

# 基于Citespace的网络舆情研究知识图谱分析

童涛\*, 陈平平, 刘睿杰

成都理工大学管理科学学院, 四川 成都  
Email: tongtao\_2021@126.com

收稿日期: 2021年5月25日; 录用日期: 2021年6月25日; 发布日期: 2021年7月8日

## 摘要

随着互联网的迅速发展, 互联网已经成为人们传播和获取各种信息的主要手段, 移动网民大规模倍增的同时, 网络也成为人们生活的重要部分之一, 舆论的主要阵地也逐渐转移到了互联网。网络舆情已然成为研究焦点, 对目前已有成果进行文献分析, 对理论性研究和实践探索具有重要意义。本研究利用Citespace对2011~2021年以“网络舆情”为关键词的研究成果进行分析, 通过知识图谱, 明晰网络舆情的研究知识结构等。国内研究文献表明: (1) 网络舆情领域研究的研究机构分布广泛, 但研究机构之间合作较弱。核心研究作者有陈福集、兰月新、黄微、王晰巍、夏一雪、曾润喜和刘冰月, 仅形成了4个主要研究团队。(2) 网络舆情研究主要分为网络舆情特征研究、网络舆情关键技术研究、网络舆情分析与预测研究三个主题。(3) 网络舆情领域内的研究呈现出多维度、多态势的趋势。在未来的研究中, 一方面要关注理论的创新, 另一方面要将理论与实际结合, 多关注应用性研究。

## 关键词

网络舆情, 可视化, Citespace

# Knowledge Mapping Analysis of Internet Public Opinion Research Based on Citespace

Tao Tong\*, Pingping Chen, Ruijie Liu

College of Management Science, Chengdu University of Technology, Chengdu Sichuan  
Email: tongtao\_2021@126.com

Received: May 25<sup>th</sup>, 2021; accepted: Jun. 25<sup>th</sup>, 2021; published: Jul. 8<sup>th</sup>, 2021

\*通讯作者。

文章引用: 童涛, 陈平平, 刘睿杰. 基于 Citespace 的网络舆情研究知识图谱分析[J]. 服务科学和管理, 2021, 10(4): 53-61. DOI: 10.12677/ssem.2021.104010

## Abstract

With the rapid development of the Internet, the Internet has become the main means for people to spread and obtain all kinds of information. While the number of mobile Internet users has doubled, the network has become an important part of people's lives, and the main position of public opinion has gradually shifted to the Internet. Network public opinion has become the focus of research. It is of great significance for theoretical research and practical exploration to analyze the existing achievements. This study uses Citespace to analyze the research results with "Internet public opinion" as the key word from 2011 to 2021, and clarify the research knowledge structure of Internet public opinion through the knowledge map. Domestic research literature shows that: (1) research institutions in the field of Internet public opinion are widely distributed, but the cooperation between research institutions is weak. The core research authors are Chen Fuji, Lan Yuexin, Huang Wei, Wang Xiwei, Xia Yixue, Zeng Runxi and Liu Bingyue, and only four main research teams have been formed. (2) The research of network public opinion is mainly divided into three topics: the characteristics of network public opinion, the key technology of network public opinion, and the analysis and prediction of network public opinion. (3) The research in the field of Internet public opinion presents a trend of multi-dimensional and multi-state. In future research, on the one hand, we should pay attention to the innovation of theory, on the other hand, we should combine theory with practice and pay more attention to applied research.

## Keywords

Internet Public Opinion, Visualization, Citespace

Copyright © 2021 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

继报纸、广播、电视之后，互联网已成为第四大媒体[1]。它已成为信息传播的主要媒介和反映社会舆论的主要载体。网络舆情是指网民对社会上各种现象和问题的主观反映，并通过网络传播公众具有一定倾向性的意见和意见[2]。由于互联网的虚拟性、随意性、传播速度快等特点[3]，在信息正常公开传播的同时，虚假信息、反动言论、恶意炒作等负面事物迅速发展[4]。扰乱社会秩序，危害网络信息安全。如果我们对这种现象放任不管，负面的舆论将对公共安全形成更大的威胁。为了维护社会稳定，防范危险，网络舆情监测和预警变得越来越重要[5]。

本文从“网络舆情”这一视角出发，在 CNKI 平台检索搜集相关研究者的学术论文，利用 Citespace 文献可视化软件进行数据处理与分析，通过其生成的知识图谱进行机构相关性分析、作者相关性分析、关键词分析等，寻求网络舆情研究的发展规律和趋势，通过网络舆情特征研究、网络舆情关键技术研究 and 网络舆情分析与预测研究等，对网络舆情研究目前的现状进行分析和阐述，进而提出科学客观的参考建议，为研究学者提供一些借鉴。

## 2. 材料与研究方法

### 2.1. 数据来源

中国知网知识发现网络平台(CNKI)是面向海内外读者提供中国学术文献、外文文献、学位论文、

报纸、会议等各类资源统一检索、统一导航和在线阅读和下载服务的平台，其具备了数据量大、数据量全的优点。因此，文本以 CNKI 为数据库进行文献数据收集：以“网络舆情”为检索词，设定检索时间为 2011~2021 年，共检索出 564 篇文献(检索时间为 2021 年 5 月 20 日)。为了保证文献的准确性，人为剔除不相关文献后，共收集到 524 篇文献。以 CKNI 为研究数据采集平台，保证了研究的客观性和科学性。

## 2.2. 研究方法

知识图谱(Knowledge Graph)，在图书情报界称为知识域可视化或知识领域映射地图，是显示知识发展进程与结构关系的一系列各种不同的图形，用可视化技术描述知识资源及其载体，挖掘、分析、构建、绘制和显示知识及它们之间的相互联系[6]。知识图谱是通过将应用数学、图形学、信息可视化技术、信息科学等学科的理论与方法与计量学引文分析、共现分析等方法结合，并利用可视化的图谱形象地展示学科的核心结构、发展历史、前沿领域以及整体知识架构达到多学科融合目的的现代理论。它能为学科研究提供切实的、有价值的参考。

Citespace 是一款着眼于分析科学分析中蕴含的潜在知识，是在科学计量学、数据可视化背景下逐渐发展起来的一款引文可视化分析软件[7]。由于是通过可视化的手段来呈现科学知识的结构、规律和分布情况，因此也将通过此类方法分析得到的可视化图形称为“科学知识图谱”[8]。本研究运用 Citespace 软件，以“网络舆情”为关键词，以 CNKI 检出出来的文献数据为研究对象，得到该研究领域的知识图谱，以求从知识图谱的角度廓清网络舆情研究领域的整体情况，客观真实地阐述其发展趋势。在知识图谱中，节点(字体)越大，表示出现的凭此越高；连线越粗，表示节点之间的关系越强。基于此，辨别节点所代表内容的学术贡献度。

## 3. 结果与分析

### 3.1. 研究机构、研究作者共现分析

探究研究领域内研究机构之间的关系，能够分析研究机构的学术探究深度以及研究机构之间的合作强度[9]。利用 Citespace 软件生成网络舆情研究的研究机构共现知识图谱(图 1)，可以通过量化的形式更加科学有效地评价研究机构的学术影响力。通过统计该领域内发表论文数量较多的研究机构，得到表 1。综上，图 1 表明网络舆情研究的核心研究机构，以及机构之间的关系，表 1 则说明核心研究机构的发文

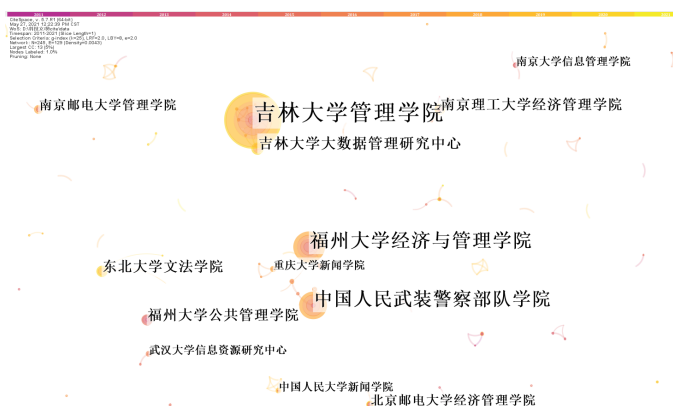


Figure 1. A co-present map of research institutions

图 1. 研究机构共现图谱

**Table 1.** Statistical table of the number of papers published by research institutions**表 1.** 研究机构论文发表数量统计表

研究机构	发表数量
吉林大学	44
福州大学	24
中国人民武装警察部队学院	22
华中科技大学	14
南京理工大学	9

数量。进一步分析可知：(1) 从总体来看：该领域研究的研究机构分布广泛，说明了网络舆情研究十分火热。核心研究机构的发表论文数量比较丰富，一定程度上说明了在该领域研究有所进展。然而，研究机构之间大多以单独节点存在，机构之间的合作较弱。(2) 从区域的角度来看：该领域研究的研究机构多分布在中国中东部，这也从一定程度上说明了发展较快的地区更能洞悉到该领域研究的重要意义。

利用 Citespace 生成网络舆情研究作者共现知识图谱(图 2)，其能反映出该研究领域的主流研究学者(核心作者)，并且能得到作者之间的互引关系以及合作强度，进而更加科学合理地评估该领域内研究学者的学术影响力。普拉斯定理[10]表明，主流研究学者的发文数量代表了学术水平值。结合网络舆情研究可以得知，发表文章数量在 7 篇以上的研究学者为核心作者，由此可以根据软件生成结果统计出研究核心作者统计表(表 2)。由此可知，网络舆情研究领域的核心作者是：陈福集、兰月新、黄微、王晰巍、夏一雪、曾润喜和刘冰月，而研究仅仅形成了 4 个主要研究团队。从作者群的角度分析可知，研究作者之间的分布为整体集中、部分发散的状态；研究团队中的作者大多来自于同一个研究机构，存在内部紧密联系，外部联系较少的情况。

**Figure 2.** A co-present map of the study authors**图 2.** 研究作者共现图谱**Table 2.** A statistical table of the core authors of the study**表 2.** 研究核心作者统计表

作者	数量
陈福集	30
兰月新	26

## Continued

黄微	17
王晰巍	17
夏一雪	13
曾润喜	9
刘冰月	8

基于此,综合研究机构共现知识图谱、研究作者共现知识图谱可知,研究核心作者来自于研究核心机构,这种现象说明了机构于作者在推动研究进程方面之间的关系:研究机构为研究学者提供的科研环境不同,研究学者在研究领域内的研究成果也不同。

### 3.2. 研究主题分析

研究主题能够反映出不同时段内研究领域的研究视角变化。与此同时,关键词代表了学术论文的核心观点,是对研究主题的高度凝练和概括。利用 Citespace 生成研究关键词共现知识图谱(图 3),解释网络舆情研究领域的研究主题。

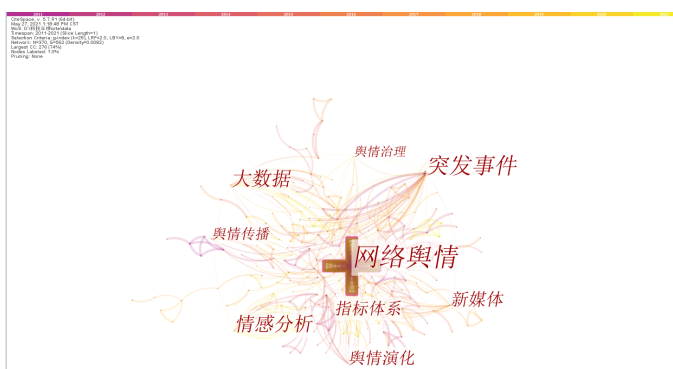


Figure 3. A co-present map of the key words  
图 3. 研究关键词共现图谱

根据 Citespace 生成结果,合并内涵相似的关键词,提取该领域内的高频关键词(表 3)。高频关键词从某种程度上揭示了研究领域的重点内容。为了更加准确地分析网络舆情的研究内容,在此基础上,生成关键词演变共现知识图谱(图 4),更加客观的反应出网络舆情的研究主题。基于知识图谱节点,进行文献梳理,将网络舆情的主要研究主题归纳为:网络舆情特征、网络舆情关键技术、网络舆情分析与预测三个方面。接下来,分别对其进行阐述。

Table 3. High-frequency words in the research literature

表 3. 研究文献高频词

关键词	频次	关键词	频次
网络舆情	428	信息传播	10
突发事件	49	情感分析	10
大数据	25	舆情传播	9
新媒体	16	知识图谱	9
指标体系	12	舆情演化	9

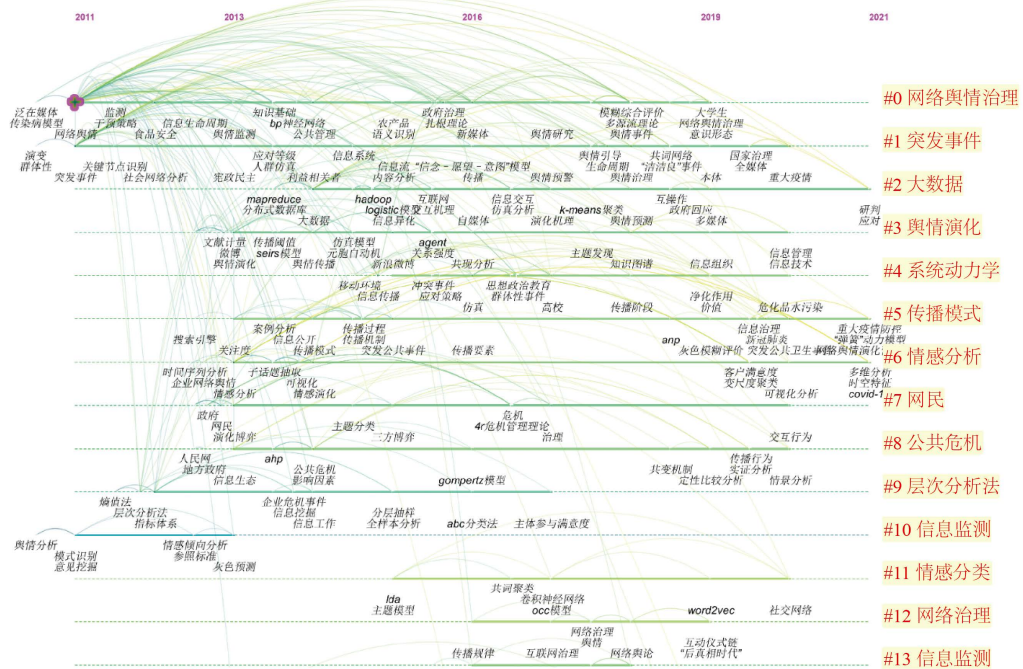


Figure 4. A co-present map of the evolution of the researchers' keywords

图 4. 研究关键词演变共现图谱

首先，在网络舆情特征研究方面：网络舆情作为社会民意的一种表达形式，是指公众针对现实生活中的一些热点问题，通过网络发表的具有较强影响力和倾向性的言论和观点[11]。通过对这方面的相关文献进行梳理，可以发现网络舆情可以通过多样化的信息和互动方式方便快捷地表达。网络的开放性和虚拟性决定了网络舆论的以下特征：(1) 直接性：网民可以通过论坛、新闻评论、博客网站、微信等即时表达自己的观点。这样民意就可以更加顺畅地表达；网络舆情也具有即时、快速传播的可能性。信息只能通过复制和粘贴来转发[12]。这一特性使得它很容易通过阻塞，让监管部门束手无策[13]。(2) 随机性和多样性：虚拟的、匿名的、无边界的、即时互动的“网络社会”使得网络舆情在价值传递、利益需求等方面呈现多元化和边缘性[14]。各种文化类型、意识形态、价值观、生活标准和道德规范都可以找到自己的位置。既有积极健康的舆论，也有庸俗灰色的舆论。总之，网络舆情的内容丰富多彩。(3) 难以预测：网络打破了时间和空间的界限。重大新闻事件在成为网络关注焦点的同时，也迅速成为舆论热点。由于网络可以实时更新，网络舆情可以以最快的速度传播[15]。(4) 隐蔽性：互联网是一个虚拟世界，说话人的身份是隐蔽的。此外，它缺乏规则和限制，以及有效的监督。自然，网络成为一些网民倾诉情感的空间[16]。(5) 衍生性：网络舆情是社会状况和民意中最活跃、最具有情感的部分。然而，网络舆情并不能等同于所有人的立场。随着互联网的普及和新闻跟进、论坛、博客的出现，网民开始拥有前所未有的话语权，可以相对自由地表达自己的观点和感受。然而，由于法律伦理在网络空间的制约作用较弱，网民缺乏自律，就会产生一些不负责任的言论，如热衷于公开他人隐私、用谣言蛊惑人心、反社会倾向、极端主义和非理性、群体盲从、冲动等[17]。

互联网让每个人都成为信息发布者，而无需经过审查。一位媒体研究者声称：“一方面，它使信息更加公开透明；同时，由于真假混杂，你无法辨别这些信息的真假。对于年轻网民来说，他们的情绪是很容易被煽动起来的。”

其次，在网络舆情关键技术研究方面，主要对两大关键技术展开研究：舆情信息搜集技术和舆情信

息提取技术。

### 1. 舆情信息收集技术[18]

由于民意的多样性和复杂性，网络舆情信息采集具有一定的难度。传统的舆情信息采集方法是人工完成的，局限性大，效率低。网络信息采集依赖于网络爬虫技术。它主要利用网页之间的关系来收集网页。爬虫技术的研究方向如下：(1) 基于全网的信息收集。它可以从一些种子 URL 扩展到整个 web 信息集合，完成整个 web 集合。它目前在实际应用中占主导地位。(2) 增量式 web 信息采集。它只收集刷新页面时已经更改或新生成的页面，而不能收集未更改的页面。(3) 基于主题的 Web 信息采集。集合是通过选择性搜索与预定义主题关联的那些页面来完成的。(4) 个性化的信息收集。它可以通过用户交互的方式收集满足用户不同需求的信息。(5) 基于 agent 的信息收集。智能代理系统是指计算机系统，它可以像人类一样在环境中独立灵活地进行活动，能够独立感知用户兴趣的变化并调整获取策略。不同信息采集方法对同一查询请求返回的处理结果不同，质量参差不齐。这主要是由于其自身的优点和缺点。在实际采集过程中发现，单独采用一种方法是不太令人满意的。它会错过很多合格的页面。所以我们通常使用两种或两种以上的方式来完成信息的获取。

### 2. 舆情信息提取技术[19]

网页内容通常包括导航、标题、文本、相关链接和广告信息。Web 信息提取就是从源文件中提取网页内容，识别与主题相关的信息。常用的提取技术有以下几种：(1) 基于网页特征的提取技术网页布局一般采用 HTML 标记。数据之间有一个层次结构，按时间顺序排列。使用结构特征的 html 页面布局，或者使用正则表达式技术可以得到相对纯文本的消息。该技术可以去除网页中的附加噪声信息，如网页广告、链接到其他页面、描述性信息页面等。它可以有效地使信息提取更加准确和纯净。(2) 基于自然语言处理的信息提取该技术将整个网页作为文本文档进行处理。因此，更适合于需要提取信息量较大的信息。该方法是对传统语言处理技术的应用，包括句法分析、语义分析和识别。提取过程还包括一些提取规则的使用。本体是一个形式化的概念。用它来描述相关领域之间的关系，以提供对知识的共同理解，最终实现知识共享。(3) 基于本体的方法利用信息描述数据本身，建立本体库，利用指定的抽取规则完成信息抽取。这种方法的优点是不依赖于页面的结构。只要正文足够大，就可以提取该领域的各种信息。(4) 基于隐马尔可夫模型的信息提取技术。隐马尔可夫模型是一种有限状态自动机，是信号处理的重要手段。目前已在语音识别、行为识别等领域取得了相当成功的应用。近年来，该模型被广泛应用于信息抽取中，目前已产生了非常显著的效果。稳健地处理新数据是可能的。

最后，在网络舆情分析与预测方面：网络舆情特征研究和网络舆情关键技术研究，为舆情的分析和预测提供了支撑。舆情分析是舆情分析体系的核心。目前，数据挖掘中的聚类和分类方法都是用来解决这类问题的[20]。其最终目的是通过民意分析来预测未来的民意趋势。网络舆情预测就是研究如何利用海量的预警数据来预测未来的网络舆情，即通过分析过去来预测未来。舆情分析与预测技术主要包括以下任务：

(1) 识别热点话题和敏感话题[21]。互联网上的海量信息包含着大量的话题。按主题领域可分为：军事、体育、娱乐、科技、社会等。根据话题的重要性，可以分为一般话题和热点、敏感话题。目前，主要是在 web 文本中发现对该话题的挖掘，然后测度相关报告对该话题的总量、增长率、波动水平、有用用户对该话题回复密集、情绪倾向等维度来识别某一时期的热点话题。同时，我们通过关键词匹配和进一步的语义分析来确定它是否属于敏感话题。

(2) 话题倾向性分析主要是指情感倾向性分析[22]。进一步延伸可以细分为政治取向分析、产品偏好分析和利益偏好分析。其主要目的是分析发表的文章、博客、回复和微博，对网络进行语义分析。它可以决定其情感倾向，如褒义、贬义或中性。

(3) 趋势预测网络舆情分析[23]。分析预测的主要焦点,目前有两个方面,一是受到舆论的关注程度的影响,主要是预测趋势报告的数字和数量的回复,另一种是矛盾的复杂性程度的公众舆论,主要趋势来预测他们的情感倾向。目前广泛应用的 BP 神经网络模型是基于自回归移动平均模型(ARIMA)、决策树模型和灰色理论相结合的算法。

## 4. 结论

网络舆情研究内容丰富,涉及的学科分布广泛。本研究梳理了网络舆情的研究脉络,展现了网络舆情研究进展,发现目前网络舆情研究的理论成果较多,而应用性研究还不够丰富。面对网络舆情,现实生活中还存在以下问题:比如机构设置、力量配备不到位;法律、法规赋予的管理权限小;对舆情管理的规律、自身的优势缺少认识等问题。近年来,突发公共卫生事件的爆发无疑对世界各国造成了沉重的影响,如何借助网络舆情引导事情向好的方向发展具有重大意义。

本研究利用信息可视化软件 Citespace 对 2011~2021 年以网络舆情为关键词的学术论文进行信息挖掘分析,绘制出知识图谱,进一步归纳总结出网络舆情领域的知识结构关系等,研究发现:

(1) 从研究现状来看,网络舆情领域研究的研究机构分布广泛,多分布在中国中东部,核心研究机构的发表论文数量比较丰富,研究机构之间合作较弱。研究作者之间的分布为整体集中、部分发散的状态;研究团队中的作者大多来自于同一个研究机构,存在内部紧密联系,外部联系较少的情况。

(2) 从研究总体脉络来看:网络舆情研究主要分为三部分:网络舆情特征研究、网络舆情关键技术研究、网络舆情分析与预测研究。

(3) 从发展趋势来看:网络舆情领域内的研究呈现出多维度、多态势的趋势。在未来的研究中,一方面要关注理论的创新,另一方面要将理论与实际结合,多关注应用性研究。

## 基金项目

四川省哲学社会科学重点研究基地区域公共管理信息化研究中心一般项目(项目编号: QGXH20-03;项目名称:突发公共卫生事件中的大数据驱动旅游地响应决策研究)、成都理工大学 2020 年度校级科技立项(项目名称:突发公共卫生事件下网络舆情特征分析及引导策略研究)。

## 参考文献

- [1] 刘晶晶. 新形势下网络媒体的比较分析[J]. 传播力研究, 2020, 4(1): 181-182.
- [2] 沈国权, 马海兵, 张煜. 建议构建军队网络舆情分析系统[J]. 南京政治学院学报, 2007, 23(3): 105.
- [3] 孟庆峰. 论新媒体管理中的网络舆情监测与引导机制建构[J]. 智库时代, 2020(5): 112-113.
- [4] 张鑫. 新媒体时代下环保网络舆情危机的应对研究[D]: [硕士学位论文]. 郑州: 郑州大学, 2017.
- [5] 郑昌兴. 关于突发事件网络舆情监测预警的探讨[J]. 电脑知识与技术, 2020, 16(33): 72-74.
- [6] 冯新翎, 何胜, 熊太纯, 等. “科学知识图谱”与“Google 知识图谱”比较分析——基于知识管理理论视角[J]. 情报杂志, 2017(1): 153-157.
- [7] Chen, C. (2014) CiteSpace II: Detecting and Visualizing Emerging Trends and Transient Patterns in Scientific Literature. *Journal of the Association for Information Science & Technology*, **57**, 359-377.
- [8] 李杰, 陈超美. Citespace: 科技文本挖掘及可视化[M]. 北京: 首都经济贸易大学出版社, 2016: 32.
- [9] 胡泽文, 孙建军, 武夷山. 国内知识图谱应用研究综述[J]. 图书情报工作, 2013, 57(3): 131-137, 184.
- [10] 刘永胜. 《情报资料工作》核心著者测评[J]. 情报资料工作, 2003(2): 49-50.
- [11] 徐海涛. 重大突发公共事件的网络舆情传播及政府治理研究——以 X 市“诬陷”舆情为例[D]: [硕士学位论文]. 郑州: 郑州大学, 2018.
- [12] 王玲玲. 大学生网络舆论特征及其引导[J]. 思想理论教育(上半月综合版), 2006(5): 36-39.



- 
- [13] 曹效阳, 曹树金, 陈桂鸿. 网络舆情的结构与网络特征分析[J]. 情报科学, 2010, 28(2): 231-234, 238.
- [14] 柯缙祖. 网络舆论特点研究[J]. 红旗文稿, 2011(9): 4.
- [15] 赵耀, 王建新. 基于多元主体共在与信息即时公开的新冠肺炎疫情网络舆情的思考[J]. 中国矿业大学学报: 社会科学版, 2020(2): 88-100.
- [16] 徐琳琳. 网络中的虚拟自我探析[D]: [博士学位论文]. 大连: 大连理工大学, 2010.
- [17] 郑雯, 桂勇. 网络舆情不等于网络民意——基于“中国网络社会心态调查(2014)”的思考[J]. 新闻记者, 2014(12): 10-15.
- [18] 孙志国. 网络涉农舆情收集与预警技术[J]. 农业网络信息, 2010(1): 80-81.
- [19] 刘华春, 王星捷. 网络舆情信息提取技术研究是实现[J]. 计算机技术与发展, 2016(9): 8-11.
- [20] 梁艺琼. 基于数据挖掘技术的舆情分析系统的设计[J]. 电脑知识与技术, 2020, 16(3): 7-8.
- [21] 张海龙, 淦文燕, 陈刚, 等. 基于词共现网络的海量文本分析与热点话题发现[J]. 计算机与数字工程, 2015(10): 1729-1735.
- [22] 李路. 新媒体时代的网络舆论现状与引导对策分析[J]. 新闻研究导刊, 2016, 6(6): 82-83.
- [23] 魏思超. 媒体报道倾向与受众态度的关联分析及其对“两个舆论场”的影响——以财经媒体报道为例[J]. 新闻研究导刊, 2020, 11(6): 175-176.