

社会化问答社区中社交关系的成因及作用 ——来自知乎的实证研究*

蒋逸尘 金悦 黄京华
(清华大学 经济管理学院, 北京 100084)

摘要 随着社会化媒体的发展, 加入社交关系的社会化问答社区开始迅速兴起。本文基于关系构建理论、知识贡献模型、UGC 质量判断相关研究, 探索社会化问答社区中用户如何增加自己的关注者数以及关注者数的增加如何影响用户的参与体验。研究结果显示, 用户的参与度和权威性是其吸引关注者的关键因素, 虽然拥有大量关注者并不会使用该用户提出的问题更易得到解决, 但却会使该用户提供的答案更易受到认可。

关键词 社会化问答社区, 社交关系, 知识贡献, UGC 质量判断

中图分类号 C931.6

1 引言

互联网改变了人们求知、交流的方式, 尤其是 Web 2.0 用户交互时代的到来, 开启了互联网产品 UGC (user generated content, 即用户生成内容) 的全新模式。而得益于 Web 2.0 技术的广泛应用与不断成熟, 自 2002 年以来, 在线问答社区不断发展成熟。2009 年, Quora 在美国推出, 开启了在线问答的一次革命。Quora 是一个社会化问答社区, 在 Quora 上用户可以互相关注, 形成一个类似微博的社交网络, 用户登录社区后可以看到其关注的用户在社区中的参与行为。利用这种理念重新组织问答内容, Quora 迅速积累了大量用户。伴随着 Quora 的成功, 国内出现了一批 Quora 模式的网站, 其中最成功的便是知乎。知乎自 2013 年开放注册以来, 两年时间, 注册用户数增长了 40 倍。知乎与传统在线问答社区最根本的区别就在于加入了社交机制, 也就是用户之间可以互相关注。对此, 本文提出以下问题: 一是哪些因素会影响用户之间社交关系 (即关注关系) 的建立? 二是社交关系的建立对社区有何影响? 具体来说, 用户拥有更多关注者会有哪些好处? 如这类用户提出的问题更易得到解决, 以及他们提供的答案更易受到认可?

目前学术界对社会化问答社区的研究还十分缺乏。Paul 等通过分析 60 个问题话题和 3 917 个用户的数据, 描述了 Quora 上的用户活动, 并通过与 10 位 Quora 用户的深度访谈对权威性、声誉和质量等问题进行了详尽阐述, 结果显示, 用户通过其他用户的过往贡献判断他们的声誉, 社会化投票可以帮助用户识别和推送优质内容, 但易于发生偏好依附^[1]。Wang 等刻画了 Quora 中话题和用户、用户和用户、问题和问题等三种不同的连接网络的特征, 指出贡献更多更好的答案的用户往往拥有更多的关注

* 基金项目: 国家自然科学基金项目 (71272028; 71490721)。

通信作者: 黄京华, 清华大学经济管理学院, 教授, E-mail: huangjh@sem.tsinghua.edu.cn。

者, 这些连接度高的用户可以获得优势, 会有更多的朋友(关注者)来回答他们的问题和赞同他们的答案^[2]。然而, 上述这些研究虽然解释了社交关系的影响因素及其对回答行为和投票行为的影响, 但是只有定性的特征描述, 没有进行定量的实证分析, 且缺乏坚实的理论基础。

针对目前研究存在的问题, 本文将在分析三方面文献的基础上, 提出研究假设, 并收集知乎的实际数据检验假设。

2 文献综述

2.1 社交关系构建影响因素相关研究

为了解哪些因素会影响社会化问答社区中用户社交关系的建立, 需要借鉴在线社区中社交关系构建影响因素的相关研究。在线社区主要分为两种不同类型: 一种是基于社会关系或情感支持的社区, 如微博、人人等社交型社区; 另一种是基于共同兴趣或信息交换的社区, 如百度知道、知乎等问答型社区。传统的社会网络分析探索了一般社会网络的关系构建机制, 许多学者对在线社区的社会网络分析主要以社交型社区为主。Granovetter 提出的关于关系构建的理论讨论了社会影响、互惠性和选择影响等在一般在线社区发展过程中的驱动作用。社会影响是指人们通过信息传递而逐渐与他人建立关系, 分为规范性社会影响和信息性社会影响; 互惠性是指人们的行为是基于他人对自己的行为所作出的反馈; 选择影响是指人们有与那些跟自己具有共同特征的人聚在一起的倾向^[3]。而在问答型社区中加入社交机制时, 必须考虑到不同性质的社区在关系构建机制上具有本质的不同。例如, 在社区关系来源方面, Granovetter 的研究表明, 在社交型社区中, 许多网络节点的加入和网络连接的形成来自成员之间社会关系的嵌入, 这种社会关系在成员进入在线社区前或许就已形成; 而在问答型社区中, 则来自成员之间的问答关系。又如, 在社区发展动力方面, 社交型社区以规范性社会影响为主, 而针对问答型社区这种用户之间以完全陌生关系为主的社区, Goette 等的研究表明, 成员社交关系的建立受到的信息性社会影响起主要作用^[4]。信息性社会影响是指当个体向其所在社区中的其他成员获取信息时受到的该成员个人信息的影响。Paul 等在对 Quora 用户的访谈研究中发现, 用户会通过访问其他用户的主页来查看他们的历史行为等信息, 并且更愿意关注提供高质量答案的用户^[1]。Wang 等在对 Quora 网络结构的研究中也发现, 贡献更多更好答案的用户往往会拥有更多关注者^[2]。

综上所述, 基于关系构建理论, 社会影响是虚拟社区中社交关系建立的一个重要影响因素。而以陌生关系为主的社区, 主要受到信息性社会影响。在社会化问答社区中, 用户之间主要是陌生关系, 并以信息交流为主, 因此, 影响社交关系建立的因素主要是信息性社会影响, 即用户信息。然而用户信息包括哪些方面, 现有研究尚未给出明确和统一的结论。

2.2 知识共享动机相关研究

为了回答本文提出的问题, 即拥有更多关注者的用户提出的问题是否更易得到解决, 需要借鉴知识共享动机的相关研究。在信息系统领域, 对知识共享动机的研究已相对成熟。Wasko 和 Faraj 构建的知识贡献模型应用社会资本理论, 将关系资本、结构资本、认知资本等社会资本与个人动机一起, 作为知识共享的动机, 经过实证检验, 发现人们贡献知识是因为感知到他们的声誉会被提高, 并且一定程度上也是因为帮助他人是愉悦的^[5]。Chiu 等则将知识共享进一步细分为数量和质量两个维度, 研究社会资本和结果预期对它们的影响, 结果显示, 与社区相关的结果预期对知识共享的数量和质量两个维度都有显著作用, 而社会互动关系、互惠和认同等社会资本则会显著提高知识共享的数量^[6]。Paul

等在对 Quora 的研究中发现, 用户会通过回答问题来建立声誉, 特别是当他关注了某个用户并且想要给其留下印象时, 会努力提供高质量的答案^[1]。Wang 等在对 Quora 的研究中也发现, 拥有更多关注者的用户会有更多的朋友(关注者)回答他提出的问题^[2]。

综上所述, 基于知识贡献模型, 社会资本中的社会互动关系是在线社区中用户贡献知识的重要影响因素, 但在社会化问答社区中, 社会互动关系的显著影响是否依然存在需要作进一步的实证研究。

2.3 答案质量影响因素相关研究

为了回答本文提出的问题, 即拥有更多关注者的用户提供的答案是否更易受到认可, 需要了解哪些因素会影响用户对答案质量的判断, 因此, 本文借鉴了以下两个方面的相关研究。首先, 我们借鉴 UGC 质量判断的相关研究, 因为社会化问答社区中的答案也是一种 UGC。Lin 等针对在线社会网络进行了研究, 提出了一个通用框架, 该框架包括反馈特征(回复频率、回复相似性等)、内容特征(内容长度、词汇复杂度等)和作者特征(权威性、等级等), 这三个方面共同影响用户对 UGC 质量的判断^[7]。Agichtein 等针对社会化媒体, 提出了利用社区反馈自动识别高质量 UGC 的方法, 该方法采用的特征指标包括用户关系和用户行为历史数据等^[8]。用户对 UGC 的投票行为反映了用户对 UGC 质量的判断。Liu 等针对产品评论质量判断的研究发现, 以往的研究中存在三类偏差: 不平衡投票偏差、赢家循环偏差和早起者偏差, 消除这三类偏差后, 再利用问答社区中的统计数据, 如用户获得积分、答案获得票数和点击次数等, 来评估答案质量^[9]。Mudambi 和 Schuff 在对亚马逊上用户提供的在线评论的研究中, 用字数来衡量评论深度^[10]。其次, 关于答案质量判断的相关研究主要有以下几个。Jurczyk 和 Agichtein 针对传统的在线问答社区, 研究了如何利用用户权威性判断答案质量, 先在用户之间通过回答问题建立起一个关联, 接着运用链接分析来计算用户的权威性^[11]。Paul 等在对 Quora 的研究中也发现了早起者现象, 发布较早的答案获得的投票可能很多且排名很高, 发布较晚的答案获得的投票可能较少而排名靠后, 就很难被投票者看到, 因此, 发布较晚的答案的后续投票会更少。此外, 投票具有社交功能, 用户会将投票用作一种引起知名用户注意的社交信号^[1]。Wang 等在对 Quora 的研究中也发现, 拥有更多关注者的用户会有更多的朋友(关注者)赞同他提供的答案, 即该答案会有更多的投票数^[2]。

综上所述, 已有研究提出了影响 UGC 质量判断的多个因素, 包括用户关系、回复频率、信息量、发布时间等, 但这些因素是否适合社会化问答社区还需要实证检验。

3 研究假设

3.1 用户信息对社交关系的影响

在社会化问答社区中, 社交关系是指用户之间的关注关系, 可用关注者数表示^[1]。

如前面文献所述, 用户信息会影响社交关系^[1~4]。Zhou 等在对 Yahoo! Answers 的研究中将用户信息具体分为三类, 包括参与度相关的信息、权威性相关的信息和等级相关的信息^[12]。由于知乎社区中并没有等级制度, 因而本文只考虑用户的参与度信息和权威性信息。

首先, 在社会化问答社区中, 用户之间是以完全陌生关系为主, 根据社区社会关系构建的研究成果, 在这种情境下, 由于用户参与社区的主要目的是信息交流, 因此, 在考虑是否与社区中的其他用户建立社交关系时, 为了确保信息来源的真实性、可靠性, 会更注重考查该用户的个人信息, 关系的构建更易受到用户所传递的个人信息主导。其次, 用户会通过访问其他用户的主页来查看他们的历史

行为，参与度高的用户会拥有更丰富的历史行为，可以贡献更多的信息，权威性高的用户更容易取得信任，可以贡献更有价值的信息，因此，人们往往会关注参与度高且提供高质量答案的用户^[1]，贡献更多更好答案的用户往往会拥有更多关注者^[2]。综上，本文提出以下研究假设：

H₁：用户的参与度越高，关注者数越多。

H₂：用户的权威性越高，关注者数越多。

3.2 社交关系对问题得到解决的影响

回答问题是在知乎社区中的主要参与行为，已有的很多研究认为，在线问答社区中回答者提供答案的行为是一种问题解决型的知识共享行为。首先，根据知识共享动机的研究成果，由于社交关系的建立会在社区参与中提高用户间的互动频率，这种互动是人们积累社会资本的一个过程，因此，社会互动关系会显著提高知识共享的数量。其次，由于用户可以通过回答问题来建立声誉，特别是当他关注了某个知名用户时，努力提供高质量的答案可以给其留下印象^[1]，因此，拥有更多关注者的用户会获得优势，即会有更多的朋友（关注者）来回答他们提出的问题^[2]。综上，本文提出以下研究假设：

H₃：用户关注者数越多，提出的问题更易得到解决。

3.3 社交关系对答案受到认可的影响

阅读答案并对答案进行“赞同”或“反对”是用户在知乎社区中的另一主要参与行为，当用户觉得答案质量好的时候，就会“赞同”该答案，而当用户觉得答案质量不好时，就会“反对”该答案。首先，根据答案质量评估的研究成果，相比于与自己没有任何关联的人，人们往往会更相信与自己存在某种社交关系的人，因此，用户关系会影响用户对答案质量的评估，并通过用户对答案的投票情况体现。其次，由于投票具有社交功能，可以起到向内容提供者示好的作用，因此，用户会将投票用作一种引起目标用户注意的社交信号^[1]，拥有更多关注者的用户会获得优势，即会有更多的关注者来赞同他们提供的答案^[2]。综上，本文提出以下研究假设。

H₄：用户关注者数越多，提供的答案更易受到认可。

综上所述，本文的研究框架见图 1。

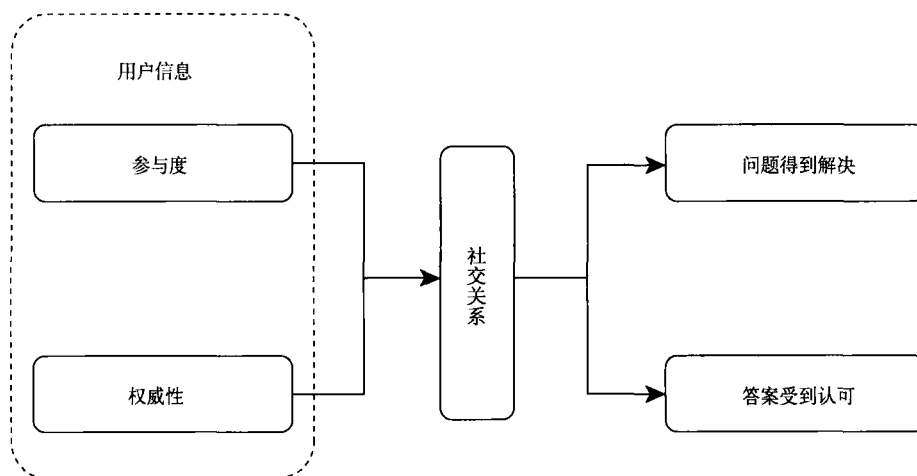


图 1 研究框架

4 研究方法 with 数据收集

4.1 变量的操作化与研究模型

关于假设 1 和假设 2，用户信息的测量如下。

用户的参与度信息方面，Zhou 等主要根据用户所得积分来衡量用户的参与度^[12]，而知乎社区并没有积分制度，考虑到在传统在线问答社区中，用户主要通过“提问”“回答”等活动获得积分，且知乎社区中加入了“公共编辑”功能，即社区成员可以共同编辑一个问题，因此，本文以用户提问数、回答数、公共编辑数来衡量用户的参与度。

用户的权威性信息方面，Zhou 等主要根据用户所贡献的答案成为最佳答案的比例来衡量用户的权威性^[12]，而知乎社区没有最佳答案指标，用户权威性主要体现在以下两个方面：一是若某个回答者提供的答案对其他用户真的很有帮助，便会获得其他用户的“感谢”；二是知乎社区根据一套独有的内部算法来计算每个用户的 PR (person rank, 即个人等级) 值，如果某个用户的 PR 值达到一定水平，知乎社区便会赋予该用户开通“专栏文章”的权利。因此，本文以用户获得感谢数、专栏文章数来衡量用户的权威性。

假设 1 和假设 2 的检验模型如下所示：

$$Ufollower_i = \beta_0 + \beta_1 Question_i + \beta_2 Answer_i + \beta_3 Edit_i + \beta_4 Thank_i + \beta_5 Article_i + \varepsilon_i \quad (1)$$

其中，Ufollower_{*i*} 为用户 *i* 的关注者数；Question_{*i*} 为用户 *i* 的提问数；Answer_{*i*} 为用户 *i* 的回答数；Edit_{*i*} 为用户 *i* 的公共编辑数；Thank_{*i*} 为用户 *i* 获得的感谢数；Article_{*i*} 为用户 *i* 的专栏文章数。

根据文献综述，假设 3 中问题获得回答数还会受到问题关注人数、发布时间、所属话题的影响，其中问题所属话题被分为六类：职业发展、体育运动、自然科学、艺术人文、日常生活、游戏娱乐，因此，关于假设 3 的检验模型如下所示：

$$GetAnswer_i = \beta_0 + \beta_1 Ufollower_i + \beta_2 Qfollower_i + \beta_3 Qtime_i + \beta_4 Qtopic_i + \varepsilon_i \quad (2)$$

其中，GetAnswer_{*i*} 为问题 *i* 获得的回答数；Ufollower_{*i*} 为问题 *i* 的提问者的关注者数；Qfollower_{*i*} 为问题 *i* 的关注人数；Qtime_{*i*} 为问题 *i* 的发布时间（分别计算每个问题与发布最早的问题发布时间的时差，单位为天）；Qtopic_{*i*} 为问题 *i* 所属的话题。

根据文献综述，假设 4 中答案获得赞同数还会受到答案评论数、发布时间、信息量的影响，其中答案信息量需要利用中国科学院分词系统的 ICTCLAS (institute of computing technology Chinese lexical analysis system, 即计算技术研究所汉语词法分析系统) 对答案进行分词处理，去除答案内容的标点符号及停用词得到概念词，概念词的数量即为答案信息量的值，因此，关于假设 4 的检验模型如下所示：

$$GetVote_i = \beta_0 + \beta_1 Ufollower_i + \beta_2 Acomment_i + \beta_3 Atime_i + \beta_4 Ainformation_i + \varepsilon_i \quad (3)$$

其中，GetVote_{*i*} 为答案 *i* 获得的赞同数；Ufollower_{*i*} 为答案 *i* 的回答者的关注者数；Acomment_{*i*} 为答案 *i* 的评论数；Atime_{*i*} 为答案 *i* 的发布时间（分别计算每个答案与其对应问题发布时间的时差，单位为天）；Ainformation_{*i*} 为答案 *i* 的信息量。

4.2 数据收集

为了检验以上假设，本文的分析主要基于从知乎网站上抓取下来的一个数据集，包括关于问题、答案和用户的多组数据，具体流程如下：进入知乎的话题广场，共有 34 个话题，先从每个话题下随机

选取 10 个子话题，然后从每个子话题下随机选取 1 个问题，随机选取该问题的 2 个相关问题，对于每个相关问题，再随机选取该问题的 2 个相关问题，接着选取每个问题的提问者，并从每个问题下随机选取 10 个回答者，最后对所有问题和用户进行去重处理，得到 1 905 个问题，111 200 个答案，14 671 个用户。

对于每个问题，抓取该问题的主页链接地址、关注人数、发布时间、提问者以及回答数。对于问题的每个答案，抓取该答案的回答者、赞同数、评论数、发布时间以及答案文本。对于问题的提问者和回答者，抓取该用户的主页链接地址、提问数、回答数、公共编辑数、获得感谢数、专栏文章数以及关注者人数。

基于 Python 程序抓取的知乎社区基本信息见表 1。

表 1 知乎社区基本信息

数据名称及类型		数据解释	数据格式
问题信息	问题主页	用来进入该问题主页的链接地址	链接地址
	关注人数	关注该问题的人数	数值型
	发布时间	用户提出该问题的时间	日期型
	提问者	提出该问题的用户	文本型
答案信息	回答数	该问题获得的回答数	数值型
	回答者	提供该答案的用户	文本型
	赞同数	该答案获得的赞同数	数值型
	评论数	该答案获得的评论数	数值型
	发布时间	用户提供该答案的时间	日期型
用户信息	答案文本	该答案的文本内容	文本型
	用户主页	用来进入该用户主页的链接地址	链接地址
	提问数	该用户提出的问题数	数值型
	回答数	该用户提供的答案数	数值型
	公共编辑数	该用户参与编辑的问题数	数值型
	获得感谢数	该用户获得的感谢数	数值型
	专栏文章数	该用户撰写的专栏文章数	数值型
关注者人数	关注该用户的人数	数值型	

5 描述性统计与假设检验

5.1 描述性统计

表 2 列出了 5.2 中实证检验所使用的变量的描述性统计。

表 2 变量的描述性统计

变量	最大值	最小值	均值	标准差
Ufollower /百人	8 622.83	0	25.135	23 097.130
Question /个	3 055	0	12.792	68.921
Answer /个	10 132	0	90.338	256.123
Edit /个	141 022	0	142.667	1 893.454
Thank /个	174 682	0	399.700	2 966.605
Article /篇	924	0	1.124	12.894

续表

变量	最大值	最小值	均值	标准差
GetAnswer /个	2 344	0	58.418	153.260
Qfollower /百人	1 455.41	0.01	16.917	6 191.815
Qtime /天	32	0	28.608	360.414
GetVote /个	35 419	0	42.174	480.567
Acomment /个	2 374	0	5.029	32.357
Atime /天	64	0	9.867	64.497
Ainformation /个	1 240	1	23.491	67.321

5.2 假设检验

我们先进行了自变量的相关性检验，结果表明，检验系数最大为 0.473，所以不存在相关性。

关于假设 1 和假设 2 的检验结果见表 3，关于假设 3 的检验结果见表 4，关于假设 4 的检验结果见表 5。

表 3 关注者数的影响因素的系数检验结果 ($n=14\ 616$)

变量	系数
Constant	- 317.639 0**
Question	45.830 2***
Answer	1.296 0**
Edit	0.538 8***
Thank	4.904 6***
Article	80.569 9***
Adjusted R^2	0.475 8
F	2 654.584 6***

***、**分别表示在 1%、5%的水平上显著

表 4 问题得到解决的影响因素的系数检验结果 ($n=1\ 687$)

变量	系数
Constant	- 1 973.687 7***
Ufollower	n.s.
Qfollower	1.487 4***
Qtime	0.048 3***
Qtopic1	—
Qtopic2	n.s.
Qtopic3	n.s.
Qtopic4	28.700***
Qtopic5	n.s.
Qtopic6	40.725***
Adjusted R^2	0.411 3
F	148.588 1***

***表示在 1%的水平上显著

表 5 答案受到认可的影响因素的系数检验结果 ($n=30\ 160$)

变量	系数
Constant	-19.263 8***
Ufollower	0.029 9***
Acomment	11.625 5***
Atime	n.s.
Ainformation	n.s.
Adjusted R^2	0.623 9
F	12 508.314 0***

***表示在 1%的水平上显著

检验结果显示,假设 1、假设 2 得到了验证。一方面,人们加入知乎这种以信息交流为主的社区,最主要的目的就是获取信息,参与度高的用户,也就是活跃用户,能够为关注者提供更多的信息,因而能够吸引更多的关注者;另一方面,信息质量也是人们选择关注者时考量的一个要素,那些权威性更高的用户,他们提供的信息更专业,更具有说服力,对其他用户的帮助更大,因而也能吸引更多的关注者。可见,除了李开复、徐小平、周源等少数即使很少出现也会拥有很多关注者的名人之外,知乎的大多数用户还是通过贡献大量高质量的答案来吸引他们的关注者的。

假设 3 未能得到验证,实际上这也是有可能的,在知乎这种用户之间以陌生关系为主的社区,人们关注其他用户,最主要的目的是从该用户那里寻求有帮助的问题答案,而不是去帮助该用户解决他所提出的问题;另外,由于能力或精力所限,人们也很有可能选择不去回答该用户的问题,即便是出于社交目的,他们也可以选择参与成本更低的方式,如给该用户提供的答案投票等。此外,从表 4 的结果可以发现,问题关注人数对问题获得回答数有显著影响,因为问题关注者越多,说明该问题的价值越高,越能吸引更多的用户予以解答;发布时间对问题获得回答数有显著影响,因为随着时间的累积,不断有新的回答者加入,问题获得的回答数也就会越来越多;以“职业发展”为基准,“艺术人文”和“游戏娱乐”话题下的问题获得回答数与其存在显著差异,这是因为有些话题更大众化,对用户能力要求不高,回答数就相对较多,有些问题更专业化,对用户能力要求较高,回答数则相对较少。

假设 4 得到了验证,可能有以下两个原因:第一,关注某个用户的人倾向于从该用户那里寻求有帮助的问题答案,因而在看到该用户回答问题的消息推送时,更有可能进行浏览并投票;第二,投票行为比回答行为所需要付出的努力显然更少,因而如果用户关注了某个用户,那么出于一些社交目的,相比于回答该用户提出的问题,更有可能对该用户提供的答案进行投票。例如,有人在知乎上提问“年轻人太早结婚是否会影响事业的发展”,李开复回答“我 21 岁结婚”,获得了大量用户的赞同,这便是典型的“明星效应”。此外,从表 5 的结果可以发现,答案评论数对答案获得赞同数有显著的正向影响,这是因为评论可以补充、丰富答案内容,帮助其他人更好地理解答案,获取更多信息,因此会使答案更容易被认可;答案发布时间对答案获得赞同数并没有显著影响,说明知乎运用一定的算法有效避免了早起者偏差;答案信息量对答案获得赞同数也没有显著的影响,实际上这也是有可能的,因为知乎社区中的答案是否精辟并不在于信息量的多少,获得赞同数很多的答案往往是那些很短的一句话回答,这也正是知乎社区与一般在线评论社区的差异之一,知乎推出的移动端应用“知乎日报”中就有一个栏目,专门汇总整理出这种简短而受到很多人认可的答案。

6 结论

本文通过实证分析得出了以下结论：第一，在知乎社区中，参与度和权威性有助于用户增加自己在社区中的关注者数；第二，用户通过增加自己在社区中的关注者数，并不能使其提出的问题更易得到解决；第三，用户通过增加自己在社区中的关注者数，能使其提供的答案更易受到认可。

本文的创新主要有以下几点：第一，本文基于关系构建理论，结合社会化问答社区的特点，从用户信息角度建立了社交关系影响因素模型；第二，基于知识贡献模型，提出了社交关系会影响知识贡献的假设；第三，基于UGC和传统问答社区答案质量的研究，提出了社交关系等因素对答案质量判断的模型；第四，与以往社会化问答社区相关研究的定性特征描述不同，本文从知乎网站上抓取了大量数据进行定量分析和建模，从而验证并丰富了以往研究得出的一些结论。

本文的发现对以知乎为代表的社会化问答社区如何加强社交功能，促进用户参与，提升网站人气等具有一定的启示作用。第一，尽管知乎一直在努力弱化声誉机制的作用，但是应该保证用户的行为历史能被方便地查看，以方便用户间社会关系的建立。用户不仅应积极参与提问和回答，撰写专栏文章，而且应提高回答的质量，这样才可以获得更多的关注者。第二，由于社交机制并不能直接提高用户的知识共享行为，社会化问答社区需要通过其他机制，如促进问题关注、将问题准确归类至合适的话题等，来吸引更多用户回答问题。第三，社交机制的出现会促进用户对答案的投票行为，即点赞，这在一定程度上可以激励用户提供高质量答案；然而，由于社交机制会影响用户对答案质量的判断及投票行为，因此，社会化问答社区需要设计合适的算法来识别并量化这一影响，从而为所有用户呈现真正有帮助的高质量的答案。

本文的不足之处在于：首先，本文使用的是截面数据，如果能获取用户的参与历史，就可以更准确地衡量用户的参与行为与社交关系之间是如何相互作用的。其次，本文只考虑了单向的关注关系，而未考虑双向的朋友关系，研究表明，关系强度对用户的参与行为也会有重要影响。再次，很多研究将问题分为信息类和对话类两种，研究表明，用户对这两类问题的参与行为有所不同，若将它们分开进行更深入细致的研究，可能会有新的发现。最后，虽然知乎社区是一个以完全陌生关系为主的社区，但本文在样本的分析上并没有作进一步的研究。

参 考 文 献

- [1] Paul S A, Hong L, Chi E H. Who is authoritative? Understanding reputation mechanisms in Quora[C]. In Proceedings of Collective Intelligence, 2012.
- [2] Wang G, Gill K, Mohanlal M, et al. Wisdom in the social crowd: an analysis of Quora[C]. In Proceedings of International Conference on World Wide Web, 2013.
- [3] Granovetter M S. The strength of weak ties[J]. American Journal of Sociology, 1973, 78 (6): 1360-1380.
- [4] Goette L, Huffman D, Meier S. The impact of social ties on group interactions: evidence from minimal groups and randomly assigned real groups[J]. American Economic Journal Microeconomics, 2012, 4 (1): 101-115.
- [5] Wasko M M, Faraj S. Why should I share? Examining social capital and knowledge contribution in electronic networks of practice[J]. MIS Quarterly, 2005, 29 (1): 35-57.
- [6] Chiu C M, Hsu M H, Wang E T. Understanding knowledge sharing in virtual communities: an integration of social capital and social cognitive theories[J]. Decision Support Systems, 2006, 42 (3): 1872-1888.

- [7] Lin C, Huang Z, Yang F, et al. Identify content quality in online social networks[J]. IET Communications, 2012, 6(12): 1618-1624.
- [8] Agichtein E, Castillo C, Donato D, et al. Finding high-quality content in social media[C]. In Proceedings of the 2008 International Conference on Web Search and Data Mining, 2008.
- [9] Liu J, Cao Y, Lin C, et al. Low-quality product review detection in opinion summarization[C]. In Proceedings of the Conference on Empirical Methods in Natural Language Processing Association for Computational Linguistics, 2007.
- [10] Mudambi S M, Schuff D. What makes a helpful online review? A study of customer reviews on Amazon.com[J]. MIS Quarterly, 2010, 34 (1): 185-200.
- [11] Jurczyk P, Agichtein E. Discovering authorities in question answer communities by using link analysis[C]. In Proceedings of the 16th ACM Conference on Information and Knowledge Management, 2007.
- [12] Zhou Z M, Lan M, Niu Z Y, et al. Exploiting user profile information for answer ranking in CQA[C]. In Proceedings of the 21st International Conference on World Wide Web, 2012.

A Study of the Construction and Benefits of Social Relationships in the Social Q&A Community: An Empirical Research from Zhihu

JIANG Yichen, JIN Yue, HUANG Jinghua

(School of Economics and Management, Tsinghua University, Beijing 100084, China)

Abstract With the development of social media, social Q&A communities that integrate social network into their basic structures are rapidly growing. Based on studies about social relationships building, knowledge contribution and quality judgement of user generated content (UGC), this paper studies how users increase their followers and how it influences their participation experience. The empirical results show that engagement and authority of users are key contributors to the increase of their number of followers, and by having more followers, users cannot get more answers for their questions, but can get more votes for their answers.

Key words social Q&A community, social relationships, knowledge contribution, quality judgement of UGC

作者简介

蒋逸尘 (1993—), 女, 清华大学经济管理学院 2015 级博士研究生, 江苏靖江人, 研究方向为信息系统、电子商务等。E-mail: jiangych.11@sem.tsinghua.edu.cn。

金悦 (1992—), 女, 清华大学经济管理学院 2013 级博士研究生, 研究方向为信息系统、电子商务等。E-mail: jinyue.09@sem.tsinghua.edu.cn。

黄京华 (1963—), 女, 清华大学经济管理学院教授、博士生导师, 研究方向为信息系统、电子商务等。E-mail: huangjh@sem.tsinghua.edu.cn。