

基于平均场理论的微博传播网络模型*

吴腾飞¹⁾ 周昌乐¹⁾²⁾ 王小华¹⁾ 黄孝喜^{1)†} 谌志群¹⁾ 王荣波¹⁾

1) (杭州电子科技大学计算机学院, 认知与智能计算研究所, 杭州 310018)

2) (厦门大学智能科学与技术系, 福建省仿脑智能系统重点实验室, 厦门 361005)

(2014年4月7日收到; 2014年8月8日收到修改稿)

微博是在通过用户关注机制建立的用户网络上分享实时信息的社交平台, 而微博消息主要通过用户的转发行为使消息在用户上传传播. 掌握微博消息的传播机制, 对研究微博上舆论谣言的传播、产品推广等具有指导作用. 本文通过对微博传播网络的结构分析来探索微博传播过程, 利用新浪微博数据, 建立微博传播网络, 分析该网络的生成机制, 使用平均场论的方法, 推导微博传播网络的度分布模型. 实验结果表明: 微博传播网络的度分布是时间相依的, 在特定时间下网络的度分布服从幂律分布.

关键词: 复杂网络, 微博传播网络, 无标度网络, 平均场理论

PACS: 05.10.-a, 89.75.-k

DOI: 10.7498/aps.63.240501

1 引言

复杂网络是研究自然界复杂系统的一种新方法, 其研究源头可追溯到18世纪数学家Euler(欧拉)对“七桥问题”的研究而开创的图论学科; 20世纪60年代, Erdos和Rényi^[1,2]及Solomonoff和Rapoport^[3]引入随机图论描述复杂网络; 1998年, Watts和Strogatz提出的小世界网络模型^[4]; 1999年, Barabási和Albert通过对含有325729个节点的WWW网络研究, 发现网络的度分布 $P(k)$ 随网络的度 k 幂律衰减, 即 $P(k) \sim k^{-\gamma}$, 网络中少数节点拥有大量连线, 而大部分节点只有少数连线, 由于节点的这种异质性, 网络的特征标度消失, Barabási和Albert称该网络为无标度网络^[5], 文献上通常将Barabási和Albert提出的无标度网络模型称为BA模型.

BA模型认为增长和择优是网络演化的两条重要机理, 二者缺一不可, 否则无法生成无标度网络.

随后一些学者在BA模型的基础上提出其他形式的无标度网络模型, 如Dorogovtsev和Mendes^[6]提出一个新连线数按幂律加速增长的模型、Krapivsky和Redner^[7]根据文献引用的情况提出复制模型, Li和Chen^[8]根据局部择优和距离择优提出局域世界演化模型等.

随着复杂网络结构研究的迅速发展, 人们逐渐发现真实网络中一些有趣的传播现象, 如: 疟疾、天花、SARS、H1N9等疾病在人类社会网络中的流行; 震荡波、灰鸽子、熊猫烧香等计算机病毒在通信网络中的扩散; 消息、谣言、产品广告等在社交网络中的蔓延; 电力故障在电力网络中的传播; 金融危机在经济贸易网络的扩散等成为人们越来越关注的主题. 这些都可以看作是服从某种特征规律的事件(或事物)在复杂网络上的传播行为^[9], 如何去描述这些事件(或事物)的传播过程, 挖掘它们的传播特性, 是复杂网络极具挑战的研究课题^[10].

复杂网络上传播动力学的分析最早是从传染病模型开始, 在典型的SIR传染病模型^[11]中, 将

* 国家自然科学基金青年科学基金(批准号: 61103101, 61202281)和教育部人文社会科学研究项目青年基金(批准号: 10YJCZH052, 12YJCZH201)资助的课题.

† 通讯作者. E-mail: huangxx@hdu.edu.cn

人群中的个体定义成三种状态: 易感状态 S (susceptible)、感染状态 I (infected) 和恢复状态 R (recovered), 建立三种状态个体数量变化的微分方程. 随后发展出了 SI , SIS , $SIRS$ 等模型, 基于传染病模型对信息传播过程的研究是近年来一大热门研究方向. Moreno 等^[12] 基于 SIR 模型将人们分为 Ignorants, Spreaders 和 Stiflers 三种状态, 建立谣言传播模型; Xiong 等^[13] 将 Twitter 网络中的用户分成 Susceptible, Infected, Contacted 和 Refractory 四种状态, 建立 SCIR 模型; Liu 等^[14] 建立无标度网络上带时滞的 SIS 模型; Chen 和 Sun^[15] 提出一种异构网络上最优控制的 $SIRS$ 模型; Gong 等^[16] 建立含个体警惕性的 SIR 模型; Lu 等^[17] 建立无标度网络上具有预防接种的 SIR 模型; 张彦超等^[18] 考虑了节点度和传播机理的影响, 建立一个基于在线社交网络的信息传播模型; 熊熙和胡勇^[19] 对社交网络上的观点传播动力学进行了研究.

本文通过对微博传播网络的结构分析来探索微博传播过程. 利用新浪微博数据, 建立微博传播网络, 分析该网络的生成机制, 使用平均场论的方法^[20], 推导微博传播网络的度分布模型. 深入挖掘微博传播过程, 对研究舆论、谣言、疾病等的传播行为具有指导作用, 进而能够对这些传播行为进行有效的预防控制.

2 微博传播规律分析

2.1 微博传播网络的构建

微博消息的传播是在微博用户网络上进行的, 如图 1(a) 所示. 微博上的注册用户 A 关注用户 B 成为 B 的粉丝, D 关注 C, C 和 A 相互关注, 构成一个用户网络.

当用户发布公开微博的时候, 该消息会实时推送到其粉丝用户的首页, 粉丝在阅读到该消息后可以采取相应的操作 (评论、转发或收藏). 如果某个粉丝对该微博消息进行了转发操作, 则该粉丝的粉丝用户同样也能实时接收到该转发微博. 这种转发功能保留了原微博消息, 防止消息在传播过程中被篡改. 从中我们可以看出, 微博消息主要通过粉丝用户的转发行为在用户网络上进行传播.

如在图 1(b) 中, t_1 时刻用户 A 发布一条公开微博 Mes (令 $t_1 = 0$), t_2 时刻 A 的粉丝 B, C 对 Mes 进行转发, t_3 时刻 D 转发 C 的微博, t_4 时刻 C 回复 D. 微博转发过程中存在相互转发 (如用户之间进行短期交流, 如 C 和 D) 及用户可转发多个不同用户 (如 E 在 t_5 时刻从 C 转发, t_6 时刻再次从 D 转发), 最终 A 发布的微博 Mes 被用户 B, C, D, E 转发传播, 依此构成微博消息 Mes 的传播网络.

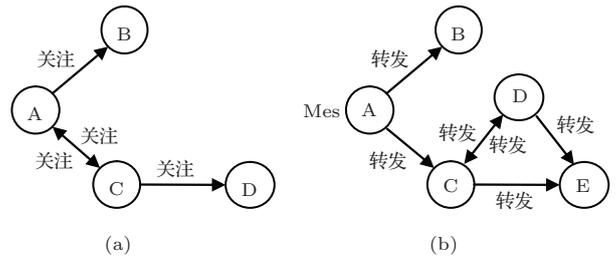


图 1 微博用户网络和微博传播网络 (a) 用户网络; (b) 微博传播网络

用 $G = (V, E, T, W)$ 表示节点数为 N 、边数为 M 的微博传播网络, 微博上 N 个不同用户群体构成网络的节点集合 $V = (v_1, v_2, \dots, v_N)$; 用户间的转发关系构成网络的边集 $E = (e_1, e_2, \dots, e_M)$, 其中任意一条边对应一对节点的二元组 $e_{ij} = (v_i, v_j)$, 表示用户 v_i 转发用户 v_j ; $T = (t_1, t_2, \dots, t_N)$ 表示节点的创建时间; $W = (w_1, w_2, \dots, w_M)$ 表示边的创建时间. 一般地, 一条微博在传播过程中随时间的推移, 转发量增加, 转发用户增加. 用 G_t 表示 t 时刻的传播网络, $G_1 \rightarrow G_2 \rightarrow \dots \rightarrow G_t$ 可以反映一条微博在不同时刻的传播状态, 刻画了信息的传播过程.

在 t 时刻结束时, 从网络中随机选择一个节点, 其度数 k 的概率记为 $P(k, t)$, 称为 t 时网络的瞬时度分布. 随着网络的演化, 如果极限 $\lim_{t \rightarrow \infty} P(k, t) = P(k)$ 存在, 则称 $P(k)$ 为网络的稳态度分布^[21].

2.2 微博传播网络的实验分析

一条微博是如何传播的呢? 微博传播网络 G 如何生成? 本文利用爬虫程序从新浪 weibo.com 上收集到 338 条微博转发信息, 其中包括这些微博的所有转发用户及用户转发微博的时间. 对每条微博数据集进行整理, 构建了 338 个微博传播网络, 作为实验的测试集 S . 对 S 上的所有传播网络节点

数 N 和边数 M 画图分析, 如图 2 所示.

图 2 中每个点表示一条微博的传播网络, 横坐标对应于该网络的节点数, 纵坐标对应于边数. 用直线拟合得到 $M \approx 1.01N$, 表明节点数和边数近似相等, 网络 G 每新增加一个节点就产生一条连边, 这也表明大部分微博用户对感兴趣的微博只转发一次, 只有极少用户对一条微博多次转发.

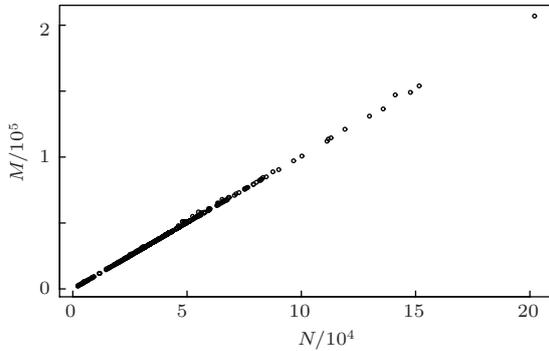


图 2 测试集 S 上的所有传播网络节点数 N 和边数 M 的关系

Barabási 在文献 [5] 中提出, 在一个复杂网络中, 当一个新建的节点想加入到网络中或者网络中的节点想与其他节点相关联, 它会倾向于和网络中已存在的优先级较高的节点相连接. 那么在微博传播网络中, t 时刻一个新增的节点 v_j 倾向于和 $t-1$ 时刻网络 G_{t-1} 中哪个节点进行连边呢? t 时刻用户 v_j 会从哪个关注用户转发微博 Mes?

2002 年 Newman 在文献 [22, 23] 引入了相关系数作为研究度-度相关性的重要测量, 定义如下:

$$r = \frac{M^{-1} \sum_{e_{ij}} k_i k_j - \left[M^{-1} \sum_{e_{ij}} \frac{1}{2} (k_i + k_j) \right]^2}{M^{-1} \sum_{e_{ij}} \frac{1}{2} (k_i^2 + k_j^2) - \left[M^{-1} \sum_{e_{ij}} \frac{1}{2} (k_i + k_j) \right]^2}, \quad (1)$$

其中, k_i 表示节点 v_i 的度, M 是网络的边数, $e_{ij} = (v_i, v_j)$ 表示节点 v_i 和 v_j 产生的连边. 显然 $-1 \leq r \leq 1$. Newman 称 $r < 0$ 的网络为异配网络, $r > 0$ 的网络为同配网络, $r = 0$ 的网络为混配网络.

一般地, 同配就是指属性相近的节点倾向于互相连接. (1) 式定义的是节点度相关性, 如果度大的节点倾向于连接度大的节点, 那么网络是同配网络, 如果度大的节点倾向于连接度小的节点, 那么

网络是异配网络. 为了分析节点时间相关性, 本文定义节点时间相关系数如下:

$$rt = \frac{M^{-1} \sum_{e_{ij}} t_i t_j - \left[M^{-1} \sum_{e_{ij}} \frac{1}{2} (t_i + t_j) \right]^2}{M^{-1} \sum_{e_{ij}} \frac{1}{2} (t_i^2 + t_j^2) - \left[M^{-1} \sum_{e_{ij}} \frac{1}{2} (t_i + t_j) \right]^2}, \quad (2)$$

其中, t_i 表示节点 v_i 的创建时间, $-1 \leq rt \leq 1$. 如果创建时间大的节点倾向于连接创建时间大的节点, 那么网络是同配网络, 这说明新建节点倾向于和最新节点进行连接; 如果创建时间大的节点倾向于连接创建时间小的节点, 那么网络是异配网络, 表明新建节点倾向于和“老”节点发生连接.

计算测试集 S 上所有传播网络的 r 值和 rt 值, 发现 r 值都小于 0, 平均值为 -0.068 ; rt 值都大于 0, 平均值为 0.15. 通过分析不同传播网络的规模 N 与 r 和 rt 的关系, 得到如图 3 所示的结果.

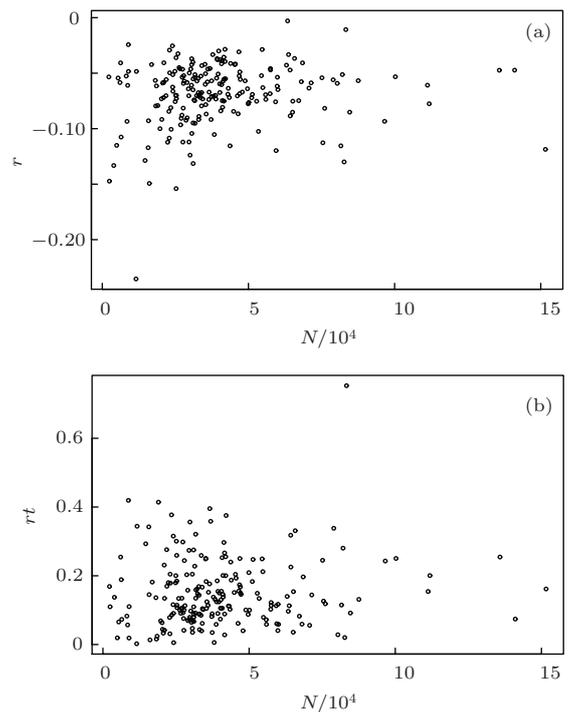


图 3 传播网络的规模 N 与 r 值和 rt 值的关系 (a) $N-r$; (b) $N-rt$

图 3 表明, 传播网络中度小的节点倾向于和度大的节点连接, 在节点度方面表现为异配性; 创建时间相近的节点倾向于相互连接, 说明在时间上表现为同配性. 传播网络的度异配性说明微博用户倾向于从转发量大的用户处转发微博, 传播网络的时

间同配性是由于新浪微博的实时推送机制, 微博用户总是选择当前系统推送的最新微博进行转发.

3 微博传播网络模型

3.1 传播网络模型的构建算法

平均场方法由Barabási和Albert^[20]提出, 主要思想是建立 t 时刻节点 i 的度 $k_i(t)$ 满足的动力学方程, 求出 $P(k, t)$ 并取极限得到 $P(k)$. 利用平均场论的方法推导微博传播网络模型, 详细过程如下.

1) 初始: 开始给定1个节点.

2) 增长: t 时刻增加 $m_t = mt^\theta$, $\theta > 0$ 个新节点, 每个新节点产生一条连边.

3) 择优: t 时刻新建节点与 l 时刻创建的节点 v_i 按择优概率 $\Pi(k_i(l, t))$ 建立连边, 不允许重复连线. 其中, 择优概率定义为

$$\Pi(k_i(l, t)) = \frac{k_i}{\sum_j k_j}$$

$t > l$, $k_i(l, t)$ 表示 l 时刻创建的节点 v_i 在 t 时刻的度数. 假设 t 时刻随机选择 l 时刻的节点服从 $P(l)$ 分布.

步骤1)表示初始网络, 一个微博传播网络由初始用户发布微博消息Mes开始逐步构建; 步骤2)体现了在不同时刻 t , 微博Mes的转发量为 m_t ; 步骤3)的择优概率 $\Pi(k_i(l, t))$ 描述的现象是用户一般倾向于从转发量大的用户转发Mes; $P(l)$ 表示用户倾向于选择当前微博系统实时推送的最新Mes, 例如, 令 $P(l) = \frac{l}{t}$, 当 l 越大时, $P(l)$ 越大, 越新的微博被选择的概率越大.

根据连续性理论, 把 $k_i(l, t)$ 看作连续动力学函数, l 时刻创建的节点 v_i 在 t 时刻的度数 $k_i(l, t)$ 近似地满足如下动力学方程:

$$\frac{\partial k_i}{\partial t} \approx m_t \Pi(k_i) = \frac{(\theta + 1)k_i}{2t}, \quad k_i(l, l) = 1, \quad (3)$$

求得(3)式的解为 $k_i(l, t) = \left(\frac{t}{l}\right)^{\frac{\theta+1}{2}}$.

3.2 独立于演化时间的传播网络模型

假设 $P(l)$ 服从均匀分布, 即 $P(l) = \frac{l}{t}$, 由(3)式的解推导出网络瞬时度分布如下:

$$P(k_i(l, t) < k)$$

$$= P\left(l > tk^{-\frac{2}{\theta+1}}\right) = 1 - k^{-\frac{2}{\theta+1}}, \quad (4)$$

$$P(k, t) = \frac{\partial P(k_i(l, t) < k)}{\partial k} = \frac{2}{\theta + 1} k^{-\frac{\theta+3}{\theta+1}}, \quad (5)$$

(5)式的结果表明传播网络瞬时度分布服从幂律分布, 且独立于时间.

那么, 真实的微博传播网络的瞬时度分布是否服从幂律分布, 是否独立于时间? 本文选择测试集 S 上的一条微博做实验分析. 该微博是一名微博用户(非加V用户)2013年6月14日16:02发布的揭示食品安全的图片长文微博, 该微博截止2013年7月03日17:37共有135810名用户参与转发, 转发量为136524. 结果如图4所示.

图4(a)右上的矩形是对 $k \in [0, 35]$ 区域的放大. 从图4(a), (b)看出, 微博传播网络的瞬时度分布服从幂律分布, 具有长尾特征, 大部分节点的度比较小, 只有少数的度大节点. 对双对数坐标系下的瞬时度分布做曲线拟合, 求出 $P(k) \sim k^{-4.074}$, 幂指数 $\gamma = 4.074$.

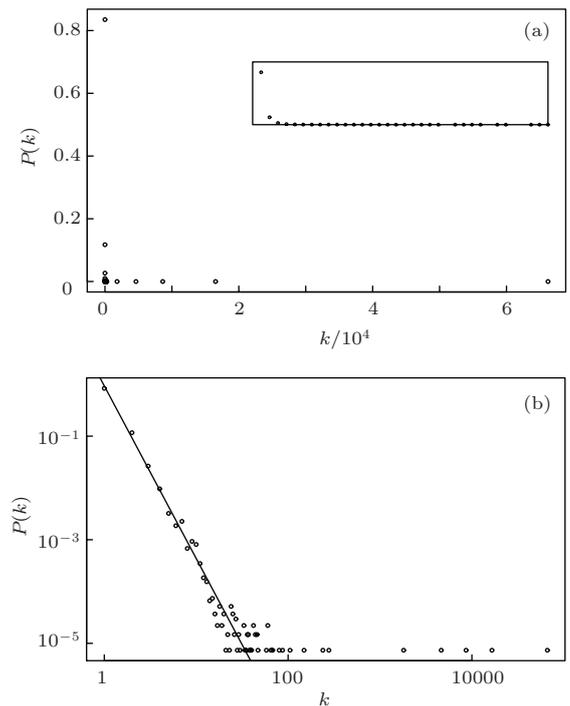


图4 传播网络在不同坐标系下的瞬时度分布 (a) 瞬时度分布; (b) 双对数坐标系下瞬时度分布

分析该条微博 $t = 700, 1100, 3000$ 时的传播网络 $G_{700}, G_{1100}, G_{3000}$, 计算它们在双对数坐标系下的度分布, 如图5所示.

显然, 微博传播网络的度分布是依赖于演化时

间的, 幂指数随时间 t 缓慢递减.

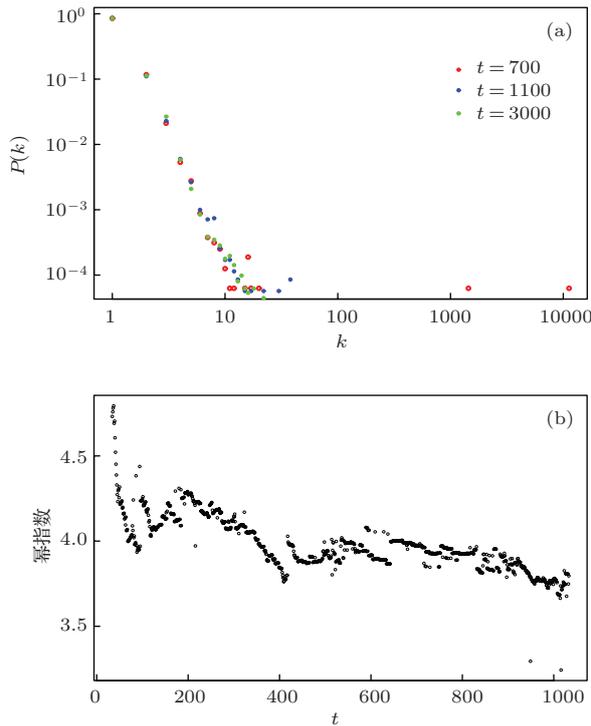


图5 传播网络的度分布与时间 t 的关系 (a) $t = 700, 1100, 3000$ min 时的度分布; (b) 不同时刻 γ 的变化

3.3 依赖演化时间的传播网络模型

文献 [24, 25] 中定义了链接的创建与节点的创建的时间关系计算公式. 本文在此基础上稍做改进, 定义新节点的创建与旧节点的创建时间关系计算公式如下:

$$P_{\text{age}}(a) = \frac{|\{v(t) : e_{ij}(t) \wedge t_i - t_j = a\}|}{\sum_t |\{n : T - t_n \geq a\}|}, \quad (6)$$

其中, $P_{\text{age}}(a)$ 表示新建一个节点 v_i 时, v_i 和时间间隔为 a 的节点 v_j 相连接的概率, $e_{ij}(t) = \{v_i(t), v_j(t)\}$ 表示 t 时刻节点 v_i 和节点 v_j 相连的边, t_i 表示节点 v_i 的创建时间. 统计 $P_{\text{age}}(a)$ 随 a 的变化, 如图 6 所示. 新建节点 v_j 与时间间隔为 a 节点 v_i 相连接的概率随 a 幂律递减, 即 $P(a) \sim a^{-\beta}$, $\beta > 0$.

显然, $P(l)$ 并不服从均匀分布, 假设 $P(l) = (t - l)^\lambda$, $\lambda < 0$, 推导网络瞬时度分布如下:

$$P(k_i(l, t) < k) = 1 - t^\lambda \left(1 - k^{-\frac{2}{\theta+1}}\right)^\lambda, \quad (7)$$

$$P(k, t) = \frac{\partial P(k_i(l, t) < k)}{\partial k} = -\frac{2\lambda}{\theta+1} t^\lambda \left(1 - k^{-\frac{2}{\theta+1}}\right)^{\lambda-1} k^{-\frac{\theta+3}{\theta+1}}. \quad (8)$$

(8) 式也验证了微博传播网络的度分布是时间相依的, 传播网络的瞬时度分布服从于幂律分布, 不同 t 时刻网络的瞬时度分布不同, 微博传播网络模型的度分布依赖于演化时间.

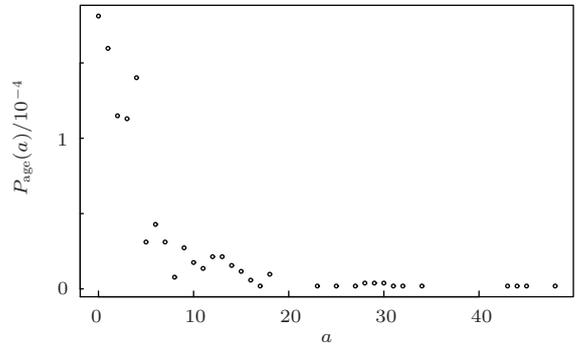


图6 $P_{\text{age}}(a)$ 随 a 的变化, 其中 $P_{\text{age}}(a) \sim a^{-1.43}$

3.4 实验统计分析

一条微博传播网络是怎样演化成幂律分布的呢? 图 7 给出了一条微博在 $t = 5, 10, 60, 120, 600$ min 时的度分布.

图 7 (a)—(e) 表明, 微博的传播方式是典型的“蒲公英式”传播, 开始一段时间内, 消息 Mes 从一个用户发布, 其他多个用户转发, 慢慢演化成以其他多个用户为中心再次扩散, 随着时间 t 增加, 中心用户数量逐渐增多, 中心用户的度数也逐渐增大, 转发微博的数量随之增大, 新增用户量也增大, 微博传播网络度分布慢慢趋近于幂律分布.

计算图 7 (c), (d), (e) 的幂指数分别为 $\gamma_c = 4.16$, $\gamma_d = 4.07$, $\gamma_e = 3.92$. 由于微博“蒲公英式”的传播方式, 每条微博都有一个阈值时间 t_r , 当 $t \geq t_r$ 时网络的瞬时度分布服从于幂律分布; 不同内容的微博, 它们的传播过程不同, 由于中心用户转发微博 Mes 的偶然性和传播过程中中心用户的数量及规模的影响, 使得每条微博的 t_r 不尽相同.

最后, 本文对测试集 S 上所有微博传播网络在不同时刻的度分布进行验证, 取 $t = 7200$ min 时的度分布进行统计, 结果如图 8 所示.

如图 8 所示, 测试集 S 上所有微博传播网络的幂指数 $\gamma \in [2.93, 7.92]$, 平均幂指数 $\hat{\gamma} = 4.92$. 显然, 当时间 t 大于或等于阈值 t_r 时, 所有微博网络的瞬时度分布服从幂律分布.

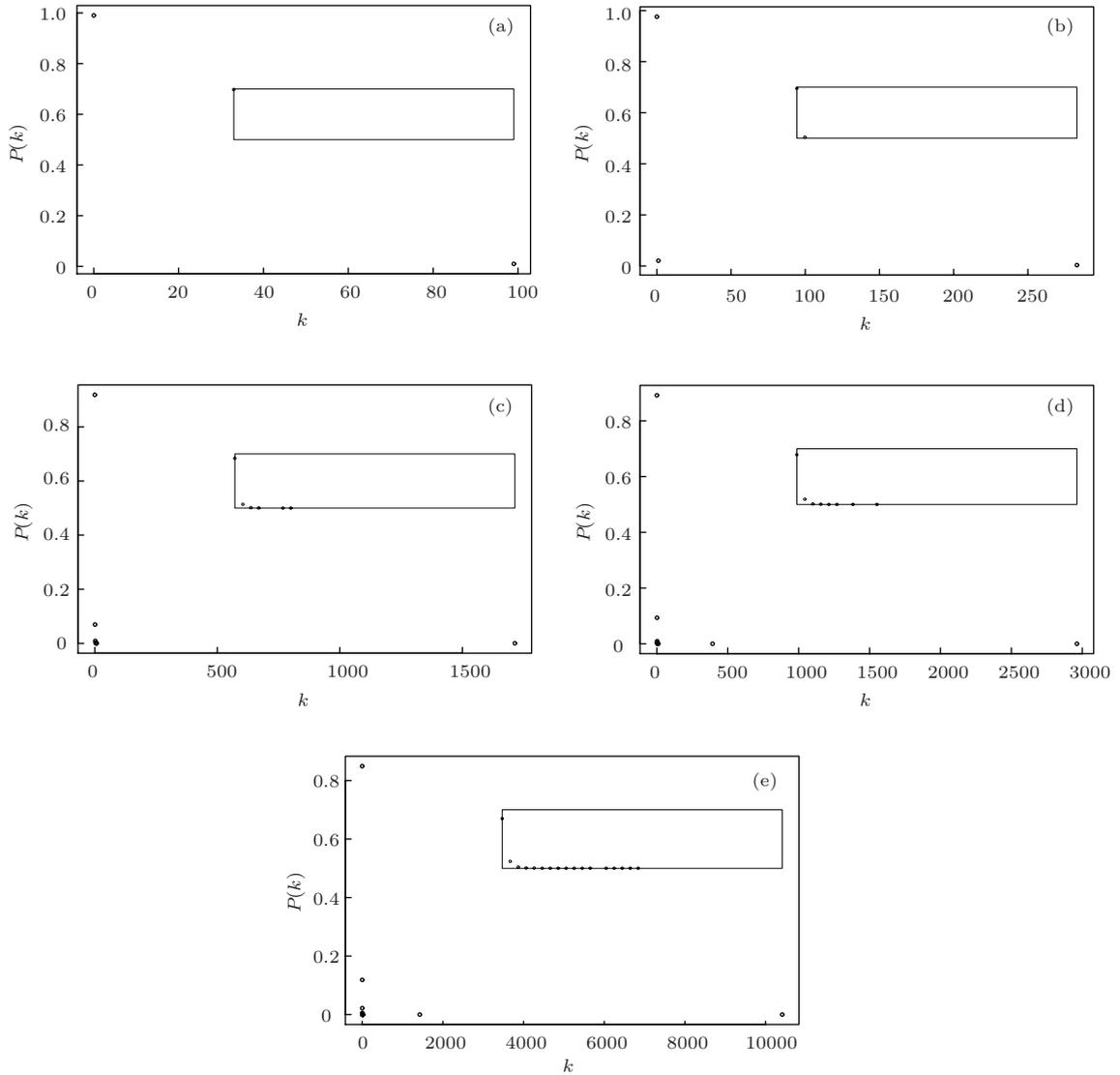


图7 不同时刻传播网络的度分布 (a) $t = 5 \text{ min}$; (b) $t = 10 \text{ min}$; (c) $t = 60 \text{ min}$; (d) $t = 120 \text{ min}$; (e) $t = 600 \text{ min}$

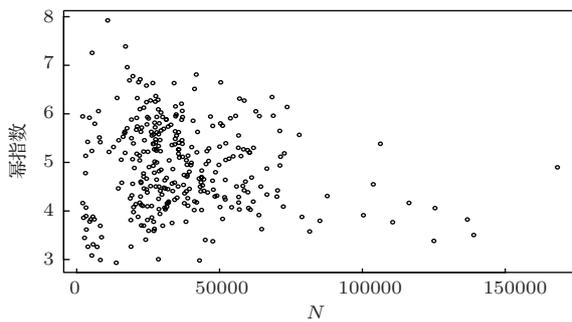


图8 测试集 S 上所有微博传播网络在 $t = 7200 \text{ min}$ 时的度分布

4 结 论

新浪微博是在通过用户关注机制建立的用户网络上分享实时信息的社交平台, 而微博消息主要通过用户的转发行为使消息在用户网络上传播. 本

文以微博上参与一条微博消息传播的所有用户作为节点集合, 用户之间的转发关系作为边集, 构建微博传播网络. 通过对测试数据集 S 实验分析, 发现用户对于一条微博 Mes 的转发行为, 其倾向于从度大的、最新转发该微博的用户转发它, 将微博传播过程中的这种特性作为微博传播网络演化的择优准则, 设定传播网络的增长特性, 从而构建了微博传播网络生成算法. 最终推导出两种微博传播网络模型: 独立于演化时间的传播模型和依赖演化时间的传播模型. 通过实验验证, 微博传播网络模型依赖于演化时间, 微博的传播方式是典型的“蒲公英式”传播, 每条微博的传播网络都有一个阈值 t_r , 当时间 t 大于或等于 t_r 时, 微博网络的瞬时度分布服从幂律分布. 与 BA 模型、复制模型、局域世界演

化模型不同, 传播网络模型具有传播的特性, 对研究信息传播动力学有指导意义. 在后续工作中, 我们将增加对不同信息主体的传播进行对比分析, 以期挖掘微博传播过程中更多的规律, 构建更加完善的传播网络模型.

参考文献

- [1] Erdos P, Rényi A 1961 *Bull. Inst. Internat. Statist* **38** 343
- [2] Erd Hos P, Rényi A 1959 *Publ. Math.: Debrecen* **6** 290
- [3] Solomonoff R, Rapoport 1951 *B Math. Biol.* **13** 107
- [4] Watts D J, Strogatz S H 1998 *Nature* **393** 440
- [5] Barabási A, Albert R 1999 *Science* **286** 509
- [6] Dorogovtsev S N, Mendes J F F 2001 *Phys. Rev. E* **63** 25101
- [7] Krapivsky P L, Redner S 2005 *Phys. Rev. E* **71** 36118
- [8] Li X, Chen G 2003 *Physica A* **328** 274
- [9] Pastor-Satorras R, Vespignani A 2001 *Phys. Rev. Lett.* **86** 3200
- [10] Watts D J 2004 *Annu. Rev. Sociol.* **30** 243
- [11] Kermack M, Mckendrick A G 1927 *Proc. R. Soc. A London*, August 1, 1927 p700
- [12] Moreno Y, Nekovee M, Pacheco A F 2004 *Phys. Rev. E* **69** 66130
- [13] Xiong F, Liu Y, Zhang Z, Zhu J, Zhang Y 2012 *Phys. Lett. A* **376** 2103
- [14] Liu Q, Deng C, Sun M 2014 *Physica A* **410** 79
- [15] Chen L, Sun J 2014 *Physica A* **410** 196
- [16] Gong Y W, Song Y R, Jiang G P 2012 *Chin. Phys. B* **21** 010205
- [17] Lu Y L, Jiang G P, Song Y R 2012 *Chin. Phys. B* **21** 100207
- [18] 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]
- [19] Xiong X, Hu Y 2012 *Acta Phys. Sin.* **61** 150509 (in Chinese) [熊熙, 胡勇 2012 物理学报 **61** 150509]
- [20] Barabási A, Albert R, Jeong H 1999 *Physica A* **272** 173
- [21] Shi D H 2011 *Theory of Network Degree Distributions* (Beijing: Higher Education Press) p60 (in Chinese) [史定华 2011 网络度分布理论 (北京: 高等教育出版社) 第60页]
- [22] Newman M E 2002 *Phys. Rev. Lett.* **89** 208701
- [23] Newman M E 2003 *Phys. Rev. E* **67** 26126
- [24] Leskovec J, Backstrom L, Kumar R, Tomkins A 2008 *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data mining* Las Vegas, NV, USA, August 24–27, 2008 p462
- [25] Allamanis M, Scellato S, Mascolo C 2012 *Proceedings of the 2012 ACM Conference on Internet Measurement Conference* Boston, MA, USA, November 14–16, 2012 p145

Microblog propagation network model based on mean-field theory*

Wu Teng-Fei¹⁾ Zhou Chang-Le¹⁾²⁾ Wang Xiao-Hua¹⁾ Huang Xiao-Xi^{1)†}
Chen Zhi-Qun¹⁾ Wang Rong-Bo¹⁾

1) (*Institute of Cognitive and Intelligent Computing, School of Computer Science and Technology, Hangzhou Dianzi University, Hangzhou 310018, China*)

2) (*Fujian Key Laboratory of the Brain-like Intelligent Systems, Cognitive Science Department, Xiamen University, Xiamen 361005, China*)

(Received 7 April 2014; revised manuscript received 8 August 2014)

Abstract

Microblog is a social medium platform allowing users to share real-time information through the user relationship networks which are established by the mechanism of “following”. Message is spread through the user relationship networks by users’ behavior of “retweeting”. The research on the spread mechanism of microblog message is very important for public sentiment analysis and rumor propagation control. In this paper, we intend to explore the process of microblog propagation through structural analysis of microblog propagation network. First, we adopt the data of Sina Microblog to establish microblog propagation network, then analyze the formation mechanism of this network, finally, derive the degree distribution model of microblog propagation network by mean-field theory. Experimental results show that the degree distribution of microblog propagation network is a time-dependent model, and it is a power-law distribution at specific time.

Keywords: complex networks, microblog propagation network, scale-free network, mean-field theory

PACS: 05.10.-a, 89.75.-k

DOI: 10.7498/aps.63.240501

* Project supported by the Young Scientists Fund of the National Natural Science Foundation of China (Grant Nos. 61103101, 61202281) and the Humanity and Social Sciences Foundation for Young Scholars of China’s Ministry of Education, China (Grant Nos. 10YJCZH052, 12YJCZH201).

† Corresponding author. E-mail: huangxx@hdu.edu.cn