

### 一种基于分支过程的信息流行度动力学模型

吴联仁 李瑾颀 齐佳音

Modeling information popularity dynamics based on branching process

Wu Lian-Ren Li Jin-Jie Qi Jia-Yin

引用信息 Citation: *Acta Physica Sinica*, 68, 078901 (2019) DOI: 10.7498/aps.68.20181948

在线阅读 View online: <https://doi.org/10.7498/aps.68.20181948>

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

---

### 您可能感兴趣的其他文章

#### Articles you may be interested in

#### 一种基于社交影响力和平均场理论的信息传播动力学模型

An information diffusion dynamic model based on social influence and mean-field theory

物理学报. 2017, 74(3): 030501 <https://doi.org/10.7498/aps.66.030501>

#### 一种基于文本互信息的金融复杂网络模型

Financial complex network model based on textual mutual information

物理学报. 2018, 75(14): 148901 <https://doi.org/10.7498/aps.67.20172490>

#### 一种基于用户相对权重的在线社交网络信息传播模型

An information spreading model based on relative weight in social network

物理学报. 2015, 72(5): 050501 <https://doi.org/10.7498/aps.64.050501>

#### 一种改进的基于信息传播率的复杂网络影响力评估算法

An improved evaluating method of node spreading influence in complex network based on information spreading probability

物理学报. 2017, 74(20): 208901 <https://doi.org/10.7498/aps.66.208901>

#### 基于用户行为的微博网络信息扩散模型

Modeling information diffusion on microblog networks based on users' behaviors

物理学报. 2016, 73(15): 158901 <https://doi.org/10.7498/aps.65.158901>

## 一种基于分支过程的信息流行度动力学模型\*

吴联仁<sup>1)</sup> 李瑾颀<sup>2)†</sup> 齐佳音<sup>1)3)</sup>

1) (上海对外经贸大学工商管理学院, 上海 201620)

2) (上海师范大学数理学院, 上海 200234)

3) (北京邮电大学, 可信分布式计算与服务教育部重点实验室, 北京 100876)

(2018年11月2日收到; 2019年1月17日收到修改稿)

Web 2.0 时代, 建模和预测在线信息流行度是信息传播中的重要问题. 本文基于社交网络系统信息传播的机制, 通过假设和简化, 提出了分支过程的概率模型, 来描述在线社交网络信息的流行度动力学过程. 对典型在线社交网络系统的信息流行度数据和网络结构数据进行了分析, 统计结果表明信息流行度衰减遵循幂律分布 (幂指数为 1.8), 微博网络的入度和出度分布也均服从幂律分布 (幂指数为 1.5). 模型仿真结果发现, 该模型能够再现真实社交网络数据的若干特征, 且信息流行度与网络结构相关. 对模型方程进行求解得到理论预测的结果与仿真分析和实际数据结果相符合.

**关键词:** 统计物理, 分支过程, 复杂网络, 信息流行度

**PACS:** 89.75.Da, 02.50.-r

**DOI:** 10.7498/aps.68.20181948

## 1 引言

社交媒体上的信息流行度演化与预测给复杂系统的研究者提出了诸多挑战, 如识别“病毒式”传播的原因、网络结构特征以及通过信息的内容和早期的关注预测未来的流行度. 虽然在线信息流行度的建模与预测已经取得了一定的研究进展<sup>[1-4]</sup>, 然而这个方向的研究大多由于可用的大规模数据缺少而受到阻碍. 近年来, 从在线社交网络获得的大规模社交数据为探索人类行为特征及其对在线内容流行的影响创造了前所未有的机会.

建模和预测信息流行度的困难在于各种混杂因素的共存<sup>[5-9]</sup>. 同时, 它给研究人员带来了许多挑战, 包括原因的识别<sup>[10-13]</sup>, 时间效应<sup>[14]</sup>和结构特征<sup>[15,16]</sup>. 一些学者对这一问题提出了不同的看法. Cetin 和 Bingol<sup>[17]</sup>认为个人注意力对信息的流

行有重要影响, 并且提出能见度的衰退和分散注意力的结合解释了为什么社交网络中的大多数信息级联不能成为流行. Weng 等<sup>[18]</sup>采用 agent-based 模型来研究模因 (memes) 之间的竞争是否会影响其流行度, 结果表明模因的流行度存在异质性, 这种现象是由模因竞争用户有限的注意力和在线网络的结构共同导致的. Gleeson 等<sup>[19]</sup>认为, 有限的用户注意力资源限制了消息的流行, 并且自然地导致一些消息变得非常流行, 其他消息只是中等流行, 或者被忽略. Yan 等<sup>[20,21]</sup>通过微博信息数据从人类动力学方面研究微博信息传播及扩散.

针对上述问题, 本文对新浪微博的信息数据和网络结构数据进行了分析, 结果发现信息流行度衰减遵循标度律. 其次, 提出了基于分支过程的概率模型, 来描述微博信息流行度变化的过程. 第三, 对所提出的模型进行数值仿真和理论求解, 发现该模型能够再现真实社交网络数据的若干特征. 此

\* 国家自然科学基金重大研究计划 (批准号: 91546121) 和国家自然科学基金青年科学基金 (批准号: 71601005) 资助的课题.

† 通信作者. E-mail: leejinjie@shmu.edu.cn

外, 信息流行度分布的幂指数与微博网络的度分布幂指数相关, 微博系统中信息流行度受网络结构的影响.

## 2 模型描述

在线社交网络用户之间相互关注形成有向网络(如 Twitter、新浪微博等). 在新浪微博系统中, 每个用户具有“微博首页”和“个人页面”两个列表. 用户关注的好友如果发出信息, 这些信息按时间先后顺序都将显示在“微博首页”这个列表上. 当用户打开微博系统查看“微博首页”列表上的信息时, 就会对感兴趣的微博进行转发或评论. 转发的微博信息同时会显示在“个人页面”这个列表上. 本文定义  $n(t)$  表示微博信息的流行度, 即从信息被生成后  $t$  时间内获得转发和评论的总数.  $q_n(t)$  表示在时间  $t$  信息获得流动度  $n(t)$  的概率.

由于“微博首页”存储列表的顶部是最新收到的信息, 之前收到的信息会逐渐淹没在列表的底部. 根据用户注意力有限的假设, 每次用户从“微博首页”列表顶部开始查看信息, 并且查看的信息是有限的. 这就会导致一些被淹没在列表底部的信息不被用户评论或转发. 不失一般性, 此处假设“微博首页”和“个人首页”存储信息的能力为 1, 即都只能保存一条信息, 新的信息到来时将覆盖掉原来保存在列表上的信息.

图 1 描述的是微博系统中信息传播的过程. 对于每个用户在当前时刻有两个动作(或状态): 1) 以概率  $\mu$  生成一条新的信息发出去; 2) 以概率  $(1 - \mu)$  转发“微博列表”上已有的信息. 如用户 1 在  $t_1$  时刻以概率  $\mu$  生成了一条信息(用圆圈表示), 同时发给其粉丝用户 2 和 3; 在  $t_2$  时刻用户 2 以概率  $(1 - \mu)$  转发这条信息给其粉丝用户 4 和 5; 在  $t_3$  时刻, 用户 3 又以概率  $\mu$  生成一条新的信息(用方框表示), 并将该信息发给其粉丝用户 6 和 7. 每当信息被传播一次, 信息的流行度加 1.

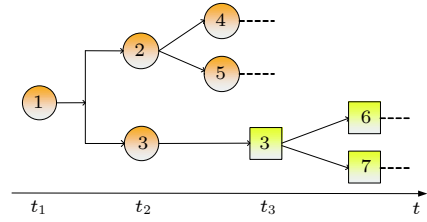


图 1 流行度模型示意图

Fig. 1. Schematic of the model.

## 3 模型方程与求解

根据第 2 节的模型描述, 本节采用分支过程来刻画信息的流行度动力学过程, 微博系统中每条消息的传播都遵循一个分支过程.

首先, 引入  $H(t, x)$  作为流行度分布的概率生成函数, 则

$$H(t; x) = xG(t, x) f(G(t, x)) = \sum_{n=1}^{\infty} q_n(t)x^n, \quad (1)$$

其中  $q_n(t)$  表示的是在  $t$  时间, 信息具有流行度  $n$  的概率. 同时, 在时间  $t = \Omega$ , 定义  $G_{jk}(t, \Omega; x)$  作为信息  $M$  流行度分布的概率生成函数, 并且信息  $M$  是由节点  $n_{jk}$  (节点  $n$  有  $j$  个出度(关注)和  $k$  个入度(粉丝)) 生成的或第一个转发的.

根据网络的度分布概率  $p_k$ , 可得到如下关系:

$$G(t, x) = \sum_k p_k G_{jk}(t, \Omega; x), \quad (2)$$

在时间  $t = \tau$ , 随机选择一个节点(用户), 那么在下一个时间  $t = \tau + \Delta t$ , 该节点(用户)“微博首页”存在四种可能的结果, 如表 1 所列.

将表 1 中的四种结果相加, 得到  $G_{jk}(t, \Omega; x)$  表达式如下:

$$G_{jk}(t, \Omega; x) = k\Delta t + \mu\Delta t + (1 - \mu)\Delta t x G_{jk}(t - \Delta t, \Omega; x) [G(t - \Delta t, \Omega; x)]^k + 1 - (k + 1)\Delta t G_{jk}(t - \Delta t, \Omega; x), \quad (3)$$

表 1 一个时间步节点(用户)“微博首页”的结果

Table 1. Single time-step outcomes of user's list.

列表 $S$	概率	$G_{jk}(t, \Omega; x)$
(a) 被新收到的信息所占据	$k\Delta t$	1
(b) 创造了一条新的信息	$\mu\Delta t$	1
(c) 以概率 $1 - \mu$ 转发信息	$(1 - \mu)\Delta t$	$xG_{jk}(t - \Delta t, \Omega; x) [G(t - \Delta t, \Omega; x)]^k$
(d) 保持原样	$1 - (k + 1)\Delta t$	$G_{jk}(t - \Delta t, \Omega; x)$

当  $\lim \Delta t \rightarrow 0$  得到  $G_{jk}(t, \Omega; x)$  的一个常微分方程, 该微分方程含有参数  $x$ :

$$\frac{\partial G_{jk}}{\partial t} = k + \mu - (k + 1)G_{jk} + (1 - \mu)xG_{jk}[G]^{k-1}. \quad (4)$$

通过乘以网络度分布概率  $p_k$ , 对节点  $n_{jk}$  可能的入度进行平均, 得到关于  $G(t, \Omega; x)$  的方程:

$$\frac{\partial G}{\partial t} = k + \mu - (k + 1)G + (1 - \mu)x \sum_k p_k G_{jk} [G]^k, \quad (5)$$

要从方程 5 中求解  $G$  需要知道在所有度  $k$  下的函数  $G_{jk}(t, \Omega; x)$ , 但如果做如下的近似:

$$\begin{aligned} \sum_k p_k (G_{jk}[G]^k) &\approx \left( \sum_k p_k G_{jk} \right) \\ \left( \sum_k p_k [G]^k \right) &= Gf(G), \end{aligned} \quad (6)$$

则 (5) 式变为

$$\frac{\partial G}{\partial t} = k + \mu - (k + 1)G + (1 - \mu)xGf(G). \quad (7)$$

将上面的推导一般化为每个用户的列表具有容量  $c$ , 即用户的“微博首页”和“个人页面”可同时保存  $c$  条微博信息:

$$c \frac{\partial G}{\partial t} = k + \mu - (k + 1)G + (1 - \mu)xGf(G), \quad (8)$$

$$\begin{aligned} m(t) &= \sum_{n=1}^{\infty} n q_n(t) = \frac{\partial H}{\partial x}(t, 1) = \\ &1 + (k + 1) \frac{\partial G}{\partial x}(t, 1), \end{aligned} \quad (9)$$

可得到信息  $M$  在时间  $t$  的平均流行度  $m(t)$  的方程为

$$c \frac{dm(t)}{dt} = (k + 1)(1 - \mu m). \quad (10)$$

令  $m(0) = 1$ , 得到平均流行度  $m(t)$  的解为

$$m(t) = \begin{cases} 1 + \frac{(k + 1)}{c} t, & \mu = 0 \\ \frac{1}{\mu} - \frac{1 - \mu}{\mu} e^{-\frac{\mu(k + 1)}{c} t}, & \mu > 0. \end{cases} \quad (11)$$

下面对流行度  $q_n(t)$  进行分析,  $q_n(t)$  表示在时间  $t$  信息获得流动度  $n(t)$  的概率, 表达式可以通过 (1) 式在  $t \rightarrow \infty$  时近似求解:

$$\lim_{t \rightarrow \infty} H(t; x) = xG_{jk}(\infty, \Omega; x) f(G_{jk}(\infty, \Omega; x)), \quad (12)$$

其中,  $G_{jk}(\infty, \Omega; x) = \lim_{t \rightarrow \infty} G_{jk}(t, \Omega; x)$ . 如果

$$H(\infty; x) = xG_{jk}(\infty, \Omega; x) f(G_{jk}(\infty, \Omega; x)) \quad (13)$$

为流行度概率  $q_n(t)$  的生成函数, 且  $G_{jk}(\infty, \Omega; 1 - w) \sim 1 - \phi(w)$  as  $w \rightarrow 0$ , 其中  $w = 1 - x$ . 则  $H(\infty; x)$  的近似形式为

$$H(\infty; 1 - w) \sim 1 - w - (k + 1)\phi(w) \text{ as } w \rightarrow 0. \quad (14)$$

因此,  $G_{jk}(\infty, \Omega; x)$  是方程 10 的解

$$k + \mu - (k + 1)G_{jk} + (1 - \mu)xG_{jk}f(G_{jk}) = 0. \quad (15)$$

如果网络度分布服从幂律分布, 即  $p_k \sim Dk^{-\gamma}$  且  $\gamma$  满足  $1 < \gamma < 3$ , 在这种情况下, 可得到当  $t \rightarrow \infty$  时

$$q_n(\infty) \sim \begin{cases} Bn^{-\frac{\gamma}{\gamma-1}} & \mu = 0 \\ Cn^{-\gamma} & \mu > 0, \end{cases} \quad (16)$$

其中因子  $B = -(k + 1) \frac{(D\Gamma(1 - \gamma))^{-\frac{1}{\gamma-1}}}{\Gamma\left(\frac{1}{1 - \gamma}\right)}$  和  $C = (k + 1) \left(\frac{1 - \mu}{\mu(k + 1)}\right)$ .

从 (11) 式可知, 当  $t \rightarrow \infty$ , 信息流行度分布  $q_n(t)$  服从幂律分布, 且幂指数为  $\frac{\gamma}{\gamma - 1}$  ( $\mu = 0$ ) 和  $\gamma$  ( $\mu > 0$ ), 幂指数等于或大于网络度分布的幂指数  $\gamma$ .

## 4 实证分析和数值模拟

### 4.1 数据描述和处理

本文分析了两个数据集, 均通过新浪微博开放平台 API 收集 ([www.weibo.com](http://www.weibo.com)), 新浪微博是目前中国最流行的微博平台之一. 第 1 个数据集是微博信息数据, 通过滚雪球采样的方法收集了某个话题在 2009/8/20-2010/9/3 之间发布的 125139 条微博, 以及这些微博被转发了 2260826 次和 1822450 条评论. 每条微博包括发出用户 ID、微博 ID、微博发布时间、微博内容、转发次数和评论次数. 此外还采集了微博的评论时间, 微博的转发时间虽无法获得, 但微博的转发流行度分布和评论流行度分布具有相同的分布特征 (如图 2 所示), 均服从幂指数约为 1.8 的幂律分布. 因此, 评论流行度的变化可近似表示微博整体流行度的变化 (评论流行度和转发流行度).

第 2 个数据是微博用户数据, 第 2 个数据集收集了参与话题讨论的 41667 个用户信息. 对每个用户, 采集了用户 ID、用户关注数和关注关系、用户

粉丝数. 从而获得微博网络用户度分布情况, 数据统计分析发现用户入度 (粉丝) 和出度 (关注) 分布均服从幂指数约为 1.5 的幂律分布 (如图 3 所示), 但出度分布在 2000 附近具有一个截断, 这是由于新浪微博系统初期允许关注的上限是 2000.

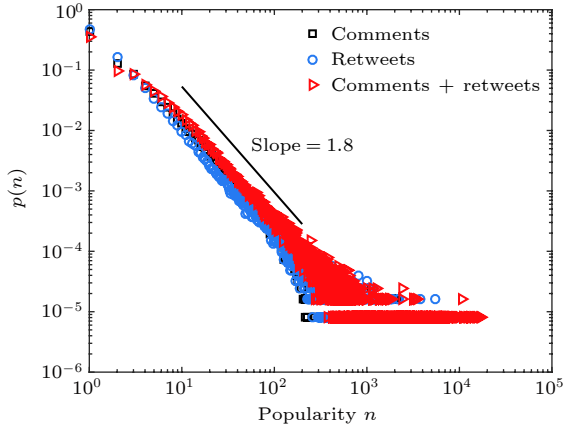


图 2 微博的流行度分布

Fig. 2. Distribution of micro-blogs popularity.

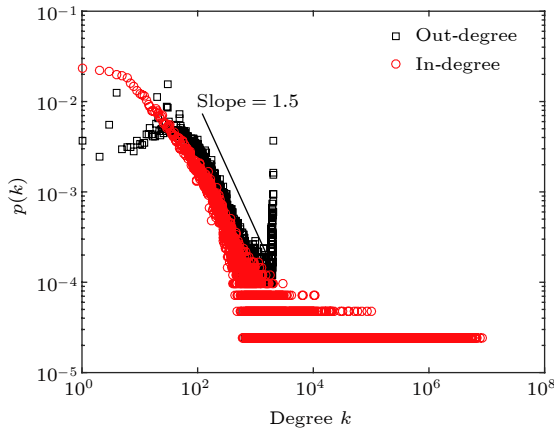


图 3 微博用户度分布

Fig. 3. Distribution of out-degree and in-degree.

## 4.2 数值模拟

为了验证分支过程近似的精确性, 并探讨网络结构与有限注意力的相互作用, 本文将模型的数值仿真结果与实际数据和理论预测结果进行比较. 首先生成具有指定度分布的有向网络, 网络的度分布服从幂律分布  $p_k \sim k^{-\gamma}$ ,  $\gamma = 1.5$ . 模型参数设定为节点总数是  $N = 10^5$ , 且

$$\mu = \frac{\text{数据集1中不重复的微博数}}{\text{数据集1中微博总数}} = \frac{125139}{2260826} \approx 0.055.$$

$\mu$  值反映的是微博系统中原创微博信息的占比, 从本数据集中的  $\mu$  值可知, 该话题下大部分微

博是重复转发的, 原创微博的占比很小.

数值仿真的微博信息的平均流行度 (蓝圆线), 与实际数据 (黑钻石线) 和方程 6 的理论预测 (红方线) 的比较如图 4 所示. 其中参数取值为微博生成概率  $\mu = 0.055$ , 平均入度  $k = 12.3$ , 微博列表存储信息能力  $c = 1$ . 因新浪微博网络中不同时间窗口、不同主题话题下微博信息流行度存在差异, 本文实证数据计算结果与理论模拟存在偏差. 从图 4 可见, 在初始阶段, 平均流行度的实际数据与理论预测和仿真偏差较大, 随着时间  $t$  不断增加, 偏差先减小后增大, 最后实际平均流行度值和理论预测值均趋向于定值, 偏差稳定. 另外在数据处理与选择时, 单个微博信息流行度时间序列  $\{n(t)\}$  数据点大于等于 10 时, 该微博信息才被纳入计算平均流行度的数据. 每个微博信息流行度的时间序列  $\{n(t)\}$  数据点不相等, 也导致了平均流行度的计算结果和理论模拟的偏差.

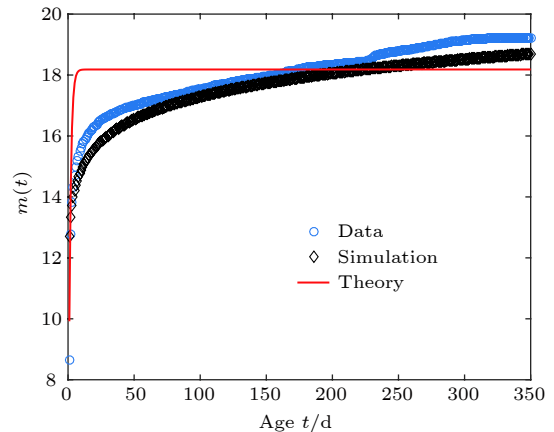


图 4 微博信息平均流行度

Fig. 4. Mean popularity of Micro-blogs.

图 5 比较了微博信息流行度模型仿真结果、实际数据和理论预测的结果. 其中参数取为  $p_k \sim k^{-\gamma}$ ,  $\gamma = 1.5$ ,  $N = 10^5$ ,  $\mu = 0.055$ . 网络结构是新浪微博中 41667 个用户数据生成的网络. 在  $n$  较小时, 模型仿真结果与实际数据及理论预测结果一致, 当  $n \rightarrow \infty$  时, 实证数据与仿真结果和理论预测之间存在一定偏差, 但大致结果是符合的. 导致偏差的主要原因是, 基于分支过程的流行度模型是建立在假设微博网络是树形结构基础上的, 但实际上, 新浪微博 41667 用户生成的网络中 34% 的链接是互惠链接. 基于树形结构理论的精确结果应用于现实世界的网络中, 因此导致了一定的偏差.

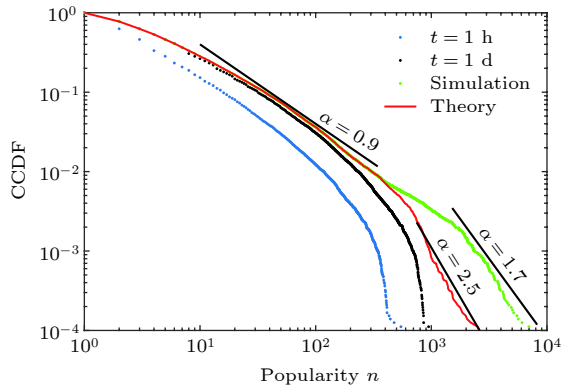


图 5 微博信息流行度的互补累积概率分布 (CCDF)

Fig. 5. Complementary cumulative distribution functions (CCDFs)—the fraction of micro-blogs with popularity  $\geq n$ .

通过对微博信息评论转发数据和网络结构数据的统计分析可知, 在微博系统中存在高度的异质性, 如流行度小于 100 的微博信息占比达到 95.8%, 而流动度超过 1 万的微博信息数量是非常少的. 粉丝数 (入度) 小于 100 的用户占比达 56.4%, 而有的用户粉丝数高达百万. 通过仿真, 重现了复杂社会系统中的高度异质性, 大量的流行度 (或关注) 都被少量的信息 (或用户) 获得, 得到了流行度依赖时间的重尾分布特征. 本文所提出的框架构成了社会传播现象的零模型, 与纯粹的实证研究或基于模拟的模型相比, 它清楚地区分了影响信息流行度的两个不同因素的作用, 即用户的记忆时间和社交网络的连接结构.

## 5 总结与讨论

在线社交媒体极大地影响了人们彼此沟通的方式. 近年来, 在线社交媒体信息流行度的预测和建模引起了众多学者的关注. 例如, 预测和建模社交媒体上的新闻流行度<sup>[22]</sup>和量化论文流行度<sup>[23]</sup>. 本文引入并分析了一种信息传播的概率模型, 该模型具有分析易处理性, 可以再现实际数据的若干特

征. 但是该模型也存在一定的局限性, 其中要求做出一些假设以获得分析结果. 在将来的研究中, 我们希望对模型进行一些可能的扩展.

## 参考文献

- [1] Wang D, Song C, Barabási A L 2013 *Science* **342** 6154
- [2] Sasahara K, Hirata Y, Toyoda M, Kitsuregawa M, Aihara K 2013 *PloS one* **8** e61823
- [3] Quattrocioni W, Caldarelli G, Scala A 2014 *Sci. Rep.* **4** 4938
- [4] Kim Y, Park S, Yook S H 2016 *Sci. Rep.* **6** 23484
- [5] Song B, Jiang G P, Song Y R, Xia L L 2015 *Chin. Phys. B* **24** 100101
- [6] Wang J R, Wang J P, He Zhen, Xu H T 2015 *Chin. Phys. B* **24** 060101
- [7] Min L, Liu Z, Tang X Y, Chen M, Liu S Y 2015 *Acta Phys. Sin.* **64** 088901 (in Chinese) [闵磊, 刘智, 唐向阳, 陈矛 2015 物理学报 **64** 088901]
- [8] Wang J L, Liu F A, Zhu Z F 2015 *Acta Phys. Sin.* **64** 050501 (in Chinese) [王金龙, 刘方爱, 朱振方 2015 物理学报 **64** 050501]
- [9] Li Y J, Yin C, Yu H, Liu Z 2016 *Acta Phys. Sin.* **65** 020501 (in Chinese) [李勇军, 尹超, 于会, 刘尊 2016 物理学报 **65** 020501]
- [10] Gleeson J P, O'sullivan K P, Banos R A, Moreno Y 2016 *Phys. Rev. X* **6** 021019
- [11] Zapperi S, Lauritsen K B, Stanley H E 1995 *Phys. Rev. Lett.* **75** 4071
- [12] Iribarren J L, Moro E 2011 *Phys. Rev. E* **84** 046116
- [13] Xu Y, Guo L P, Ding N, Wang Y G 2010 *Chin. Phys. Lett.* **27** 078901
- [14] Zhu J F, Han X P, Wang B H 2010 *Chin. Phys. Lett.* **27** 068902
- [15] Li J J, Wu L R, Qi J Y, Sun Q M 2017 *Chin. Phys. Lett.* **34** 068901
- [16] Yang L, Song Y R, Li Y W 2018 *Acta Phys. Sin.* **67** 190502 (in Chinese) [杨李 宋玉蓉 李因伟 2018 物理学报 **67** 190502]
- [17] Cetin U, Bingol H O 2014 *Phys. Rev. E* **90** 032801
- [18] Weng L, Flammini A, Vespignani A, Menczer F 2012 *Sci. Rep.* **2** 335
- [19] Gleeson J P, Ward J A, O'sullivan K P, Lee W T 2014 *Phys. Rev. Lett.* **112** 048701
- [20] Yan Q, Yi L L, Wu L R 2012 *Physica A* **391** 1540
- [21] Yan Q, Wu L R, Zheng L 2013 *Physica A* **392** 1712
- [22] Miotto J M, Altmann E G 2014 *PloS one* **91** e111506
- [23] Sinatra R, Wang D S, Deville P, Song C Y, Barabási A L 2016 *Science* **354** 6312

# Modeling information popularity dynamics based on branching process\*

Wu Lian-Ren<sup>1)</sup> Li Jin-Jie<sup>2)†</sup> Qi Jia-Yin<sup>1)3)</sup>

1) (*School of Management, Shanghai University of International Business and Economics, Shanghai 201620, China*)

2) (*College of Mathematics and Science, Shanghai Normal University, Shanghai 200234, China*)

3) (*Key Laboratory of Trustworthy Distributed Computing and Service, Beijing University of Posts and Telecommunications, Beijing 100876, China*)

( Received 2 November 2018; revised manuscript received 17 January 2019 )

## Abstract

In the age of Web 2.0, modeling and predicting the popularity of online information was an important issue in information dissemination. Online social medium greatly affects the way we communicate with each other. However, little is known about what fundamental mechanisms drive the dynamical information flow in online social systems. To address this problem, we develop a theoretical probabilistic model based on branching process to characterize the process in which micro-blog information gains its popularity. Firstly, the data of information popularity and network structure of micro-blog network are analyzed. The statistical results show that the attenuation of information popularity follows a scaling law whose exponent is 1.8, and in-degree and out-degree of micro-blog network each also obey a power law distribution whose exponent is 1.5. The results of power law distribution show that there is a high-degree heterogeneity in a micro-blog system. The proportion of micro-blog information with popularity less than 100 is 95.8%, while the amount of micro-blog information with popularity more than 10, 000 is very small. The number of fans (in-degree) less than 100 accounts for 56.4%, while some users have millions of fans. Secondly, according to the design mechanism of the Weibo system, we assume that each user has two lists, i.e. a "home page list" and a "personal page list". Meanwhile, each user has two states at each moment: generating a new message with probability  $\mu$  to be sent out; 2) or forwarding the information already on the "personal page list" with probability  $(1 - \mu)$ . Based on the assumptions, the information popularity model is proposed. Finally, the model is simulated. The simulation results show that the model can reproduce some features of real social network data, and the popularity of information is related to the network structure. By solving the model equation, the results of theoretical prediction are consistent with the simulation analyses and actual data.

**Keywords:** statistical physics, branching process, complex network, information popularity

**PACS:** 89.75.Da, 02.50.-r

**DOI:** 10.7498/aps.68.20181948

\* Project supported by the Major Research Plan of the National Natural Science Foundation of China (Grant No. 91546121) and the Young Scientists Fund of the National Natural Science Foundation of China (Grant No. 71601005).

† Corresponding author. E-mail: [leejinjie@shnu.edu.cn](mailto:leejinjie@shnu.edu.cn)