



青年学者文库

虚拟社会网络下集群行为 感知与规律研究

胡昌龙 著



WUHAN UNIVERSITY PRESS
武汉大学出版社

教育部人文社会科学研究《虚拟社会网络环境下微博的集群行为感知与规律研究》12YJC610000项目成果



青年学者文库

虚拟社会网络下集群行为 感知与规律研究

胡昌龙 著



WUHAN UNIVERSITY PRESS

武汉大学出版社

图书在版编目(CIP)数据

虚拟社会网络下集群行为感知与规律研究/胡昌龙著. —武汉: 武汉大学出版社, 2016. 11

青年学者文库

ISBN 978-7-307-18782-5

I. 虚… II. 胡… III. 虚拟网络—社会人类学—研究 IV. C912.4

中国版本图书馆 CIP 数据核字(2016)第 272334 号

责任编辑:叶玲利

责任校对:汪欣怡

版式设计:马 佳

出版发行: **武汉大学出版社** (430072 武昌 珞珈山)

(电子邮件: cbs22@whu.edu.cn 网址: www.wdp.com.cn)

印刷: 虎彩印艺股份有限公司

开本: 787 × 1092 1/16 印张: 12.25 字数: 289 千字 插页: 1

版次: 2016 年 11 月第 1 版 2016 年 11 月第 1 次印刷

ISBN 978-7-307-18782-5 定价: 45.00 元

版权所有, 不得翻印; 凡购我社的图书, 如有质量问题, 请与当地图书销售部门联系调换。

前言

大数据不仅是一种应用性很强的工具，更是一种重要的思维方法。大数据思维就是将传统的思维方法转变到以大数据为基础，以大数据分析为手段的数据驱动思维方法，即以大数据来认识事物、把握事物。另外，大数据也是一种系统思维观，要求在数据证据的基础上观大局，大数据的思维要求认识大数据、认可大数据、使用大数据，按大数据揭示的事物发展规律办事。大数据技术的战略意义不在于掌握庞大的数据信息，而在于在大数据思维的指导下，利用能够处理大数据的系统对大数据进行专业化分析处理，实现大数据的战略意义。大数据的思维除了带来思维方式的转变，还必然带来方式和流程的转变，现有的用户网络消费行为的聚合管理与决策研究显然不能适应大数据的应用。大数据不仅是一种应用，同时也是一项很重要的技术，涉及数据采集技术、数据挖掘技术、数据可视化技术、相应的软件技术以及机器学习、统计分析等相关学科的知识。

网络社会系统有着复杂多样的特质，加之系统内部和外部存在的异质性作用，使得虚拟和谐以及社会系统的建构过程变得异乎寻常的艰难。为了实现网络社会较快速度的合理建构，政府和企业应该发挥作用，同时需要对社会群体行为的感知与引导进行研究，使网络社会在一个有序的机制下不断发展和完善。就网络群体性事件的信息本质而言，它是在一定的信息环境中，人们的一种具有显著特征的复杂信息行为。而网络群体性事件的源信息处理，对各级党政机关预防与解决网络群体性事件具有极大的现实意义，是找出适合基于目前移动互联网特点的网络群体性事件的防范策略、舆情疏导、网络群体性事件的处置、健康网络环境的营造以及网络突发应急处理平台构建的首要前期任务。

基于作者主持的教育部人文社科青年基金项目“虚拟社会网络环境下微博的集群行为感知与规律研究(12YJC870009)”，作者构建了一套拥有语义网存储、推演计算能力的语义网格平台，可以支持将大规模的语义网分布地存储在不同地点、不同类型的数据库中，处理查询时牵涉多个子语义网以及对应的多个数据库，并关联不同语义网里面的各种语义对象。在此语义网格平台下如同每个数据库都有自己的模式(Schema)，每个语义网也有自己的本体(Ontology)来规定自己网络中节点以及连接的类型。现正在研究不同语义网本体的关联和数据库模式的映射，研究语义对象出现的环境信息的表示，包括出现的时间、空间属性以及与其他对象之间的关联。但是，在时空匹配以及用户特性的表示方面还有很多不足，还需要研究在语义信息管理过程中如何保护个体隐私的技术，并研究将语义网格平台向云计算平台上的功能迁移升级。多维大数据有关的语义信息上国内外还没有较好的大数据表示处理方式，未来的研究项目中或许可以涉及并尝试实现。

关于用户的集群仿真主要研究基于主体的仿真建模工具 SWARM、Netlogo、大数据分析工具 KNIME 等，并已经具备构建了一定数据级用户的系统“伪实时”仿真模型能力

(10MB/s), 希望在结合现有的数据研究仿真算法后构建仿真的软件系统, 在现有的可视化大数据平台中实时展示。但现在社会网络用户行为方面, 大型和小微型用户研究都有很多方式, 对于中型规模的用户行为研究方面, 国内外都缺乏可行的处理方式。

本书从集群行为感知、模式和规律挖掘、管理引导机制三个层次的基础行为感知和建模仿真展开, 详细介绍了多种移动大数据环境下的数据分析、处理工具在实际中的应用, 并将研究过程中的技术分章节描述介绍。

本书共有 6 章, 各章的主要内容如下:

第 1 章介绍了虚拟社会网络的微博集群行为有序管理所涉及感知、分析、引导三个阶段的关键技术以及需解决三个关键研究内容的方法论。

第 2 章介绍了作为网络行为获取基础的聚集搜索引擎的结构原理和实现。聚集搜索引擎是针对某一个行业的专业搜索引擎, 是搜索引擎的细分和延伸。本章对于聚焦搜索的算法、系统架构、功能、工作原理都做了描述说明。

第 3 章介绍了使用 NetLogo 实现网络用户集群行为动力学演化多种模型的动态仿真模拟, 提供了网络中的社会影响模型、网络中的正负口碑模型、有界自信下的连续观点动力学模型、社会网络上的新闻传播模型以及网络上的文化传播模型的 NetLogo 实现。

第 4 章 Ontology 数据管理是一个突出的问题, 特别是语义研究对本体的存储和管理等各方面提出了更高的要求, 需要构建一个独立于具体应用、查询效率高、具有良好的通用性、互操作性的本体存储管理系统。本章论述了基于网络本体语言的图形数据库并如何存储语义网信息。

第 5 章提出了基于兴趣图谱、社交网络和现实世界的多网复合模型, 通过在不同层面的映射来感知用户行为, 并提出了一个数据整合系统, 包括了系统的设计与实现, 并且做出了实例分析。

第 6 章分析了知识网络的发展现状, 从文本分析的角度阐释知识网络的获取, 并从现有的可视化技术中得到启示, 利用数据挖掘、数据质量、数据清洗等相关理论及技术来实现知识网络的可视化, 介绍了在经典的 DIKW 模型基础上进行扩展, 重新界定知识范畴, 构造适合可视化的相关知识模型的过程。

本书包含丰富的图表、程序, 其中大部分是作者根据多年实践总结出来的, 图示简练准确、易于理解, 并附有解决问题的具体步骤方法和相应程序。

本书可以作为从事互联网行为感知、大数据分析技术人员的参考资料。由于时间仓促, 书中难免有不足之处, 恳请广大读者指正。

目 录

1 基于虚拟社会网络下集群行为感知介绍	1
1.1 研究基于虚拟社会网络下集群行为感知	1
1.1.1 虚拟社会网络下集群行为感知	1
1.1.2 国内外研究现状	2
1.2 基于虚拟社会网络下集群行为感知研究和方法	4
1.2.1 研究的内容	4
1.2.2 研究方法	10
2 中文搜索引擎的构建	12
2.1 搜索引擎现状	12
2.1.1 搜索引擎发展	12
2.1.2 国内外研究现状	13
2.2 搜索引擎的基本结构	14
2.2.1 搜索引擎组成	14
2.2.2 搜索引擎机制	15
2.2.3 网络信息检索工具的类型	17
2.2.4 搜索引擎的检索功能	18
2.2.5 搜索引擎体系结构	19
2.3 中文分词关键技术	20
2.3.1 中文分词原理	20
2.3.2 中文分词算法	21
2.3.3 自然语言处理	22
2.4 系统架构设计	25
2.4.1 Input 组件	26
2.4.2 Filter 组件	27
2.4.3 Output 组件	28
2.4.4 Util 组件	29
2.5 系统实现	30
2.5.1 UI 设计	30
2.5.2 系统 API	31
2.5.3 聚焦搜索引擎算法	33

2.5.4	代码实现	34
2.6	聚焦搜索系统测试	39
2.6.1	JUnit 特性	39
2.6.2	框架组成	40
2.6.3	测试示例	40
2.6.4	中文分词测试	41
2.7	总结及展望	46
2.7.1	本系统总结	46
2.7.2	研究展望	46
3	网络用户集群行为动力学演化	49
3.1	网络用户集群行为动力学演化背景	49
3.1.1	研究的背景及理论意义	49
3.1.2	国内外研究现状	50
3.1.3	本课题研究内容和研究方法	52
3.2	网络用户集群行为动力学演化介绍	54
3.2.1	动力学演化研究方法	54
3.2.2	集群行为	55
3.2.3	复杂性科学与复杂系统理论	56
3.2.4	网络信息的知识发现和数据挖掘	57
3.2.5	复杂网络	57
3.2.6	演化模型	60
3.3	仿真平台与模型介绍	62
3.3.1	仿真平台	62
3.3.2	模型框架	64
3.4	人类行为动力学模型	65
3.4.1	社会网络算法	65
3.4.2	有界自信模型	67
3.4.3	影响因素	69
3.5	动态仿真模型实现	72
3.5.1	网络中的社会影响模型	72
3.5.2	网络中的正负口碑模型	74
3.5.3	有界自信下的连续观点动力学模型	76
3.5.4	社会网络上的新闻传播模型	79
3.5.5	网络上的文化传播模型	81
3.5.6	整体概述	83
3.6	网络用户集群行为仿真总结	84

4 语义网存储系统.....	85
4.1 语义网存储系统研究目的、意义及发展现状.....	85
4.2 语义网理论基础.....	87
4.2.1 语义网介绍.....	87
4.2.2 本体的定义.....	91
4.2.3 本体的描述语言.....	91
4.2.4 本体的存储方法.....	92
4.2.5 图形数据库.....	93
4.3 语义网存储系统的概要设计.....	94
4.3.1 系统基本需求分析.....	94
4.3.2 系统功能图.....	95
4.4 语义网存储系统的详细设计.....	96
4.4.1 OWL 本体存储模式设计	96
4.4.2 关系数据库模式与本体间映射实现.....	97
4.4.3 关系数据库转换图数据库实现	103
4.4.4 系统类图	104
4.5 系统界面展示	108
4.5.1 Web 界面展示及使用说明	108
4.5.2 后台管理界面展示及使用说明	110
4.6 应用实例展示	112
4.6.1 基于社会基本邻里关系的发现	112
4.6.2 电影数据库	113
4.7 全章总结及研究展望	116
5 网络用户行为感知	118
5.1 网络用户行为感知导论	118
5.1.1 网络用户行为感知背景及意义	118
5.1.2 网络用户行为感知国内外研究现状	119
5.1.3 研究内容和研究方法	120
5.2 网络用户及网络用户行为	120
5.2.1 网络用户	120
5.2.2 网络用户行为	121
5.3 网络用户行为分析的研究	124
5.3.1 网络用户行为的应用场景	124
5.3.2 网络用户行为需求分析	125
5.3.3 网络用户行为分析的方法	130
5.3.4 行为分析相关理论	132
5.4 用户行为感知系统的实现	136

5.4.1	系统简介	136
5.4.2	系统架构特点	137
5.4.3	数据结构	137
5.4.4	节点	137
5.5	系统相关界面及效果展示	139
5.5.1	系统界面	139
5.5.2	节点状态	140
5.5.3	端口	141
5.5.4	实例 1：访问日志处理	141
5.5.5	实例 2：静态 IP 地址的定位	146
5.6	全章总结及研究展望	151
5.6.1	全章总结	151
5.6.2	研究展望	152
6	知识网络可视化呈现	153
6.1	知识网络背景和意义	153
6.1.1	知识网络可视化研究现状	153
6.1.2	国内外知识网络可视化研究现状	155
6.2	相关理论与技术综述	157
6.2.1	DIKW 体系及扩展	157
6.2.2	文本可视化	159
6.2.3	数据挖掘技术	161
6.2.4	数据质量	162
6.2.5	文本挖掘	163
6.3	知识网络可视化系统概要设计	166
6.3.1	知识网络可视化系统需求分析	166
6.3.2	系统实现技术数据挖掘	167
6.3.3	知识网络可视化系统功能模块介绍	169
6.4	知识网络可视化系统详细设计	170
6.4.1	知识网络可视化系统架构设计	170
6.4.2	业务流程图	172
6.4.3	信息抽取过程设计与实现	173
6.5	知识网络可视化系统实现	175
6.5.1	知识网络可视化系统 UI 设计	175
6.5.2	模型构建	178
6.5.3	工作流设计	179
6.6	总结与展望	183
	参考文献	184

1 基于虚拟社会网络下集群行为感知介绍

1.1 研究基于虚拟社会网络下集群行为感知

1.1.1 虚拟社会网络下集群行为感知

虚拟社会网络与现实相对应，它带有假设、虚幻、不真实的意味，被称为第四媒体的互联网以及互联网的无限延伸：电子信箱、BBS、个人主页、即时通信工具、手机短信等给现实社会实际存在的个人、组织带来了虚拟的生存空间，诸多的交往活动因技术的强大支持构成了虚拟社会。因此，虚拟社会网络是一种源于计算机网络技术不断普及和发展而形成的一种新的社会形态。

中国互联网信息中心 CNNIC 第 38 次《中国互联网络发展状况统计报告》中指出：截至 2016 年 6 月，我国网络新闻用户规模为 5.79 亿，网民使用比例为 81.6%。与 2015 年年底相比，用户规模增加 1 487 万，半年增长率为 2.6%。其中，手机网络新闻用户规模为 5.18 亿，占移动网民的 78.9%，较 2015 年年底规模增加 3 635 万，增长率为 7.5%。移动端已经成为网民获取新闻的最主要渠道，而移动互联网发展带来的信息膨胀和碎片化，则加速了网络用户对于个性化、垂直化新闻资讯的需求。同时，移动互联网的媒体属性日益增强，对新闻媒体也提出了更高的要求。

《纽约时报》在评论国内微博热时表示，“这是一个观察中国正在发生什么的实时检测系统”。如今微博越来越成为公众意见表达的渠道，连接着现代社会的神经。上海交大人文艺术研究院副院长谢耘耕教授说：“社会舆情就是中国社会的体温计，它可以测量社会的方方面面，测量社会机体是否健康。”中央组织部党建研究所世界政党研究室主任韩久根表示，互联网等新兴技术的兴起，使社会舆情发生了很多变化。这包括参与人数的剧增，也包括人员结构的多样化。以前年轻人是网络舆情的主体，而现在科技发展以后，手机上网很方便，很多老年人也加入其中。

微博的快速崛起已经成为网络舆情热点的重要推动力量。微博对网络舆情热点形成的巨大推动作用，不仅表现在它拓展了热点事件的信息源头，更主要的是它大大加速了网络舆情的形成发展进程。微博实时、持续的移动传播方式，以人际关系网络为基础的传播机制，使得事件的广而告之与舆情的酝酿、传递过程几乎是同时启动的，信息传递的也不仅仅是事实性的信息，人们的意见、情绪甚至谣言等通过转发或评论等方式也随之得以传播，在信息传播达到一定规模后，舆情热点也随之形成。如 2011 年 6 月的官员微博开房事件，2011 年 7 月的郭美美事件等，都是典型的基于虚拟社会网络下微博的集群行为。

可以预见，快速发展的微博平台对未来的网络舆情也必将产生更加广泛、更加深远的影响。

中国社科院 2011 年《舆情蓝皮书》建议：政府应该加强舆情信息的监测与分析，提升舆情的预警与研判能力；转变舆情管理观念，合作而非对抗；提升运用新媒体的能力，拓宽信息传播渠道；完善舆情的利益协调机制，注重与受害方的沟通；提高议题管理能力，有效引导舆论；重视社会舆情热点事件之后的形象修复与管理。政府部门在应对社会舆情事件的时候，已经有了一些尝试，比如人民网推出“直通中南海——中央领导人和中央机构留言板”。截至 2011 年 7 月底湖北省已经建立基层舆情信息直报点总共有五个单位，分别是中共麻城市委宣传部、中共应城市委宣传部、中共当阳市委市委宣传部、中共赤壁市委宣传部。

当下的虚拟社会网络并没有与现实世界泾渭分明。它与现实世界同步、互动等特征，有时表现得极其强烈。应该说，舆情的产生、表达和效果是一个社会进行社会建设和社会管理的正常范畴。舆情与民情相连，在现代社会中指民众的社会政治态度。传统社会条件下，人际传播受到传播速度和范围的制约，出现社会矛盾时，社会群体的组织动员一般会局限在一定的地域内，但现在的网络 and 手机短信解决了技术上的阻隔，可以迅速地将松散的、潜在的行动者动员组织起来。以往社会矛盾发生时，经常是与直接利益者产生冲突，而非直接利益者则很少介入，但在网络环境下，既有直接利益者的群体冲突事件，也有非直接利益者大量参与的群体冲突事件。从这两年由网络舆论引发的社会事件看，酿成大规模冲突的事件往往由小事情引发，由个体矛盾引发。

在舆情力量不断成长的今天，既要充分发挥新闻媒体的舆论主导作用，形成正面舆论宣传强势，同时应该更加关注和反映民众的诉求和声音，畅通社情民意表达渠道。把舆情引导工作贯穿于舆情生成、舆情表达、舆情扩散、舆情影响、舆情消退等舆情产生发展的全过程。就网络群体性事件的信息本质而言，它是在一定的信息环境中人们的一种具有显著特征的复杂信息行为，从网络群体性事件的源信息处理，对各级党政机关预防与解决网络群体性事件具有极大地现实意义；是找出适合基于目前互联网特点的网络群体性事件的防范策略，为舆情疏导、网络群体性事件的处置、营造健康网络环境，构建网络突发应急处理平台的首要前期任务。所以，基于虚拟社会网络下微博的集群行为感知与规律研究有效解决当前网络群体性事件的，特别是针对现在由微博引发的集群行为的应对处理提供理论上和实际上辅助，具有非常重要的实际应用价值。

1.1.2 国内外研究现状

目前国内外主要是对于社交网络的感知进行分析、研究。但是在微博同以往的社交网络所使用的 SNS 系统来说，其具有了很多新的特点。

国外，Boyd 和 Ellison 研究了社交网络的演化历史，给出了社交网络的定义，分析了社交网络的研究热点。Acquisti 和 Gross 以 Facebook 为例，探讨了社交网络中的感知、信息共享和隐私问题，指出用户对社交网络性质、结构和个人隐私设置等内容感知的重要性。Schmidt 和 Gellersen 通过整合用户浏览 Web 页面时的信息，感知个人页面的浏览者。Steiny 和 Harri 从社交网络搜索、创新和知识分享的角度分析了社交网络感知，认为社交

网络感知是一个多层次的感念,包括用户对周围网络环境的感知、感知的策略和实现感知策略的过程及工具。之后,Steiny 又引入了社会上下文的概念,通过社交网络感知可以找到用户的位置,发现用户的行为、态度、世界观及其社会关系。Misook Heo 提出了在线学习环境中基于关系的社会感知,学生可根据感知信息保护个人隐私。Susanne 和 Ellen 研究了弹性工作环境下计算机支持的社会感知问题。Buder 和 Bodemer 将群体感知引入到 CSCL 领域以支持在线讨论,并实现了感知工具。Newma 从论文的合著关系入手,对来自数据库 MEDLINE, NCSTR 等的论文集进行分析,统计了作者数量、平均论文数、合著者数量等,发现了科研合作网中合著者之间的聚类关系,分析了科研合作网的结构。Miki 等通过研究者之间的超链接来表示学术社区中的知识关系,并建立了本体描述,发现了学术社区中研究群体的结构。Matsuo 等基于搜索引擎建立了一个社交网络提取系统 POLYPHONET,通过 Google 搜索引擎实现了关系的提取、群体结构的检测。Mika 建立了 Flink 系统,从网页、FOAF 个人资料等信息资源提取个人信息,并进行基于语义的推理,实现了在线社会网络的提取、聚集、分析和可视化。但西方的信息环境、自然语言和网络群体性事件的自身特征都与我国现实情况大不相同,很难把国外的做法直接应用到国内。

而国内在网络群体性事件方面的研究还处于起步阶段。北京师范大学师曾志博士研究了网络群体性事件形成机制的理论基础。山东省委党校杨久华博士论述了网络群体性事件发生的模式原因及其防范方法。南京大学新闻传播学院杜骏飞教授对网络群体事件的类型进行了辨析。南京大学政府管理学院汪建昌博士分析了网络群体性事件的舆论生成与政府决策。上海交通大学彭博在其硕士论文中,对网络群体性事件的规律和防范策略进行了较深入研究。

但这些研究基本上是以公共管理和传播的专业角度对网络群体性事件进行的研究,缺乏对网络群体性事件的信息本质的研究。四川大学计算机学院杨频博士提出了一种网络舆情的定量分析方法。上海交通大学信息安全工程学院金晓鸥在其硕士论文中,对互联网舆情信息获取与分析展开了研究。中国人民公安大学刘剑宇基于 Web 挖掘技术,在网络舆情预警中如何应用提出了方法。中山大学资讯管理系黄晓斌教授研究了文本挖掘在网络舆情信息分析中的应用。华东师范大学信息学系王伟基于聚类对网络舆情热点发现进行了分析。中国科技大学郑魁博士对网络舆情热点信息的自动发现提出了方法。国防大学刘常昱博士后基于小世界网络和不对称人际影响对舆论涌现模型进行了研究。中国科学院方薇采用元胞自动机来构建网络舆情的传播模型。

国内外学者与课题相关的研究较多,但是移动宽带互联网背景下集群行为的语义聚融、集群行为的动力学演化、集群行为有序引导的综合集成决策支持以及大尺度跨网跨媒体社会集群行为感知、网络社会海量集群行为信息的聚融与管理、网络社会集群行为的社会结构挖掘与演化、网络社会集群行为社会学研究及定量指标、移动互联网环境下网络社会集群行为有序引导机制、网络社会集群行为感知与引导的应用平台研究尚有很大空间。

以上研究较好地分析了网络群体性事件的信息技术成因与处理方法,但是都比较缺乏对网络群体性事件的信息产生过程进行深入研究,特别是对网络群体性事件的源信息没有很好的界定,没有对网络群体性事件的源信息特征进行分析和建模,源信息生命周期的演化规律以及如何对基于虚拟社会网络下微博的集群行为感知与规律的特征挖掘方法与演化

算法，这些关键问题是目前研究所没有触及的，这也是本项目研究的立足点。

1.2 基于虚拟社会网络下集群行为感知研究和方法

1.2.1 研究的内容

为突破基于虚拟社会网络的微博集群行为有序管理所涉及感知、分析、引导三个阶段的关键技术，需解决三个关键研究内容，见图 1-1。

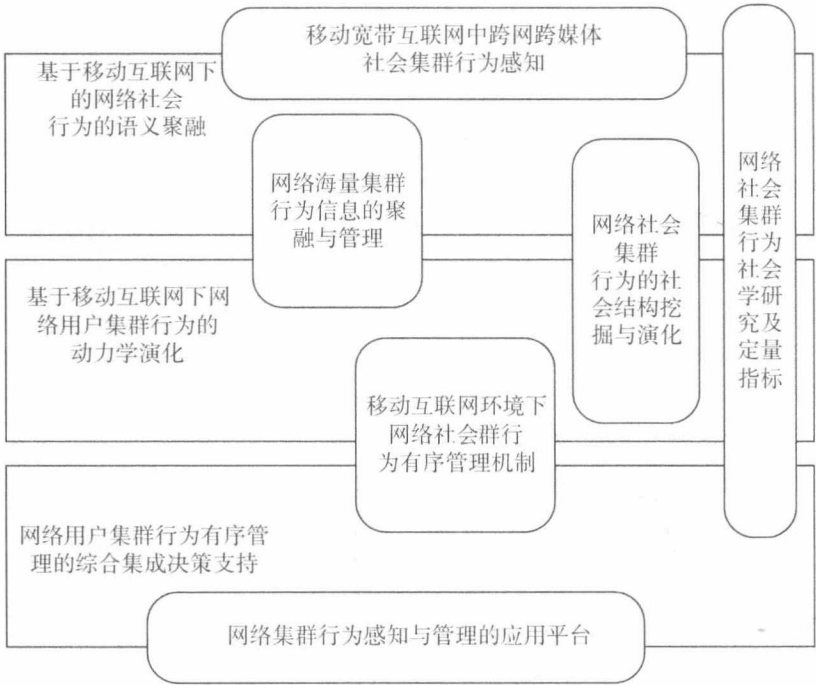


图 1-1 三大研究内容同六个子研究内容间包容关系图

研究内容一：虚拟社会网络下微博集群行为的语义聚融。该研究旨在揭示虚拟社会网络下的微博集群行为信息海量异构数据之间的内在语义关联与层次，为后续的分析和调控提供支撑。微博中包含了文本、图像、视频、运动和空间等呈现多源、异构和海量特性的信息，这些信息承载的主体行为发生的空间分布广、持续时间长、冗余信息多。为了有效实现虚拟社会网络下微博的海量异构噪音信息掩蔽下的社会集群信息感知，需要结合数据库、统计学习、信息论等领域的成果，研究大尺度跨媒体集群行为感知、海量异构信息的聚融与管理两项核心内容，实现集群行为特征描述、实体之间多层次时变关联、信息一致性和完整性加工、分布式海量异构数据的综合管理与检索等技术，拓展选择性注意机制驱动下人类行为信息获取的基础理论，建立大时空尺度的虚拟社会网络下微博社会安全信息智能获取与感知框架，见图 1-2。

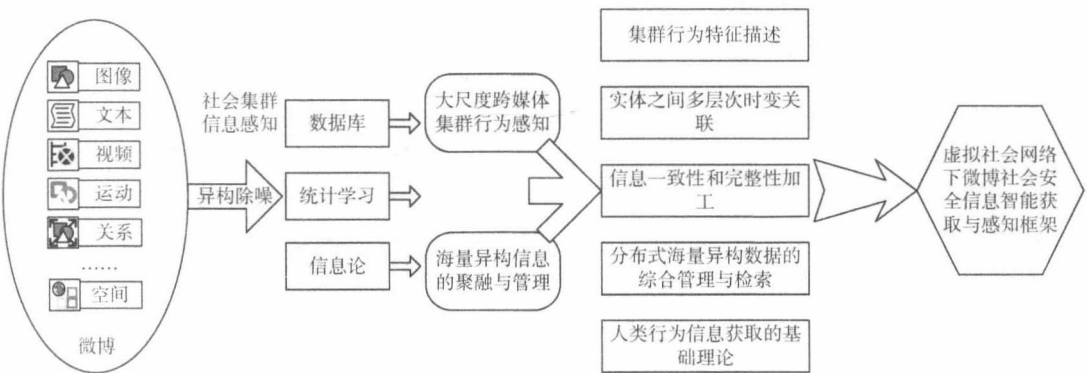


图 1-2 虚拟社会网络下微博集群行为的语义聚融

研究内容二：微博用户集群行为的复杂网络动力学演化。该研究旨在揭示虚拟社会网络下使用微博的社会集群的人类行为学特性、虚拟社会网络与社会集群模式结构映射规律、社会集群行为动力学演化机理。突发涌现的与事件紧密相关的社会集群行为的生命周期本质上是社会集群的结构形成与演化的过程。为把握社会集群行为的形成和演化规律，需要运用信息技术挖掘出隐藏在数据与表象下的集群行为社会结构，运用系统模型刻画集群行为社会结构的运行状态与演化过程。集群行为不仅具有明显的人类行为学特性，而且具有复杂网络结构特性。利用人类行为学特性，我们可以检测并识别集群行为。利用其复杂网络结构特性，我们可以把握异常集群行为形成和演化机理。在复杂网络中，异常活跃的个体行为遵循自身价值增加最大化的原则寻找关联目标，高价值个体更易成为共同的关联目标，并伴随高集聚系数的结构特征涌现出来，形成集群行为本征系统。集群行为对应的本征社会结构的部分个体在特定任务目标驱动下，经动态演化形成复杂网络，并在偶然突发因素的诱导下，涌现出特定的集群行为。集群行为的本征社会结构及动力学演化是目前主要用于理解混合网社会集群行为规律的形成、演化、平息和反弹的内在机理的关键方法。集群行为复杂网络的形成与个体的行为有着密切关系，Agent 个体行为选择是集群行为演化的动力机制。因此，研究人类行为学表征方法，发现微博集群行为的社会结构和基于 Agent 作用的微博集群动力学演化规律，是虚拟社会网络下微博社会集群行为研究的关键问题，见图 1-3。

研究内容三：微博用户集群行为有序管理的综合集成决策支持。该研究旨在揭示面向社会集群行为有序管理，构建社会科学综合知识库及定性定量模型、在不确定性与信息不对称等条件下社会集群行为有序管理的机制设计与综合集成决策支持的理论体系。社会集群行为是很多社会、安全、经济事件的主体，它在虚拟社会网络中形成与演化是一种极其复杂的信息传递与社会心理传播过程。只有相关的计算工具和优化方法映射了社会学语义，系统优化的结果才具有明确的实际意义。社会集群行为的有效管理涉及社会集群行为复杂系统的多目标分析与优化，其中的群体协作的资源共享、综合集成环境中的会商决策、情境感知的资源部署等系统管理方法学需要能够适应不同的优化目标。有的情况下需

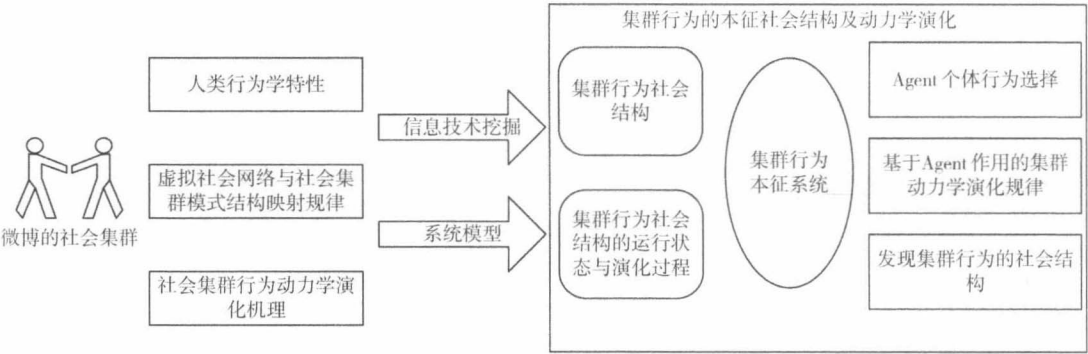


图 1-3 微博用户集群行为的复杂网络动力学演化

要对社会集群行为进行衰减，如有害信息的传播、异常组织的扩散、异常集群行为模式的出现等，有的情况下需要对社会集群行为进行强化，特别是在网络技术不断改进的背景下，通过探索网络社会行为变化的规律，采取相应的引导措施，促进其社会经济效应不断演化，不断对模型和规则进行更新学习。因此，为了使各领域各部门能够适应信息社会中社会集群的管理，有必要在社会集群行为相关的“信息-社会”综合知识基础上，建立微博社会行为有序管理的综合集成决策支持体系。在日常状态下，在组织与制度框架下，引导个体、群体、组织向着全社会整体收益最大化、社会进步的方向发展。在危机状态下，提升政府乃至全社会的突发应急能力、统筹协调能力，见图 1-4。

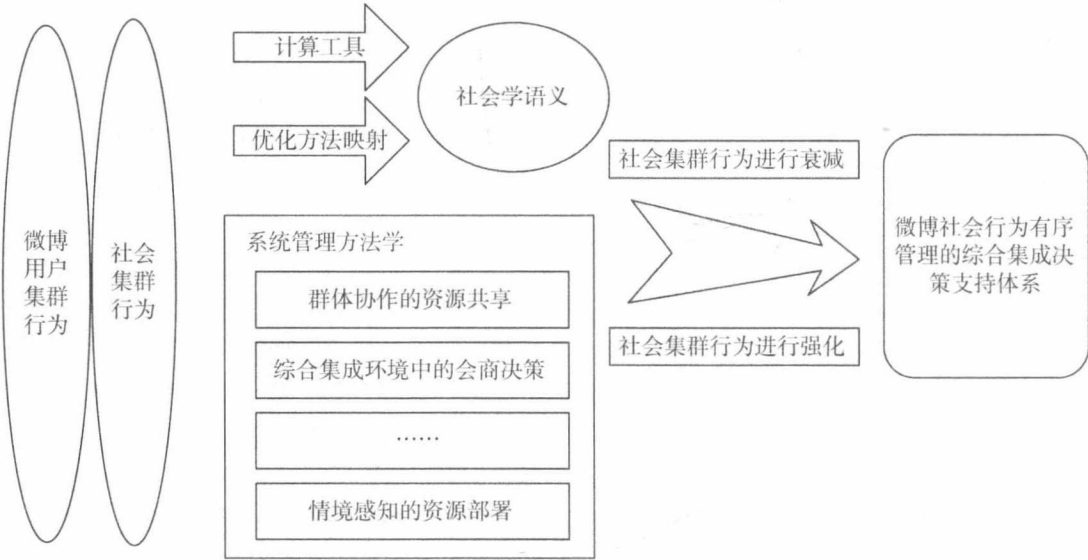


图 1-4 微博用户集群行为有序管理的综合集成决策支持

围绕上述三项研究内容，下设六个子项研究内容，其中五项开展基础理论及关键技术

研究,另一项开展应用平台建设、综合集成关键技术、集中展示系统概念:

子研究内容 1:大尺度跨微博媒体社会集群行为感知。借鉴信息论中的 CEO 问题、网络编码问题以及信号处理中的多源信号处理及估计理论等,研究如何从分散的、“有噪”的信息中,对事件进行可靠表征和优化汇聚;结合非线性流形、本体论、概率图、不变性、空间推理等理论,对感知微博对象的时空行为进行语义理解、知识发现、综合检索,提取出有明确景象意义的微博集群行为主体的语义要素,进一步推断出大时空尺度景象中的微博集群行为,见图 1-5。



图 1-5 大尺度跨微博媒体社会集群行为感知

子研究内容 2:微博海量集群行为信息的聚融与管理。以承载微博用户集群行为的数据为研究对象,借鉴数据工程、统计机器学习、信息论研究的已有成果,通过研究微博用户集群行为信息之间的关联性与互补性,建立语义关联,在保护隐私的前提下,设计基于虚拟社会网络下微博的集群行为信息快速关联与索引算法,建立基于虚拟社会网络下微博的集群行为信息聚融与管理的理论框架,见图 1-6。

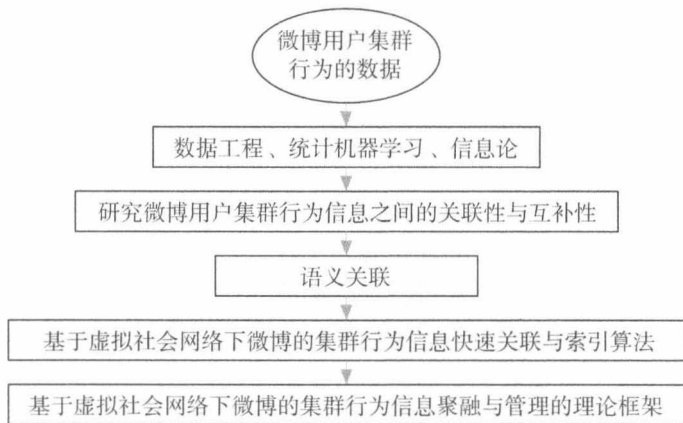


图 1-6 微博海量集群行为信息的聚融与管理

子研究内容 3：微博集群行为的社会结构挖掘与演化。首先，以微博聚融后的数据为来源，以基于虚拟社会网络下微博的集群模式及结构为主要研究对象，分析挖掘相应的微博用户社会集群模式结构以及社会集群与事件的关联，建立基于虚拟社会网络下微博用户集群行为的模式分析和结构挖掘的理论框架。其次，需要将虚拟社会网络中的不同环节在一定层面上抽象成为可以进行理论研究的系统模型。再次，(社会)群体中的个体特性信息通过虚拟社会网络各种媒介的传播在一定条件下有可能引发社会集群行为的涌现，有必要研究集群行为动力学演化和临界现象。最后，针对网络拓扑结构与节点动力学特性，提出有效的协调控制方法，用于模拟对微博集群动力学行为的调控与引导，见图 1-7。

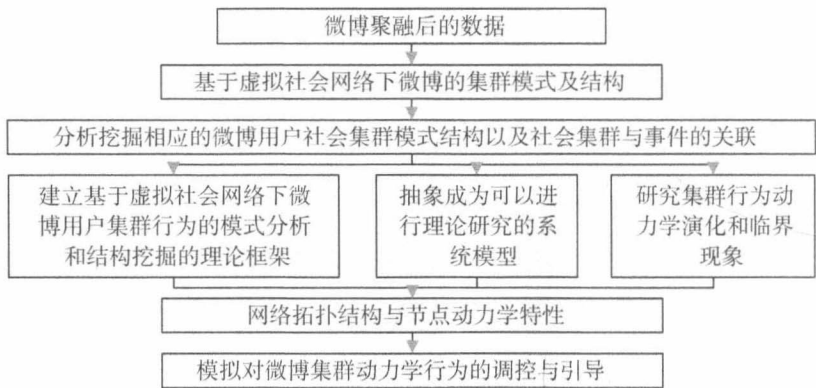


图 1-7 微博集群行为的社会结构挖掘与演化

子研究内容 4：微博集群网络行为社会学研究及定量指标。只有微博对集群行为感知方法与模式规律赋予国情、制度、文化等社会科学语义的约束与属性，才能使相关的信息、知识、规律服务于集群行为的有序管理。为此，需建立定量和定性相结合的多学科综合交叉的指标体系，为网络社会行为感知与管理提供社会科学综合知识支撑，这是信息科学与社会科学之间综合交叉的重要结合点，也是定量研究社会研究内容的关键技术，见图 1-8。

子研究内容 5：微博环境下集群行为有序管理机制。一方面，设计微博环境下集群行为有序管理机制，在不确定性与信息不对称条件下，通过在制度、规范、传统等约束下的相容激励，以最低的信息成本，使个体利益与社会整体利益相符合，微博用户在我国虚拟社会网络下更多地呈现常态稳定的有序状态。另一方面，微博集群行为事件具有突发性强、冲击力度大等特点，容易形成公共社会安全事件，有必要研究在危机情境下的集群行为有序管理的决策理论与方法，建立支撑应急决策的综合集成的模型、方法和技术，以便决策者在应对突发事件预案的基础上，在对各种可能充分考虑的基础上，识别决策对象和决策环境，避免决策误区，迅速做出正确决策，见图 1-9。

子研究内容 6：微博集群行为感知与管理的应用平台。针对国情，在微博集群行为表现充分、数据容易获得的应用场景展开实证研究，验证研究的理论、方法和技术。