



社会计算丛书

量化社会

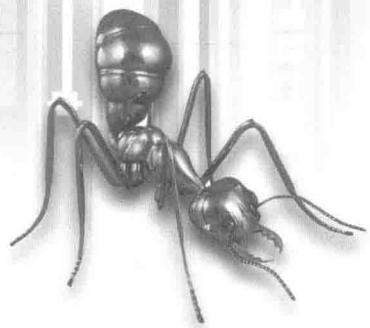
——大数据与社会计算

LIANGHUA SHEHUI
DASHUJU YU SHEHUI JISUAN

安俊秀◎著



西南交通大学出版社



社会计算丛书

量化社会

——大数据与社会计算

LIANGHUA SHEHUI
DASHUJU YU SHEHUI JISUAN

安俊秀◎著

西南交通大学出版社
· 成 都 ·

图书在版编目 (C I P) 数据

量化社会：大数据与社会计算 / 安俊秀著. —成都：西南交通大学出版社，2016.11
ISBN 978-7-5643-5105-2

I. ①量… II. ①安… III. ①数据收集—研究 IV.
①TP274

中国版本图书馆 CIP 数据核字 (2016) 第 266982 号

量化社会

——大数据与社会计算

· 安俊秀 著

责任编辑	宋彦博
封面设计	何东琳设计工作室
出版发行	西南交通大学出版社 (四川省成都市二环路北一路 111 号 西南交通大学创新大厦 21 楼)
发行部电话	028-87600564 028-87600533
邮政编码	610031
网 址	http://www.xnjdcbs.com
印 刷	四川煤田地质制图印刷厂
成品尺寸	185 mm × 230 mm
印 张	13.5
字 数	195 千
版 次	2016 年 11 月第 1 版
印 次	2016 年 11 月第 1 次
书 号	ISBN 978-7-5643-5105-2
定 价	40.00 元

图书如有印装质量问题 本社负责退换
版权所有 盗版必究 举报电话：028-87600562

前 言

社会计算是一门现代计算技术、社会学与心理学之间的交叉学科。有学者将其定义为：面向社会活动、社会过程、社会结构、社会组织和社会功能的计算理论和方法。本书作者认为，社会计算是研究计算机以及信息技术在社会科学中的应用，从而影响传统的社会行为的过程。随着 Web2.0 的兴起，许多大公司设立了专门的小组着力研究该领域，如 Microsoft、IBM、Intel、HP、Google 等诸多公司和研究机构都参与其中，开发了诸如 Wallop、Sapphire 等大量的实验项目。

传统的社会学研究往往使用调查、问卷、面谈、参与者观察与统计等形式获取数据，所使用的数据规模较小，并且难以得到个体完整的行为记录。因此，传统研究的成果更多地来源于直观认识，缺乏基于大规模真实数据的实验验证。而社交网络给人们提供了一个研究人类社会的新平台。计算社会学认为：网络上的大量信息，如博客、论坛、聊天、消费记录、电子邮件等，都是现实社会的人或组织的行为在网络空间的映射。这些网络数据可以用来分析个人或者群体的行为模式，从而深化我们对社会的了解。就像大规模基因数据催生了生物信息学一样，海量的社会数据催生了社会计算，即以计算手段研究社会学中的定性问题并解决传统社会学中的实验问题。

本书作者在社会计算、社区发现与挖掘方面进行了深入研究，不仅熟稔社会计算的理论研究，而且具有非常丰富的社会计算应用经验。本书是对作者多年研究成果的总结，针对社会计算的基本理论、主要方法及典型应用进行了系统的阐述、分析和讨论，旨在让读者了解什么是社会计算，社会计算的研究方向和发展趋势是什么，如何获取并分析社会媒体数据，以及如何使用社会计算分析社会网络，以此来构建一个更健康、更和谐的社会网络环境。全书共有 10 章，第 1、2 章介绍社会计

算的研究背景、发展前景，以及何为社交媒体，旨在让读者先了解一些关于社会计算所必需的知识。第3、4、5章讲解进阶知识，如线性阈值模型、观点挖掘方法、社区等。第6、7、8、9章是本书的灵魂。第6章讲解了社交网络模型的概念，介绍了几种较为常用的模型，并且提出了基于传统模型而改进的社交网络模型。第7章讲解什么是克隆选择算法，以及如何运用克隆选择算法对网络中的虚假信息进行防范及阻止。第8章分析了微博网络节点影响力因素，并提出了微博网络节点影响力算法 UserRank。第9章介绍如何对网络群体采用 HowNet 和 Naive Bayes 相结合的观点挖掘及情感倾向分析，从而引导网络舆论积极向上发展。第10章是对本书的总结，并对社会计算的未来进行了展望。可以预见，通过社会计算，动态分析社会网络，将会更有效地解决网络化社会新问题。

在本书的撰写过程中，四川师范大学的靳宇倡、郭英老师以及西南民族大学的王鹏老师给予了大力支持，曹书哲、黄琴琴、张宇博、丁冠中、陶冶、胡逸芃、郭嘉懿和谭宇等同学针对本书提出的理论及算法改进做了大量实践，提出了很多宝贵的修改意见，在此一并表示诚挚的感谢！同时，本书的编写和出版还得到了国家自然科学基金项目“社交网站使用对个体主观幸福感影响机制的研究（71673032）”，国家社会科学基金项目“移动社交网络对大学生交往方式的影响研究（15BSH025）”和“网络化时代青少年社会认同的研究：基于社交网站的分析（12XSH019）”，四川省社科规划项目“大数据时代人际交往方式变化的社会心理研究（SC14B016）”，四川网络文化研究中心2013年度项目“网络社会群体的趋势分析研究（WLWH13-01）”的支持。

由于时间仓促，加之水平有限，书中难免存在不妥之处，敬请广大读者批评指正。如果有任何问题或建议，可发送至电子邮箱 86631589@qq.com。

安俊秀

2016年于成都信息工程大学

目 录

第 1 章	社会计算概述	1
1.1	社会计算的研究背景及定义	2
1.2	社会计算的研究现状	5
1.3	社会计算的应用领域	8
1.4	社会计算的发展趋势	14
第 2 章	社交媒体概述	22
2.1	社会媒体的相关概念	23
2.2	社会媒体的特点	29
2.3	社交媒体面临的挑战	32
第 3 章	节点，联系与影响	34
3.1	社会网络	35
3.2	节点的重要性	36
3.3	联系的强度	41
3.4	社会影响建模	44
第 4 章	社交媒体数据获取与分析	49
4.1	社会传感器网络	50
4.2	观点挖掘	54
4.3	情感倾向分析概述	57
第 5 章	社区发现	65
5.1	社区的基本概念	66
5.2	社区发现的算法	67
5.3	社区评价	74



第 6 章 社交网络建模与分析	77
6.1 社交网络基本理论	78
6.2 社交网络信息传播研究现状	86
6.3 当前社交网络的信息传播模型	89
第 7 章 社交网络中虚假信息传播特点及控制算法研究	105
7.1 社交网络中虚假信息传播的特点	106
7.2 社交网络信息控制的研究现状	107
7.3 克隆选择算法概述	109
7.4 改进的克隆选择信息控制算法	112
7.5 改进的克隆选择信息控制算法实验及分析	120
第 8 章 微博网络节点影响力因素及度量算法分析	127
8.1 研究背景及意义	128
8.2 国内外研究现状	130
8.3 微博网络影响力关键因素分析	134
8.4 微博网络节点影响力关键因素度量研究	139
8.5 LeaderRank 算法与 UserRank 算法分析	147
第 9 章 HowNet 和 Naive Bayes 相结合的网络社会评论倾向性分析	165
9.1 网络社会评论倾向性分析概况	166
9.2 网络群体心理趋势智能分析模型架构	170
9.3 基于 HowNet 的情感词语资源的情感特征识别	181
9.4 基于朴素贝叶斯理论的分类器构造	191
9.5 HowNet 和 Naive Bayes 相结合的网络社会评论倾向性分析 结构设计及实现	196
9.6 实验与结果分析	199
第 10 章 结论与展望	205
10.1 社会计算面临的挑战	206
10.2 社会计算的发展方向	209

第1章 社会计算概述

- 1.1 社会计算的研究背景及定义/2
- 1.2 社会计算的研究现状/5
- 1.3 社会计算的应用领域/8
- 1.4 社会计算的发展趋势/14



1.1 社会计算的研究背景及定义

1.1.1 社会计算的研究背景

社会科学作为独立的学科是在 19 世纪末才出现的^[1]，这是对工业化所带来的挑战做出的回应。由于大规模的社会结构、社会组织出现，人们的社会联系越来越频繁和多样，导致社会冲突加剧，社会管理和控制的难度空前增大。此时，传统的靠少数社会精英“拍脑袋”决策管理社会的方式已经过时，而社会科学的产生和发展成为历史的必然。目前，社会科学的主要学科有社会学、教育学、经济学、政治学等。社会学的研究对象是市民社会以及以市民社会为基础形成的社会组织，其研究内容是社会良性运行和协调发展的条件以及机制^[2]。

目前，社会科学、自然科学和人文科学只是学术建制意义上的区分，它们之间有着密切的联系。自然科学的根本目的在于发现自然现象背后的规律。20 世纪 50 年代之前，社会科学与自然科学相对独立，跨学科交叉研究较少。如图 1-1 所示^[3]，社会科学与自然科学由一条学科河分隔，河的左岸是以科学计算为核心研究范式，以系统科学、控制论、人工智能等作为研究方法的自然科学；河的右岸是心理学、经济学、传播学、社会学、政治学等社会科学。

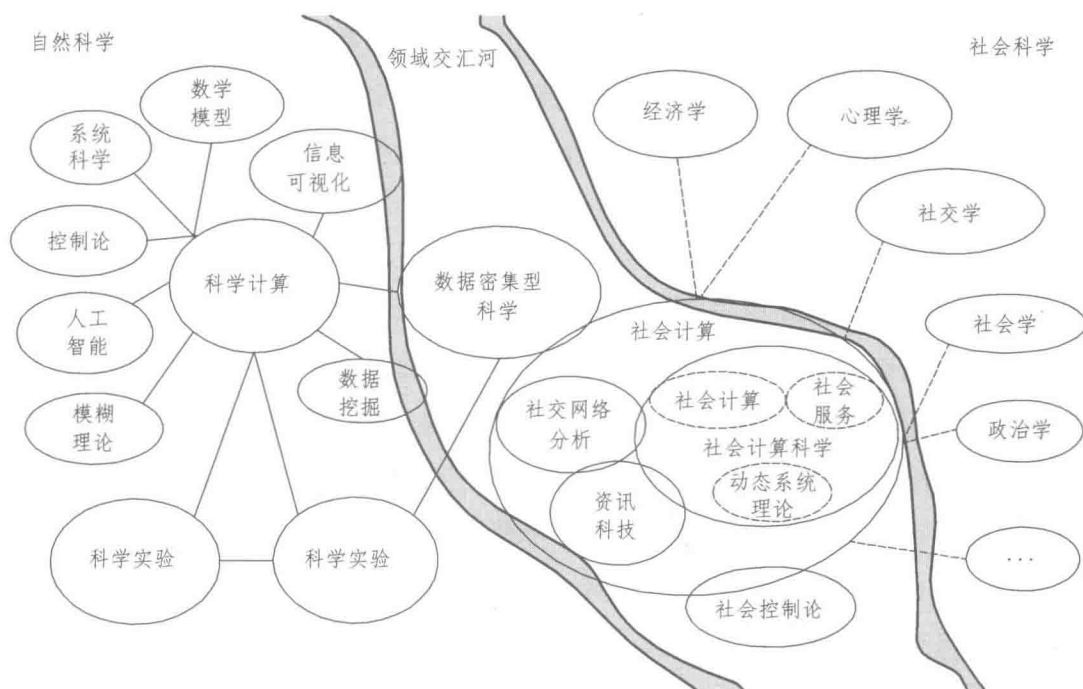


图 1-1 学科分类

20 世纪 70 年代,随着计算科学技术的发展,人们开始注意到经济与社会系统中的复杂现象。以圣菲研究所 (Santa Fe Institute, SFI) 为代表的研究机构,开创了复杂性科学这一全新的领域。为了研究复杂性现象,他们提出了“复杂自适应系统 (Complex Adaptive System, CAS)”理论,该理论用计算机作为从事复杂性研究的最基本工具,用计算机模拟相互关联的繁杂网络,观察复杂性事物适应系统的涌现行为。相关的研究引发了“人工社会”“人工科学”等诸多相关领域的兴起,形成了一系列研究复杂性的科学方法。进入 21 世纪之后,“9·11”恐怖袭击事件进一步推动了对社会计算这种宏观层面的研究的需求。人们进一步意识到,政府应当积极利用各种信息技术及相应软件实现对社会影响的合理有效的控制,并且结合信息技术和社会变化情况制定相宜的政策。然而,目前我们对信息技术在不同文化、不同社会结构中的影响的了解程度,还不足以确保我们制定出正确的政策。因此,使用

计算机模拟测试和验证社会经济政策的效果，成为公共政策领域的迫切需求。另一方面，恐怖袭击这种非对称威胁也引发了社会公共安全研究的新需求。人们迫切需
要开发出新的信息处理方法，从而更有效地分析海量的信息内容，以保障社会公共安全。

各种迹象表明，继物理计算和生物计算之后，社会计算（Social Computing）可能成为科学计算研发的新焦点，并产生新的方向和领域^[4]。形成这一趋势的根本原因是互联网在社会的各方面的不断深入和普及。而在过去十年中，Web 迅速发展，涌现出了很多用户参与的 Web 应用程序和社会信息网络，包括微博、论坛、社交网络等，它们被统称为社会媒体（Social Media）。与传统的 Web 应用相比，其最显著的不同就是：广大用户既是内容、信息的消费者，同时也是内容、信息的生产者。由大量用户提供的海量的社会行为数据，为我们开展社会计算研究提供了有利条件。同时，社会媒体最大的特点——充分的用户交互，也体现了前所未有的复杂性。然而社会计算的发展时间至今仍旧很短暂，虽然其在一些领域已经获得了一些理论上的研究成果，但由于社会系统的复杂性，其在理论和应用方面都还存在许多难以解决的问题。我们仍然需要深入研究如何有效地将社会科学理论知识与计算技术结合，最终达到科学规划社会发展的目的。

1.1.2 社会计算的定义

目前，学界对社会计算还没有一个明确和公认的定义。1994 年，社会计算的概念被首次提出，Schuler 认为“社会计算可以是任何一种类型的计算应用，它是以软件作为媒介建立社交关系的应用”。Dryer 等人将社会计算描述为一种理论概念，包括科学和技术两方面，即“人类使用计算技术进行的社交行为和交互行为所产生的相互作用”。王飞跃等人认为，社会计算是面向社会活动、社会过程、社会结构、社



会组织和社会功能的计算理论和方法。Musser D.等人认为,社会计算是指使用信息系统作为社会交互的场所,并使用信息系统作为数据收集和处理的^{空间};社会计算是在虚拟场所中的感知、交流和协作;社会计算需要把计算设备作为人与人之间交流的媒介,需要将人机交互设定为一个社会实践的环境,将理解社会过程作为交互系统的一部分工作。

从以上定义可知,社会计算是一门现代计算技术与社会科学之间的交叉学科。一方面,社会计算是研究计算机以及信息技术在社会中的应用,从而影响传统的社会行为的过程。这类研究多限于微观和技术的层面,从人机交互(Human Computer Interaction, HCI)等相关研究领域出发,研究改善人使用计算机和信息技术的手段。另一方面,社会计算是基于社会科学知识、理论和方法,借助计算技术和信息技术处理数据的方法,来帮助人类认识和研究社会科学的各种问题,提升人类社会活动的效益和水平。这类研究试图从宏观的层面来观察社会,凭借现代计算技术的力量,解决以往社会科学研究中使用经验方法和数学方程式等手段难以解决的问题。

1.2 社会计算的研究现状

我们所调研的社会计算领域的主要学术会议包括:第一届至第七届全国社会计算研讨会(CCSC, 2009—2015),第十二届IEEE社会计算国际会议(SocialCom, 2015),第一届至第六届社会计算、行为建模和预测国际研讨会(SBP, 2008—2014),第五届中国计算机学会服务计算学术会议(NCSC, 2014),以及AAAI社交媒体国际会议(ICWSM, 2010—2015)。此外,我们还调研了近年来《科学》(*Science*)、《自然》(*Nature*)、IEEE《智能系统》(*Intelligent Systems*)、IEEE《计算机》(*Computer*)、《ACM 通讯》(*Communications of the ACM*)等主流学术期刊上的相关论文和专刊,

从而对社会计算的现状和研究领域有了比较深入的了解。

1. 个体与群体的社会建模

个体与群体的社会建模包括构建社会群体或个体的认知、行为、心理模型以及对其行为特点的分析,还包括对交互模式、社区结构、个体间的社会关系等的建模。许多社会科学的理论模型都与群体和个体的社会建模相关。如社会动力学^[5]研究人类社会发展的动态过程及其演化规律;社会心理学^[6]揭示社会认知与心理的形成机制及其发展的基本规律;社会物理学^[7]研究社会稳定的机理,以及人类行为模式与社会稳定的关系。从计算角度研究社会个体与群体的工作大多是基于文本数据,近期工作的趋势是面向多媒体数据和群体行为特点进行分析与建模。2010年,IEEE《计算机》杂志发表了封面文章《社会多媒体计算》^[8]。该杂志同期的另一篇封面文章分析了中国特有的“人肉搜索”现象中的群体行为特点^[9]。社会网络是刻画个体间社会交往与互动关系的主要手段,社会群体的识别主要通过网络节点间的链接关系来发现潜在的社会群体^[10-11]。

2. 社会文化建模与分析

社会文化建模与分析包括基于智能体的人工社会建模、基于社会文化因素建模、计算实验分析、人工社会系统与计算实验平台设计等。利用计算技术来研究文化冲突和变迁,分析不同文化背景下的国家或组织的决策过程,探寻其行为所依赖的社会文化因素,已成为社会计算建模的重要研究方向^[12-13]。由于社会事件的出现往往具有突发性和不可重复性,因此采用传统方法对其演化过程进行实验分析和评估是十分困难的。所以,针对复杂社会系统的实验分析困境,我国学者提出了以社会学基本模型为基础的“人工社会(Artificial Societies)+计算实验(Computational Experiments)+平行执行(Parallel Execution)”的ACP方法^[14-17]。人工社会是一种自下而上的基于智能体的建模方法,适于动态刻画社会事件中的涌现行为。计算实

验是利用人工社会中实验的可重复性和可设计性特点,通过对人工系统设计不同的实验方案,按不同指标体系对复杂社会事件的演化规律进行可量化的实验分析。

3. 社会交互及其规律分析

社会交互及其规律分析是针对人群交互行为的特点及社会事件演化规律进行分析,也包括对社会网络结构、信息扩散和影响、复杂网络与网络动态性、群体交互和协作等的分析。计算社会学^[18]认为:网络上(如贴吧、论坛、微博、微信、电子邮件等)的大量信息,都是现实社会的人或组织的行为在网络空间的映射,网络数据可用来分析个人和群体的行为模式,从而深化我们对生活、组织和社会的理解。计算社会学的研究涉及人们的交互方式、社会群体网络的形态及其演化规律。社会事件演化规律分析主要是对事件产生、发展、激化、维持和衰减的过程和机理进行分析和评估。如在分析群体活动演化规律方面,研究者借助社会动力学,基于对数万个参与研究的移动用户终端的长期跟踪监测,对人类时空运行轨迹的规律进行分析,发现人们的运行方式遵循可重复的模式^[19]。此外,研究者已采用多种模型来分析信息在网络中的传播与扩散规律及其影响因素^[20-23]。

4. 社会数据感知与知识发现

对社会数据的获取和对规律性知识的挖掘,包括社会学习、情感及观点挖掘、社会媒体信息分析与挖掘、行为的识别和预测等。社会数据的主要形式包括文本、图像、音频和视频等,其来源除了网络媒体信息外,还包括专用网络、传统媒体和应用部门的闭源数据。为了有效利用数据源所隐含的社会化信息的结构特征,研究者提出仿照物理学传感器的原理,构建社会传感网络^[24],通过对重要节点信息的动态监控,实现对社会数据的全方位感知。基于社会数据,以及对社会个体或群体的心理和行为进行分析与挖掘发现,多种学习算法已用于预测社会组织的行为^[25-26]。规划推理方法在行为预测的基础上,还可以识别行为的目标和意图等深层信息^[27]。



社会群体心理分析主要面向文本信息(包括语音识别后转化成的文本),通过分析大量社会媒体信息,挖掘网民群体的观点及情感倾向^[28-29]。

5. 决策支持及应用

社会计算在社会安全与经济等领域的应用包括向组织管理者和社会提供决策支持、应急预警、政策评估和建议等。近年来,社会计算取得了长足的发展,并已得到广泛应用。由于网络社会媒体能够充分体现人们的价值取向和真实意愿,因此往往可以做出比传统媒体更为快速、准确的反应。开源的信息在辅助决策和应急预警中发挥出重要作用^[30-31]。在社会与公共安全领域,中国科学院自动化研究所情报安全信息学研究团队与国家相关业务部门合作,基于 ACP 方法研发了大规模的开源情报获取与分析处理系统,对社会情报进行实时的监控、分析、预警^[32],还可提供决策支持,目前已经应用于相关部门的实际业务和安全相关领域的实战中。社会文化计算已开始应用于安全和反恐决策预警中。此外,考虑到社会系统的复杂性,开展大规模社会计算研究需要特定的计算环境和平台支持,包括云计算平台及各种建模、分析、应用、集成工具和仿真环境等。

1.3 社会计算的应用领域

目前,国际上的许多研究动向表明,社会计算已成为继科学计算、生物生命计算之后新的国际前沿研究和应用的焦点与方向。我们认为,以认知科学、智能科学和复杂性科学为基础,开展社会计算研究和应用,已成为确保国家的安全、和谐社会的建设和科学发展观的落实所刻不容缓的任务^[33]。

社会计算可以在许多重大工程和社会问题的解决中得到应用,特别是在数字化



和网络化社会状态与趋势的动态分析中的应用,更是当务之急。在此,我们简单讨论一些社会计算的应用领域。

(1) 复杂工程系统的社会计算。即在人工交通、人工电网、人工制造、人工生产、人工农业等系统中加入人、社会和自然生态的影响,在更高的层次上考虑整体性的问题,为简化复杂工程系统的管理与操作,提高效益,提供科学的手段。

(2) 人口系统的社会计算。即通过构造人工人口系统,描述人口的动态变化、个体或整体人口状态;构造人口平行系统,用于国家人口综合规划和人口政策研究,并对人口进行动态管理和控制。

(3) 复杂生态问题的社会计算。即通过建立各种人工生态系统,并将人的生态观念和认识嵌入其中的计算与决策之中,评估不同生态政策对自然环境和人类社会在不同周期内的影响。

(4) 复杂经济系统的社会计算。国外已有大量关于各种人工经济系统的研究,如人工股市、经济模型、人工能源系统等。通过平行系统方法,可进一步利用这些成果,深入地研究社会经济系统的动态行为并评估不同经济政策的效果。

(5) 模拟战争系统中的社会计算。即通过人工军事系统和相应的平行系统方法,提供更有效和逼真的战争模拟,并对不同军事战略的效果、应变能力、社会经济影响、国内外政治影响进行评估。

(6) 复杂社会系统的社会计算。即通过建立各种人工社会,构造相应的平行系统,为“以人为本,全面、协调、可持续”的科学发展观提供一种可行的分析和评估方法,并应用于实际复杂社会系统的管理与控制,为将要到来的数字化社会和数字化政府管理奠定基础。

社会计算不仅是一种技术,而且是一种社会现象的浓缩。社会实践所处的环境决定了技术系统的社会特点。社会计算目前有两种发展趋势:一种是面向社会科学的社会计算,另一种是面向技术应用的社会计算。这两种发展趋势同时存在且相互影响。

1.3.1 面向社会科学的社会计算

1. 社会网络分析

社会科学中网络的概念非常广泛，主张采用多理论多层次的方法研究社会信息传播以及其他形式的组织和社会网络^[34]。Faust K.等人^[35]研究了包括人在内的4个不同动物种群的42个网络以及多种关系内涵、多种社群规模，尝试研究这些网络“表面存在差异，结构是否相似”这一问题。社会网络分析（Social Network Analysis, SNA）^[36-37]的目标是研究大型动态复杂网络，特别是人类社会系统的结构和交互模式等。社会网络分析研究的主题包括社会流动、健康行为、疾病扩散的关键节点、连锁董事、在线社区分析等。社会网络分析所采用的方法主要是基于代理的模型、理论物理、图论等。在Milgram^[38]、Watts^[39]等人开创的小世界（Small World）现象研究的基础上，Barabási等人在无标度网络（Scale-Free Network）^[40]特性上有多个发现。他们发现复杂网络中的连接符合幂律分布（Power-Law Distribution），有些节点的连接稠密，而大多数节点的连接很稀疏。这些发现说明复杂网络尽管结构复杂但并不是随机的，符合一定的宇宙主宰原则，在混沌中存在着秩序。社会网络分析领域的经典研究成果还有强弱关系（Strong or Weak Relations）^[41]、结构洞（Structural Holes）^[42]、信息级联（Information Cascades）^[43]等。

2. 计算社会科学

社会学中早就有一个计算社会学（Computational Sociology）^[44]分支，它不仅与传统的系统科学、控制论和复杂性科学交叉，并且跨越社会科学多个领域，如经济学、生态学、社会网络组织、人口学中的小群体动力学、环境以及城市规划等。计算社会学的主要研究领域是社会模拟（Social Simulation），即用计算机建立人工实验环境，研究复杂社会系统。其模拟方法包括基于方程的模型和计算模型。其主要计算方式是在个体代理或多代理系统观点下进行社会模拟^[45]。