

考虑社会选择作用的社交网络生成模型

刘业政, 李玲菲, 孙春华

(合肥工业大学管理学院, 安徽 合肥 230009)

摘要: 针对现有社交网络生成模型在解释网络中观结构和微观特性上的不足, 提出了考虑社会选择作用的 BA 模型(SSBA). 该模型基于社交网络用户的特征分布和同质性形成的内在机理, 在建模过程中考虑了社会影响和社会选择的共同作用. 在仿真数据与真实数据集上的实验表明, SSBA 模型能很好地刻画不同类型社交网络在宏观特性和中观结构上的特点. 发现不同作用机制下网络表现出相异的统计特性, 社会选择作用较强时, 网络度分布逐渐偏离幂律分布, 且网络逐渐向同配网络转变.

关键词: 生成模型; 社交网络; 社会选择; 社会影响

中图分类号: O157.5; TP393.09

文献标识码: A

文章编号: 1000-5781(2019)05-0587-11

doi: 10.13383/j.cnki.jse.2019.05.002

Social selection-aware social network generation model

Liu Yezheng, Li Lingfei, Sun Chunhua

(School of Management, Hefei University of Technology, Hefei 230009, China)

Abstract: To address the weakness of the existing generation models in explaining the mediostructure and microcosmic mechanisms of real social networks, this paper proposes a social network generation model named social selection-aware BA (SSBA). Considering the characteristic distribution of users in social networks and the internal mechanism of homogeneous, the SSBA model combines the mechanisms of social selection and social influence in the modeling process. Experiments on both simulated and real world datasets show that the SSBA can capture the macroscopic perspective and mediostructure in a variety of networks. The experimental results also indicate that networks under different mechanisms show diverse statistical characteristics: when social selection is stronger, the degree distribution gradually represents a departure from power law distribution, and the whole network gradually transforms into the an assortative network.

Key words: generation model; social networks; social selection; social influence

1 引言

近年来, 在线社交网络得到了蓬勃的发展, 涌现出了诸如微博、微信、Facebook、Twitter 以及 Youtube 等众多社交平台, 这使得每个用户都能成为信息的创造者, 并实时地与好友进行信息的共享和交流. 社交网络这种高效的信息传播能力正促使传统的新闻、营销等领域产生深刻的变革, 社交网络的生成与演化因此受到极大的关注. 但现实中, 不同类型的社交网络往往具有不同的生成机制, 而微观生成机制的差异会导致不

收稿日期: 2016-10-18; 修订日期: 2017-05-08.

基金项目: 国家自然科学基金重大资助项目(71490725); 国家自然科学基金重大研究计划重点资助项目(91846201); 国家自然科学基金创新研究群体资助项目(71521001).

同的网络结构,使得信息的产生、传播与扩散模式在不同社交网络上呈现出差异化的特征.研究社交网络的生成机制,有助于深刻认识社交网络的内在规律和本质特征,对于挖掘不同社交网络的商业价值与营销模式具有重要的理论及现实意义.

网络生成机制的研究主要分为实证研究与仿真研究两类.在实证研究方面,现有结论发现网络的生成与演化是外生因素、社会影响以及社会选择共同作用的结果^[1].其中外生因素包括线下的亲缘关系、友缘关系和共同的外部刺激等,具有不可控性.社会影响又称为偏好依附或流行性,即人们选择网络中影响力(度数)大的节点做朋友的倾向^[2,3].社会选择也被称为同质性,表现为具有相似特征的个体选择彼此作为朋友的倾向^[4,5].选择与影响机制间存在明显的反馈效应,这种反馈效应使得社会影响与社会选择很难完全区分开^[6,7].研究者通过在真实社交网络中进行大规模实验,探索网络中社会影响与社会选择的作用^[1,8-16].例如,Zhang等^[15]使用链接预测与似然分析度量大型社交网络的多重演化机制,发现Facebook与Flickr上体现出更强的社会选择作用,在Youtube、ScienceNet以及Epinions等网络上社会影响的作用更强,而在DBLP论文合作网络中两种影响机制都存在.Wang等^[16]在论文合作网络中研究了新节点的连接机制,结果表明学科方向同质性(38%),社会影响(36%)和机构同质性(26%)等机制具有不同的解释能力.

在仿真研究方面,Bzrabási等^[17]首次根据偏好依附提出BA无标度模型,随后许多学者以该模型为基础进行了多方面的扩展,如边重连^[18]、边删除^[19]、局部演化^[20]和多阶段择优连接^[21]等,形成了众多变体模型.国内胡海波等^[22]基于Wealink以及Facebook分析了社交网络中三种优先连接行为,并构建了网络增长模型以模拟网络的演化.上述模型重点分析了社交网络生成过程中的社会影响机制,在诸如无标度性等网络宏观特性上给出了很好的解释^[18-21].Boguná等^[23]与Wong等^[24]则将每个节点对应一个多维特征向量,分别建立了不同的空间随机图模型,并基于网络节点数固定的假设,重点分析了社交网络形成的社会选择机制.Marsili等^[25],Kumpula等^[26]以及李倩倩等^[27]分别基于社交网络中诸如朋友的朋友更有可能成为你的朋友等微关系,构建了社交网络生成模型,这些模型主要基于拓扑结构构造三角关系,不考虑三角关系形成的内在机制.Li等^[28]在离散空间中用相同的节点颜色代表节点间的同质性与亲密关系,基于择优连接提出homophily/kinship模型.Papadopoulos等^[29]开创性地在双曲空间中构造了网络生成的popularity $\times$ similarity(PS)模型,将节点到中心的距离作为节点的流行性,用节点间的角距离解释节点间的相似性,根据节点间的双曲距离进行边连接.该模型复现出了互联网、代谢网络等复杂网络的小世界、无标度等性质.Zuev等^[30]在此基础上假设新节点进入双曲空间的位置与该区域的密度成正比,提出几何择优连接模型GPA.Ferretti等^[31]证明了基于偏好依附生成的网络与双曲空间中的静态随机网络是对偶的,且BA模型就是具有无穷曲率的双曲空间中的静态随机网络.

虽然网络生成与演化已有众多研究成果,但现有的模型仍然存在以下不足:1)现有的生成模型在解释某些网络的宏观特性上具有很好的效果,但在中观结构和微观特性上解释能力较弱,对于诸如引文网络等真实网络中度分布存在明显偏离幂律分布这一宏观现象^[32],BA模型和PS模型及其扩展模型也都无法仿真重现.虽然Li等^[20]提出的模型能拟合这种偏离现象,但对偏离现象产生的原因解释不足;2)基于社会选择理论,用户间的相似性或同质性是用户在社会交往中择优连接的重要影响因素,它是用户内在特征的显式体现,但无论是一维的颜色还是二维空间中的角距离都无法完全体现多维特征下用户间的相似度.因此,如何根据网络连接的微观机制构建网络生成模型,使其能够拟合真实网络中的特征,并揭示网络生成机制与网络特征之间的关系成为本文研究的核心问题.

鉴于此,本文随机抽取了微博用户的标签数据,通过实证的方法分析了社交网络用户的特征分布,并基于用户特征分布仿真构建用户的多维特征向量;随后考虑社会选择作用的产生机理,根据用户特征相似度描述用户间的社会选择作用,并综合考虑社会影响和社会选择两种机制,构建了网络生成的SSBA(social selection-aware BA)模型.该模型可通过参数调整网络生成过程中社会选择与社会影响作用的强弱,从而构建不同类型的社交网络.实验研究表明,与BA和PS等社交网络生成模型相比,SSBA模型能很好地刻画不同机制下生成的社交网络的统计特性,并从网络生成的微观机制上解释了网络度分布偏离幂律分布以及网

络同配性形成的原因. 本文结论对有效辨别社交网络类型, 进而制定针对性的管理策略具有指导意义.

2 用户特征分布的构建

作为客观存在的个体, 社交网络中的用户往往具有某些特定的特征, 这些特征将不同的用户区分开来, 但是用户特征所服从的分布是未知的. 在真实的在线社交网络中, 用户的特征通常能在用户的个人信息中体现. 以微博为例, 每个微博用户可以给自己添加标签, 使得具有相同标签的用户能更快地找到对方. 考虑到标签的设置是非强制性的, 本文认为微博用户一旦为自己添加了标签, 那么这个标签就是用户真实特征的反映, 因此可以将用户的标签看成用户的特征, 用户的标签数则对应该用户具有的特征个数. 本文随机抽取了 15 万微博用户的个人信息, 删除没有设置标签的用户后, 得到 82 578 条用户标签记录. 统计每个用户的标签数量及所有标签的分布情况, 结果如图 1 所示.

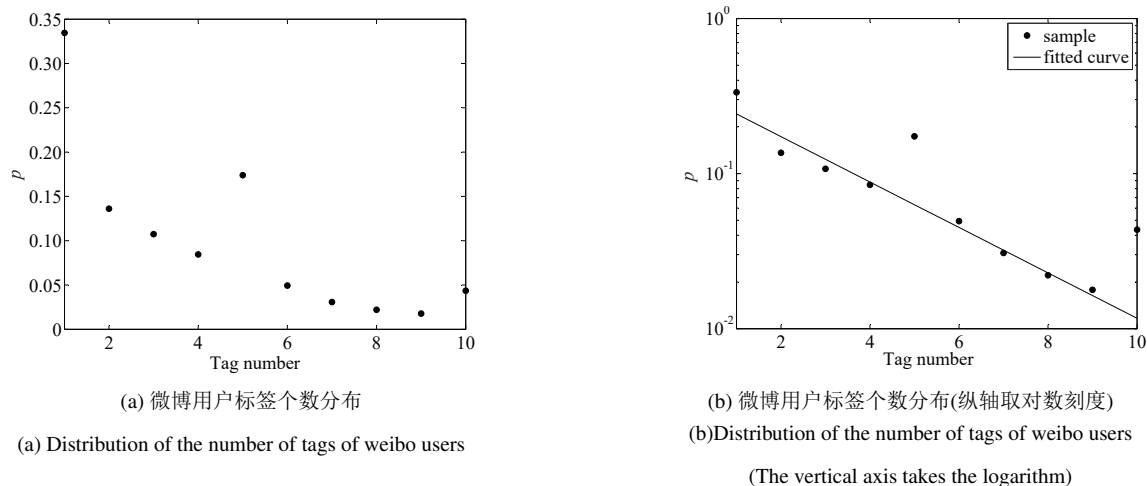


图 1 微博用户标签个数分布

Fig. 1 Distribution of the number of tags of weibo users

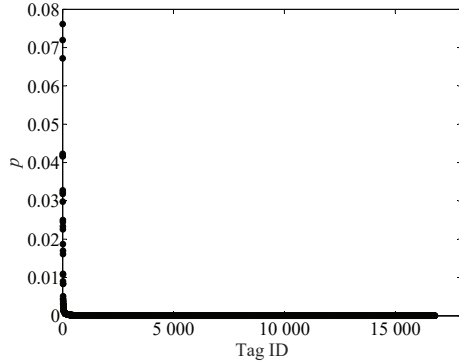
由图 1(a)可知, 微博用户具有 i 个标签(特征)的概率随着 i 的增大是近似递减的, 但在 $i = 5$ 及 $i = 10$ 两处概率值异常高. 这可能是微博对标签的设置规则导致的. 当新用户注册微博账号时, 用户可以选择至多 5 个标签, 系统将根据设置的标签为新用户推荐相应的用户, 因此标签数为 5 的用户特别多. 在用户使用微博的过程中, 用户可以添加或删除标签, 但标签的上限是 10 个, 由于截断的存在, 标签数为 10 的用户也较标签数为 9 的用户多. 为进一步探讨标签个数服从的分布, 本文将图 1(a)中 y 轴转换成对数刻度, x 轴保持线性刻度(图1(b)), 发现忽略异常点 $i = 5$ 及 $i = 10$, 得到的散点图是近似线性的(黑色实线为线性拟合函数), 拟合优度检验 $R^2 = 0.9712$. 对指数分布 $p(x) = \lambda e^{-\lambda x}$ 两边取对数得

$$\ln(p(x)) = \ln(\lambda) - \lambda x. \quad (1)$$

指数分布在 y 轴转换成对数刻度, x 轴保持线性刻度下是线性的. 从理论上说, 与人的认知能力将允许人类拥有约 150 人的稳定人际关系一样^[33], 由于人的认知能力有限, 每个人所具有的特征数不可能无限多, 且特征的数量越多, 个体所需的能力越强, 因此个体具有 i 个特征的概率是关于 i 的减函数, 且递减的趋势应强于线性关系, 同时与幂律分布相比又应具有更稳定的均值与方差. 由此, 本文假设用户特征个数的分布近似服从指数分布.

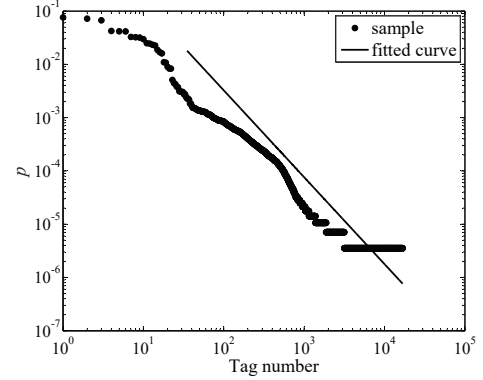
另一方面, 由于每个用户的兴趣不同, 特征的总量往往非常大, 但是有些特征比较大众, 如喜欢音乐、跑步; 有些特征则相对小众, 如研究深度学习、社交网络分析. 这种极度不均匀性会使得特征的分布呈现出明显的长尾效应. 而幂律分布在社会、经济学领域被广泛用于拟合这种长尾效应^[34], 因此在理论上用幂律分布拟合用户的特征分布是合理的. 图 2(a) 给出了标签(特征)的分布, 多数标签很少被选择, 但有少数标签被选

中的概率很大. 将用户标签分布取双对数坐标(如图 2(b)), 得到的散点图近似线性, 使用 MLE 方法^[35]估计幂指数, 黑色实线即为拟合曲线. 因此, 本文假设特征分布服从幂律分布.



(a) 微博用户标签分布

(a) Tag distribution of weibo users



(b) 双对数坐标下微博用户标签分布

(b) Tag distribution of weibo users in double logarithmic coordinates

图2 微博用户标签分布

Fig. 2 Tag distribution of weibo users

综上, 用户特征分布构建过程如下:

步骤1 假设网络中共有 X 种特征, 每个用户的特征个数不大于 Y 个, 初始化任意用户 n_i 的特征分布 $\mathbf{I}_i$ 为 1 行 X 列的向量;

步骤2 对每个用户 n_i , 从参数为 λ 的指数分布 $f(y) = \lambda e^{-\lambda y}$ 中随机抽取一个样本 y , 若 $y < Y$, 设该用户特征数为 y , 否则该用户的特征数量为 Y ;

步骤3 从参数为 γ 的幂律分布 $g(x) = Cx^{-\gamma}$ 中随机抽取 y 个不相等的样本 $x_1, x_2, \dots, x_y$ 作为用户具有的特征的编号, 任意 $x_i \in [1, X], i = 1, 2, \dots, y$;

步骤4 对每个特征 x_i , 在 $[0, 1]$ 上取随机数 z_i 作为用户对特征 x_i 的偏好程度, 并将该值代入用户特征分布的第 x_i 列中;

步骤5 将特征分布归一化, 使得用户对所有特征的偏好程度之和为 1, 从而得到用户 n_i 的特征分布为 $\mathbf{I}_i = (a_{i1}, a_{i2}, \dots, a_{iX})$, 其中 $\sum_{j=1}^X a_{ij} = 1$.

3 考虑社会选择作用的社交网络生成模型(SSBA 模型)

在构造完每个用户的特征分布后, 用户间的同质性可以用用户间的特征相似度表示. 结合网络生成机制相关实证分析的结果, 本文在择优连接中考虑社会影响和社会选择的共同作用, 假设新节点连接到旧节点的概率是旧节点影响力(社会影响作用)与新旧节点相似度(社会选择作用)的线性加权, 提出考虑社会选择作用的社交网络生成模型 SSBA. 模型的构造方法如下.

假设每个节点在进入网络时都通过第 2 节的方法构建了一个对应的特征分布, 初始时刻设定网络中有 m_0 个节点, 任意节点 n_i 的特征分布为 $\mathbf{I}_i$, 网络中的 m_0 个节点全连接. 在每个时间步内, 引入一个新节点. 对任意新节点 n_t , 该节点的特征分布为 $\mathbf{I}_t$. 新节点有 m 条边($m \leq m_0$)与现有节点相连. 当选择现有节点与新节点连接时, 新节点连接到任意节点 n_i 的概率 π 满足如下等式, 即

$$\pi_{ti} = \alpha \bar{S}_{ti} + (1 - \alpha) \bar{F}_i, \quad (2)$$

其中 α 为节点相似度(社会选择作用)在连接过程中的权重, $\bar{F}_i$ 为节点 n_i 标准化的影响力, $\bar{S}_{ti}$ 为 n_t 与 n_i 间标准化的特征相似度.

用节点 n_i 的度数 d_i 代表节点的影响力, 则 $\bar{F}_i = d_i \left(\sum_{j=1}^{t-1} d_j \right)^{-1}$. 本文使用余弦相似度计算新节点 n_t 与任意旧节点 n_i 的特征相似度 S_{ti} . 假设节点 n_t 的特征分布 $\mathbf{I}_t = (a_{t1}, a_{t2}, \dots, a_{tX})$, 节点 n_i 的特征分布 $\mathbf{I}_i = (a_{i1}, a_{i2}, \dots, a_{iX})$, 那么

$$S_{ti} = \cos(n_t, n_i) = \frac{\mathbf{I}_t^T \mathbf{I}_i}{\|\mathbf{I}_t\| \|\mathbf{I}_i\|} = \sum_{j=1}^X a_{tj} a_{ij} \left(\sum_{j=1}^X a_{tj}^2 \right)^{-1/2} \left(\sum_{j=1}^X a_{ij}^2 \right)^{-1/2}, \quad (3)$$

则节点 n_t 与所有现有节点的标准化特征相似度 $\mathbf{S}_t = (\bar{S}_{t1}, \bar{S}_{t2}, \dots, \bar{S}_{t,t-1})$, 其中

$$\bar{S}_{ti} = S_{ti} \left(\sum_{j=1}^{t-1} S_{tj} \right)^{-1}. \quad (4)$$

4 实验

本节控制其他参数不变, 通过改变社会选择作用在连接过程中所占权重 α , 发现社会选择和社会影响作用的不同比例对网络结构特征的影响, 同时与 popularity $\times$ similarity(PS)模型^[29] 及真实社交网络数据进行对比, 验证 SSBA 模型能否通过仿真网络的微观生成机制, 更好地解释网络的宏观特性及中观结构.

实验过程中涉及到的参数及其设置如下: X 为网络中的特征总数, 假设 $X = 100$; Y 为用户特征数的上限, 假设 $Y = 10$; λ 为用户特征个数所服从的指数分布的参数, 结合微博标签个数的分布情况, 假设 $\lambda = 1/3$, 此时每个用户的特征数的均值为 $1/\lambda = 3$; γ 为特征总体所服从的幂律分布的参数, 考虑到本文选取的特征总数为 100, 为了控制第 100 号特征出现的次数(即控制抽取的样本大于 100 从而发生截断的情况), 同时防止样本总是集中在前面若干个特征上, 本文通过多次实验最终选取特征分布参数 $\gamma = 2.5$; m_0 为初始时刻网络中节点数, 假设 $m_0 = 6$; m 为新节点连接的边数, 假设 $m = 5$, 此时网络的平均度为 $2m = 10$; N 为网络中节点总数, 假设 $N = 10\,000$.

4.1 宏观特性

度分布、聚类系数及度相关性是描述社交网络宏观特性的三大特征量. 许多真实社交网络的度分布近似服从幂律分布, 但在另一些网络中, 度分布则偏离了幂律分布^[32], 现有研究尚无法解释这种偏离现象的产生机理. SSBA 模型尝试对其进行解释. 在 SSBA 模型下改变社会选择作用在连接过程中所占的权重 α , 得到 SSBA 模型下网络的度分布如图 3 所示.

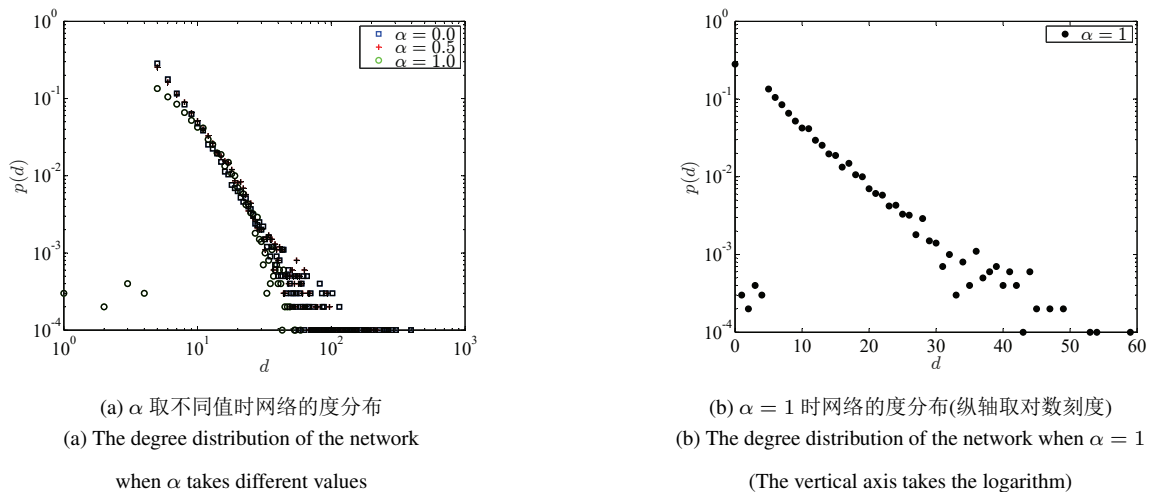


图3 SSBA 模型下网络的度分布

Fig. 3 Degree distribution for SSBA model

当 α 分别取 0, 0.5, 1 时, 节点度分布的变化趋势总体上一致(图 3(a)). 随着 α 的逐渐增大, 社会选择作用不断增强, 度分布在高度数部分逐渐表现出更陡的下降趋势. 当 $\alpha = 1$ 时, 网络的度分布已偏离幂律分布, 更接近于指数分布(图 3(b))(小度数位置的离群点是部分节点特征过于小众导致的). 综上, SSBA 模型表明网络度分布对幂律分布的偏离是由社会选择作用的不断增强导致的.

为验证上述实验结论, 本文选取斯坦福大学大规模网络数据集中的两种典型社会网络 YouTube 及 DBLP 论文合作网络进行研究. YouTube 是一个流行的视频分享网站, 注册用户可以在上传原创视频、关注感兴趣的用户并观看这些用户发布的视频. 该数据集有约 113 万节点, 299 万条边. DBLP 是计算机科学中论文作者的关系网络, 若两个作者共同合作了至少一篇论文, 这两个作者间将建立连边. 该数据集有约 32 万节点, 105 万条边. 这两类网络的度分布如图 4 所示.

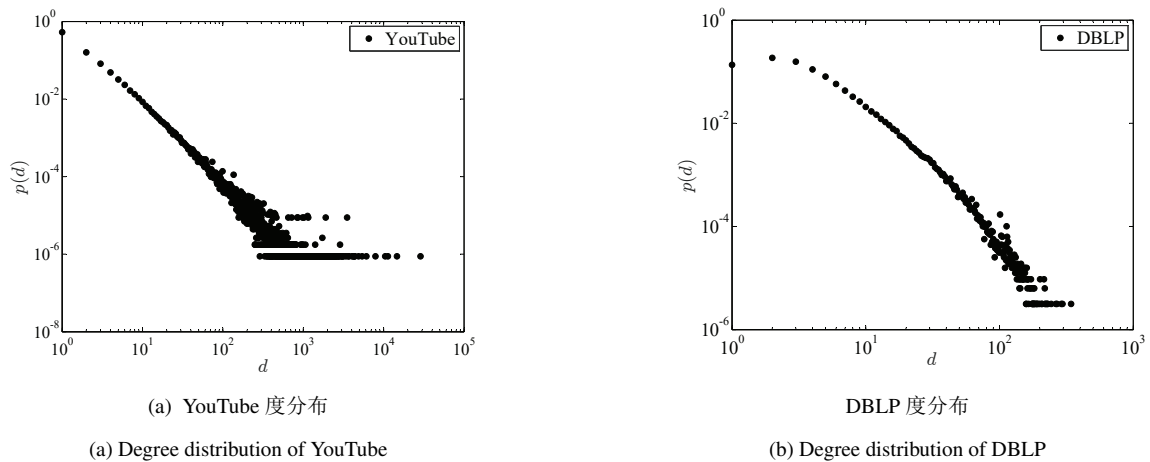


图4 两种大规模社会网络度分布

Fig. 4 Degree distribution of two large-scale social networks

YouTube 的度分布服从幂律分布; DBLP 虽然也基本服从幂律分布, 但在高度数部分表现出更陡的下降趋势. 实证研究表明, YouTube 上社会影响的作用更强^[15], 而 DBLP 主要受社会选择机制的影响^[16]. 这一度分布的趋势有力地支撑了 SSBA 模型关于社会选择作用将使网络逐渐偏离幂律分布的论述. PS 模型同样考虑了社会影响与社会选择的共同作用, 并使用参数 β 调整社会影响作用的强弱, 该模型的度分布服从参数为 $1 + 1/\beta$ 的幂律分布(如图 5), 不能解释度分布对幂律分布的偏离.

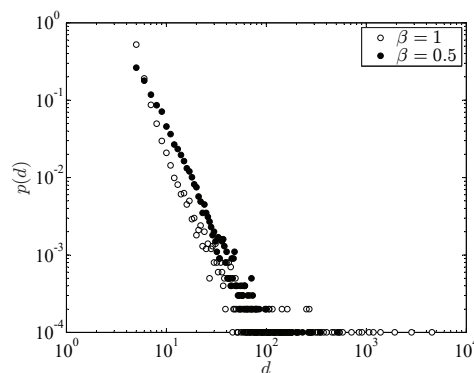


图5 PS 模型下网络的度分布

Fig. 5 Degree distribution of network under PS model

聚类系数体现了网络中节点的聚集程度, 图 6 给出了 SSBA 模型及 PS 模型下网络的 k 度节点平均聚类系数分布. SSBA 模型中, 随着社会选择作用的不断增强, 高度数节点的聚类系数呈逐渐增加的趋势. 这种尾部聚类系数的增长在 Flickr 中也存在^[36]. 另一方面, Flickr 上具有更强的社会选择作用^[15]. 因此, SSBA 模

型能有效地描述强社会选择作用下网络的聚类系数. 但在 PS 模型中, 聚类系数随着节点度数的增加而单调递减, 且该趋势与参数的取值独立, 因此 PS 模型无法对社会选择作用占优的网络的聚类系数分布进行有效解释.

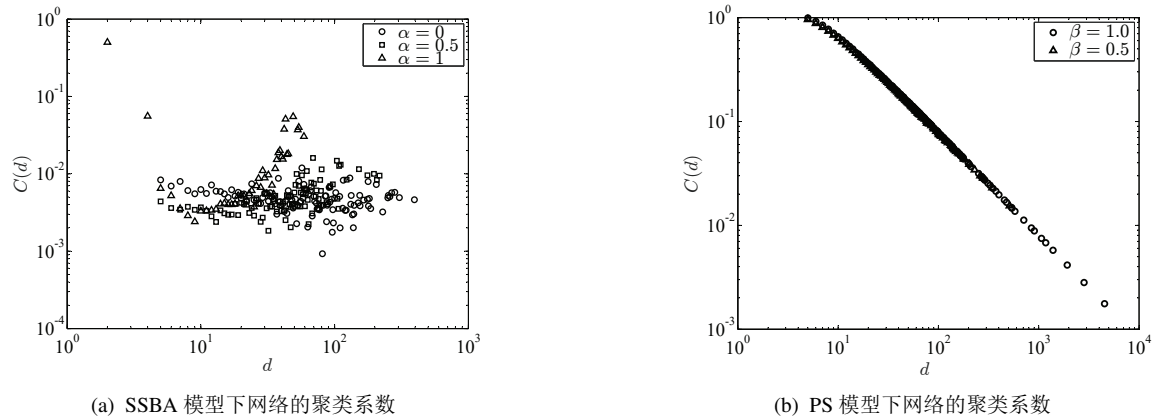


图 6 SSBA 模型及 PS 模型下网络的聚类系数

Fig. 6 Clustering coefficient of networks under SSBA and PS models

度相关性(k_{nn})体现了节点的连接倾向, 表现为度为 k 的节点的邻居的平均度, 两种模型下的度相关性如图 7 所示.

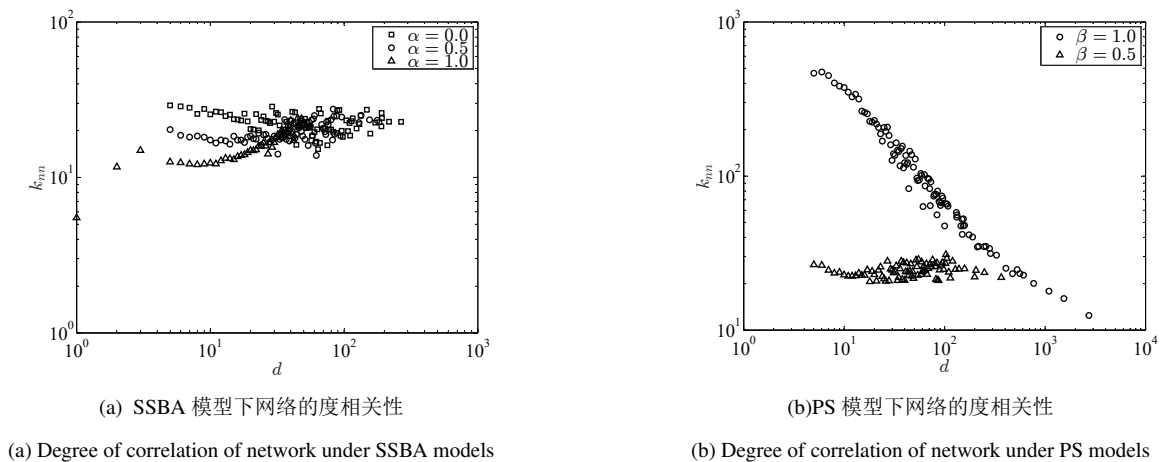


图 7 SSBA 模型及 PS 模型下网络的度相关性

Fig. 7 Degree of correlation of network under SSBA and PS models

SSBA 模型中, 随着社会选择作用的增强, 度越大的节点越倾向于连接度数大的节点, 网络的同配性逐渐增强. 在 PS 模型中, 随着参数的增大, 网络的异配性逐渐增强. 实证研究表明, 多数在线社交网络属于同配网络^[36], 即高影响力人群之间更容易选择相互连接. 从这点看, SSBA 模型更接近真实的社交网络情况.

4.2 中观结构一致性

社区结构是社交网络的典型特征^[37]. 网络中每个小社区可以看成是相对独立的小系统, 这些小系统相互交叠最终形成了不同的网络. 因此社区这种介于微观个体与宏观整体之间的中观结构所具有的特征将影响整个网络的结构与特性. 若某个社区的形成过程中社会影响起主导作用, 称这个社区是影响型的; 若社会选择起主导作用, 则称该社区是选择型的. 当社区偏向于选择型时, 社区内节点的度分布会逐渐偏离幂律分布而向指数分布转变. 虽然分布的形式发生了改变, 但从双对数坐标下的度分布图来看, 社会选择

作用占比的增大会导致尾部下降趋势更陡,因此使用 MLE 方法^[35]估计的幂指数将不断增大(SSBA 模型在 $\alpha = 0, 0.5, 1$ 时估计的幂指数分别为 2.78, 3.02, 4.23; YouTube, DBLP 的幂指数分别为 2.14, 3.26)。

基于上述认知,为探索不同真实网络的中观结构,并验证仿真网络在中观结构上与真实网络是否一致,本文以幂指数为指标判断社区的类型,计算 YouTube、DBLP、SSBA 和 PS 四种网络中每个社区内度分布的幂指数。在社区划分上,YouTube 和 DBLP 中社区的划分是已知的,仿真网络中则使用 Newman 等^[38]提出的重叠社区发现算法将网络划分为 100 个社区,并将每个节点划分到最具可能性的社区中。不同网络下幂指数的分布如图 8 所示。

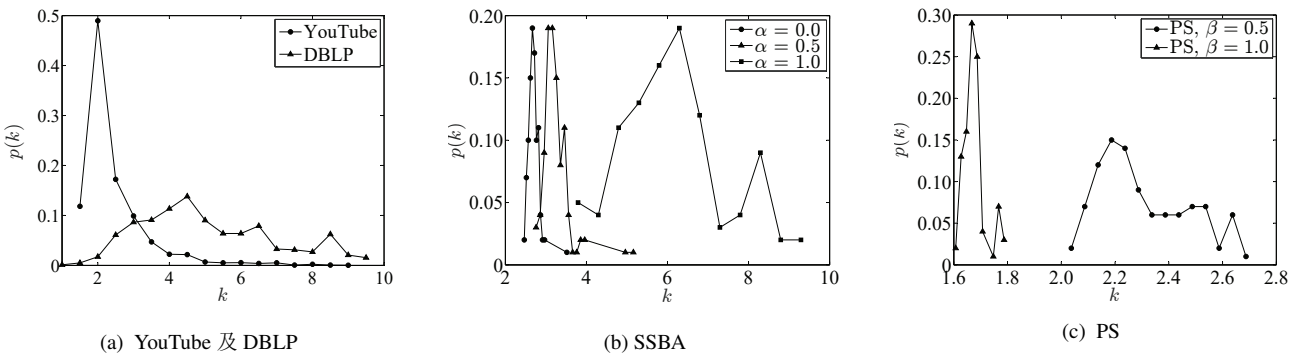


图 8 不同网络以社区为单位计算度分布后幂指数的分布

Fig. 8 Community-level power exponent distribution for different networks

从图 8(a)可知, YouTube 中社区的幂指数多数集中在图的左侧,在右侧形成长尾,表明 YouTube 中社区以影响型为主。DBLP 中幂指数的峰值右移,表明网络中社区以选择型为主。上述结果与前述实证分析结果一致,同时也证明了不同类型的社交网络在中观结构上存在差异。SSBA 模型下网络的幂指数分布如图 8(b)所示,随着社会选择作用占比逐渐增大,幂指数峰值逐渐右移,即网络中选择型社区的比重不断增加。上述仿真结果与真实网络中的实验结果一致。PS 模型下网络的幂指数分布如图 8(c)所示,该模型下幂指数的分布很集中,不能体现中观结构的差异性。因此本文认为 SSBA 模型除了能够拟合真实网络中的度分布等宏观特征外,在以社区为单位的中观结构上也具有很好的拟合效果。这是传统的 BA 模型以及 PS 模型无法实现的。

从管理与营销的角度,社会选择作用下构建的选择型社区与社会影响作用下构建的影响型社区在网络结构、信息传播、交互方式上存在很大的不同,因此不同类型的网络下应该选用不同的营销策略,而本文的结论提供了一个简单的方法识别网络及社区的类型,能为进一步的营销策略研究提供理论支持。例如,当企业在一个社交网络上进行产品推广时,如果是社会影响机制起主导作用时,可通过寻找高影响力的种子节点进行传播;反之,如果是社会选择机制起主导作用时则可以采用广而告之的模式。

SSBA 模型从用户同质性产生的内在机理出发,仿真用户的多维特征分布,从而获得用户间的相似度。这种完全内生性的相似度获取机制使得用户因为社会选择作用而连接的现象更符合真实网络中的因为共同特征而建立关系的情况,因此 SSBA 在网络的宏观特性及中观结构上都能很好地拟合主要受社会选择机制影响的真实社交网络。

4.3 敏感性分析

为验证 SSBA 模型下网络的主要特征对网络规模 N 、新节点连接的边数 m 以及特征分布的参数 λ 和 γ 的敏感性,本文以度分布为例进行敏感性分析。

在网络规模方面,本文在网络增长的过程中截取了节点数量分别为 1 000, 5 000, 10 000 的三个截面数据,分别计算 $\alpha = 0, 0.5, 1$ 时的度分布,结果如图 9。

无论 α 的取值如何,网络的度分布在不同的网络规模下具有一致性,因此节点个数的增长不会改

变 SSBA 模型生成的网络结构.

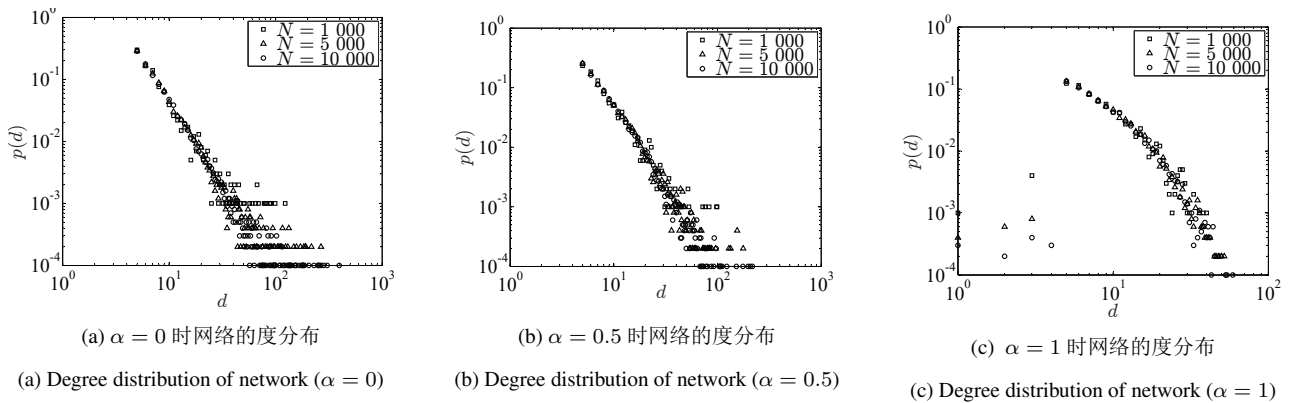


图9 节点数分别为 1 000, 5 000, 10 000 时网络的度分布

Fig. 9 Degree distribution when $N = 1\,000, 5\,000, 10\,000$

在 α 为 0, 0.5, 1 时分别计算 $m = 3, m = 5$ 时网络的度分布, 结果如图 10 所示. 虽然边数 m 发生变化, 但在同样的 α 下网络度分布的趋势是一致的, 因此度分布对 m 不敏感.

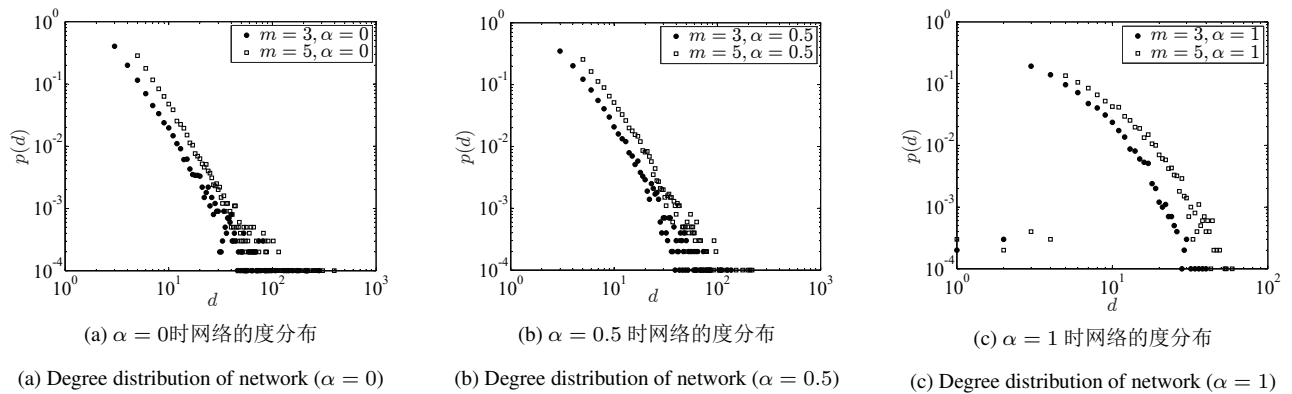


图 10 m 对网络度分布的影响

Fig. 10 The effect of m on degree distribution of network

在构造用户特征分布时, SSBA 模型假设了用户特征个数服从参数为 λ 的指数分布, 特征总体的分布服从参数为 γ 的幂律分布. 为验证 SSBA 模型对上述参数的敏感性, 以 $\lambda = 1/3, \gamma = 2$ 为基础, 设计对比实验 $\lambda = 1/3, \gamma = 3$ 以及 $\lambda = 1/4, \gamma = 2$. 每个实验组固定网络规模 $N = 5\,000$, 在 α 分别取 0, 0.3, 0.5, 0.7, 1 时得到 15 组实验结果, 用每组实验度分布的幂指数描述网络的度分布情况, 结果如图 11 所示.

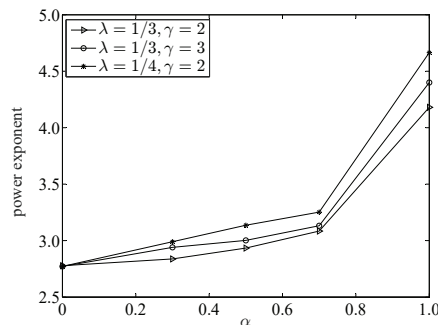


图 11 SSBA 对 λ, γ 的敏感性分析

Fig. 11 Sensitivity analysis of SSBA model to λ and γ

不同参数设置下网络度分布的幂指数的变化趋势是一致的。 λ 越小, 指数分布的均值越大, 平均每个个体所具有的特征数越多; γ 越大, 幂律分布越陡, 兴趣的分布越集中, 这两种变化都会导致用户间同质的可能性增大, 在度分布上体现为幂指数的增长。

5 结束语

本文结合现有社交网络形成的微观机制研究成果, 针对现有相似性度量方法的不足, 根据每个用户特征个数服从指数分布, 特征总体分布服从幂律分布, 内生性地构造了用户的特征分布, 并据此计算用户间的相似度. 在此基础上提出了考虑社会选择作用的社交网络生成模型——SSBA模型. 该模型假设边连接的过程中, 新节点与旧节点连接的概率是新旧节点特征相似度与旧节点影响力的加权和. 仿真实验与实证分析的结果表明, 本文所提的 SSBA 模型不仅能很好地解释以社会选择作用为主的社交网络中诸如度分布、集聚性、度相关性等宏观特性, 在以社区为单位的中观结构上也具有很好的解释效果. 本文结论有助于进一步理解不同生成与演化机制下生成的网络的特征, 对研究网络中的信息传播机制, 制定有效的社会化营销策略等具有良好的理论指导意义. 深入研究社会选择与社会影响在网络生成过程中的作用机理, 提出有效的度量指标区分影响型(社会影响占主导作用)网络 and 同质型(社会选择占主导作用)网络是本文后续研究的主要方向.

参考文献:

- [1] Aral S, Walker D. Identifying influential and susceptible members of social networks. *Science*, 2012, 337(6092): 337–341.
- [2] Backstrom L, Huttenlocher D, Kleinberg J, et al. Group formation in large social networks: Membership, growth, and evolution // *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York: ACM, 2006: 44–54.
- [3] Leskovec J, Backstrom L, Kumar R, et al. Microscopic evolution of social networks // *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York: ACM, 2008: 462–470.
- [4] Anagnostopoulos A, Kumar R, Mahdian M. Influence and correlation in social networks // *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York: ACM, 2008: 7–15.
- [5] Ma L, Krishnan R, Montgomery A L. Latent homophily or social influence. An empirical analysis of purchase within a social network. *Management Science*, 2014: 61(2), 454–473.
- [6] Crandall D, Cosley D, Huttenlocher D, et al. Feedback effects between similarity and social influence in online communities // *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data mining*. New York: ACM, 2008: 160–168.
- [7] Aral S, Muchnik L, Sundararajan A. Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proceedings of the National Academy of Sciences*, 2009, 106(51): 21544–21549.
- [8] Denrell J. Sociology: Indirect social influence. *Science*, 2008, 321(5885): 47–48.
- [9] Bond R M, Fariss C J, Jones J J, et al. A 61-million-person experiment in social influence and political mobilization. *Nature*, 2012, 489(7415): 295–298.
- [10] Aral S. Social science: Poked to vote. *Nature*, 2012, 489(7415): 212–214.
- [11] Muchnik L, Aral S, Taylor S J. Social influence bias: A randomized experiment. *Science*, 2013, 341(6146): 647–651.
- [12] Lewis K, Gonzalez M, Kaufman J. Social selection and peer influence in an online social network. *Proceedings of the National Academy of Sciences*, 2012, 109(1): 68–72.
- [13] Godinhodematos M, Ferreira P, Krackhardt D. Peer influence in the diffusion of iPhone 3G over a large social network. *Management Information Systems Quarterly*, 2014, 38(4): 1103–1133.
- [14] Muchnik L, Pei S, Parra L C, et al. Corrigendum: Origins of power-law degree distribution in the heterogeneity of human activity in social networks. *Scientific Reports*, 2015, 5: 15932.
- [15] Zhang Q M, Xu X K, Zhu Y X, et al. Measuring multiple evolution mechanisms of complex networks. *Scientific Reports*, 2015, 5: 10350.

- [16] Wang Z Z, Zhu J J. Homophily versus preferential attachment: Evolutionary mechanisms of scientific collaboration networks. *International Journal of Modern Physics C*, 2014, 25(05): 1440014.
- [17] Barabási A L, Albert R. Emergence of scaling in random networks. *Science*, 1999, 286(5439): 509–512.
- [18] Albert R, Barabási A L. Topology of evolving networks: Local events and universality. *Physical Review Letters*, 2000, 85(24): 5234.
- [19] Holme P, Kim B J, Yoon C N, et al. Attack vulnerability of complex networks. *Physical Review E*, 2002, 65(5): 056109.
- [20] Li X, Chen G. A local-world evolving network model. *Physica A: Statistical Mechanics and Its Applications*, 2003, 328(1): 274–286.
- [21] Liu J G, Dang Y Z, Wang Z T. Multistage random growing small-world networks with power-law degree distribution. *Chinese Physics Letters*, 2006, 23(3): 746–749.
- [22] 胡海波, 刘 璇. 在线社会网络增长中的优先连接. *系统工程学报*, 2014, 29(3): 289–298.
Hu H B, Liu X. Preferential linking in the growth of online social networks. *Journal of Systems Engineering*, 2014, 29(3): 289–298. (in Chinese)
- [23] Boguná M, Pastor-Satorras R, Díaz-Guilera A, et al. Models of social networks based on social distance attachment. *Physical review E*, 2004, 70(5): 056122.
- [24] Ling H W, Pattison P, Robins G. A spatial model for social networks. *Physic A: Statistical Mechanics & Its Applications*, 2006, 360(1): 99–120.
- [25] Marsili M, Vegaredondo F, Slanina F. The rise and fall of a networked society: A formal model. *Proceedings of the National Academy of Sciences of the United States of America*, 2004, 101(6): 1439–1442.
- [26] Kumpula J M, Onnela J P, Saramaki J, et al. Emergence of communities in weighted networks. *Physical Review Letters*, 2007, 99(22): 228701.
- [27] 李倩倩, 顾基发. 用户行为驱动的在线社交网络建模. *系统工程学报*, 2015, 30(1): 9–15.
Li Q Q, Gu J F. Activity driven modelling of online social network. *Journal of Systems Engineering*, 2015, 30(1): 9–15. (in Chinese)
- [28] Li A, Li J, Pan Y, et al. Homophily/kinship model: Naturally evolving networks. *Scientific Reports*, 2015, 5: 15140.
- [29] Papadopoulos F, Kitsak M, Serrano M A, et al. Popularity versus similarity in growing networks. *Nature*, 2012, 489(7417): 537–540.
- [30] Zuev K, Boguñá M, Bianconi G, et al. Emergence of soft communities from geometric preferential attachment. *Scientific Reports*, 2015, 5: 09421.
- [31] Ferretti L, Cortelezzi M, Mamino M. Duality between preferential attachment and static networks on hyperbolic spaces. *Europhysics Letters*, 2014, 105(3): 38001.
- [32] Redner S. How popular is your paper: An empirical study of the citation distribution. *The European Physical Journal B-Condensed Matter and Complex Systems*, 1998, 4(2): 131–134.
- [33] Dunbar R I M. Neocortex size as a constraint on group size in primates. *Journal of Human Evolution*, 1992, 22(6): 469–493.
- [34] Newman M E J. Power laws, pareto distributions and Zipf's law. *Contemporary Physics*, 2005, 46(5): 323–351.
- [35] Clauset A, Shalizi C R, Newman M E. Power-law distributions in empirical data. *SIAM Review*, 2009, 51(4): 661–703.
- [36] Mislove A, Marcon M, Gummadi K P, et al. Measurement and analysis of online social networks // *Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement*. New York: ACM, 2007: 29–42.
- [37] Girvan M, Newman M E. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, 2002, 99(12): 7821–7826.
- [38] Newman M E, Leicht E A. Mixture models and exploratory analysis in networks. *Proceedings of the National Academy of Sciences*, 2007, 104(23): 9564–9569.

作者简介:

刘业政(1965—), 男, 安徽和县人, 教授, 博士生导师, 研究方向: 电子商务与商务智能, 决策理论与方法, Email: liuyezheng@hfut.edu.cn;

李玲菲(1992—), 女, 江西兴国人, 博士生, 研究方向: 在线社交网络分析, Email: lilinglei@126.com;

孙春华(1977—), 女, 安徽合肥人, 副教授, 研究方向: 消费者行为, 互联网营销, Email: ahecrocky@163.com.