

在线社交网络分析与信息传播丛书

# 基于社交网络的行为分析与挖掘

李磊 李爱平 汪萌 吴信东 著



科学出版社

在线社交网络分析与信息传播丛书

# 基于社交网络的行为分析与挖掘

李 磊 李爱平 汪 萌 吴信东 著

科学出版社

北 京

## 内 容 简 介

随着社交网络应用的快速发展,用户在线行为越来越复杂,并呈现出与以往不同的内容和形式。本书对社交网络行为进行系统性分析,主要从社交网络行为成因、社交网络行为表现和社交网络行为影响三方面简略分析社交网络行为分析的研究进展。首先介绍社交网络行为成因,包括最重要的三个原因:用户采纳、用户忠诚和用户信任。有了这些原因,用户才在社交网络中发生行为。然后主要介绍用户行为、行为类别预测和行为主体的话题分析等社交网络行为表现。最后,对社交网络行为的影响力和对用户行为引导的相关工作进行分析。

本书内容丰富,涉及计算机科学、社会学、管理学等多个学科领域。通过算法示例演示社交网络行为分析方法,可读性较强。本书既可作为各高等院校及科研院所本科生和研究生的教材,也可供从事社交网络行为分析的科研人员使用。

---

### 图书在版编目(CIP)数据

基于社交网络的行为分析与挖掘/李磊等著. —北京:科学出版社, 2018.5

(在线社交网络分析与信息传播丛书)

ISBN 978-7-03-054829-0

I. ①基… II. ①李… III. ①互联网络—影响—人际关系学—行为分析 IV. ①C912.11-39

中国版本图书馆CIP数据核字(2017)第253366号

---

责任编辑:赵艳春 / 责任校对:郭瑞芝

责任印制:师艳茹 / 封面设计:迷底书装

科 学 出 版 社 出 版

北京东黄城根北街16号

邮政编码:100717

<http://www.sciencep.com>

文林印务有限公司 印刷

科学出版社发行 各地新华书店经销

\*

2018年5月第 一 版 开本:720×1 000 B5

2018年5月第一次印刷 印张:13

字数:243 000

定价:78.00 元

(如有印装质量问题,我社负责调换)

## 《在线社交网络分析与信息传播丛书》编委会

顾 问:

李国杰 梅 宏

主 编:

方滨兴

副主编:

贾 焰

编 委:

许 进 李建华 黄河燕 齐佳音

张宏莉 吴信东 周 斌 杨善林

胡长军 郭 莉 程学旗 廖湘科

秘 书:

王忠儒

# 前言

随着在线社交网络的蓬勃发展和线上用户的急剧增加,以交友、信息共享等为目的的社交网络迅速发展为人们传播信息、推销商品、表述观点、产生影响力的理想平台。截至2013年4月,就有约一半以上的中国网民通过社交网络沟通、交流、分享信息,社交网络已成为覆盖用户最广、传播影响最大、商业价值最高的Web 2.0业务。在线社交网络的这种蓬勃发展,一方面改变了人们的日常生活方式,改善了人们的生活质量,另一方面给国家和社会带来了一定的负面隐患。

在线社交网络已经逐步渗透到人们生活的方方面面。在经济方面,网上购物已经成为主流购物方式之一,超过70%的社交网络成人活跃用户使用淘宝等社交网络进行网上购物;2016年双十一购物节,淘宝的总成交额就达到1207亿元。在文化方面,社交网络正改变着人们的生活方式,网民利用社交网络可以不出家门进行交友、游戏、互动,形成了所谓的宅男、宅女。

在线社交网络在给人们生活带来便利的同时,也给社会各方面带来了一些负面影响。在政治方面,一些不法分子蓄意制造和传播有损国家和社会利益的谣言,刻意制造群体事件,影响社会稳定。例如,我国民众受微博谣言蛊惑而发生的抢盐风波,严重影响了社会的稳定;在2011年伦敦骚乱中,犯罪分子通过Twitter或Blackberry提供的免费信息服务协调各种犯罪行为<sup>①</sup>,导致骚乱持续扩大。在经济方面,不法分子通过聊天工具发布虚假信息,利用在线购物平台欺诈顾客;在文化方面,低俗团伙利用社交视频、即时通信等渠道传播网络色情内容;邪恶势力通过社交网络教唆青年人崇尚暴力,鼓吹破坏性的个人英雄主义。特别是,青少年更容易受到社交网络上的负面影响,研究指出89%的青少年曾经在社交网络上受到不同程度的冒犯<sup>②</sup>。

社交网络的本质在于它能够影响人们的思想,影响人们的世界观、认知观、价值观和人生观,并能够迅速形成群体行为。而负面群体行为又会给社会稳定带来巨大的影响。那么,如何有效掌控社交网络这一新型的战略资源,维护国家安全与社会稳定?如何了解民意,推广先进文化,引导网络群体行为?如何避免负面群体行为对于社会和经济的影响?这些都是国家安全与社会发展所面临的巨大挑战,亟待开展基于社交网络的行为分析和挖掘方法研究。

---

① <http://www.reuters.com/article/2011/08/08/us-britain-riot-idUSTRE7760G820110808>.

② 2013年中国电子商务研究中心报告. <http://b2b.toocle.com/detail--6048263.html>.

作为社交网络分析的关键内容,本书针对社交网络的行为分析和挖掘,首先介绍社交网络行为的成因,包括最重要的3个原因:用户采纳、用户忠诚和用户信任。有了这些原因,用户才在社交网络中发生行为,本书主要介绍用户行为、行为类别预测和行为主体的话题分析。最后,我们对社交网络行为的影响力和对用户行为的引导进行了研究。通过对社交网络行为成因、社交网络行为表现和社交网络行为影响的分析,我们能从社会学的角度加深对社交行为的理解,并为公共决策和舆论导向提供理论依据,同时促进政治、经济和文化的交流和沟通信息,具有重要的社会意义和应用价值。

具体来说,第1章是社交网络行为分析和挖掘的综述,对社交网络行为相关的研究作了系统性分析;主要是从社交网络行为成因、社交网络行为表现和社交网络行为影响三方面简略分析社交网络行为分析的研究进展。

第2章针对用户采纳介绍基于技术接受模型、社会影响理论和网络外部性理论构建的在线社交网络用户采纳行为影响因素模型及其实证检验;同时,针对用户忠诚介绍基于期望确认模型、感知价值理论和心流体验理论构建的在线社交网络用户忠诚意愿影响因素模型及其实证检验。

第3章主要介绍信任研究的概况。从概念上提出信任的一般结构、信任的基础及在不同学科中信任的概念。此外,在社交网络中引入典型的信任评估方法。我们还研究以信任为导向的社交网络环境中采用提取信任网络的方法。在此基础上,在以信任为导向的社交网络中分析现有的三类信任传播模型的运用和各自的优缺点。

第4章主要阐述社交网络的用户使用行为及异常使用行为检测,在此基础上,分析垃圾用户和网络水军发现方法。在异常使用行为检测方面,主要分析常规异常行为和交友使用行为。在垃圾用户发现方面,研究如何根据统计特征与信任双向传播方法利用四个优先级规则逐步排除的定性方法对用户分类。在网络水军发现方面,讨论在社交网络中如何检测网络水军及网络推手。

第5章首先介绍社交网络行为分类的意义;然后描述社交网络行为分类的问题描述与分类的理论基础;接下来介绍社交网络行为分类的研究现状;最后介绍两种具有代表性的社交网络行为分类方法,分别是基于社会维度的社交网络行为分类方法和基于关系分类模型的社交网络行为分类方法。

第6章主要介绍社交网络行为主体的话题分析,从话题的发现、摘要、演化等方面介绍话题分析技术与相关研究工作。在话题发现方面,主要介绍以突发性方法、主题模型方法、聚类方法为代表的三类话题检测模型。话题摘要技术部分,我们从关键词提取与关键句提取两方面进行介绍。在话题演化部分,我们主要从内容与强度两方面介绍话题演化模型。

第7章主要介绍影响强度计算和影响力个体分析两方面。其中,社交网络个体

之间的影响强度计算早期研究仅考虑网络拓扑结构；随后，相关学者提出了基于个体行为的影响强度计算模型和基于话题的影响强度计算模型。而影响力个体分析早期的计算方法主要基于网络拓扑结构；最后，学者研究了基于随机游走的影响力个体分析方法和基于个体行为与话题层次的影响力个体分析方法。

第8章介绍社交网络中行为的诱导与引导的相关内容。首先介绍社交网络中行为的诱导，主要从诱导交互技术、诱导技术框架——三元功能、诱导技术的应用以及计算机信任四方面介绍了社交网络中行为的诱导。然后主要介绍社交网络中行为的引导，主要是基于社交网络中的群体语义行为提取方法，具体介绍行为空间映射、信任智能体的特征选择及基于信任智能体的行为引导方法。

本书各章之间并不互相依存，读者可以根据需要选择任意章节研究，并不影响阅读。

本书是集体智慧的结晶。李磊副研究员主要负责第1章和第8章内容的撰写工作，王佳淼和何建萍协助。汪萌教授负责第2章和第5章内容的撰写工作，代宝和张赞协助。吴信东教授负责第3章和第6章内容的撰写工作，李磊、楚喻棋和介飞协助。李爱平教授负责第4章和第7章内容的撰写工作，李树栋协助。全书最后由李磊副研究员统稿，何进和楚喻棋负责校稿。

本书得到了国家重点研发计划项目（2016YFB1000901）、国家重点基础研究发展计划（2013CB329604）和国家自然科学基金（61503114）的资助，特此向支持和关心作者研究工作的所有单位与个人表示衷心的感谢。书中部分内容参考了国内外有关单位或个人的研究成果，均已在参考文献中列出，在此一并致谢。

另外，由于作者水平所限，虽几经改稿，书中难免存在不足之处，恳请广大读者不吝赐教。

作者

2017年9月于斛兵塘

# 目 录

## 前言

|                                  |    |
|----------------------------------|----|
| 第 1 章 社交行为分析简介 .....             | 1  |
| 1.1 引言 .....                     | 1  |
| 1.2 社交网络中的行为 .....               | 1  |
| 1.2.1 用户行为与语义行为 .....            | 2  |
| 1.2.2 个体行为与群体行为 .....            | 2  |
| 1.3 社交网络的行为成因 .....              | 2  |
| 1.3.1 社交网络中的用户采纳 .....           | 3  |
| 1.3.2 社交网络中的用户忠诚 .....           | 3  |
| 1.3.3 社交网络中的用户信任 .....           | 4  |
| 1.4 社交网络的行为表现 .....              | 6  |
| 1.4.1 一般使用行为 .....               | 6  |
| 1.4.2 内容创建行为 .....               | 8  |
| 1.4.3 内容消费行为 .....               | 9  |
| 1.5 社交网络的行为影响 .....              | 9  |
| 1.5.1 社交网络影响力的衡量 .....           | 9  |
| 1.5.2 社交网络中的行为引导 .....           | 10 |
| 1.6 社交网络中社交行为提取及其实现 .....        | 12 |
| 1.6.1 Java 环境变量配置 .....          | 12 |
| 1.6.2 Java 爬虫 WebCollector ..... | 16 |
| 1.6.3 分词工具 IKAnalyzer .....      | 24 |
| 1.6.4 机器学习 MALLET .....          | 30 |
| 1.6.5 根据 LDA 结果修改分词词典 .....      | 36 |
| 1.6.6 控制台修改 .....                | 38 |
| 1.7 本章小结 .....                   | 40 |
| 参考文献 .....                       | 40 |
| 第 2 章 社交网络中的用户采纳与忠诚 .....        | 45 |
| 2.1 引言 .....                     | 45 |
| 2.2 社交网络用户采纳 .....               | 46 |

|                      |    |
|----------------------|----|
| 2.2.1 理论基础           | 46 |
| 2.2.2 研究模型           | 47 |
| 2.2.3 实证检验           | 49 |
| 2.3 社交网络用户忠诚         | 52 |
| 2.3.1 理论基础           | 52 |
| 2.3.2 研究模型           | 53 |
| 2.3.3 实证检验           | 55 |
| 2.4 本章小结             | 58 |
| 参考文献                 | 59 |
| 第3章 社交网络中的信任         | 62 |
| 3.1 引言               | 62 |
| 3.2 信任的构成、基础及定义      | 62 |
| 3.2.1 信任的构成          | 62 |
| 3.2.2 信任的基础          | 64 |
| 3.2.3 信任的定义          | 66 |
| 3.3 社交网络中的信任衡量       | 67 |
| 3.4 社交网络中信任网络的提取     | 68 |
| 3.4.1 洪泛搜索           | 69 |
| 3.4.2 随机游走搜索         | 70 |
| 3.4.3 高权搜索           | 70 |
| 3.4.4 在线社交网络中的信任路径选择 | 71 |
| 3.5 社交网络中信任的传递       | 71 |
| 3.6 本章小结             | 73 |
| 参考文献                 | 73 |
| 第4章 社交网络中的用户行为       | 76 |
| 4.1 引言               | 76 |
| 4.2 用户使用行为概述         | 76 |
| 4.3 异常使用行为检测         | 77 |
| 4.3.1 常规异常使用行为       | 77 |
| 4.3.2 交友使用行为         | 78 |
| 4.4 基于使用行为的异常用户发现    | 81 |
| 4.4.1 基于使用行为的垃圾用户发现  | 81 |
| 4.4.2 基于使用行为的网络水军检测  | 89 |
| 4.5 本章小结             | 96 |

|                          |     |
|--------------------------|-----|
| 参考文献                     | 96  |
| 第 5 章 社交网络行为类别预测         | 99  |
| 5.1 引言                   | 99  |
| 5.2 社交网络行为标签分类的问题描述与理论基础 | 99  |
| 5.3 社交网络行为标签分类的研究现状      | 100 |
| 5.4 基于社会维度的社交网络行为分类      | 101 |
| 5.5 基于关系分类模型的社交网络行为分类    | 107 |
| 5.6 本章小结                 | 110 |
| 参考文献                     | 110 |
| 第 6 章 社交网络行为主体的话题分析      | 112 |
| 6.1 引言                   | 112 |
| 6.2 话题发现                 | 112 |
| 6.2.1 基于突发性的检测方法         | 113 |
| 6.2.2 基于主题模型的检测方法        | 115 |
| 6.2.3 基于聚类的检测方法          | 119 |
| 6.3 话题摘要                 | 120 |
| 6.3.1 关键词/短语摘要           | 120 |
| 6.3.2 整句摘要               | 122 |
| 6.4 话题演化                 | 123 |
| 6.4.1 内容演化               | 123 |
| 6.4.2 强度演化               | 126 |
| 6.5 本章小结                 | 126 |
| 参考文献                     | 127 |
| 第 7 章 社交网络行为的影响力分析       | 130 |
| 7.1 引言                   | 130 |
| 7.1.1 影响强度表示与计算          | 130 |
| 7.1.2 影响力个体分析            | 130 |
| 7.2 影响力强度数学模型            | 132 |
| 7.2.1 基于社交网络属性的模型        | 132 |
| 7.2.2 基于用户行为的模型          | 133 |
| 7.2.3 基于话题内容的模型          | 135 |
| 7.3 影响力个体分析              | 141 |
| 7.3.1 基于社交网络属性的影响力分析     | 141 |

|                                 |            |
|---------------------------------|------------|
| 7.3.2 基于传统排名算法的影响力分析 .....      | 144        |
| 7.3.3 基于用户行为的影响力分析 .....        | 148        |
| 7.3.4 基于话题内容的影响力分析 .....        | 151        |
| 7.3.5 基于扩散能力的影响力分析 .....        | 152        |
| 7.4 本章小结 .....                  | 158        |
| 参考文献 .....                      | 159        |
| <b>第8章 社交网络行为的诱导与引导 .....</b>   | <b>161</b> |
| 8.1 引言 .....                    | 161        |
| 8.2 社交网络中行为的诱导 .....            | 161        |
| 8.2.1 诱导交互技术 .....              | 161        |
| 8.2.2 诱导技术框架：三元功能 .....         | 164        |
| 8.2.3 诱导技术的应用 .....             | 168        |
| 8.2.4 信任 .....                  | 169        |
| 8.3 社交网络中群体语义行为提取 .....         | 173        |
| 8.3.1 群体语义行为提取过程 .....          | 174        |
| 8.3.2 实验结果分析 .....              | 178        |
| 8.4 社交网络中行为的引导 .....            | 182        |
| 8.4.1 社交网络中行为引导的相关研究内容 .....    | 182        |
| 8.4.2 基于信任智能体的行为引导 .....        | 184        |
| 8.4.3 基于信任智能体的行为引导方法的仿真实验 ..... | 186        |
| 8.5 本章小结 .....                  | 190        |
| 参考文献 .....                      | 190        |

# 第1章 社交行为分析简介

## 1.1 引言

近年来,社交网络类应用,如在线社交网络、微博、博客、BBS和维基出现并迅速发展起来。这导致互联网使用方式的重大改变,即从简单的信息检索和网页浏览到社交关系的构建与维护,以及基于社交关系的信息创建、交流与共享。基于互联网的社交网络成为社交关系维护和信息共享的重要渠道和载体<sup>[1]</sup>。一般来说,社交网络是一种由多个节点和节点之间关系构成的社会结构<sup>[2]</sup>。其中,一个节点表示一个人或者社交网络的一个参与者。利用节点间的链接关系,社交网络可以链接各种各样的社会关系,不仅包括偶然相识说过几句话的陌生人关系,而且包括家庭成员之间的紧密血缘关系。

和传统的网络应用<sup>[3]</sup>或者信息媒体相比,在线社交网络具有独特的性质:在通过社交网络简单发布或者收到信息之后,社交网络的每个参与者都有权利发表自己的观点,不同的观点每时每刻都在社交网络中发布;一旦信息发布出去,就有可能以“核爆炸”的方式几何扩散出去,从而促成具有类似兴趣的虚拟社区的形成<sup>[2,4,5]</sup>。因此,社交网络可以快速组织线下人员到现实的社会群体活动,从而形成了一定的社会动员力。随着社交网络的快速发展,在线社交网络和线下社交网络的融合对人类社会的行为影响越来越大。因此,对社交网络,特别是社交网络中的行为进行研究具有重大的理论价值、应用价值和实际意义。

为了解决国家安全和发展的重大需求问题,如网络舆情分析和引导、突发事件监控和响应以及在线群体交互等,本章调研并简单分析了社交网络中社交行为成因、社交行为的表现及社交行为的影响等社交行为分析与挖掘的相关理论及模型,并仔细描述了如何从在线社交网络中提取行为数据。

## 1.2 社交网络中的行为

随着在线社交网络的快速发展和在线用户的快速增加,以交友和信息共享为目的的社交网络成为信息传播、商品推荐、观点表达和影响力产生的理想平台。对于不同类型的社交网络技术、服务和应用,网络用户选择合适的行为以满足他们对社交、娱乐和信息获取的需求。

### 1.2.1 用户行为与语义行为

社交网络的用户行为是指基于个人需求、社会影响力和社交网络技术等因素影响用户意愿的因素而采用某种社交网络服务的行为及各种相关活动的总和。典型的社交网络用户行为包括一般用户行为、内容创建行为和内容消费行为。这也是基于社交网络行为分析的主要研究内容。例如,我们可以研究内容创建或者内容消费的模式,从而发现下次内容创建或者内容消费的时间。

另外一个基于社交网络行为分析的重要研究方向是关于社交网络的语义行为。社交网络中的语义行为是指基于社交网络中的用户行为,一般是基于社交网络用户发布的内容,推导出其在现实生活中的用户行为。例如,基于一个用户发布的图片的背景,我们可以推断出用户的家庭住址、每天的行进路线或者个人爱好。这些信息可以被不法分子利用社会工程学中的方法来攻击用户,从而导致用户的经济损失。

社交网络的用户行为是社交网络行为的狭义形式,这也是当今社交网络行为研究的核心,大部分关于社交网络行为研究的论文都是关于这方面的。社交网络的语义行为是社交网络行为的广义形式,这是社交网络行为研究的未来。这是因为语义行为可以覆盖更多的信息,可以对用户的工作和生活有更大的影响。但是,由于社交网络语义行为可以较容易地伪造<sup>[6,7]</sup>,导致得出信息的可信度比较低<sup>[8,9]</sup>,所以我们需要引入新的理论和方法来确定社交网络语义行为的可信度,获得社交网络语义行为的真实信息,从而获得社交网络用户在真实生活中的行为信息。

### 1.2.2 个体行为与群体行为

社交网络的个体行为分析是指针对社交网络中的一个参与者基于其个人需求、社会影响和社交网络技术综合分析衡量所采用的服务和所有相关行为<sup>[10]</sup>。而社交网络的群体行为是由某种条件激发,之后多个个体加入一个群体中并彼此产生行为交互<sup>[11]</sup>(如共享信息),以达到某种程度的一致,从而表现出一种群体行为<sup>[12]</sup>。简单来说,社交网络的个体行为是指分析一个人的行为,而社交网络的群体行为是指分析一群人的行为。因此,个体行为研究是群体行为研究的基础,相对得到了更广泛的研究。所以,本章主要从微观角度分析社交网络的个体行为,基本不涉及宏观角度的社交网络群体行为分析。

## 1.3 社交网络的行为成因

对于社交网络行为成因,即什么导致了社交网络的行为,我们基于社交网络的技术性质和用户的心理经验研究了用户采纳、用户忠诚和用户信任对用户行为的影响。

### 1.3.1 社交网络中的用户采纳

社交网络中的用户采纳是指用户基于他们自己的需求、社会影响和社交网络技术等原因来采纳某种社交网络服务的意愿和行动。到目前为止,对社交网络用户的采纳行为的原理的研究已经有了很多理论和模型。在实际应用中,我们用得比较多的有技术接受模型(technology acceptance model, TAM)和计划行为理论(theory of planned behaviour, TPB)。

对于技术接受模型, Kwon 和 Wen<sup>[12]</sup>研究了在线社交网络的用户接受行为,并证实了临场感(telepresence)和利他主义(altruism)通过可用性感知和有用性感知对容易易感性有着间接作用,而社会角色对有用易感性和鼓励易感性有着直接的显著作用。Nikou 和 Bouwma<sup>[13]</sup>在基础 TAM 上增加了一些新的变量,包括移动性、关键群体、使用背景、社会影响力和习惯。而 Sledgianowski 和 Kulviwat<sup>[14]</sup>在基础 TAM 上增加了有趣易感性、规范压力易感性、信任和关键群体作为使用意愿的直接影响因素。实验结果证实了这四个变量都对使用意愿有着直接影响,并且有趣易感性对使用意愿有着显著的直接影响。

对于计划行为理论, Baker 和 White<sup>[15]</sup>研究了在线社交网络中的用户采纳,并观察到态度和行为控制易感性都对青少年在社交网络的使用意愿有着显著的直接作用。同时,主观规范对青少年在社交网络的使用意愿没有显著的直接作用。但是,这和早期关于 TPB 的研究结果相矛盾。这很可能是由于在衡量社会规范性对于采用的影响方面,这篇论文新引入的社团规范性比早期研究中使用的主观规范性更有效。这可以由当社团规范性进入回归模型的时候主观规范性对使用意愿的影响从显著变化为微弱这个事实证实。

在 TPB 的基础之上, Pelling 和 White<sup>[16]</sup>增加了新的变量,如自我认知、从属性、性别、年龄和过去使用史,构建了一个新的社交网络用户使用意愿预测模型。实验结果显示态度、主观规范性和自我认知全都对使用意愿有着显著的正面作用。Chang 和 Zhu<sup>[17]</sup>利用 TPB 模型聚焦了中国的社交网络采用行为,并构建了 5 种使用意愿模型。分析结果显示,信息、娱乐、交友和一致性都对社交网络使用意愿有着显著影响,并且用户态度、行为控制易感性和主观规范都对社交网络采用意愿有着显著影响。

### 1.3.2 社交网络中的用户忠诚

在用户开始使用在线社交网络之后,如何使用户继续使用成为社交网络服务提供商的一个核心难题。随着新的社交网络服务产生压力的逐渐增大,保持用户的忠诚度变得越来越困难。因此,我们需要引入理论和方法来研究社交网络中用户忠诚的原理。到目前为止,关于社交网络用户忠诚的研究中,期望确认理论(expectation

confirmation theory, ECT) 和心流体验理论 (flow theory) 是在实际应用中使用最多的理论。

对于期望确认理论, Kang 等<sup>[18]</sup>聚焦了韩国最大的社交网络赛我网 (Cyworld) 的用户行为, 试图发现社交网络中用户忠诚的关键因素。实验结果揭示了期望确认对满意度没有显著的作用; 满意度对持续意愿有着微弱的作用; 后悔对满意度没有显著的作用, 这可能是由于负面作用只在期望确认值比较低的时候显现, 而 Cyworld 的数据集中期望确认值很高; 利用期望确认理论, Chen 和 Shao<sup>[19]</sup>引入了容易易感性、有趣易感性和变换代价易感性等。他们证实了容易易感性、有用易感性、兴趣易感性和变换代价易感性都对社交网络的持续意愿有着直接显著作用。Shin 和 Hall<sup>[20]</sup>利用创新扩散理论中的创新易感性代替期望确认理论中的容易易感性, 并观察到创新易感性对社交网络的持续使用有着显著影响。

基于心流体验理论, Zhou 等<sup>[21]</sup>研究了移动社交网络的用户忠诚。分析结果显示, 移动社交网络的信息质量和系统质量都对信任和心流体验经历有相关作用, 信任对心流体验经历和用户忠诚有相关作用, 心流体验经历对用户忠诚有显著作用。除此之外, 和信息质量相比, 系统质量对用户信任有重大作用。Hsu 和 Wu<sup>[22]</sup>结合技术接受和使用统一理论 (unified theory of acceptance and use of technology, UTAUT)、期望失验理论 (expectation disconfirmation theory, EDT) 和心流体验理论, 针对 Facebook 用户提出了一个持续意愿模型。在分析了 482 个样本数据之后, 他们得出心流体验经历不仅通过满意度对持续使用有着间接作用, 而且对持续意愿有着直接显著作用。另外, Wu 等<sup>[23]</sup>和 Chang<sup>[24]</sup>采用心流体验理论解释了社交网络中社交游戏或者娱乐应用的持续行为。基于心流体验理论, Wu 等<sup>[23]</sup>验证了兴趣易感性是用户对社交网络娱乐应用变得忠诚的直接原因, 并且兴趣易感性可以被社交网络的娱乐应用的设计影响 (如内在社会性、应用之间的交互)。Chang<sup>[24]</sup>利用心流体验理论来验证交互作用、用户价值和满意性是心流体验理论的先验变量, 并且心流体验经历和满意度是社交网络游戏中持续行为的先验变量。

### 1.3.3 社交网络中的用户信任

在社交网络平台上, 参与者进行各种活动, 如寻找员工和电影推荐<sup>[25]</sup>。在这些活动中, 信任是参与者决策时最重要的指标之一<sup>[26,27]</sup>。然而在线社交网络中, 大多数参与者先前并没有直接交互 (例如, 雇主和雇员之间以前并没有互动, 在电影推荐者和接受推荐的人之间以前并没有交互)。因此, 在两名未知参与者之间评估信任变得重要和必要。在介绍具体的应用方法之前, 我们首先介绍如何利用各种和信任相关的不同信息源来确定信任。

研究人员尝试确定引导信任出现的原因, 包括心理、社会和组织等因素, 这些因素会影响个人对他人的信任和对交易中行为的期望<sup>[1,26]</sup>。

(1) 倾向性信任: 每个人对不同的人所表现的信任是不同的, 具有一定的倾向性<sup>[26]</sup>, 并且这种倾向相对稳定<sup>[28]</sup>。

(2) 基于历史的信任: 人们选择信任的人很大程度上依赖于对他过去的了解<sup>[29]</sup>。

(3) 基于第三方渠道的信任: 最简单的基于第三方渠道的信任的例子就是八卦和口碑, 这些渠道可以为我们提供关于他人重要的二手消息。

(4) 基于类别的信任: 人们更倾向于给予其他群内成员正面属性特征<sup>[28,30]</sup>, 这将导致人们与组内的其他成员之间建立一种组织信任。

(5) 基于角色的信任: 基于角色的信任侧重于对待估者在组织中的特定角色来进行信任评估, 而不是信托人对待估者的性格、意图和动机都有特定认知之后进行信任评估<sup>[27]</sup>。

(6) 基于规则的信任: 基于规则信任不是有意识地对后果进行估算, 而是对适当行为规则的共同理解。关于规则对自我认识和对组织中其他成员预期的影响, 它可以在组织中创建和保持一个高水平的信任<sup>[31]</sup>。

为了评价社交网络中两个未知参与者之间的信任, 已有文献提出了以下几种信任管理方法。在信任传播的研究中, Golbeck 等<sup>[32]</sup>通过添加合适人选的办法来扩大一个朋友的朋友列表 (friend of a friend, FOAF), 该方法是利用参与者过去交往的经验来判断一个人是否可信。这个功能已经被应用在 FilmTrust 数据集上, 其中每个参与者都可以对另一个参与者指定一个 1 (表示非常低的信任) ~10 (表示非常高的信任) 的信任度, 而信任度的值是基于参与者电影推荐的质量。然后他们提出了建立信任关系的信任推理机制, 该机制在源参与者和与其没有直接交互的目标参与者之间使用<sup>[33]</sup>。在此模型中, 被源节点所信任的那些邻居节点的态度 (已获得源节点的高信任值) 在推断他们的邻居节点的信任值时要考虑进去。他们平均这些由源节点的邻居赋给他们邻居的信任值, 此过程将重复直到目标节点。这种信任推理模型也在 FilmTrust 上得到了进一步的应用。Guha 等<sup>[34]</sup>提出了一种信任传播模式来推断两个没有直接交互的参与者之间的信任或不信任。在他们的模型中, 传播的跳数以及在源节点和目标节点之间社会信任路径上中间节点的信任情况都要考虑。他们采用了加权求和的方法来合计以上每个部分对信任传播的影响力, 由此计算并推断出信任值。

在面向信任的推荐系统研究中, Walter 等<sup>[35]</sup>提出了一个面向社交网络的推荐系统。在他们的模型中, 参与者可以给推荐人一个信任值, 该值是根据推荐人过去的推荐行为给出的。这种信任值可见且可为其他参与者的选择提供参考建议。参与者通过将源节点和目标节点之间的信任值乘以中间任何在信任路径上的参与者的信任值以计算目标参与者的信任值。Jamail 等<sup>[36]</sup>针对在线社交网络的卖家和买家提出了一个随机游走模型。在他们的模型中, 买家在社交网络中沿着他的购买路径执行固定跳数的随机游走来得到评分, 该值是由买家评分给卖方的, 用以评价该卖方的产

品。而卖方的置信度是由随机游走路径的数量、跳数及卖方在各路径上的评分来计算的。

上述信任评价方法考虑了社会信任路径中间节点间的信任值，它可以帮助在两个未知参与者之间建立信任关系。

## 1.4 社交网络的行为表现

基于社交网络行为的表现，我们对社交网络中不同的行为进行了分类，并使用数据网络技术对行为表现进行了研究。

### 1.4.1 一般使用行为

随着在线社交网络应用的发展，利用社交网络中用户的大规模数据已经成为可能<sup>[37]</sup>。举例来说，Goyal 等<sup>[1]</sup>收集了 420 万用户发布的 3.62 亿条日志信息，并且分别以每天和每月作为时间单位分析社交网络用户的日常行为。Gyarmati 和 Trinh<sup>[38]</sup>分析了常用的社交网络，包括拜博网（Bebo）、聚友网（MySpace）、Netlog、Tagged 等中的 80000 个用户，并发现社交网络中用户的使用时间满足 Weibull 分布，社交网络中用户的持续使用时间满足幂律分布。现在我们以这份工作为例详细解释我们对社交网络用户行为研究的思路。首先，我们描述了收集 80000 个用户的方法。然后，社交网络用户的使用时间的模式可以被挖掘出来。最后，我们可以指出社交网络用户的持续使用时间的模式。其中，社交网络中一个用户的持续使用时间是指一个用户从登录系统到使用结束花费的总时间。通常有两种方法结束使用行为：①用户可以自己退出系统；②用户可以关闭浏览器。

#### 1) 数据收集

因为一般来说很难从社交网络服务提供商处直接获得数据，Gyarmati 从社交网络网站，包括拜博网、聚友网、Netlog 和 Tagged 爬取了 80000 个用户的公共信息（包括用户的个人信息和发布的状态信息）。对应的时间段是 2009 年 3 月 15 日～5 月 2 日。

在数据收集过程中，我们需要不时地下载用户的公共信息。因此，选择合适的时间窗口以监控用户公共信息的变化是很重要的。在总资源有限（只有 500 个节点用于爬取）的约束下，在时间窗口较小的情况下，可以获得细粒度的数据，同时下载到更多的用户公共信息网页。但是，这只能监控到有限数量的用户。而更大的时间窗口对应更多的被监控用户。根据经验，时间窗口取 1 分钟。每隔 1 分钟，社交网络中一个用户的持续使用时间可以通过下载网页的变化估计出来。举例来说，如果一个网页发生了变化，那么可以估计用户仍然在线。