

由式(2)可知, 在 $[1, \infty]$ 上单调递减。

(2) 恢复概率。社交网络信息传播过程中的另一个重要因素就是用户影响力, 影响力大的用户发布的信息往往具有较长的持续时效。影响力越大, 用户对其邻居施加的持续影响力越大, 相对应模型中传播态变为疲劳态的时间就越长, 其函数关系如式(3)所示:

$$r_i = \exp(-b \cdot C_i), \quad (3)$$

其中, r_i 为节点 i 由传播态变为疲劳态的转移概率, $b \in [0, \infty)$ 为影响力差异系数, 用于刻画不同用户节点间影响力差异程度。当 $b = 0$ 时, 用户间影响力不存在差异, $r_i = \exp(0) = 1$, 传播态节点只对邻居节点转发一次信息便失效, 每条边只使用一次, 类似于信息快速更新的微博网络。 C_i 为节点 i 的影响力强度, 这里研究采用直接反映社交网络用户人际关系的点度中心度 (degree centrality) 来表示节点直接影响力; $\text{deg}(i)$ 为节点 i 的度, 则节点 i 的影响力强度如式(4)所示:

$$C_i = \text{deg}(i), \quad (4)$$

将式(3)代入式(2)得:

$$r_i = \exp(-b \cdot \text{deg}(i)). \quad (5)$$

2.3 传播建模

模型中, 假定传播过程开始前网络上所有节点均处于未知状态。当社交网络上的用户发布或转发信息的同时, 会扩散给其所有的好友, 即随机选取一个种子节点, 这个种子节点在每次给其所有的邻居发送信息后都以 r_i 的概率置为疲劳态。在每个时间步内, 如果一个节点(未知态或积蓄态)收到信息, 都将立刻变为积蓄态, 并以概率 $P(x)$ 变为传播态, 这里 x 是指个体已经接收到信息的次数。当 $x \leq \alpha$ 时, $P(x)$ 的值随信息接收次数 x 的增加而单调递增, 并在 $x = \alpha$ 时达到最大。当 $x > \alpha$ 时, $P(x)$ 的值随信息接收次数 x 的增加而单调下降。因此, 转发概率的大小不仅依赖接收信息的次数, 也取决于信息本身的特性。如果一个节点在某个时间步内转变为传播态, 就将在下一个时间步把信息发送其所

有的邻居,同时以概率 r_i 变为疲劳态。如果一个处在积蓄态的节点,在当前的时间步没有接收到任何信号,无论已经接受到信息的次数 x 为多大,都将不会发生任何状态上的改变。详细步骤如下。

输入: 网络拓扑 $G = (V, E)$ 、巅峰概率 β 、累积峰值 α

输出: 终态的疲劳态节点集 R

Step 1 将所有 N 个节点初始化为未知态。

Step 2 从 N 个未知态节点中随机选取一个节点置为传播态。

Step 3 所有传播态节点向其所有邻居转发消息,这些邻居节点中处于未知态的节点变为积蓄态,曝光度 $x = 1$;积蓄态个体曝光度 $x = x + 1$ 。

Step 4 检查所有积蓄态节点,若随机数 $random[0, 1] < p(x)$,则该节点置为传播态。

Step 5 检查所有传播态节点,若随机数 $random[0, 1] < r(k)$, k 为该节点的度,则该节点置为疲劳态。

Step 6 循环 Step 3 ~ Step 5,直至网络中不存在传播态节点。

3 仿真实验

3.1 数据集

在线社交网络中存在许多高聚集性的社区,朋

友圈等。社区内部紧密相连,类似规则网络。而不同的社区间又通过一些共享用户随机地链接在一起。这使得在线社交网络兼具高聚集系数与短平均路径。已有模型如 BA 模型^[11]、WS 模型^[12]及一些拓展模型,这些模型虽然宏观上表述了在线社交网络的小世界现象与无标度特性,但都不能完全表征在线社交网络的拓扑特征。Centola 的电子邮件实验从宏观的角度说明了相比于具有较短平均路径的随机网络,高聚集系数的规则网络更有利于行为在人类社会中的传播。为了验证这一结论,研究将首先在规则网络与随机网络两种均匀网络拓扑上验证本文模型的传播演变过程。

此外,本文选取国内较流行的大学生社交网络人人网的用户数据作为非均匀网络数据,该网络上的节点为注册用户,用户间的好友关系表示为边。研究建立了一个基于网页解析方式的爬虫程序,首先选取多个同一所学校的用户作为种子节点,递归地爬取每个种子用户的好友关系,及其好友的好友关系。然后将爬取得到的多个好友网络进行拼接与去重,并且对不属于种子节点注册学校的其它节点以及边的关系进行删除,最终得到一个包含 8 102 个节点、45 776 条边的最大联通子图。各网络的拓扑参数见表 2。

表 2 各网络的结构参数

Tab. 2 Structural parameters of each network

数据集	网络类型	节点数	边数	平均度	聚集系数	平均路径长度
随机网络	均匀网络	100~8 102	550~44 560	11.0	$\langle k \rangle / N$	$L \propto \ln(N) / \ln(\langle k \rangle)$
规则网络	均匀网络	100~8 102	550~44 561	11	0.675	$N/2k$
人人网	非均匀网络	8 102	45 776	11.3	0.133	4.7

3.2 仿真结果

分别选取节点数 $N = 8 000$,平均度 $k = 11$ 的规则网络与随机网络,影响力差异系数 $b = 0.2$,迭代次数 $T = 500$ 。图 3 中列出了随着巅峰概率 β 取不同值时,规则网络与随机网络上疲劳态节点的比例随累积峰值 α 变化的情况,对每组数据的结果都是进行 500 次独立重复模拟求均值得到的。当累积峰值 $\alpha = 1$ 时,对应具有时效衰减类型记忆效应的即时新闻,由图 3(a) ~ (c) 可以看出,当 $\alpha = 1, \beta$ 取 0.1, 0.2, 0.3 时,随机网络的 R 值均略大于规则网络,这是由于规则网络的聚集性强,时效衰减的效果更容易积累,阻碍了传播。同时,研究还可以观察到,随着峰值概率 β 的上升,规则网络与随机网络的扩散规模均呈现增加趋势。当 $\beta = 0.8$ 时,图 3(d) 中规则网络与随机网络的传播规模都几乎覆盖了全部节

点。由此可以认为,当峰值概率的值上升到一定程度时,聚集系数与平均路径长度这两个网络拓扑特征已不能左右信息的传播,信息都将扩散至全部网络。

疲劳态节点密度差值比例与 α 的变化关系如图 4 所示。由图 4 可见,当 $\beta = 0.1, N = 8 000$,累积效应峰值 $\alpha > 1$ 时,规则网络的传播范围 R_{random} 均大于随机网络的传播范围 $R_{regular}$ 。联合图 3(a) 进一步分析可发现, $1 < \alpha < 5$ 时, $R_{regular} - R_{random}$ 呈明显递增趋势,规则的传播优势更加明显;当 $5 \leq \alpha \leq 10$ 时, $R_{regular} - R_{random}$ 呈现波动性变化;当 $\alpha > 10$ 后, $R_{regular} - R_{random}$ 几乎维持不变。

此外,可以从图 3 中发现,总是存在临界值 α_c ,使得当 $1 \leq \alpha < \alpha_c$ 时, $R_{random} > R_{regular}$; $\alpha > \alpha_c$ 时,则 $R_{random} \leq R_{regular}$ 。保持节点数量 $N = 8 000$ 不变,临界值 α_c 相对于巅峰概率 β 变化规律如图 5 所示。

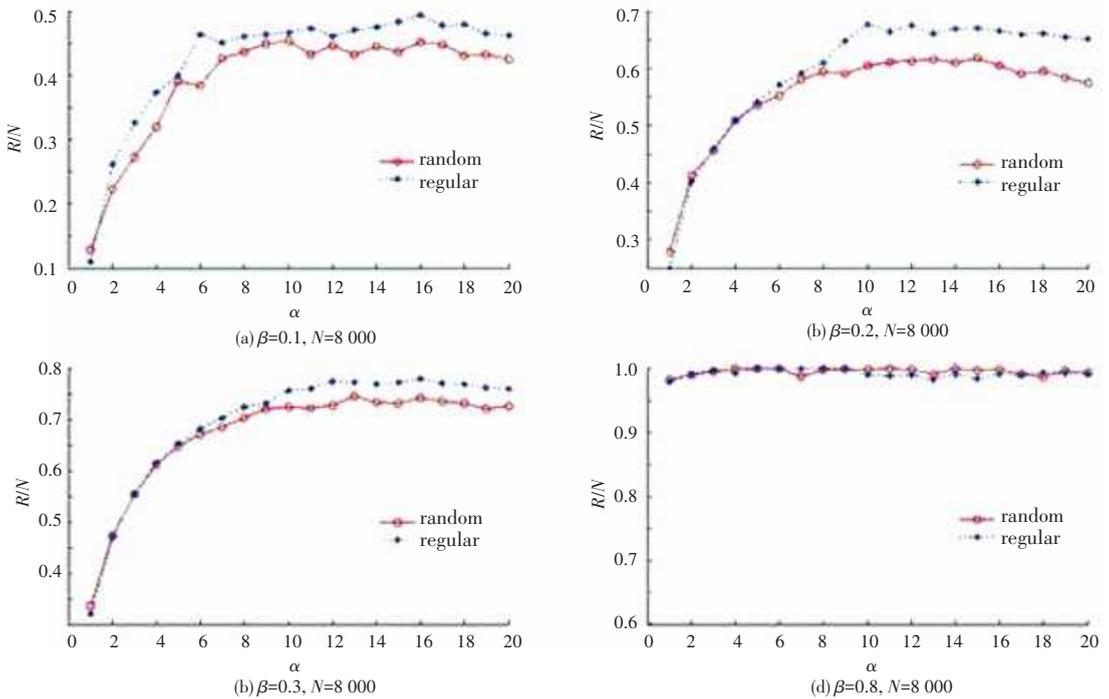


图3 疲劳态节点最终密度随累积峰值 α 变化的关系

Fig. 3 Relationship between final density of fatigue state node and cumulative peak value α

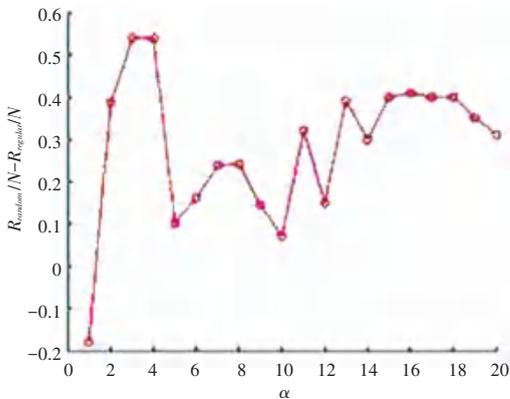


图4 疲劳态节点密度差值比例与 α 的变化关系

Fig. 4 The relationship between the ratio of the difference in fatigue state node density and α

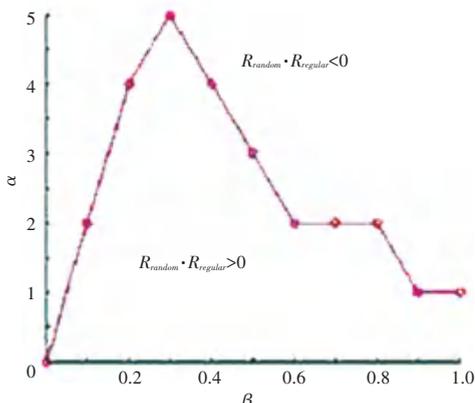


图5 临界值 α_c 随巅峰概率 β 的变化关系

Fig. 5 The relationship between the critical value α_c and the peak probability β

由图 5 可以看出 α_c 随着 β 先增大后减小, 当 $\beta=0.3$ 时, $\max \alpha_c = 5$ 。这说明当记忆效应呈现出兴趣累积特征时, 若这种特征不明显, 随机网络也有可能比规则网络更适合传播, 而 Centola 的规则网络优于随机网络的结论在这种场景下并不适用。

除了传播过程中的巅峰概率 α 与累积效应峰值 β 外, 另一个重要的影响因素是网络的规模 N 。研究中保持巅峰概率 $\beta = 0.1$, 网络平均度 $k = 11$ 不变, 在 100~8 000 间不同网络规模 N 上分别对 $\alpha = 1$, $\alpha = 10$, $\alpha = 20$ 这三种取值进行实验模拟。如图 6 所示。研究分析发现当 $N \leq 500$ 时, 随机网络总是优于规则网络。这是由于规模小的网络, 节点彼此间收到信息次数差异不明显。而在网络规模 $N > 500$ 后, $\alpha = 10$ 与 $\alpha = 20$ 对应的曲线总体上都呈现上升趋势。这是因为网络规模 N 上升, 随机网络的聚集系数 $\langle k \rangle / N$ 随之下降, 而规则网络聚集系数恒定的优势逐渐显现出来, 这使得单个节点的信息接收次数更易积累, 这与 Centola 的行为传播实验结论是一致的。

在线社交网络的度分布是不均匀的, 分别选取人人网数据集度最大的节点 $k_{\text{max}} = 221$ 与度最小的节点 $k_{\text{min}} = 1$ 为初始传播节点。当 $\alpha = 1, N = 8102$ 时, 在线社交网络中处于疲劳态 R 的节点占总数量 N 的比值随时间变化情况用 $R(t)/N$ 表示, 如图 7 所示。可以发现, 网络规模均为 $N = 8102$ 时, 在线社

交网络上的信息传播速度要快于随机网络与规则网络,这是由于社交网络中存在一些度较大的节点,使信息得以快速扩散。还可以发现,初始节点的选择对信息传播的规模存在影响,度较大则达到的最终传播规模更大,但这种规模并没有超越随机网络,而是介于随机网络与规则网络之间,这说明在聚集系数低的随机网络上记忆效应更不容易累积。

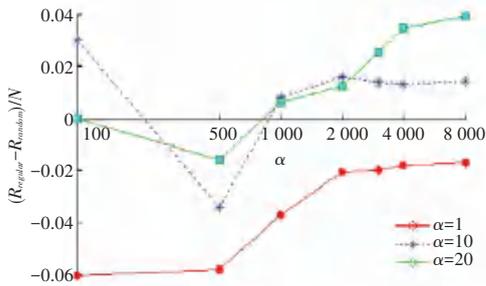


图6 疲劳态节点数量差值比例与网络规模的关系

Fig. 6 Relationship between the ratio of the difference of fatigue state nodes number and the network scale

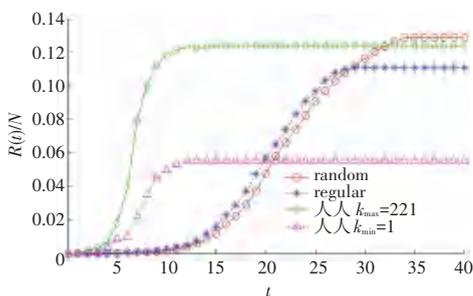


图7 初始节点度最大与度最小情况下的传播情况

Fig. 7 Propagation of the maximum degree and the minimum degree of initial node

4 结束语

本文首先讨论了在线社交网络上信息传播的记忆效应,及其不同表征过程。然后针对社交网络上传播行为的记忆效应与个体影响力对传统传染病模型进行改进,并在此基础上设计提出了基于记忆效应的在线社交网络传播模型。再由模拟网络与人人网数据集搭建的实验网络中对信息传播进行仿真,实验结果表明,该模型能够表征在线社交网络的传播特性,并且信息在不同结构网络上的传播效果存在差异,对此可表述如下。

(1)具有时效衰减特征的信息在随机网络上进行传播更广,拓展了Centola的实验结论。

(2)谣言、爱好等具有兴趣累加类型记忆效应的信息更容易在规模较大规则网络上传播,如学校的在线社区,且这种趋势会随着网络规模与累积峰值的增大而更加显著。

本文研发提出的模型针对的是单一信息的独立传播,而真实社交网络中信息是多种类且并行传播的,会产生互相影响,从而与独立信息的传播规律不同。研究多信息传播的规律并建立验证模型将是今后的研究方向。

参考文献

- [1] KERMACK W O, MCKENDRICK A G. Contributions to the mathematical theory of epidemics[J]. Proceedings of the Royal Society London, 1927, 115(772): 700-721.
- [2] DALEY D J, KENDALL D G. Epidemics and rumours [J]. Nature (S0028-8036), 1965, 204(9): 1118.
- [3] ABDULLAH S, WU Xindong. An epidemic model for news spreading on twitter [C]//2011 23rd IEEE International Conference on Tools with Artificial Intelligence (ICTAI). Boca Raton, FL, USA: IEEE, 2011: 163-169.
- [4] LIU Dechun, CHEN Xi. Rumor propagation in online social networks like Twitter -- A simulation study [C]// 2011 Third International Conference on Multimedia Information Networking and Security (MINES). Washington, DC: IEEE, 2011: 278-282.
- [5] CENTOLA D. The spread of behavior in an online social network experiment[J]. Science (New York, N. Y.), 2010, 329(5996): 1194-1197.
- [6] ROMERO D M, MEEDER B, KLEINBERG J. Differences in the mechanics of information diffusion across topics: idioms, political hashtags, and complex contagion on twitter [C]//Proceedings of the 20th International Conference on World Wide Web. New York: ACM, 2011: 695-704.
- [7] NEWMAN M E J. The structure and function of complex networks [J]. SIAM review, 2003, 45(2): 167-256.
- [8] BAO P, SHEN H W, CHEN W, et al. Cumulative effect in information diffusion: empirical study on a microblogging network[J]. PloS one, 2013, 8(10): e76027.
- [9] MYERS S A, ZHU Chenguang, LESKOVEC J. Information diffusion and external influence in networks [C]//Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Beijing, China: ACM, 2012: 33-41.
- [10] LÜ Linyuan, CHEN Duanbing, ZHOU Tao. The small world yields the most effective information spreading [J]. New Journal of Physics, 2011, 13(12): 123005.
- [11] ALBERT-LASZLO B, REKA A. Emergence of scaling in random networks [J]. Science, 1999, 286(5439): 509-512.
- [12] DUNCAN J W, STROGATZ S H. Collective dynamics of 'small-world' networks [J]. Nature, 1998, 393(6684): 440-442.