

用户行为驱动的在线社交网络建模

李倩倩^{1,2}, 顾基发^{2,3}

(1. 中国科学院科技政策与管理科学研究所, 北京 100190;

2. 中国科学院自然与社会交叉科学研究中心, 北京 100190;

3. 中国科学院数学与系统科学研究院, 北京 100190)

摘要: 基于节点威望和节点相似性这两个驱动在线社交网络连接形成的潜在机制, 从用户连接行为入手, 本文提出一个在线社交网络生成模型. 采用优先连接刻画网络无标度特性; 从微博好友推荐机制建模节点趋同性. 该网络模型考虑了社交网络形成的一些重要驱动因素, 如同质性和聚类性. 从网络连接的微观交互中寻求刻画网络宏观特性的连接机制是本文研究的出发点. 模型有力刻画了在线社交网络的统计特性, 如度分布、聚类系数、度相关性. 结果表明节点威望吸引和节点相似性是在线社交网络结构形成的内在机理.

关键词: 在线社交网络; 优先连接; 相似性

中图分类号: TP273

文献标识码: A

文章编号: 1000-5781(2015)01-0009-07

doi: 10.13383/j.cnki.jse.2015.01.002

Activity driven modelling of online social network

Li Qianqian^{1,2}, Gu Jifa^{2,3}

(1. Institute of Policy and Management, Chinese Academy of Sciences, Beijing 100190, China;

2. Center for Interdisciplinary Studies of Natural and Social Sciences, Chinese Academy of Sciences, Beijing 100190, China;

3. Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China)

Abstract: Popularity and homophily were the latent mechanisms for the link formation of online social network(OSN). This paper proposed an OSN link formation model from the perspective of user behavior. The authors used preferential attachment to characterize scale free property of OSN. By means of friendship recommendation, they could depict high clustering and assortativity of OSN. The fundamental factors for link formation in OSN were included: homophily and clustering. This paper started from the microscopic interaction process of people making friends in an OSN to characterize OSN's macroscopic statistical properties. This model reproduces degree distribution, clustering and degree correlation of OSN. Simulation results show that popularity and similarity are the important internal mechanisms of OSN structure formation.

Key words: online social network; preferential attachment; homophily

1 引言

随着Web2.0的迅速普及和高覆盖率, 在线社交网络(online social network, OSN)成为信息传播的主力军.

收稿日期: 2013-08-13; 修订日期: 2014-02-27.

基金项目: 国家重点基础研究发展计划(973计划) 资助项目(2010CB731500); 国家自然科学基金资助项目(91024010; 11001106; 91324009); 国家自然科学基金青年基金资助项目(71403262); 中国科学院科技政策与管理科学研究所重大科研资助项目(Y201201Z06).

潮流趋势、产品营销、创新推广、谣言等的扩散都离不开OSN的身影. 与传统Web网页以内容为主体吸引用户不同, OSN以人为中心, 吸引兴趣相投的人先建立连接, 再以“人人都是自媒体”的形式向朋友推送信息. 所以OSN的网络结构与传统Web网页不同^[1], 而网络结构对社会传播行为具有重要影响^[2,3], 研究网络拓扑结构对理解和掌握社会传染过程具有现实意义.

近年来, OSN的结构建模吸引了研究者的兴趣, 涌现出一批对网络的动态性和结构探测的OSN生成模型. 胡海波等从连接建立的双向交互行为入手, 建模中国若邻网(Wealink)的时序演化过程^[4]. Leskovec等将网络生成分解为三个过程^[5]: 1) 新增节点到达; 2) 边初始化过程; 3) 边的目标节点选择过程. 并利用最大似然估计评估网络模型对现实网络的最大逼近参数. 项荣璟等进一步将OSN中连接分为素未谋面和密友关系, 建模连接的强度^[6]. 文献[7]有一章分析Flickr和雅虎360问答两个群体内容分享网络中的社团结构呈现: 1) 单体结构(singleton); 2) 孤立集团(isolated communities); 3) 大连通图(giant component). 他们将网络中的节点划分为消极参与者、活跃者和组织者三类, 采取有偏优先连接算法建模网络中的社团结构. 文献[8]综述了部分社交网络模型, 将其分为基于网络演化模型(network evolution models, NEMs)和基于节点属性的网络模型(nodal attribute models, NAMs). 2012年, Lev Muchnik等认为优先连接机制不能完全解释度分布的幂律特性. 他采用因果推理, 分析社交网络中无标度幂律分布的形成, 发现胖尾效应可能由人类活动规律的偏态分布决定, 个体连接度的形成遵循“最大熵连接”模型^[9].

对国内外大型社交网站Flickr、Livejournal^[11]、Facebook^[10]和新浪微博^[11]的统计研究表明OSN结构普遍具有的特性是: 1) 度为幂律分布, OSN中绝大多数节点的度很低, 只有少部分节点的度比较大, 网络呈现无标度特性; 2) 具有较高聚类特性, 即节点的邻居之间成为朋友的可能性很大, 随着节点度增加, 聚类系数出现减小的趋势. 3) 同配性, 表现为度大的节点更倾向于和度大的节点连接.

OSN结构的度分布是网络的全局特性, 聚类系数和同配性是网络的局部特性. 本文试图从个体微观交互入手, 刻画OSN呈现的上述统计特性. 从在线社交网络中用户的行为中提炼出两种OSN连接形成的动力机制: 名人效应(popularity)和趋同性(homophily).

富者越富, 好者越富是名人效应的重要实现机制, 表达了网络度分布的幂律特性, 体现了节点威望吸引, 很多实际的网络都具有该特点. 优先连接^[12]机制及其变体^[13-16]很好地诠释了网络的无标度特性.

节点的相似性是社交网络产生的原动力. 最初LinkedIn、人人网等都是好友、同学之间的连接, 需要一方发起邀请, 另一方接受邀请才能建立连接^[4], 是将现实社会中的熟人关系移植到互联网中. 社会学研究表明同质性(homophily)在社交网络中具有重要作用^[8], 体现在人们更愿意和有相似性的用户建立关系. 新浪微博中采用“微关系”即用户之间朋友的联系作为好友推荐的依据, 正是考虑了用户的局部相似性, 而这点将引导网络连接朝向较高的网络聚类系数以及同配性发展.

2 在线社交网络生成模型

结合OSN的基本统计特性, 本文提出的OSN生成模型一方面利用优先连接机制刻画网络度的幂律分布; 另一方面根据社交网络好友推荐机制(图1, 表1)建模OSN的局部相似性. OSN中有三种主要的好友推荐机制: 1) 基于共同邻居的好友推荐, 即“两个节点之间共同拥有的朋友越多就越可能建立连接”, 于是两个度大的节点之间拥有较多共同邻居的概率越大, 这正是网络同配性产生的驱动力; 2) 基于你朋友的朋友也可能是你朋友的推荐, 即“我关注的人也关注他”, 体现“物以类聚, 人以群分”的网络特性, 是网络高聚类性产生的驱动力. 3) 基于网络中“三度影响力”理论^[17], 认为我们所做或所说的任何事情, 都会在网络上泛起涟漪, 影响我们的朋友(一度), 我们朋友的朋友(二度), 甚至我们朋友的朋友的朋友(三度). 如果超出三度分隔, 我们的影响就逐渐消失. 同样, 我们也深受三度以内朋友的影响, 但一般来说, 超出三度的朋友就影响不到我们了. 相距三度以内的人之间是强连接关系. 好友推荐中“他的粉丝还关注了谁”, 体现了“三度影响”在用户连接上的应用, 也是对网络聚类特性的刻画.

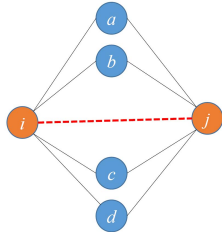
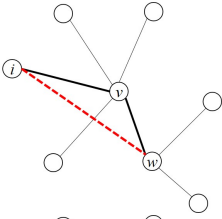
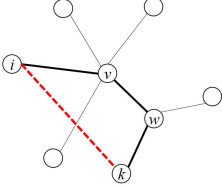


图 1 微博用户推荐—“微关系”

Fig. 1 Weibo friendship recommendation mechanism based on micro relationships

表 1 基于微关系的网络连边机制—模型解释

Table 1 Link formation in network model based on micro relationships—model explanation

微关系	模型解释	连接图示
“共同关系”	基于共同邻居的连接：基于节点之间共同邻居的数目，共同邻居越多，节点间建立连接的可能性越大。	 $\Gamma(i) = \{a, b, c, d\}$ $\Gamma(j) = \{a, b, c, d\}$ $CN(i, j) = \Gamma(i) \cap \Gamma(j) = 4$
“我关注的人也关注他”	基于三元形成的连接：基于“朋友的朋友也是我的朋友”这一共识，节点将和邻居的邻居之间建立连接。	
“他的粉丝还关注了”	基于三度影响力的连接：基于“三度影响力”作用，节点也可能和邻居的邻居的邻居之间建立连接。	

本文提出的在线社交网络结构生成算法引入如下连边机制:

1) 优先连接(preferential attachment): 一个新节点与一个已经存在的节点 v 相连接的概率 Π_v 与节点 v 的度 k_v 成正比: $\Pi_v = \frac{k_v}{\sum_{j=1}^N k_j}$, N 为系统中节点总数;

2) 基于共同邻居的连接(link based on common neighbors): 计算网络中未连接节点对之间的共同邻居的数量, $CN(i, j) = |\Gamma(i) \cap \Gamma(j)|$, 以概率 $p_{ij} \propto CN(i, j)$, 建立节点 i 和节点 j 之间的连接;

3) 基于三元形成的连接(link based on triad formation): 新增节点 i 随机选择邻居的邻居节点 w , 建立节点 i 和 w 之间的连接;

4) 基于三度影响力的连接(link based on three degrees of influence rule): 新增节点 i 随机选择邻居的邻居的邻居节点 k , 建立节点 i 和 k 之间的连接.

具体来讲, 模型网络生成过程为: 从具有 m_0 个节点的网络开始, 每次引入一个新的节点并且连接到 m 个已经存在的节点上, 这里 $m \leq m_0$; 新增节点 i 第一步执行优先连接, 对其余 $(m-1)$ 条边以概率 p_p 执行优先连接, 以概率 p_c 记录进行基于共同邻居的连接数目, 以概率 p_{tf} 执行基于三元形成的连接, 以概率 p_{td} 执行基于三度影响力的连接. 最后, 对平均 $(m-1)p_c$ 执行基于共同邻居的连接. 这里需要说明的是, 基于共同邻居的连接是从网络全局考虑的. 模型中假设 $p_p + p_c + p_{tf} + p_{td} = 1$. 当 $p_p = 1$ 时, 退化为BA无标度网络模型^[12]; 当 $p_p \neq 0, p_{tf} \neq 0, p_c = 0, p_{td} = 0$ 时, 退化为可变聚类系数网络模型^[18].

3 在线社交网络模型仿真

3.1 基本统计指标

1) 度分布(degree distribution)

网络中节点度的分布情况可用分布函数 $P(k)$ 描述. $P(k)$ 表示网络中度为 k 的节点占有所有节点数的比例, 也就是一个随机选定的节点度恰好为 k 的概率. 许多网络的度分布可以用幂律形式 $P(k) \propto k^{-\gamma}$ 来描述, 这表明网络的无标度特性.

2) 聚类系数(clustering coefficient)

社交网络中节点的聚类系数体现了该用户的朋友之间也是朋友的可能性. 定义节点 i 的聚类系数为

$$C_i = \frac{2E_i}{k_i(k_i - 1)}, \quad (1)$$

其中 $k_i(k_i - 1)/2$ 表示节点 i 的 k_i 个邻居节点之间最多可能存在的连边数量. 整个网络的聚类系数 C 就是所有节点 i 的聚类系数 C_i 的平均值, 即 $C = \frac{1}{N} \sum_{i \in N} C_i$. 显然, $0 \leq C \leq 1$.

3) k 近邻(k nearest neighbor)

度相关性(degree correlation)表示一个节点的度与其邻居节点度值之间的相关性. 网络中, 如果度大(小)的节点倾向于连接度大(小)的节点, 则称网络是正相关的, 称之为同配网络. 反之, 则称网络是负相关的, 称为异配网络. 度相关性常用联合度分布、同配系数、 k 近邻等表示. k 近邻指标定义为

$$k_{nn}(k) = \frac{1}{N_k} \sum_i \delta_{k, k_i} k_{nn, i}, \quad (2)$$

$$k_{nn, i} = \frac{1}{k_i} \sum_{j \in v(i)} k_j, \quad (3)$$

其中 $v(i)$ 表示节点 i 的邻居节点集合, k_j 表示节点 j 的度, $k_{nn, i}$ 表示节点 i 的最近邻平均度, δ_{k, k_i} 是克罗内克符号(Kronecker symbol), 当 $k = k_i$ 时, $\delta_{k, k_i} = 1$; 否则, $\delta_{k, k_i} = 0$.

如果函数 $k_{nn}(k)$ 是递增的, 那么网络呈现同配性, 因为表现为度大的节点更倾向于和度大的节点连接. 反之, 如果函数是递减的, 那么网络呈现异配性, 表现为度大的节点更倾向于和度小的节点连接.

3.2 仿真结果

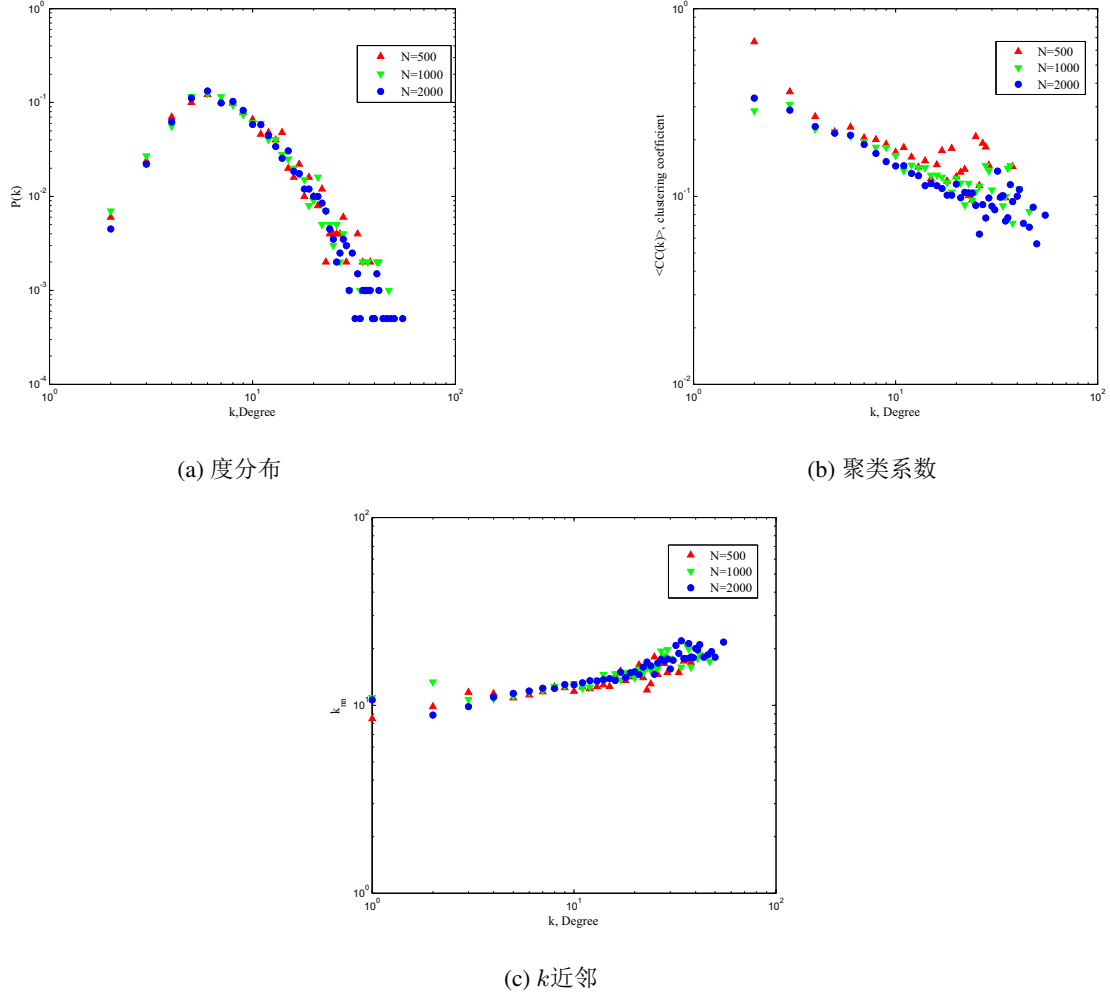


图 2 不同规模下网络模型的统计特性

Fig. 2 Statistical properties of network model under different N

首先, 分析了不同网络规模 $N = 500$, $N = 1000$, $N = 2000$ 情况下, 网络的统计特性(图2). 图2(a, c)看出网络规模对度分布、同配性影响并不显著, 图2(b)中, 当网络规模较小($N = 500$)时, 度大的节点的网络聚类系数变化较大; 当网络规模较大($N = 1000$, $N = 2000$)时, 网络聚类系数之间的差距不大. 为了避免网络规模对实验结果的影响, 本文设置 $N = 1000$. 其它参数设置为 $m_0 = 2m$, $p_p = 0.25$, $p_c = 0.25$, $p_{tf} = 0.25$, $p_{td} = 0.25$. 在 $m = 3, m = 5, m = 7, m = 9$ 四种情况下, 分析模型网络的度分布(图3)、聚类系数(图4)、同配性(图5)这三个基本特征. 网络的统计特征为10次模拟的平均值.

模型网络度分布符合幂律分布形式, 由于网络的平均度与 m 成正比, 随着 m 的增加, 度大的节点概率呈现显著增加趋势.

该模型下, 网络呈现较高的聚类特性. 图4显示了节点度和聚类系数之间的关系. 总体来讲, OSN中节点的度越高, 聚类系数越小. 在 $m = 3, 5, 7, 9$ 条件下, 网络聚类系数 $C = 0.2505, 0.1825, 0.1461, 0.1273$. 随着 m 的增加, 网络聚类系数呈现减小的趋势. 特别地, 当 m 较小时, 聚类系数和节点度呈现线性负相关关系; 当 m 较大时, 度大的节点的聚类系数在尾部出现局部增加. 这一结果与在线社交网站Flickr的研究发现

一致^[1].

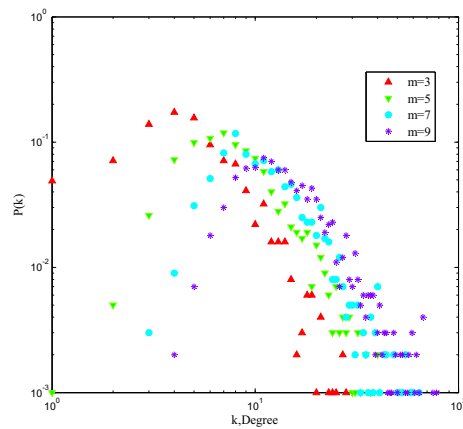


图3 网络模型度分布

Fig. 3 Degree distribution of network model

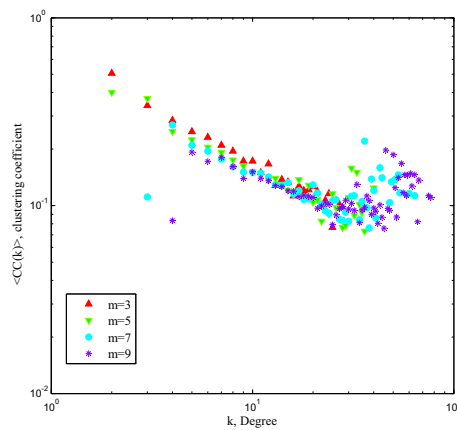


图4 网络模型的聚类系数

Fig. 4 Clustering coefficient of network model

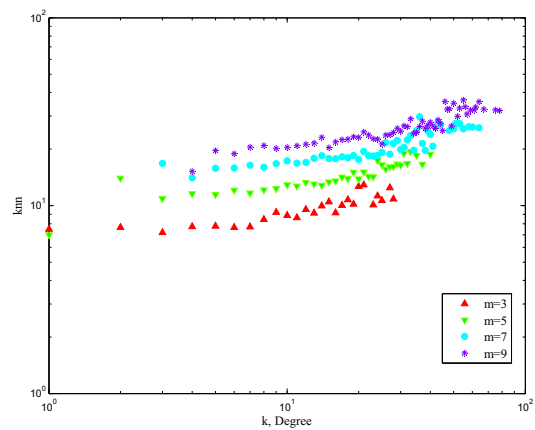


图5 网络模型的 k 近邻

Fig. 5 K nearest neighbor of network model

k_{nn} 刻画了度为 k 的节点邻居的平均度. 图5中 k_{nn} 随着节点度的增加而增加, 显示该模型对OSN同配性的较好模拟. 在 $m = 3, 5, 7, 9$ 条件下, 网络的同配系数 $r = 0.2056, 0.2128, 0.2214, 0.2217$. 随着 m 的增加, 网络的同配性也随之增加.

4 结束语

本文同时从宏观和微观层面提出一种在线社交网络演化机制. 一方面, 优先连接是对网络富者越富特性的宏观体现; 另一方面, 用户连接行为引发三种连边机制: 基于共同邻居的连接, 基于三元形成的连接和基于三度影响力的连接. 这是对OSN同配性、聚类效应从微观层面的很好解读.

参考文献:

- [1] Mislove A, Marcon M, Gummadi K P, et al. Measurement and analysis of online social networks[C]// Proceedings of the 7th ACM Special Interest Group on Data Communication Conference on Internet Measurement. San Diego, 2007: 29–42.
- [2] Centola D. The spread of behavior in an online social network experiment[J]. Science, 2010, 329(5996): 1194–1197.

- [3] 周涛, 汪秉宏, 韩筱璞, 等. 社会网络分析及其在舆情和疫情防控中的应用[J]. 系统工程学报, 2010, 25(6): 742–754.
Zhou Tao, Wang Binghong, Han Xiaopu, et al. Social network analysis and its application in the prevention and control of propagation for public opinion and the epidemic[J]. Journal of Systems Engineering, 2010, 25(6): 742–754. (in Chinese)
- [4] Hu H, Wang X. How people make friends in social networking sites: A microscopic perspective[J]. Physica A: Statistical Mechanics and Its Applications, 2012, 391(4): 1877–1886.
- [5] Leskovec J, Backstrom L, Kumar R, et al. Microscopic evolution of social networks[C]// Proceedings of the 14th ACM Special Interest Group in Knowledge Discovery and Data Mining International Conference on Knowledge Discovery and Data Mining. Las Vegas, 2008: 462–470.
- [6] Xiang R J, Neville J, Rogati M. Modeling relationship strength in online social networks[C]// Proceedings of the 19th International Conference on World Wide Web. Raleigh: North Carolina, 2010: 981–990.
- [7] Han J, Faloutsos C, Philip S Y. Link Mining: Models, Algorithms, and Applications[M]. New York: Springer, 2010: 337–357.
- [8] Toivonen R, Kovanen L, Kivela M, et al. A comparative study of social network models: Network evolution models and nodal attribute models[J]. Social Networks, 2009, 31(4): 240–254.
- [9] Muchnik L, Pei S, Parra L C, et al. Origins of power-law degree distribution in the heterogeneity of human activity in social networks[J]. Scientific Reports, 2013, 3: 1783.
- [10] Viswanath B, Mislove A, Cha M, et al. On the evolution of user interaction in facebook[C]// Proceedings of the 2nd ACM Workshop on Online Social Networks. Barcelona, 2009: 37–42.
- [11] Chen Z, Liu P, Wang X, et al. Follow whom: Chinese users have different choice[DB]. <http://arxiv.org/abs/1212.0167>, 2012–12–1.
- [12] Barabási A L, Albert R. Emergence of scaling in random networks[J]. Science, 1999, 286(5439): 509–512.
- [13] Fortunato S, Flammini A, Menczer F. Scale-free network growth by ranking[J]. Physical Review Letters, 2006, 96(21): 218701.
- [14] Valverde S, Cancho R F, Sole R V. Scale-free networks from optimal design[J]. Europhysics Letters, 2007, 60(4): 512–517.
- [15] Baiesi M, Manna S. Scale-free networks from a hamiltonian dynamics[J]. Physical Review E, 2003, 68(4): 047103.
- [16] Dorogovtsev S, Mendes J, Samukhin A. Size-dependent degree distribution of a scale-free growing network[J]. Physical Review E, 2001, 63(6): 062101.
- [17] Christakis N A. Connected: Amazing Power of Social Networks and How They Shape Our Lives[M]. New York: Brown Company, 2010.
- [18] Holme P, Kim B J. Growing scale-free networks with tunable clustering[J]. Physical Review E, 2002, 65(2): 026107.

作者简介:

李倩倩 (1983—), 女, 河南焦作人, 博士, 助理研究员, 研究方向: 舆论建模与仿真, E-mail: lqqcindy@gmail.com;
顾基发 (1935—), 男, 浙江宁波人, 苏联副博士, 研究员, 研究方向: 系统理论方法, E-mail: jfgu@amss.ac.cn.