

一种基于用户相对权重的在线社交网络信息传播模型

王金龙 刘方爱 朱振方

An information spreading model based on relative weight in social network

Wang Jin-Long Liu Fang-Ai Zhu Zhen-Fang

引用信息 Citation: *Acta Physica Sinica*, 64, 050501 (2015) DOI: 10.7498/aps.64.050501

在线阅读 View online: <http://dx.doi.org/10.7498/aps.64.050501>

当期内容 View table of contents: <http://wulixb.iphy.ac.cn/CN/Y2015/V64/I5>

您可能感兴趣的其他文章

Articles you may be interested in

推荐重要节点部署防御策略的优化模型

Recommendation of important nodes in deployment optimization model of defense strategy

物理学报.2015, 64(5): 050502 <http://dx.doi.org/10.7498/aps.64.050502>

基于平均场理论的微博传播网络模型

Microblog propagation network model based on mean-field theory

物理学报.2014, 63(24): 240501 <http://dx.doi.org/10.7498/aps.63.240501>

Frenkel-Kontorova 模型中基底势振动的影响

Effect of the oscillation of substrate potential in driven Frenkel-Kontorova chains

物理学报.2014, 63(22): 220502 <http://dx.doi.org/10.7498/aps.63.220502>

带有层级结构的复杂网络级联失效模型

A cascading failure model of complex network with hierarchy structure

物理学报.2014, 63(22): 220501 <http://dx.doi.org/10.7498/aps.63.220501>

社交网络中信息传播的稳定性研究

Stability of information spreading over social network

物理学报.2014, 63(18): 180501 <http://dx.doi.org/10.7498/aps.63.180501>

一种基于用户相对权重的在线社交网络信息传播模型*

王金龙^{1)2)†} 刘方爱¹⁾ 朱振方³⁾

1)(山东师范大学信息科学与工程学院, 济南 250014)

2)(山东省分布式计算机软件新技术重点实验室, 济南 250014)

3)(山东交通学院信息科学与电气工程学院, 济南 250357)

(2014年8月11日收到; 2014年10月10日收到修改稿)

根据在线社交网络信息传播特点和目前社交网络传播模型研究中存在的问题, 本文定义了网络用户之间的相互影响力函数, 在此基础上提出了一种基于用户相对权重的社交网络信息传播模型, 并对网络中的传播路径及传播过程进行了分析, 讨论了不同路径的信息传播影响力. 为验证模型的有效性, 将传统的SIR模型和本文模型在六类不同网络拓扑下进行了仿真实验. 仿真结果表明, 两类模型在均匀网络中没有明显差异, 但在非均匀网络中本文模型更能体现真实网络特点, 实验同时验证了节点的地位影响着信息的传播, 并且发现英文社交平台 Twitter 和中文社交平台新浪微博在拓扑结构上具备一定相似性.

关键词: 复杂网络, 在线社交网络, 信息传播

PACS: 05.10.-a

DOI: 10.7498/aps.64.050501

1 引言

社交网络服务 (social network service, SNS) 又称社会化网络服务, 是基于 Web2.0 体系下的互联网技术应用. 在社交网络中, 用户通过网络途径将视频、图片、观点等信息传播给周围用户, 进而实现个体社交范围的逐步扩大, 最终形成一个大型的社交网络. 随着社交网络不断普及和规模的不断扩大, 社交网络相关研究吸引了广大研究人员的注意. 在这些研究中, 如何描述信息在社交网络中的传播行为、揭示它们的特性和传播规律, 具有非常重要的理论研究和实际应用价值, 同时也是当前的研究热点之一.

国内外学者就网络传播及其应用已经开展了大量研究工作, 如计算机病毒在计算机网络上的传播^[1-3]、谣言在社会网络中的扩散^[4-9]、传染病在

人群中的传播^[10-12], 信息在虚拟社区和博客中的传播^[13,14]等, 这些都可看成是服从某种规律的信息传播. 而在有关在线社交网络传播的相关研究中, Zhao等^[4]将信息传播模型SIR中的传播节点分成三类, 并且分析了模型在新型网络上的传播阈值; 王辉等^[5]提出了一种移动社交网络下的谣言传播模型, 验证了该模型在均匀网络中传播速度更快; 张彦超等^[13]模拟了在线社交网络中的信息传播过程, 分析了不同类型节点在网络中的行为规律; 王超等^[14]在社交网络的信息传播中引入了遏制机理和遗忘机理, 并在理论上证明了平衡点的渐近稳定性; Zheng等^[15]则从社会加强的角度分析了在线社交网络中的信息传播问题. 上述研究中, 多数传播模型是基于SIR模型^[4,10,11,14], 拓扑结构也大多是基于规则网络或者仿真网络^[5,13,15-17]; 而在信息传播过程中, 采用的传播概率始终固定^[4-6], 这同实际网络相比具有一定差距, 研究结论难以直接

* 山东省科技发展计划 (批准号: 2013YD01045)、山东省优秀中青年科学家奖励基金 (批准号: BS2013DX033) 和教育部人文社科基金 (批准号: 14YJC860042) 资助的课题.

† 通信作者. E-mail: wjl_1979lx@163.com

应用到实际网络中.

通过研究分析我们发现, 信息在真实网络的传播过程要受到信息发布者的网络地位及其自身行为的影响. 信息能否从一个节点传播到另一个节点并使得后者成为传播者, 既与传播用户的影响力有关, 又同后者用户的接受能力有关, 而目前研究成果则很少考虑以上因素. 针对该问题, 本文在传统 SIR 传播模型基础上, 对在线社交网络中的信息传播过程进行了详细分析, 将信息传播路径按照节点用户特点进行分类, 提出一种基于用户相对权重的网络传播模型, 并在不同的网络模型下进行了数值仿真研究.

2 基于用户相对权重的网络传播模型

RWSIR 模型

2.1 SIR 传播模型

在线社交网络用户的传播行为可描述为: 某用户发布一条信息给其邻居用户; 邻居用户以一定的概率接受并转发该消息, 不感兴趣的用户则忽略该消息; 如果邻居用户已经提前获得该消息则不再传播; 如果用户已经将信息发送给全部好友, 则停止传播. 不失一般性, 结合传染病的传播模型 SIR, 本文将社交网络上的单个用户定义为节点, 用户间的好友关系抽象为节点之间的边, 信息沿节点之间的边传播. 根据节点状态, 将社交网络上的节点分为三类: 具备传播信息能力的传播节点 I、从未听说过消息的健康节点 S 和对消息不感兴趣的免疫节点 R. 传播规则具体定义如下:

- 1) 如果一个健康节点与一个传播节点接触, 则健康节点会以概率 p_1 成为传播节点;
- 2) 如果一个免疫节点与一个传播节点接触将不会产生传播行为;
- 3) 一个传播节点在传播过程中以概率 p_2 失去传播兴趣成为免疫节点.

SIR 信息传播模型的动力学方程^[18]为

$$\begin{aligned} \frac{dS(t)}{dt} &= -p_1 S(t)I(t), \\ \frac{dI(t)}{dt} &= p_1 S(t)I(t) - p_2 I(t), \\ \frac{dR(t)}{dt} &= p_2 I(t), \end{aligned} \quad (1)$$

其中, N 表示网络节点总数, $S(t)$ 表示 t 时刻健康节点总数, $I(t)$ 表示 t 时刻传播节点总数, $R(t)$ 表示 t

时刻免疫节点总数, $S(t) + I(t) + R(t) = N$; 参数 p_1 表示健康节点与传播节点接触后转化为传播节点的概率; 参数 p_2 表示传播节点在传播过程中转化为免疫节点的概率.

现实社会中, 在信息的传播过程中, 人们更愿意接受和传播权威人士的观点, 信息的传播不仅仅依赖于信息本身的流行度, 同时也受到信息来源的影响. 信息能否从传播节点成功传到健康节点, 并使后者变为新的传播节点, 要受到双方相对地位的影响. SIR 模型中, 传播概率 p_1 是影响传播过程的一个重要参数, 在整个传播过程中保持不变, 其用于描述刻画信息在社会网络中的复杂传播行为显得过于简单, 基于此, 本文提出一种改进的传播模型.

2.2 基于用户相对权重的网络传播模型 RWSIR

为描述方便, 以图 1 所示案例来定义 RWSIR 传播模型的网络拓扑结构. 不失一般性, 网络中节点的邻居越多, 则其在网络中的影响力也越大, 该类节点被认为是权威节点 (如 1 号、9 号节点); 反之则认为普通节点 (如 5 号、17 号、18 号节点). 根据节点的传播影响力, 定义四类信息传播路径:

- Pa: 权威节点到普通节点, 如有向路径 $1 \rightarrow 5$;
- Pb: 普通节点到权威节点, 如有向路径 $5 \rightarrow 1$;
- Pc: 权威节点到权威节点, 如双向路径 $1 \leftrightarrow 9$;
- Pd: 普通节点到普通节点, 如双向路径 $17 \leftrightarrow 18$.

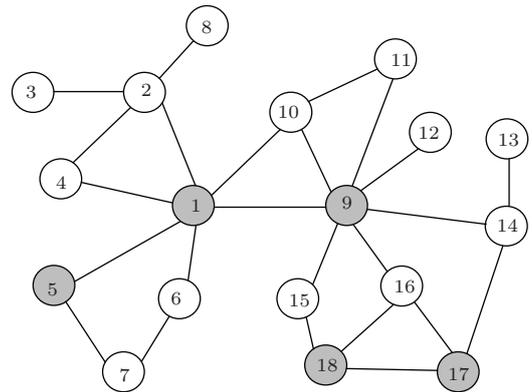


图 1 网络拓扑示例

2.2.1 相对影响力函数 $Rw(g_i, g_j)$

如前所述, 在通信过程中, 节点自身与邻居节点的相互影响力对信息是否成功转发关系重大. 令

g 表示网络中的一个节点, 取值为网络节点编号, 下面定义网络的节点影响力参数.

定义 1 节点的度 $k(g)$. 即节点 g 的邻居个数. 如 $k(1) = 6$.

定义 2 节点的邻居集合 $\Gamma(g)$. 即节点 g 的所有邻居节点的集合. 如 $\Gamma(1) = \{2, 4, 5, 6, 9, 10\}$.

定义 3 节点影响权重 $f(g_i, g_j)$. 设节点 $g_i \in \Gamma(g_j)$ 为节点 g_j 的一个邻居, 则 g_i 对 g_j 的影响力定义为

$$f(g_i, g_j) = \frac{k(g_i)}{\sum_{g_l \in \Gamma(g_j)} k(g_l)}, \quad (2)$$

其中 $\sum_{g_l \in \Gamma(g_j)} k(g_l)$ 为节点 g_j 的所有邻居的度数之和. 显然, 节点 g_i 的度 $k(g_i)$ 越大, 其对节点 g_j 影响力 $f(g_i, g_j)$ 也越大. 表 1 给出了节点 1 的邻居集合、所有邻居的度及所有邻居对节点 1 的影响力.

表 1 邻居节点集合、度及对节点 1 的影响力

g_i	2	4	5	6	9	10
$k(g_i)$	3	2	2	2	7	3
$f(g_i, 1)$	0.158	0.105	0.105	0.105	0.368	0.158

定义 4 节点 g_i 对节点 g_j 的相对影响权重 $Rw(g_i, g_j)$. 设节点 g_i 与 g_j 互为邻居, 结合定义 3, 相对影响权重 $Rw(g_i, g_j)$ 定义为

$$Rw(g_i, g_j) = 2 \frac{f(g_i, g_j)}{f(g_i, g_j) + f(g_j, g_i)}. \quad (3)$$

需要说明, $Rw(g_i, g_j)$ 一般不等于 $Rw(g_j, g_i)$, 这是由节点 g_i 和 g_j 的相对权威性决定的. 下面举例说明以上定义, 图 1 中, $\Gamma(5) = \{1, 7\}$, $k(1) = 6$, $k(7) = 2$, 则 $\sum_{g_j \in \Gamma(5)} k(g_j) = 8$, 因此有 $f(1, 5) = \frac{6}{8} = 0.75$, 由表 1 知 $f(5, 1) = 0.105$. 由定义 4 得, $Rw(1, 5) = 1.754$, $Rw(5, 1) = 0.246$. 同理可得 $Rw(9, 1) = 1.102$, $Rw(1, 9) = 0.898$, $Rw(17, 18) = 1.06$, $Rw(18, 17) = 0.94$. 显然, $Rw(g_i, g_j)$ 作用是将连接 g_i 和 g_j 边的权重根据双方的相对地位重新作了调整.

传统 SIR 模型中, 健康节点 g_S 与传播节点 g_I 接触后, 无论节点双方地位如何, 节点 g_S 转化为传播节点的概率是固定的, 即 $p_1 = \lambda$, 显然不符合真实社会信息传播的特点. 而考虑了相对影响力的传播模型 RWSIR 中, 信息传播要受到传播路径类型的影响, 转化概率将缩小或放大为原来的

$Rw(g_I, g_S)\lambda$ 倍. 如图 2 所示, 信息在 Pa 类路径上概率 p_1 被急剧放大, 在 Pb 类路径上概率被急剧缩小, 在 Pc 类和 Pd 类上则变化不明显. 模型体现了信息在传播过程中, 当两个节点权威性不对等时 (Pa 类、Pb 类), 传播概率 p_1 取决于权威性 $Rw(g_I, g_S)\lambda$ 而不是流行度 λ , 权威性大的节点很容易说服权威性小的节点接受其观点, 而反之则较为困难; 当两个节点权威性相差不大时 (Pc 类、Pd 类), 该模型对其影响不大, 传播概率 p_1 主要取决于信息的流行度 λ 而不是权威性 $Rw(g_I, g_S)\lambda$.

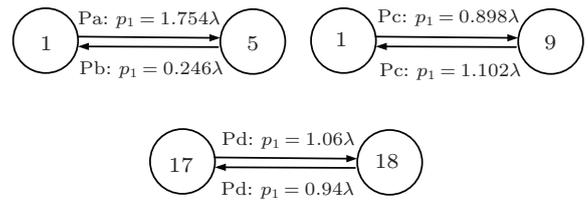


图 2 不同路径的传播概率

在 SIR 模型基础上, 结合节点相对权重 $Rw(g_I, g_S)$ 和社交网络的信息传播特点, 给定基本传播概率 λ ; 同时考虑, p_1 作为一个概率不能超过 1, 而 $Rw(g_I, g_S)\lambda$ 的值在某些特定条件下可能超过 1, 如果超过 1 时我们规定 $p_1 = 1$, 表示健康节点一定会成为传播节点. 令 $p_1(g_I, g_S)$ 为传播节点 g_I 对健康节点 g_S 的传播概率, 可表示为

$$p_1(g_I, g_S) = \begin{cases} Rw(g_I, g_S)\lambda, & Rw(g_I, g_S)\lambda < 1, \\ 1, & Rw(g_I, g_S)\lambda \geq 1. \end{cases} \quad (4)$$

2.2.2 基于用户相对权重的信息传播模型

基于以上研究, 本文提出基于用户相对权重的信息传播模型 RWSIR, 首先引入如下记法:

N_k 表示节点连接度为 k 的所有节点数量, $\sum_k N_k = N$;

$S_k(t)$ 表示 t 时刻节点连接度为 k 健康节点数量, $\sum_k S_k(t) = S(t)$;

$I_k(t)$ 表示 t 时刻节点连接度为 k 的传播节点数量, $\sum_k I_k(t) = I(t)$;

$R_k(t)$ 表示 t 时刻节点连接度为 k 的免疫节点数量, $\sum_k R_k(t) = R(t)$, 显然, $S_k(t) + R_k(t) + I_k(t) = N_k$;

$G_k = \{g_{k1}, g_{k2}, \dots, g_{kN_k}\}$ 表示节点连接度为 k 的节点集合, g_{ki} 为连接度为 k 的节点;

$V_m(g_{ki}) = \{\Gamma_{m1}(g_{ki}), \Gamma_{m2}(g_{ki}), \dots, \Gamma_{ms}(g_{ki})\}$ 表示节点 g_{ki} 所有 m 个邻居的子集的集合, 其中 $g_{ki} \in G_k, m = \{0, 1, 2, \dots, k\}, s = \binom{k}{m}$, 子集 $\Gamma_{ml}(g_{ki})$ 表示节点 g_{ki} 恰有 m 个邻居节点集合, 可表示为 $\Gamma_{ml}(g_{ki}) = \{g_{ki1}, g_{ki2}, \dots, g_{kim}\}$, 其中 $l = \{0, 1, 2, \dots, s\}$.

设节点 $g_{Sk} \in G_k$ 为连接度为 k 的一个健康节点, 假设有 m 个邻居为传播节点, 其集合为 $\Gamma_{ml}(g_{Sk}), p_1(g_{lki}, g_{Sk})$ 表示 g_{Sk} 与一个传播节点 $g_{lki} \in \Gamma_{ml}(g_{Sk})$ 接触后被感染的概率, 那么 g_{Sk} 不被感染的概率为 $1 - p_1(g_{lki}, g_{Sk})$. p_k^{SS} 表示 g_{Sk} 接触所有 m 个传播节点邻居后不被感染的概率, 则

$$p_k^{SS} = \prod_{g_{lki} \in \Gamma_{ml}(g_{Sk})} (1 - p_1(g_{lki}, g_{Sk})).$$

进一步, 对于任意的 $m = 0, 1, 2, \dots, k$, 任意一个连接度为 k 的健康节点 $g_{Sk} \in G_k$, 在下一个时间步长中不被感染的概率为

$$p_k^{SS} = \sum_{m=0}^k \sum_{\Gamma_{ml}(g_{Sk}) \in V_m(g_{Sk})} \prod_{g_{lki} \in \Gamma_{ml}(g_{Sk})} (1 - p_1(g_{lki}, g_{Sk})). \quad (5)$$

设 m 是一个随机变量, 且服从二项分布

$$\Pi(m, t) = \binom{k}{m} q(k, t)^m (1 - q(k, t))^{k-m},$$

其中 $q(k, t)$ 表示 t 时刻一个度为 k 的健康节点恰好连接一个传播节点的概率, 可表示为

$$q(k, t) = \sum_{k'} p(k'|k) \frac{I_{k'}(t)}{N_{k'}},$$

其中 $p(k'|k)$ 为度相关函数, 表示度为 k 的节点与度为 k' 的节点相连的条件概率. \bar{m} 表示一个健康节点 g_{Sk} 的邻居中传播节点个数 m 的数学期望, 则

$$\bar{m} = E(m) = kq(k, t) = k \sum_{k'} p(k'|k) \frac{I_{k'}(t)}{N_{k'}}. \quad (6)$$

$$\text{令 } \bar{R}_k = \frac{1}{kN_k} \sum_i \sum_{g_j \in \Gamma(g_{ks})} R w(g_j, g_{Sk}), \text{ 表示}$$

连接度为 k 的节点的所有邻居对该节点的平均相对权重, 给定信息流行度 λ , 邻居中所有传播节点对健康节点 g_{Sk} 的平均传播概率可表示为

$$\bar{p}_{1k} = \begin{cases} \bar{R}_k \lambda, & \bar{R}_k \lambda < 1, \\ 1, & \bar{R}_k \lambda \geq 1. \end{cases} \quad (7)$$

p_k^{SI} 表示健康节点 g_{Sk} 在下一个时间步长中转化为传播节点的概率, 综合 (5), (6), (7) 式, 有

$$\begin{aligned} p_k^{SI} &= 1 - p_k^{SS} \\ &= 1 - \sum_{m=0}^k \sum_{\Gamma_{ml}(g_{Sk}) \in V_m(g_{Sk})} \prod_{g_{lki} \in \Gamma_{ml}(g_{Sk})} (1 - p_1(g_{lki}, g_{Sk})) \\ &\approx 1 - (1 - \bar{p}_{1k})^{\bar{m}} \approx 1 - (1 - \bar{m}\bar{p}_{1k}) \\ &= \bar{m}\bar{p}_{1k} = \bar{p}_{1k} k \sum_{k'} p(k'|k) \frac{I_{k'}(t)}{N_{k'}}. \end{aligned} \quad (1)$$

已知整个网络中 t 时刻度为 k 的健康节点数为 $S_k(t)$, 则在下一个时间步长中, 转化为传播节点的平均个数为 $S_k(t)p_k^{SI}$, 即

$$S_k(t) \bar{p}_{1k} k \sum_{k'} p(k'|k) \frac{I_{k'}(t)}{N_{k'}}.$$

因此根据平均场方法, 网络中节点连接度为 k 的节点的传播动力学方程可表示为

$$\begin{aligned} \frac{dS_k(t)}{dt} &= -S_k(t) \bar{p}_{1k} k \sum_{k'} p(k'|k) \frac{I_{k'}(t)}{N_{k'}}, \\ \frac{dI_k(t)}{dt} &= S_k(t) \bar{p}_{1k} k \sum_{k'} p(k'|k) \frac{I_{k'}(t)}{N_{k'}} - p_2 I_k(t), \\ \frac{dR_k(t)}{dt} &= p_2 I_k(t), \end{aligned} \quad (8)$$

其中, \bar{p}_{1k} 为 (7) 式给出; 由文献 [13] 知, $p(k'|k) = k'P(k')/\bar{k}$, 其中 $P(k')$ 为度分布函数, \bar{k} 为网络节点平均度; 同时, 假定传播节点 g_{lki} 在传播过程中不会变成免疫节点, 只有当传播节点将消息传给所有的邻居之后或者传播节点的邻居都已经知道该信息时, 则停止传播行为并变成免疫节点, 因此方程 (8) 中我们令 $p_2 = 1$.

3 仿真模拟

3.1 相关网络拓扑及特性

我们选取比较流行的在线社交网络 Twitter [19], 新浪微博 [20] 和 Epinions [21] 作为本文的三个网络拓扑, 其相关数据均来自于互联网资源 [22-24]. 同时本文利用相关算法生成了三个仿真网络, 分别是 ER 随机网络 [25], NW 小世界网络 [26], BA 无标度网络 [27]. 我们假设每个网络的边均是无向无权的, 各网络的相关拓扑特性参数见表 2, 度分布如图 3 所示 (坐标为双对数坐标).

表2 各网络的相关特性参数

网络名称	节点个数	边数	节点平均度	节点最大度数	聚类系数	度相关系数
ER 随机网络	5000	25348	5.06	14	0.00089	-0.075
NW 小世界网络	5000	35094	7.02	17	0.5063	0.0412
BA 无标度网络	5000	29970	5.99	178	0.00402	-0.0501
Epinions	22437	212970	9.49	2031	0.9717	0.052
Twitter	145942	203152	1.392	7079	0.00014	-0.1114
新浪微博	146091	205408	1.406	2000	0.00024	-0.2446

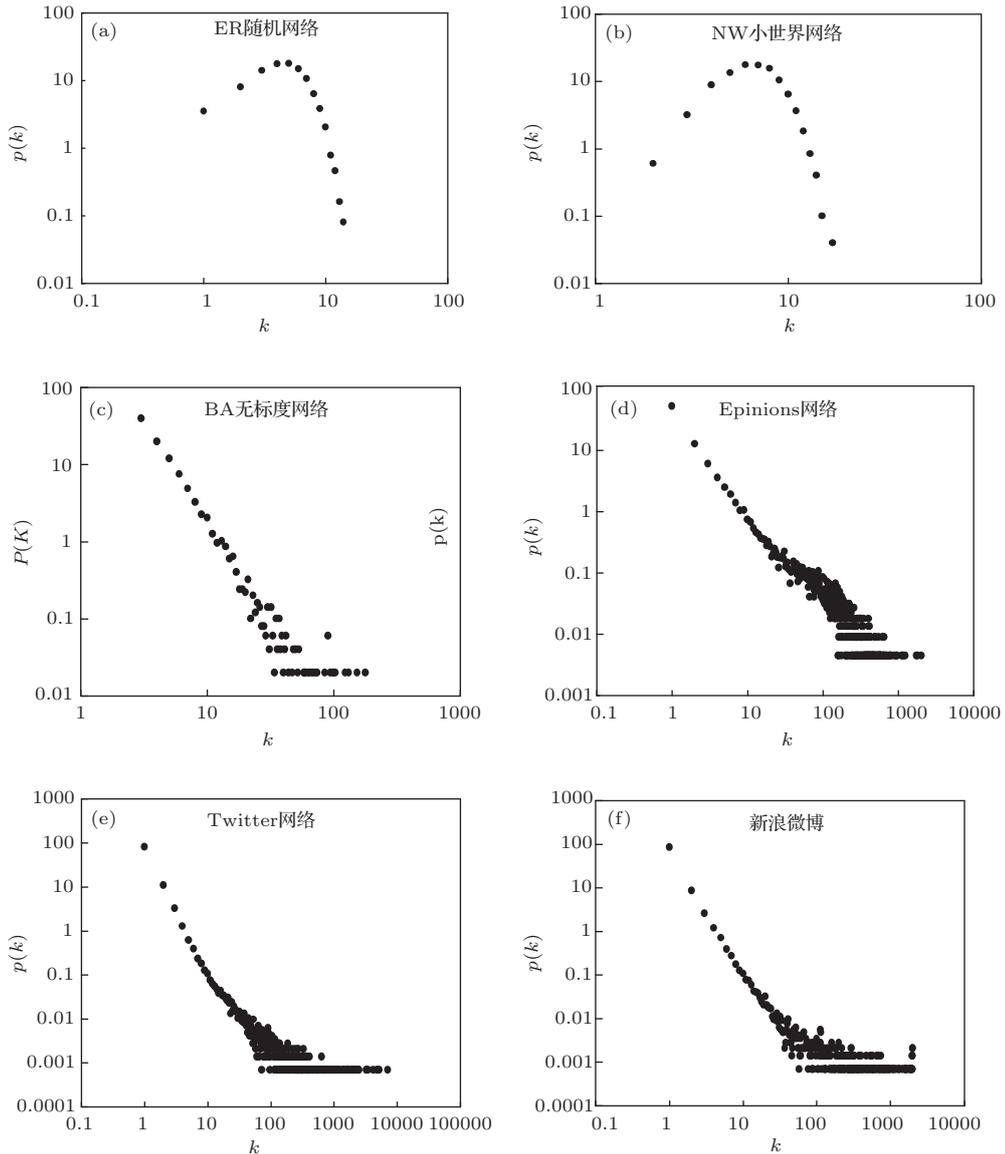


图3 各网络的度分布 $P(k)$ (a) ER 随机网络; (b) NW 小世界网络; (c) BA 无标度网络; (d) Epinions 网络; (e) Twitter; (f) 新浪微博

结合表2和图3,我们发现,ER随机网络的度分布服从泊松分布,聚类系数非常低,度相关性为负值,属于异配性均匀网络;NW小世界网络的度分布服从泊松分布,聚类系数较高,度相关性为正

值,属于同配性均匀网络;BA网络考虑了网络的增长特性和优选特性,度分布服从幂律分布,聚类系数较低,度相关性为负,属于异配性非均匀网络.在三个真实在线社交网络中,度分布基本上服从

幂律分布, 但是Epinions网络的聚类系数较高, 而Twitter和新浪微博的度相关性却是负值, Epinions属于同配性非均匀网络, Twitter和新浪微博则属于异配性非均匀网络.

3.2 RWSIR 信息传播模型

模型均选取每个网络拓扑中节点度数最大的一个节点作为传播节点, 其余节点均为健康节点, 信息的流行度 λ 取值为0.5, 当健康节点 g_S 与传播节点 g_I 相连接时, 健康节点以 p_1 的概率变成传播节点, 当传播节点将信息传播给所有邻居之后以 $p_2 = 1$ 的概率变为免疫节点. 图4给出了上述六个网络中健康节点、传播节点和免疫节点随时间 t 的

变化曲线 $S(t), I(t), R(t)$.

图4可以看出, 在给定初始传播节点之后, 由于网络节点总数不变, 随着健康节点不断被转变成传播节点, $S(t)$ 不断降低; 传播节点完成所有的传播过程之后转变为免疫节点, 从而 $R(t)$ 不断增长; 传播节点在传播初期速度增加较慢, 当节点数量达到一定规模后迅速增加而达到最大值, 由于受到网络规模和传播概率的影响, 传播节点转化为免疫节点的速度最终大于健康节点转化为传播节点的速度, 从而使得 $I(t)$ 曲线由开始增长达到最大值, 进而衰减, 最终趋于零. 三类节点在每个网络拓扑中表现出部分相同的规律, 同时在仿真网络和真实网络中又存在一定区别.

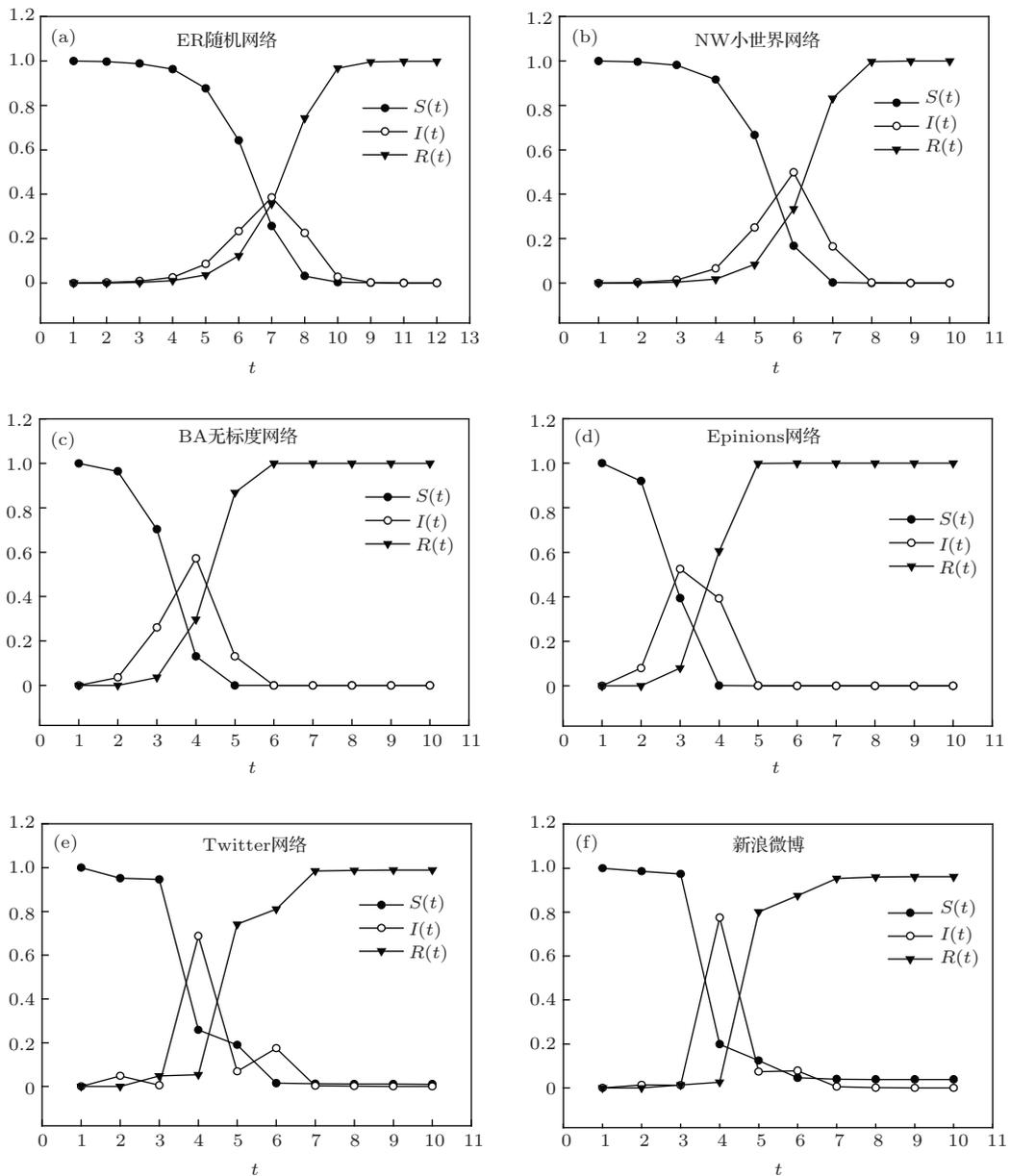


图4 各网络的RWSIR模型节点时间曲线 (a) ER随机网络; (b) NW小世界网络; (c) BA无标度网络; (d) Epinions网络; (e) Twitter; (f) 新浪微博

1) ER随机网络和NW小世界网络两个均匀网络中, $I(t)$ 到达最大峰值的时间分别为第6、7步; 而后面四个非均匀网络, 无论是BA仿真网络还是三个真实社交网络, 尽管它们的度不尽相同, 但是都是在第3或第4步达到了最大峰值, 因此在RWSIR模型中非均匀网络到达传播峰值所需的时间要少于均匀网络, 可以认为非均匀网络传播速度要快于均匀网络. 2) $I(t)$ 的峰值, 在真实网络规模中也都高于仿真网络, 同时免疫节点所占的最终比例在前五个网络都达到了1, 即传播结束之后, 信息基本上覆盖整个网络, 仅有新浪微博的值略低于1. 3) 传播节点的曲线在仿真网络上变化比较平滑, 图形比较规则, 而在真实社交网络中变化比较剧烈, 图形表现出一定的不规则性, 体现了真

实网络的信息传播突然爆发, 快速消退, 生命周期短的特点. 综上可见, 不同的网络的信息传播的规律不同, 同时真实网络与仿真网络传播规律也不相同.

4 RWSIR 和 SIR 的比较

本节对SIR和RWSIR两个模型在上述六个网络拓扑进行仿真模拟. 每个网络选取连接度最大的一个节点作为传播节点, 各网络节点最大连接度可见表2, 其余节点均为健康节点. 信息基本流行度 λ 分别取0.05, 0.1, ..., 0.45, 0.5十个数值, 每个网络模型实验各做100次, 然后取其平均值, 以下从三个方面进行比较和分析数据结果.

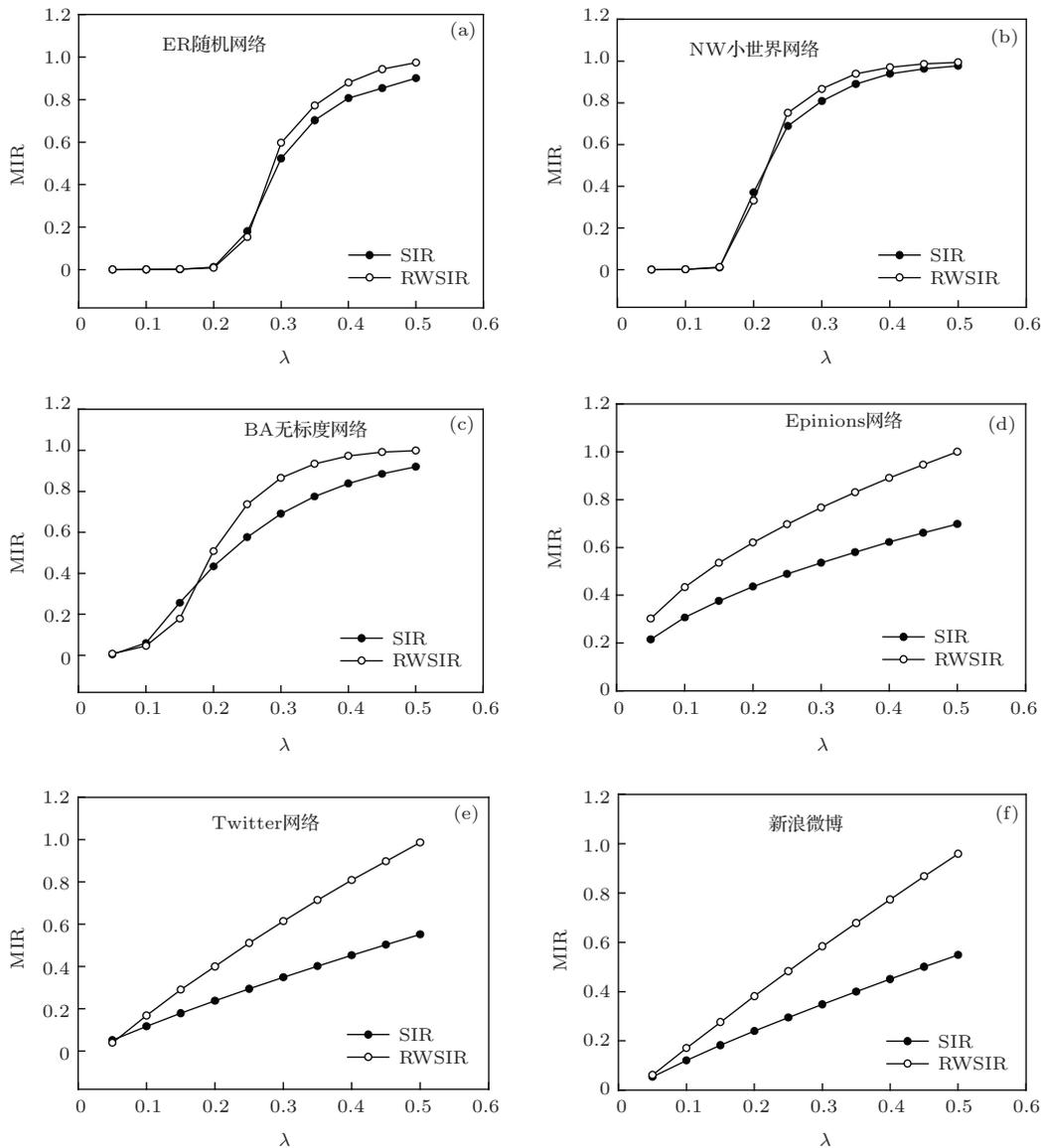


图5 最终感染率MIR的比较 (a) ER随机网络; (b) NW小世界网络; (c) BA无标度网络; (d) Epinions网络; (e) Twitter; (f) 新浪微博

1) 最大感染率MIR (max infection rate): 传播过程结束后, 网络中被感染过的节点数量占网络节点总数的比例. 由于模型中所有感染节点最终都会变成免疫节点, 因此我们可以把最终的免疫节点的比例看成最终感染率. MIR用来刻画传播节点最终对整个网络的影响范围.

如图5所示, 从整体来看, MIR值都是随着 λ 的增大而增大, 即信息流行度越高, 网络覆盖率越大. ER随机网络、NW小世界网络、BA无标度网络, 当 λ 取0.5时, MIR基本上为1, 说明此时信息能够覆盖整个网络. 而Epinions, Twitter和新浪微博中, 由于网络节点规模较大, 同时网络连接比较稀疏, SIR模型中MIR介于0.4—0.6之间, 信息不能覆盖整个网络; 而RWSIR模型中MIR接近1, 信息基本覆盖整个网络, 说明真实网络中权威节点发

布的信息很容易覆盖整个网络.

RWSIR和SIR两种模型在ER随机网络和NW小世界网络中, MIR没有明显差别, 仅在当 λ 较大时, RWSIR略高于SIR模型. 当 λ 较小时, BA无标度网络中的MIR值相差不大, 但是当 λ 较大时, RWSIR的MIR值高于SIR; 三个真实网络中, RWSIR的MIR值明显高于SIR中的值.

2) 最大感染峰值MIP (max infection peak): 传播过程中, 传播节点数量到达最大时占网络节点总数的比例. 该指标可以用来表示传播节点的最大传播速度, 显然, MIP值越大, 表明单位时间内信息的传播力度越大.

如图6所示, 两种模型在六种不同网络结构中均表现为随着 λ 增大, 最大感染峰值不断增大, 即传播概率越大, 信息传播的速度越大. ER随机网络

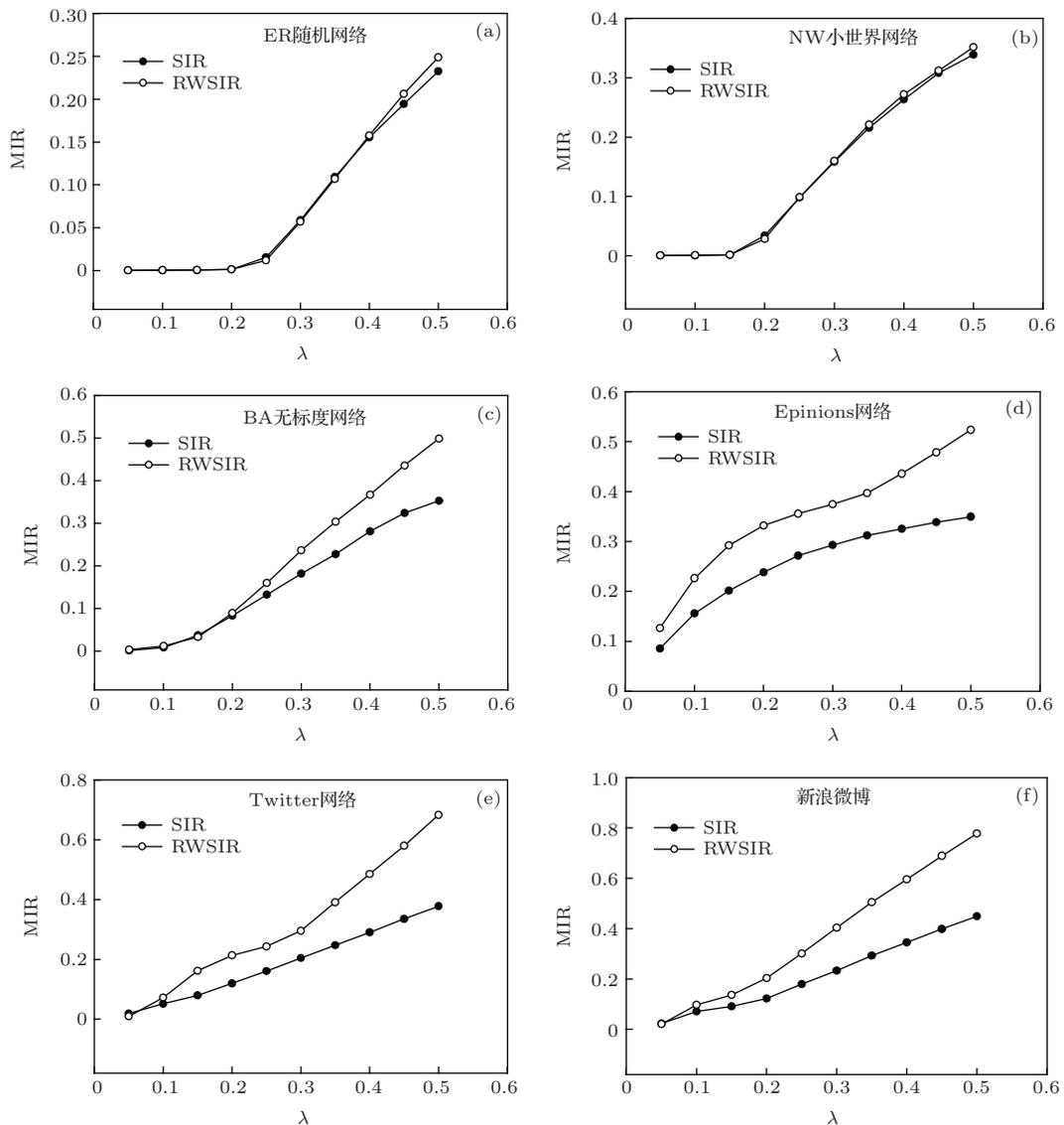


图6 最大感染峰值MIP的比较 (a) ER随机网络; (b) NW小世界网络; (c) BA无标度网络; (d) Epinions网络; (e) Twitter; (f) 新浪微博

和NW小世界网络中两种模型的MIP值没有明显差别;而在BA网络,Epinions网络, Twitter和新浪微博中, RWSIR明显高于SIR, 并且随着 λ 增大两个模型的差值在呈扩大趋势. 数据表明, RWSIR模型中真实网络中的权威节点发布的信息传播的速度明显快于SIR模型.

3) 传播生存周期SLC (spreading life cycle): 传播过程从一个传播节点开始, 到最后整个网络不存在传播节点, 整个过程所耗费的时间步长定义为传播生存周期. 传播生存周期可以用来刻画整个传播过程的所花费的时间, 即信息在网络中的生存时间.

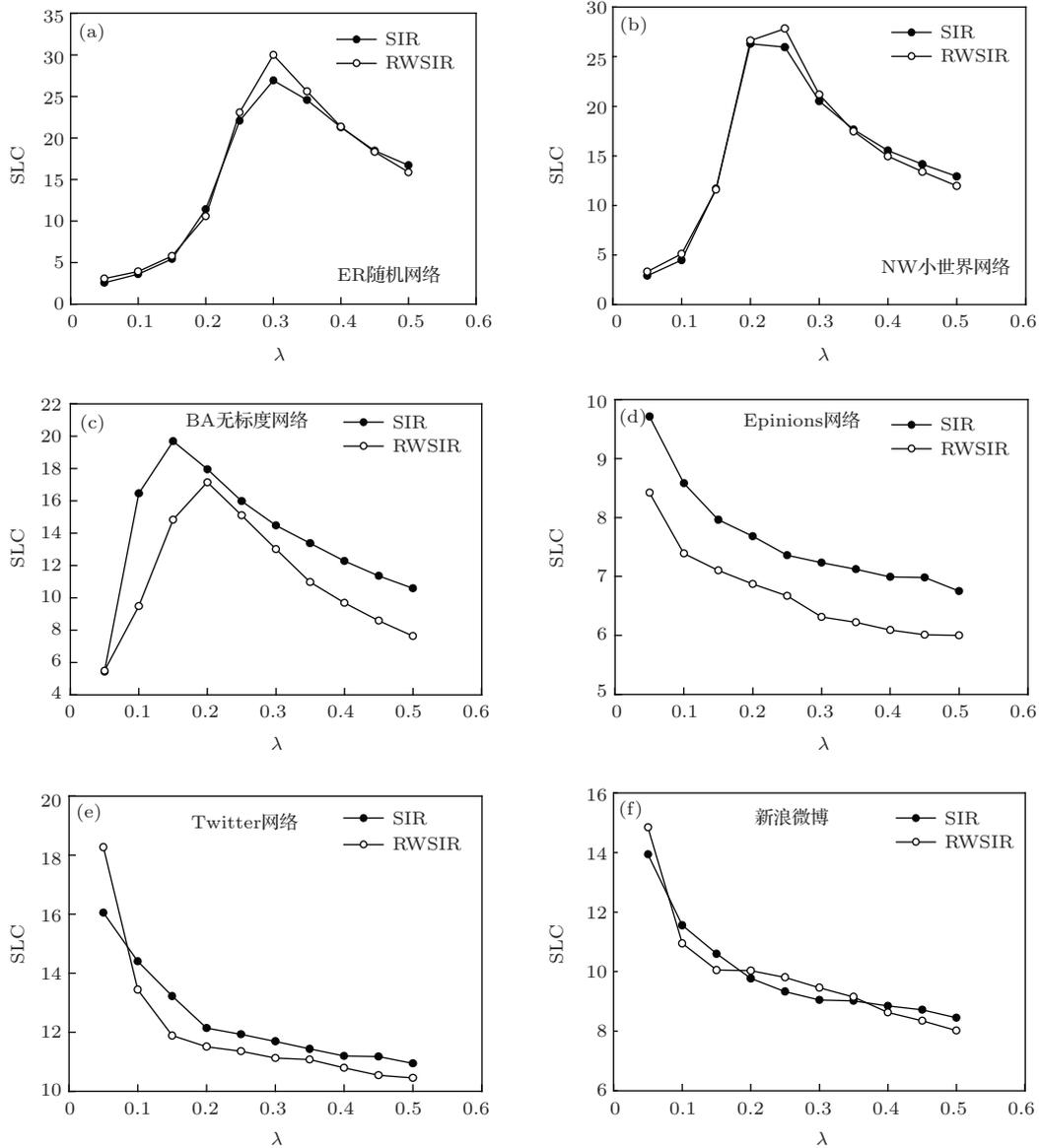


图7 传播生存周期SLC的比较 (a) ER 随机网络; (b) NW 小世界网络; (c) BA 无标度网络; (d) Epinions 网络; (e) Twitter; (f) 新浪微博

如图7所示, 三个仿真网络表现为: 随着 λ 增大, 两种模型的SLC先增大在减小, 出现一个峰值. 这是因为在传播机理中, 传播节点完成传播动作之后则转变为免疫节点, 显然, 如果 λ 很小, 信息还没来得及大范围传播就湮没, 生存周期会很短; 同时, 网络规模不变时, 随着不断增大 λ , 信息覆盖范围增大, 传播速度变快, 生存周期也越短. 而三个真

实网络中, SLC没有峰值, 随着 λ 增大, SLC不断减小. 这表明在真实网络中, 即使 λ 很小, 权威节点发布的信息也会很容易传播, 同时 λ 越大, 传播速度越快.

ER随机网络和NW小世界网络中, 两个模型中SLC值没有明显差别; BA网络和Epinions网络中, SLC在RWSIR中明显低于SIR; Twitter中,

RWSIR略低于SIR; 新浪微博的SLC在两种模型中差异不明显.

5 结 论

本文采用了六个不同的网络模型, 对SIR和RWSIR两种信息传播模型进行了仿真模拟, 结合第4节实验结果分析结论如下:

1) 信息本身的受欢迎程度, 即基本流行度 λ , 在信息传播过程中起着非常重要的作用. 图5—图7中可以看出, 无论是在均匀网络还是非均匀网络, 无论是经典的SIR模型还是本文提出的RWSIR模型, 传播过程中 λ 越大, 信息的网络覆盖率MIR和传播速度MIP越大, 相应的传播生存周期SLC越小. 这与第二节理论分析中, (4)式定义的传播概率 p_1 与 λ 成正比关系是一致的, 同时也描述了现实社会中爆炸性新闻在较短的时间内得以迅速传播、迅速消退、覆盖范围大的传播现象.

2) 数据结果同时表明, ER随机网络和NW小世界网络中, 两种传播模型的三个指标没有显著区别; 而在后四种非均匀网络中则表现出了明显差异. 这是由于SIR模型只考虑信息基本流行度 λ , 而RWSIR模型考虑了节点的相互影响力, 是基于节点的连接度提出的一种模型, 因此, RWSIR模型中的信息传播要受到网络拓扑结构的影响. 在均匀网络拓扑中, 由于节点的连接度分布为泊松分布, 大部分节点的连接度接近网络平均度, 节点之间的相互影响力基本上一致, 传播路径大多以Pc类和Pd类为主, 意味着大部分的影响力函数 $Rw(g_i, g_j)$ 都接近于1, 因此两个模型结果差异较小; 而在非均匀网络中, 度分布为幂律分布, 节点度数相差极大, 传播路径多以Pa类和Pb类为主, 影响力函数 $Rw(g_i, g_j)$ 分布极不均匀, RWSIR模型中, 权威节点发布信息时, 信息传播路径的权值被放大, 加速了信息的传播, 这与真实社交网络权威人士的观点很容易传播的现象符合. 实验结果与第二节的理论分析一致.

3) 三个仿真网络与三个真实网络的数据比较可以看出, 仿真网络和真实网络在信息传播过程中存在一定的差异. 三个真实网络度分布尽管服从幂律分布, 但是传播模型数据与BA无标度网络还存在很大差异. 数据显示, Twitter和新浪微博仿真结果比较一致, 而Epinions网络具有自己的特点, 这能够在一定程度上反映出网络的拓扑结构, 即三个真实网络中, Twitter和新浪微博在网络拓扑结

构上比较接近, 而Epinions网络与之有一定区别. 实际上, Epinions网络是一个基于信任的推荐评论网, 而Twitter和新浪微博则是基于媒体信息传播的网络.

本文研究了在线社交网络中的信息传播问题, 提出了一个基于用户相对影响力的信息传播模型. 该模型考虑节点之间相互影响力对信息传播的作用, 能够恰当的描述权威节点在信息传播中的影响, 改进了信息传播动力学模型, 使得该模型更加符合现实世界中的信息传播. 仿真数据结果与理论分析较为一致, 模型结果显示权威节点会加速信息的传播, 能够在一定程度上反映真实在线社会网络信息传播特点. 本文工作有助于进一步理解信息在真实网络中传播时的行为规律以及网络拓扑结构对信息传播的影响, 为微观的传播行为对宏观网络的影响研究奠定基础. 本文工作同时存在不足, 仅考虑了静态网络中信息传播的过程和权威节点对信息传播的影响, 如何在动态网络中对传播行为进行建模和分析普通节点对信息传播的影响将成为今后的研究工作.

参考文献

- [1] Newman M E J, Forest S, Balthrop J 2002 *Phys. Rev. E* **66** 035101
- [2] Lloyd A L, May R M 2001 *Science* **292** 1316
- [3] Yang L X, Yang X, Liu J, Zhu Q, Gan C 2013 *Applied Mathematics and Computation* **219** 8705
- [4] Zhao L J, Cui H X, Qiu X Y, Wang X L, Wang J J 2013 *Phys. A* **392** 995
- [5] Wang H, Han J H, Deng L, Cheng K Q 2013 *Acta Phys. Sin.* **62** 110505 (in Chinese) [王辉, 韩江洪, 邓林, 程克勤 2013 物理学报 **62** 110505]
- [6] Gu Y R, Ge L L 2012 *Acta Phys. Sin.* **61** 238701 (in Chinese) [顾亦然, 葛玲玲 2012 物理学报 **61** 238701]
- [7] Doerr B, Fouz M, Friedrich T 2012 *Communications of the ACM* **55** 70
- [8] Chierichetti F, Lattanzi S, Panconesi A 2011 *Theoretical Computer Science* **412** 2602
- [9] Doerr B, Fouz M, Friedrich T 2011 *Proceedings of the 43rd annual ACM symposium on Theory of computing*, San Jose, California, USA, June 6—8 2011 p21
- [10] Gong Y W, Song Y R, Jiang G P 2012 *Chin. Phys. B* **21** 010205
- [11] Lu Y L, Jiang G P, Song Y R 2012 *Chin. Phys. B* **21** 100207
- [12] Yuan X P, Xue Y K, Liu M X 2013 *Chin. Phys. B* **22** 030207
- [13] Zhang Y C, Liu Y, Zhang H F, Cheng H, Xiong F 2011 *Acta Phys. Sin.* **60** 050501 (in Chinese) [张彦超, 刘云, 张海峰, 程辉, 熊菲 2011 物理学报 **60** 050501]

- [14] Wang C, Liu C Y, Hu Y P, Liu Z H, Ma J F 2014 *Acta Phys. Sin.* **63** 180501 (in Chinese) [王超, 刘骋远, 胡远萍, 刘志宏, 马建峰 2014 物理学报 **63** 180501]
- [15] Zheng M H, Lü L Y, Zhao M 2013 *Physical Review E* **88** 012818
- [16] Lu L Y, Chen D B, Zhou T 2011 *New Journal of Physics* **13** 123005
- [17] Liu C, Zhang Z K 2014 *Communications in Nonlinear Science and Numerical Simulation* **19** 896
- [18] Ma Z E, Zhou Y C, Wang W D 2004 *The mathematical theory of infectious diseases and its applications* (Beijing: Science Press) pp4-5 (in Chinese) [马知恩, 周义仓, 王稳地 2004 传染病动力学的数学建模与研究 (北京: 科学出版社) 第4—5页]
- [19] <http://twitter.com/>
- [20] <http://weibo.com/>
- [21] <http://www.epinions.com/>
- [22] <http://snap.stanford.edu/data/>
- [23] <http://www.datatang.com/>
- [24] <http://www.nlp.ir.org/>
- [25] Erdős P, Rényi A 1960 *Publication of the Mathematical Institute of the Hungarian Academy of Sciences* **5** 17
- [26] Newman M E J, Watts D J 1999 *Physics Letters A* **263** 341
- [27] Barrat A L, Albert R 1999 *Science* **286** 509

An information spreading model based on relative weight in social network*

Wang Jin-Long^{1)2)†} Liu Fang-Ai¹⁾ Zhu Zhen-Fang³⁾

1) (School of Information Science and Engineering, Shandong Normal University, Jinan 250014, China)

2) (Shandong Provincial Key Laboratory for Distributed Computer Software Novel Technology, Jinan 250014, China)

3) (School of Information Science and Electric Engineering, Shandong Jiaotong University, Jinan 250357, China)

(Received 11 August 2014; revised manuscript received 10 October 2014)

Abstract

In this paper, we first introduce a mutual influence function among network nodes based on characteristics of information spreading in online social network. Then we put forward an information spreading model based on relative weight, analyze the propagation path and process of the network, and discuss the influence on different paths. Finally, the simulation experiments of the traditional SIR model and the model in this paper are conducted with six different network topologies. Results show that the two models have no significant difference in homogeneous networks, but there are significant differences in inhomogeneous networks. This result shows that the information spreading is influenced by the status of spreading nodes, and also shows that the real networks like Twitter and Sina Microblog have certain similarity in topological structure.

Keywords: complex network, online social network, information spreading

PACS: 05.10.-a

DOI: 10.7498/aps.64.050501

* Project supported by the Shandong Province Foundation for Development of Science and Technology, China (Grant No. 2013YD01045), the Shandong Province Outstanding Young Scientist Award Fund, China (Grant No. BS2013DX033), and the Science Foundation of Ministry of Education of China (Grant No. 14YJC86004).

† Corresponding author. E-mail: wjl_1979lx@163.com