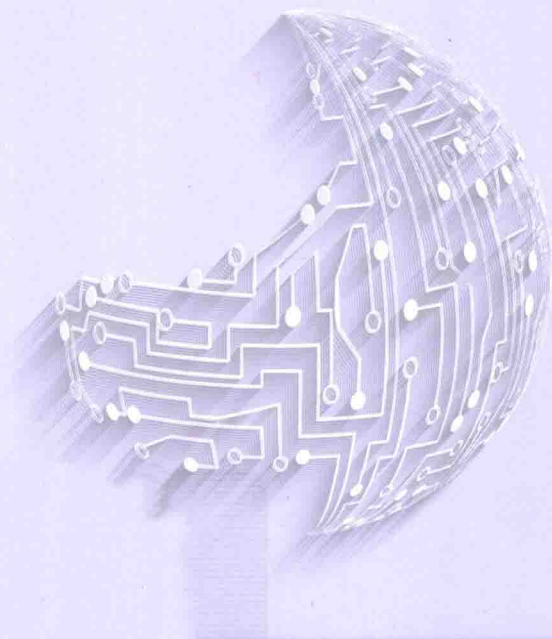


在线社交网络分析

方滨兴◎等著



Online Social Network Analysis



电子工业出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
<http://www.phei.com.cn>

在线社交网络分析

方滨兴 许 进 李建华 齐佳音 张宏莉 吴信东 周 斌
杨善林 胡长军 郭 莉 贾 焰 程学旗 廖湘科 著

(作者次序按笔画排列, 排名不分先后)

電子工業出版社

Publishing House of Electronics Industry

北京·BEIJING

内 容 简 介

互联网正逐步演变为无处不在的计算平台和信息传播平台。在线社交网站、微博、博客、论坛、维基等丰富多彩的社交媒体应用的出现和迅猛发展,使得人类使用互联网的方式产生了深刻变革——由简单信息搜索和网页浏览转向网上社会关系结构的构建与维护、具有共同行为特征的网络群体的聚集与交互,以及基于社会关系的信息创造、交流与传播。

本书从在线社交网络的“结构与演化—群体与互动—信息与传播”三个方面展开,系统、深入地阐述了在线社交网络分析中的基础理论、关键方法和技术,主要内容包括社交网络的结构特性与演化机理分析、社交网络群体行为的形成与互动规律,以及社交网络的信息传播模型及演化规律。本书第一次比较全面地揭示了社交网络的“结构、群体、信息”三个要素之间的复杂交互关系和互动规律,为开展社交网络分析与信息传播研究提供了重要的理论和技术支撑。

本书内容十分丰富,涉及计算机科学、社会学、管理学、心理学等多个学科领域,通过大量实例展示算法分析,可读性很强,既可作为国内各高等学校、科研院所本科生、研究生的教材,也可供国内外从事社交网络分析方面的研究人员、工程人员使用。

未经许可,不得以任何方式复制或抄袭本书之部分或全部内容。
版权所有,侵权必究。

图书在版编目(CIP)数据

在线社交网络分析/方滨兴等著. —北京:电子工业出版社,2014.11

ISBN 978-7-121-23508-5

I. ①在… II. ①方… III. ①互联网络—应用—人际关系学—研究 IV. ①C912.1-39

中国版本图书馆CIP数据核字(2014)第125583号

责任编辑:徐蔷薇

特约编辑:王 纲 劳嫦娟

印 刷:北京京师印务有限公司

装 订:北京京师印务有限公司

出版发行:电子工业出版社

北京市海淀区万寿路173信箱 邮编 100036

开 本:787×1092 1/16 印张:24.5 字数:572千字

版 次:2014年11月第1版

印 次:2014年11月第1次印刷

定 价:79.00元

凡所购买电子工业出版社图书有缺损问题,请向购买书店调换。若书店售缺,请与本社发行部联系,联系及邮购电话:(010)88254888。

质量投诉请发邮件至 zlts@phei.com.cn, 盗版侵权举报请发邮件至 dbqq@phei.com.cn。

服务热线:(010)88258888。

前言

21 世纪，是人类高度依赖数据、深度融入信息社会的世纪。在信息社会中，在线社交网络构建了庞大的功能平台。人们在 Twitter、Facebook、LinkedIn、新浪微博、微信等社交网络上表达观点、交友互动，每天都产生数以亿计的信息，让人们快捷地获取海量的信息。在线社交正在改变着人们的行为模式和社会形态，而在线社交网络数据也正在成为最成熟的大数据，通过运用大数据技术，人们有望对在线社交网络大数据背后的用户行为、社会现象的理解达到空前的深度。

在线社交网络分析涉及计算机科学、社会学、管理学、心理学等多个学科领域，本人作为“973”项目“社交网络分析与网络信息传播的基础研究”的首席科学家，在从事在线社交网络的研究过程中，深感本领域缺少一本从多学科的角度，系统论述在线社交网络分析概念、理论和技术的专著。为此，本人组织了“973”项目团队成员，包括国防科大、上海交大、合肥工大、北邮、中科院计算所、北大、北科大、中科院信工所和哈工大，基于这些团队的研究成果，在系统梳理国内外相关理论和技术的基础上，编著了本书，以为相关研究者提供理论性、系统性、工具性的研究指导。

本书从在线社交网络分析的三个核心要素，即“结构与演化”、“群体与互动”和“信息与传播”出发进行编写，共分 12 章。第 1 章绪论，统领全书，由本人负责执笔。第 2~4 章围绕第一个核心要素“结构与演化”展开，其中，第 2 章社交网络结构特征分析及建模，由许进教授和张宏莉教授负责执笔；第 3 章虚拟社区发现技术与方法，由李建华教授负责执笔；第 4 章虚拟社区的演化分析，由程学旗研究员负责执笔。第 5~8 章围绕第二个核心要素“群体与互动”展开，其中，第 5 章用户行为分析，由杨善林院士负责执笔；第 6 章社交网络情感分析，由周斌教授和本人负责执笔；第 7 章个体影响力分析及技术，由贾焰教授和本人负责执笔；第 8 章群体聚集及影响机制，由齐佳音教授负责执笔。第 9~12 章围绕第三个核心要素“信息与传播”展开，其中，第 9 章面向社交网络的信息检索，由郭莉研究员级高工负责执笔；第 10 章社交网络信息传播规律，由胡长军教授负责执笔；第 11 章话题的发现与演化，由吴信东教授负责执笔；第 12 章影响力最大化计算方法，由廖湘科研究员负责执笔。

感谢参与本书资料收集、内容整理，以及成果贡献的专家学者（排名不分先后，以姓氏笔画为序）：丁兆云、王晓萌、王斌、刘业政、刘晓东、李生红、李爱平、李磊、杜诗雨、吴鹏、陈秀真、陈巍、杨扬、张鲁民、时鹏、姜元春等。

感谢在本书编著过程中精心协调安排的李树栋副教授！感谢韩伟红、杨树强在本书审查校对阶段的辛勤工作！

目 录

第 1 章 绪论	1
1.1 社交网络及其发展	1
1.1.1 社交网络的起源	1
1.1.2 从社会学角度看社交网络的发展历程	1
1.1.3 从人类学角度看社交网络的发展历程	3
1.2 在线社交网络发展	4
1.2.1 在线社交网络的概念	4
1.2.2 在线社交网络的特点	5
1.2.3 在线社交网络的发展	5
1.2.4 在线社交网络对人们生活的影响	6
1.3 在线社交网络分析的背景和意义	7
1.4 在线社交网络分析的科学问题	8
1.4.1 在线社交网络分析面临的挑战	8
1.4.2 三个科学问题与相关研究内容	10
1.5 本书的组织安排	18
参考文献	20
第 2 章 社交网络结构特征分析及建模	22
2.1 引言	22
2.2 示例	22
2.3 社交网络的统计特性	23
2.3.1 度分布	24
2.3.2 平均路径长度	24
2.3.3 网络密度	25
2.3.4 聚集系数	26
2.3.5 介数	27
2.4 社交网络特性分析	27
2.4.1 小世界现象	28

2.4.2	无标度特性	30
2.4.3	同配性	34
2.4.4	互惠性	37
2.5	社交网络结构建模与生成	37
2.5.1	WS 模型	38
2.5.2	WS 模型的扩展	40
2.5.3	BA 模型	41
2.5.4	BA 模型的扩展	44
2.5.5	其他模型	46
2.6	本章小结	48
	参考文献	49
第 3 章	虚拟社区发现技术与方法	52
3.1	引言	52
3.2	虚拟社区发现技术理论基础	52
3.2.1	虚拟社区的定义	52
3.2.2	虚拟社区发现算法的发展进程	54
3.2.3	虚拟社区算法评价的准确度指标	55
3.2.4	虚拟社区算法计算复杂度	58
3.2.5	用于虚拟社区发现算法测试的典型数据集	59
3.3	虚拟社区静态计算发现算法	63
3.3.1	模块度最优化算法	63
3.3.2	多目标优化算法	66
3.3.3	基于概率模型的算法	69
3.3.4	信息编码算法	72
3.4	虚拟社区动态计算发现算法	75
3.4.1	派系过滤算法	75
3.4.2	基于相似度的聚合算法	78
3.4.3	标签传播算法	81
3.4.4	局部扩展优化算法	84
3.5	本章小结	87
	参考文献	87
第 4 章	虚拟社区的演化分析	90
4.1	引言	90
4.2	虚拟社区的涌现	90

4.2.1 虚拟社区涌现的周期闭包	91
4.2.2 虚拟社区涌现的偏好连接	92
4.2.3 虚拟社区涌现的老化因素	96
4.3 虚拟社区的演化	98
4.3.1 虚拟社区演化的累积效应	98
4.3.2 虚拟社区演化的结构多样性	101
4.3.3 虚拟社区演化的结构平衡性	105
4.4 演化虚拟社区的发现	106
4.4.1 基于相邻时刻相似度直接比较的演化虚拟社区发现	106
4.4.2 基于演化聚类分析的演化虚拟社区的发现	107
4.4.3 基于拉普拉斯动力学方法的演化虚拟社区发现	108
4.4.4 基于派系过滤算法的演化虚拟社区发现	109
4.4.5 基于节点行为趋势分析的演化虚拟社区发现	110
4.5 本章小结	110
参考文献	111
第 5 章 用户行为分析	114
5.1 引言	114
5.2 在线社交网络用户采纳与忠诚	115
5.2.1 在线社交网络用户采纳	115
5.2.2 在线社交网络用户忠诚	122
5.3 用户个体使用行为	130
5.3.1 一般使用行为	130
5.3.2 内容创建行为	136
5.3.3 内容消费行为	144
5.4 用户群体互动行为	150
5.4.1 群体互动的关系选择	150
5.4.2 群体互动的内容选择	154
5.4.3 群体互动的的时间规律	156
5.5 本章小结	159
参考文献	160
第 6 章 社交网络情感分析	165
6.1 引言	165
6.1.1 情感分析历史	165
6.1.2 情感定义及分类	166

6.1.3	情感分析应用	168
6.2	文本情感分析技术	169
6.2.1	基于语义规则的情感分析技术	169
6.2.2	基于监督学习的情感分析方法	172
6.2.3	基于话题模型的情感分析技术	177
6.3	社交网络情感分析技术	178
6.3.1	面向短文本的情感分析技术	178
6.3.2	基于群体智能的情感分析技术	182
6.3.3	社交网络的垃圾意见挖掘技术	183
6.4	情感分析技术的延伸与变形	184
6.4.1	情感摘要技术	185
6.4.2	基于迁移学习机制的情感分析技术	186
6.5	本章小结	187
	参考文献	187
第7章	个体影响力分析及技术	191
7.1	引言	191
7.2	影响强度计算	193
7.2.1	基于网络结构的影响强度计算	193
7.2.2	基于行为的影响强度计算	194
7.2.3	基于话题的影响强度计算	196
7.3	影响力个体发现	198
7.3.1	基于网络结构的个体影响力计算	198
7.3.2	PageRank 算法	201
7.3.3	基于行为的个体影响力计算	204
7.3.4	基于话题的个体影响力计算	207
7.4	本章小结	208
	参考文献	208
第8章	群体聚集及影响机制	210
8.1	引言	210
8.2	群体智慧产生机理	212
8.2.1	群体智慧	212
8.2.2	自我决定论与群体智慧	213
8.2.3	群体智慧产生的条件	214
8.2.4	群体智慧产生的影响因素	215

8.2.5 群体智慧的分析模型	217
8.2.6 社交网络上的群体智慧仿真	222
8.3 群体极化产生机理	228
8.3.1 群体极化	228
8.3.2 社会比较理论与群体极化	229
8.3.3 群体极化产生的条件	231
8.3.4 群体极化产生的影响因素	232
8.3.5 群体极化分析的主要模型	233
8.3.6 无社交网络结构影响的社交网络群体极化仿真	240
8.3.7 有社交网络结构影响的社交网络群体极化仿真	244
8.4 本章小结	251
参考文献	252
第9章 面向社交网络的信息检索	256
9.1 引言	256
9.2 社交网络内容搜索	259
9.2.1 经典信息检索及相关反馈模型	259
9.2.2 微博搜索中的查询表示研究	266
9.2.3 微博搜索中的文档表示研究	270
9.2.4 微博检索模型研究	273
9.3 社交网络内容分类	277
9.3.1 短文本分类中的特征处理研究	278
9.3.2 短文本分类的算法研究	280
9.4 社交网络推荐	281
9.4.1 社会化推荐简介	283
9.4.2 基于记忆的社会化推荐	284
9.4.3 基于模型的社会化推荐	288
9.5 本章小结	293
参考文献	294
第10章 社交网络信息传播规律	302
10.1 引言	302
10.2 社交网络中信息传播影响因素	303
10.2.1 社交网络结构	303
10.2.2 网络群体	304
10.2.3 信息	304

10.3	基于网络结构的传播模型	305
10.3.1	线性阈值模型	305
10.3.2	独立级联模型	306
10.3.3	相关扩展模型	307
10.4	基于群体状态的传播模型	308
10.4.1	经典传染病传播模型	308
10.4.2	社交网络中的传染病信息传播模型	310
10.4.3	基于影响力的传播模型	311
10.5	基于信息特性的传播模型	311
10.5.1	多源信息传播分析	312
10.5.2	信息竞争传播现象	312
10.6	热度预测方法	314
10.6.1	基于历史热度的预测模型	314
10.6.2	基于网络结构的预测模型	315
10.6.3	基于用户行为的预测模型	316
10.6.4	基于时间序列法的预测模型	317
10.7	信息溯源技术	324
10.7.1	信息溯源概述	324
10.7.2	基于中心度测量的溯源方法	325
10.7.3	通过统计推理框架溯源	327
10.7.4	基于反向传播与节点分区的多源信息溯源技术	330
10.8	本章小结	332
	参考文献	333
第 11 章	话题的发现与演化	336
11.1	引言	336
11.2	话题发现的模型和算法	337
11.2.1	基于主题模型的话题发现	338
11.2.2	基于向量空间模型的话题发现	348
11.2.3	基于词项关系图的话题发现	351
11.3	话题演化的模型和算法	354
11.3.1	朴素话题演化	354
11.3.2	基于主题模型的话题演化	356
11.3.3	基于相邻时间片关联的话题演化	358
11.4	本章小结	359
	参考文献	359

附录	362
第 12 章 影响力最大化计算方法	365
12.1 引言	365
12.2 基本概念与理论基础	365
12.3 影响力最大化度量标准	367
12.4 影响力最大化算法分类	368
12.5 影响力最大化贪心算法	369
12.5.1 贪心算法基础概念	369
12.5.2 BasicGreedy 算法	369
12.5.3 CELF 算法	370
12.5.4 MixGreedy 算法	371
12.5.5 其他贪心算法	372
12.5.6 贪心算法小结	373
12.6 影响力最大化启发式算法	373
12.6.1 DegreeDiscount 启发式	373
12.6.2 PMIA 启发式	374
12.6.3 LDAG 启发式	374
12.6.4 其他启发式	375
12.6.5 启发式算法小结	376
12.7 影响力最大化问题延伸与变形	376
12.7.1 影响力最大化问题延伸	376
12.7.2 影响力最大化问题变形	377
12.8 本章小结	378
参考文献	379

第1章 绪 论

1.1 社交网络及其发展

1.1.1 社交网络的起源

自从人类诞生以来，就在一起共同耕种、狩猎、劳作，从而形成了社会。随着社会发展和交流的不断深入，人们不可避免地与他人产生着关系，社会关系从简单的血缘、亲属关系发展为朋友关系、生产关系、劳动关系、社会交往关系等。社会成员之间由于在工作、学习、生活、娱乐等活动中的相互作用而逐渐形成了某种稳定的关系，进而构成了社交网络。正如 Mickenberg（米肯伯格）和 Dugan（杜甘）在 1995 年所说的那样，“We all connect, like a net we cannot see”^[1]。

在维基百科中，社交网络（Social Network）被定义为：“由许多节点构成的一种社会结构。节点通常是指个人或组织，而社交网络代表着各种社会关系。在社交网络中，成员之间因为互动而形成相对稳定的关系体系，这种关系体系可以是朋友关系、同学关系，也可以是生意伙伴关系抑或种族信仰关系。通过这些关系，社交网络把从偶然相识的泛泛之交到紧密结合的家庭关系，再到社会活动中的各种人们组织串联起来”^[2]。由于社交网络中存在着各种社会关系，社会组织或个人之间的社交图形结构往往是非常复杂的^[2]。复杂的关系结构影响着成员之间的互动和联系，进而影响着人们的社会行为。

从历史维度来看，社交网络是把人与互联网深度融合的中坚力量。随着工业化、城市化的进行和新兴通讯技术的兴起，社会呈现出越来越网络化的趋势，2012 年，Lee Rainie（李·雷尼）和 Barry Wellman（巴里·韦尔曼）在其新著《网络化：新的社会操作系统》（*Networked: The New Social Operating System*）中将社会网络革命、移动革命与互联网革命并列为新时期影响人类社会的三大革命^[3]。当前，互联网作为一个能够相互交流、相互沟通、相互参与的互动平台，其发展早已超越了当初 ARPANET 的军事和技术目的。而社交网络几乎涵盖了以人类社交为核心的所有网络服务形式，这使得互联网从研究部门、学校、政府/商业应用平台扩展成一个供人类社会建立和发展关系、相互交流的工具。

1.1.2 从社会学角度看社交网络的发展历程

1842 年，法国社会学家和实证主义哲学家 Auguste Comte（奥古斯特·孔德，1798—

1857 年)提出了“社会学”这一术语^[4],明确了社会学领域研究的两个主要方面:社会静力学和动力学。他首次提出根据社会行动者之间的相互联系来考察社会。在 Auguste Comte 看来,个人是社会最基本的构成要素,而个人的特性影响着社会的特性。Auguste Comte 的这一贡献推动了社会学成为了一门科学。

法国社会学家 Gustave Le Bon (古斯塔夫·勒庞, 1841—1931 年)主张从群体视角来观察社会成员之间的关系^[5],并重点关注了信息在群体成员之间的流动。Gustave Le Bon 指出,当个人变成群体中的一员时,他们将失去其个体身份。作为群体中的成员,人们模仿其周围的人,使得群体的思想和行为广泛传播并深刻影响着个体的思想和行为。

从社会学视角来看,社交网络发端于德国社会学家 Georg Simmel (格奥尔格·齐美尔, 1858—1918 年)的“社会学”理论^[6]。20 世纪 60 年代,随着冷战的开始和西方普遍出现的社会动乱,Georg Simmel 的“社会学”理论在西方广为发展,并在 20 世纪 70 年代趋于成熟。在半个世纪的发展历程中,社会学中的“社会结构”理论^[7]在心理学、社会计量学、社会学、人类学、数学、统计学、概率论等不同领域得到广泛应用,逐渐形成了一套系统的理论、方法和技术,成为一种重要的社会结构研究范式。

“社交网络”这一概念的兴起,正是源于其对社会关系互动的恰当描述。一个多世纪以来,社会学家都在使用“社交网络”这一隐喻表示不同尺度上的各种复杂的社会关系。然而,直到 1950 年,这一词汇才被系统化地表示一种不同于传统意义上的有边界的社会群体(比如村庄和家庭)和将人看做分离的个体的社会类别(比如性别与种族)。例如,如果将咖啡馆里的人、一起工作的同事或者在互联网上互动的人认为是一个有边界的社会群体,就会错误地认为他们因为相互认识而对共同群体有归属感。现实情况是人们不断地进入或退出一个社交网络,而这种社交网络中又具有复杂的结构。

1988 年,加拿大著名社会学家 Barry Wellman (巴里·韦尔曼)提出了较为成熟的社交网络定义。Barry Wellman 认为,社交网络是由某些个体间的社会关系构成的相对稳定的系统^[8],即把“网络”看做联结行动者的一系列社会联系或社会关系,它们相对稳定的关系模式构成了社会结构。随着应用范围的不断拓展,社交网络的概念已超越了人际关系的范畴,网络的行动者既可以是个人,也可以是家庭、部门和组织等集合单位。

早期的社交网络主要是指个人之间通过相识关系或工作关系建立起来的线下“社交网络”,如科研合作关系网和演员合作网等关系网络。其中,社会学家 Wayne Zachary (韦恩·扎卡里)在 20 世纪 70 年代所构建的某大学空手道俱乐部 34 位成员的社会关系网络是早期社交网络的典型代表^[9]。

随着互联网的发展,由于网络结构所反映的地域性因素减弱了,使得传统线下社交网络中的地域约束越来越减弱,跨地域的线上社会关系成为社交网络的重要形式。2003 年之后,随着 Web 2.0 技术的发展,在线社交网络媒体受到人们的广泛关注,人们相继在以 Facebook (脸谱)、Twitter (推特)、博客、社交网站等为典型代表的在线社交媒体平台上建立账号,并相互添加线下认识的好友。由此开始,线下社交网络开始推广到网络环境,这使得在线社交网络成为人们网络生活不可缺少的沟通工具。

1.1.3 从人类学角度看社交网络的发展历程

从人类学角度来看,人类学关于社交网络的早期研究主要分为非工业社会与工业社会两种社交网络模式研究。

首先,在非工业社会的社交网络研究方面,美国人类学家 Lewis Henry Morgan (路易斯·亨利·摩尔根,1818—1881年)的亲属关系研究最具代表性^[10]。在研究易洛魁(Iroquois)部落的过程中,Lewis Henry Morgan 就曾发现易洛魁人亲属称谓与现代美国社会的亲属称谓截然不同,而其他印第安人与易洛魁人的亲属称谓制度却大体相同。Lewis Henry Morgan 于1871年出版了《人类家族的血亲和姻亲制度》一书,提出亲属称谓并不是普遍的,不同的文化具有不同的亲属称谓体系。Lewis Henry Morgan 解决了文化与亲属称谓体系之间的关系问题。

英国人类学家 Radcliffe-Brown (拉德克利夫-布朗,1881—1955年)继承和发展了 Lewis Henry Morgan 的相关理论。他指出,所谓的亲属制度就是社会关系的网络,它是总的社会关系网络的一个组成部分,这种总的社会关系网络被称为社会结构。Alfred Radcliffe-Brown 主张利用社交网络分析的方法来分析亲属关系,并逐步形成结构—功能理论^[11],促成了让社交网络概念成为英国人类学界的主导性概念。

然而,传统的亲属关系研究存在两大缺陷:一是只看到亲属关系中单个的成分,忽略了成分之间的相互关系;二是关注亲属关系的起源和历史演进,忽略了横向的结构研究。为此,法国人类学家 Lévi-Strauss (列维·斯特劳斯,1908—2009年)提出从结构性的视角研究亲属关系的方法^[12],总结了八元四组关键亲属:夫妻,兄妹,父子,甥舅的二元对立关系,他的这种方法勾勒出了亲属关系背后深层的、普遍的社会网络结构。

在亲属关系研究之外,随着殖民主义的瓦解和初民社会的渐行渐远,人类学的研究重心开始转向广大的农业社会和工业社会。关于社会关系的研究也从亲属关系延伸至乡村、城市、企业、组织中的各种社会关系。1929年,美国人类学家 William Lloyd Warner (威廉·劳埃德-沃纳,1847—1928年)组织了“扬基城”研究^[13],尝试将自己以前研究澳大利亚土著的方法运用于美国城镇的研究。William Lloyd Warner 提出了一种关注社会分层、个体互动和人际网络的方法,该研究影响巨大,对后来的研究者产生了深远的影响。

其次,在工业社会模式下,英国人类学史上重要的人类学家 Max Gluckman (马克斯·格拉克曼,1911—1975年)的研究最具代表性^[14]。Max Gluckman 对5个工厂进行了考察,他一方面秉承了 Radcliffe-Brown 对社会结构的关注,另一方面开始关注车间所处的更广泛的社会情境,并以此作为研究的重点。Max Gluckman 研究发现,5个工厂均存在非正式组织,虽然工人与管理者之间的关系模式存在差异,但不像之前研究所认为的那样是自发合作的社会关系,而是存在着诸多冲突。

除了现代工业社会,社交网络也被用于对第三世界国家都市化的研究。澳大利亚人类学家 Bruce Kapferer (布鲁斯·卡普菲勒,1940—)研究了一家采矿企业里的劳工冲突^[15]。他的研究考虑了劳工之间的互动关系以及劳工组成的社交网络等因素,分析了劳工互动

和劳工社交网络与劳工罢工等工厂重大事件的联系。

1.2 在线社交网络发展

1.2.1 在线社交网络的概念

随着互联网技术的飞速发展，人们将早期社会性网络的概念引入互联网，创立了面向社会性网络服务的在线社交网络（Social Networking Services, SNS，翻译为社会性网络服务或社会化网络服务）。在线社交网络的含义中包括硬件、软件、服务及应用四个方面，由于四字构成的词组更符合中国人的构词习惯，因此人们习惯上用社交网络来代指 SNS。

根据欧盟关于社会计算的研究报告^[16]，在线社交网络可分为四类：

(1) 即时消息类应用，是一种提供在线实时通信的平台，如 MSN、QQ、AIM、飞信、微信等，其具有双向认证与实时推送的特点；

(2) 在线社交类应用，是一种提供在线社交关系的平台，如 Facebook、Google+（谷歌）、人人网、开心网、QQ 空间等，其具有双向认证与非实时获取的特点；

(3) 微博类应用，是一种提供双向发布短信息的平台，如 Twitter、新浪微博、腾讯微博、网易微博、搜狐微博等，其具有单向认证与实时推送的特点；

(4) 共享空间等其他类应用，是其他可以相互沟通但结合不紧密的 Web 2.0 应用，如论坛、博客、BBS、视频分享、社会书签、在线购物等，其具有单向认证与非实时获取的特点。

在线社交网络是一种在信息网络上由社会个体集合及个体之间的连接关系构成的社会性结构。这种社会结构主要包含关系结构、网络群体与网络信息三个要素。其中，社交网络的关系结构是社会个体成员之间通过社会关系结成的网络系统。个体也称为节点，可以是组织、个人、网络 ID 等不同含义的实体或虚拟个体；而个体间的相互关系可以是亲友、动作行为、收发消息等多种关系。基于这些关系，社交网络中的个体自组织成各种各样的虚拟社区，虚拟社区是社交网络的一个子集，具有虚拟社区内节点之间关联密切、不同虚拟社区的节点间关联稀疏等特点。在上述关系的基础上，社交网络中个体与个体之间、个体与群体之间、群体与群体之间传递着各种各样的信息，这种信息传递的不断迭代便是社交网络中的信息传播。受网络结构和信息传递的影响，个体就某个事件在某个虚拟空间聚合或集中，相互影响、作用、依赖，有目的地以类似方式进行行动，这便形成了社交网络的群体行为。

社交网络是 Web 2.0 时代的典型应用，也是 Web 2.0 时代社会性、主动性特征的典型代表。Web 2.0 时代的互联网正在进行着巨大的改变——从一系列网站到一个成熟的为最终用户提供网络应用的服务平台^[17]。这些平台中的内容因为每位用户的参与而产生，参与所产生的个人化内容，借由人与人的分享，形成了现在 Web 2.0 的世界。Facebook 和

Twitter 等社交网站就是 Web 2.0 时代用户主动参与思想的杰作。Facebook 上线不足 8 年, 已拥有超过 14 亿用户, Twitter 用户数也已超过 5 亿。根据各自官方网站的报告, 截至 2013 年 3 月, 新浪微博用户数已超过 5.56 亿。

伴随着 Web 2.0 应用鼎盛期的到来, Web 3.0 也开始崭露头角。互联网之父 Tim Berners-Lee (蒂姆·伯纳斯-李) 指出^[18], 当互联网发展成为一整张语义网涵盖着大量数据, 人们可以访问这难以置信的数据资源, 即为 Web 3.0。2010 年 11 月 16 的网络高峰会上 Mary Meeker (玛丽·米克) 指出, Web 3.0 由“社交网络、移动设备和搜索 (Social Networking, Mobile and Search)” 所组成^[19]。综合来看, Web 3.0 的特点包括: 将互联网转化为数据库, 赋予搜索引擎以智能, 实现语义网络与面向服务的架构 (SOA), 将互联网转化为一系列三维空间 (如人、时间、信息) 等。可以这样说, 谁能够引领 Web 3.0, 谁就是网络的下一任主角。

1.2.2 在线社交网络的特点

与传统的 Web 应用及信息媒体应用相比, 在线社交网络主要具有以下新特点。

(1) 迅捷性: 信息的发布和接收异常简便、迅速。用户可以通过手机和浏览器随时随地发布和接收信息。

(2) 蔓延性: “核裂变” 式的信息传播。消息一经发布即被系统推送到所有关注者。一旦被其转发, 又立即传播到下一批关注者, 呈现“核裂变” 式的几何级数扩散态势, 为普通网民创造了意见扩散的渠道。

(3) 平等性: 人人都有机会成为意见领袖。与传统媒体的非对称信息发布与信息受体相比, 社交网络服务中的广大网民都有机会通过社交网络形成意见领袖, 并在突发事件的产生、发酵、传播、爆炒等环节中起到重要作用。

(4) 自组织性: 呈现自媒体形态、能够快速形成虚拟社区。这是由于社交网络中的个体都具有提供、发布信息的手段和渠道, 并依靠社交网络快速传播信息、快速形成网上的虚拟社区。

1.2.3 在线社交网络的发展

从一定意义上来看, 在线社交网络的发展其实是源于人们网络社交的需要。图 1-1 给出了在线社交网络的发展脉络。1838 年, Samuel F.B. Morse (萨缪尔·摩尔斯) 发明了莫斯电码, 电报成为长距离沟通的途径, 使得人类社交网络的在线实时远距离交往成为可能。1876 年, 贝尔发明了可以用于电信网络的线路交换方式, 第一台电话正式产生。1969 年, 美国发明 ARPANET, 世界上第一个分组交换网络正式运营。1971 年, 美国高级研究计划署研究员发出第一封电子邮件。1990 年, 欧洲核子研究中心研究员开发出万维网 (WWW) 新协议, 标志着现代互联网的正式诞生。1994 年, 雅虎地球村是第一个互联网在线社区, 网站中的用户可以聊天及分析咨询。1999 年, 腾讯 QQ 上线。2002 年,