

分类号：
学号：20232108060

密级：公开
单位代码：10759

石河子大学

硕士学位论文



基于文本挖掘的高校网络舆情分析

学位申请人	王鑫
指导教师	徐学彬
申请学位门类级别	图书情报专业硕士
学科、专业名称	图书情报
研究方向	大数据与信息管理
所在学院	信息科学与技术学院

中国·新疆·石河子
2026年5月

分类号：
学 号：20232108060

密 级：公开
单位代码：10759

石河子大学

硕 士 学 位 论 文



基于文本挖掘的高校网络舆情分析

学 位 申 请 人	王鑫
指 导 教 师	徐学彬
申请学位门类级别	图书情报专业硕士
学 科、专 业 名 称	图书情报
研 究 方 向	大数据与信息管理
所 在 学 院	信息科学与技术学院

中国·新疆·石河子

2026年5月

**Text Mining-Based Analysis of Universities
Network Public Opinion**

A Dissertation Submitted to
Shihezi University
In Partial Fulfillment of the Requirements
for the Degree of
Master of Library and Information Science

By

Wangxin
(Big Data and Information Management)

Dissertation Supervisor: Xu Xue-bin

May, 2026

石河子大学学位论文独创性声明及使用授权声明

学位论文独创性声明

本人所呈交的学位论文是在我导师的指导下进行的研究工作及取得的研究成果。据我所知，除文中已经注明引用的内容外，本论文不包含其他个人已经发表或撰写过的研究成果。对本文的研究做出重要贡献的个人和集体，均已在文中作了明确的说明并表示谢意。

研究生签名： 王鑫

时间： 2026年5月25日

使用授权声明

本人完全了解石河子大学有关保留、使用学位论文的规定，学校有权保留学位论文并向国家主管部门或指定机构送交论文的电子版和纸质版。有权将学位论文在学校图书馆保存并允许被查阅。有权自行或许可他人将学位论文编入有关数据库提供检索服务。有权将学位论文的标题和摘要汇编出版。保密的学位论文在解密后适用本规定。

研究生签名： 王鑫

时间： 2026年5月25日

导师签名： 杨芳彬

时间： 2026年5月25日

摘要

网络社交媒体使高校舆情生成与演化日益集中，而高校网络舆情因常涉及教育公平、学术伦理等敏感议题，其科学分析与有效治理已成为推进高等教育治理现代化的重要课题。本文以北京协和医学院 4+4 临床医学教育改革试点项目引发的网络舆情事件为案例，以新浪微博为数据获取平台，综合运用文本挖掘、LDA 主题建模、情感分析等研究方法，结合生命周期理论与焦点事件理论，分析高校网络舆情在不同发展阶段的演化规律，为高校网络舆情的科学管理提供实证参考样本。

研究采集 2025 年 4 月至 8 月微博平台与研究案例相关的舆情数据，经清洗与预处理后，构建包含 28179 条有效博文的结构化数据集。基于发文量、转发数、评论数三项核心指标构建综合热度指数，采用熵权法确定各指标权重，将该高校网络舆情事件划分为响应启动期、平稳过渡期、结论发布与恢复期三个特征显著的阶段。在此基础上，从热度演化、主题分布、情感倾向三个维度，对舆情全周期及各发展阶段展开多维度实证分析，研究发现高校网络舆情热度呈现爆发式启动、阶梯式衰减、节点式激活的时序特征，日内传播规律与公众日常作息高度契合；舆情主题随生命周期发展呈现层级跃迁特征，最终在官方通报发布后，升级至对教育制度系统性修复的建设性反思，整体演化趋势体现为核心稳定、外延拓展、理性升级；舆情情感从启动期的对抗分化状态，逐步过渡为平稳收敛的格局，正向情绪随官方关键信息的发布逐步占据主导，消极情绪则从初期的集中宣泄转向后期克制的监督与反思。

基于实证分析结果，本文从政府、高校、教师、学生四大核心主体出发，提炼高校网络舆情的多主体协同治理启示。构建高校网络舆情的系统性治理制度体系，提出强化学风建设筑牢舆情治理根基、完善过程干预机制把控舆情演化节奏、建立长效保障制度推动舆情常态化治理的具体路径，推动高校网络舆情治理从被动应对转向主动预防，从零散处置转向系统治理。

关键词：高校网络舆情；舆情生命周期；文本分析

Abstract

The in-depth popularization of online social media has made the cyberspace a core field for the generation and evolution of university-related online public opinion. As such public opinion often involves sensitive issues such as educational equity and academic ethics, its scientific analysis and effective governance have become a key task in advancing the modernization of higher education governance. Taking the online public opinion incident triggered by the 4+4 pilot program for clinical medicine education reform at Peking Union Medical College as a case study, this thesis uses Sina Weibo as the data acquisition platform, comprehensively adopts research methods including text mining, LDA topic modeling and sentiment analysis, and combines the life cycle theory and the focal event theory to analyze the evolutionary laws of university-related online public opinion at different development stages, so as to provide an empirical reference sample for the scientific management of university-related online public opinion.

Relevant public opinion data of the research case was collected from Sina Weibo from April to August 2025. After cleaning and preprocessing, a structured dataset containing 28,179 valid microblogs was constructed. A comprehensive heat index was built based on three core indicators, namely the number of posts, reposts and comments, with the entropy weight method adopted to determine the weight of each indicator. Accordingly, the university-related online public opinion incident was divided into three distinct stages: the triggering stage, the stable evolution stage, and the official notice release and recovery stage. On this basis, a multi-dimensional empirical analysis was carried out on the whole life cycle and each development stage of the public opinion from the three dimensions of heat evolution, topic distribution and sentiment tendency. The results show that the heat of university-related online public opinion presents a temporal characteristic of explosive initiation, stepwise attenuation and node-based activation, and its intraday communication law is highly consistent with the daily schedule of the public, featuring concentration in the daytime and dispersion at night. The topics of public opinion show a hierarchical transition characteristic with the development of the life cycle, and eventually rise to constructive reflections on the systematic improvement of the education system after the release of the official notice, with the overall evolutionary trend characterized by stable core, extended connotation and rational upgrading. The sentiment of public opinion gradually transitions from a state of confrontation and differentiation in the initiation stage to a pattern of stability and convergence; positive emotions gradually take a dominant position with the release of key official information, while negative emotions shift from concentrated venting in the early stage to restrained supervision and reflection in the later stage.

Based on the empirical analysis results, this thesis extracts the enlightenment for the multi-subject collaborative governance of university-related online public opinion from the perspective of four core subjects: the government, universities, teachers and students. It constructs a systematic governance system for university-related online public opinion, and puts forward specific paths including strengthening the construction of academic atmosphere to lay a solid foundation for public opinion governance, improving the process intervention mechanism to control the evolution rhythm of public opinion, and establishing a long-term guarantee system to promote the normalized governance of public opinion, so as to transform the governance of university-related online public opinion from passive response to active prevention, and from scattered disposal to systematic governance.

Key words:Universities online public opinion; Public opinion life cycle; Text analysis

目录

摘要.....	I
Abstract	II
第 1 章 绪论.....	1
1.1 研究背景与研究意义.....	1
1.1.1 研究背景.....	1
1.1.2 研究意义.....	2
1.2 国内外研究现状.....	4
1.2.1 关于高校网络舆情的研究.....	4
1.2.2 关于文本挖掘应用的研究.....	6
1.2.3 关于舆情生命周期的研究.....	9
1.2.4 研究述评.....	11
1.3 研究内容、方法、技术路线.....	12
1.3.1 研究内容.....	12
1.3.2 研究方法.....	12
1.3.3 技术路线.....	13
1.4 研究创新点.....	14
第 2 章 基本概念和理论基础.....	16
2.1 基本概念.....	16
2.1.1 网络舆情的概念.....	16
2.1.2 网络舆情的特征.....	16
2.1.3 高校网络舆情的概念.....	17
2.1.4 高校网络舆情的特征.....	17
2.2 理论基础.....	18
2.2.1 生命周期理论.....	18
2.2.2 焦点事件理论.....	19
2.3 本章小结.....	19
第 3 章 数据获取与数据预处理.....	20
3.1 数据获取.....	20
3.1.1 案件选取.....	20

3.1.2 案件回顾.....	21
3.1.3 数据来源.....	23
3.1.4 数据获取.....	24
3.2 数据获取及预处理.....	25
3.2.1 数据清洗.....	25
3.2.2 数据预处理.....	26
3.3 舆情生命周期划分量化依据.....	27
3.3.1 周期划分的依据.....	27
3.3.2 周期划分的量化指标.....	28
3.3.3 周期划分结果.....	29
3.3.4 划分结果的验证.....	29
3.4 本章小结.....	30
第4章 高校网络舆情演化分析.....	31
4.1 热度演化的阶段分析.....	31
4.1.1 事件响应启动期热度演化分析.....	31
4.1.2 事件平稳过渡期热度演化分析.....	32
4.1.3 事件结论发布与恢复期热度演化分析.....	34
4.2 主题变迁的阶段分析.....	35
4.2.1 事件响应启动期主题建模分析.....	36
4.2.2 事件平稳过渡期主题建模分析.....	39
4.2.3 事件结论发布与恢复期主题建模分析.....	42
4.3 情感特征的阶段分析.....	46
4.3.1 事件响应启动期网络舆情的情感特征分析.....	47
4.3.2 事件平稳过渡期网络舆情的情感特征分析.....	48
4.3.3 事件结论发布与恢复期网络舆情的情感特征分析.....	50
4.4 全周期舆情演化的综合分析.....	52
4.4.1 热度演化节律分析.....	52
4.4.2 主题演化路径分析.....	54
4.4.3 情感演化序时分析.....	56
4.5 本章小结.....	58
第5章 高校网络舆情启示.....	59
5.1 高校网络舆情分析的多主体治理启示.....	59
5.1.1 政府统筹协调与制度保障.....	59
5.1.2 高校从被动应急到主动治理.....	60

5.1.3 教师专业发声与科普沟通.....	60
5.1.4 学生理性参与与责任意识.....	61
5.2 高校网络舆情治理的制度构建.....	61
5.2.1 强化学风建设筑牢舆情治理根基.....	61
5.2.2 完善过程干预机制把控舆情演化节奏.....	62
5.2.3 建立长效保障制度.....	62
5.3 本章小结.....	62
第6章 结论与展望.....	63
6.1 研究结论.....	63
6.2 研究不足与展望.....	64
6.2.1 研究不足.....	64
6.2.2 研究展望.....	64
参考文献.....	66

第1章 绪论

1.1 研究背景与研究意义

1.1.1 研究背景

当下数字化浪潮正在席卷全球，互联网已经不仅仅是一种前沿技术，更是支撑着当前社会生活与社会经济发展的重要工具。目前，国内居民互联网的参与度以及参与规模宏大。根据中国互联网络信息中心（CNNIC）发布的第57次《中国互联网络发展状况统计报告》，截至2025年12月，我国网民总数已达11.25亿，互联网普及率攀升至80.1%^[1]，稳居世界首位。更为关键的是，其中有99.6%的网民是通过智能手机等移动设备接入网络，这意味着我国已全面进入高度移动化、密集化的网络时代。这一庞大而活跃的数字人口基础，不仅催生了繁荣的数字经济，更深刻地改变了信息传播、社会交往与公共参与方式的多种渠道，使网络空间成为与现实社会平行交织、即时互动的第二社会，而网络空间的意见表达、情感流动与共识形成，对现实社会进程产生着不可忽视的影响^[2]。

在此背景下，网络社交媒体的高度普及重塑了舆情的生成与演化机制。以新浪微博为代表的平台，集成了大众传播的广度、人际传播的深度与即时通信的速度，构建了一个高度复杂、动态开放的线上广场^[3]。在这个广场上，信息传播呈现出去中心化、裂变化和情感驱动的鲜明特征^[4]。任何局部事件都可能被瞬间置于公共关注之下，信息传播与群体互动中快速发酵、变形、放大，形成裹挟巨大社会能量的网络舆情。这种舆情不仅快速反映社情民意，更可能设置公共议程、施加舆论压力、甚至左右事件走向，对政府治理、企业运营和机构声誉构成严峻挑战。因此，对网络舆情的科学监测、精准分析与有效引导，已成为国家治理体系和治理能力现代化的重要组成部分。

高校作为思想文化的高地、知识创新的源泉和青年人才的摇篮，其网络舆情生态具有独特的敏感性、复杂性和影响力^[5]。大学生群体是互联网的原住民和社交媒体最活跃的用户之一，他们思维敏捷、权利意识强、表达意愿高，同时又处于世界观、人生观、价值观定型的关键时期^[6]。校园内部涉及教育公平、学术伦理、师德师风、管理服务乃至社会热点事件的观点碰撞，极易在高度同质化、互动密集的学生社群中引发强烈共鸣^[7]。由于高校本身的社会关注度高，加之学生群体与家庭、社会联系紧密，校园舆情极易突破学生群体，从校内热议迅速演变为全民关注的公共事件。若应对失当，不仅会损害高校的声誉与公信力，影响正常教学科研秩序，还可能激化社会矛盾，关乎社会稳定

与和谐。因此,深入研究高校网络舆情的发生发展规律,构建与之相适应的分析响应体系,是维护高校稳定、促进高等教育健康发展、进而引领积极向上网络文化的必然要^[8]。

面对微博等平台产生的海量、非结构化、实时涌流的舆情数据,传统依赖人工巡查、经验判断和定性描述的舆情监测方法已经难以及时掌控发展动态,难以实现全景式把握、前瞻性预警和深层动因挖掘。因此,文本挖掘技术作为自然语言处理与数据科学交叉融合的前沿领域,可以为解决这一困境提供强大的方法论工具。利用文本挖掘技术能够自动、高效地从大规模文本数据中提取特征、发现规律、聚类主题并研判情感,将杂乱无章的数据转化为可视化的信息^[9]。将文本挖掘技术应用于高校网络舆情分析,可以超越对表面热度与情绪宣泄的简单追踪,转而深入洞察舆情的核心议题分布、观点演化路径、群体情感结构及关键传播节点,从而实现从被动应对到主动洞察、从表象描述到机理阐释的跨越^[10]。

综上所述,本研究选题在文本挖掘的支撑下,对高校网络舆情进行探析,正是在宏观数字化生存已成常态、中观社交媒体重塑舆论生态、微观高校稳定面临新挑战、方法论上技术革新提供新可能的多重背景下确立的。研究旨在通过构建一个整合文本挖掘技术的综合分析框架,以代表性的微博平台数据为样本,对高校网络舆情进行一场深入肌理的解析,以期揭示其内在的、可量化的规律,为构建科学、精准、高效的高校网络舆情分析体系,贡献具有实证基础的参考。

1.1.2 研究意义

在当前高度媒介化与深度数字化的社会语境下,网络舆情不仅是现实社会议题的映照,更是推动社会进程、影响政策制定的重要力量,其有效治理直接关系到社会和谐稳定与意识形态安全^[11]。

高校网络舆情极易演变为社会舆情甚至引起更大的反响。高校作为培育国家未来栋梁、孕育前沿思想、汇聚高知青年群体的特殊场域,其网络舆情生态独具特点:议题往往与教育公平、学术诚信、青年价值观等深层社会关切紧密相连;传播因师生群体的高媒介素养与密集社交网络而异常迅速;影响则因其与家长、关联单位的关联及社会对高等教育的高度期待而被急剧放大。这使得高校网络舆情治理,成为国家社会治理现代化进程中一个极具挑战性的前沿课题。大数据、人工智能等前沿技术的迅猛发展,为实现对海量、复杂舆情信息的精准感知、智能解析与科学干预提供了革命性工具。然而,当前研究在涉及高校的网络舆情对以上技术的研究和利用还不多,严重制约了治理效能。因此,本文的分析研究具有重大意义。

在理论层面,本研究的意义在于深化与拓展现有高校网络舆情研究的方法。传统舆情研究多侧重直接引用舆情生命周期理论进行划分,对于高校舆情发布调查结果或者处

理通告后仍会引发舆论争议加以考虑。本研究通过引入生命周期理论并系统应用文本挖掘技术,旨在实现三个方面的理论推进:第一,通过对微博平台上海量高校舆情文本进行周期划分,可以超越主观归纳,客观、量化地识别出高校师生群体长期、稳定关注的核心议题谱系,揭示不同舆情生命周期的议题最容易触发校园舆论共振。第二,通过主题聚类、情感分析,可以精准测绘出在不同事件中师生群体的情绪演化轨迹与观点分歧结构,从而从情感动力学视角深化对舆情发酵与平复机制的理解。第三,结合传播路径分析,可以揭示信息在传播之间的交互规律^[12]。这些基于实证数据的发现,能够为构建更具解释力、更贴合高校情境的网络舆情理论模型提供坚实的数据基石,弥补当前研究中对高校舆论场域关注不足的理论缺憾。

在方法层面,本研究的意义体现在推动研究方式从定性经验判断向定量数据驱动的推进。文本挖掘技术作为自然语言处理与数据科学的交叉领域,其价值在于能够将非结构化的、看似杂乱的文本数据,转化为可计算、可分析的结构化信息。本研究以微博这一具有代表性的开放式舆论广场为数据来源,获取多个维度的文本数据,将文本数据转化为可视化的信息,从词频、共现、语义关联等维度发现不易察觉的潜在模式,从而实现舆情分析的规模化、客观化与前瞻性^[13]。这一套方法体系的探索与验证,可为后续相关领域的研究提供一个可参考、可拓展的技术框架,具有显著的方法示范价值。

在实践应用层面,本研究的意义在赋能高校网络舆情治理体系的现代化上提供一些参考,提升其精准性、效率与韧性。研究的结果致力于形成可直接或间接服务于高校治理实践的参考模板^[14]。其应用价值具体体现在:第一,服务于精准化风险预警。通过构建基于文本挖掘的实例,可对后续构建预警机制,实现微博等平台上的校园相关议题进行早期扫描、热度评估与情感倾向判断,帮助管理者在海量信息中精准定位潜在风险点,争取宝贵的处置窗口期。第二,服务于科学化舆情研判。当事件发生后,系统可快速生成主题演化图、情感波动曲线和观点集群报告,帮助管理者穿透纷繁复杂的表面言论,把握舆情的核心矛盾、主流民意与关键少数意见,从而避免因误判而采取不当应对措施。第三,服务于效果评估与策略优化。在发布应对措施后,可通过持续的情感与主题分析,量化评估干预措施的网络反响与舆论引导效果,为策略的动态调整提供即时反馈。这实质上是推动高校治理从依赖个人经验的传统,迈向基于数据洞察的科学,全面提升治理能力的现代化水平。

最后,在更广泛的社会价值层面,本研究的意义还在于筑牢高校稳定防线、营造清朗网络空间并贡献社会治理智慧。通过提升高校相关的网络舆情治理能力,能够有效防范个体诉求演化为群体事件、学术争议升级为公共危机,切实维护校园的政治安全、意识形态安全与和谐秩序^[15]。同时,健康的网络舆情生态本身就是最好的育人环境。此外,理解并引导好高校网络舆论,有助于在青年学生中培育理性、平和、积极向上的讨论文化,这本身就是立德树人的重要一环,也是构建和谐网络空间的重要环节。高校作为知

识密集、技术前沿的微型社会，其在网络舆情治理数字化转型中的探索、经验与模式，无论是成功的经验还是遇到的挑战，都能为更广泛的社会组织、社区乃至城市层面的网络社会治理，提供具有价值的先行探索，升华为对国家治理体系和治理能力现代化的有益贡献。

1.2 国内外研究现状

本文在研究高校网络舆情的过程中运用了文本挖掘技术和舆情生命周期理论，下面将针对这三个领域的国内外研究现状分别进行分析，从而为本论文的研究奠定基础。

1.2.1 关于高校网络舆情的研究

以CNKI为检索平台，设置主题为“高校网络舆情”，选择“主题”作为检索选项，设置检索时间为2016年1月1日到2025年12月31日，近十年来国内有关高校网络舆情的相关文献数量为3175篇，文献的发布趋势如图1-1所示：

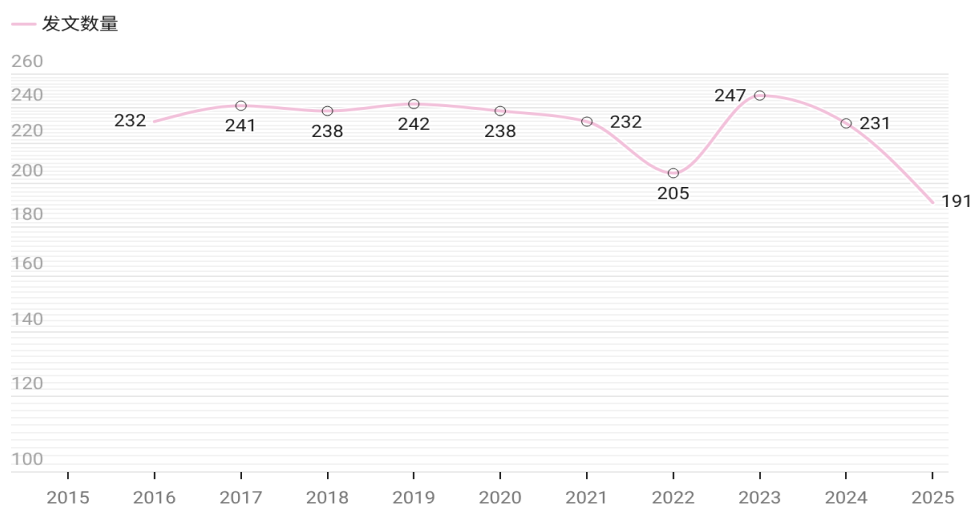


图1-1 国内关于高校网络舆情近十年发文量

国内的现有研究主要集中在两个维度：一是舆情特征与成因的归纳，二是治理策略的构建，但两者之间的理论衔接不足。陶佳苹（2022）分析了高校网络舆情呈现出的极高的集群性、群体的非理性化、扩散的不可控性、传播的裂变性等特点，并提出了有效引导机制^[16]。李文辉（2023）认为高校网络舆情的成因是学生校园热点共鸣、信息传播渠道众多、社会关注度高、高校监督管控不足等四个方面^[17]。然而，上述特征归纳多基于个案观察，缺乏大规模数据的实证检验。丁然（2023）以“安徽审计职业学院百度贴吧”为例，爬取贴吧内的文本数据，利用分词、词频统计、词云图制作等技术进行数据分析，为校园相关部门舆情引导提供帮助和新的思路^[18]。王茜，何轲等（2024）以“苏

州大学学生造女生黄谣事件”为例，对获取的微博数据进行时间序列分析，在划分舆情发展时期的基础上，使用 LDA 主题模型和 Bi-LSTM 神经网络结构模型对网络舆情主题和情感进行融合分析，结合事件发生过程中的新闻事件探究网民关注热点和情感演化情况^[19]。值得注意的是，“管理者向引导者转变”的命题虽具理论吸引力，但如何操作化实施，现有研究尚未提供清晰的阶段划分标准。李之泓,翟崑等（2025）以算法时代为背景，针对高校学生网络舆情呈现的情感极化、圈层共振等新特征，进行了传播学视阈下的治理分析。提出了从静态管控转向动态引导、构建多元共治格局、依托数据驱动智慧平台、重塑信任沟通关系等治理策略，为构建理性对话、凝聚共识的校园网络空间提供了理论框架与实践路^[20]。刘鸿滨,孙海清（2025）立足新时代背景，审视了高校网络舆情治理在意识形态渗透、传播生态泛政治化、技术环境变革及学生能力局限等方面面临的多重风险，并在此基础上，提出应通过处理好守正与创新、出击与预防、虚拟与现实、快速与稳妥四对关系，系统构建网络空间新格局、新生态、新场域与新底色，从而全方位提升高校网络舆情治理能力^[21]。

以 Web of science 为检索平台，设置主题为“University Online Public Opinion”，选择“主题”作为检索选项，设置检索时间为 2016 年 1 月 1 日到 2025 年 12 月 31 日，近十年来国外有关高校网络舆情的相关文献数量为 360 篇，文献的发布趋势如图 1-2 所示：

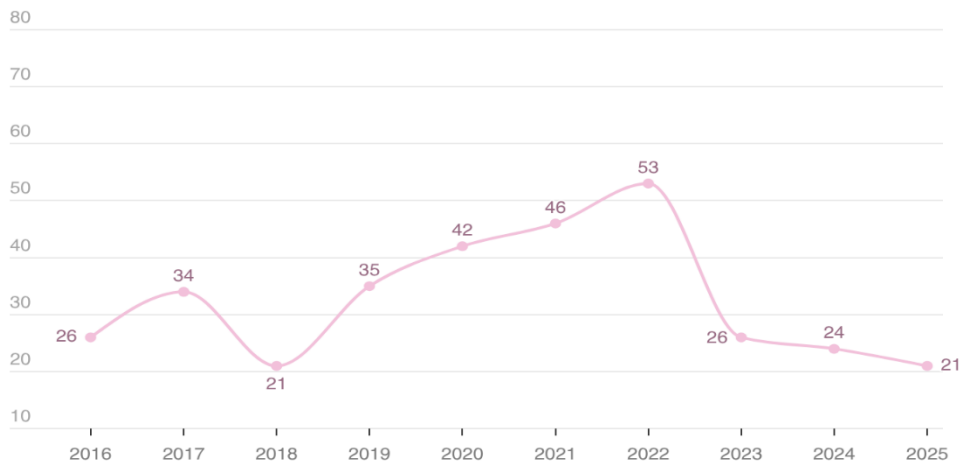


图 1-2 国外关于高校网络舆情近十年发文量

国外对于高校网络舆情方面，美国学者 Philip H.Coombs（1985）的著作《The World Crisis In Education: The View From The Eighties》中首次提到“高校危机”，他认为，学生数量的激增、学生学习需求和动机的增强、教育基础设施的不足以及教育不公平等因素都是导致高校危机的原因。这本著作被普遍认为是较早关于高校危机管理的系统研究之一。《School crisis response combat guide》是由美国学者勒纳（1996）等人所著的一本书。这本书在国外高校应对校园危机事件中具有相当大的影响力和权威性。书中详细介

绍了校园意外伤亡、自杀暴力、寻衅滋事、性骚扰等危机，并提出了相应的应对策略。勒纳指出，高校应当加强教育培训，并不断提高应对突发事件的管理水平。然而，上述危机管理框架形成于前社交媒体时代，其“控制—修复”逻辑与当前去中心化传播环境存在明显张力。近年来，Wei-na He 等（2022）以大学网络舆情动态监测系统为案例，运用网络爬虫与 MapReduce 并行计算技术，结合机器学习聚类与概率神经网络异常检测方法，构建了高精度、短周期的舆情监测模型。该模型在信息处理能力、并发性能及响应时间上均优于传统方法，为维护校园安全稳定提供了技术支撑，其可扩展性设计可应用于其他领域网络舆情监测，具有显著的实践应用价值与学术创新意义^[22]。该研究的技术路径虽具参考价值，但其“监测—预警”的单一功能定位，难以回应高校舆情治理中“何时回应、如何回应”的决策需求。

国内外研究的共同局限在于，多将高校舆情视为一般公共事件的子集，对其“官方回应引发二次波动”的特殊演化规律关注不足。当前国内高校网络舆情研究呈现治理和政策导向趋势，国外的学者更早地提出了高校网络舆情这一概念，关注点更加聚焦于高校网络舆情的识别与监测方面。

1.2.2 关于文本挖掘应用的研究

文本挖掘技术的研究起源于国外信息检索与自然语言处理领域的交叉探索。20世纪90年代，随着互联网文本数据的爆发式增长，国外学者率先开展了基于统计模型的文本特征提取与主题识别研究。文本挖掘这一概念最早在1995年由乌萨马·M·费尔德曼（Usama M. Feldman）与同事首次正式提出“文本挖掘”（Text Mining）这一概念，又称“文本中的知识发现”，Knowledge Discovery from Text, KDT 的术语，并将其定义为“从非结构化文本文档中抽取用户感兴趣的、重要模式或知识的过程”。

以 CNKI 为检索平台，设置主题为“文本挖掘”，选择“主题”作为检索选项，设置检索时间为2016年1月1日到2025年12月31日，研究层次选“应用研究”近十年来国内有关文本挖掘应用的相关文献数量为2380篇，文献的发布趋势如图1-3所示：