

引用格式:廉莹,罗峥,沈浩,马宁,董雪璠.数据驱动融合与创新:新闻传播学领域大数据研究综述(2013年-2023年)[J].信息传播研究,2024,31(06):24-41.

文章编号:2097-4930(2024)06-0024-18

## 数据驱动融合与创新:新闻传播学领域 大数据研究综述(2013年-2023年)

廉莹<sup>1\*</sup>,罗峥<sup>1</sup>,沈浩<sup>2</sup>,马宁<sup>3</sup>,董雪璠<sup>4</sup>

(1. 中国传媒大学新闻学院,北京 100024; 2. 中国传媒大学媒体融合与传播国家重点实验室,北京 100024;  
3. 中国科学院科技战略咨询研究院,北京 100190; 4. 北京工业大学经济与管理学院,北京 100124)

**摘要:**本文基于中国知网期刊数据库和Web of Science核心合集数据库,系统梳理了2013年-2023年间大数据技术在新闻传播领域的应用与影响。通过文献计量方法和网络分析方法,系统研究了文献的时间分布、高被引论文、国家与机构贡献、期刊分布、关键词演变及作者合作网络。研究结果揭示了新闻传播学在数据驱动下的核心趋势和研究热点,并指出其理论、研究对象、方法论和伦理层面面临的新挑战与机遇。本文提出未来研究应聚焦于数字化媒体和数据新闻的理论框架构建、大型语言模型在新闻分析和推荐中的应用、以及数据隐私保护和算法透明度的伦理问题。此研究为新闻传播学领域的研究者提供了理论、方法与伦理层面的重要参考,助力更好地应对智能时代的挑战。

**关键词:**新闻传播学;大数据;文献计量分析;数字媒体

**中图分类号:**G210.7 **文献标识码:**A

## Data-driven integration and innovation: a review of big data research in journalism and communication (2013-2023)

LIAN Ying<sup>1\*</sup>, LUO Zheng<sup>1</sup>, SHEN Hao<sup>2</sup>, MA Ning<sup>3</sup>, DONG Xuefan<sup>4</sup>

(1. School of Journalism, Communication University of China, Beijing 100024, China; 2. State Key Laboratory of Media Convergence and Communication, Communication University of China, Beijing 100024, China; 3. Institutes of Science and Development, Chinese Academy of Sciences, Beijing 100190, China; 4. College of Economics and Management, Beijing University of Technology, Beijing 100124, China)

**Abstract:** Based on the China National Knowledge Infrastructure and Web of Science Core Collection databases, the application and impact of big data technology in the field of journalism and communication from 2013 to 2023 were systematically reviewed in this paper. Using bibliometric and network analysis methods, an in-depth analysis of the temporal distribution of literature, highly cited papers, contributions from different countries and institutions, journal distribution, keyword evolution, and author collaboration networks were provided. The findings reveal core trends and research hotspots in journalism and communication in the era of data-driven approaches, highlighting new challenges and opportunities in terms of theory, research subjects, methodology, and ethics. It was suggested that future

**基金项目:**国家自然科学基金(72404250;72274010;72074206)

**作者简介(\*为通讯作者):**廉莹(1989-),女,博士,教授,主要从事计算传播研究。E-mail: lianying@cuc.edu.cn;罗峥(2001-),女,硕士研究生,主要从事计算传播研究。E-mail: luozheng@163.com;沈浩(1963-),男,博士,教授,主要从事大数据与人工智能、舆情分析相关研究。E-mail: shenhao@cuc.edu.cn;马宁(1986-),女,博士,副研究员,主要从事舆情风险治理研究。E-mail: maning@casisd.cn;董雪璠(1991-),男,博士,研究员,主要从事信息传播研究。E-mail: dongxf@bjut.edu.cn

research should focus on constructing theoretical frameworks for digital media and data journalism, applying large language models for news analysis and recommendation, and addressing ethical considerations related to data privacy and algorithm transparency. Valuable theoretical, methodological, and ethical insights in the field of journalism and communication were offered for scholars and practitioners, helping them better adapt to the challenges of the digital age.

**Keywords:** journalism and communication; big data; bibliometric analysis; digital media

## 1 引言

自2013年起,大数据技术在学术界引发广泛关注,促成了新闻传播领域的根本变革。这一时期,大数据技术得到了迅猛发展,全面覆盖了数据采集、存储、分析和可视化等关键环节,同时,由于分析软件用户界面的简化和操作流程的优化,非数据领域的研究者也可以应对管理和分析大规模数据。这种易用性的提高不仅加快了数据的收集和发布过程,也使得研究者能够更快地识别数据趋势、分析受众反馈并制定相应的策略,极大地增强了处理和理解信息的能力<sup>[1]</sup>。在此背景下,越来越多的新闻传播领域研究者开始利用大数据探究新闻传播学领域的关键科学问题,主要涉及研究内容、研究方法和研究伦理三个维度。

研究内容层面,已有相关研究借助大数据分析技术聚焦在新闻话题的演变、信息传播机制及其对公众影响多样性等方面。其中,新闻话题的演变分析主要集中在利用大数据技术跟踪新闻话题的演变过程,包括了解特定话题的热度如何随时间变化,以及不同新闻媒体如何报道和呈现相同话题,并揭示媒体报道的倾向性和观众兴趣的变化<sup>[2-3]</sup>。例如,郑路等<sup>[3]</sup>从社会、时间和空间三个维度对移动新闻客户端用户的阅读行为社会关注特征和影响因素展开了系统性分析;申楠等<sup>[2]</sup>基于推特(Twitter)平台,探讨了“一带一路”相关话题,以了解国际舆论对该议题的态度和反应,进而提出中国应对策略。同时,数据新闻的实践应用也为新闻传播领域提供了抓手,帮助新闻媒体更好地理解其受众和提高内容质量<sup>[4-6]</sup>。信息传播机制方面,已有研究倾向选用社交媒体平台作为主要的研究工具,以此探索信息的传播路径<sup>[7-8]</sup>,并利用大规模社交媒体数据进行实时分析,量化评估新闻信息的传播速度与范围<sup>[9]</sup>。此外,通过深入研究社会网络的结构及其动态变化,研究者能够揭示信息如何在网络中扩散<sup>[10-11]</sup>,以及这些网络结构如何影响信息的流通速度和覆盖面<sup>[12-13]</sup>。受众反应和情感分析主要聚焦在利用大数据技术分析社交媒体上的评论及其背后情感变化

等<sup>[14-15]</sup>。通过这种方式,研究者可以了解新闻报道如何影响受众情感和态度,以及如何改善报道以更好地满足受众需求<sup>[16-17]</sup>。同时,在广告领域,大数据分析也被用于评估广告效果,包括广告点击率、受众反应、广告投放时机和渠道的选择,从而帮助提高广告投放边际效用<sup>[18]</sup>。

研究方法层面,机器学习和自然语言处理技术的快速发展极大地推动了新闻传播学研究的深化<sup>[19-20]</sup>。这些技术为处理和分析庞大数据集提供了高效的工具和方法,从而加深了研究者对大数据在新闻传播中应用的理解。例如,机器学习算法,特别是在模式识别和预测建模方面的应用,极大地增强了新闻内容的定量分析能力<sup>[21]</sup>。通过对大规模新闻数据集的深入分析,机器学习不仅能够揭示新闻报道的趋势和模式,还能识别新闻事件的发展轨迹,并预测未来新闻的可能方向<sup>[22]</sup>。数据驱动的分析增强了新闻报道的深度和广度,为新闻编辑提供了强大的决策支持。此外,自然语言处理技术,在文本挖掘和情感分析方面的应用,为解析和理解新闻内容开辟了新视角<sup>[23-24]</sup>。这项技术使研究者能够从大量新闻文本中提取关键信息,如关键词、实体及情感倾向,深入探讨新闻语境及其潜在的社会影响。同时,该技术还能揭示报道中的偏见和立场<sup>[25-26]</sup>,为社会议题和公众舆论提供更为深刻的借鉴。在此基础上,学术界也在积极探讨大数据技术在新闻传播中的具体应用及其对行业的影响,包括新闻采集、编辑和分发流程的变革<sup>[27-28]</sup>。大数据技术的应用不仅使个性化新闻和定制化内容的制作成为可能,还引发了对新闻内容质量和信息泡沫问题的关注。

研究伦理层面,随着技术红利的享受越来越普遍,对大数据殖民主义和算法霸权的研究显得日益重要,研究者主要关注信息和技术之间的权力关系及其对社会和媒体的深远影响<sup>[29-31]</sup>。大数据殖民主义方面的研究探讨了大数据如何在全球范围内,特别是在全球南部国家,被用于收集、分析和掌握信息<sup>[32-33]</sup>。研究者们关注数据的收集过程、所有权与控制权如何构成

了新的全球信息权力结构。这一领域的研究强调了信息访问和控制的平等问题,揭示了发达国家与发展中国家在数据获取和利用方面的巨大差距<sup>[34]</sup>。算法霸权的研究则着重考察算法如何控制和影响信息流,包括社交媒体算法如何决定信息的筛选和用户体验,以及算法如何决定内容的可见性和广泛传播<sup>[35]</sup>。此类研究揭示了算法可能对信息环境造成的操纵和塑形作用,如何促成信息过滤、信息孤岛现象以及加剧社会偏见<sup>[36]</sup>。在伦理和权力问题方面,学者们探讨了大数据分析和算法应用中的伦理挑战,如隐私权保护、数据滥用、社交媒体平台的审查以及算法的广泛社会影响<sup>[37]</sup>。研究重点在于如何在新闻传播领域中找到数据分析潜力与伦理权利间的平衡点。此外,政策和监管方面的研究也变得至关重要。随着大数据和算法在新闻传播中的应用日益普遍,政府和监管机构必须制定相应的政策和法规以确保公共利益、信息自由和个人隐私的保护。学术界致力于分析和推动有效政策的制定,以应对技术发展带来的挑战,保障信息公平性和社会正义<sup>[38]</sup>。

在这十一年的时间里,大数据技术在新闻传播领域引起了革命性的变革,它不仅重新定义了新闻的采集、处理和分发流程,还深刻地影响了传播理论与实践的多个层面。这种技术的进步显著地扩展了新闻报道的速度和范围,同时也提高了新闻内容的个性化和目标化。然而,尽管取得了诸多进展,当前研究对大数据如何在结构和功能上综合影响新闻与传播的系统性和综合性理解仍然处于初级阶段。

因此,本文基于中国知网期刊数据库及 Web of Science 核心合集数据库,力争探索 2013-2023 年间在新闻传播领域中大数据技术的应用与影响。本综述不仅考察了文献的时间分布,也深入分析了高被引论文、不同国家与机构的贡献度、期刊分布、关键词的演变以及作者合作网络,以揭示大数据在新闻传播学研究中的核心趋势和研究热点。通过对已有研究成果的综合评述和分析,本文期望为新闻传播领域中学者和实践者围绕大数据分析技术提供理论、方法与伦理层面上更多的参考。

## 2 研究方法

本文分别在中国知网期刊数据库及 Web of Science 核心合集数据库进行文献搜索。基于中国知网期刊数据库,以“大数据”为关键词,检索时间范围设置为 2013-2023 年,来源类别包括北大核心、CSSCI、

CSCD 核心期刊,文献类型选择学术期刊,学科选择“新闻与传媒”,共检索出 3610 篇文献,经人工判断,最终将 3435 篇文献纳入研究范围;基于 Web of Science 核心合集数据库,以“big data”为关键词,检索时间范围设置为 2013-2023 年,语言设定为英语,文献类型为论文(Article),WoS Categories 选择“communication”,共检索出 1359 篇文献,经过人工判断,该 1359 篇文献均属于本文研究范畴。

数据处理方面,首先对数据进行人工清洗,清洗工作包括两个方面:一是对同一作者在不同文献中以不同形式出现的情况进行整合;二是通过查找文献发表官方信息对漏采或缺失信息进行补充。基于上述数据,利用中国知网内嵌的计量可视化分析软件对中国知网文献展开分析,利用 Thomson Data Analyzer 软件从时间分布、国家分布、机构分布、领域分布和期刊分布五个部分对 Web of Science 文献展开分析,同时利用网络分析方法对作者的合作网络以及关键词网络进行研究。

## 3 结果与分析

### 3.1 时间分布

2013-2023 年 Web of Science 核心合集数据库和中国知网期刊数据库发文量随时间的分布情况如图 1 所示。一方面,在中国知网期刊数据库中,文章发表数量于 2015 年前快速增长,2015-2018 年间新闻传播领域大数据研究热度保持在较高水平,其中 2016 年发文量最多,达到 472 篇。随着大数据技术在新闻与传播领域的持续运用,产生了一系列研究成果。2018 年后相关研究热度开始逐年下降。另一方面,在 Web of Science 核心合集数据库中,文章发表数量在 2013-2020 年间一直处于增长态势,自 2021 年起有所下降,并且一直持续到 2023 年。

通过对中国知网期刊数据库与 Web of Science 核心合集数据库近十一年的文献发表量进行比较,可以得出以下结论:首先,中国知网期刊数据库的发文量在这一时期内一直高于 Web of Science 核心合集数据库。尽管如此,国外相关研究在 2013-2020 年间持续稳步增长,于 2020 年达到峰值 216 篇,与同期中国知网期刊数据库的发文量相差无几。到了 2023 年,两者的发文量几乎持平。其次,近年来,国内外的发文量都呈现出明显的下降趋势。这一现象主要由以下几个因素引起。



一是随着大数据技术的进一步细分,相关研究领域也开始从宏观的概念性讨论转向更加细分和专业化的应用探索。研究的焦点逐渐下沉至具体的细分领域,例如数据新闻和媒体融合等。这种转变导致传统意义上关于大数据技术的研究数量减少,而具体应用的研究成果则分散在更广泛的学术领域中。

二是随着大数据相关技术的持续发展和进化,新闻传播学领域对大数据的应用越来越倾向于深度融合和跨学科研究。结合人工智能、机器学习等先进

技术进行内容生产、舆情分析和用户行为研究的复杂度显著增加。这种深化的学科融合过程要求新闻传播学领域的研究者具备更丰富的跨学科知识背景和技能。

三是随着大数据与新闻传播融合的初期热潮逐渐平息,研究面临创新研究点不足的问题,导致领域发展出现了一定程度的停滞。尤其在计算传播和计算广告等专业方向中,如何深度融合专业技能与大数据技术,突破现有的理论和实践限制,成为了亟待解决的问题。

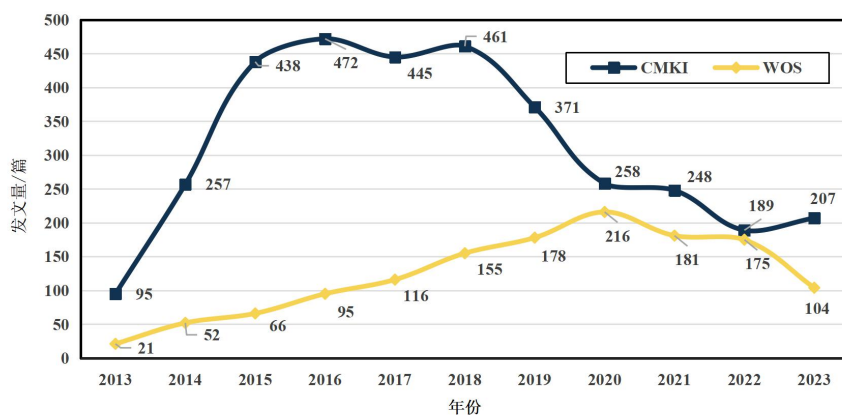


图1 2013-2023年Web of Science核心合集数据库和中国知网期刊数据库新闻传播学大数据领域文献发文量统计

2013-2023年Web of Science核心合集数据库的前十篇高被引文献相关信息如表1所示。其中,于2014年发表在期刊Journal of Communication上的一篇名为“Echo chamber or public sphere? predicting political orientation and measuring political homophily in Twitter using big data”的文章被引用次数最高,该文第一作者是Elanor Colleoni,被引次数为591。该文主要基于推特平台的媒体数据,使用大数据挖掘方法,结合机器学习和社交网络分析探究共和党和民主党的政治同质性问题<sup>[39]</sup>。综合来看,这些高被引文献共同突显了社交媒体数据在解析公众政治倾向、情感流动、社会事件和媒介叙事结构中的独特价值<sup>[40-41]</sup>,展示了大数据分析技术如何加深对诸如回音室效应、政治同质性、议程设定理论及新兴议题如数据殖民主义、算法社会影响力等传统与现代议题的理解。此外,这些文献探讨了大数据如何改变信息传播方式,影响政治参与形式,并通过新型监管模式如“超级劝导”介入公共生活<sup>[42]</sup>。在假新闻的议程设定能力分析中,也凸显了大数据时代对信息真实性的挑战及对传统媒介角色的重构需求。

通过评估数据新闻学和计算新闻学的兴起,这些研究不仅阐明了新闻传播领域的量化转向,也为理解新闻作为数据驱动叙事如何影响公众知识和政治偏好提供了新的视角。

2013-2023年中国知网期刊数据库的前十篇高被引文献相关信息如表2所示。其中被引用频次最高的文章是2013年发表于期刊“国际新闻界”的“全球视野下的‘数据新闻’理论与实践”,第一作者为方洁,被引用679次。该文通过对国内外关于数据新闻的阐释文献的梳理,结合主流媒体运用数据新闻的经典案例剖析,探索数据新闻理念的内涵与特征,分析数据新闻产生的背景及原因并呈现数据新闻实践的成果<sup>[50]</sup>。综合来看,所识别的十篇高被引文献主要集中在数据新闻的理念和实践,以及其如何形成传播学研究中的新范式。这些文献反映了学术界从全球视角与本土实践相结合的双重关注,突出了在大数据时代,理念创新与内容生产的关键性。此外,场景理论的探讨及其在移动传播和广告领域的应用<sup>[51-53]</sup>,揭示了新媒体环境下内容框架的演变及对现有困境的策略反思。值得一提的是,这

表1 2013—2023年Web of Science核心合集数据库新闻传播学大数据领域高被引文献(前十)

排名	题目	期刊	第一作者	年份	被引用次数
1	Echo chamber or public sphere? predicting political orientation and measuring political homophily in Twitter using big data <sup>[39]</sup>	Journal of Communication	Colleoni, Elanor	2014	591
2	Affective publics and structures of storytelling: sentiment, events and mediality <sup>[43]</sup>	Information Communication & Society	Papacharissi, Zizi	2014	453
3	Data colonialism: rethinking big data's relation to the contemporary subject <sup>[44]</sup>	Television & New Media	Couldry, Nick	2015	334
4	The dynamics of public attention: agenda-setting theory meets big data <sup>[45]</sup>	Journal of Communication	Neuman, W. Russell	2017	307
5	Imagined affordance: reconstructing a keyword for communication theory <sup>[41]</sup>	Social Media + Society	Nagy, Peter	2014	296
6	The social power of algorithms <sup>[46]</sup>	Information Communication & Society	Beer, David	2015	294
7	'Hyperjudge': big data as a mode of regulation by design <sup>[42]</sup>	Information Communication & Society	Yeung, Karen	2013	287
8	Every tweet counts? how sentiment analysis of social media can improve our knowledge of citizens' political preferences with an application to Italy and France <sup>[47]</sup>	New Media & Society	Ceron, Andrea	2013	283
9	Clarifying journalism's quantitative turn: a typology for evaluating data journalism, computational journalism, and computer-assisted reporting <sup>[48]</sup>	Digital Journalism	Coddington, Mark	2015	275
10	The agenda-setting power of fake news: a big data analysis of the online media landscape from 2014 to 2016 <sup>[49]</sup>	New Media & Society	Vargo, Chris J.	2020	253

些文献不仅审视了技术进步如何推动新闻生产方式的变革,还关注了大数据如何支持网络意识形态的治理和提升社会治理能力<sup>[54-55]</sup>。通过这些高被引文献的评估,不仅能够洞察数据新闻学的发展轨迹,也为理解新闻作为数据驱动的叙事如何影响社会治理和公众知识提供了宝贵的视角。

相比较而言,国外文献在大数据研究中通常关注其在全球政治和社会结构中的作用,强调大数据对公

共领域和政治动态的影响,以及在处理假新闻和公共情绪等方面的应用。这种关注体现了国际学术界对大数据社会影响力的深度挖掘,及其对技术伦理和政治参与新模式的持续探讨。相对而言,国内文献则更专注于大数据技术在新闻传播实践中的具体应用,包括数据新闻的创新、媒体融合过程及其对新闻行业的直接影响。这种研究重点揭示了我国学术界对技术如何改变新闻生产流程和内容消费方式的浓厚兴趣,

表2 2013—2023年中国知网期刊数据库新闻传播学大数据领域高被引文献(前十)

排名	题目	期刊	第一作者	年份	被引用次数
1	全球视野下的“数据新闻”:理念与实践 <sup>[50]</sup>	国际新闻界	方洁	2013	679
2	场景理论的内容框架与困境对策 <sup>[51]</sup>	当代传播	郜书锴	2015	380
3	镶嵌、创意、内容:移动互联广告的三个关键词——以原生广告的操作路线为例 <sup>[52]</sup>	新闻与写作	喻国明	2014	347
4	大数据时代的数据新闻报道——以英国《卫报》为例 <sup>[56]</sup>	现代传播(中国传媒大学学报)	文卫华	2013	269
5	传播学研究:大数据时代的新范式 <sup>[57]</sup>	管理评论	喻国明	2013	238
6	大数据环境下多媒体网络舆情传播要素及运行机理研究 <sup>[54]</sup>	图书情报工作	黄微	2016	236
7	场景理论:开启移动传播的新思维 <sup>[53]</sup>	新闻界	郜书锴	2015	234
8	大数据助力网络意识形态治理及提升路径 <sup>[55]</sup>	马克思主义研究	付安玲	2016	233
9	大数据时代的数据新闻生产:现状、影响与反思 <sup>[58]</sup>	现代传播(中国传媒大学学报)	刘义昆	2014	196
10	大数据与新闻理念创新——以全球首届“数据新闻奖”为例 <sup>[59]</sup>	编辑之友	王斌	2013	187

并体现了对增强媒体监管能力和优化舆情引导的系统探索。从理论建构的角度看,国外文献倾向于通过案例分析和大数据分析等方法,探索大数据与现有社会理论的关联,寻求理论框架的扩展。相比之下,国内研究更多从实际应用的角度出发,讨论大数据如何服务于新闻传播的具体场景,并关注技术应用对解决现实问题的具体贡献,同时对传统传播理论在大数据环境下的适应性和创新进行深入思考。

### 3.2 国家与机构分布

对 Web of Science 核心合集数据库文献的国家与机构进行分析。2013-2023 年发文量国家排名情况如表 3 所示,前十位依次为:美国、西班牙、英国、德国、澳大利亚、荷兰、加拿大、中国、丹麦和比利时。其中,美国发文量 425 篇,远高于其余各国,占比 32.79%;西班牙发文量 156 篇,占比 12.04%;英国发文量 138 篇,占比 10.65%。美西英三国发文占比总和超 50%。中国位居第 8 位,发文量 58 篇,占比 4.48%。

表 3 2013-2023 年 Web of Science 核心合集数据库新闻传播学大数据领域文献发文量国家排名(前十)

排名	国家	发文量(篇)	占比
1	美国	425	32.79%
2	西班牙	156	12.04%
3	英国	138	10.65%
4	德国	85	6.56%
5	澳大利亚	78	6.02%
6	荷兰	77	5.94%
7	加拿大	67	5.17%
8	中国	58	4.48%
9	丹麦	33	2.55%
10	比利时	27	2.08%

资料来源:数据来自 Web of Science 核心合集数据库。

注:表 3 中存在一篇文献具有多个国家的情况。

就机构方面来说,2013-2023 年新闻传播学大数据研究发文量排名前十均为国外研究机构(表 4),其中伦敦大学和阿姆斯特丹大学分别名列第一与第二,发文量占比明显高于其他机构。其余机构发文量占比均小于 2%。

表 4 2013-2023 年 Web of Science 核心合集数据库新闻传播学大数据领域文献发文量机构排名(前十)

排名	机构名称	发文量(篇)	占比
1	UNIVERSITY OF LONDON	41	5.69%
2	UNIVERSITY OF AMSTERDAM	40	3.086%
3	UNIVERSITY OF CALIFORNIA SYSTEM	24	1.852%
4	UNIVERSITY OF ILLINOIS SYSTEM	24	1.852%
5	LONDON SCHOOL ECONOMICS POLITICAL SCIENCE	20	1.543%
6	UNIVERSITY OF TEXAS SYSTEM	20	1.543%
7	CARDIFF UNIVERSITY	19	1.466%
8	COMPLUTENSE UNIVERSITY OF MADRID	19	1.466%
9	UNIVERSITY OF OXFORD	19	1.466%
10	UNIVERSITY OF WISCONSIN SYSTEM	19	1.466%

资料来源:数据来自 Web of Science 核心合集数据库。

注:表 4 中存在一篇文献具有多个机构的情况。

### 3.3 期刊分布

发文量排名前十的期刊如表 5 所示。在 Web of Science 核心合集数据库中,新闻传播学领域大数据研究相关文献主要发表在人文科学领域下属传播学和社会学类期刊,包括 New Media & Society (发文量 79 篇,占比 6.10%)、Information Communication & Society (发文量 74 篇,占比 5.71%)、International Journal of Communication (发文量 753 篇,占比 4.09%)、Digital Journalism (发文量 46 篇,占比 3.55%) 等。Professional De La Informacion (发文量 54 篇,占比 4.17%)

和 Telecommunications Policy (发文量 27 篇,占比 2.08%) 等管理学类期刊也是新闻传播大数据领域文章发表的热门期刊。

在中国知网期刊数据库中排名前十的期刊发文量总体高于 Web of Science 核心合集数据库中排名前十期刊的发文量。其中,新闻研究导刊发表的文献量高达 272 篇,占比 8.12%,发文量超过 100 篇的期刊共 5 个,还包括西部广播电视(发文量 171 篇,占比 5.11%)、科技传播(发文量 110 篇,占比 3.28%)、传播力研究(发文量 108 篇,占比 3.22%) 和中国传媒科技(发文量 106 篇,占比 3.17%)。



表5 2013-2023年Web of Science核心合集数据库和中国知网期刊数据库  
新闻传播学大数据领域文献发文量期刊排名(前十)

排名	Web of Science 核心合集数据库			中国知网期刊数据库		
	期刊名称	发文量(篇)	占比	期刊名称	发文量(篇)	占比
1	New Media & Society	79	6.10%	新闻研究导刊	272	8.12%
2	Information Communication & Society	74	5.71%	西部广播电视	171	5.11%
3	Profesional De La Informacion	54	4.17%	科技传播	110	3.28%
4	International Journal of Communication	53	4.09%	传播力研究	108	3.22%
5	Digital Journalism	46	3.55%	中国传媒科技	106	3.17%
6	Social Media+Society	43	3.318%	新闻传播	86	2.57%
7	Policy and Internet	33	2.55%	中国报业	71	2.12%
8	Internet Policy Review	30	2.32%	新媒体研究	64	1.91%
9	Telecommunications Policy	27	2.08%	新闻与写作	62	1.85%
10	Convergence the International Journal of Research into New Media Technologies	23	1.78%	新闻战线	56	1.67%

### 3.4 关键词分布

关键词方面,本文分别收集整理了Web of Science核心合集数据库与中国知网期刊数据库2013-2023年新闻传播学大数据领域相关文献的关键词数据,绘制关键词关系网络。同时,为进一步从时间维度探索关键词词频及关系变化情况,本文将2013-2023年分成五个时间区间,即2013-2014年、2015-2016年、2017-2018年、2019-2020年、2021-2023年,根据每个区间的文献关键词数据分别绘制各时期关键词语义网络,所有语义网络均为无向加权网络,权重代表两个关键词之间的共线次数。

结果如图2、图3所示,其中网络节点大小代表关键词词频大小,边代表关键词间存在联系,越粗表示两关键词出现在同一篇文章中的频率越高。同时,需要说明的是,由于本文所采用的CNKI数据集是按照以“大数据”为关键词进行搜索而得到的,导致关键词“大数据”在整个语义网络中的影响过大,因此,本文后续在进行语义网络绘制以及重要关键词提取时将“大数据”删除,进而更好地综合性展现分析结果。

在Web of Science核心合集数据库中,新闻传播领域大数据研究文献主要关键词有“social media”、“artificial intelligence”、“twitter”、“algorithms”等,“social media”、“privacy”、“surveillance”、“text mining”等关键词与“big data”联系紧密,表明随着人工智能算法和数据挖掘方法的发展,社交媒体研究不仅拓宽了分析范围和深度,同时对隐私和监控的政策讨论也变得尤为重要。从Web of Science核心合集数据库中关键词关系子网络随时期的变化可以看出,关键词关系分布从分散逐渐趋向于紧密,节点由大小均等逐渐转变为个别节点权重突出,如

“social media”、“algorithms”、“artificial intelligence”和“twitter”,并形成子团体中心。其中,“cov-19”从2019-2020年开始出现并在2021-2023年迅速成为主要关键词,这不仅反映了全球性卫生事件对新闻传播领域的影响,也突显了大数据在公共健康危机响应和信息传播中的关键作用<sup>[60-62]</sup>。同时,该发现也预示着未来研究将更加聚焦于如何利用先进技术应对社会挑战,以及这些技术在实现信息自由流通与保护个人隐私之间必须平衡的复杂议题。

在中国知网期刊数据库中,新闻传播领域大数据研究文献最主要关键词有“新媒体”、“人工智能”、“数据新闻”、“媒体融合”,且均位于网络中心,与其他众多关键词如“网络舆情”、“社交媒体”、“可视化”、“云计算”等联系紧密。该发现揭示了技术进步对行业的深远影响,同时也指出了在这一过程中需要解决的关键问题,包括技能培养、伦理隐私保护、社会影响力管理以及新闻呈现方式的创新。同时,从5个时间区间的关键词分布可以看出,“新媒体”、“数据新闻”在近十一年中均为高频关键词,说明学界对新闻媒体数字化的关注热度居高不下,也反映了公众对于透明度、准确性和参与度的期待上升<sup>[63-64]</sup>;“媒体融合”的词频在近十一年间大体呈现增长趋势,暗示了在数字时代背景下,跨平台内容生产和多媒介互动成为新的常态<sup>[65]</sup>,这不仅改变了新闻传播的路径,也为理解多元化媒介生态提供了新的视角;“人工智能”关键词在近6年中尤其突出,标志着人工智能在内容创作、分发优化、受众分析等方面的应用日益成熟<sup>[66]</sup>,同时也对新闻行业的从业者提出了新的技能要求,挑战了传统的新闻生产逻辑和伦理标准<sup>[67]</sup>。

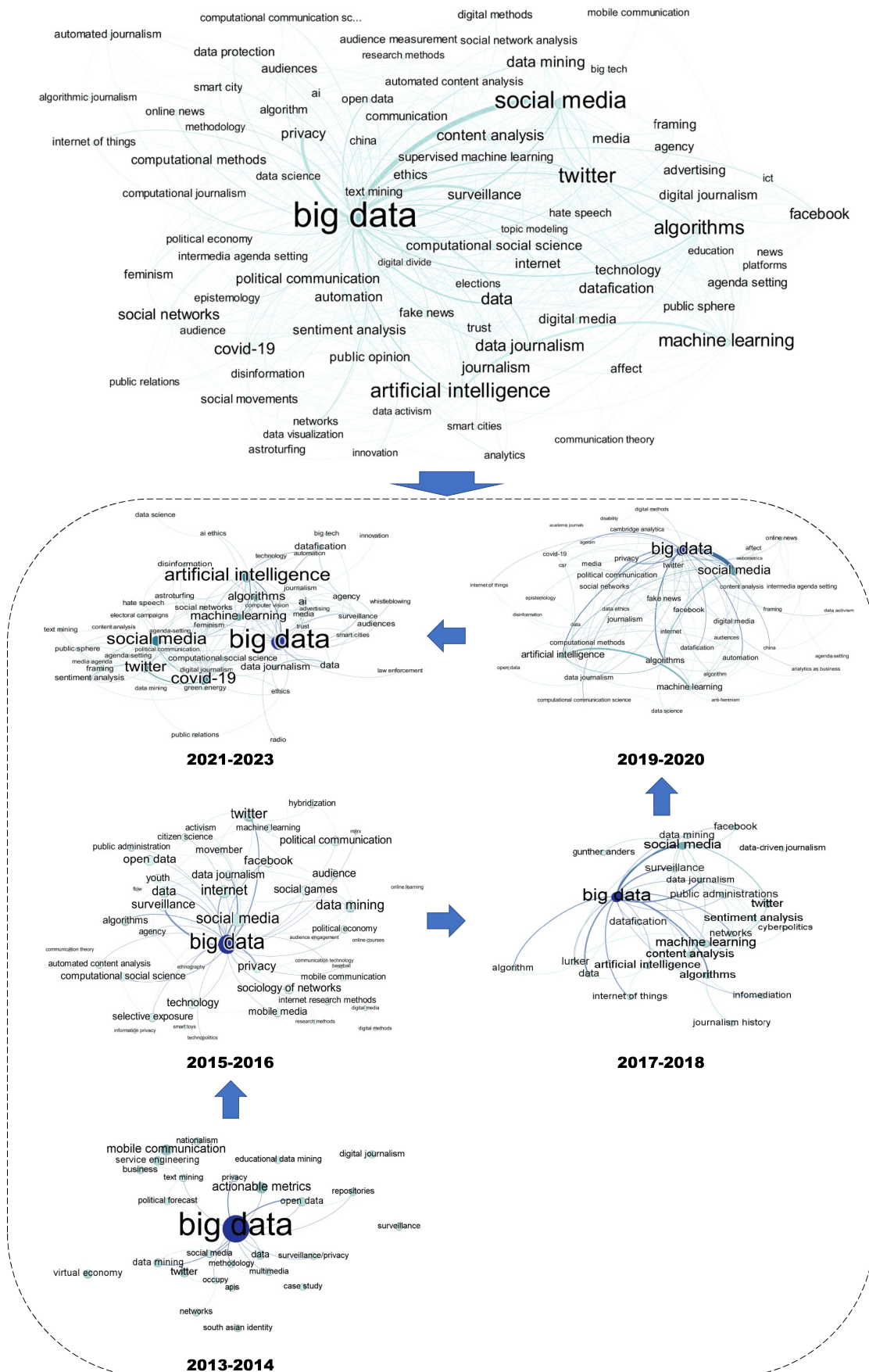


图2 2013-2023年 Web of Science 核心合集数据库新闻传播学大数据领域文献关键词关系网络



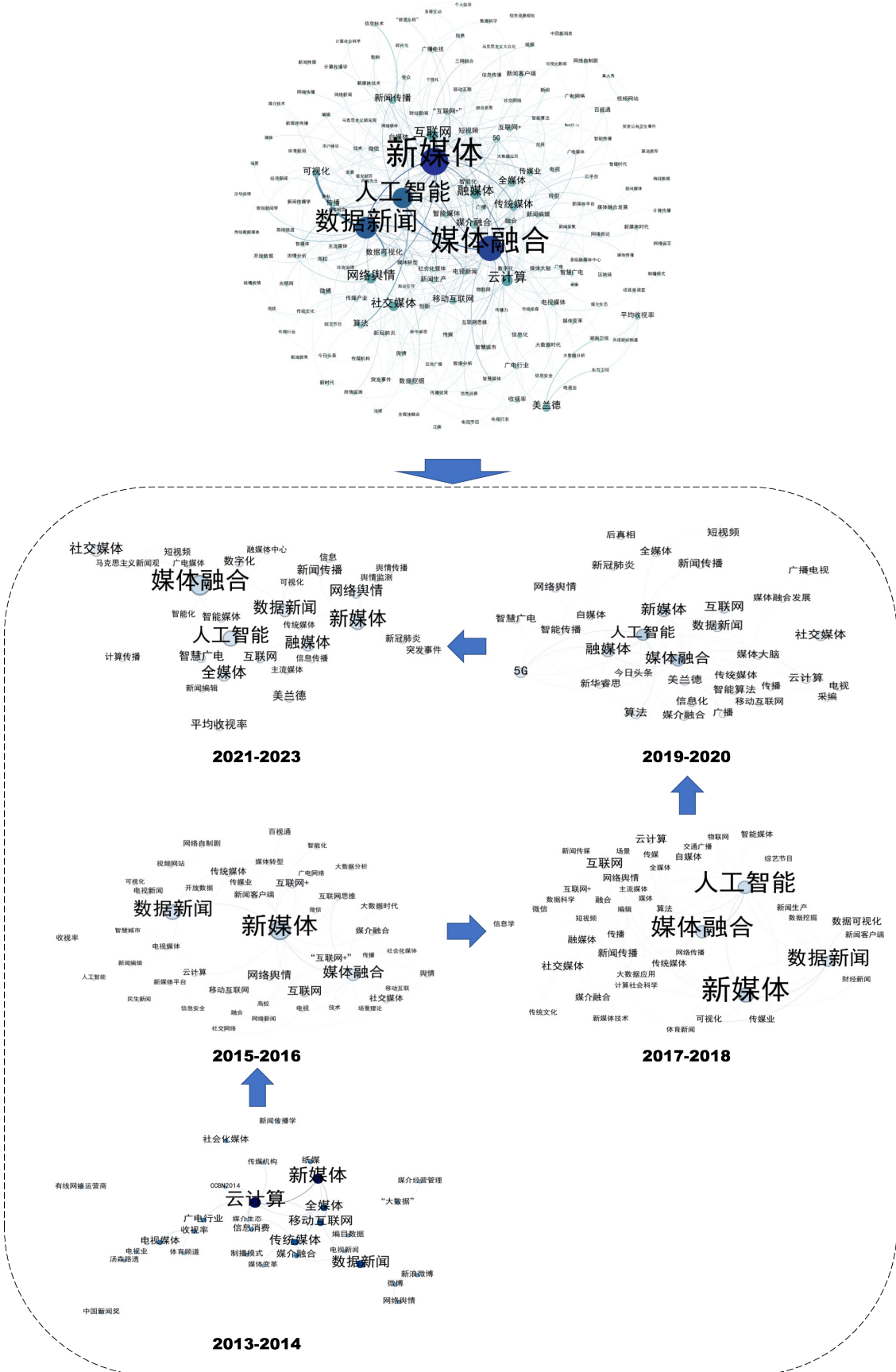


图3 2013—2023年中国知网期刊数据库新闻传播学大数据领域文献关键词关系网络

本文计算了各网络的图密度和平均聚类系数以进一步分析数据库中关键词之间的关系,如表6和表7所示。根据结果可以得知,在中国知网期刊数据库中,关键词网络的图密度一直处于较低的水平,并且自2015年起几乎没有变化,而平均聚类系数在2014年后大幅上升,随后缓慢下降。这说明新闻传播大数据研究领域的研究方向虽一直相对分散,但在2014年前后,各方向之间的聚集程度明显开始趋于紧密。在Web of Science核心合集数据库中,关键词网络的图密度呈现下降趋势,平均聚类系数在2016年前缓慢上升,随后下降,说明该领域研究方向越来越分散,各方向间的聚集程度先趋于紧密,在2016年后逐年趋于分散。

表6 2013-2023年中国知网期刊数据库新闻传播学大数据领域文献关键词网络图指标

时间区间	图密度	平均聚类系数
2013-2014年	0.001	0.032
2015-2016年	0.002	0.423
2017-2018年	0.002	0.422
2019-2020年	0.002	0.395
2021-2023年	0.002	0.323

表7 2013-2023年Web of Science核心合集数据库新闻传播学大数据领域文献关键词网络图指标

时间区间	图密度	平均聚类系数
2013-2014年	0.008	0.074
2015-2016年	0.004	0.156
2017-2018年	0.003	0.105
2019-2020年	0.002	0.092
2021-2023年	0.002	0.071

此外,本文根据关键词网络图分别统计各关键词节点的加权重度与中介中心度数值,得到降序排列。中国知网期刊数据库与Web of Science核心合集数据库各时间区间的关键词网络中,指标排名前十的关键词节点排序如表8和表9所示。其中,加权重度表示对网络中节点重要性的度量,加权重度数值越大,说明其邻居节点越多,影响力相对更大;中介中心度则以经过该节点的最短路径数目来刻画节点的重要性,节点的中介中心度越大,网络中经过该节点的最短路径数占最短路径总数之比越大,该节点的中介调节效应越大。

就中国知网期刊数据库而言,2014年前后,“云计算”作为一个技术热点,其在关键词网络中的加权重度与中介中心度均居高不下。随后,“数据新闻”迅速崛起,替代“云计算”成为研究热潮,并连续四年保持加

权重最高。到了2017年,“媒体融合”逐渐显现其不可忽视的重要性,并在接下来的四年中成为加权重度最高的关键词。在2017至2020年间,“媒体融合”与“融媒体”分别成为中介中心度最高的关键词,反映出传统媒体与新兴媒体融合发展成为了新闻传播大数据研究的核心议题。值得关注的是,近年来新出现的关键词“美兰德”,作为数据发布源,在2021-2023年具有最高的中介中心度,并在加权重度中排名第八。这一现象表证着特定数据来源在新闻传播学研究与实践中的日益重要性,突显了数据资源在新闻生产和传播过程中的核心作用。从“云计算”的技术热潮到“数据新闻”的兴起,再到“媒体融合”及“融媒体”的突出,这一变迁不仅映射了技术发展如何重塑新闻传播实践和研究焦点,还反映了新闻传播领域对技术创新与媒介变革适应性的深度探索。特别是“媒体融合”和“融媒体”的关键词成为中介中心度最高的标识,凸显了探索传统媒体与新兴媒体融合的路径和模式成为当前及未来重要的研究方向<sup>[68]</sup>。这种融合不仅关涉内容生产和传播方式的革新,也涉及媒介组织结构和运营机制的创新。未来新闻传播研究将进一步关注数据资源的开发与应用,探索数据驱动的新闻生产模式,并通过数据深化新闻内容的个性化和精准化,从而推动新闻传播学科的理论创新和实践变革。

而Web of Science核心合集数据库的关键词网络中,数据源与数据处理方法的关键词占据较大比例。除了检索关键词“big data”外,“twitter”和“social media”在近十一年中一直是该领域研究的焦点,显示国际学术界在新闻传播的大数据研究中倾向于关注社交媒体,尤其是Twitter等平台提供的数据。近两年来,“artificial intelligence”已成为关键词网络中加权重度与中介中心度均最高的关键词,这标志着人工智能领域在新闻与传播大数据研究中的地位越发重要。“twitter”和“social media”这两个关键词的持续重要性揭示了国际学术界在探索大数据对新闻传播影响时,高度依赖社交媒体平台作为主要的数据来源和研究对象<sup>[69-70]</sup>。这种依赖反映出社交媒体在当代社会中传播信息<sup>[71]</sup>、塑造公众舆论<sup>[72]</sup>、影响政治动态<sup>[73]</sup>的核心作用,同时也突显了研究者在从海量社交媒体数据中提取有价值信息的技术需求和面临的挑战。随着“artificial intelligence”成为近两年的关键研究焦点,人工智能的方法论和技术在处理、分析及理解大规模数据集方面的重要性日益增强。这不仅预示着人工智能技术将在未来新闻传播研究中扮演关键角色,提高研





究的效率和深度,还暗示未来研究将更深入探讨人工智能如何影响新闻的生产、分发和消费过程,以及这些技术如何在确保数据隐私、保障信息真实性等方面提供新的解决策略。因此,未来的新闻传播大数据研究将更加集中于利用先进的人工智能技术来解析社交媒体的动态,以及评估这些技术在新闻传播领域应用的伦理和社会影响。

### 3.5 作者合作网络分析

为进一步对合作关系进行分析,本文可视化了该领域2013-2023年Web of Science核心合集数据库和中国知网期刊数据库作者合作无向网络,如图4所示。同时,本文统计了加权重度前十的节点(作者)及所属的国家和单位,如表10、表11所示。其中,在Web of Sci-

ence核心合集数据库中,来自复旦大学的中国学者L Guo表现出最大的合作度,其加权重度达到17。另外,合作程度最高的十名作者中所属国家的分布较为分散,有3名来自美国,2名来自英国,2名来自中国,其余3名分别来自墨西哥、荷兰和西班牙。

在中国知网期刊数据库中,近十一年来研究者之间合作关系比较稀疏(图5),只呈现出众多规模较小的子团体,并且加权重度高的节点之间大多没有边,表明国内新闻传播大数据研究领域研究者联系较为疏远。在合作网络中,合作程度最突出的作者有兰月新、喻国明、夏一雪、黄微和许志强,其中,兰月新与夏一雪属于同一个子团体,均来自中国人民警察大学,其余三人分别来自北京师范大学、吉林大学和浙江

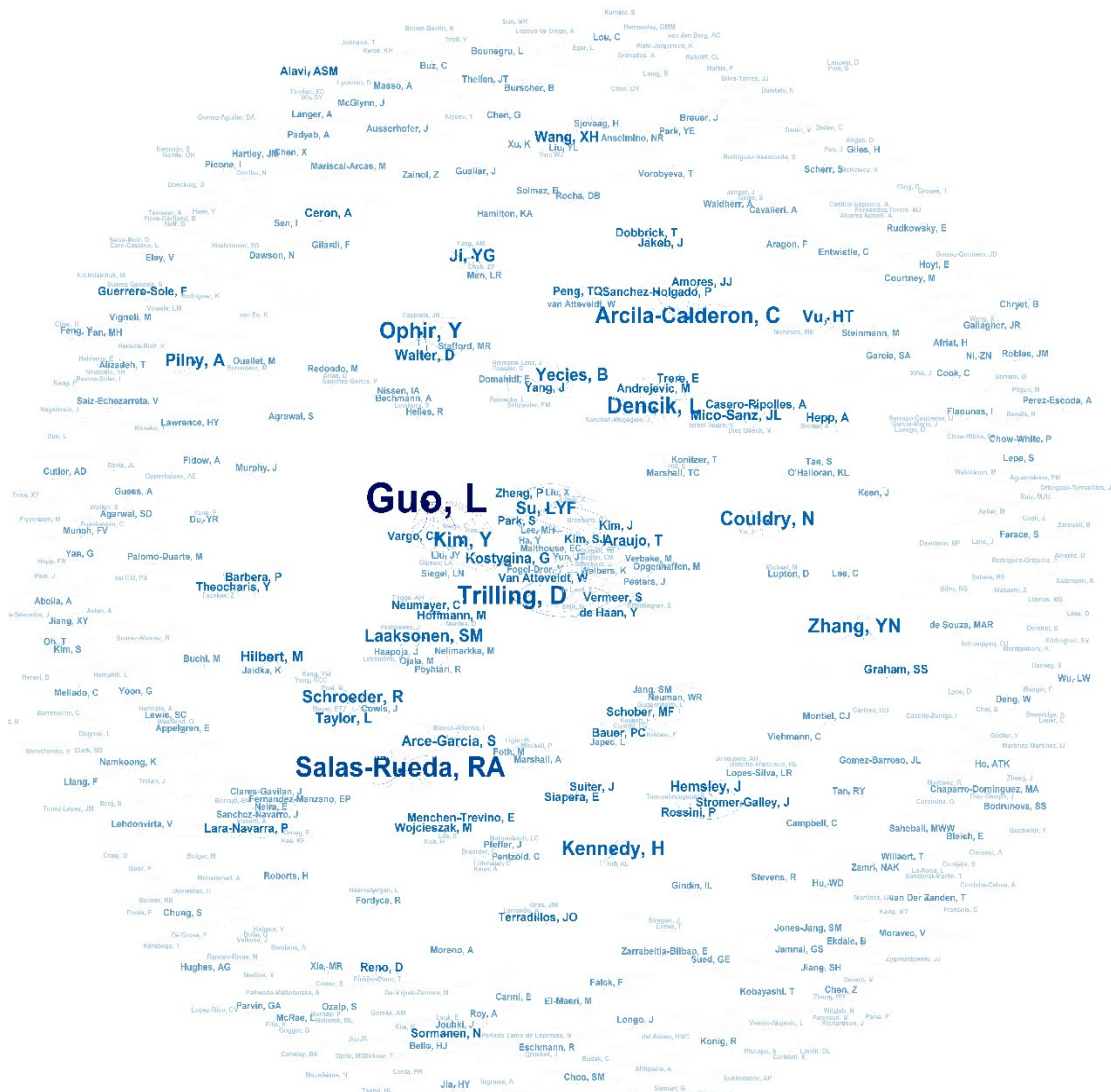


图4 Web of Science核心合集数据库新闻传播学大数据领域文献作者合作网络图

表 10 2013–2023年 Web of Science 核心合集数据库作者合作活跃度情况排名(前十)

排名	作者	加权重	所属国家	所属单位(最新)
1	Guo, L	17	中国	复旦大学
2	Salas-Rueda, RA	14	墨西哥	墨西哥国立自治大学
3	Trilling, D	10	荷兰	阿姆斯特丹大学
4	Arcila-Calderon, C	9	西班牙	萨拉曼卡大学
5	Dencik, L	9	英国	卡迪夫大学
6	Ophir, Y	8	美国	纽约州立大学布法罗分校
7	Couldry, N	8	美国	纽约州立大学奥斯威戈分校
8	Zhang, YN	8	美国	纽约州立大学布法罗分校
9	Kennedy, H	8	英国	利兹大学
10	Kim, Y	7	中国	美国中康涅狄格州立大学

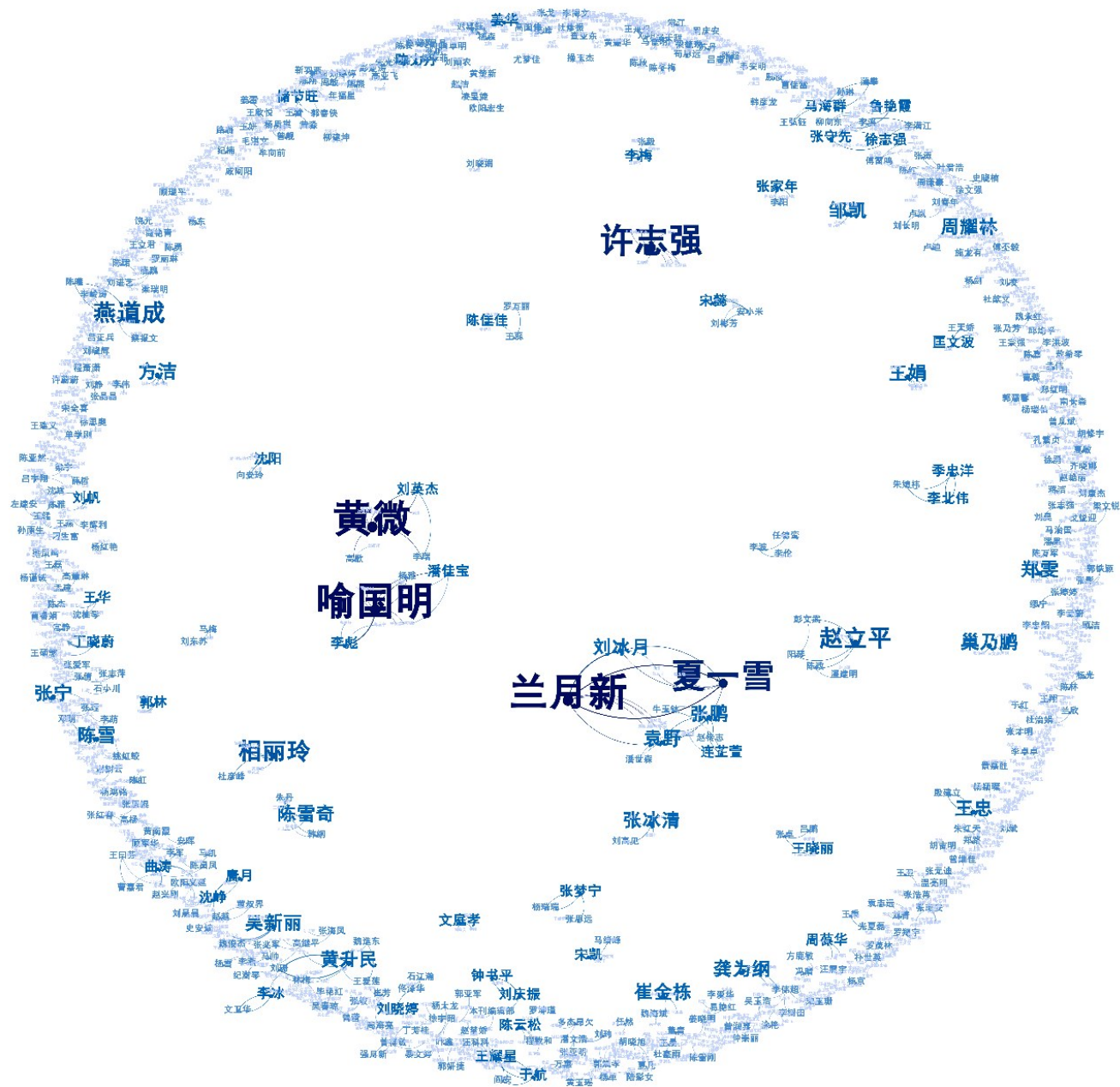


图 5 中国知网期刊数据库新闻传播学大数据领域文献作者合作网络图

表 11 2013-2023年中国知网期刊数据库作者合作活跃度情况排名(前十)

排名	作者	加权重	所属单位(最新)
1	兰月新	11	中国人民警察大学
2	喻国明	9	北京师范大学
3	夏一雪	9	中国人民警察大学
4	黄薇	8	吉林大学
5	许志强	7	浙江大学
6	黄升民	6	中国传媒大学
7	丁晓蔚	6	南京大学
8	吴新丽	6	新华网
9	方洁	5	中国人民大学
10	燕道成	5	湖南师范大学

#### 4 现状评析

在当前的新闻传播大数据研究领域中,国际与国内的学术探索各展特色,均面临着新技术革命和媒介生态变迁的复杂挑战。国际学术界对社交媒体如 Twitter 的持续关注,不仅反映了公共讨论空间的数字化转移,也突显了社交媒体数据在捕捉社会脉动、塑造公众议程中的关键作用。这种关注持续升温,显示出全球学者在理解数字时代下公民行为、政治参与以及文化流变方面的不懈追求。随着技术的演进和媒介形式的多样化,新兴平台如短视频和播客逐渐成为学术研究的新焦点,尽管这些平台在数据获取与分析上提出了更大的技术挑战。这一趋势不仅预示着对新兴媒介特性和影响力的深入探索,也迫切要求学术界发展与之相适应的先进技术和方法,例如,采用计算机领域更加前沿的计算模型和算法来处理非文本型数据。此外,传统的将社交媒体作为一整体进行研究的视角正逐步让位于对不同平台特性的精细化研究。这种细分研究方法能更准确地把握各平台在用户构成、使用习惯等方面的差异,从而提升研究的精确度和实用价值。综合来看,新闻传播学领域的研究者正逐渐适应这一媒介环境的变化,通过创新研究方法和技术手段,深化对大数据如何重塑信息传播方式、影响公众舆论以及促进社会动态理解的研究。

与此同时,我国新闻传播大数据研究的轨迹体现了与国际研究不同的独特特色。从对“云计算”技术的早期探索,到对“数据新闻”创新实践的深入关注,再到“媒体融合”现象的系统研究,中国新闻传播学术界展现了对技术革新和媒介融合趋势的紧密关注,更偏向于大数据技术对新闻传播实践的影响研究。特别是在“媒体融合”的研究领域,我国学者不仅关注数字技术如何推动

传统媒介与新兴媒介界限的模糊化,还深入探讨了如何在这一过程中构建更加互动、多元和高效的新闻传播生态系统。这些研究揭示了媒体融合背后的动力机制,探索了融合过程中的策略和实践,及其对新闻行业结构和运作方式的影响。此外,随着“美兰德”数据的引入,大数据技术在内容分析、受众研究等方面的应用正日趋深入。这一趋势不仅为我国新闻传播大数据研究提供了新的视角和工具,也标志着研究方法的进一步创新和优化。通过这些数据资源,研究者能够更精确地分析受众行为,更细致地解读媒体内容与公众互动的复杂关系,从而更好地理解 and 预测媒介传播效果。

关键词网络分析揭示了国际与国内新闻传播大数据研究的发展方向 and 焦点之间存在显著差异。国际研究领域的扩散趋势表明了研究主题和方法的广泛多元化,反映出全球学者在探索大数据对新闻传播影响的过程中,采取了包容各种理论视角和方法论的策略。例如,国际学术界趋向于从跨文化的视角分析数据新闻的全球流动性,以及社交媒体数据在全球政治事件中的角色,展示了一种对新兴技术和全球化媒介环境相适应的研究开放性。相比之下,国内研究的聚集趋势显示了在特定领域如媒体融合、数据新闻的制作与应用等方面的深入探索和整合。中国学者通常聚焦于如何在国内的特定社会环境下,有效利用大数据技术来优化新闻生产流程和提高新闻内容的质量与受众针对性。这种研究趋势揭示了对本土化应用和技术整合的强烈关注,同时也体现了在传统媒体和数字平台融合过程中的策略性思考。这些差异不仅提供了对当前新闻传播与大数据研究状态的深刻洞见,也为未来的研究方向提供了重要的指引。在技术与媒介形式不断演变的背景下,全球学术界需要更有效地利用大数据技术来深化对社会传播机制的理解,探索数据如何在不同文化和政治背景下影响新闻的生成和消费。同时,研究应致力于在保障数据伦理和隐私安全的前提下,推动新闻传播实践的创新与发展。这包括开发新的数据分析工具,以及制定相应的政策和法规,以确保技术应用的透明度和公正性,从而在全球范围内促进新闻传播学科的理论创新和实践变革。

#### 5 结论与展望

本文基于 Web of Science 核心合集数据库和中国知网期刊数据库,采用文献计量方法,对 2013-2023 年新闻传播学大数据领域相关文献进行了系统梳理,从时间、国家、机构、期刊、研究内容以及作者等方面展开分析,并采用网络分析方法对作者合作网络和关键词关联网



络进行研究。结果显示,新闻传播学作为一个不断进化的领域,在数据驱动的时代背景下,其理论、研究对象、方法论以及伦理都将面临新的挑战 and 机遇。适应这些变化,不仅要求学者们具备跨学科的视野和创新能力,也需要对现有的理论和方法进行深入的反思和更新。因此,本文提出新闻传播领域可能的主流研究方向:

1)理论层面:须构建新的理论框架,以全面理解这些新兴媒介所引发的变革,并通过案例研究和实证研究来验证和深化这些理论。这种方法将使新闻传播学更好地适应数字时代的挑战,提升所提供的新闻内容的质量和社会影响力。

未来新闻传播学的研究应深入关注媒体融合、数据新闻,以及社交媒体与公共舆论之间的动态交互。需要系统探讨数字化媒体如何重塑传统新闻的叙述结构和流程。随着数字技术的演进,新闻组织已转型为多元互动平台,包括用户生成内容、跨平台故事叙述及实时更新的新闻报道。这种媒体融合不仅革新了新闻生产方式,也重新定义了观众参与新闻制作的模式。例如,研究可以探索用户如何通过社交媒体平台对新闻事件进行评论和分享,以及这些互动如何反过来影响新闻机构的报道策略和内容布局。

数据新闻的崛起引领了结合数据分析、可视化和叙述的理论创新研究体系。该体系的应用要求研究者不仅精通数据处理技能,更需具备将复杂数据信息转化为公众易于理解的故事的能力。从理论层面而言,研究应深入探讨如何利用大数据分析技术揭示公众对特定事件的看法和情绪变化,进而将这些数据结论转化为具有深度和洞见的新闻内容。同时,理论研究应进一步探索数据新闻如何重塑新闻叙事的框架和方式。例如,未来相关理论探讨可以涉及数据新闻如何帮助新闻制作者创新报道形式,通过数据驱动的深度报道来揭示和解释复杂事件的多层面因素。同时,这种理论探索应包括评估数据新闻对提升新闻报道事实基础的效果,以及它如何改进报道的透明度和公众的理解度。此外,理论研究还应关注数据新闻在新闻生产流程中的具体应用,包括数据采集、处理、分析及其在新闻叙述中的整合方式。研究应考察这些技术如何影响新闻的内容质量和受众接受度,以及它们如何帮助新闻机构在竞争激烈的媒体环境中保持其新闻的时效性和相关性。

对微博、Twitter等社交媒体在塑造公共舆论中的作用的研究成为一个不断扩展的重要领域。这些平台不仅仅作为信息传播的渠道,更成为公众表达意见、进行讨论和影响议程设置的核心舞台。因此,未来的研究需

要深入分析这些社交媒体平台如何塑造信息传播的路径,并探索这些路径如何影响公共舆论和社会议程。具体而言,理论研究应探讨社交媒体算法如何管理和过滤信息,影响信息的可见性和隐匿性,以及这些算法如何塑造或扭曲公众对事件的认知和响应。这包括分析算法如何基于用户行为和偏好来推荐内容,以及这种内容定制如何在无形中构建和巩固特定的舆论氛围和政治立场。研究者应评估这些技术介入对民主参与和公共讨论的长远影响,特别是在关键政治和社会事件中的作用。此外,理论研究还需关注社交媒体如何成为政治行动和社会动员的工具。例如,研究可以探索在线运动如何通过社交媒体发起和扩散,以及这些运动如何通过数字联结实现社会变革。这涉及到理解社交媒体平台在全球政治抗议、公民权利运动以及环境倡议中的角色。

2)方法层面:随着技术的进步,传统的数据挖掘、文本分析算法正被大型语言模型的应用所补充和替代。因此,大型语言模型的广泛应用应成为未来新闻传播学研究中方法层面的重要创新载体,其不仅限于高效、多维度的数据处理,更应扩展到信息筛选和个性化推荐,以及通过仿真模型进行场景模拟等方面。

利用大型语言模型快速分析和处理包括文本、图像及视频在内的大规模多模态数据集。尽管已有研究尝试通过人工编码的方式处理图像等非文本数据,但这种方法在处理大规模数据集时往往显得力不从心,难以满足当前新闻传播研究的需求。相比之下,大型语言模型的应用不仅极大提升了处理速度和效率,还能够深入挖掘和整合来自不同数据源的信息,实现对新闻内容的全面和深入分析<sup>[74]</sup>。这种能力显著增强了新闻内容分析的广度和深度,使研究者能够从多角度、多层面新闻事件进行综合考察。例如,在使用大型语言模型分析政治新闻事件时,研究者不仅可以处理来自各大新闻网站的文本报道,还能同时分析社交媒体上的图像和视频评论,以及电视新闻的视频片段。这样的多模态分析使得研究更为全面,有助于揭示新闻事件的多维度影响和公众反应的全貌。此外,大型语言模型在处理和解析这些多模态数据时,还能应用先进的机器学习技术如深度学习,自动识别和学习数据中的模式和趋势,而无需依赖繁琐的人工编码。这种技术进步不仅提高了研究的效率,也大大提升了分析的准确性和可靠性,为新闻传播学研究提供了新的视角和深度。

采用大模型 Agent 进行仿真的方法为新闻传播学领域带来了革命性的研究工具。该方法在虚拟环境中模拟新闻传播过程及公众反应,为研究者提供了一个无风

险且高度可控的实验平台。这种仿真技术允许研究者以一种高度系统化的方式预测和分析不同新闻报道策略的效果,深入评估各种传播手段如何影响公众意见形成和舆论动态。通过这种方法,研究者能够在完全控制的环境中调整和测试新闻传播的各种变量,如信息的呈现方式、使用的媒介渠道、消息的时序和频率等。这不仅极大提高了实验的灵活性和扩展性,还保证了研究的安全性,避免了在现实世界中进行社会实验可能引发的伦理和法律问题。此外,大模型 Agent 的使用使得理论假设的验证过程更为准确和高效。仿真实验可以提供详细的数据支持,帮助研究者精确测量和解析新闻传播策略的具体效果,从而在理论建构和实证研究之间建立更为坚实的桥梁。这种高度数字化的方法论途径不仅加快了研究的步伐,也推动了新闻传播研究方法的现代化,为未来探索新闻传播的复杂动态提供了强大的技术支持。

大型语言模型在优化信息筛选和个性化新闻推荐方面展示了显著的应用潜力。通过分析用户的互动数据,这些模型能够推荐与用户兴趣更贴近的新闻内容,有效提升用户体验和参与度。这种技术的应用不仅增强了新闻机构在数字时代的竞争力,也极大地提升了新闻内容的针对性和相关性。因此,未来研究应深入探讨大型语言模型在新闻内容的自动化生成、信息筛选和个性化推荐系统中的应用效果及其对整个新闻生产流程的系统性影响。这包括评估这些模型如何提升新闻报道的质量、效率和多样性,以及这些技术如何改善新闻的传播精准度。研究者需要开展实证研究,通过数据驱动的方法来验证大型语言模型在实际新闻环境中的效能,并探索这些模型如何在不同新闻传播场景下被有效地应用。此外,研究应探索大型语言模型处理和推荐内容的机制,特别是如何通过算法优化来确保内容推荐的广度和深度。这涉及到模型如何处理复杂数据集,如何识别和适应用户行为变化,以及如何在保持新闻内容质量的同时,实现高度个性化的内容推荐。

3) 伦理层面:虽然大数据技术的应用为新闻传播领域研究提供了前所未有的便利,但也伴随着诸多伦理和技术挑战,尤其是在个人隐私保护、真实性与算法透明度等方面。未来的研究需要系统地应对这些挑战,以确保新闻传播的科技进步能够促进而非破坏媒体的公信力和社会责任。

随着新闻机构日益依赖个人数据来定制内容和推广信息,确保这些数据的安全性和保密性成为了一个重要的研究议题。未来的研究需要聚焦于如何在侵犯

个人隐私的前提下有效利用这些数据,探索和发展先进的数据匿名化和去标识化技术,以防止个人信息的泄露和滥用。此外,构建符合国际和国内法律法规的数据收集和处理机制也是未来研究的重要方向。研究者需深入探讨如何在数据收集和使用过程中,严格遵守相关隐私保护法规,例如欧盟数据保护条例和中国的个人信息保护法。这包括研究如何设计透明的数据处理流程和机制,以及如何实施有效的数据安全政策和措施。同时,研究数据保护的最佳实践也是构建公众信任的关键。未来研究应探讨如何通过教育和培训提高新闻机构工作人员的数据保护意识,研究如何利用技术和管理措施来保护消费者数据免受未授权访问和滥用。此外,研究还应评估不同数据保护策略的效果,以识别和推广那些能够最有效保障个人隐私同时支持新闻业务发展的实践和技术。

大数据技术的应用带来了数据准确性、真实性与算法透明度等一系列挑战。确保技术应用不损害数据的完整性与准确性,同时防止虚假新闻和误导信息的传播,成为当前亟需关注和解决的问题。为此,面对数据准确性、真实性和算法透明度等挑战,未来研究应着重于开发和验证确保新闻内容可靠性的技术和方法,例如,如何通过算法审计、数据验证流程等手段有效识别和纠正技术应用中可能出现的偏差和错误,确保新闻报道的真实性和公正性。研究还应该关注人工智能技术在新闻传播中的伦理问题,特别是在处理虚假新闻和误导信息传播的风险方面,探索如何利用人工智能辅助新闻机构开发高效的虚假信息识别和过滤系统,以及如何构建算法决策的透明度和可解释性,增强公众对新闻内容的信任。其次,由于算法设计往往依赖于数据输入,而这些数据本身可能反映或加剧已有的社会偏见,因此如何设计公正无偏的新闻推荐系统成为挑战。未来研究应着力于开发新的算法策略,通过引入多样性和公正性的度量标准来评估和改进算法的决策过程,同时需研究如何通过增加算法的透明度来减少偏见,确保算法的决策过程对新闻从业者和公众都是开放和可理解的。此外,鉴于自动生成的新闻内容和个性化推荐系统可能引发的信息茧房问题,未来研究应探索如何设计算法来打破信息的孤岛,推动信息的多元化传播,例如探索如何调整推荐系统的参数,以促进观点的多样性和平衡,以及如何利用算法推荐不同但具有建设性的观点,以促进更广泛的社会对话和理解。



## 参考文献(References):

- [1] 张开平, 孟天广, 黄种滨. “软宣传”的兴起、特征与效果——基于2009-2023年主流媒体与政务新媒体的大数据分析[J]. 新闻与传播研究, 2023, 30(12): 86-103.
- [2] 申楠, 苏怡丹, 马凯. “一带一路”议题全球舆论话语图景与中国应对——基于2013-2023年全球社交媒体平台X的大数据研究[J]. 情报杂志, 2024, 43(06): 153-159.
- [3] 郑路, 吴文兵, 葛天任, 等. 社会关注度的新闻大数据分析——社会、时间与空间的维度[J]. 浙江学刊, 2018(03): 82-90.
- [4] 林子青. 《解码十年》: 基于“大数据+”的多维数据新闻创新实践[J]. 传媒, 2024(04): 63-64.
- [5] 李国光, 马东丽. 5G时代的数据新闻价值: 技术与想象的耦合[J]. 电视研究, 2023(11): 68-70.
- [6] 吕宇翔, 方格格, 华伊然. 数据新闻发展的全球规律和本土实践——基于2015-2021年获奖作品的分析[J]. 新闻与写作, 2022(06): 97-109.
- [7] 刘念, 朱婧. 谁掀动了网络情绪? ——网络舆情中愤怒情绪传播的关键节点[J]. 广州大学学报(社会科学版), 2023, 22(01): 172-182.
- [8] Liu Y, Wang B, Wu B, et al. Characterizing super-spreading in microblog: an epidemic-based information propagation model[J]. Physica A, 2016, 463: 202-218.
- [9] Xing Y, He W, Cao G, et al. Using data mining to track the information spreading on social media about the COVID-19 outbreak[J]. Electronic Library, 2022, 40(1/2): 63-82.
- [10] Niu Y, Qu C, Wang G, et al. Information spreading with relative attributes on signed networks[J]. Information Sciences, 2021, 551: 54-66.
- [11] 王家坤, 孟祥宇, 郭筱彤, 等. 耦合情绪与收益的网络舆情传播及治理研究[J]. 情报理论与实践, 2024, 47(10): 140-150.
- [12] Namtirtha A, Dutta A, Dutta B. Weighted kshell degree neighborhood: a new method for identifying the influential spreaders from a variety of complex network connectivity structures[J]. Expert Systems with Applications, 2020, 139: 112859.
- [13] Ally A F, Zhang N. Effects of rewiring strategies on information spreading in complex dynamic networks[J]. Communications in Nonlinear Science & Numerical Simulation, 2018, 57: 97-110.
- [14] 唐雯云. 融媒体环境下报纸新闻编辑面临的机遇、挑战及应对策略探究[J]. 新闻研究导刊, 2024, 15(07): 98-101.
- [15] 朱贺. 社会突发事件网络舆情的传播管理研究[J]. 新闻世界, 2024(04): 62-64.
- [16] 张雅慧. 数学方法在新闻数据分析中的应用及实践研究[J]. 新闻研究导刊, 2024, 15(06): 123-125.
- [17] 刘立华, 孟凡荣. 评价理论视角下非洲媒体中国精准扶贫的话语呈现研究[J]. 北京科技大学学报(社会科学版), 2024, 40(03): 78-88.
- [18] 姚曦, 商超余. 中国广告产业发展的内生动力、首要任务及创新路径[J]. 武汉大学学报(哲学社会科学版), 2022, 75(01): 68-78.
- [19] 谢海涛, 肖倩. 面向语义缺失场景的社交媒体中热门新闻识别方法研究[J]. 现代情报, 2019, 39(09): 28-40.
- [20] 钟智锦, 王童辰. 大数据文本挖掘技术在新闻传播学科的应用[J]. 当代传播, 2018(05): 12-18.
- [21] Alaql A A, Alqurashi F, Mehmood R. Data-driven deep journalism to discover age dynamics in multi-generational labour markets from linkedIn media[J]. Journalism and Media, 2023, 4(1): 120-145.
- [22] Pinto M C, Barbosa S O. Artificial intelligence (AI) in Brazilian digital journalism: historical context and innovative processes[J]. Journalism and Media, 2024, 5(1): 325-341.
- [23] Villar-Rodríguez G, Souto-Rico M, Martín A. Virality, only the tip of the iceberg: ways of spread and interaction around COVID-19 misinformation in Twitter[J]. Communication & Society, 2022, 35(2): 239-256.
- [24] 吴小坤, 赵甜芳. 自然语言处理技术在社会传播学中的应用研究和前景展望[J]. 计算机科学, 2020, 47(06): 184-193.
- [25] Balla S N, Pani A, Sahu P K, et al. Examining shifts in public discourse on electric mobility adoption through Twitter data[J]. Transportation Research Part D: Transport and Environment, 2023, 121: 103843.
- [26] Ljajić A, Prodanović N, Medvečki D, et al. Uncovering the reasons behind COVID-19 vaccine hesitancy in Serbia: sentiment-based topic modeling[J]. Journal of Medical Internet Research, 2022, 24(11): e42261.
- [27] 谢菊莲. 人工智能技术在新闻生产中的运用策略探究[J]. 新闻研究导刊, 2023, 14(24): 66-68.
- [28] 晏辉, 陈丹. 人工智能技术下的新闻伦理问题研究[J]. 采写编, 2021(01): 112-113.
- [29] 闫婷. 大数据时代新媒体的应用与发展趋势研究[J]. 新闻研究导刊, 2023, 14(20): 99-102.
- [30] 刘海明, 马源. 消融的边界: 新闻实践的泛化表征与主流再塑[J]. 青年记者, 2023(16): 13-16.
- [31] 王仕勇, 樊文波. 向善向上: 基于良性互动的算法新闻治理伦理研究[J]. 重庆大学学报(社会科学版), 2021, 27(02): 225-236.
- [32] Mishra K. Data as a national asset[J]. Technology and Regulation, 2024(2024): 93-102.
- [33] Mann M, Daly A, Milan S, et al. (Big) Data and the North-in-South: australia's informational imperialism and digital colonialism[J]. Television & New Media, 2019, 20(4): 379-395.
- [34] Kukutai T. Indigenous data sovereignty—a new take on an old theme[J]. Science, 2023, 382(6674): 14664.
- [35] 谢波, 曹亚男. 生成式人工智能意识形态安全风险探析[J]. 国家安全研究, 2024(01): 102-127.
- [36] 杨慧. ChatGPT的意识形态属性、风险及治理[J]. 理论导刊, 2024(01): 102-108.
- [37] 邓观鹏, 顾友仁. 算法殖民: 西方意识形态算法化统治的审视与批判[J]. 新疆社会科学, 2023(05): 20-31.
- [38] 邹开亮, 刘祖兵. 论类 ChatGPT 通用人工智能治理——基



- 于算法安全审查视角[J]. 河海大学学报(哲学社会科学版), 2023, 25(06): 46-59.
- [39] Colleoni E, Rozza A, Arvidsson A. Echo chamber or public sphere? predicting political orientation and measuring political homophily in Twitter using big data: political homophily on Twitter[J]. *Journal of Communication*, 2014, 64(2): 317-332.
- [40] Papacharissi Z. Affective publics and structures of storytelling: sentiment, events and mediality [J]. *Information, Communication & Society*, 2016, 19(3): 307-324.
- [41] Nagy P, Neff G. Imagined affordance: reconstructing a keyword for communication theory[J]. *Social Media+Society*, 2015, 1(2): 362329746.
- [42] Yeung K. 'Hypernudge': big data as a mode of regulation by design[J]. *Information, Communication & Society*, 2017, 20(1): 118-136.
- [43] Papacharissi Z. Affective publics and structures of storytelling: sentiment, events and mediality [J]. *Information Communication & Society*, 2016, 19(3): 307-324.
- [44] Couldry N, Mejias U A, Milan S, et al. Data colonialism: rethinking big data's relation to the contemporary subject[J]. *Television & New Media*, 2019, 20(4): 336-349.
- [45] Neuman W R, Guggenheim L, Jang S M, et al. The dynamics of public attention: agenda-setting theory meets big data[J]. *Journal of Communication*, 2014, 2(64): 193-214.
- [46] Beer D. The social power of algorithms [J]. *Information, Communication & Society*, 2017, 20(1): 1-13.
- [47] Ceron A, Curini L, Iacus S M, et al. Every tweet counts? how sentiment analysis of social media can improve our knowledge of citizens' political preferences with an application to Italy and France[J]. *New Media & Society*, 2013, 16(2): 340-358.
- [48] Coddington M. Clarifying journalism's quantitative turn: a typology for evaluating data journalism, computational journalism, and computer-assisted reporting[J]. *Digital Journalism*, 2015, 3(3): 331-348.
- [49] Vargo C J, Guo L, Amazeen M A. The agenda-setting power of fake news: a big data analysis of the online media landscape from 2014 to 2016[J]. *New Media & Society*, 2018, 20(5): 2028-2049.
- [50] 方洁, 颜冬. 全球视野下的“数据新闻”: 理念与实践[J]. *国际新闻界*, 2013, 35(06): 73-83.
- [51] 部书锴. 场景理论的内容框架与困境对策[J]. *当代传播*, 2015(04): 38-40.
- [52] 喻国明. 镶嵌、创意、内容: 移动互联广告的三个关键词——以原生广告的操作路线为例[J]. *新闻与写作*, 2014(03): 48-52.
- [53] 部书锴. 场景理论: 开启移动传播的新思维[J]. *新闻界*, 2015(17): 44-48.
- [54] 黄薇, 李瑞, 孟佳林. 大数据环境下多媒体网络舆情传播要素及运行机理研究[J]. *图书情报工作*, 2015, 59(21): 38-44.
- [55] 付安玲, 张耀灿. 大数据助力网络意识形态治理及提升路径[J]. *马克思主义研究*, 2016(05): 105-112.
- [56] 文卫华, 李冰. 大数据时代的数据新闻报道——以英国《卫报》为例[J]. *现代传播(中国传媒大学学报)*, 2013, 35(05): 139-142.
- [57] 喻国明, 王斌, 李彪, 等. 传播学研究: 大数据时代的新范式[J]. *新闻记者*, 2013(06): 22-27.
- [58] 刘义昆. 大数据时代的数据新闻生产: 现状、影响与反思[J]. *现代传播(中国传媒大学学报)*, 2014, 36(11): 103-106.
- [59] 王斌. 大数据与新闻理念创新——以全球首届“数据新闻奖”为例[J]. *编辑之友*, 2013(06): 16-19.
- [60] Luo P, Wang C, Guo F, et al. Factors affecting individual online rumor sharing behavior in the COVID-19 pandemic[J]. *Computers in Human Behavior*, 2021, 125: 106968.
- [61] Han P K J, Scharnetzki E, Scherer A M, et al. Communicating scientific uncertainty about the COVID-19 pandemic: online experimental study of an uncertainty-normalizing strategy[J]. *Journal of Medical Internet Research*, 2021, 23(4): e27832.
- [62] Krawczyk K, Chelkowski T, Laydon D J, et al. Quantifying online news media coverage of the COVID-19 pandemic: text mining study and resource [J]. *Journal of Medical Internet Research*, 2021, 23(6): e28253.
- [63] 李国光, 马东丽. 5G时代的数据新闻价值: 技术与想象的耦合[J]. *电视研究*, 2023(11): 68-70.
- [64] 李香帅. 时政类数据新闻可视化叙事策略研究——以澎湃《美数课》栏目为例[J]. *新闻研究导刊*, 2023, 14(17): 119-121.
- [65] 许向东, 尹艺斐. 媒体融合时代传统主流大报的“守正创新”[J]. *新闻爱好者*, 2021(05): 23-27.
- [66] 张宏瑜, 杨孔雨. 人工智能应用于新闻生产流程带来的伦理失范问题及对策[J]. *新闻世界*, 2023(08): 40-43.
- [67] 张鑫. 人工智能与数字化技术在新闻中的应用探析[J]. *新闻研究导刊*, 2023, 14(15): 5-7.
- [68] 王晓娜. 媒体融合发展的创新策略探析[J]. *新闻研究导刊*, 2023, 14(13): 87-89.
- [69] Sutton J, Gibson C B, Phillips N E, et al. A cross-hazard analysis of terse message retransmission on Twitter[J]. *Proceedings of the National Academy of Sciences-PNAS*, 2015, 112(48): 14793-14798.
- [70] Wang Y, Hao H, Platt L S. Examining risk and crisis communications of government agencies and stakeholders during early-stages of COVID-19 on Twitter[J]. *Computers in Human Behavior*, 2021, 114: 106568.
- [71] Cinelli M, Cresci S, Quattrociochi W, et al. Coordinated inauthentic behavior and information spreading on Twitter[J]. *Decision Support Systems*, 2022, 160: 113819.
- [72] Chen Q, Jacobson T. A process model of the public sphere: a case of municipal policy debates on Sina Weibo[J]. *Policy and Internet*, 2022, 14(2): 485-502.
- [73] Holt K, Shehata A, Strömbäck J, et al. Age and the effects of news media attention and social media use on political interest and participation: do social media function as leveller? [J]. *European Journal of Communication*, 2013, 28(1): 19-34.
- [74] 朱鸿军. ChatGPT对新闻传播系统的颠覆性重构[J]. *探索与争鸣*, 2023(05): 29-32.