

虚拟社区中人际交互行为的统计分析研究*

司夏萌[†] 刘云

(北京交通大学电子信息工程学院,北京交通大学通信与信息系统北京市重点实验室,北京 100044)

(2010年8月23日收到;2010年9月29日收到修改稿)

本文对互联网论坛中用户的交互行为进行实证分析,建立基于隐式交互行为的虚拟社区网络,通过考察其统计特性,发现该虚拟社区网络是有向、不对称、异配的无标度网络,且具有分层性质和社团结构. 用户的浏览量与其入度正相关,二者均可描述用户的影响力和受关注程度. 本文从定量的角度发现,论坛中具有“贡献驱动”特性的权威用户在信息传播中的作用不容小觑;网络论坛的成功需要提高大多数惰性用户的活跃度;用户要想在虚拟社区网络中占据权威地位,需要发表高质量的帖子或形成个人品牌.

关键词: 统计物理, 复杂网络, 人际交互, 幂率分布

PACS: 89.75.-k, 29.85.-c, 89.75.Da

1. 引言

微观个体行为和宏观动力演化一直以来都是物理学研究的重要方向之一^[1,2]. 其中,人际交互行为与社会群体现象的涌现越来越受到关注^[3-5]. 最初,物理学家试图通过数学建模和计算机仿真来解释社会现象^[6],但多数关注模型变量多于实际现象,造成了模型与实际的脱离^[7]. 随着复杂网络理论研究的兴起^[8-10],根据现实交互关系构建复杂网络、分析复杂网络特性及其上负荷的相互影响成为探究人类行为和群体动力的新的研究手段^[11-13]. 然而,由于获取大批量数据的困难,数据依靠人工收集和处理,早期的研究只能局限在分析小范围人群之间的交互行为上^[14].

互联网和通信技术的发展使得网络在日常生活中扮演着越来越重要的角色,成为人们获取信息、交流沟通的重要渠道. 结合复杂网络理论,数据库技术和大规模数据的高速处理能力,为研究网络上人类交互行为和群体动力提供了条件,关于虚拟社区及其用户行为分析的研究也日渐增多. 虚拟社区是人们在网络上从事公众讨论,经过一段时间的互动后形成的一种人际关系网络,是一种新型的社

会组织. 此类研究大多针对社会网络网站 (social networking sites) 和博客 (blog) 两种虚拟社区形式展开. 前者是模拟现实社会的网上社交平台,网络基于好友关系,信息只能在好友间流通;后者是个人发表、传播自己思想的出版方式,虽可以不受好友关系限制地传播信息,但仍带有很大程度的私人性.

文献[15]根据国内某论坛的实际网络数据统计与分析,提出了虚拟社区网络的演化机理和构造算法,来得到与真实网络具有相似演化过程的模型. 文献[16]分别分析了新浪博客和校内网中用户好友关系网络的统计特征,验证了小世界和无标度特性的普遍性,认为博客网络是有向、异配网络,在线社会网络 (online social network) 是无向、同配网络. 文献[17]研究了 Wealink 的复杂网络特性随着时间的变化,发现网络的增长规模成 S 形,网络密度、簇系数、异质性和同配性与时间是非单调关系. 文献[18]认为在线社会网络节点度服从幂率分布,而文献[19,20]则发现在线社会网络中节点度分布由幂率和指数两种分布共同组成.

不难看出,目前关于虚拟社区及其用户行为分析的研究尚没有统一结论,但并不意味着针对性地分析某个虚拟社区网络没有意义. 因为,对不同的虚拟社区网络 and 用户行为的了解是总结虚拟社区

* 国家自然科学基金 (批准号:60972012),北京市自然科学基金 (批准号:4102047),中央高校基本科研业务费专项资金 (批准号:2009YJS007),教育部哲学人文社会科学重大课题 (批准号:08WL1101),科技人员服务企业项目 (批准号:2009GJA00048),北京市教育委员会学科建设与研究生建设项目和2010丰台科技型中小企业创新基金 (批准号:CXJJ1007) 资助的课题.

[†] E-mail: sixiameng@gmail.com

网络一般特性和网路用户行为普遍规律的前提基础. 相比社会网络网站和博客, 互联网论坛或电子布告栏(bulletin board system)在信息传播与观点扩散中发挥着更加重要的作用, 并凭借其用户可以不受好友关系限制自由获取信息的优势, 已经成为国内舆论突发事件的主要集中地. 对网络论坛中用户的行为特性分析, 不仅有助于我们了解网络论坛用户的行为特点、推进信息扩散, 而且对于改进作为信息传播重要场所的网络论坛平台、提升 Web 服务质量具有重要意义. 然而, 该方面的研究却未能引起充分关注.

在互联网论坛中, 用户间的关系是通过回复建立的, 是不需要双方认可的隐式关系. 那么, 这种隐式的回复网络具有怎样的复杂网络特性, 隐式的回复网络与外显的基于好友关系的社会网络有何不同, 网络论坛中的用户具有怎样的看帖回帖习惯? 本文以国内知名互联网论坛——天涯为例(<http://www.tianya.cn/>), 分析了该虚拟社区的用户特点、用户的活跃度和影响力分布. 研究发现隐式的网络论坛回复网络是不对称、异质、有向网络. 除了帖子的传播程度, 网络论坛中帖子的受关注度、用户的影响力和活跃度均服从幂率分布. 不同于基于好友关系的社交网络, 网络论坛基于用户回复关系的虚拟社区网络与 Web 网一样具有异配性, 且具有分层性质和社团结构. 网络论坛中存在非常活跃的贡献者, 也存在很多无名的普通用户. 少数权威用户存在惰性但却吸引了社区中大部分的关注, 大多数用户都默默无闻但依然活跃.

2. 数据集及概念

互联网论坛是一种基于内容的在线讨论网站, 是人们发布、共享信息和交流沟通的重要场所^[21]. 其具有如下特点: 1) 表达的自由; 2) 信息的公开; 3) 一对多的传播; 4) 成员间互动的自主性. 互联网论坛中的用户可以发表和回复帖子, 并通过回复根帖与根帖的发表者建立链接. 由于阅读和回复帖子不需要对方的确认, 因此用户间的回复链接建立起来更自由, 信息传播也更方便. 作为信息传播的主要场所、网络突发事件的主要聚集地, 互联网论坛中用户的发帖和回帖都是公开透明的, 这些便于自动获取的回复关系为研究网上人际交互行为、信息的扩散创造了条件.

本文分析的数据来自于国内目前最大的网络论坛之一——天涯社区. 利用垂直网络爬虫沿网页链接采集了一个版块从 2002 年 9 月到 2009 年 6 月的信息, 获取了其 Web 页面中的主要数据, 并经过预处理去掉了页面中包含的冗余信息. 版块中的话题一般称为帖子, 帖子构成了网络论坛中的信息单元. 根据帖子的发布时间和回复关系不同可细分为根帖、回帖. 根帖用于发起一个新的话题, 引起大家的注意和关注, 并展开讨论和交互; 回帖用于回复根帖、延续该话题, 与根帖互动. 在基于回复的虚拟社区网络中, 用户 A 回复 B 意味着 A 与 B 之间有一条由 A 指向 B 的边.

在该时间区间内, 共有 412,782 个用户, 5,397 个根帖, 3,056,936 个回帖. 将数据分为根帖和回帖两种格式存储, 根帖的数据格式为: 用户 i 根帖 p_i 时间 t_i 阅读量 m 回复量 n , 表示用户 i 在时间 t_i 发表了根帖 p_i , 并被阅读过 m 次, 被回复过 n 次; 回帖的数据格式为: 用户 j 根帖 p_i 用户 i 时间 t_j 次数 l , 表示用户 j 在时间 t_j 首次回复了用户 i 的根帖 p_i , 共回复了 l 次. 将用户作为节点, 将从一个用户给另一个用户的回复作为边, 回复次数作为权重, 根据以上数据构造有向、加权的回复网络. 去掉自链(用户自己回复自己的边), 该回复网络共有 410,177 个节点, 975,305 条有向边.

为便于下文讨论, 有必要对复杂网络基本概念或术语进行介绍:

直径 表示网络中任意两个节点间距离的最大值.

平均路径长度 表示任意两个节点间距离的平均值.

入度 k_{in} 表示回复过该用户的用户数, 描述用户的影响力. 入度越大, 影响力越大.

出度 k_{out} 表示该用户回复过的用户数, 描述用户的活跃度. 出度越大, 用户越活跃.

活跃度 反映某个用户回复和响应他人多少的指标, 本文由节点出度决定.

簇系数 描述用户的邻居间的连通度, 即其邻居间互有回复的可能性, 反映节点的集聚特性. 用户 i 的簇系数 C_i 定义为它 k_i 个邻居之间的连接数 E_i 占可能的最大连接数的比例, 即

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

异配 指网络中度大的节点倾向于连接度小的节点, 反之亦然, 反映网络节点的连接倾向. 通常

可用节点度相关函数、Pearson 相关系数等指标判断,本文利用节点度相关函数判断基于回复的虚拟社区网络的异配性.

度相关函数 是节点 i 出度到该出度所有节点邻居的入度平均值的映射,即

$$K_{in}(k_{out}) = \frac{\sum_{j \in \Gamma(i)} k_{j,in}}{k_{out}},$$

其中 $\Gamma(i)$ 为节点 i 指向的所有节点集合. 它反映了网络中节点的连接倾向,是倾向与同类节点连接,还是倾向与异类节点连接. 单调递减的度相关函数表明度大的节点倾向于连接度小的节点,网络具有异配性;否则,网络具有同配性.

长尾效应 指统计量在自变量较大时,分布广泛且具有很大的差异性,通常表现为曲线前部突起,尾部细长平缓.

3. 人际交互行为统计分析

在互联网论坛中,用户通过发帖、回帖完成信息的发布与共享、观点的交互与传播. 用户个体与帖子构成了网络论坛的两种元素. 本节首先以帖子为研究对象分析帖子的属性分布;其次,以用户个体为对象统计分析研究用户个体交互的行为差异;最后,忽略回复网络中不对称的边,删除孤立的节点后得到一个双向或无向网络,研究该无向网络中的集聚特性.

帖子是网络论坛中的信息单元,包括根帖和回帖,根帖及其下面的回帖构成了一个帖子链. 根帖的浏览量 R 和被回复量 S_{in} 能够分别体现该帖子所含信息的传播程度和受关注度. 图 1 显示了二者统计区间为 100(a) 和 1(b) 的分布情况. 当帖子浏览量和被回复量每增加 100 分别统计一次的时候,可以看出,帖子的传播程度和受关注度均服从幂率分布,即

$$P(x) \sim x^{-\gamma}, \text{ wherein } \begin{cases} \gamma = 1.09, & x = R, \\ \gamma = 1.3, & x = S_{in}. \end{cases} \quad (1)$$

但是,当帖子浏览量和被回复量每增加 1 分别统计一次的时候,帖子的被回复量依然服从幂率分布,而帖子浏览量却意外地呈现弯曲线状.

由图可知,帖子浏览量在 30 左右的帖子比例最大,在此之前,帖子的比例竟然随其浏览量呈现上升趋势. 通常,浏览量小于等于 1 的帖子基本不存在,浏览量为个位数的帖子也少之又少,正常帖子

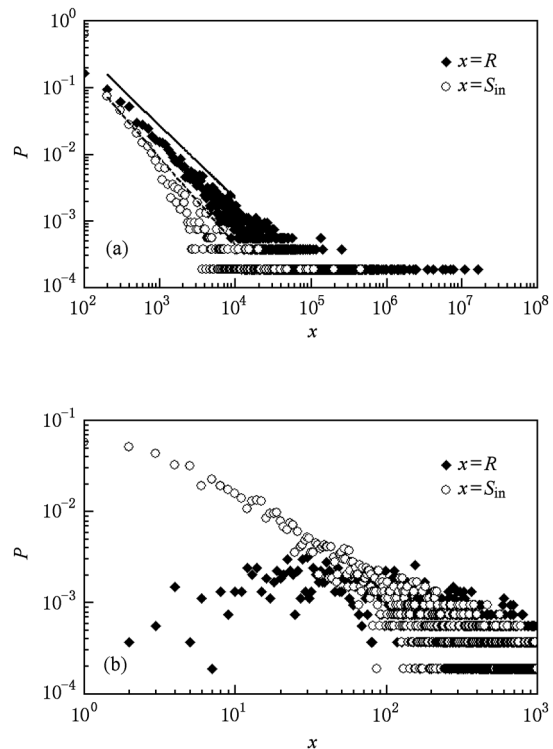


图 1 帖子的浏览量 R 和被回复量 S_{in} 的分布图 (a) 统计区间为 100; (b) 统计区间为 1

的浏览量都不会太低. 这就造成了帖子的浏览量 R 在 30 左右存在临界点,低于该值帖子比例随浏览量增长;高于该值帖子比例随浏览量减小,且服从幂率分布. 这说明网络论坛的信息传播性能优异,绝大多数帖子都有一定的传播范围,且存在少数传播极端广泛的帖子. 另一方面,帖子的被回复量服从幂率分布,少数帖子被回复量极高,但大部分帖子难以引起回复,说明帖子的受关注度具有异质性,也说明少数热帖吸引了网络论坛中大多数用户的注意力. 查询数据后发现这些广受关注的热帖大多是连载小说或当前最流行的热点话题,证明了研究结论的合理性.

在近百万条帖子中,根帖只有 5,397 个,发过根帖的用户只有 4,611 个,有 2,329 个用户只发帖不回帖,有 408,264 个用户只回帖不发帖. 单从这些数据就能看出用户的发帖回帖行为具有差异性,下面利用复杂网络理论研究网络论坛用户行为分布和差异. 根据网络爬虫从天涯社区的一个版块获取的数据,建立了一个基于网络论坛用户回复行为的虚拟社区网络,表 1 显示了该网络的基本数据. 可以看出,节点的平均入度很小,最大入度很大,说明论坛中用户的影响力差异很大,绝大多数用户默默无闻,不受

关注;节点的平均出度比平均入度大而最大出度相对最大入度很小,说明论坛中用户活跃度相比影响力较为平均,大量不受关注的用户依然很活跃.

表1 虚拟社区网络的基本数据

平均入度	最大入度	平均出度	最大出度	直径	平均路径长度
1.547	14319	2.378	379	15	4.572

BBS 虚拟社区网络是有向网络,有向链接有利于研究信息传播过程,定位信息流通中的内容,例如,搜索算法 PageRank 将从源网页到目的网页的链接作为源网页对目的网页的支持,从而评判网页的影响力,并对其进行排序^[22]. 同样,在虚拟社区网络中,用户 A 对用户 B 的回复也说明用户 A 对用户 B 的关注,根据得到的回复数可以分辨权威用户. 为了进一步深入分析论坛用户交互行为间的差异,该虚拟社区网络的节点入度、出度分布如图 2 所示. 对图 2 中的点集进行拟合发现虚拟社区网络的节点入度、出度符合幂率分布,即

$$P(k) \sim k^{-\gamma}, \text{ wherein } \begin{cases} \gamma = 1, & k = k_{in}, \\ \gamma = 2.4, & k = k_{out}, \end{cases} \quad (2)$$

其中,入度分布的幂率指数 $\gamma = 1$, 出度分布的幂率指数 $\gamma = 2.4$. 之前的实证分析对社会网络和自然生物网络等复杂网络的节点度都服从幂率分布,即 $P(k) \sim k^{-\gamma}, \gamma \in [1, 3]$ ¹⁰. 基于用户回复行为的虚拟社区网络的节点度分布情况与此符合,表明它是无标度网络.

由图 2 知,虚拟社区网络中少量用户的热帖吸引了绝大多数用户的注意力,具有很强的影响力,绝大多数用户无法引起关注;少数用户很活跃,但大多数用户存在惰性. 用户喜欢回复有影响力的用户,喜欢看浏览量高的帖子,恰恰符合了 BA 无标度网络“穷者越穷、富者越富”的规则. 值得注意的是,相比出度,入度分布具有更明显的长尾效应,说明用户的影响力分布广泛,具有很大的差异性,具有吸引力和影响力的用户不多,但这些用户的号召力和影响力不容小觑,他们在信息传播和普及中起着至关重要的作用. 而出度分布的尾巴很短,说明用户间活跃度差异不大. 即便是活跃度很高的用户与普通用户的活跃度相差也不很多,即活跃度值域不大,表明了网站的成功不能仅依靠少数活跃用户,而应提高大多数普通惰性用户的兴趣和注意力.

那么,用户的影响力与其自身的活跃度有何关系呢? 图 3 显示了虚拟社区网络节点入度、出度比

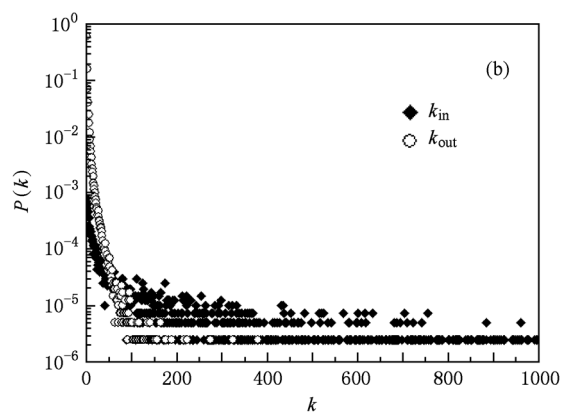
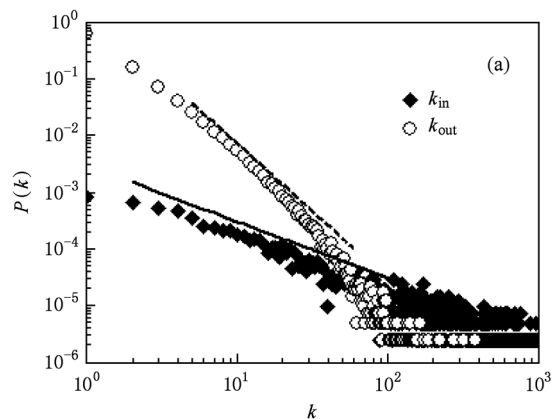


图2 虚拟社区网络的入度、出度分布 (a) 双对数坐标; (b) 半对数坐标

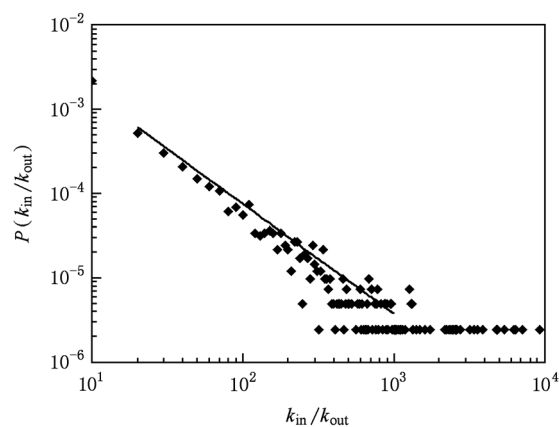


图3 节点入度与出度比的分布图

也服从幂率分布,幂率指数 $\gamma = 1.3$. 不难看出,大部分用户的影响力与活跃度相差在百倍以内,说明他们之所以具有影响力是因为其活跃回帖吸引了大量的关注者;但少数用户的影响力是活跃度的几百倍,说明这些用户并非靠频繁回帖来取得关注,其

自身凭借名人效应或有趣的帖子内容已经成为虚拟社区网络中有影响力的权威节点. 查询数据后发现, 这些用户大多是连载小说作家或网络论坛红人. 由此可知, 用户要想在虚拟社区网络中占据领导地位, 光靠灌水是远远不够的, 还需要发表高质量的帖子或形成个人品牌.

为了进一步研究网络论坛用户影响力和活跃度的形成机理, 考察用户加入论坛后的行为特征, 本文分别选取入度和出度最大的前 5 名用户, 图 4 分别显示了入度(a)和出度(b)随时间的演化过程. 其中, 有 4 个用户的入度随时间逐渐增加, 有 1 个用户的入度在某一时间后骤然增加. 这说明, 在加入论坛后影响力的增加分为缓慢递增和骤然猛增两种, 即有些用户的知名度和影响力是日复一日积累的, 而有些用户可以依靠一个热点话题瞬间成为网络论坛中具有影响力的人. 用户出度演化没有表现出太大差异, 用户的活跃度随时间循序渐进地增加, 说明任何用户的活跃度都是随时间积累得到的.

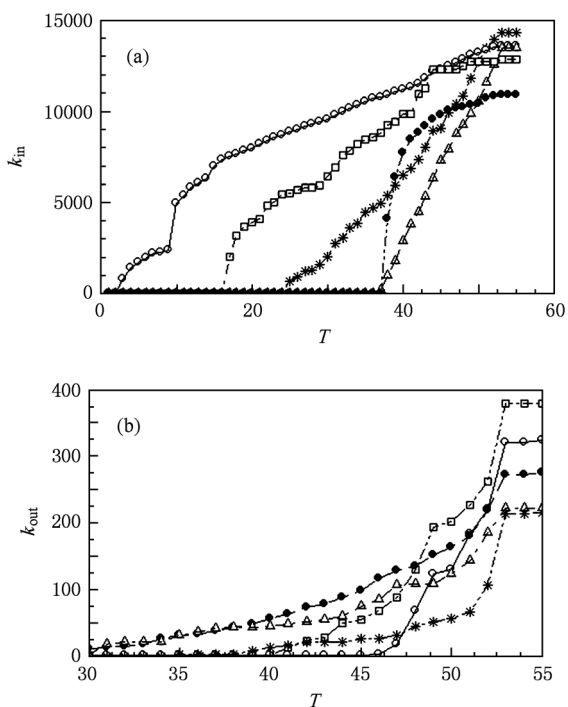


图 4 前五名用户的入度、出度随时间演化图 (a)入度;(b)出度

图 5 显示了用户浏览量及其入度的关系. 其中, 用户浏览量 $R(k_{in})$ 为所有度为 k_{in} 的用户的总浏览量的平均值. 可以看出, 用户的浏览量正相关于该用户的入度, 说明用户的浏览量和入度一样可以反映该用户的影响力和受关注程度. 该结论是合理的, 因为大的节点入度可以增加用户的知名度, 从

而增加浏览量; 相反, 大的浏览量也同样容易使帖子引起新的用户回复.

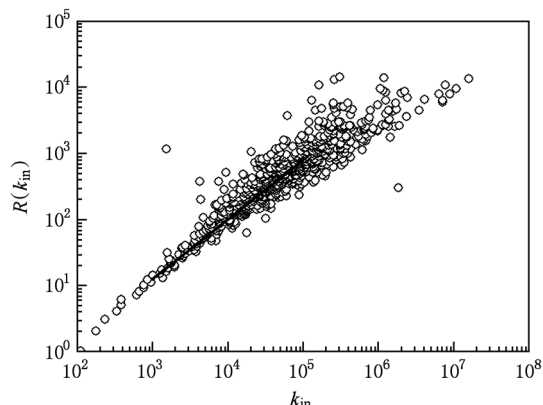


图 5 用户浏览量与入度的关系

在该回复网络的近百万有向连接中, 双向连接的比例还不到 1%, 说明虚拟社区网络是一个相当不对称的网络. 为了定量分析虚拟社区网络的不对称性, 考察用户的回复偏好, 节点度相关函数 K_{in} 与出度 k_{out} 关系如图 6 所示. 明显地, 节点出度越大, 其度相关函数越小, 说明复杂网络中度大的节点喜欢与度小的节点相连. 网络论坛中活跃用户喜欢回复惰性用户, 即普通用户都愿意回复知名度高的用户, 相反, 知名度高的用户之间、普通用户之间的回复互动行为相对较少. 由此可见, 与现实社会网络和基于好友关系的网上社会网络不同, 网络论坛基于用户回复关系的虚拟社区网络与 Web 网和博客网络一样具有异配性^[23]. 这是由于网络论坛具有很强的“贡献驱动”特性, 即论坛中少数极度受欢迎的用户仅靠发表热帖引起大量普通用户的回复, 而非依靠活跃频繁回复他人吸引注意力. 这种异配性反映了网络论坛中不同种用户的混合连接.

基于用户回复行为的虚拟社区网络是一个有向网络, 为了考察论坛中用户的集聚特性, 即其邻居间的回复行为, 需要考察无向网络的簇系数. 忽略网络中所有不对称(单向)的边, 删除因为删边而被隔离的孤立节点, 得到该虚拟社区网络的双向(无向)网络. 该双向网络仅有不到 500 个节点, 尚不足虚拟社区网络的 1%, 平均节点度为 1.57. 整个网络的平均簇系数为 0.111, 远远大于与其具有相同规模的随机网络的簇系数 $C_{rand} = 1.57/463 = 0.0034$.

图 7 显示了双向网络中节点簇系数与度的关系. 随着节点度的增加, 簇系数的变化表现出不同

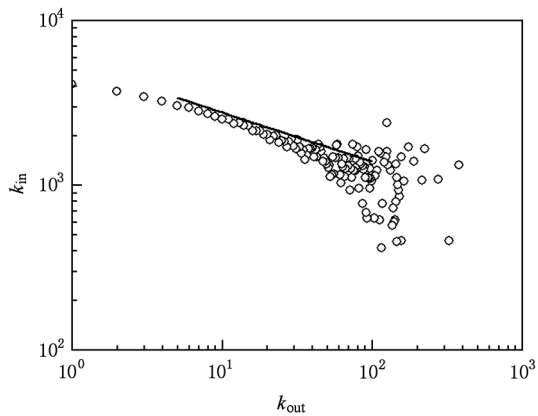


图6 节点度相关函数与出度关系图

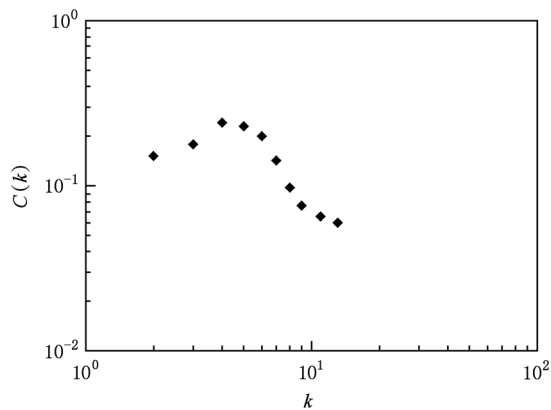


图7 双向网络中节点簇系数与度的关系图

的变化趋势. 不同于文献[16]中簇系数与节点度的幂率关系, 基于用户回复行为的虚拟社区网络中簇系数与节点度表现出两种特性. 当 $k < 4$ 时, 簇系数随节点度的增加而增加; 当 $k > 4$ 时, 簇系数随节点度的增加而减小. 这是因为, 当节点度很小 ($= 2$) 时, 用户只与极个别用户有交互, 这样的用户处于网络中的边缘, 其邻居间的交互很少, 称为网络边缘节点. 随着节点度增加至 4, 用户的邻居逐渐增多, 邻居间交互的可能性增加, 这种用户多属于内部连接较好的小社团, 从而簇系数增加到最大, 称为社团中心节点. 随着节点度的进一步增加到 10 以

上, 用户在网络论坛中的重要性逐渐凸显, 很多在虚拟社区网络中起着枢纽作用, 称为接口节点. 作为不同社团间的接口, 度最大的节点同时属于很多不同的小社团, 其属于不同社团的邻居间鲜有交互, 造成了度最大的节点簇系数最小的结果. 节点簇系数的不同说明基于回复的虚拟社区网络具有分层性质和社团结构, 社团间通过簇系数很小的接口节点连接, 社团内具有簇系数很大的社团中心节点和簇系数较小的网络边缘节点.

4. 结 论

本文对互联网论坛中用户的交互行为进行实证分析, 建立基于隐式交互行为的虚拟社区网络, 通过考察该复杂网络的统计特性, 如入度、出度、度相关系数和簇系数等, 发现隐式的虚拟社区网络是有向、不对称、异配的无标度网络. 网络论坛中帖子的传播程度并不服从幂率分布, 而帖子的受关注度、用户的影响力和活跃度均服从幂率分布; 用户的浏览量正相关于该用户的入度; 基于回复关系的虚拟社区网络具有分层性质和社团结构, 社团间通过簇系数很小的接口节点连接, 社团内具有簇系数很大的社团中心节点和簇系数较小的网络边缘节点. 由此可知, 网络论坛的信息传播性能优异, 绝大多数帖子都有一定范围的传播; 论坛中具有“贡献驱动”特性的权威用户在信息传播中的作用不容小觑; 用户要想在虚拟社区网络中占据权威地位, 光靠灌水是远远不够的, 还需要发表高质量的帖子或形成个人品牌. 部分结论具有一般性, 但由于目前对虚拟社区网络及用户行为尚无定论, 并不能通过某些特性的一般性来断定每个虚拟社区及其用户行为的特性. 因此, 本文通过实证分析研究网络论坛虚拟社区及其用户回复行为特性, 并从定量的角度证实了社会学定性分析的结论, 具有实际意义. 下一步, 可以根据基于回复行为的虚拟社区网络的基本性质, 建立网络论坛人际交互模型, 以期模拟现实虚拟社区网络中个体交互行为.

[1] Castellano C, Fortunato S, Loreto V 2009 *Rev. Mod. Phys.* **81** 591

[2] Wang L, Fu F, Chen X J, Wang J, Wu B, Chu T G, Xie G M 2008 *CAAI Transactions on Intelligent Systems* **3** 95 (in Chinese)

- [王 龙、伏 锋、陈小杰、王 靖、武 斌、楚天广、谢广明 2008 智能系统学报 **3** 95]
- [3] Si X M, Liu Y, Xiong F, Zhang Y C, Ding F, Cheng H 2010 *Physica A* **389** 3711
- [4] Si X M, Liu Y, Zhang Z J 2009 *Int. J. Mod. Phys. C* **20** 2013
- [5] Shi X M, Shi L, Zhang J F 2010 *Chin. Phys. B* **19** 038701
- [6] Stauffer D 2005 *AIP Conf. Proc.* **779** 56
- [7] Sobkowicz P 2009 *J. Artif. Soc. Soc. Sim.* **12**
- [8] Watts D J, Strogatz S H 1998 *Nature* **393** 440
- [9] Barabai A L, Albert R 1999 *Science* **286** 509
- [10] Newman M E J, Barabási A L, Watts D J 2006 *The Structure and Dynamics of Networks* (Princeton: Princeton Univ. Press).
- [11] Wang Y, Zheng Z G 2009 *Acta Phys. Sin.* **58** 4421 (in Chinese) [王 延、郑志刚 2009 物理学报 **58** 4421]
- [12] He M H, Zhang D M, Wang H Y, Li X G, Fang P J 2010 *Acta Phys. Sin.* **59** 5175 (in Chinese) [何敏华、张端明、王海艳、李小刚、方频捷 2010 物理学报 **59** 5175]
- [13] Feng C F, Guan J Y, Wu Z X, Wang Y H 2010 *Chin. Phys. B* **19** 060203
- [14] Porter M A, Mucha P J, Newman M E J, Friend A J 2007 *Physica A* **386** 414
- [15] Zhang L, Liu Y 2008 *Acta Phys. Sin.* **57** 5419 (in Chinese) [张 立、刘 云 2008 物理学报 **57** 5419]
- [16] Fu F, Liu L H, Wang L 2008 *Physica A* **387** 675
- [17] Hu H B, Wang X F 2009 *Phys. Lett. A* **373** 1105
- [18] Mislove A, Marcon M, Gummadi K P, Druschel P, Bhattacharjee B 2007 *Proc. 7th ACM SIGCOMM Conf. Int. Meas.*, San Diego, USA, ACM Press, 29
- [19] Grabowski A 2007 *Physica A* **385** 363
- [20] Ahn Y Y, Han S, Kwak H, Moon S, Jeong H 2007 *Proceedings of the 16th international conference on World Wide Web*, Banff, Canada, ACM Press, 835
- [21] http://en.wikipedia.org/wiki/Internet_forum
- [22] <http://en.wikipedia.org/wiki/PageRank>
- [23] Newman M E J 2003 *Phys. Rev. E* **67** 026126

Empirical analysis of interpersonal interacting behavior in virtual community*

Si Xia-Meng[†] Liu Yun

(School of Electronic and Information Engineering, Beijing Jiaotong University, Key Laboratory of Communication and Information Systems (Beijing Jiaotong University), Beijing Municipal Commission of Education, Beijing 100044, China)

(Received 23 August 2010; revised manuscript received 29 September 2010)

Abstract

Internet forum is the place where online mass incidents usually happen, and is very important in information diffusion and opinion spreading. We empirically analyze the interacting behavior of internet forum users, construct the virtual community network based on implicit interaction behavior. Through inspecting the statistical properties, we find that this virtual community network is directional, asymmetric and disassortative scale free network, and it has hierarchy and community structure. Moreover, user page view is positively correlated with his in-degree, so both can describe the influence power and popularity. From quantitative view, the results show that the effects of influential user with ‘contribute-driven’ property on information spreading can not be ignored; the success of internet forum needs to increase the activity of the most inert users; if user wants to be authoritative in virtual community networks, he needs to propose interesting message or has a personal brand.

Keywords: statistical physics, complex networks, interpersonal interaction, power law distribution

PACS: 89.75.-k, 29.85.-c, 89.75.Da

* Project supported by the National Natural Science Foundation of China (Grant No. 60972012), the Beijing Natural Science Foundation (Grant No. 4102047), the Fundamental Research Funds for the Central Universities (Grant No. 2009YJS007), the Major Program for Research on Philosophy & Humanity Social Sciences of the Ministry of Education of China (Grant No. 08WLI101), the Service Business of Scientists and Engineers Project (Grant No. 2009GJA00048), the Academic Discipline and Postgraduate Education Project of Beijing Municipal Commission of Education, and the 2010 Fengtai Small and Medium-Size High-tech Enterprises Technological Innovation Fund (Grant No. CXJJ1007).

[†] E-mail: sixiameng@gmail.com