

在线股票社区用户参与行为研究 ——基于股吧的实证分析

高娇娇, 张仁杰

(西南交通大学 经济管理学院, 成都 610031)

摘要 在线股票社区是股票投资者进行交流的重要渠道, 他们通过发帖和回帖行为实现信息的传播, 而信息的传播对投资者情绪有重要影响。本文以“股吧”论坛的“股市实战”版块为研究对象, 研究了论坛用户发帖和回帖行为的特征。结果表明: 用户多在股市开市期间进行发帖, 且大多数用户是信息的“接收者”而非“发送者”; 通过构建指数随机图模型发现互惠性在用户回帖行为中较为活跃; 该回帖网络存在一定程度的小团体特征; 高社会资本的用户是论坛中发回帖行为最积极的用户, 他们的言论会对其他用户产生重要影响。

关键词 在线股票社区, 社会网络, 指数随机图模型, 发回帖行为

中图分类号 G206.2

1 引言

根据中国互联网络信息中心 (China Internet Network Information Center, CNNIC) 2018 年 1 月发布的第 41 次《中国互联网络发展状况统计报告》^[1], 截至 2017 年底, 中国网民规模已达到 7.72 亿, 即时通信、搜索引擎、网络新闻和社交作为基础应用, 其用户规模保持平稳增长。伴随着互联网的发展, 越来越多的股票投资者倾向于从互联网获取相关信息来指导自己的决策行为。在线社区因其自由、开放的特点, 逐渐成为用户表达意见、进行知识共享或者交换信息的平台^[2]。他们常使用在线社区去获取相关的情感支持和信息资讯。我国股票在线社区种类繁多, 许多社区的日均发帖量都达到了上千条。其中, “股吧”作为最为活跃的股票论坛之一, 是许多学者研究的对象^[3-6]。为了探讨股民间交流网络的特点, 本文将以此论坛中活跃度最高的“股市实战”版块为基础进行后续研究。

在线社区本质上是一种特殊的社会网络, 用户通过发回帖来建立关系, 进而形成回帖关系网络。在回帖关系网络中, 节点表示发帖人和回帖人, 网络中的有向边则表示用户间的回复关系^[7]。研究用户的回帖行为, 对维系社区的持续发展有重要作用。但是, 目前对在线股票社区的相关研究主要集中在社区信息对股票价格的影响, 并没有从用户发回帖行为的角度分析在线股票社区信息传播的特点。理解在线股票社区的社会网络机制不仅有助于了解投资者情绪在用户之间的传播特点, 还能帮助论坛管理者更科学地运营和管理社区。因此, 对在线社区用户发帖和回帖行为进行研究具有重要意义。结合上述分析, 本文将通过进一步的研究回答下列问题: 用户的发帖偏好是什么? 他们更喜欢在哪些时间段发帖? 什么样的帖子更容易得到别人的关注? 什么样的用户更容易回复他人的帖子?

基于上述问题, 本文通过搜集“股市实战”版块的相关信息, 利用网络爬虫技术分析用户的发帖行为特征并通过构建社会网络模型来研究用户的回帖行为。

2 文献综述

2.1 在线社区用户参与行为研究

在线社区用户参与行为可以定义为“用户访问在线社区并留下痕迹”^[8], 如用户在社区中发帖、回帖、评论、点赞、私信及添加好友等行为。按照用户参与的目的不同, 将其归结为查询信息、社交互动、需求社会支持。此外, 有研究者从不同视角对在线社区参与行为进行了研究。Zhou 等^[9]和 Cetto 等^[10]从知识管理的角度出发, 将用户的在线参与行为划分为知识贡献与知识接受。秦敏和梁溯从社会行为理论的视角来分析用户参与贡献与用户识别机制的关系^[11]。赵蓉英和曾宪琴通过研究微博用户的参与行为发现用户参与可以促进信息的扩散^[12]。而 Mislove 等通过研究 Orkut、Flickr 和 YouTube 社交网络社区的人际网络, 发现网络用户具有幂律特征、小世界特征和无标度特征^[13]。杨化龙则将视角放在在线医疗社区上, 研究医生与患者的交互过程^[14]。Wu 等从营销学的角度, 来研究在线社区用户参与行为与口碑传播的关系^[15]。Wellman 等从社会的角度来研究用户在线行为对社会网络结构变化的影响^[16]。

尽管在线社区用户参与行为的作用得到了广泛认同, 获得不少研究者的关注与研究, 但鲜有研究探析用户的某类具体行为类型, 如用户在社区内的发帖与回帖行为。

2.2 在线股票社区用户参与行为研究

已有对在线股票社区的研究主要体现在网络舆论、投资者情绪、投资者关注度等对股价的影响。在网络舆论上, Liu 等研究了 2009~2016 年百度与 360 搜索引擎中股票的搜索指数与股价的关系^[17]。Oliveira 等利用 Twitter 数据来分析网络舆论对股票价格的影响^[18]。金秀等对股吧上证指数观测样本进行分析, 预测投资者情绪对股票收益的影响^[4]。随着信息技术的普及, 学者们开始着重将投资者情绪与投资者舆论结合起来。Wu 和 Lin 整理 2001~2014 年中国台湾地区台湾经济新报 (Taiwan Economic Journal) 数据库搜集的关于上市公司的新闻报道, 来研究媒体舆论对投资者行为和股票价格的影响^[19]。Ho 等则研究社交媒体情绪与股票收益之间的关系^[20]。Checkley 等使用 Twitter 和 StockTwits 上的数据来提取公众的情绪指标预测股票市场的反应^[21]。缪杰利用百度指数的关键词的统计, 以此为基础构建关注度来分析投资者的关注与股价的关系^[22]。虽然, 已有研究已经对在线股票社区进行了广泛研究, 但关注点多放在股票价格与论坛内容之间的联系, 而论坛内信息的传播特点依旧值得我们进一步深究。

2.3 网络发回帖行为研究

目前, 使用“社会网络”概念来研究在线论坛机理的文章基本以论坛的回帖网络为基础来分析论坛用户的发回帖行为特征, 主要使用随机指数模型、小世界模型和无标度模型这 3 种基本模型。许多学者在这些模型的基础之上, 对不同的网络社区进行了实证研究。例如, 模拟用户的回帖行为、分析用户回帖行为的分布特性、统计分析用户的回帖行为等。Himmelboim 研究了用户间的回帖结构及影响因素^[23]。Kimmerle 等对在线医疗社区用户回帖的内容进行分析, 来研究咨询式的帖子对回帖的影响^[24]。Meng 等采用封闭的网络, 发现用户评论的网络结构对不同社会支持接受的影响^[25]。Wu 等为研究用

户回帖关系,提出了一个回帖网络演化模型^[26]。而 O’Dea 等研究 Twitter 上关于自杀的帖子,发现帖子的回复率与回复速度之间的关系,对于预测自杀率有一定的策略性帮助^[27]。Yu 等则采用统计方法分析用户的评论和回帖行为,发现评论数和回帖数服从不同幂指数的幂律分布^[28]。

上述研究多是基于数理统计分析法或社会网络分析法,且大部分没有考虑用户的个体属性对其发帖与回帖行为的作用。由于用户的个体属性对其发回帖行为有重要的影响,将个体属性纳入研究中很有必要。

2.4 指数随机图模型

指数随机图模型(exponential random graph models, ERGM)是以关系为中心的随机社会网络分析模型,也是用于研究网络结构的统计模型。ERGM 在社会网络结构的研究中得到广泛的使用与发展。Box-Steffensmeier 等利用 ERGM 对一个企业集团进行角色分析,了解了企业内部的社会结构变化^[29]。Windzio 对 1990~2013 年全球网络的移民数据使用 ERGM 进行分析,来分析各国间的移民网络结构^[30]。Hossain 等为了在发生森林大火时,提高管理人员紧急调度不同机构的协调能力,使用 ERGM 构建了急救管理人员协作网络^[31]。Wimmer 和 Kewis 则通过 ERGM 研究了 Facebook 中朋友关系网络里存在的种族同质性现象^[32]。Shen 和 Monge 使用 ERGM 发现在线社区中的用户在参与交互活动中存在着互惠性^[33]。

基于上述分析可以发现,目前关于股票在线社区的研究主要是基于文本挖掘技术来研究用户情绪与股价的影响,并没有从社会网络出发来研究用户的发帖和回帖行为。本文以此为出发点,首先,通过网络爬虫技术揭示该版块用户的发帖行为特征;其次,使用 ERGM 构建该版块的回帖网络以分析用户的回帖行为;最后,对该在线股票社区的用户发回帖行为进行总结,给出相应的结论和建议。

3 用户发帖行为特征

使用 Python 爬虫框架采集“股市实战”版块 2017 年 10 月到 2018 年 3 月共半年时间的帖子,帖子总数为 135 086 条,其中有回帖的主题帖共 25 716 条,占总帖数的 19%。

通过统计单日内各时间段平均发帖量、半年内每日发帖量、主题帖回帖量、主题帖查看量及主题帖正文长度,可以发现用户发帖行为具有较强的规律性。结果如图 1 所示。

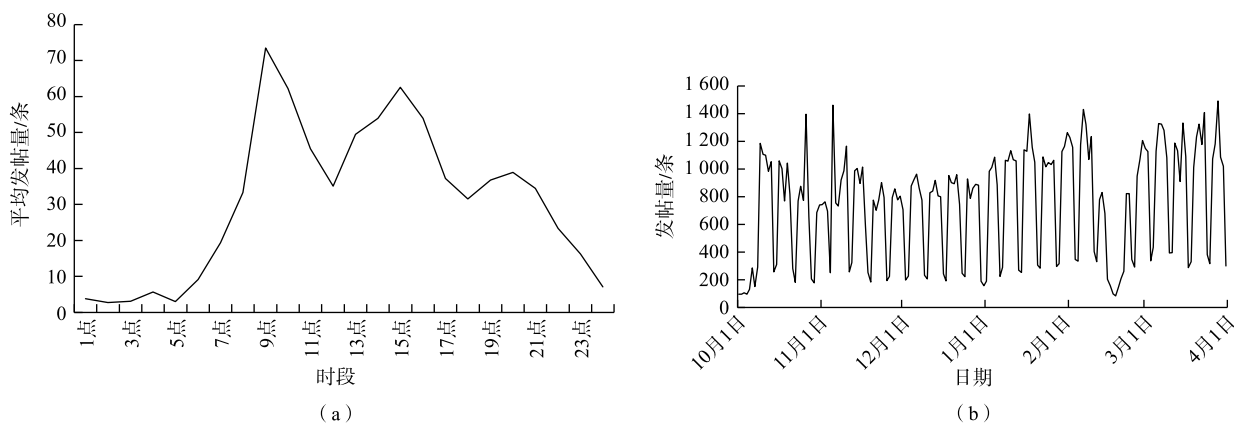


图 1 日内日间发帖量特征

图 1 展示了用户在时间维度上的发帖行为特征。图 1 (a) 为用户日内的平均发帖量,从中可以看

出，一天内包括三个峰值，分别为 9 点、16 点和 20 点。根据股市 9:30~11:30、13:00~15:00 的开市时间判断，当股市开市前半小时，与股票相关的信息量达到峰值；下午则是在股市闭市后信息量才达到峰值；晚间会在 8 点左右迎来又一次小高峰。因此，相关时段内的用户活跃程度较高，此时的帖子能为股票投资者提供大量的信息并对股市产生一定的影响。值得注意的是，中午时段的发帖量较少，说明在上午闭市后，并不会大量帖子对上午的股市进行讨论。

图 1 (b) 为半年内每日的发帖量统计。从图 1 (b) 中可以看出发帖量呈有规律的波动，具体分析发现各峰值均出现在工作日，而发帖量下降的日子均为节假日，如 2 月 1 日到 3 月 1 日间的谷底正值春节放假。这说明股票论坛的用户只会在股市开市的时段关注论坛内的资讯，在节假日期间，相比继续关心股票信息，他们更愿意将精力放在其他事情上。

图 2 统计了主题帖的部分特征，其中图 2 (a) 为回帖数统计、图 2 (b) 为查看数统计、图 2 (c) 为文章字数统计，三张图均为长尾结构。图 2 (a) 和图 2 (b) 说明多数帖子都没有人进行回帖且被查看数较少。对比回帖数和查看数可以发现帖子的被查看数远远大于被回帖数，这说明论坛中的大多数用户更愿意查看他人发的帖子而不愿意发表自己的观点想法。图 2 (c) 的特征说明大多数帖子以简短的文字为主，或发表某一观点，或进行简短的评述。但也有部分用户会发表字数大于 1 000 的帖子，通过人为查看，发现这些帖子的主题包括对某只股票的分析、对未来一周股市的预测和论坛活动介绍等内容。

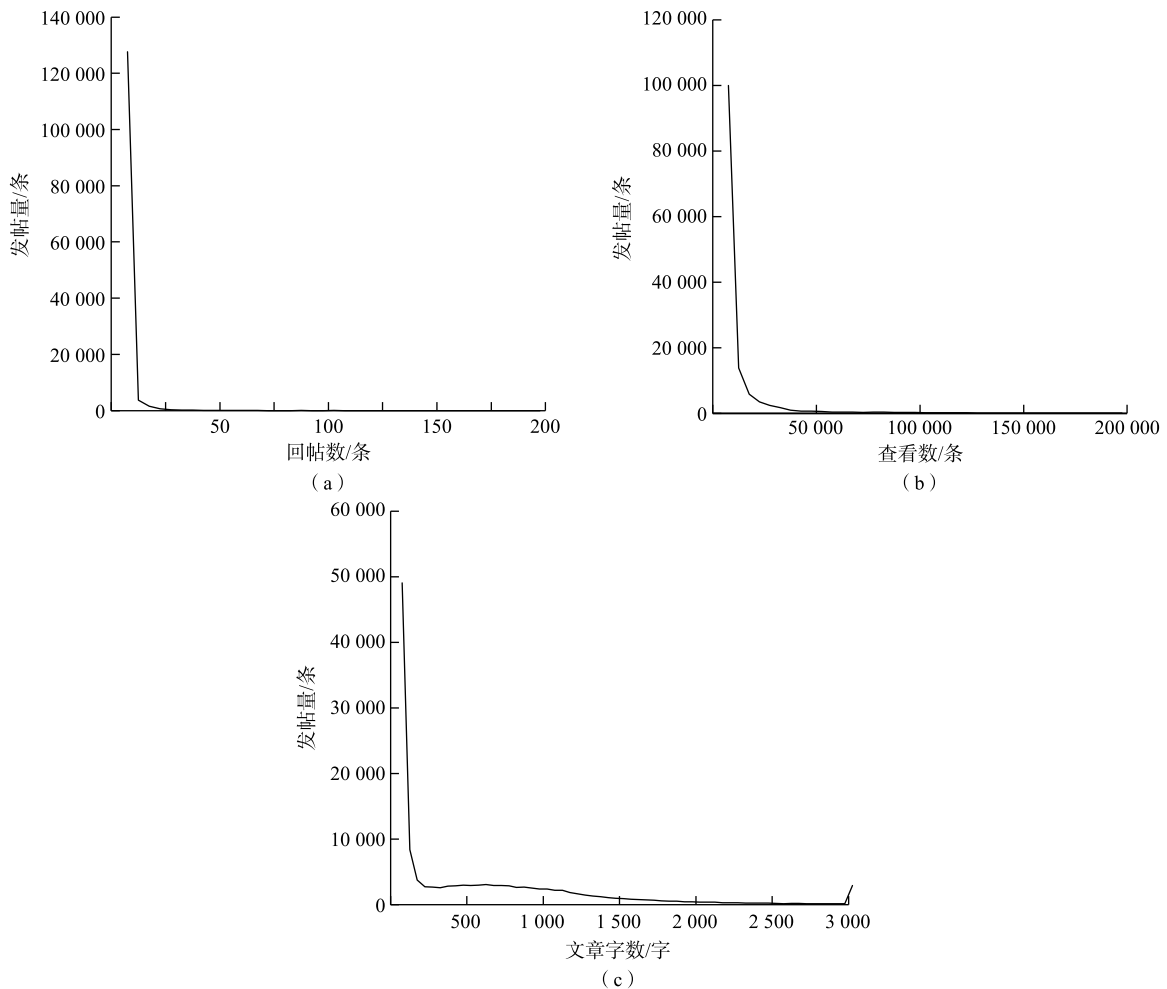


图 2 主题帖部分特征

通过上述分析可以得出以下结论：①论坛活跃度较高的时段主要集中在股市开市的时段，其余时间论坛用户更倾向于做别的事；②论坛内的多数用户一直处于“潜水”状态，相比于发表自己的意见，更愿意查看他人的言论；③帖子主要以简短的评论类帖子为主，针对股票或股市现状进行综合分析的帖子数量较少。上述结论说明多数股民是信息的接受者，那些查看数高、回帖数高的帖子能传递大量信息给论坛用户以影响他们的决策判断。对于论坛管理者而言，应当提高用户回帖的积极性、提高主题帖质量以保证论坛能高质量地运行下去。

4 回帖关系网络特征

通过使用 ERGM，本部分将进一步分析“股市实战”版块用户间的网络特征，以探求何种用户更容易受到其他用户的关注、何种用户更愿意与他人互动。由于半年内，该版块的用户节点数多达 2 万多个，而现有研究表明过高的网络节点数不利于模型的构建^[34, 35]，为了保证结论的有效性，本部分将以该版块 2018 年 3 月的回帖网络作为基准进行模型的构建和分析，后续会通过稳健性检验来验证所得结论的普遍性。3 月回帖网络中共包含 5 526 个节点数及 8 204 条边，使用 Gephi 软件绘制的 3 月份关系网络图如图 3 所示。

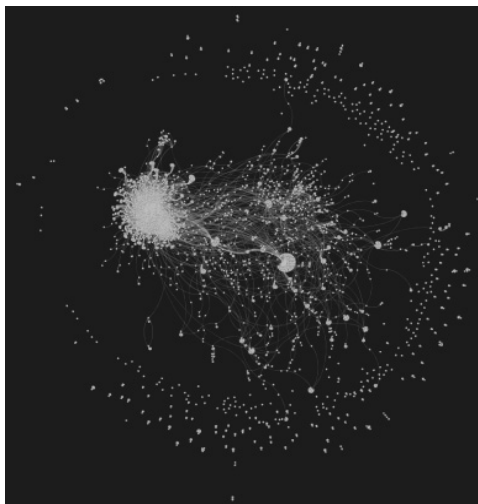


图3 关系网络图

图3中外游离的点表示一些孤立的关系对，即这些用户仅对少数用户进行了回帖。图3左侧有一块面积较大的高亮区域，说明在这块区域内的用户间有频繁的互动行为，彼此间的回帖关系紧密。图3中多个亮度较高的圆形区域表示一些得到较多回帖的用户，他们的观点更容易对他人产生影响。除此以外，这些高亮区域间也有很强的联系，说明同一个用户会在多个用户间进行回帖。

4.1 回帖网络基本特征

通过使用 Gephi 软件进行基本网络特征的统计，该社会网络的图密度为 0.001、平均聚类系数为 0.006。这说明股票在线社区的回帖网络是一个非常松散的社会网络，网络中的成员会围绕部分核心用户建立回帖关系，但大多数用户间互动不频繁。不难分析造成这一现象的原因：股票社区日均发帖量可以达到上千条，多数用户仅会参与讨论度高、关注度高的帖子，造成用户间的高互动仅停留在少

部分的帖子上。同时,根据第三部分得到的结果,多数用户更愿意查阅帖子而不发表意见,愿意主动发帖的用户更是少之又少,这就导致多数用户间无法建立起发回帖的关系。在上述两个原因的作用下,松散的回帖网络便形成了。

此外,该社会网络的模块度指数为 0.718,说明该网络的社区结构强度较高,划分社区的质量也较好。该网络具有较强的社团性。

4.2 研究假设与 ERGM

为了进一步研究股票社区用户回帖行为的特征,在前人研究的基础上,从网络结构和用户属性两方面提出了相应假设。此外,本文还从该网站的特点出发,将认证股评员和认证投资机构对用户回帖行为的影响纳入假设中。

互惠性是社会网络的重要特征^[36]。其含义为若 i 向 j 发出一条关系, j 也会向 i 发出一条关系。在回帖关系网络中,互相回帖是互惠性的表现。较高程度的互惠性意味着用户会相互选择,信息在用户间得到传递。此外,相互回帖也是满足“关系”的途径,彼此间的“礼尚往来”反映了人际的互动^[37]。这种互动可以增加用户间意见交换的频率、丰富帖子的内容、增强社区间成员的关系。因此,提出以下假设。

H₁: 股票社区论坛中存在互惠性特征。

传递性闭合是社会网络中的另一个重要特征。较为明显的传递性闭合表明在该社会网络中存在明显的小团体特征。有学者通过对知乎网络社区的社会网络进行研究,发现存在明显的传递性闭合特征。这种特征会使信息在部分群体间封闭式传播,导致媒体功能的弱化^[38]。与知乎社区平台类似,股吧作为股票信息传播和分享的平台,也很有可能存在传递性闭合特征。因此,提出以下假设。

H₂: 股票社区论坛存在传递性闭合特征。

社会资本能够帮助个体增加获得地位的机会^[39]。在线社区中,用户的影响力、粉丝数可以被看作社会资本。所以,影响力高、粉丝数多的用户在论坛中更容易获得较高的地位,成为别人关注的对象。根据意见领袖理论,这些用户往往具有引导舆论的重要作用,是社会网络中信息和影响的重要来源^[40]。他们的言论能够对其他用户产生影响。基于此,提出以下假设。

H_{3a}: 影响力高的人更能得到他人的回帖。

H_{3b}: 粉丝数多的人更能得到他人的回帖。

何黎等的研究表明,在微博社区中,用户关注数、发文数和被关注数之间存在较高的相关性,用户关注数高表明其活跃程度较高,所发的内容更容易得到他人的关注与回应^[41]。刘行军和王伟军的研究也得到了相似的结论^[42]。在同为在线社区的股票社区中,用户的关注列表表明其感兴趣用户的多寡。关注列表人数多的用户其活跃程度可能更高,就有更大的概率受到他人的关注、得到更多的回应。因此,提出以下假设。

H₄: 关注数多的用户更容易得到他人的回帖。

拥有高势位的用户,往往把握着质量较高的知识产物,信息的传递也常从这些用户流向低势位的用户^[43]。刘璇等的研究进一步验证了这一观点^[44]。在股票论坛中,影响力高的用户是论坛中的高势位群体,这些用户会更愿意对他人的帖子进行回帖,使信息向低势位群体流去。因此,提出以下假设。

H₅: 影响力高的用户更愿意回复他人的帖子。

除了上述用户的社会资本属性外,用户的人口统计学特征也较为重要。在论坛内,官方为用户开通了认证通道,类似于微博的加 V 认证。用户通过上传个人的真实信息及其他相关资料,经由股吧官

方审核通过后可以获得官方认证。这类用户通常是股评师或者投资机构。而他们的言论往往能对其他用户产生影响并刺激他们回帖。因此，提出以下假设。

H₆: 认证股评师和认证机构能得到更多人的回帖。

为了检验以上假设，本文使用了 ERGM。ERGM 的原理是将模拟生成的随机图与实际网络图相比，观察实际网络图中的相关网络特征是否与随机产生的网络图有显著区别，以此确定相应特征是否明显。根据文献^[36]的描述，ERGM 的一般数学公式如下：

$$\Pr(X = x) = P_{\theta}(x) = \left(\frac{1}{k}\right) \exp\left\{\sum_n \theta_n z_n(x)\right\} \quad (1)$$

其中， X 表示模型拟合出的网络； x 表示实际生活中的网络； k 为一个标准化参数，保证式 (1) 是一个合理的概率分布； n 为观测网络中所有网络结构的集合，包含了本文中选取的数个结构； θ_n 为对应结构 n 的参数； $z_n(x)$ 为对应结构 n 的网络统计值，若结构 n 存在于网络 x 中则 $z_n(x)=1$ ，否则为 0。模型的估计方法有最大似然估计法 (maximum likelihood estimation, MLE) 和马尔可夫链蒙特卡洛 (Markov chain Monte Carlo, MCMC) 最大概率估计法，由于前者已被证明估计有偏误，因此本文使用了 MCMC 最大概率估计法对模型进行估计^[45]。

针对每一个假设，表 1 给出了各假设相对应的变量名称及其结构示意图。

表 1 变量名称、假设说明与示意图

变量名称	对应假设	示意图
Mutual	H ₁ : 股票社区论坛中存在互惠性特征	
AT-T	H ₂ : 股票社区论坛存在传递性闭合特征	
Influence_I	H _{3a} : 影响力高的人更能得到他人的回帖	
Fans_I	H _{3b} : 粉丝数多的人更能得到他人的回帖	
Follow_I	H ₄ : 关注数多的用户更容易得到他人的回帖	
Influence_O	H ₅ : 影响力高的用户更愿意回复他人的帖子	
VIP_I	H ₆ : 认证股评师和认证机构能得到更多人的回帖	

本文使用的用户属性数据均来源于该网站的用户个人页面。其中，用户影响力是分类变量，它是用户在股吧里的人气指数，每半星为一级，0~10 共 11 级。影响力等级与用户注册时长、发帖的被评论量、被转发量和被赞量等信息有关。如某用户为三星半，则其影响力记为 8，数字越高表示其影响力越高；用户关注数是连续变量，表明其关注人数的多少；用户粉丝数是连续变量，表明其粉丝数的多少；认证用户数为二分变量，是认证股评师或认证投资机构记为 1，否则记为 0。

根据 ERGM 相关变量对数据结构的要求，为了准确验证相关属性的发出者效应和接收者效应，需要明确定义发出节点和接收节点。由于部分原始数据不满足相应条件，故将相关属性处理成二分变量^[46, 47]。为此，本文作如下处理：用户影响力以 5 为分界点，影响力小于 5 的取 0，其余取 1，表示高影响力用户；关注数和粉丝数以各自的平均数为分界点，低于平均数的用户取 0，其余取 1，表示高关注数、高粉丝数的用户。

表 2 是 5 526 名用户个体属性描述性统计。从表 2 中可以看出影响力的平均值为 0.23，标准差为 0.42，说明影响力高的用户在论坛中占比较小；关注数的均值是 0.21，标准差为 0.41，与影响力类似；粉丝数的平均值仅为 0.09，说明在该论坛内，粉丝数高的用户占比很小，绝大多数用户没有太多的粉丝。认证用户均值为 0.01，标准差为 0.08，说明该论坛内的认证用户人数较少。上述结论与实际生活中的情况相符。

表 2 个体属性描述性统计

属性	观察数	平均值	标准差
影响力	5 526	0.23	0.42
关注数	5 526	0.21	0.41
粉丝数	5 526	0.09	0.29
认证用户数	5 526	0.01	0.08

表 3 给出了个体属性相关性检验。从表 3 中可以看出，关注数和认证用户数与其他变量的相关程度较低；影响力与粉丝数的相关程度略高，说明同为社会资本的这两个属性有一定联系，在后续稳健性检验中将剔除其中一个属性来检验模型的稳健性。

表 3 个体属性相关性检验

属性	影响力	关注数	粉丝数	认证用户数
影响力	1.000	0.086	0.425**	0.038
关注数	0.086	1.000	0.038	-0.021
粉丝数	0.425**	0.038	1.000	0.044
认证用户数	0.038	-0.021	0.044	1.000

**表示 $p < 0.05$

4.3 回归结果

通过使用 R 语言的 ERGM 包，将收集到的数据进行模型模拟。表 4 展示了“股市实战”版块 3 月回帖关系网络的估计结果。参数大于标准误差的两倍以上即认为该变量显著，若参数为正表示该种网络特征与随机图相比出现的概率更高，反之说明出现这种网络特征的概率较低^[48]。参数不显著或符号为负都说明实际网络不支持提出的假设。

表 4 回归结果

变量名称	假设	参数	标准误差	p 值	收敛统计量
Mutual	H_1	2.876***	0.183	< 0.001	-0.071
AT-T	H_2	1.494***	0.278	< 0.001	-0.079
Fans_I	H_{3a}	2.171***	0.027	< 0.001	0.052
Influence_I	H_{3b}	0.947***	0.029	< 0.001	0.107
Follow_I	H_4	-0.060*	0.026	0.021	-0.084
Influence_O	H_5	0.578***	0.023	< 0.001	0.065
VIP_I	H_6	0.646***	0.074	< 0.001	-0.077

***表示 $p < 0.001$ ，*表示 $p < 0.05$

注：收敛统计量 = (观测值 - 样本均值) / 标准差

根据模拟结果可以看出，多数假设的收敛统计量绝对值均小于 0.1，而假设 H_{3b} 的收敛统计量虽高于 0.1，但仅比 0.1 多出 0.007，尚在可接受范围^[36]。因此模型完全收敛，拟合结果可信^[48]。进一步进行稳健性分析以确定回归结果是否普遍适用，其中， M_0 为基准模型； $M_1 \sim M_3$ 分别为将高用户属性的

用户选取规则变为所有用户的前 40%、前 25% 和前 10%，用于检验数据处理方法对结论是否有影响；M4~M5 是原模型分别去掉粉丝属性和影响力属性后的回归结果，用于检验变量相关性对结论的影响；M6~M10 分别使用相同方法对 2017 年 10 月到 2018 年 2 月的数据进行了模拟分析，用于检验时间维度对结论的影响。最终结果见表 5。

表 5 稳健性检验

对应假设	M0	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	结论
	三月	前 40%	前 25%	前 10%	去粉丝数 (入)属性	去影响力 (入)属性	二月	一月	十二月	十一月	十月	
H ₁	2.876*** (0.183)	3.119*** (0.173)	2.864*** (0.169)	3.218*** (0.162)	2.875*** (0.172)	3.003*** (0.171)	2.114*** (0.274)	1.795*** (0.308)	2.723*** (0.199)	2.328*** (0.243)	1.819*** (0.262)	支持
H ₂	1.494*** (0.278)	2.104*** (0.211)	1.513*** (0.271)	2.210*** (0.222)	1.573*** (0.262)	1.721*** (0.254)	0.101 (0.524)	-0.456 (0.693)	1.496*** (0.316)	1.252*** (0.337)	1.267*** (0.300)	不支持
H _{3a}	2.171*** (0.027)	2.422*** (0.053)	1.160*** (0.024)	0.532*** (0.036)		2.741*** (0.022)	2.966*** (0.031)	2.366*** (0.025)	2.115*** (0.022)	2.603*** (0.027)	2.161*** (0.027)	支持
H _{3b}	0.947 (0.029)	1.164 (0.029)	1.888 (0.027)	1.126 (0.026)	2.002 (0.023)		0.821 (0.032)	0.436 (0.026)	1.044 (0.023)	0.411 (0.028)	0.873 (0.029)	支持
H ₄	-0.060 (0.026)	-1.402*** (0.052)	-0.235*** (0.025)	0.330*** (0.037)	-0.041 (0.025)	0.002 (0.026)	-0.000 (0.025)	-0.269*** (0.023)	-0.010 (0.022)	-0.124*** (0.024)	-0.135*** (0.024)	不支持
H ₅	0.578*** (0.023)	0.402*** (0.022)	0.546*** (0.023)	0.406*** (0.030)	0.576*** (0.023)	0.582*** (0.023)	0.451*** (0.023)	0.487*** (0.020)	0.416*** (0.020)	0.488*** (0.024)	0.612*** (0.021)	支持
H ₆	0.646*** (0.074)	0.879 (0.073)	0.530 (0.074)	1.173 (0.074)	0.811 (0.076)	0.548 (0.075)	1.041 (0.067)	0.825 (0.069)	2.299*** (0.032)	0.499 (0.072)	0.796 (0.060)	支持

***表示 $p < 0.001$, *表示 $p < 0.05$

注：“去粉丝数(入)属性”指模型去除变量“Fans_I”后的回归结果，“去影响力(入)属性”指模型去除变量“Influence_I”后的回归结果；括号内为标准误差

综合表 4 和表 5 的结果可以得到以下结论：

(1) H₁ 的参数为 2.876，显著且通过稳定性检验，H₁ 得到验证。这说明在“股市实战”版块中存在互惠性且这种性质是活跃的^[35]，论坛中的成员愿意彼此进行意见交换和知识共享。从时间趋势上来看，该变量的系数维持在 1.795 以上，且 11 个模型中有 9 个模型的系数达到了 2 以上，一直处于高水平状态。可见互惠性特征是该网络的重要特征之一，这种模式能够帮助用户间建立更加紧密的关系，使论坛形成良性的社交生态、稳定持续地向前发展。

(2) H₂ 的参数为 1.494，结果显著，但未通过稳健性检验，故 H₂ 没有得到验证。从时间趋势上来看，1 月与 2 月不存在明显的传递性闭合，故小团体特征在该回帖网络中并不稳定。但是，除 1 月和 2 月外，其他月份的系数稳定维持在 1.2 以上，属较高水平。因此，虽然这种特征并不稳定，但论坛管理者仍然需要想出办法，如大型炒股经验分享活动、有奖互动等方式，来促进论坛成员间的交流。

(3) H_{3a} 的参数是 2.171，显著且通过稳定性检验，H_{3a} 得到验证。说明粉丝数多的用户确实能得到更多人的回帖。EXP(2.171) ≈ 8.75，说明与其他用户相比，粉丝数多的用户能够得到将近 9 倍回帖量。高粉丝数的用户在论坛中会受到更多人的关注，他们确实会成为社区论坛中的“中心点”。因此，这些用户的言论会对其他用户产生重要的影响。从时间趋势上来看，6 个月的系数虽有波动但均大于 2，说明粉丝数对一个人是否能得到回帖起关键性作用。

(4) H_{3b} 的参数是 0.947，显著且通过稳健性检验。EXP(0.947) ≈ 2.55，说明影响力高的用户能得到更多的回帖，该假设得到验证。但是，与粉丝数相比，影响力对回帖行为的影响稍弱。从时间趋势上观察亦是如此。除 12 月(M8)外，其余月份(M0、M6、M7、M9、M10)的系数均在 1 以下，且还有两个月(M7、M9)降至 0.4 左右。说明影响力并没有预期的那么重要，粉丝数才是该论坛社

会资本中最重要的一环。

(5) H_4 虽然结果显著,但由于其参数是 -0.060 ,说明用户关注数的提高并不会为其带来更多他人的关注。稳健性检验中, $M3$ 的系数符号虽然发生变化,但仅集中在前10%的用户群体中。考虑到 $M1$ 、 $M2$ 的结论与主模型一致,故用户想要通过增加关注数来提高被关注度并非一条简单易行的办法,主模型结论依然可信。 $M4$ 、 $M5$ 、 $M6$ 和 $M8$ 的回归结果虽然不显著,但其表达的含义是实际网络与随机网络相比无显著差别,关注数高的用户并不会获得更多的回帖,拒绝假设,与基准模型的结论并不冲突,因此 H_4 没有得到验证。造成这一现象的原因:在微博社区中,用户可以通过“转发”的方式快速提高自己的发帖量,进而得到更多人的评论。同时,微博官方也会通过“个性推荐”等手段增加用户间的互动行为^[49]。但该股票社区在这方面做得还不够完善,因此与微博社区不同,在该社会网络中关注数并不能促进用户间回帖关系的形成。从时间趋势上看,该变量的系数不论显著与否,基本为负值且稳定在 -0.1 左右。进一步证明了该假设并不适用于股票社区。

(6) H_5 的参数为 0.578 ,显著且通过稳健性检验, H_5 得到验证。 $EXP(0.578) \approx 1.76$,说明与其他用户相比,影响力高的用户更愿意回复他人的帖子。从时间趋势上来看,该变量系数在 0.5 左右徘徊,最高达到了 0.612 。虽然有些许波动,但总体上较为稳定。

(7) H_6 的参数为 0.646 ,显著且通过稳健性检验, H_6 得到验证。 $EXP(0.646) \approx 1.89$,说明普通用户的确更关心认证用户所发的帖子,也更愿意与这些认证用户进行互动。从时间趋势上来看,该假设的波动较大,11月($M9$)系数只有 0.499 ,12月($M8$)突然增至 2.299 。但考虑到社会网络特征是动态变化的,且该系数的均值为 1.01 ,得到官方认证依旧是提高被关注度的有效途径。

5 管理启示

通过分析在线股票论坛用户发帖行为和回帖关系网络特征,能够为在线股票社区的管理者提供相关实践启示,进而提升在线股票社区的服务质量。以下是具体建议:

(1) 从用户的发帖规律可以看出,用户使用论坛的时段主要集中在工作日期间,周末或节假日论坛的访问量大幅下降。为了更好地提高用户对于网站的黏着度,论坛管理者可以在周末或节假日举办一些经验分享活动、有奖征集活动内容,提高周末或节假日用户的访问量。

(2) 从用户的回帖特征来看,大多数用户是信息的“接收者”而非“发送者”。因此,提高用户间的活跃度、鼓励用户积极发回帖也是论坛管理者需要思考的问题。目前,用户发回帖量的多寡只与用户影响力相关,而用户影响力对普通用户并没有太多吸引力。因此,管理者可以考虑扩大发回帖量对用户的影响,如提高发回帖获得积分的数量、所得积分能够兑换一些在线服务或产品等。通过不同的方法提高用户的发回帖积极性。

(3) 从回帖网络的网络结构特征可以看出,该论坛存在较为明显的互惠性和一定程度上的小团体特征。论坛管理者应当采取相应措施进一步鼓励用户间的互动行为,如前文所提的部分举措。用户间的互动频率变高不仅能加快用户间信息的传播速度和效率,还能提高网站帖子的质量。从长远来看,是维持网站稳定繁荣的重要因素。

(4) 从回帖网络的用户属性特征来看,高粉丝数、认证用户的帖子往往是论坛热门帖的来源。因此,论坛管理者应当重视对这部分用户的管理。例如,对这部分用户的帖子进行“加精”、对帖子进行置顶操作、向其他用户推荐这些帖子等措施。通过这些举措以达到提高用户参与度、促进用户互动的目的。

6 结束语

本文通过搜集“股吧”论坛的“股市实战”版块 2017 年 10 月到 2018 年 3 月的帖子及用户数据,分析了该版块内用户的发帖行为和回帖关系网络特征。通过分析发帖行为的特征发现该版块内的用户通常在工作日内进行大量发帖,股市上午开盘前、下午收盘后及晚 8 点是用户发帖的高峰时段;多数用户是论坛信息的“接收者”而非“发送者”,他们更倾向于查看帖子而不进行评论;发帖用户通常发表简短评述类的帖子,字数较多、内容较丰富的帖子不多。进一步使用 ERGM 对回帖网络关系进行分析以验证提出的 7 个假设。研究发现该版块具有较高的互惠性,回帖网络中用户互动频繁;该网络一定程度上存在小团体特征;粉丝数高、影响力高的用户更容易得到他人的回帖,且粉丝数起主导作用,社会资本能有效提高一个人的被关注度;股票论坛与微博社区不同,关注数高的用户并不会得到更多人的关注;影响力高的用户更愿意对他人进行回帖,说明拥有高势位优势的用户会向低势位的用户传递更多的信息;此外,得到官方认证的用户也能获得更多人的关注。

本文通过采用多个理论,如互惠性理论、社交资本理论等,丰富了股票在线社区的相关研究,从社会网络视角进一步探索了股票社区对股票投资者的影响。但是,本文仅针对“股市实战”版块进行研究,没有考虑与其他版块的联系。因此,可以从多个网络间的交互作用着手,进行更进一步的分析;受限于 ERGM 的特点,本文对用户属性进行了人为处理,导致一定程度的信息损失,在后续研究中可以使用其他方法进行更为精准的分析;此外,本文主要针对回帖关系网络特征揭示何类用户发的帖子更易受到他人的关注,没有从文本情感的角度分析其对回帖行为的影响,这些有待后续进一步的验证。

参 考 文 献

- [1] 中国互联网络信息中心. CNIC 第 41 次中国互联网发展状况统计报告[EB/OL]. http://www.cac.gov.cn/2018-01/31/c_1122347026.htm, 2018-01-31.
- [2] 周军杰, 左美云. 线上线下互动、群体分化与知识共享的关系研究——基于虚拟社区的实证分析[J]. 中国管理科学, 2012, 20(6): 185-192.
- [3] 熊熊, 罗春春, 张焯. 股吧和交易: 股吧中的信息内容研究[J]. 系统科学与数学, 2017, 37(12): 2359-2374.
- [4] 金秀, 姜尚伟, 苑莹. 基于股吧信息的投资者情绪与极端收益的可预测性研究[J]. 管理评论, 2018, 30(7): 16-25.
- [5] 金雪军. 股吧发帖能否影响股价?[J]. 清华金融评论, 2017, (3): 97-98.
- [6] 叶德磊, 姚占雷, 刘小舟. 公司新闻、投资者关注与股价运行——来自股吧的证据[J]. 华东师范大学学报(哲学社会科学版), 2017, 49(6): 136-143.
- [7] 叶作亮, 王雪乔, 王仙玲, 等. 一类 BBS 网络统计特性实证分析[J]. 复杂系统与复杂性科学, 2010, 7(1): 52-58.
- [8] 刘鹏. 网络用户行为分析的若干问题研究[D]. 北京邮电大学博士学位论文, 2010.
- [9] Zhou J, Yu Y, Zuo M, et al. How fundamental and supplemental interactions affect users' knowledge sharing in virtual communities? A social cognitive perspective [J]. Internet Research, 2014, 24(5): 566-586.
- [10] Cetto A, Klier M, Richter A, et al. "Thanks for sharing" – Identifying users' roles based on knowledge contribution in enterprise social networks[J]. Computer Networks, 2018, 135(22): 275-288.
- [11] 秦敏, 梁溯. 在线产品创新社区用户识别机制与用户贡献行为研究: 基于亲社会行为理论视角[J]. 南开管理评论, 2017, 20(3): 28-39.
- [12] 赵蓉英, 曾宪琴. 微博信息传播的影响因素研究分析[J]. 情报理论与实践, 2014, 37(3): 58-63.
- [13] Mislove A, Marcon M, Gummadi K P, et al. Measurement and analysis of online social networks[C]//ACM SIGCOMM Conference on Internet Measurement. San Diego: DBLP, 2007: 29-42.
- [14] 杨化龙. 在线医患交互过程中患者生成信息的作用机理研究[D]. 哈尔滨工业大学硕士学位论文, 2017.

- [15] Wu J, Fan S, Zhao J L. Community engagement and online word of mouth: an empirical investigation[J]. *Information & Management*, 2018, 55 (2) : 258-270.
- [16] Wellman B, Quan-Haase A, Boase J, et al. The social affordances of the internet for networked individualism[J]. *Journal of Computer-Mediated Communication*, 2006, 8 (3) : 1-18.
- [17] Liu P, Xia X, Li A. Tweeting the financial market: media effect in the era of big data[J]. *Pacific-Basin Finance Journal*, 2018, 51: 267-290.
- [18] Oliveira N, Cortez P, Areal N. The impact of microblogging data for stock market prediction: using Twitter to predict returns, volatility, trading volume and survey sentiment indices[J]. *Expert Systems with Applications*, 2017, 73: 125-144.
- [19] Wu C H, Lin C J. The impact of media coverage on investor trading behavior and stock returns[J]. *Pacific-Basin Finance Journal*, 2017, 43: 151-172.
- [20] Ho C S, Damien P, Gu B, et al. The time-varying nature of social media sentiments in modeling stock returns[J]. *Decision Support Systems*, 2017, 101: 69-81.
- [21] Checkley M S, Higón D A, Alles H. The hasty wisdom of the mob: how market sentiment predicts stock market behavior[J]. *Expert Systems with Applications*, 2017, 77 (1) : 256-263.
- [22] 缪杰. 基于百度指数的投资者关注度对于股票市场表现的影响[D]. 厦门大学硕士学位论文, 2014.
- [23] Himelboim I. Reply distribution in online discussions: a comparative network analysis of political and health newsgroups[J]. *Journal of Computer-Mediated Communication*, 2008, 14 (1) : 156-177.
- [24] Kimmerle J, Bientzle M, Cress U. “Scientific evidence is very important for me” : the impact of behavioral intention and the wording of user inquiries on replies and recommendations in a health-related online forum[J]. *Computers in Human Behavior*, 2017, 73: 320-327.
- [25] Meng J, Chung M, Cox J. Linking network structure to support messages: effects of brokerage and closure on received social support [J]. *Journal of Communication*, 2016, 66 (6) : 986-1006.
- [26] Wu M, Li H, Zhang K, et al. An evolutionary model of reply networks on bulletin board system[C]. *IEEE 2011 International Conference on Information Technology, Computer Engineering and Management Sciences (ICM)*, 2011.
- [27] O'Dea B, Melinda R, Mark E, et al. The rate of reply and nature of responses to suicide-related posts on Twitter[J]. *Internet Interventions*, 2018, 13: 105-107.
- [28] Yu J, Hu Y, Yu M, et al. Analyzing netizens' view and reply behaviors on the forum[J]. *Physica A: Statistical Mechanics and Its Applications*, 2010, 389 (16) : 3267-3273.
- [29] Box-Steffensmeier J M, Campbell B W, Christenson D P, et al. Role analysis using the ego-ERGM: a look at environmental interest group coalitions[J]. *Social Networks*, 2018, 52: 213-227.
- [30] Windzio M. The network of global migration 1990-2013: using ERGMs to test theories of migration between countries[J]. *Social Networks*, 2018, 53: 20-29.
- [31] Hossain L, Hamra J, Wigand R T, et al. Exponential random graph modeling of emergency collaboration networks[J]. *Knowledge-Based Systems*, 2015, 77: 68-79.
- [32] Wimmer A, Kewis L. Beyond and below racial homophily ERG models of a friendship network documented on Facebook[J]. *American Journal of Sociology*, 2010, 116 (2) : 583-642.
- [33] Shen C, Monge P. Who connects with whom? A social network analysis of an online open source software community[J]. *First Monday*, 2011, 16 (6) : 1-13.
- [34] Stivala A D, Koskinen J H, Rolls D A, et al. Snowball sampling for estimating exponential random graph models for large networks[J]. *Social Networks*, 2016, 47: 167-188.
- [35] An W. Fitting ERGMs on big networks[J]. *Social Science Research*, 2016, 59: 107-119.
- [36] 鲁谢尔 D, 科斯基宁 J, 罗宾斯 G. 社会网络指数随机图模型: 理论、方法与应用[M]. 杜海峰, 任义科, 杜巍, 等译. 北京: 社会科学文献出版社, 2016.
- [37] 陈向东, 徐之路. 在线知识交流回帖动因的个案研究[J]. *图书情报工作*, 2010, 54 (10) : 30-34.
- [38] 刘雨农, 刘敏榕. 虚拟知识社区的社会网络结构及影响因素——以知乎网为例[J]. *图书情报工作*, 2018, 62 (4) : 89-96.
- [39] 林南, 俞弘强. 社会网络与地位获得[J]. *马克思主义与现实*, 2003, (2) : 46-59.
- [40] Watts D J, Dodds P. Influentials, networks, and public opinion formation[J]. *Journal of Consumer Research*, 2007, 34 (4) : 441-458.
- [41] 何黎, 何跃, 霍叶青. 微博用户特征分析和核心用户挖掘[J]. *情报理论与实践*, 2011, 34 (11) : 121-125.
- [42] 刘行军, 王伟军. 微博用户类型及行为特征的实证分析——基于信息传播行为视角[J]. *情报科学*, 2014, 32 (9) :

- 130-136.
- [43] Lavie D, Drori I. Collaborating for knowledge creation and application: the case of nanotechnology research programs[J]. *Organization Science*, 2016, 23 (3): 704-724.
- [44] 刘璇, 汪林威, 李嘉, 等. 在线健康社区中用户回帖行为影响机理研究[J]. *管理科学*, 2017, 30 (1): 62-72.
- [45] Geyer C, Thompson E A. Constrained Monte Carlo maximum likelihood for dependent data[J]. *Journal of the Royal Statistical Society*, 1992, 54 (3): 657-699.
- [46] Wu B, Jiang S, Chen H. The impact of individual attributes on knowledge diffusion in web forums[J]. *Quality & Quantity*, 2015, 49 (6): 2221-2236.
- [47] 吴硕默. 社会网络视角下区域环境合作治理演化研究——以长三角城市群为例[D]. 电子科技大学硕士学位论文, 2017.
- [48] Robins G, Snijders T, Wang P, et al. Recent developments in exponential random graph (p) models for social networks[J]. *Social Networks*, 2007, 29 (2): 192-215.
- [49] 闫强, 吴联仁, 郑兰. 微博社区中用户行为特征及其机理研究[J]. *电子科技大学学报*, 2013, 42 (3): 328-333.

Research on User Participative Behaviors in Online Stock Forum: Empirical Research from Guba Forum

GAO Jiaojiao, ZHANG Renjie

(School of Economics and Management, SWJTU, Chengdu 610031, China)

Abstract Online stock forum is an important channel for stock investors to share information; they spread information through posting and replying behaviors. The information dissemination has major influence on investor sentiment. This article analyzed the feature of forum users' posting and replying behaviors based on "stock actual combat" part in "Guba" community. The result reveals that most users prefer to post posts when stock market is opening and most of them are "receiver" rather than "sender"; we found that mutual feature is positive in the forum by employing exponential random graph model; there is a certain degree of small group characteristics in the social network; users who have more social capitals are more likely to post and reply and their opinions have vital effect on others.

Keywords Online stock forum, Social network, Exponential random graph model, Posting & replying behavior

作者简介

高娇娇(1988—), 女, 西南交通大学经济管理学院 2017 级博士研究生, 研究方向为在线社区用户行为分析。E-mail: 279067287@qq.com。

张仁杰(1994—), 男, 西南交通大学经济管理学院 2017 级硕士研究生, 研究方向为社会网络、数据挖掘。E-mail: zrz9497@163.com。