

第 1 章

绪论

1.1 研究背景及意义

1.1.1 研究背景

(1) 理论背景

① 引入复杂系统理论建模群体观点演化复杂现象,是对现有观点动力学研究的深化完善

观点演化是观点扩散的基础,观点演化过程中往往伴随着新观点的产生和传播。社交媒体用户之间的频繁互动促进了观点在网络中的传播和扩散。社交媒体用户观点演化是一个多用户交互过程,不是简单的单向线性过程,具有复杂性、动态性和交互性^①,探索复杂问题需要引入复杂系统理论。观点动力学就是研究群体中观点随时间的演变及个体间的相互作用对观点演化

^① 安璐,张思宇.支持或反对:社交媒体用户观点的形成机理[J/OL].图书馆论坛:1-13[2023-09-12].<http://kns.cnki.net/kcms/detail/44.1306.g2.20230203.1046.002.html>.



过程的影响^①。但现有观点动力学研究聚焦于群体观点演变的理论建模与仿真,所用数据往往为随机模拟数据,针对实证视角下的群体观点演化机制,群体意见的强化和争议的消减,观点的扩散与群体决策之间的联系等相关研究缺乏。从应用角度来看,对观点演化进行研究的目的是将成果应用于舆情治理、市场营销、公共管理等诸多实践领域,现有的观点动力学成果主要集中于模型的精细设计。针对用户观点在演化过程中的群体交互规律分析以及基于观点演化结果辅助网络社会治理决策的现有研究尚不充分,这不利于准确掌握在线群体观点的演化规律,进而为网络社会治理提供参考。

② 采用未来网络科学中的高阶网络剖析多个体交互,提供了群体观点演化分析客观科学的新方法

加州大学洛杉矶分校数学教授 Mason A. Porter 在 *Nonlinearity + Networks: A 2020 Vision* 一书中指出了未来复杂网络研究的 4 个热点,分别是:时序网络、网络上的动力学过程、自适应网络和高阶网络^②。其中的高阶网络考虑的就是多个节点间的连接。高阶交互指的是作用在网络的高阶结构上且涉及多个体的交互行为^③。相较于二元交互,高阶交互可用于描述系统中涉及多个体的交互行为。在网络空间中,个体并非孤立存在,在群体成员的相互影响下,个体了解与观察形成观点的过程中势必会受到群体内其他成员的影响。基于高阶交互的研究,能够从新的角度去解释传播现象,研究网络高阶结构与功能之间的关系,发现二元交互视角下缺失的动力学性质等。因此,为准确刻画舆情事件下在线群体观点演化过程的相互影响,本书拟采用复杂网络中的前沿技术即高阶交互网络,剖析舆情事件中在线群体观点的形成、突变和预测过程。

(2) 实践背景

① 伴随着社交媒体的发展,群体观点演化逐步成为网络舆情治理分析的核心问题

全球社交媒体用户总数在新冠疫情期间增长了近 30%,将近 48.8 亿用

^① Monti C, Morales G D, Bonchi F. Learning opinion dynamics from social traces[C]. 26th ACM SIGKDD Conference on Knowledge Discovery and Data Mining(KDD'20), New York, 2020: 764-773.

^② Porter M A. Nonlinearity+Networks: A 2020 Vision[J]. Emerging Frontiers in Nonlinear Science, 2020, 32: 131-159.

^③ Iacopini I, Petri G, Barrat A, et al. Simplicial models of social contagion [J]. Nature Communications, 2019, 10: 2485.

户活跃在社交网络上,占全球人口的比例已经高达 60.6%^①。同时,中国互联网信息中心(China Internet Network Information Center,CNNIC)的报告显示,截至 2023 年 8 月,我国网民规模达 10.79 亿,互联网普及率达 76.4%^②。在此背景下,越来越多的用户热衷于通过社交媒体发表言论、表达诉求,由信息传播的受众变成了信息传播的主体。在这些言论和诉求中,往往蕴含着用户针对某个事件和组织的看法、观点。对于热点事件,每个人都有着自己的观点。有限理性的个体在已有观点驱动下与社交媒体中的其他用户碰撞融合后形成了群体观点,群体观点推动和决定着群体行为的方向。违背主流价值的群体观点不仅会使公众对舆情关注点发生错位,甚至挑动社会不安情绪,不断积聚形成强大舆论势能,给社会安定造成威胁。因此,群体观点作为群体态度、情感与意见的凝集,在舆情演化过程中具有重要的地位,有效探究群体观点的演化,对促进舆情合理治理、明确意见演变机理具有重要作用。

② 圈群文化蔓延,群体认知分析是深耕数据价值助力精准决策的现实需求

网络圈群是新媒体环境下的在线用户由于共同的兴趣、地缘或工作业务而利用微信、微博和 QQ 等自媒体平台形成的一种虚拟聚合空间^③。舆情事件的爆发,导致作为社会感知器的用户在社交媒体平台上生成大量关于事件的方方面面信息,并吸引在线圈群中聚集的好友、粉丝加入事件讨论,产生了“围观”“吃瓜”“友军”等描述舆情集聚现象的热词。在此背景下,社交媒体中汇集了在微观个体的认知与交互基础上形成的中观群体对宏观事件的态势感知数据,但其具有高度碎片化、价值密度低等特点,难以被直接利用。在全球数字经济加速发展的大趋势下,党的二十大报告提出“建设网络强国、数字中国”“发展数字经济”和“数字经济和实体经济深度融合”等战略性目标。因此,在挖掘并利用社交媒体中群体用户对舆情事件的态势感知数据基础上,精细化提炼社交媒体群体用户发表的关于舆情事件的多维动态观点,有助于相关部门理解舆情事件的中观群体观点演化过程和发展逻辑,为应急响应提供情报支持,提升舆情事件预警能力。

① Data reportal,Meltwater,We are social. Digital 2023 global overview report [R/OL]. (2023-01-26) [2023-09-12]. <https://datareportal.com/reports/digital-2023-global-overview-report>.

② 中国互联网信息中心. 第 52 次《中国互联网络发展状况统计报告》[R/OL]. (2023-08-28) [2023-09-12]. <https://www.cnnic.net.cn/6/86/88/index.html>.

③ 王仕勇,余佳琦. 网络圈群现象研究综述[J]. 情报杂志,2022,41(5): 86-91.



1.1.2 研究意义

(1) 理论意义

① 丰富并完善了社交媒体群体认知研究和观点动力学研究的理论框架

针对舆情事件的多元观点出现以及观点间的交互融合,使得在线群体观点演化逐步呈现出一致、分叉和极化等不同现象。现有社交媒体群体认知研究构建了用户观点演化的复杂网络动力学模型,但忽略了群体观点演化过程中的高阶交互社交关系。属于统计物理学的观点动力学方法更适合于内外部因素稳定的传统网络环境,无法模拟舆情事件下复杂多变的群体观点演化。社交舆情背景下在线群体观点频繁的分歧和反转,迫切需要一种能够揭示各种非线性特征潜在机制的方法。本研究综合考虑社交媒体用户观点演化的多重交互属性,研究高阶交互网络中社交媒体用户观点演化机制,揭示在线群体用户观点演化的本质和机制,更加准确真实地反映在线群体观点演化的现实情况,是对现有社交媒体群体认知研究和观点动力学理论体系的补充,为后续在线群体认知的研究提供了参考借鉴。

② 拓展并深化了高阶交互网络在观点演化和舆情引导方面的应用场景

探索网络问题还需借助网络的力量。高阶交互网络作为一种多层次社会网络分析方法,可用于研究不仅关注节点间的直接关系,而且关注节点间间接关系的复杂网络结构。本书通过将高阶交互网络引入在线群体观点演化这一复杂情景,构建在线群体观点演化高阶交互网络以刻画网络中各节点(网络用户)之间的互动关系(如转发、评论和点赞等行为),表征网络节点间特有的多对多特性和“群体性”结构特征,挖掘在线群体观点随群组耦合产生的认知演化规律,突破了以往二元交互视角下网络舆情引导研究中忽视社交媒体群体行为多元交互性的局限,深化了对在线网络结构和社交关系特征的分析,验证了高阶交互网络在刻画社交网络复杂传播现象方面的有效性。

(2) 实践意义

① 精准刻画在线群体的观点网状传播规律和复杂决策模式

在线平台为用户提供可视化、可互动、自主发声空间的同时,也为群体观点极化及意见领袖意识压迫提供了方便。本书通过挖掘社交网络-群体互动网络-观点演化网络的结构特征、观点演化规律,探寻了社交舆情背景下在线

群体集聚后的群体观点在形成、突变和预测阶段的耦合机制,理解了在线群体观点在演化过程中的动力学规律,并从过程化、交互式和跨层次的舆情演化角度描绘了数智时代下的“大众观点”,有助于克服先前把握社交媒体用户观点演化规律时单一时段、单一层次、个体割裂的囿限。此外,本书以群体观点演化分析为基础,能够及时掌握群体用户在舆情事件中的异常行为和极端反馈,对提升网络观点信息的正面影响效果,降低网络内容的负面引导,深入理解社会事件中在线群体认知、态度、情感、行为等的形成演化规律具有典型示范效果。

② 科学辅助监管部门的社交媒体生态治理和舆情管理

社交媒体作为用户观点扩散的重要媒介,对其干预可以抑制网络谣言等有害信息的蔓延,促进疾病防控、创新思想等有利信息的扩散。先前研究主要从两个个体间分析观点的扩散并从关键节点如意见领袖角度实现舆情引导,忽视了舆情演化系统中的各动力系统结构之间交互耦合的事实。本书以网络舆情为研究主题,以情报智慧赋能多学科交叉的理论方法和实践问题展开研究,融合网络舆情与复杂科学专业知识,重点分析在线群体观点交互机制与演化规律,厘清在线群体观点的形成、突变与预测机制,并基于观点演化数据分析结果制定了更精准的差异化群体观点引导策略,这有助于监管部门提高舆情管理的科学性和网络社会治理的实效性,改善网络群体分级管理制度,进而设计舆情研判监管策略,为我国社交媒体的健康发展和网络文化生态环境治理提供参考。

1.2 国内外研究现状

本书主要是借助复杂网络领域中的前沿技术准确刻画舆情事件爆发后,由围观舆情的微观个体集聚交互引发的在线群体观点演化机制。这一机制本质上是从舆情扩散的角度对在线群体观点动力学进行的科学研究。因此,文献综述主要围绕“复杂社交网络”“观点演化”“舆情扩散”和“在线群体”4个方面展开,主要内容如下。



1.2.1 复杂社交网络研究现状

社交网络是一种典型的复杂网络，遵循小世界、无标度等复杂网络特征^①。复杂社交网络是复杂网络理论在社交网络中的应用。图 1-1 和图 1-2 分别是国内外此领域关键词时线图谱，据此可分析近年来国内外此领域的研究热点和发展趋势。由图看出，国内外针对复杂社交网络研究主要集中在以下几个方面。

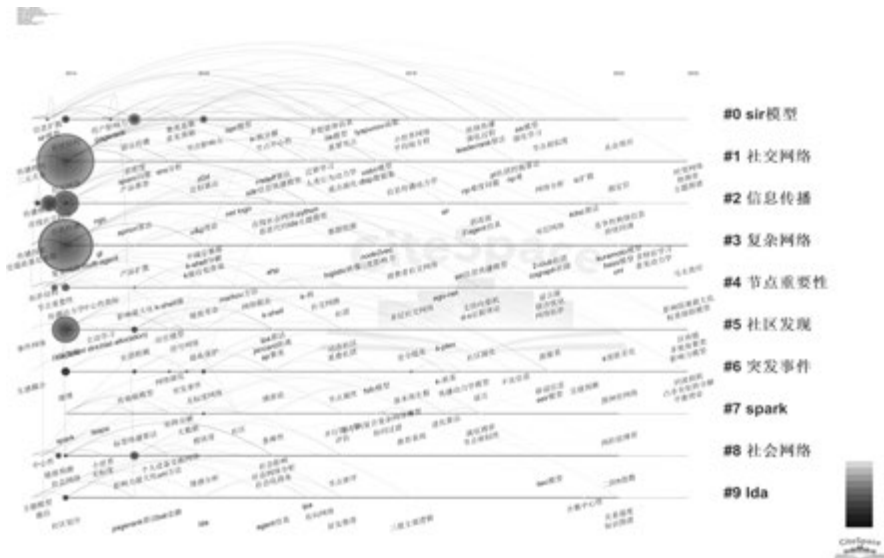


图 1-1 国内复杂社交网络研究热点

(1) 复杂社交网络中的网络结构与演化分析研究

演化分析主要通过节点的加入和退出、连接的建立和断裂等关注复杂社交网络的结构和网络的形成、发展和变化规律。国内学者韩菁等^②(2022)构建并分析了新产品的两阶段博弈模型的推广网络规模、网络演化速度等网络

^① Watts A. Complex social networks[J]. Journal of Economic Literature, 2008, 46(1): 192-193.

^② 韩菁, 蔡寻, 滕新玉. 社交网络中新产品推广策略两阶段博弈演化分析[J]. 系统工程学报, 2022, 37(3): 317-329+343. DOI: 10.13383/j.cnki.jse.2022.03.003.

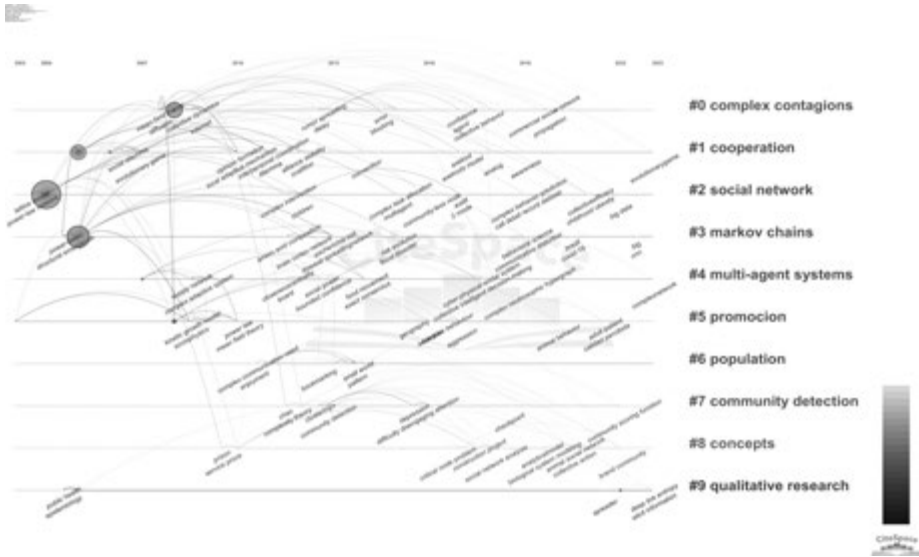


图 1-2 国外复杂社交网络研究热点

结构参数对普通网民选择策略的影响。李赵兴等^①(2021)借助群体智能算法的思想和社交网络的拓扑结构,设计了一种离散粒子群优化算法以求解社交网络结构平衡问题,挖掘了复杂社交网络的社区结构。梁勤等^②(2016)分析了社交网络中边的生命周期对复杂社交网络中的度、网络直径和平均聚类系数等结构参数的影响。国外学者 Singh 等^③(2023)则基于信息科学理论与控制网络理论构建了一个集成框架以量化和控制复杂网络中节点之间的信息流,并阐述了社交网络中网络拓扑结构与功能模式之间的影响关系。Nian 等^④(2022)为揭示大规模社交网络中的社区结构和社区演化规律,提出了一种大众吸引力驱动力的社区演化分析方法,实验表明该方法能有效检测和识别

① 李赵兴,陈莉,刘琼海. 基于粒子群优化算法的社交网络结构平衡的实现[J]. 电子设计工程, 2021,29(14): 140-144. DOI: 10.14022/j.issn1674-6236.2021.14.030.

② 梁勤,李磊,刘冠峰. 基于链接寿命的社交网络结构演化分析[J]. 计算机工程与科学,2016,38(10): 2025-2037.

③ Singh M S, Pasumarthy R, Vaidya U. On quantification and maximization of information transfer in network dynamical systems[J]. Scientific Reports,2023,13(1): 5588(2023).

④ Nian F Z, Luo L, Yu X L. Crowd attraction-driven community evolution on social network[J]. International Journal of Modern Physics C,2022,33(01): 2250009.



出大规模社交网络中的社区结构。Zang 等^①(2018)发现微信社交网络中节点与边的时间幂律增长规律,提出了基于随机生成器模拟节点和边相互作用的动力学模型以预测网络演化过程。

(2) 复杂社交网络中的行为动力学与信息传播研究

基于复杂社交网络动力学的信息传播研究侧重于节点之间的互动行为、信息传播和影响力传播等。动力学分析可以通过建立节点间的代理模型、传染病模型、传播模型等来模拟和预测节点的行为和信息传播的过程。国内学者沈旺等^②(2023)使用超图刻画豆瓣的网络拓扑结构,利用改进的 SIR 模型制定信息传播规则构建了在线社交网络信息传播模型。刘小洋等^③(2020)提出了在线社交网络的竞争性非线性动力学信息传播模型,求解了模型中传播动力学微分方程,为解决在线社交网络上不同类型信息的竞争性传播类问题提供了一种新方法。王家坤等^④(2019)基于复杂网络理论和传播动力学方法,构建了双层社交网络上的舆情传播模型,并提出企业舆情传播的控制策略。国外学者 Wang 等^⑤(2021)建立了突发公共卫生事件中基于 SIR 模型的网络谣言反转模型,以控制网络谣言的传播并减少其负面影响。Zhu 等^⑥(2019)提出了一种基于阈值模型的加权双层社交网络信息传播模型,分析了真实社交网络中不同层次的节点间亲密度对信息传播的影响。Su 等^⑦(2016)提出了一种进化博弈论框架来模拟多个传染事件同时传播时的扩散过程,这个框架为如何做出转发决策提供了新的见解。

(3) 复杂社交网络中的链路预测与应用研究

作为理解和预测复杂社交网络中的交互行为、信息传播和社区发现的基

① Zang C, Cui P, Faloutsos C, et al. On power law growth of social networks[J]. IEEE Transactions on Knowledge and Data Engineering, 2018, 30(9): 1727-1740.

② 沈旺, 时倩如, 王俊尧, 等. 基于超图的在线社交网络信息传播模型研究[J]. 情报学报, 2023, 42(03): 354-364.

③ 刘小洋, 何道兵. 在线社交网络的竞争性非线性动力学信息传播模型研究[J]. 计算机学报, 2020, 43(10): 1842-1861.

④ 王家坤, 王新华. 双层社交网络上的企业舆情传播模型及控制策略研究[J]. 管理科学, 2019, 32(01): 28-41.

⑤ Wang X W, Li Y Q, Li J X, et al. A rumor reversal model of online health information during the Covid-19 epidemic[J]. Information Processing & Management, 2021, 58(6): 102731.

⑥ Zhu X Z, Yang Q W, Tian H, et al. Contagion of information on two-layered weighted complex network[J]. IEEE Access, 2021, 7: 155064-155074.

⑦ Su Y, Zhang X, Liu L, et al. Understanding information interactions in diffusion: an evolutionary game-theoretic perspective[J]. Frontiers of Computer Science, 2016, 10(3): 518-531.

基础性工作,复杂社交网络中的链路预测是基于已知的节点和连接信息,预测未来可能出现的连接或关系。国内学者刘春江等^①(2021)指出基于邻居节点、共有邻居和共有社区的三种图算法是未来复杂网络链路预测任务的主流。顾秋阳等^②(2021)提出一种基于高阶路径相似度的复杂网络链路预测方法,提升了链路预测方法的精度和效率。张子柯^③(2017)述评在线社交网络信息传播机制与动力学研究时指出,复杂网络的链路预测算法大致可以分为基于结构相似性的预测算法和基于似然分析的预测算法。国外学者 Yin 等^④(2023)通过关注两个群组实体之间的链路预测,提出了互多视图注意力网络 MMAN 模型。Uthayasuriyan 等^⑤(2023)通过实验证明,将差分进化算法与 XG-Boost 结合使用可以更好地预测复杂网络中的链路。Daud 等^⑥(2020)在基于复杂社交网络链路预测的应用方面认为推荐系统、异常检测、影响力分析和社区检测等是复杂社交网络中基于链路预测常见的应用。

(4) 复杂社交网络中的社群发现研究

复杂社交网络中的社群发现任务是将复杂社交网络中的节点划分为若干小组,保证小组的内部节点比外部节点更相似或关系更密切。国内学者余韦等^⑦(2022)基于符号网络社团检测模型挖掘了舆情数据中的语义社团,以推演群众对某事件的普遍看法。张海涛等^⑧(2019)以中美间贸易摩擦事件话

① 刘春江,李姝影,胡汗林,等.图数据库在复杂网络分析中的研究与应用进展[J].数据分析与知识发现,2022,6(07):1-11.

② 顾秋阳,吴宝,池仁勇.基于高阶路径相似度的复杂网络链路预测方法[J].通信学报,2021,42(07):61-69.

③ 张子柯.在线社交网络信息传播机制与动力学研究综述[J].情报学报,2017,36(04):422-431.

④ Yin G Z, Wang X, Zhang H L, et al. Beyond individuals: Modeling mutual and multiple interactions for inductive link prediction between groups[C]. Proceedings of the Sixteenth ACM International Conference on Web Search and Data Mining(WSDM'23), Singapore, 2023: 751-759.

⑤ Uthayasuriyan A, Ramya G R, Jeyakumar G. Effective link prediction in complex networks using differential evolution based extreme gradient boosting algorithm [C]. 2nd International Conference on Advanced Network Technologies and Intelligent Computing(ANTIC), Varanasi, 2023, 1797: 149-163.

⑥ Daud N N, Hamid S H A, Saadoon M, et al. Applications of link prediction in social networks: A review[J]. Journal of network and Computer Applications, 2020, 166: 02716.

⑦ 余韦,章金楠,朱梦丽,等.基于符号网络社团检测的舆情情报分析方法研究[J].情报杂志,2022,41(05):55-60+161.

⑧ 张海涛,刘雅姝,张泉慧,等.基于模块度的话题发现及网民情感波动研究——以新浪微博“中美间贸易摩擦”话题为例[J].图书情报工作,2019,63(04):6-14. DOI: 10.13266/j.issn.0252-3116.2019.04.001.



题为例,通过社群发现算法来识别基于事件发展的子事件评论网络中的话题社群,动态跟踪网民意见及情感波动。张锴琦等^①(2016)提出了一种基于进化算法的重叠社群探测方法,并对重叠社群结构进行了准确的划分。国外学者 Roghani 等^②(2023)提出基于一种局部平衡标签扩散的快速社群发现算法,在真实世界网络上证明了优越性。Yuan 等^③(2022)从最大化群落结构的马尔可夫稳定性角度出发,提出了一种基于图卷积网络的重叠群落检测方法,发现了复杂网络中的社团结构。Ullah 等^④(2022)提出了一种新的基于相关性的信息交互模型 RIIM,并基于局部和全局拓扑知识识别复杂网络中的社区,以降低网络计算和处理的成本。复杂社交网络中的社群发现对理解社交网络的组织结构、发现潜在的社群关系、预测节点的行为等具有重要的应用价值。

(5) 复杂社交网络中的关键节点与异常行为识别研究

识别复杂社交网络中的关键节点可以帮助我们发现网络中的重要节点,而重要节点可能在网络中具有特殊的行为模式,分析这些特殊的行为模式是复杂社交网络中的重要内容。在这方面,国内学者杨瑞仙等^⑤(2022)采用复杂网络方法对学术虚拟社区核心用户间的知识流动特征进行了分析,多维度揭示了核心用户间的知识流动特征,对后续研究知识交流表现形式及交流机制提供了参考。李纲等^⑥(2015)以 MetaFilter 为例,通过统计节点网络中的中介中心度和聚集系数,探究了“小众专家”特征识别方法。国外学者 Rai 等^⑦(2023)认为当前复杂社交网络有影响力节点识别的方法分为基于结构的方法、基于向量的方法、基于 MCDM 的方法和基于机器学习的方法。

① 张锴琦,杜海峰,何晓晨. 基于进化算法的重叠社群结构探测[J]. 管理工程学报,2016,30(01): 221-227. DOI: 10.13587/j.cnki.jieem.2016.01.028.

② Roghani H, Bouyer A. A fast local balanced label diffusion algorithm for community detection in social networks[J]. IEEE Transactions on Knowledge and Data Engineer, 2023, 35(6): 5472-5484.

③ Yuan S J, Zeng H F, Zuo Z Y, et al. Overlapping community detection on complex networks with Graph Convolutional Networks[J]. Computer Communications, 2022, 199: 62-71.

④ Ullah A, Wang B, Sheng J F, et al. A novel relevance-based information interaction model for community detection in complex networks[J]. Expert Systems with Applications, 2022, 196: 116607.

⑤ 杨瑞仙,郭孟含. 学术虚拟社区核心用户知识流动特征研究[J]. 现代情报,2022,42(08): 50-60.

⑥ 李纲,叶光辉,张岩. “小众专家”特征识别——基于 MetaFilter 的实证分析[J]. 现代图书情报技术,2015(06): 71-77.

⑦ Rai K A, Machkour M, Antari J. Influential nodes identification in complex networks: a comprehensive literature review[J]. Beni-Suef University Journal of Basic and Applied Sciences, 2023, 12(1): 18.

1.2.2 观点演化研究现状

个体或群体的观点会随着时间的推移发生变化,在观点演化过程中,人们可能会在吸收新信息后接受新的观点或者调整他们的立场。图 1-3 和图 1-4 分别是国内外此领域关键词时区图谱,可以分析出近年来国内外针对观点演化的研究主要聚焦于以下几个方面。

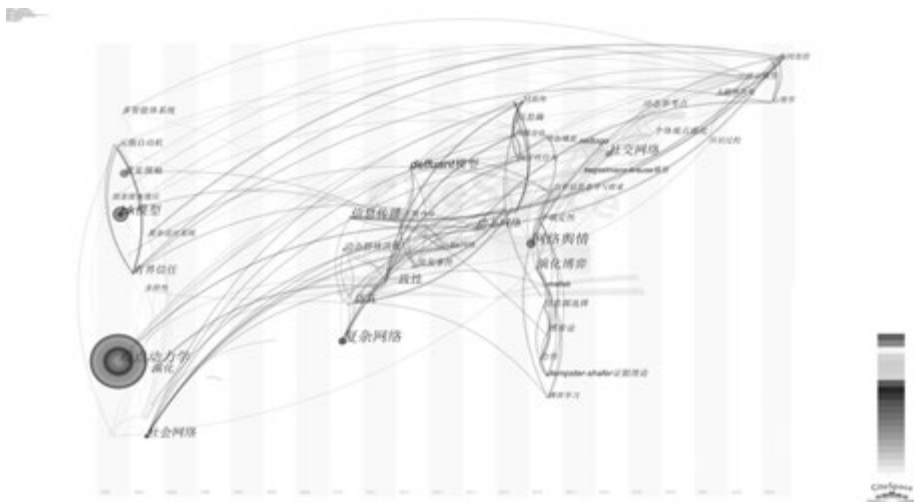


图 1-3 国内观点演化研究热点

(1) 基于观点动力学的观点演化与影响因素建模仿真

基于观点动力学的观点演化与影响因素建模仿真是一种基于观点动力学理论,利用计算模型和仿真技术来研究和模拟观点演化过程以及影响观点演化的因素的方法。该类研究假定观点是可以通过相互作用和影响而发生变化。例如,魏静等^①(2022)提出一种基于有权网络的边权演化 Deffuant 模型,解析了个体之间的观点传播和联系强度的演化特点。朱晓霞等^②(2019)基于观点动力学引入有界信任模型,探索了多层网络中意见领袖及跟随者在

^① 魏静,张耀曾,朱恒民,等. 基于改进 Deffuant 模型的舆情观点演化解析[J]. 管理评论,2022, 34(09): 147-157. DOI: 10.14120/j.cnki.cn11-5057/f.2022.09.022.

^② 朱晓霞,孟建芳. 多层网络中意见领袖及跟随者观点演化研究——基于有界信任模型[J]. 情报科学,2019,37(06): 30-36. DOI: 10.13833/j.issn.1007-7634.2019.06.005.

销领域,岳芳等^①(2023)为帮助企业更有效地利用协同数据,基于马尔可夫链建立了群体观点演化模型,以分析任意网络结构和规模的群体观点演化过程。面向情境应用与决策支持的观点演化分析提供了一种系统性的方法以支持决策者在不确定和复杂的情境中做出决策。

(3) 观点演化过程中的异常现象分析研究

观点共识、观点冲突、观点突变、观点反转与观点极化是观点演化中经常被学者研究的特殊现象。就观点一致或观点共识而言,Xie等^②(2017)基于群体观点动力学模型分析了固执个体的观点在群体观点演化中的动力学,发现固执个体的存在会促使观点共识的达成。安璐等^③(2023)从新冠疫情下的“新冠疫苗接种”话题语料中识别突发事件情境下的网络冲突观点,并提取其中的重叠共识,为网络舆情冲突治理提供了方法与借鉴。胡斌等^④(2023)建立了网络社区群体观点突变的弹性指数预测模型,对互联网社区群体观点的突变现象进行了分析。

1.2.3 舆情扩散研究现状

舆情扩散是涉及公众关注和讨论的事件、话题、观点等在社会中广泛传播和传递的过程。图 1-5 和图 1-6 分别是近年来国内外针对舆情扩散的关键词聚类图谱,从图中可以看出国内外对舆情扩散的相关研究主要集中在以下几个方面。

(1) 舆情扩散的前因与扩散的驱动力分析

舆情扩散的前因与扩散的驱动力分析目的在于事前做好预案,通过相互作用的因子分析,防范负面舆情扩散,推动正向舆情传播。基于上述目的,

① 岳芳,樊茂瑞,高子雅,等.开放式知识协同平台中的“回音室效应”研究——以Bilibili“新能源汽车”视频评论为例[J/OL].情报理论与实践:1-11[2023-09-17].<http://kns.cnki.net/kcms/detail/11.1762.G3.20230908.1056.002.html>.

② Xie H,Li G J,Yan Y J,et al. Evolution of bounded confidence opinion in social networks[J]. Discrete Dynamics in Nature and Society,2017,2017: 3173016.

③ 安璐,周凡倩.突发事件情境下冲突观点与重叠共识研究[J].情报理论与实践,2023,46(09): 69-78. DOI: 10.16353/j.cnki.1000-7490.2023.09.009.

④ 曾紫琳,胡斌,王智超,等.网络社区群体观点演化的弹性与突变机制[J/OL].中国管理科学:1-16 [2023-09-17].<https://doi.org/10.16381/j.cnki.issn1003-207x.2023.0247>.



图 1-5 国内舆情扩散研究热点

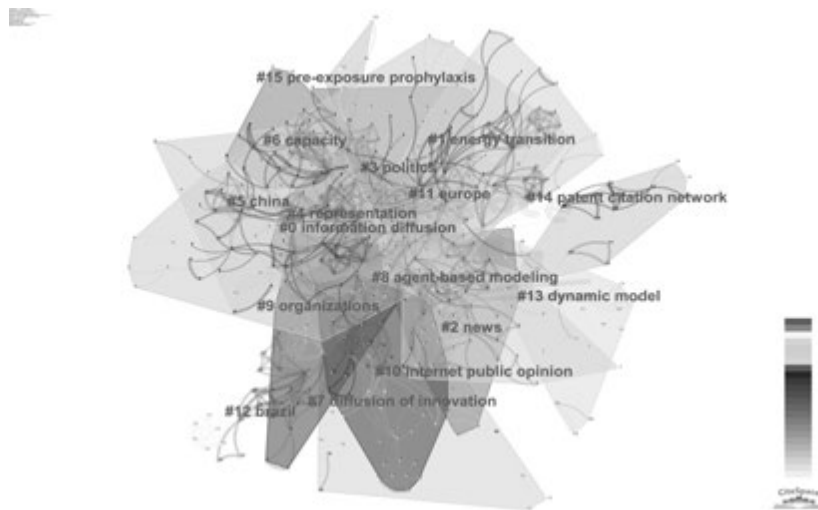


图 1-6 国外舆情扩散研究热点

Wang 等^①(2022)从内部和外部两个层面探讨了舆论在两种类型的社交网络中的传播规律,分析了影响其演变过程的关键因素。陈玉萍等^②(2023)的研究结果表明,体育赛事危机事件网络舆情演化的动力因素包含主体因素(网民、媒体、意见领袖、当事人、政府)、载体因素(平台载体)、客体因素(赛事危机事件)。董娜^③(2022)认为舆情主体、本体、媒体、空间等构成的要素互关联和彼此影响为用户生成的网络视频舆情的传播提供了必要支持。

(2) 舆情扩散的过程及扩散机制模型构建

舆情扩散的过程及扩散机制模型构建目的在于事中有有效干预,通过分析舆情扩散过程中各要素的交互机制与关键节点,及时对舆情做出有效应对。例如,朱光等^④(2023)以“钟薛高雪糕事件”为例,构建了主题-情感融合分析模型以探寻不同群体视角下舆情演化的传播机制。温志韬等^⑤(2022)以“杭州女子失踪案”为例构建了网络舆情态势演化模型,以可视化结果分别突显出了网络舆情态势演化的静态特征和动态特征。Liu 等^⑥(2018)提出了突发公共事件信息传播模型和意见领袖分级模型,发现意见领袖在微博信息传播中具有较强的传输能力,同时表明意见领袖可以影响那些对某些公共紧急情况不确定的微博用户。

(3) 舆情扩散的后果及扩散应对策略提出

舆情扩散的后果及扩散应对策略提出侧重于事后合理应对,旨在评估舆情传播扩散带来的不良后果及影响程度,并针对舆情的扩散提供相应的建议,辅助决策者及时制定合理的策略。例如,Bermes^⑦(2021)指出虚假新闻使

① Wang J K, Li Y. Research on the propagation and governance of public opinion information under the joint action of internal and external factors[J]. *Aslib Journal of Information Management*, 2022, 75(2): 193-214.

② 陈玉萍,刘嘉毅,冯加付. 体育赛事危机事件网络舆情演化的动力因素、演化机理与应对策略[J]. *成都体育学院学报*, 2023, 49(04): 48-55. DOI: 10.15942/j.jcsu.2023.04.007.

③ 董娜. 基于用户生成内容的短视频网络舆情传播生态系统构建[J]. *图书馆*, 2022(04): 73-81.

④ 杨沁蓉,朱光,雍华贵. 基于主题-情感融合模型的网络舆情演化研究——以“钟薛高雪糕事件”为例[J]. *情报探索*, 2023(08): 111-119.

⑤ 温志韬,夏一雪. 基于演化建模的突发事件网络舆情态势感知分析[J]. *情报杂志*, 2022, 41(09): 71-78.

⑥ Liu X Y, Liu C. Information diffusion and opinion leader mathematical modeling based on microblog[J]. *IEEE Access*, 2018, 6: 34736-34745.

⑦ Bermes A. Information overload and fake news sharing: A transactional stress perspective exploring the mitigating role of consumers' resilience during COVID-19[J]. *Journal of Retailing and Consumer Services*, 2021, 61: 102555.



消费者产生了诸如社交媒体倦怠、社交媒体迷失、睡眠问题、感知过载、网络懈怠、强迫性社交媒体使用等一系列不良后果。在应对策略方面,童玉珍等^①(2019)针对决策者在时间紧急及信息不完备的情况下给出评估值时可能会出现犹豫不决的情况,提出基于犹豫模糊集的网络舆情突发事件应急群决策方法。

1.2.4 在线群体研究现状

长期以来,学者们试图了解个人完成集体行动的机制。随着各种在线社交媒体的快速发展,详细的人类行为数据被大规模记录下来,为研究在线群体行为提供了巨大的机会和挑战。图 1-7 和图 1-8 分别是国内外关于在线群体相关研究成果的关键词共现图谱。由图可以看出,国内外关于在线群体的研究主要侧重于以下几个方面。



图 1-7 国内在线群体研究热点

(1) 在线群体行为与动力学分析

研究者通过挖掘群体行为数据构建了一系列数学模型,用于揭示在线群

^① 童玉珍,王应明.基于犹豫模糊集的网络舆情突发事件应急群决策方法[J].计算机系统应用,2019,28(09):9-17. DOI:10.15888/j.cnki.csa.007068.

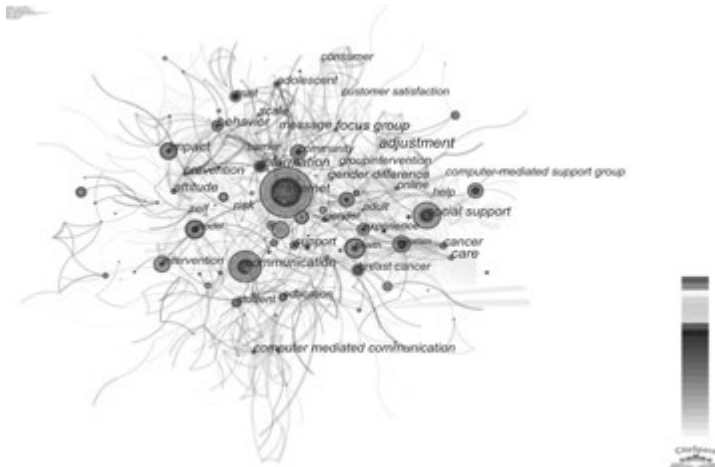


图 1-8 国外在线群体研究热点

体的行为特征和动力学规律。例如, Tang 等^①(2012)提出了一种以边为中心的聚类方案以提取稀疏的社交维度,并通过该维度预测社交媒体群体行为。Candia 等^②(2008)从手机记录中发现了群体呼叫活动的行为模式。Banerjee 等^③(2012)根据博客上的观察数据通过引入一种自然启发理论建模了群体行为。Bahramian 等^④(2021)提出了 Petri 网和第二层耦合神经元组成的两层网络,为受网络连接变化影响的神经元群体行为提供了新的见解。Xiao 等^⑤(2019)将人工蜂群方法与优化理论结合,评估了网络中的同步现象,并在找出合适的网络结构基础上结合新浪微博数据建立了基于 Kuramoto 模型的群

① Tang L, Wang X F, Liu H. Scalable learning of collective behavior[J]. IEEE Transactions on Knowledge and Data Engineer, 2012, 24(6): 1080-1091.

② Candia J, Gonzalez M C, Wang P, et al. Uncovering individual and collective human dynamics from mobile phone records[J]. Journal of Physics A-Mathematical and Theoretical, 2008, 41(22): 224015.

③ Banerjee S, Agarwal N. Analyzing collective behavior from blogs using swarm intelligence [J]. Knowledge and Information Systems, 2012, 33(3): 523-547.

④ Bahramian A, Parastesh F, Pham V T, et al. Collective behavior in a two-layer neuronal network with time-varying chemical connections that are controlled by a Petri net[J]. Chaos, 2021, 31(3): 033138.

⑤ Xiao R B, Li J, Chen T G. Modeling and intelligent optimization of social collective behavior with online public opinion synchronization [J]. International Journal of Machine Learning and Cybernetics, 2019, 10(8): 1979-1996.



体行为网络舆情同步演化模型。

(2) 在线群体认知与群体决策分析

在线群体认知与群体决策分析是在相互依赖、相互作用、相互影响的多个体社交媒体环境中,围绕着共同的话题内容进行交流,在产生群体共识的基础上实现最优决策。在该领域,张涛等^①(2022)以同伴互评和群体认知为基础,采用认知网络分析方法刻画了在线同伴互评中群体认知的特征、网络与发展轨迹。裘江南等^②(2016)指出在线知识社区协同知识建构中影响群体共识的五个因素是成员参与度、知识异质性、认知对象复杂度、沟通协调和群体结构,并构建了群体共识影响模型。

(3) 特殊群体与特殊情境下的在线群体分析

针对老年在线群体数字融入活动存在的困境,朱庆华等^③(2022)探索了老年群体数字融入活动的微观形成机制和宏观变化规律,并针对该群体数字融入困境提供了应对策略。赵宇翔等^④(2019)借助心智模型理论,以网易云音乐为实验平台对高校学生群体开展实验,旨在为音乐平台中的音乐分类体系优化和相应检索模式设计提出建议。为提升知识付费情境下的付费阅读服务水平,王晓光等^⑤(2022)从群体角度精确计算并分析了不同用户群体的感知价值阈值区间,识别了用户群体的差异性。在健康知识推荐情境下,李贺等^⑥(2020)为丁香园糖尿病患者群体设计了一种基于模糊认知图的健康知识推荐模型。

① 张涛,张烁,张思,等. 在线同伴互评中群体认知的特征、网络与发展轨迹分析[J]. 现代教育技术, 2022, 32(09): 82-90.

② 刘丰军,裘江南,张野. OKC协同知识建构中群体共识影响因素研究——群体结构的调节作用[J]. 科学学研究, 2016, 34(10): 1448-1457. DOI: 10.16192/j.cnki.1003-2053.2016.10.002.

③ 薛翔,马海云,赵宇翔,等. 活动理论视角下的老年群体数字融入结构与路径研究[J]. 图书情报知识, 2022, 39(06): 78-88. DOI: 10.13366/j.dik.2022.06.078.

④ 薛翔,赵宇翔. 音乐平台中音乐分类体系的用户心智模型研究——以高校学生群体为例[J]. 数据分析与知识发现, 2019, 3(02): 1-12.

⑤ 姜崇,王晓光,简华. 在线付费阅读中群体感知价值差异研究[J]. 信息资源管理学报, 2022, 12(04): 105-120. DOI: 10.13365/j.jirm.2022.04.105.

⑥ 李贺,刘嘉宇,沈旺,等. 基于模糊认知图的在线健康社区知识推荐研究[J]. 数据分析与知识发现, 2020, 4(12): 55-67.

1.2.5 研究述评

通过对上述研究成果的梳理可以看出,众多学者从不同学科不同视角采用不同方法对复杂社交网络、观点演化、舆情扩散和在线群体进行了研究。总体研究特点和研究趋势述评如下。

从**群体视角**来看,过去的大量研究主要从微观的社交个体或宏观的网络舆情视角探测社交媒体用户的观点交互过程和舆情演化逻辑。社交媒体特别是网络舆情的爆发放大了人与人之间的关联,在线用户仅凭微弱的交互就可以建立联系。中观群体视角为理解群体行为和认知是如何从个人信息生成和群体交互中产生的提供了新框架。群体观点分析能更好地了解在线群体的网络构成、交互特点和共识态度,得到一个更全面和准确的群体用户画像,更好地把握群体的想法信念,从而预测理解群体的行为认知模式。鉴于此,本书拟从群体视角出发,旨在理解社交媒体中的成员在社交舆情下是如何从交互中形成“一种暂时又十分明确的集体心理”,且随着时间推移这种集体心理是如何受群体环境影响发生演变的。

从**技术方法**来看,现有研究只考虑了成对交互关系的普通网络,但现实的社交网络系统的个体之间不仅存在二元交互,也广泛存在多个体同时交互的高阶交互现象。以往基于二元交互结构下的复杂网络或基于仿真建模的观点动力学已无法适应社会事件快速传播,舆情事件频繁反转,人与人互动便捷的网络环境。尽管一些群体观点动力学研究试图捕捉复杂社交系统中群体相互作用的动力学机制,但其受制于研究范式的阻碍,难以真实刻画复杂社交系统中在线群体观点演化的高阶交互特性,在策略的可行性上仍有一定缺陷。综上,本书在分析在线群体观点演化时,拟构建能反映社交媒体多元交互现象的在线群体观点演化高阶交互网络。这种高阶交互网络具有多层次、多维度、多节点特征,可以用来描述和表达用户之间的交互作用和影响,它为研究社交媒体中多用户的相互作用和高丰富度的社交背景提供了新工具。

从**观点演化**来看,既往群体观点演化的研究多是选择群体中的特殊个体、外部环境、用户行为和观点本身探讨上述因素对经典群体观点动力学模型的影响机制,最终结果是基于模型改进的理论层面提出建议措施,这掩盖了观点和意见的强化和争议的消减机制的探究。通过群体观点演化分析理解多个体观点相互作用机理,明晰群体观点随舆情进展的强化和争议的消减



对舆情引导具有启发意义。作为经典的态势感知模型,情景意识理论为本书分析群体观点演化提供了理论支撑。综上,本书拟借助情景意识理论将在线群体观点演化的过程定义为群体用户对舆情情景状态信息的获取、加工及对预测未来的过程。从群体观点的形成、突变和预测实现过去、现在和未来的全生命周期式分析,为深入理解群体观点的动态过程和作用机制提供更全面的洞见。

从舆情引导来看,先前研究在舆情引导策略制定过程更多的是从网络媒体、网民、意见领袖或监管机构等主体入手,依赖这些主体的力量在此消彼长中达到平衡实现网络生态治理,这本质上是一种两方或三方博弈观念。这种依赖仿真模拟得出的引导策略往往建立在对现实世界简化抽象基础上,导致网络社会治理面临网络主体识别困难、网络结构关系复杂和网络生态场景不明等问题。为解决上述问题,本研究拟提出真实数据驱动下的基于实证的在线群体观点演化与引导机制研究,通过采集真实社会事件下的在线用户属性及交互数据,构建数据驱动的在线群体观点演化高阶交互网络,实现舆情背景下群体观点演化分析,并基于演化结果设计在线群体观点差异化引导策略,这在一定程度上克服了以往基于观点动力学的舆情引导策略建议可操作性弱、预测能力不足等问题。

1.3 本书内容与特色之处

1.3.1 研究内容

舆情爆发激化了网络受众在社交媒体使用过程中遭受的“信息茧房”“观点极化”等认知偏差现象,使其产生的危害进一步放大,网络社会治理面临网络主体识别困难、网络结构关系复杂和网络生态场景不明等问题。依据不同主体在不同时空情境下的不同情绪倾向提出差异化引导策略是解决上述问题的重要一环。鉴于此,本书提出了量化网络社会治理中的群体观点演化,主要内容如下。

第1章 绪论。本章主要介绍了研究背景、研究意义和研究现状等内容,同时概括了全文内容、技术路线及创新点等。其中,研究背景以互联网环境

下社交舆情引发的在线用户聚集现象和网络圈群效应等实际问题为切入点,还分析了现有观点动力学研究和高阶交互网络在解释群体观点演化方面的理论缺陷与技术优势,这为本书从未来网络科学前沿角度分析群体观点演化进而解决网络社会治理中三大问题的研究目标与意义做出了铺垫。在文献综述时,本书选择了复杂社交网络、观点演化、舆情扩散、在线群体等与研究内容相关的几个方向,并述评了上述领域的研究进展与不足,这为本书从研究视角、研究方法、研究内容和应用方向上的创新提供了支撑。最后,本章还介绍了研究过程使用的理论模型和技术方法。综上,第1章内容为后文研究的开展提供了基本骨架。

第2章 量化网络社会治理中群体观点演化的新方法:高阶交互网络。该章首先辨析了网络社会治理所面临的主要挑战,探讨了群体观点演化分析如何能够帮助应对挑战。接着,章节深入讨论了传统量化群体观点演化中观点动力学的相关概念、应用领域及经典模型,承认观点动力学在解释群体观点演化的核心作用。随后,本章引入了复杂网络研究的最新进展即高阶交互网络,明确了高阶交互网络的基础概念、应用领域以及技术优势。高阶交互网络因其在处理多主体交互数据中的高效性和灵活性,被本书认为是解释群体观点演化的强有力工具。最后,章节通过探讨高阶交互网络与超图的关系,进一步阐述了基于超图的高阶交互网络是如何提出的,以及超图技术方法在其中的应用,上述内容拓展了高阶交互网络的应用范围,也为分析网络社会治理中的群体观点演化提供了新的方法论。总体而言,本章为理解和应用高阶交互网络在量化网络社会治理中提供了技术铺垫。综上,第2章内容为后文研究的开展提供了技术方法。

第3章 量化网络社会治理中群体观点演化的机理模型。该章旨在从理论层面探究社交舆情背景下在线群体的集聚交互,揭示个体观点到群体观点的涌现路径及群组与群组耦合的观点演化过程,通过梳理社交媒体在线群体观点的形成机制和演化逻辑,构建了量化网络社会治理中群体观点演化的机理模型,为后文研究提供理论框架。具体来说,本章首先解析了在线群体的观点演化现象,界定了在线群体观点演化的内涵、特征、构成要素及影响因素,从上述几方面阐述了在线群体观点演化的理论基础。其次,本书首次提出了社交舆情背景下在线群体观点演化过程存在的高阶交互特性,并借鉴群体动力理论分析了在线群体观点演化的高阶交互动因,明晰了在线群体观点演化的高阶交互类型和关系。在深化对在线群体观点交互理解的同时,为后文基于高阶交互网络分析在线群体观点演化提供了技术支撑。最后,借助情



景意识理论的三维度,在理论解释基础上将在线群体的观点演化过程划分为在线群体的观点形成、突变和预测阶段,将群体观点演化与舆情演变的生命周期全过程对应,说明基于高阶交互网络的在线群体观点演化内容维度。综上,第3章内容为后文研究的开展提供了理论依据。

第4章 量化网络社会治理中群体观点演化的高阶交互网络构建。本章以社交舆情背景下在线群体观点演化的高阶交互性分析结果为依据,通过构建量化网络社会治理中群体观点演化的高阶交互网络,旨在准确刻画社交舆情下的群体集聚现象和网络圈群效应。该章首先基于社会影响理论分析了在线群体观点演化的高阶交互网络构建依据,从信息性影响和社会性影响描述了在线群体观点演化时的高阶交互动力来源。其次,在网络构建环节,本书创新性获取发帖—评论—回复这一能体现舆情事件下用户高阶交互数据来源,通过在线群体用户的个人静态交互数据(粉丝数、好友数等)和动态交互数据(帖子转发数、回复数等),以帖子下集聚的发帖、评论和回复用户为超点,以该帖为超边构建了基于加权超图的在线群体观点演化的高阶交互网络,同时融合静态交互和动态交互数据计算了网络中超边和超点的权重。然后,从节点超度、接近中心度、超边超度、超边重叠度和超边连接度等指标计算了构建的高阶交互网络拓扑结构,并以可视化方式呈现并验证了舆情事件中在线群体的交互集聚网络圈群效应。最后,本章采集了新浪微博平台中“父母犯罪子女能不能考公”等相关6个话题下的动静态交互数据对上述过程进行了实证分析。综上,第4章内容为后文研究的开展提供了方法支撑。

第5章 网络社会治理中群体观点演化的形成机制量化。本章主要是对在线群体观点演化过程第一个阶段即观点形成的解析,对应情景意识理论第一个维度。舆情事件爆发,促使带有先验观点的个体在社交属性驱动下主动或被动通过社交媒体与其他用户互动交流,并在多个体交互过程中经观点融合博弈最终达成了群体共识。为了探讨多个体用户是如何在交互因素的影响下形成群体共识的,本章逆向解析了在线群体观点的形成机制。我们从群体共识识别入手,在直接将体现观点形成最原始交互的帖子、对帖子评论和对评论回复的用户视为一个群组基础上,利用(超)图表示学习建模交互行为并对多交互群组、交互成员和交互文本所反映的观点、偏好等实现了基于图的表示学习,并提出自适应融合机制在多源观点融合基础上识别出了群体共识。群体共识体现的是群组大多数成员对舆情事件的普遍看法态度,忽略了部分妥协用户的个人观点。为了区分群体共识与群组中个体用户的观点差异,本书采用余弦相似度计算了个体用户与群体共识的观点文本语义差异,

并依据语义差异结果实现了基于全脑模型的在线群体分级。此外,本书还从动静交互数据中构建了观点形成阶段影响群体用户分级的特征集,并基于SHAP可解释机器学习对在线群体分级结果进行了分类,感知影响在线群体用户分级分类的环境因素,为后文基于主体差异的分级引导策略制定提供支撑。

第6章 网络社会治理中群体观点演化的突变机制量化。本章主要是对在线群体观点演化过程第二个阶段即观点突变的解析,对应情景意识理论的第二个维度。群体用户会随着舆情事件的推进或新用户的进入,在掌握新信息后推动观点演化系统的突变。网络社会治理中群体观点演化的突变机制量化,首先应识别突变阶段哪些用户的观点产生了突变。因此,本章从信息生态视角构建了影响用户观点突变的特征维度,并采用HGNN分类算法识别了观点突变群体,验证了基于HGNN识别观点突变群体的效果,也证明了社交舆情背景下在线群体观点演化中观点突变现象的存在。其次,本书基于突变理论和经典的尖点突变模型,拟合了舆情事件下在线群体的观点突变现象,并通过指标验证了模型拟合的效果。然后,为直观有效地评估在线群体观点演化过程的群体观点突变现象,本书选择弹性指数构建了在线群体观点突变评估模型,通过分析观点突变、弹性指数和弹性损失三者之间的一致性关系证明了在线群体观点突变评估模型的合理性。最后,将群体观点突变评估模型应用在观点突变和弹性指数的关系分析、观点突变临界点识别以及对情绪变化、弹性指数和观点突变三个指标关系的仿真模拟,为后文基于情绪差异的弹性引导策略制定提供支撑。

第7章 网络社会治理中群体观点演化的预测机制量化。本章主要是对在线群体观点演化过程第三个阶段即观点预测的解析,对应情景意识理论的第三个维度。该章首先基于三元交互理论和人工观点数据标注结果,从交互群组环境、交互成员行为和交互文本体现的认知维度设计了预测阶段在线群体的认知特征集。基于该特征集,利用改进的HG-SDCN识别出了预测阶段的不同观点用户群体(赞同、反对和中立)。然后,分别以3类观点群体文本发布时间为依据,设定5分钟为时间步长,将在该步长内发布文本的个体视为同一个子群组,并利用情感词典的方式衡量了该子群组的情感观点,实现了群组交互序列分析和情感词典分析基础上的在线群体序列观点历史数据构建及预处理。最后,将预处理数据输入LSTM以预测群体观点,并将预测结果与基线模型对比评价模型的观点预测效果性能,为后文基于集聚差异的推拉引导策略制定提供支撑。



第8章 面向网络社会治理问题的群体观点引导。本章是在前文第5章至第7章演化分析结果基础上,根据每个阶段挖掘出的差异化结果,制定在线群体观点精准引导策略。差异化是精准决策的必备条件,是提升舆情引导的重要举措。通过提出面向网络社会治理问题的群体观点引导路径,本章在厘清数据挖掘、知识发现和情报决策的关系基础上,剖析了在线群体观点演化过程中观点形成阶段的主体差异、突变阶段的情绪差异和预测阶段的集聚差异,并针对差异性内容分别制定了分级、弹性和推拉的在线群体观点引导的具体策略。

第9章 研究结论与研究展望。本章是对全书研究内容的总结和研究过程的反思。最后,提出了未来研究工作中可能的改进途径和下一步研究方向。

1.3.2 技术路线

本书以提出网络空间安全治理的新机制和新策略为目标导向,立足社交舆情背景下中观群体用户视角,构建量化网络社会治理中群体观点演化的机理模型,应用高阶交互网络等相关理论与方法,围绕在线群体观点演化过程中的“在线群体观点形成、在线群体观点突变、在线群体观点预测”等具体过程,探索并科学制定舆情事件下在线群体观点的差异化引导策略,以解决网络社会治理面临的网络主体识别困难、网络结构关系复杂和网络生态场景不明等现实问题。本书总体遵循“观察舆情现象→发现研究问题→建构理论框架→构建高阶网络→分析演化过程→总结研究结果→制定引导策略”的研究逻辑,技术路线如图1-9所示。

1.3.3 本书特色

本书从中观群体视角研究社交舆情背景下在线群体的观点演化,通过构建量化网络社会治理中群体观点演化的高阶交互网络,刻画社交媒体在线群体观点演化过程,发现社交媒体在线群体观点演化规律,理论解析在线群体交互特征、社交网络结构特征、观点动力学特征,为在线群体观点的演化提供有效方法,为科学引导在线群体观点提供有力依据。本书的特色与创新之处体现在以下四个方面。

(1) 从中观层面群体集聚交互探讨了在线群体观点的耦合演化,形成了区别以往仅从宏观层面探讨网络舆情形形成演变机制或微观层面剖析个体观

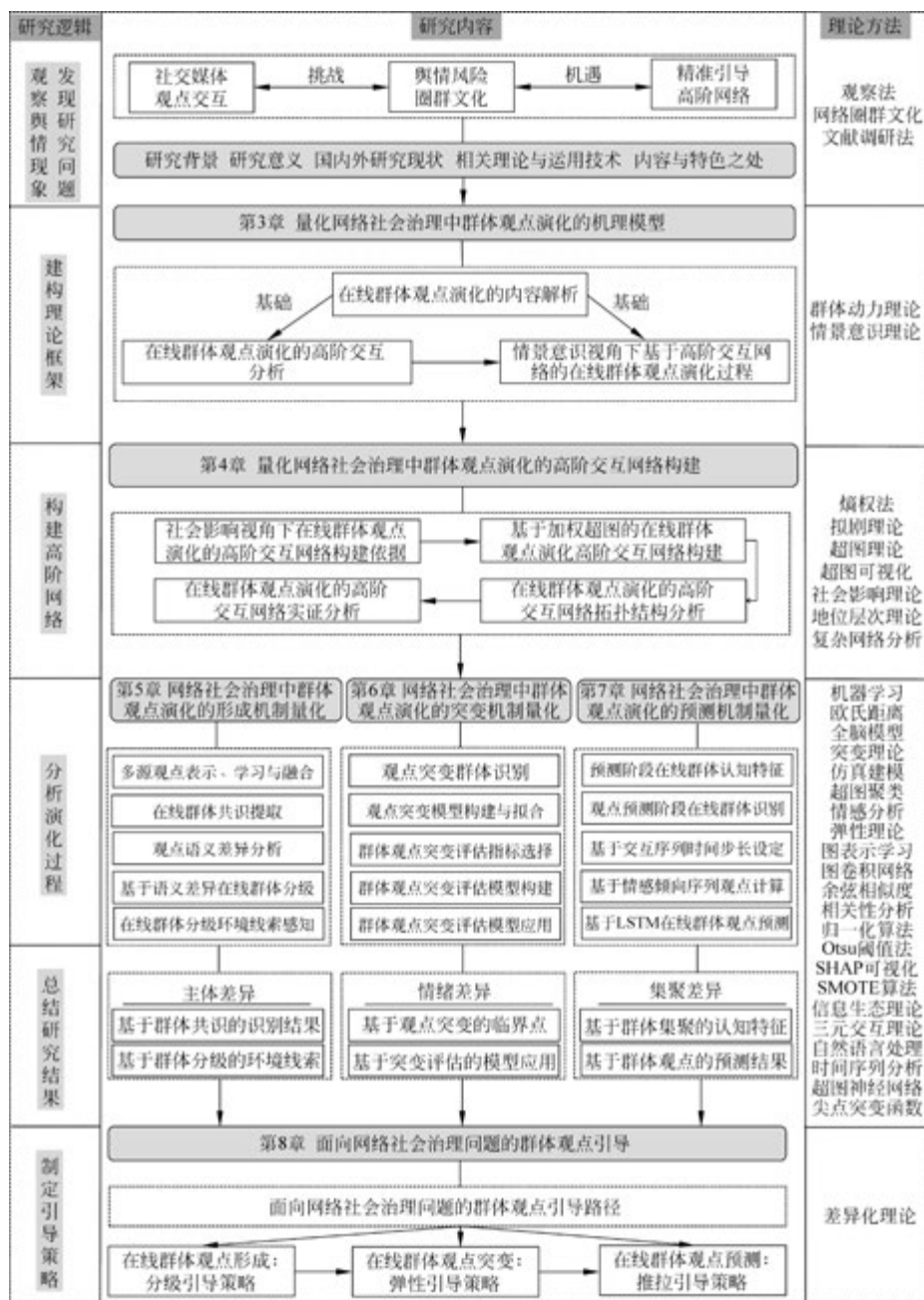


图 1-9 技术路线



点社交影响的新视角。

舆情事件刺激在线个体集聚并引发群体性观点涌现,群体观点在社交媒体中交互融合又影响着宏观舆情的发展走向。个体意识—群体共识—社会舆论构成了舆情事件涉及主体的完整观点体系。立足中观视角分析社交舆情背景下在线群体观点演化,能综合考虑社交媒体系统中各动力学要素(个体、个体—个体、个体—群体、群体—群体、群体—平台环境、个体—平台环境)的认知规则和交互方式,进而更全面准确地描述在线群体观点的演化过程,更好地预测和引导群体观点的发展趋势。基于此,本书聚焦于中观层面的在线群体观点演化,通过构建基于加权超图的在线群体观点演化高阶交互网络,充分利用网络中的超边—超点保存了在线群体中个体—个体、个体—群体、群体—群体间的交互作用关系。

(2) 从复杂网络前沿中的高阶交互网络提出了捕捉社交主体多元结构的在线群体观点演化分析方法,提供了更准确刻画群体观点交互演化的新工具。

以往的复杂网络假设成对交互是系统单元之间连接的主干,忽视了社会系统、生态学和神经网络系统存在的群体互动现象。现实社交媒体系统主要为多元交互关系,舆情的爆发放大了这种社交关系。基于二元交互的复杂关系网络表示难以很好地描述出舆情事件下集群用户的复杂交互。高阶网络(主要是超图和单纯复形)通过描述三个或更多群体的相互作用,弥补了先前复杂网络存在的缺陷^①。因此,本研究利用高阶交互网络构建群体观点涌现和基于观点涌现的群体观点演化模型,能更精确地刻画群体观点交互及观点演化过程。

(3) 从情景意识理论的态势感知定义了在线群体观点演化的形成、突变和预测阶段,构建了群体观点随环境感知而动态演化的生命周期式新框架。

不同于既往频繁研究的观点一致、观点反转和观点极化等观点演化过程表现出的特殊现象,本书借助情景意识理论从个体用户对环境感知产生的自我表征态势角度,提出了在线群体观点的形成、突变和预测。情景意识理论作为经典的认知心理学理论,强调人的思维和知识不是独立于环境的,而是在与环境互动中依靠情境中的具体经验来构建和发展的^②。类似的是,观点

^① Majhi S, Perc M, Ghosh D. Dynamics on higher-order networks: A review[J]. Journal of The Royal Society Interface, 2022, 19(188): 20220043.

^② Endsley M R. Toward a theory of situation awareness in dynamic systems[J]. Human Factors, 1995, 37(1): 32-64.

作为个体主观意愿的表现形式,其演化过程也应具有动态特征。舆情事件下的群体认知过程是根据舆情事件获取到的新信息和所处的网络群组环境不断变化的。因此,本书将在线群体观点演化过程视为动态交互的发展过程,梳理出一种面向生命周期式的在线群体观点演化研究的新逻辑。

(4) 从差异化视角的不同维度制定了真实数据驱动下的在线群体观点引导策略,弥补了先前观点动力学在研究范式上模拟结果与可行措施割裂的应用缺陷。

传统的观点动力学模型是基于代理的,即它们是由网络互连的一组代理根据预先确定的机制进行交互的模拟。该研究范式存在的最显著的弊端在于“实证证据和理论建模之间存在明显的不平衡”^①。因此,探索观点演化过程中发现的机制规律与现实世界的联系,从实证角度实现观点演化分析结果支撑群体决策是目前观点演化分析存在的挑战。本书采用了实证方式实现了特定舆情事件下在线群体观点的演化分析,并基于分析结果的差异设计了分级、弹性和推拉引导策略,为网络舆情监管和群体观点引导提供了真实数据驱动下的可行性措施建议。

1.4 本书涉及相关理论

1.4.1 群体动力理论

群体动力理论(group dynamics theory)为本书第3章的核心理论,目的在于基于该理论实现社交舆情背景下在线群体的高阶交互互动因分析,为后文的高阶交互网络构建提供理论依据。

群体动力理论由美国社会心理学家库尔特·勒温提出,是一个描述群体成员之间相互作用和影响及群体结构和动力的理论模型^②。勒温受物理学中

^① 刘举胜,何建佳,韩景侗,等. 观点动力学研究现状及进展述评[J]. 复杂系统与复杂性科学, 2021,18(02): 9-20. DOI: 10.13306/j.1672-3813.2021.02.002.

^② Lewin K. Principles of topological psychology[M]. New York: McGraw-Hill Book Company, 1936.



场论的启发,提出个体及其所属的环境会形成社会场。在该社会场中,个体的行为会受到社会场中力场的影响。力场是指社会场中各种力量和影响的总和,包括个体之间的相互作用、彼此之间的关系、角色和权力结构等。力场的平衡与不平衡会影响个体的行为和决策。勒温的上述思想用公式表达为

$$B = f(P, E) \quad (1-1)$$

其中, B 表示行为, P 表示个体, E 表示环境。同样,勒温认为群体和群体的环境也会形成社会场。

根据对上述理论含义的分析和金秀平等^①(2021),群体动力理论的基本观点有:第一,群体不仅仅是个体的简单累加,而是一个复杂的有机组合。组合中的个体之间相互影响和相互作用,具有与个体独处时不同的行为表现。第二,群体动力理论包含群体凝聚力、群体驱动力和群体耗散力三种动力要素。群体的发展受到这三种力的综合作用,它们的综合影响可以决定群体的行为和决策。第三,群体目标和群体规范会对群体内部的成员产生影响并对群体成员起到约束作用。

综上,理解群体内个体之间的相互影响和相互作用对于理解群体行为和决策的动态和复杂性非常重要。这种相互影响可以通过群体动力理论来研究和解释,帮助我们更好地理解群体的行为和决策过程^②。

1.4.2 情景意识理论

情景意识理论(situation awareness theory)为本书第3章的又一核心理论,利用该理论旨在确定基于高阶交互网络的在线群体的观点演化过程,表明群体随着对舆情环境交互收集、处理和解释信息以形成对不同情境的认知理解。

1995年由Endsley提出的情景意识理论目的在于研究并改善人们在复杂和动态环境中的决策和行为^③。情景意识指的是一种认知状态,是人们对周围环境中的情境和事件的感知和理解程度。该理论包括三个层次:感知层

^① 金秀平,宗春燕.群体动力理论视角下大学生就业共同体构建[J].创新创业理论与实践,2021,4(22):196-198.

^② 付稼嵘.群体动力理论视域下提升小学生协作能力的教学活动设计研究[D].济南:山东师范大学,2022. DOI: 10.27280/d.cnki.gsdusu.2022.000551.

^③ Endsley M R. Toward a theory of situation awareness in dynamic systems[J]. Human Factors,1995,37(1):32-64.

次、理解层次和预测层次。其中,感知层次即对环境要素的线索直觉,是指对环境中的情境和事件的感知和检测能力。该层次涉及收集和感知环境中的信息,如观察、听觉和触觉等,目标是获取准确和完整的环境信息。理解层次即对整合信息的意义理解,主要是对环境中的情境和事件的理解和解释,以形成对当前情境的认识和理解。该层次的目标是整合感知到的信息,个体形成对环境的准确而全面的认识。预测层次即对其未来状态的预测,是预测环境中情境和事件的未来发展和结果走向。该层次在当前认识和理解上,通过预测未来可能发生的情境和事件而帮助个体做出合理和准确的决策行为。此外,王丽丽等^①(2023)在分析突发公共事件中强关系社交网络谣言扩散机制时,认为突发公共卫生事件等复杂的情景会触发受众的情景意识,环境、个体和信息三要素相互作用形成谣言的扩散意愿。这表明情景意识理论还强调个体会整合各类环境情境状态与信息以及个体因素加工形成对复杂情景的知觉水平与理解^②。

总之,情景意识可以帮助人们更好地适应和应对复杂和动态的环境,提高决策和行为的准确性和效果。它涉及对环境信息的感知、理解和预测,以及对环境中可能发生的情况和事件意识和应对能力。情景意识在许多领域都具有重要的应用价值,如航空、交通、医疗、军事等。

1.4.3 社会影响理论

社会影响理论(social influence theory)为本书第4章的核心理论,基于该理论旨在探讨个体用户在社交网络中加入群体的驱动因素和高阶交互动力形成机制。就本文而言,社会影响理论是对群体动力理论的解释补充。群体动力理论表明群体会在社会场的吸引下产生和个体独处时不一样的行为,且群体交互受到一种驱动力。社会影响理论通过信息性影响和规范性影响将这种驱动力细化为两类,为构建在线群体高阶交互网络提供了动力来源解释。就个体与群体交互后的融合机制而言,社会影响理论通过认同、内化和顺从解释了个体间的交互是如何形成群体性观点的。

^① 王丽丽,张玲玲.突发公共事件中强关系社交网络谣言扩散机制研究——基于情景意识理论[J].情报杂志,2023,42(05):112-118+136.

^② Al-Mekhla A A, Isha A S N, Abdulrab M, et al. Moderating effect of safety culture on the association inter work schedule and driving performance using the theory of situation awareness[J]. Heliyon, 2022, 8(11): e11289.



社会影响是指人们在社会交往中受到他人影响,从而改变自己的行为、态度、观点的过程^①。Deutsch 和 Gread 将社会影响划分为信息性社会影响和规范性社会影响,用来解释人们在社会交往中受到他人影响的不同方式^②。其中,信息性社会影响是指接受从他人那里获得的信息作为现实证据的影响,主要涉及个体对他人的认同和信任。规范性社会影响是指符合他人对群体期望的影响,主要涉及个体对社会关系和群体认同的需求。当个体获取到与决策相关的事实、证据或其他形式的信息时,这些信息会影响决策的治理。上述论述表明信息性社会影响会导致个体重新评估他们的立场。当个人受到规范性社会影响时,无论他们对行为的信仰和态度如何,他们都会感受到更高水平的社会压力,要求他们做出或不做出行为。

此外,Kelman 认为个体的态度、信仰以及随后的行动通过顺从、认同和内化三个阶段受到他人的影响。这种影响差异可以归因于个人接受社会影响的过程差异。顺从是指被影响的个体为避免惩罚与获得奖励而做出符合他人的行为。认同是指被影响的个体为了与所在的群体建立良好的关系而产生的一种归属感。内化是指群体接受了所在群体的价值观并将其视为自己的信仰,并自愿采取行动实现整个群体的目标。由于社会影响可以塑造人的态度、信仰和行动,因此社会影响对信息系统接受和使用的影响已被广泛研究。

1.4.4 全脑模型

全脑模型(whole brain model)为本书第 5 章的核心理论。本文在对用户观点和群体共识的文本语义差异分析上,依据该理论实现了观点形成阶段群体用户的分级,以剖析在线用户在观点形成阶段针对舆情事件的认知风格。

全脑模型是由美国认知心理学家 Herrmann 于 20 世纪 70 年代提出的一种心智模型^③。心智模型塑造了一个人看待世界的方式,它是个体对于外部

^① Chou C H,Wang Y S,Tang T I. Exploring the determinants of knowledge adoption in virtual communities: A social influence perspective[J]. International Journal of Information Management, 2015,35(3): 364-376.

^② Li C Y. Persuasive messages on information system acceptance: A theoretical extension of elaboration likelihood model and social influence theory[J]. Computers In Human Behavior, 2013, 29(1): 264-275.

^③ Herrmann N. The creative brain[J]. The Journal of Creative Behavior, 1991, 25: 275-295.

世界的理解和思维方式的总结和归纳。基于人体大脑每个半球(左脑、右脑、左脑边缘区和右脑边缘区)的特征,Herrmann 提出了大脑的四象限模型,如图 1-10 所示。该模型基于大脑的四个主要功能区域,将人们的认知偏好划分为四个象限^①:

第一象限是左脑区域(A),属于这个象限的人被描述为“理性我”,他们理性、逻辑性强,好分析,重事实,善于强调量化。因此,属于这个象限的人会对某一特定情况做出逻辑上的、基于事实的判断。

第二象限是左脑边缘区(B),属于这个象限的人被描述为“守成我”,他们占主导地位,厌恶风险,保守,重规则,重细节。这些人根据可靠或详细的数据做出保守的判断。

第三象限是右脑边缘区(C),属于这个象限的人被描述为“感觉我”,他们具有直觉、善交际,重感觉,重运动感觉,情绪主导等特征。这些人在思考某一特定情况时,会非常专注于他们的所见所闻和感受。

第四象限是右脑区域(D),属于该区域的人被描述为“探索我”,他们善于用直觉的,整体的,融会贯通的,演绎推理的认识与处理问题。每当思考一种现象时,这些人都会将收集到的信息或数据概念化或形象化。

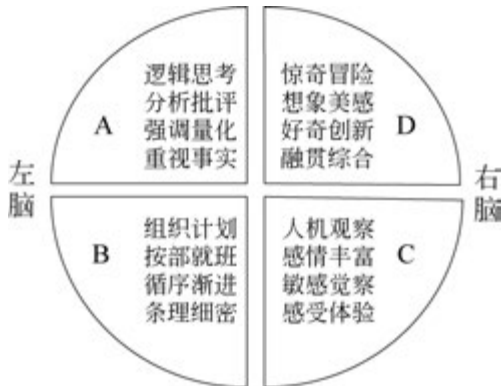


图 1-10 全脑模型

全脑模型认为每个人都有上述四种思维方式,但个体可能存在偏好某种思维方式的倾向。通过了解在线个体的认知偏好,可以更好地理解个体在思

^① Ham J, Lee K, Kim T, et al. Subjective perception patterns of online reviews: A comparison of utilitarian and hedonic values[J]. Information Processing & Management, 2019, 56(4): 1439-1456.



考、决策和解决问题时的观点倾向。

1.4.5 突变理论

突变理论(catastrophe theory)为本书第6章的核心理论,为在线群体观点演化过程中的观点突变提供了理论依据。1972年,法国数学家勒内·托姆在《结构稳定性和形态发生学》一书中系统阐述了突变理论,是对自然界不连续变化现象的系统总结^①。

突变理论也称灾变论,核心思想是表示复杂系统从一种稳定状态跃迁到另一种稳定状态的现象规律^②。该理论认为复杂系统可以具有稳定的平衡状态,但当某些控制变量越过临界阈值时,系统可能会发生突然且不连续的变化。突变理论认为控制变量影响系统的状态和行为,当控制变量处于某个范围内时,系统可能保持稳定的平衡状态。一旦控制变量越过临界阈值,系统可能会遭受突变^③。通过研究控制变量和系统行为间的关系,我们可以理解复杂系统中的突变。

突变理论根据系统所处的时空维度(三维空间和一维时间)将控制变量的数量限定在4以内,以表征系统具体状态的势函数^④。当控制变量小于4时,常见的突变模型包括折叠突变模型、尖点突变模型、燕尾突变模型和蝴蝶突变模型四种。表1-1中的 l, u, v, b, w 分别表示控制变量,势函数 $f(x)$ 为研究对象, x 为状态变量,归一化公式表示控制变量的不同质态^⑤。

突变理论已广泛应用于各个领域,包括物理学、生物学、心理学和经济学,被用来解释相变、动物行为、决策过程和市场崩盘等突变现象,是理解复杂系统动态变化的一种有价值工具。

① 程津,李桂明. 基于改进突变理论的钻井作业风险评价[J]. 化工安全与环境, 2023, 36(10): 31-34.

② 侯广辉,廖桂铭,王刚. 基于突变级数的颠覆性技术识别模型构建及实证研究[J]. 情报杂志, 2021, 40(10): 7-14.

③ 曾紫琳,胡斌,王智超,等. 网络社区群体观点演化的弹性与突变机制[J/OL]. 中国管理科学: 1-16[2023-09-24]. DOI: 10.16381/j.cnki.issn1003-207x.2023.0247.

④ Hu B, Xia N. Cusp catastrophe model for sudden changes in a person's behavior[J]. Information Sciences, 2015, 294: 489-512.

⑤ 程津,李桂明. 基于改进突变理论的钻井作业风险评价[J]. 化工安全与环境, 2023, 36(10): 31-34.

表 1-1 常见突变模型及其势函数、归一化公式表

突变模型	控制变量数	势函数	归一化公式
折叠	1	$f(x) = x^3 + ux$	$x_a = \sqrt{a}$
尖点	2	$f(x) = x^4 + ux^2 + vx$	$x_a = \sqrt{a}, x_b = \sqrt[3]{b}$
燕尾	3	$f(x) = x^5 + ux^3 + bx^2 + wx$	$x_a = \sqrt{a}, x_b = \sqrt[3]{b}, x_c = \sqrt[4]{c}$
蝴蝶	4	$f(x) = x^6 + lx^4 + ux^3 + vx^2 + wx$	$x_a = \sqrt{a}, x_b = \sqrt[3]{b}, x_c = \sqrt[4]{c},$ $x_d = \sqrt[5]{d}$

1.4.6 弹性理论

弹性理论(resilience theory)为本书第6章的又一核心理论。本文旨在利用弹性指数这一指标,从复杂系统内部快速评估在线群体观点演化过程中的突变现象,更直观反映舆情事件下观点演化系统内部的稳定与受众的情绪累积能力,为缓解群体观点突变现象提出针对性引导策略。

不同的研究领域对弹性有不同的定义。但从其初始含义看,弹性都表示了一种反弹的能力,即个体或系统在外部压力或逆境中恢复的能力。根据 Reggiani 等^①(2002),弹性也可以用于描述系统从一个稳定状态到另一个稳定状态。基于此,弹性在描述系统的突变现象领域得到了广泛的应用。在复杂科学领域,对弹性的研究重点在系统弹性的测度^②。相关研究人员提出了“弹性指数”以衡量复杂系统中的累积压力,并预测系统发生质变向另一个状态的转换^③。

弹性理论和突变理论结合,允许决策者可以直接依据弹性指数判断系统的突变情况,而无须剖析系统的复杂结构和影响因素,对提升决策效率、提高决策效果具有积极意义。例如,李根等^④(2023)采用突变级数模型测算了

① Reggiani A, Graaff T D, Nijkamp P. Resilience: An evolutionary approach to spatial economic systems[J]. Networks & Spatial Economics, 2002, 2(2): 211-229.

② 危小超,余其平,聂规划. 基于随机突变的电商平台信用监管弹性研究[J/OL]. 管理工程学报: 1-15[2023-09-24]. <https://doi.org/10.13587/j.cnki.jieem.2024.01.020>.

③ Hu B, Wang Z C, Du Y X, et al. A resilience index of online group opinion[J]. Soft Computing, 2022, 27(3): 1265-1287.

④ 李根,刘欣雨,刘家国,等. 产业协同集聚如何影响制造业绿色韧性? ——基于突变级数与双固定动态空间杜宾模型的实证分析[J/OL]. 中国管理科学: 1-11[2023-09-24]. <https://doi.org/10.16381/j.cnki.issn1003-207x.2022.2766>.



2005—2020 年我国 30 个省市的制造业绿色弹性,并从产业协同集聚角度对提高制造业绿色弹性提出了建议。陆可晶等^①(2022)探索了环境突变的不同阶段中组织韧性构建的关键因素及其干预发挥作用的动态过程,为完善中国情境下的组织韧性理论提供了参考。

总的来说,与量化复杂系统的内部结构实现系统稳定状态衡量相比,弹性理论中的弹性指数可以直观地通过指数的变化实现复杂系统发展演化趋势探究,是在系统突变基础上衡量系统状态的常见指标。

1.4.7 三元交互理论

三元交互理论(triadic theory)为本书第 7 章的核心理论,目的是从交互用户个体特征、交互平台网络环境和交互文本行为三个角度构建群体观点预测阶段的在线群体认知特征集。该特征集与超图聚类算法结合,可识别预测阶段的在线群体。

班杜拉提出的三元交互模型的核心思想是人的内部因素、行为和环境影响三者之间互为决定因素,三个决定因素间通过连续不断交互作用解释了社会情境中人的行为是如何起源的^②。该理论是对斯金纳环境决定论的彻底批判,通过强调观察学习的重要性拓展了社会学习理论。班杜拉的三元交互理论注重解释个体特别是认知因素(P)、环境(E)和行为(B)之间的交互机制。其中,个体是指个体的认知、情感、动机等内在因素;环境是指个体所处的社会、文化、物理环境等外部因素;行为是指个体的行动和反应^③。具体包含以下三个方面的交互关系,如图 1-11 所示。

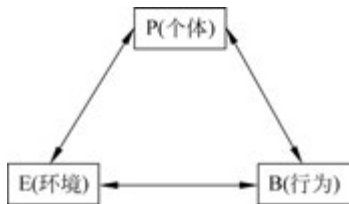


图 1-11 三元交互模型

① P—E 的交互关系:个体会根据自己的认知和情感状态对环境做出反应和行动。反过来,人的信念、认知等发展变化是社会影响力的结果。环境决定人的同时,人也决定着环境。

① 陆可晶,罗仲伟.环境突变下中小制造企业的组织韧性研究[J].价格理论与实践,2022(04):20-27. DOI: 10.19851/j.cnki.cn11-1010/f.2022.04.118.

② 张莉,孙晓宁,朱庆华.基于三元交互模型的社会化搜索用户行为研究[J].情报理论与实践,2021,44(04):40-49. DOI: 10.16353/j.cnki.1000-7490.2021.04.007.

③ 金帅岐,李贺,沈旺,等.用户健康信息搜寻行为的影响因素研究——基于社会认知理论三元交互模型[J].情报科学,2020,38(06):53-61+75. DOI: 10.13833/j.issn.1007-7634.2020.06.008.

②P—B的交互关系：这一交互关系说明人的信念、意向等支配着人的行为，即人的思维决定了行为；人的行为指向一定的主体，行为及其结果能够反映出主体思维的内容与形式。③E—B的交互关系：个体通过行为这一中介改变环境使之符合自己生存的需要。环境潜移默化地影响人的行为，在某些情况下环境对人的行为起着决定性作用。

1.4.8 差异化理论

差异化理论(differentiation theory)为本书第8章的核心理论。该理论旨在连接基于高阶交互网络的在线群体观点演化引导路径中的数据挖掘层和情报智慧层,通过制定差异化引导策略实现决策辅助,为社交媒体生态和数字治理生态的平衡提供差异化对策建议。

差异化理论脱胎于迈克尔·波特提出的差异化竞争战略,是指企业在营销过程中应对从技术、功能、整体形象和个性化服务上全面提高产品的价值,形成产品独特的市场地位。后来,企业管理领域相关学者直接采用了差异化理论这一概念,以分析具体情境下如何提升产品竞争性问题。例如,冯玲玲等^①(2022)以产品差异化理论为基础,在分析喀拉峻景区产品差异化中存在问题基础上,探索了喀拉峻景区旅游产品差异化的策略。郭富红等^②(2016)从产品差异化程度越高市场势力越大的视角,综述了产品差异化理论的相关研究。

为帮助企业在市场中建立独特的品牌形象和市场定位,吸引更多的消费者选择自己的产品或服务,企业可以通过提供独特、与众不同的产品或服务来满足消费者的特定需求。基于上述思想,本书拟借鉴这一理论制定在线群体观点演化的差异化引导策略,从在线群体观点的形成阶段、突变阶段和演化阶段提出有区别性的群体观点引导策略,避免一概而全的引导致使策略丧失价值和意义。

① 冯玲玲,王友文,刘文. 基于产品差异化理论的草原旅游竞争策略研究——以新疆喀拉峻国际生态旅游区为例[J]. 经济研究导刊, 2020(04): 151-157.

② 郭富红,陈艳莹. 产品差异化理论研究综述——基于产品差异化程度越高市场势力越大视角[J]. 现代管理科学, 2016(09): 97-99.



1.5 本书运用相关技术

1.5.1 图相关技术方法

超图建立在图模型上,研究超图首先得理解图模型。从定义看,图模型是一种由点和线组成的用以描述系统的数据结构。从功能作用看,图模型不仅能够作为数据挖掘解决方案的一种补充,提供基于语义的思考框架,而且能从宏观上考虑各数据仓储的相互联系。从本书的研究看,我们采用基于超图的高阶交互网络实现了群体观点演化分析和群体社交结构挖掘。书中相关的图模型技术包括:

(1) 图表示学习

图表示学习是指通过学习将图中的节点或边表示为低维稠密的向量化表示方法,以便于后续的图分析和机器学习任务^①。从算法层面来看,图表示学习可以分为基于随机游走的图表示学习、基于矩阵分解的图表示学习和基于深度学习的图表示学习。其中,基于随机游走的图表示学习以采样方式对图中的节点关系和图的拓扑结构学习。基于矩阵分解的图表示学习是一种将图结构转化为矩阵形式,并通过矩阵分解的方法来学习节点的表示方法。该类算法常见的有非负矩阵分解、谱分解等算法。基于深度学习的图表示学习主要通过神经网络模型来学习节点和图的表示。

(2) 图神经网络(graph neural networks,GNN)

本书在图表示学习中运用的是基于 GNN 的图表示学习。GNN 是一类基于神经网络来学习非结构化数据,并提取和发掘图结构数据特征和模式的算法总称^②。作为深度学习算法和图算法的结合,GNN 的优势在于以下三个方面:第一,适用于结构不规则的非欧几里得空间图数据;第二,具备推理能

^① Li Z F,Zhang Q,Zhu F F,et al. Knowledge graph representation learning with simplifying hierarchical feature propagation[J]. Information Processing & Management,2023,60(4): 103348.

^② 肖国庆,李雪琪,陈玥丹,等. 大规模图神经网络研究综述[J/OL]. 计算机学报: 1-31[2023-09-24]. <http://kns.cnki.net/kcms/detail/11.1826.tp.20230817.0856.002.html>.

力,通过图拓扑结构的分析完成关系推理;第三,高效利用样本实例间的结构特性。常见的 GNN 模型包括图循环神经网络、图卷积神经网络、图注意力网络等,这些模型在图数据的表示学习中取得了良好效果,并在实际应用中得到了广泛应用。

(3) 图卷积神经网络(graph convolutional network,GCN)

本书在第 5 章实现多源观点表示学习时,采用了 GCN 算法。作为 GNN 的一种,GCN 通过在图上进行卷积操作来捕捉节点之间的关系和拓扑结构。在神经网络中,卷积操作是提取输入数据特征的常用操作。GCN 是通过卷积操作对相邻节点信息的吸收获取后来表示节点的^①。作为 GNN 的代表,图卷积神经网络本质上仍属于特征提取器,是卷积神经网络在非结构化图数据上的延伸。GCN 主要包含两类:一类是基于空域的,另一类是基于频域的。基于空域卷积的方法可以直接将卷积操作定义在节点的连接关系上;基于频域卷积的方法是一种利用图信号的频域特征进行卷积操作的深度学习模型。

1.5.2 自然语言处理技术方法

自然语言处理技术是利用计算机对自然语言的形、音、义等信息进行加工处理的过程,即对自然语言的基本构成单位字、词、句以及篇章的识别、分析、理解、生成的操作。自然语言处理技术可以应用于文字识别、分词、句读、实体识别、情感分析和信息检索等多个方面。书中在对用户的交互文本数据进行处理时,用到了两种应用广泛的自然语言处理工具。

(1) 语言探索和词频统计(linguistic inquiry and word count,LIWC)

LIWC 是由 Pennebaker 等开发的一种用于文本分析的计算机软件,在心理学和认知行为领域得到了广泛的应用,其信效度也得到了认可。通过对文本包含的词汇进行分析,该软件提供了对文本中的语言和心理特征的量化评估^②。软件依赖一个内部默认词典,利用词典规则可以对用户发布的文本内容进行分词、归类,进而计算出用户在文本内容中表达的认知、态度和情感等人格特质。该软件的工作原理是会在附带的字典中搜索词语匹配项,如果出

^① 唐雪梅,苏祺,王军,等.基于图卷积神经网络的古汉语分词研究[J].情报学报,2023,42(06):740-750.

^② 张凯亮,臧国全,路杭霖,等.社交网络用户自我披露水平与其人格特质之间的关联研究[J].现代情报,2022,42(06):49-56.



现匹配项,则该词对应的类别比例将增加。当搜索过程结束时,将向每个类别添加一个最终分数,表示文本样本中与该类别匹配的关联词的百分比。

(2) 自然语言处理与信息检索技术(natural language processing and information retrieval, NLP/IR)

由张华平团队开发的 NLP/IR 提供了自然语言处理、文本挖掘和网络搜索技术的 13 项功能,是数智时代下大数据语义智能分析平台^①。NLP/IR 可以应用于各种中文文本处理的任务,包括信息检索、文本挖掘、舆情分析、机器翻译等。它提供了简单易用的接口和丰富的功能,可以帮助用户高效地处理中文文本数据。书中我们利用该工具分析了用户发布文本中的词性特征、语言学特征等。

1.5.3 机器学习技术方法

机器学习是一门多领域交叉学科,专门研究计算机怎样模拟或实现人类的学习行为,以获取新的知识或技能,重新组织已有的知识结构使之不断改善自身的性能^②。机器学习技术的应用可以通过算法、模型,训练计算机完成自主学习,进而实现信息资源的自动化识别、组织等。本书在实验过程中用到的关于机器学习的方法如下。

(1) 特征工程

特征工程是机器学习中一个重要环节,它涉及对原始数据进行处理和转换,以便更好地适应机器学习算法的输入要求和提取更有用的特征。特征工程的目的是通过选择、构造和转换特征,使得模型能够更好地理解数据,并提高模型的性能。

机器学习中常见的特征工程技术包括下面几种:①特征选择,即选择对目标变量有重要影响的特征,排除无关或冗余的特征。常用的特征选择方法包括相关性分析、卡方检验、互信息等^③。②特征构造,即通过对原始特征进行组合、衍生或转换,生成新的特征。③缺失值处理,即处理数据中存在的缺

① 张华平,商建云. NLP/IR-Parser: 大数据语义智能分析平台[J]. 语料库语言学, 2019, 6(01): 87-104.

② 刘景江,郑畅然,洪永森. 机器学习如何赋能管理学研究? ——国内外前沿综述和未来展望[J]. 管理世界, 2023, 39(09): 191-216. DOI: 10.19744/j.cnki.11-1235/f.2023.0106.

③ Li Z M, Zhang Q, Du X Y, et al. Social media rumor refutation effectiveness: Evaluation, modelling and enhancement[J]. Information Processing & Management, 2021, 58(1): 102420.

失值。常见的方法包括删除法、填充法(如均值、中位数、众数填充)或使用模型预测缺失值。④标准化和归一化:将特征数据进行缩放,使其具有相似的尺度和分布。常见的方法包括Z分数(Z-score)法、最大—最小(max-min)法等。⑤特征编码:将非数值特征转换为数值特征,以便机器学习算法能够处理。常见的方法包括独热编码(one-hot)、标签编码等。⑥特征降维,即通过降低特征的维度,减少数据的复杂性和冗余性,以提高模型的效率和泛化能力。

机器学习中特征工程的选择和实施需根据具体问题和数据集进行,它可以帮助提取更有效特征,改善模型的性能,并提高算法的准确性和效果。

(2) 不平衡数据处理

分类、聚类是机器学习最常见的任务场景。但在许多真实数据集中,很难收集到平衡的数据集实现模型训练,这就导致了数据不平衡的问题即不同类别的样本数不同^①。数据不平衡会显著影响模型性能,许多在平衡数据集上表现良好的模型在遇到不平衡数据集时性能往往表现不佳。

计算机领域解决分类任务中存在的的天数据不平衡问题往往采取以下方法:①重采样,即通过欠采样或过采样的方式来平衡数据集中不同类别的数量。欠采样主要是移除数据量较多类别中的部分数据,但数据删除易引发信息丢失问题。过采样主要是对小类的数据样本进行采样来增加小类数据样本个数。②增加权重,即对不同类别的样本引入不同的权重,以使模型更加关注少数类别。③数据增强,即通过增加数据集中数量较少的类别样本数量,以使其数量与其他类别的样本数量更加接近。基于增强数据的多样性可以分成释义、噪声和采样三类。

为克服书中应用真实数据但存在的类不平衡情况,我们借鉴了计算机领域解决类不平衡的思路以保证机器学习算法结果的准确性。

1.5.4 情感分析技术方法

情感分析又称意见挖掘,是基于自然语言处理和文本挖掘技术来识别和分析文本、语音或图像中蕴含的情感、意见或情绪^②。从粗粒度来看,情感分

^① 刘嘉宇,李贺,谷莹,等.不平衡数据集上在线评论有用性识别研究[J].情报理论与实践,2023,46(11):119-125+153. DOI:10.16353/j.cnki.1000-7490.2023.11.016.

^② 叶富豪,许燕,朱一杰,等.网民对“人祸”事件的道德情绪特点——基于微博大数据研究[J].心理学报,2016,48(03):290-304.



析的结果可以是积极、消极或中性,并可进一步细分为不同的情感类别如喜欢、愤怒、悲伤等。但现有的情感分析侧重于细粒度情感挖掘,如从不同层次粒度、从知识抽取的角度实现情感分析等^①。

从不同层次粒度的情感分析来看,情感分析任务可以划分为句子级情感分析、方面级情感分析和篇章级情感分析。其中,句子级情感分析是通过分析句子中的词语、语义、语境等信息来确定句子的情感倾向,在用户评论、社交媒体帖子、新闻标题等情感分析任务中,帮助人们了解他人对特定主题或事件的情感态度。方面级情感分析旨在明确特定方面相关的情感,一般包括方面术语、观点词和情感极性三要素,方面级情感分析可以通过识别和提取文本中的关键词、实体或短语来确定分析方面。篇章级情感分析是指对整篇文本或篇章进行情感分析的方法。与句子级情感分析或方面级情感分析不同,篇章级情感分析关注的是整体的情感倾向和情感表达。

从知识抽取的角度实现情感分析主要是对(实体、方面/属性、观点、观点持有者、情感)五元组的挖掘抽取,深度学习、词典匹配是抽取过程中常用技术。

1.5.5 时间序列分析技术方法

随机事件按照顺序将其发展变化的过程记录下来就构成了一个时间序列。在此基础上对时间序列观察研究,构建数学模型挖掘其发展变化规律,预测未来趋势和模式就是时间序列分析^②。时间序列分析提供了一种能够从数量上刻画某一事物发展变化动态规律的方法论,也通过阐述现象间的内在数量关系帮助决策者预测和控制现象的未来行为状态。按照发展历程,时间序列分析大致可以分为基于统计的时间序列分析阶段、基于机器学习的时间序列分析阶段和基于深度学习的时间序列分析阶段^③。

常见的基于统计的时间序列分析模型方法包括自回归移动平均模型

^① 谭翠萍. 文本细粒度情感分析研究综述[J]. 大学图书馆学报, 2022, 40(04): 85-99+119. DOI: 10.16603/j.issn1002-1027.2022.04.011.

^② 李海林, 邬先利. 基于时间序列聚类主题发现与演化分析研究[J]. 情报学报, 2019, 38(10): 1041-1050.

^③ Pierre A A, Akim S A, Semeno A K, et al. Peak electrical energy consumption prediction by ARIMA, LSTM, GRU, ARIMA-LSTM and ARIMA-GRU Approaches[J]. Energies, 2023, 16(12): 4739.

(auto-regressive moving average model, ARMA)、差分整合移动平均自回归模型(auto-regressive integrated moving average model, ARIMA)和季节性差分自回归移动平均模型(seasonal auto-regressive integrated moving average, SARIMA)等。支持向量回归(support vector regression, SVR)、随机森林(random forest, RF)等是基于机器学习的时间序列分析中常用的模型。基于深度学习的时间序列分析主要是深度学习模型在时间序列分析任务中的应用^①。例如,常见的 LSTM 和门控循环单元是基于循环神经网络(recurrent neural network, RNN)时间序列分析模型的变体。

时间序列分析已经被广泛应用于电力系统监测、交通流量预测、血糖检测、股票预测等领域。它可以帮助人们做出更准确的未来预测,为决策提供参考和支持。

^① 张艳丰,彭丽徽,洪闯. 在线用户追评行为时间序列关联特征实证研究——以京东商城手机评论数据为例[J]. 情报理论与实践, 2019, 42(03): 139-145. DOI: 10.16353/j.cnki.1000-7490.2019.03.024.