分类号 C8/276　　密级
U D C　　编号 10741

兰州财经大学
LANZHOU UNIVERSITY OF FINANCE AND ECONOMICS

硕 士 学 位 论 文

(专业学位)

论文题目 突发公共卫生事件网络舆情演化规律及情感倾向时空特征研究

研 究 生 姓 名：金亚亚

指导教师姓名、职称：刘明　教授

学 科、专 业 名 称：统计学　应用统计专业硕士

研 究 方 向：大数据分析

提 交 日 期：2021 年 6 月 6 日

独创性声明

本人声明所呈交的论文是我个人在导师指导下进行的研究工作及取得的研究成果。尽我所知，除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。

学位论文作者签名：金亚亚 签字日期：2021.6.6

导师签名：刘明 签字日期：2021.6.6

导师(校外)签名：荣康骥 签字日期：2021.6.6

关于论文使用授权的说明

本人完全了解学校关于保留、使用学位论文的各项规定，同意（选择“同意”/“不同意”）以下事项：

1.学校有权保留本论文的复印件和磁盘，允许论文被查阅和借阅，可以采用影印、缩印或扫描等复制手段保存、汇编学位论文；

2.学校有权将本人的学位论文提交至清华大学“中国学术期刊（光盘版）电子杂志社”用于出版和编入 CNKI《中国知识资源总库》或其他同类数据库，传播本学位论文的全部或部分内容。

学位论文作者签名：金亚亚 签字日期：2021.6.6

导师签名：刘明 签字日期：2021.6.6

导师(校外)签名：荣康骥 签字日期：2021.6.6

Research on the evolution of network public opinion and emotional temporal and spatial characteristics of public health emergencies

Candidate : Jin Yaya

Supervisor: Liu Ming

摘要

突发公共卫生事件引发的网络舆情难以预测，且易在社会层面内引发负面情绪，导致舆论环境恶化，进而对管理决策和应对措施产生不利影响。因此，挖掘突发公共卫生事件网络舆情演化规律及情感时空变化特征，可以为政府及相关管理部门在处理相关事件时提供参考，及时识别舆情演化的周期阶段和情感倾向时空特征，更能为引导舆论方向和差异化防控提供科学依据。

新冠肺炎疫情对经济、社会及人民生活产生了巨大的影响。本文以新冠肺炎疫情为案例，研究突发公共卫生事件网络舆情演化规律及情感时空变化特征。首先从微博中爬取有关新冠肺炎疫情的文本数据。其次，结合生命周期理论和微博转发数划分出网络舆情传播阶段，接着结合 TF-IDF 算法与 LDA 主题模型，挖掘事件在时序上的舆情主题演化规律和映射出的情感倾向。最后将用户地理信息和文本中蕴含的地理信息引入网络舆情分析中，进行网络舆情空间分析，获得不同区域网民对新冠肺炎疫情事件的关注程度与情感倾向。

研究发现，在时间维度上，网络舆情传播周期一般可以分成四个阶段，分别为酝酿期、爆发期、波动期及长尾期。不同阶段下网络舆情主题演化呈现出主题关联、主题易变的规律。在舆情主题演变的背景下情感倾向变化主要可以分为两个阶段，前期的情感主要为紧张、慌恐和焦虑的负面情绪，后期的情感主要为振作、自信、平复及积极的正面情绪；在空间层面上，突发公共卫生事件网络舆情具有较强的空间特性，不仅具有全域覆盖性和全民参与性，而且与突发公共卫生事件本身空间分布特性整体具有相似一致性，但是存在局部差异性，呈现出的区域情感值与疫情事件的严重程度为负向相关性。

关键词：突发公共卫生事件 网络舆情 新冠肺炎 主题演化规律 情感倾向 空间分布

Abstract

Internet public opinion caused by public health emergencies is difficult to predict, and it is easy to trigger negative emotions at the social level, leading to a deterioration of the public opinion environment, which has a negative impact on management decision-making and response measures. Therefore, mining public opinion evolution rules and emotional spatiotemporal characteristics of public health emergencies can provide a reference for the government and relevant administrative departments when dealing with related events, and timely identify the evolutionary cycle stages of public opinion and the temporal and spatial characteristics of emotional tendencies, which can be more guidance the direction of public opinion and differentiated prevention and control provide a scientific basis.

The new crown pneumonia epidemic has had a huge impact on the economy, society and people's lives. This paper uses the new crown pneumonia epidemic as a case to study the evolutionary law of public opinion on the Internet and the characteristics of emotional temporal and spatial changes in public health emergencies. Firstly, the paper obtain text data about the novel crown pneumonia epidemic from Weibo. Secondly, combining The Life Cycle Theory and the number of Weibo reposts to divide the network public opinion dissemination stage, and then combining the TF-IDF algorithm and the LDA topic model to mine the

evolution law of the public opinion theme and the emotional tendency mapped out in the event. Finally, the user's geographic information and the geographic information contained in the text are introduced into the online public opinion analysis, and the spatial analysis of the online public opinion is carried out to obtain the degree of attention and emotional orientation of netizens in different regions to the new crown pneumonia epidemic.

The study found that in the time dimension, the network public opinion dissemination cycle can generally be divided into four stages, namely the incubation period, the outbreak period, the volatility period and the long tail period. At different stages, the evolution of online public opinion topics shows the laws of topic relevance and variability. In the context of the evolution of public opinion themes, the change of emotional tendency can be divided into two stages. The early emotions are mainly negative emotions such as tension, panic and anxiety, and the later emotions are mainly refreshing, confident, calm and positive positive emotions. At the spatial level, public opinion on public health emergencies has strong spatial characteristics. It not only has global coverage and public participation, but also has similar consistency with the spatial distribution characteristics of public health emergencies, but there are local differences. The regional emotional value presented is negatively correlated with the severity of the epidemic event.

Keywords: Public health emergencies; Network public opinion; COVID-19 epidemic；Topic evolution；Emotional tendency; The spatial distribution

目录

1 绪论

1.1 研究背景及意义

1.1.1 研究背景

在互联网日益普与网络强国持续建设的背景下，各种新兴媒体应运而生，且对社会稳定、经济发展及政治文化等各方面影响日益深入。2020 年 4 月 28 日，CNNIC 发布的第 45 次《中国互联网络发展状况统计报告》中显示，2020 年初，由于新冠肺炎疫情影响巨大，大多数网络应用平台的用户数量大幅度增长，网民数量突破 9 亿，较 2018 年底增长 7508 万，互联网普及率达 64.5%[55]。互联网的发展不仅拓宽了民众交流表达意见的空间，而且加快了网络舆情酝酿、发酵和传播的速度。网络舆情是互联网快速发展而形成的，是公众就各种社会问题在网络空间中发表的不同意见和态度的信息[56]。近些年，舆论生存环境和传播方式发生了较大的变化，习近平总书记曾说，要善于运用网络传播规律，不断创新改进宣传方式，弘扬主旋律[24]。网络舆情传播速度快和影响范围广的特点，既能在突发事件发展过程中起到重要的导向作用，加快突发事件有效解决的速度，也能引发社会范围内负面情绪激增，影响社会稳定和经济发展，甚至破坏党和政府在人民心中的形象。

近几十年，突发公共卫生事件频发（SARS，H1N1 流感，新冠肺炎疫情等），这种事件由于控制难度大、暴发性强及应对周期长等特点，极易在社会层面内引发负面情绪，导致舆论环境恶化，进而引发社会局势动荡和经济发展受阻等问题，然而政府及相关管理部门对突发公共卫生事件网络舆情爆发后的应对措施上存在时间和空间上的滞后性，对管理决策和应对措施的产生不利影响，容易引发次生危机。新冠肺炎疫情爆发于 2020 年初，冲击力极强，对社会稳定、经济发展及人民幸福生活产生较大的影响，其是对我国突发公共卫生事件防控管理的一次考验，更是对突发公共卫生事件背景下的网络舆情管理的一次重大挑战[26]。因此，本文以新冠肺炎疫情为案例，深层挖掘突发公共卫生事件网络舆情演化规律及情感倾向时空变化特征，及时识别舆情演化的周期阶段，在舆情演化周期研判的基

础上，实现舆情预警、引导和差异化管控。同时，可以为政府及相关管理部门在处理相关事件时提供参考，进一步提高我国网络舆情治理能力，构建可持续的突发公共卫生事件网络舆情应对管理体系提供参考意见[48]。

1.1.2 研究目的及意义

突发公共卫生事件的突发性和未知性使得其引发的网络舆情不同于经济社会常态化下的网络舆情，经济社会常态化下的网络舆情通常具有演化周期短，区域及群体覆盖面小的特点，而突发公共卫生事件网络舆情往往是演化周期长且呈现全域全民覆盖的特点。突发公共卫生事件网络舆情在传播生命周期中一般会伴随时间推移而发生主题演化，在舆情主题演化的背景下，公众所表达的情感会存在较大的差异，其中所表达的负面情感会对事件的传播速度和舆情走向产生直接影响，进而影响社会正常治安秩序，打乱人民的正常生活。此外，突发公共卫生事件通常具有地域性，其引发的网络舆情也存在空间差异性，如果不能差异化制定防控管理政策，会引发区域群体性负面情绪，使网络舆情发展势头恶化，极易导致次生危机，使本来难以应对的突发公共卫生事件愈发困难。因此，本文以新冠肺炎疫情事件为例，探析舆情传播随时间变化而发生的主题演化与情感倾向变化，结合各个区域疫情病例数据与实际严重程度，同时量化不同区域的网络用户对疫情发展的关注程度与在此过程中呈现出的情感特征，分析区域层面上网络舆情空间分布特性，进而进行网络舆情社区发现分析。深层挖掘突发公共卫生事件所引发的舆情在网络环境下随时间和空间变化而变化的发展历程、传播规律，在时间层面上，为政府及相关部门及时识别网络舆情演化阶段，有效监控与预警以及制定相应决策提供参考；在空间层面上，根据区域差异性为相关部门提供制定差异化精准防控管理措施的理论支撑；最后，结合网络舆情演化与情感倾向时空特征，及时识别舆情演化的周期阶段，实现舆情预警、引导和差异化管控的管理流程，为突发公共卫生事件网络舆情的应对管理提供较完备的思路与方法。同时，及时回顾突发公共卫生事件网络舆情的发展历程，有助于发现在应对管理过程中存在的短板与纰漏，总结经验教训，形成突发公共卫生事件网络舆情应对管理系统，为未来有效应对突发情况做足准备，实现合理降低社会治理成本。

研究突发公共卫生事件网络舆情具有理论意义与现实意义。

（1）理论意义

提出了一个对网络舆情研究有价值的研究框架，通过分析典型案例，将时间和空间因素同时引入到突发公共卫生事件网络舆情的传播周期演变与情感倾向特征方面的研究。在时间层面上，对突发公共卫生事件舆情演化规律与情感倾向变化进行分析，挖掘舆情传播周期各阶段主题演化规律，分析不同阶段下主题与情感有何变化，结合舆情事件本身探析网络舆情发生演变的原因，整理出网络舆情传播的普遍规律，为政府及相关部门及时的识别出舆情演化阶段，更加有效的制定决策提供依据；在空间层面上，发掘微博网络用户和蕴含在文本数据的地理位置信息，将其纳入网络舆情分析中，获得不同区域网络用户对突发公共卫生事件网络舆情的关注程度与情感倾向，分析疫情严重程度不同的区域之间网络用户的关注程度与情感值的差异，根据区域差异性对于重点区域制定差异化精准防控治理策略。最后，结合时空特征，为未来突发公共卫生事件舆情应对提供较完备的思路与方法。

（2）现实意义

新冠肺炎疫情传染性较强，破坏力大，对社会经济和人民的健康生活产生了巨大影响，疫情爆发时期正逢春节，面对新冠肺炎疫情的未知性和高传染性，居家隔离成为全民抗击疫情的重要措施之一。如此一来，互联网媒体平台成为公众获取外界信息及表达交流意见的重要途径，面对疫情的严峻性，网络舆情一时呈井喷式爆发，短时间内给网络舆情应对和疫情防控等带来了多重考验和压力。因此，研究新冠肺炎疫情的发展过程中网络舆情在不同时间段中主题演化和情感倾向变化，以及不同区域公众对新冠肺炎疫情的关注度和情感倾向，有助于政府及相关部门采取针对性的控制舆情措施和防疫措施，提升政府公信力和官媒影响力，避免滋生二次舆情。

1.2 研究文献综述

本文以 Web of Science 数据库（包含 SCI、CSSCI）和中国知网（包含 SCI、EI、CSSCI 和中文核心期刊等）为文献来源，对国内外有关网络舆情分析的相关文献进行检索，进而梳理总结出本文的研究思路。

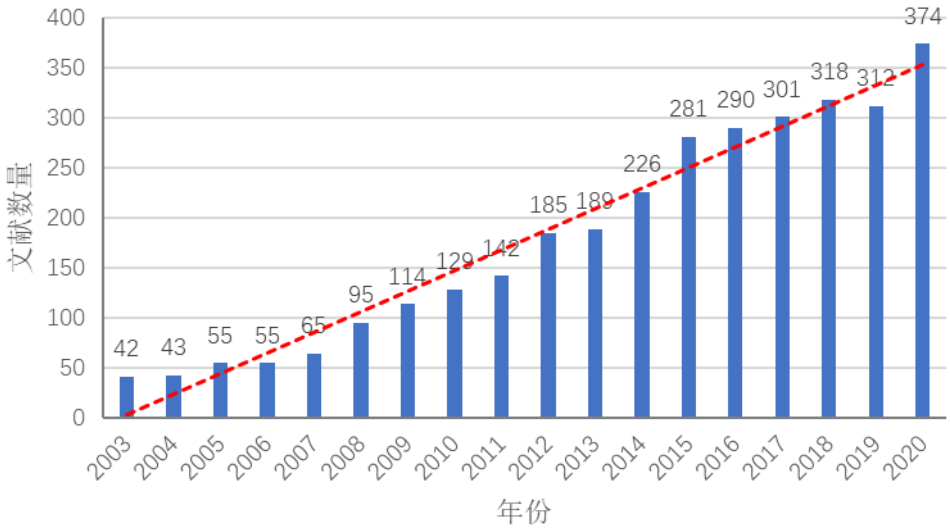


图 1.1 国外网络舆情研究文献统计图

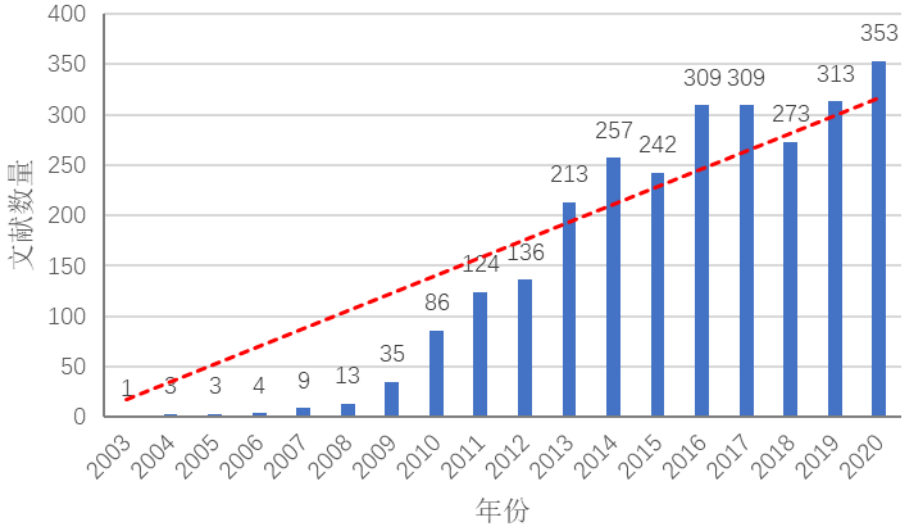


图 1.2 国外网络舆情研究文献统计图

以“online public opnion”&“internet public opinion”等为关键词在 Web of Science 数据库中对国外网络舆情研究进行文献检索，检索 2003 年至 2020 年文献数量总共为 3216 篇，并绘制国外相关研究的文献发表年份与文献数量图，如图 1.1 所示；以“网络舆情演化”&“网络舆情情感分析”&“微博网络舆情”&“网络舆情分析”为关键词在中国知网中对国内网络舆情研究进行文献检索，检索出 2003 至 2020 年文献数量总共为 2683，并绘制国内相关研究的文献发表年份与文献数量图，如图 1.2 所示。

由图 1.1 和图 1.2 可以发现国内外对网络舆情研究基本起源于 2003 年前后，其中，国内相关研究相对滞后，但是国内外相关研究均呈线性增长态势，2018

年和 2019 年有小幅下降，但是 2020 年的文献数量显示继续呈正增长。自 SARS 爆发之后，舆情研究逐渐进入到我国学者的视野，但主要是以理论研究为主[51]。2008 年之后，学者从不同视角对网络舆情展开分析研究，但是多数属于定性分析，定量分析成果较少[16]。2010 年之后，国内学者有关突发公共卫生事件引起的网络舆情的研究步入正轨，国外学者则已经将网络舆情研究平台转向新兴媒体平台[46]。其中国外新兴媒体平台主要集中于 Twitter、Facebook，国内的主要集中于微博、微信及各大论坛等[7]。

总的把握国内外相关研究趋势之后，本文从网络舆情演化及情感分析的内容和技术与方法这两个方面对近几十年来的国内外相关文献进行梳理总结。

1.2.1 突发公共卫生事件网络舆情的研究内容

通过梳理文献可以发现，近几年，学者们对突发公共卫生事件背景下的网络舆情演化机理和传播规律进行了研究，但是缺乏典型案例。网络舆情演化是舆情发展过程中公众所热议的话题及话题中所表达出的情感倾向的发展和演变，目前，网络舆情演化分为演化周期研究和演化过程研究[33]。

在演化周期研究方面，主要遵循生命周期理论，生命周期表示各种事物从发生、发展到消亡的整个过程，网络舆情也是这样的一个过程[42]。这方面的研究已经较为成熟，网络舆情生命周期划分方法尚未形成完全一致的标准，但是依据的原理基本一致。将网络舆情传播周期划分为四个阶段是最普遍的方法，如宋海龙等人依据网民情绪划分突发事件网络舆情的发展过程，最终划分为形成期、高涨期、波动期和淡化期[44]。也有学者将网络舆情的周期划分为潜伏期、发生期、持续期和恢复期这四个阶段，如梁冠华、鞠玉婷等[37]。赵岩等将利用生命周期理论（酝酿期、发生期、持续期和消退期这四个阶段）对网络舆情典型案例进行分析，划分为不同阶段并且分析各阶段的差异性，为解决该类事件提供较为完备[54]。也有部分学者运用其他理论对网络舆情演化过程进行研究，如兰月新等人通过构建微分方程模型来识别舆情扩散过程中的三个关键时间节点及舆情演变的四个阶段[30]。wang、康维等人运用社会网络分析方法，研究传播路径、传播速度和传播范围受舆情网络传播结构的影响情况[29]、[12]。

在演化过程研究方面，主要是在演化周期划分的基础上融入深层研究公众的评论主题文本及所表达处的情感倾向等因素。如 Zhang L 等利用复杂网络中的情感传染研究公众情感随着时间变化的演化规律[14]。如安璐等利用 LDA 模型和 SOM 方法挖掘 Ebola 病毒的微博热点演化模式和时序趋势有什么不同[15]。蒋知义等以“罗一笑”事件为案例，用情感倾向分析模型，统计出该事件微博文本的情感倾向与网络舆情演化阶段，探究出网络舆情演化特征与规律[28]。国外学者 Gomide J 等人对 Twitter 上登革热病流行期间的舆情信息进行分析，主要研究民众慌恐情绪及健康状态[7]。

1.2.2 突发公共卫生事件网络舆情的研究方法

研究突发公共卫生事件网络舆情主题演化及情感倾向时空特征，有助于政府及相关部门及时了解网络舆情发展态势，从而做出及时准确的舆论引导工作。本文对网络舆情主题演化和情感倾向分析所用到的相关方法进行归纳总结。学者们研究突发公共卫生事件网络舆情演化和情感倾向时主要关注时序上的变化，用到的方法主要是主题模型与情感分析，而忽略了突发公共卫生事件的地理特性。董悦等指出研究某一事件网络舆情时，不能将情感分析和主题建模分割开，要综合使用分析[23]。

通过主题提取识别，对大量文本数据进行处理分析，帮助用户快速了解信息内容，发掘信息主题[38]。刘铭等利用 TF-IDF 方法构造了多条词链，用来表达文章的叙事线索，也就是获取特征词向量，进而构造出多条反映主题信息的词汇链[39]。随着以贝叶斯为基础的函数分布概念引入文本预料分析，以 LDA 为代表的主题模型被引入研究中；如 Mark G 等人为刻画出战争主题的时间线，使用主题模型来提取阿拉伯语及英语的博客主题[9]；Dey L 等人利用主题模型对多起突发公共卫生事件（SARS、RVF 和 VEE）进行验证事件过程中指示与预警分级模型[4]；如 Griffiths T L 等在原始时间轴上表示主题演化[8]；曾子明等利用 LDA 主题模型，提取出有关红绿蓝幼儿园虐童事件评论的主题演化特征[18]。目前 LDA 在主题识别中有着良好的表现且对于微博这类短文本的主题识别有着良好的效果[32]。

情感分析是根据对某一主题内容的了解和分析，得出其反映出的情感态度，包括时序情感变化情况[27]。网络舆情研究中情感分析是极为重要的一部分，通过不断地发展，其反映出的情感特征更为准确。Esuli 等人为了判断词语的情感特征，对主观词的注释信息进行了定量分析[5]；Neppalli VK 利用朴素贝叶斯模型对 12 年“桑迪”事件的网民情感规律进行了呈现[10]；陈福集等利用 HowNet 情感词典对复旦投毒案件进行情感分析，对情感进行正负两极的情感分类[20]。

将突发事件网络舆情与空间因素结合，利用 GIS 平台及其他空间可视化方法的优势，将网络舆情和地理空间完备结合，揭示网络舆情的时空分布特征，更好的表达网络舆情与现实之间的深层关系[35]。对网络舆情情感倾向在空间层面上的分析的相关研究较少。CHOI S 等人针对社会大数据，将 Twitter 用户地理位置纳入研究，进行了灾情演化的分析[3]；徐迪等认为网络舆情起源于多变的空间中，网络舆情在时空中以动态呈现，空间可视化研究网络舆情最为直观[50]。新冠肺炎疫情爆发后，有部分学者逐渐将空间因素考虑进网络舆情研究中，比较有代表性的有陈兴蜀等人以省级行政区为空间单元将全国各省的民众情感值等进行空间可视化，反映出疫情传播的空间差异性[21]。

1.2.3 文献评述

根据上述相关文献综述来看，掌握网络舆情的演化规律，识别舆情传播演化周期，把握公众对网络舆情的关注程度及情感倾向特征是政府及相关部门实施舆情正向引导和舆情管控的重要依据，网络舆情主题演化及情感倾向特征在不同时间阶段呈现不同的特征。突发公共卫生事件多为地理事件，首发于某一区域，由其引发的网络舆情一般具有空间分布上的差异。研究突发公共卫生事件网络舆情在时空维度下的演化规律和情感倾向特征显得极其重要。

目前，在突发公共卫生事件网络舆情主题演化及情感倾向特征的研究方面已经成效显著，但是仍存在以下的几点不足：在分析案例方面，现有研究缺少典型案例的实证研究，目前对新冠肺炎疫情事件进行网络舆情主题演化及情感时空特征挖掘分析的案例更少；在分析维度方面，现有研究大多数只对舆情事件在时间维度上进行舆情传播规律和情感倾向分析，缺少从空间维度对突发公共卫生事件进行挖掘分析。

基于此，本文主要构建了研究突发公共卫生事件网络舆情演化规律及情感倾

向时空特征的理论框架。在大数据背景下，利用文本挖掘技术、主题模型、情感分析方法、空间可视化技术和舆情网络社区发现算法，以新冠肺炎疫情为案例，从时空维度进行主题演化规律与情感倾向变化特征探索与实证分析，分析网络舆情随时间变化而发生的主题演化情况与情感变化情况，深层挖掘网络舆情主题演化的原因。最后结合新冠肺炎疫情本身数据，在空间维度上将微博用户地理位置信息与文本数据中蕴含的地理位置信息引入网络舆情分析，获得不同区域公众对于新冠肺炎疫情的关注程度与情感倾向的差异情况。为政府及相关部门针对网络舆情实施有效舆情演化周期识别，舆情预警监控、制定合理决策提供必要的理论支撑，同时根据网络舆情的空间差异性为制定差异化精准防控措施提供重要的科学依据。此外，有效拓展了关于突发公共卫生事件的研究领域。

1.3 研究内容及框架

1.3.1 研究内容

本文以新冠肺炎疫情引发的网络舆情为例，研究突发公共卫生事件网络舆情演化规律及情感时空变化特征。首先从微博中爬取有关新冠肺炎疫情的微博及评论文本数据，接着进行数据清洗与文本规范化处理。其次，结合生命周期理论与微博转发数划分舆情传播阶段，在此基础上，结合 TF-IDF 算法与 LDA 主题模型对处理完的文本数据挖掘事件在时间序列上的舆情主题演化规律，然后分析舆情事件在主题演化背景下的情感倾向，结合事件本身发展情况探究网络舆情发生变化的深层原因。最后结合新冠肺炎疫情确诊病例数据，在空间层面上挖掘网络用户所属地理位置和文本中蕴含的地理位置信息，对确诊病例数和网络用户空间分布情况进行对比，同时计算出不同区域网络用户情感平均值，并且进行网络舆情社区发现分析，获得不同区域网络用户对网络舆情的关注程度与情感倾向的差异情况。基于此，文章研究内容可以分为六个部分：

第一部分为绪论。主要介绍本文研究背景与研究意义，其次通过查阅国内外有关突发公共卫生事件网络舆情的相关文献资料，梳理出国内外研究现状。

第二部分为相关技术理论介绍。主要构建研究突发公共卫生事件网络舆情演化及情感时空特征的理论框架，在时间维度层面，主要涉及 TF-ID 算法、LDA

主题模型、文本情感分析；在空间维度层面，介绍相关的空间统计分析方法与舆情社区挖掘算法。

第三部分为数据采集与预处理。本文以“新冠肺炎疫情”这件典型的突发公共卫生事件为案例，选择新浪微博采集关键词为“新冠肺炎疫情通报”的原微博博文、点赞数、评论数、转发数、评论文本与评论者地址等。首先对采集到的信息进行预处理，包括去空去重处理等，其次进行文本规范化处理。

第四部分为网络舆情主题演化分析，结合生命周期理论与微博转发数划分新冠肺炎网络舆情事件传播阶段，在划分阶段的基础上对处理完的文本数据结合TF-IDF 算法与 LDA 主题模型挖掘事件在时间走向上的舆情演化规律，然后分析网络舆情在主题演化背景下的情感倾向变化，最后深层挖掘网络舆情发生变化的原因。

第五部分为网络舆情空间分布分析，通过挖掘微博用户的地理位置和蕴含在文本信息中的地理位置，将其纳入研究中，结合新冠肺炎疫情病例数据，在空间层面上挖掘不同区域公众对于新冠肺炎疫情的关注程度与情感倾向的差异。

第六部分为结论与启示，首先整理出本文研究结论，分析研究中存在的不足，同时为未来的突发公共卫生事件应对提供较完备的思路和方法。最后提出关于突发公共卫生事件网络舆情管理工作的启示与对网络舆情主题演化与情感倾向时空变化相关领域的研究展望。

1.3.2 研究框架

研究框架如图 1.3 所示，主要分为三个部分，分别是数据采集处理层、舆情分析层和结果验证分析层。其中数据采集层主要使用网络爬虫获取微博官方媒体“人民日报”下有关“新冠肺炎疫情通报”的微博及评论信息，数据预处理层包含数据清洗、中文分词、去停用词等。舆情分析层分为三部分，包括舆情主题演化挖掘分析、情感倾向时序变化分析与空间舆情网络社团发现分析。结果验证分析层，结合事件本身发展历程对舆情主题演化、情感时序变化及网路舆情空间分布等进行结果验证与解读。

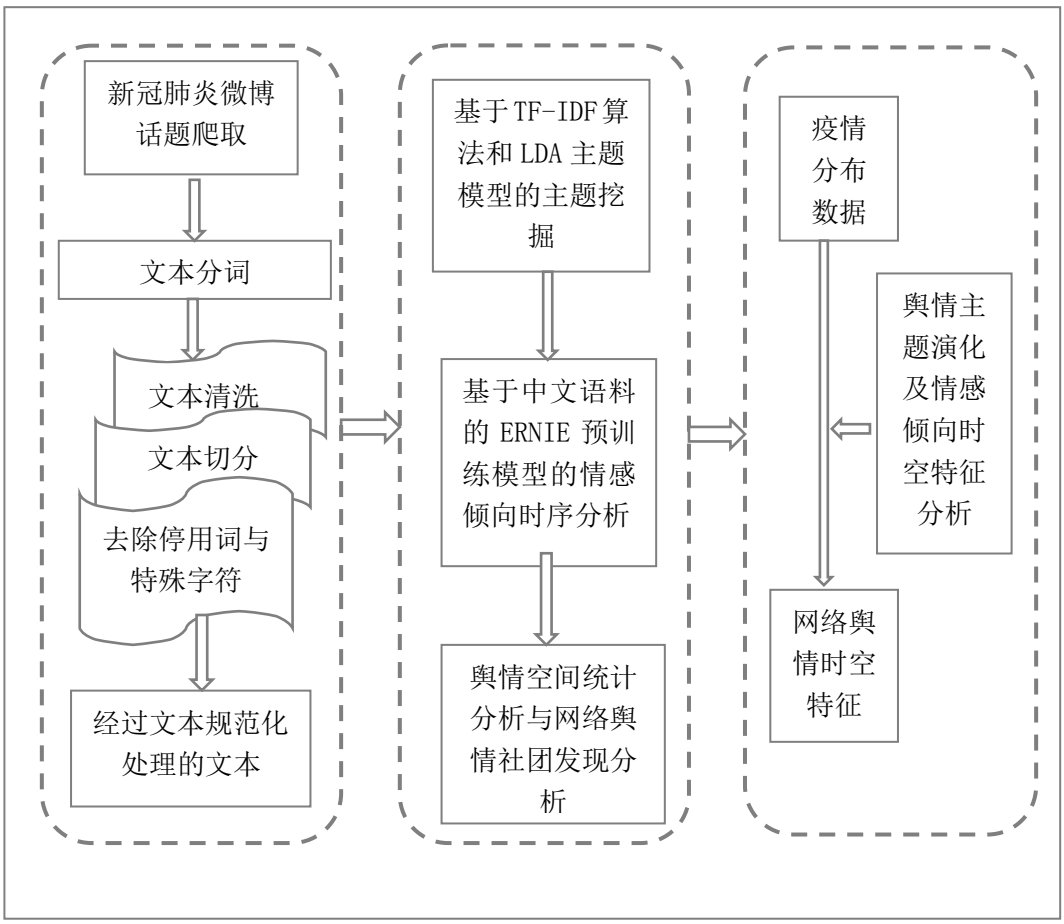


图 1.3 研究技术路线图

1.4 创新点

本文的创新之处有以下几个方面：

（1）引入时间与空间两个因素对网络舆情进行全面分析，归纳出一套较为完备的用于分析突发公共卫生事件网络舆情演化规律的方法体系。从时间层面上分析突发公共卫生事件网络舆情传播主题演化与情感时序变化；从空间层面上挖掘微博用户的地理位置和蕴含在文本信息中的地理位置，深层次挖掘不同区域公众对舆情事件的关注程度与情感倾向。有助于政府及相关部门有效识别网络舆情演化周期，掌握情感倾向时空特征，分区域把握舆情态势，差异化制定防护措施。

（2）以典型的突发公共卫生事件“新冠肺炎疫情”为案例，对突发公共事件网络舆情演化规律进行探索分析，归纳了新冠肺炎疫情网络舆情演化过程中的规律和差异，使得研究成果具备普遍意义，拓宽了本领域的研究视野和素材范围。

2 网络舆情演化规律及情感时空特征研究理论

这一部分是相关技术理论的介绍，主要构建研究突发公共卫生事件网络舆情演化及情感时空特征的理论框架，在时间维度层面，主要涉及 TF-ID 算法模型、LDA 主题模型、文本情感分析；在空间维度层面，介绍相关的空间统计分析方法与舆情社区挖掘算法。

2.1 TF—IDF 方法

TF-IDF（Term Frequency-Iinverse Document Frequency）是一种在文本挖掘过程中广泛使用的确定文本特征词的方法，用来衡量词的重要性。它由 IDF—逆文档频率，TF—词频两个部分组成。

IDF 认为，只有在少数文本中出现的词汇，才更有可能代表这些文本中存在的独特信息。当存在由 N 个文本组成的文档集时，文档集中包含某词汇 t 的文本数量 n(t)越少，则认为该词汇的 idf 值越大。因为该词汇区分度由 idf 值代表，当词汇的区分度较大时，这个词汇成为主题词的可能性也较大。

$$idf\left(t\right)=\log\frac{N}{n(t)+1}$$

但是，在整个文档集中也有部分词汇多次重复出现，虽然其 idf 值较小，但该词汇由于出现频率较高，也应该作为主题词进行分析，因此需要引入 tf 值，TF 认为，在文本中出现频率较高的词汇更有可能代表该文本存在的主要信息，tf 值的含义为特定词汇在特定文本中出现的频率大小，tf 值越大则说明该词汇对该文本的代表性越好。经过 idf 值和 tf 值作为双重标准确定的词汇即可作为主题词，用于区分不同的文本，因此将 idf 值与 tf 值相乘所得到的值作为判断一个词汇是否能作为主题词的标准。

$$f\left(t\right)=tf\left(t\right)\times idf\left(t\right)=tf\left(t\right)\times\log\frac{N}{n(t)+1}$$

2.2 LDA 主题模型

LDA 主题 （Latent Dirichlet Allocation）模型由 D.M. Blei 等人在 2003 年

提出提出，是基于概率图的 3 层架构贝叶斯模型，包括文档集层、主题层及特征词层，实质就是利用文本的特征词的共现特征挖掘文本的 topic，层次较清晰[2]。其结构图如图 2.1 所示。该模型认为不仅同一词可以在不同主题出现，而且生成主题的分布不唯一，均服从多项分布，每个词依据一定的概率选择某一主题，主题以一定的概率选择某词[19]。

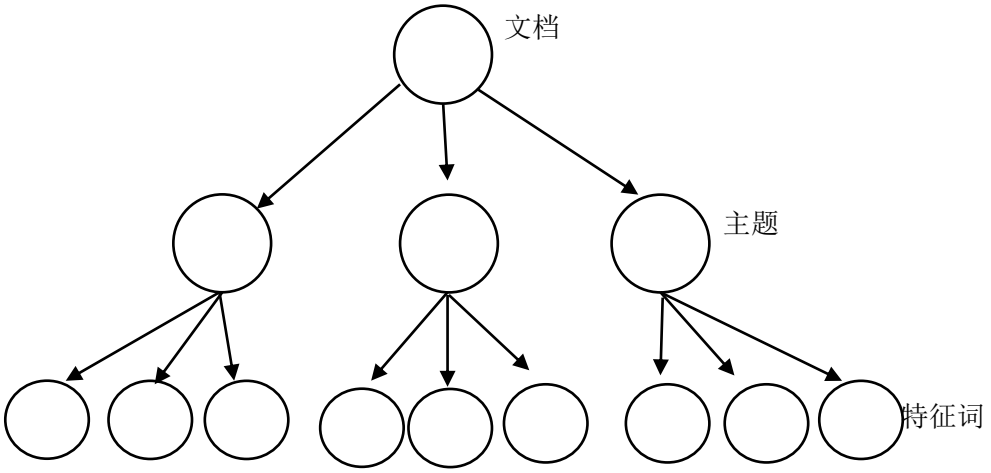


图 2.1 LDA 主题模型结构图

LDA 模型的生成过程为：

（1）抽取主题比例 $\theta \sim \mathrm{dir}(\alpha)$

（2）对于文档集中中的每个单词 $w_n, n \in \{1,2,3,...,N\}$

（3）计算主题分布 $p(z_n \mid \theta) \sim Mult(\theta)$

（4）计算单词分布 $p(w_n \mid z_n) \sim Mult(\pi_{z_n})$

LDA 的图模型如图 2.2 所示，其中的参数说明见表 2.1 所示。

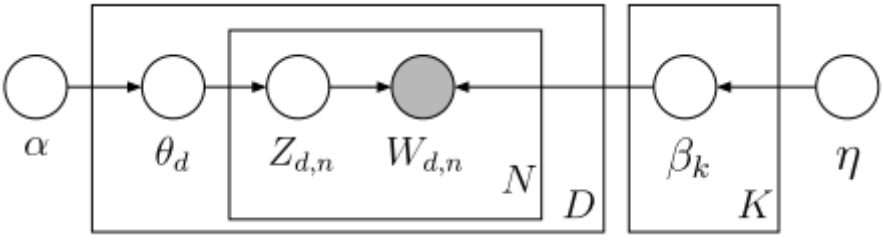


图 2.2　LDA 模型的结构示意图

表 2.1　LDA 模型参数说明

参数名称	参数说明
α	文档中主题分布的 Dirichlet 先验参数
β	狄利克雷先验参数，v 维向量
V	词袋的长度
N	文档中的词语数量
K	主题数量
D	文档数量
θ_d	文档 d 的主题概率分布
β_k	主题 k 词分布，v 维向量
$Z_{d,n}$	模型最终生成的词
$W_{d,n}$	对于$W_{d,n}$ 的主题分布

由于本文爬取的样本数据为微博评论数据，均为短文本数据。经过多位学者验证，LDA 模型对于微博这类短文本的主题识别有着良好的效果。下面则是 LDA 主题模型的短文本数据主题提取流程，大体可以分为四部分，流程图见图 2.3 所示：

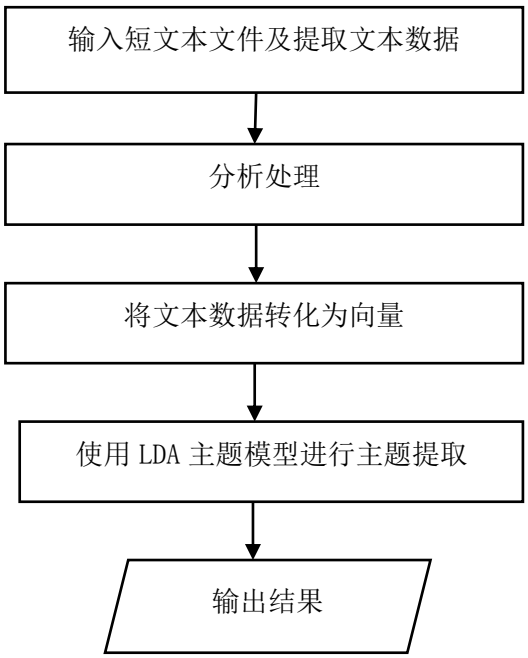


图 2.3 基于 LDA 模型的短文本主题提取流程图

2.3 情感分析理论

网络舆情研究领域常见的情感分析方法主要有基于情感词典方法、机器学习方法及深度学习方法。

目前，大多数情感词典物是用手工标记记构建的，需要消耗大量的人力和物力资源，并且充满主观因素，缺乏明确的判别标准，此外，还有三种自动构造情感词典的主要形式，分别为：基于语义字典的方法、基于语料库的方法及这两种方法的组合。基于语义字典的方法，主要是使用现有的情感标记词汇去计算类似词的相似性，从而给定词的情感偏向。例如，通过计算特定词汇与正负面词汇之间的相似度，哪种相似度大，就表现为那种。雷蒙德等人分析汉语单词和短语的倾向，并且其准确性和召回率大大提高[31]。但是这一方法存在较多的局限性，比如需要提前构建合适的情感词典，并且要考虑上下文本中程度副词对情感词的影响，还有整个上下文语义造成的情感词的含义偏离等，在现有的一些著名开源项目中被证实实证效果并不理想。

机器学习方法则需要大量的标注集进行训练，然而标注的过程不仅耗时长，而且限制分类数据的平衡性，且该方法情感分类极性主要为政、负两极分类研究，但是突发公共卫生事件期间所映射出的情感是更加复杂的[57]。此外，由于本文数据集涉及到事实是最近发生的，所以暂时没有公开标注的数据集供训练，因此，本文不采用这一方法。

基于神经网络的深度学习模型往往能克服上述两种方法的缺点，并且在一些著名的开源项目中被证实效果远胜于传统方法。举例来说，表 2.2 为百度的 Senta 模型在各□公开数据集上的表现。

但是深度学习模型也存在弊端，其往往需要大量的训练时间以及大量的训练语料。基于种种考量，在比较了诸多开源模型之后，最适合的方法是使用百度开源的基于中文语料的 ERNIE 预训练模型来直接预测情感倾向。这是目前中文预训练模型中效果最好的一个。百度的 senta 模型的本质是 ERNIE，ERNIE 模型是基于 BERT 改进的，该模型的优点大致有两个方面：首先具有更强通用性和可扩展性，BERT 模型，其 masking 策略是基于字的，在训练时学习到的更多是字与字之间的关系，例如[蒙]在[内]和[古]之间的局部关系，在中文语言模型其的策略

不利于知识信息的学习，但是 ERNIE 模型增强了模型语义表示能力，通过学习词与实体的表达，使得模型除了能学到上面的关系之外，还能学习到[内蒙古]与[呼和浩特]、[省会]等之间的关系，而这就是所谓的知识信息；其次 ERNIE 模型引入了更多优质语料[11]。需要注意的是，该 ERNIE 模型采用的仍是基于字特征的输入建模，只不过 mask 的粒度大小有所变化。

表 2.2 实验效果对比

任务	数据集合	语言	指标	原 SOTA	SKEP
句子级情感分类	ChnSentiCorp	中文	ACC	95.8	96.5
	NLPCC2014-SC	中文	ACC	78.72	83.53
	SST-2	英文	ACC	97.5	97.6
	Amazon-2	英文	ACC	97.37	97.61
评价对象级的情感分类	Sem-L	英文	ACC	81.35	81.62
	Sem-R	英文	ACC	87.89	88.36
	AI-challenge	中文	F1	72.87	72.9
	SE-ABSA16_PHNS	中文	ACC	79.58	82.91
	SE-ABSA16_CAME	中文	ACC	87.11	90.06
观点抽取	COTE_BD	中文	F1	82.17	84.5
	COTE_MFW	中文	F1	86.18	87.9
	COTE_DP	中文	F1	84.33	86.3
	MPQA-H	英文	b-F1/p-F1	83.67/77.12	86.32/81.11
	MPQA-T	英文	b-F1/p-F1	81.59/73.16	83.67/77.53

2.4 网络社团模型

社区的定义因研究领域的不同而定义方法不同，但界定社区的标准唯一。直观来看，社团指网络中某些密集群体，社区内部结点之间联系紧密，社区与社区之间的连接相对稀疏，如图 2.4 所示。

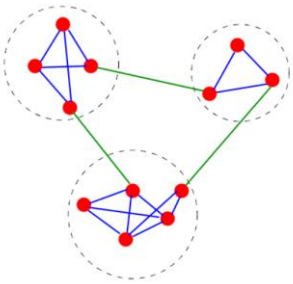


图 2.4 蛋白质网络社区结构

社区划分结果是否合理常见的评价指标有：聚类系数、强/弱社区、模块度[34]。综合比较，本文选择使用模块度来评价社区划分结果，其本质是划分出的子网络结构与该网络中节点随机构成的网络结构差异越大越好[36]，定义如下式所示:

$$Q=\frac{1}{2}\sum_{i,j}[A_{ij}-\frac{k_i k_j}{2m}]\delta(C_i,C_j)$$

其中，k_i 和 k_j 分别表示节点 i 和 j 的度（节点的弧维条数加弧头条数），C_i 表示某节点 i 所在的社区，m 表示复杂网络包含的总边数，A_{ij} 表示节点 i 和节点 j 之间边的权值，$\delta(C_i,C_j)$ 判断节点 i 和节点 j 是否属于同一个社区，若 i 和 j 属于同一个社区结构，当 $C_i=C_j$，$\delta(C_i,C_j)=1$，时，节点 i 和节点 j 属于同一个社区结构，否则，$\delta(C_i,C_j)=0$。模块度的值一般在 0 到 1 之间，当越靠近 1 时，表示社区结构质量越好，通常，当 Q=0.3 时，网络会出现明显的社区结构。

社区可称为聚类等，社团结构是复杂网络的一个重要拓扑结构特征。社区发现算法主要有图分割理论、GN 算法、Newman 快速算法和 hLouvain 算法等。衡量比较后，由于 Louvain 算法效率很高，其是一种实现最大化模块度社区发现算法，该算法可以划分为两个阶段，反复迭代计算。Louvain 算法具体流程如下:

（1）将每个节点看成一个独立的社区，初始社区数目与省份数目相同（34）；

（2）遍历任意一个节点 i，尝试将其加入邻居节点 j，同时计算模块度增量 ΔQ 的变化，若 $\max\Delta Q<0$，则将其放回原社区，否则将其加入到模块度增量最大的社区；

（3）重复步骤（2），直到模块度达到局部最大值，所有节点所处社区不再变化；

（4）上述步骤可以得到一个新的社区划分，将每一个社区对应的节点缩成超节点，超节点带包一个社区中若干节点。对于边，将一条连接同一社区的边变成超节点的自环边；将连接不同社区的一条边连接两个超节点。接着计算全局模块度值；

（5）重复步骤（1）直至社团结构不再发生改变。

本文中舆情网络社团挖掘，实质上是一组由省份点位与有向共现省份链接组

成的集合，省份间联系紧密且存在社团化或群组化的结构。社团结构发现是指根据共现省份词链，将省份节点一个个划分到社团中的过程，社团内部节点存在某种特质。

3 数据采集与处理

3.1 数据采集

3.1.1 数据选取

国内重要的互联网社交媒体之一是新浪微博，其是热点话题聚积地之一，也是网络舆情酝酿爆发的源头之一，是网络舆情演化的重要平台[25]。2020 年 10 月 21 日，《2019-2020 年中国移动社交行业年度研究报告》发布，报告中显示，微博月活跃用户达 5.23 亿，日活跃用户达 2.29 亿[1]。此外，微博 2020 年第一季度财报数据中显示，新冠肺炎疫情期间，用户对资讯得需求让微博活跃用户创下历史新高，微博月活跃用户增 8500 万。由此可见，微博拥有庞大的用户体系，且在新冠肺炎疫情期间，微博是重要的信息中转枢纽，其产生的文本数据具有较强的代表性，其用户行为能够反映新冠肺炎疫情网络舆情传播周期演化。因为疫情期间微博信息量庞大，呈井喷式发布态势，微博信息发布者众多，包括政务媒体、自营媒体及个人等。经过比较分析，为了体现数据的代表性和可说明性，选择爬取政务媒体“人民日报”的法人官方微博下的有关新冠肺炎的微博数据，人民日报的微博数据显示，相比 2019 年同期，2020 年 2 月 9 日 0 时至 2 月 11 日 24 时，日均发博量增加 21.5 倍，平均点赞数增加 6.4 倍，平均评论数增加 1.1 倍。

2020 年 1 月 1 日，湖北省武汉市发布有关“不明原因肺炎”的通告，打响了疫情防疫战的第一枪。2020 年 4 月 9 日，历时 76 天“封城”的武汉市成功解除“封城”。在新冠肺炎疫情发展的过程中，微博中几乎被有关新冠肺炎疫情的话题所覆盖，网民参与讨论的活跃度相当高，产生了大量的文本数据，受限于文章篇幅，为实现本文网络舆情演化规律和情感时空特征研究，选择有代表性的政务媒体“人民日报”官方微博为爬取对象，疫情期间“人民日报”官方微博每日发布的疫情报道下产生了大量的评论内容。因此，本文选择 2020 年 1 月 1 日至 4 月 9 日之间 “人民日报”官方微博下有关全国新冠疫情通报的微博评论为爬取对象，关键词为“新冠疫情通报”，以每日为一个时间节点，共 100 个时间节点，每条微博按照评论热度高低爬取前 1000 条评论文本。每条微博爬取的字段包括微博发布时间、微博内容 ID、微博内容、评论数、微博转发数和点赞数等。每

条微博评论爬取的字段包括评论者 ID、评论内容、评论者性别、评论者地址和评论日期等。

3.1.2 基于网络爬虫的微博数据爬取

目前获取微博数据的方式主要有三种，分别是基于 API 的数据获取、基于网络爬虫的数据获取及各种开源爬虫工具。由于基于 API 的数据获取对个人爬取而言存在一定的局限性，开源爬虫工具获取数据量有限且比较死板。因此本文选择使用基于网络爬虫获取微博数据。利用 Python3.8 来爬取本文所需文本数据，具体获取方式如下所示：

步骤一，模拟账号登录。利用 cookie 浏览器实现账号登录。

步骤二，采集微博数据和用户基本信息。首先要从微博搜索页面内搜索出人民日报官方微博，进入人民日报官方微博中，进入微博高级搜索页面，搜索关键词，设置爬取时间范围，筛选出满足条件的数据。

图 3.1　微博高级搜索页面

步骤三，分析设置的关键词和时间范围与高级搜索页面的 URL 之间的关系，构造页面 URL 进行页面解析，抓取页面内容。

步骤四，解析上述步骤获得的页面内容，使用 Xpath 语法获取微博内容 ID、微博内容，转发数、评论数及点赞数等信息，以 csv 格式存储到本地。

步骤四，实现微博评论数据及评论用户信息数据等信息的获取，构造页面 URL 实现数据获取。

步骤五，将获取的微博数据和其对应的评论数据存储到本地数据库中，主要有两张表，一张是存储微博内容数据，转发数、评论数及点赞数等数据，另一张表存储评论文本数据、评论用户地理位置，性别，昵称等信息。第二张表中每一条评论数据都对应着第一张表中微博内容 ID 字段，也就是每一天的微博下的评论与该条微博对应。

新冠肺炎疫情事件微博数据爬取的工作流程图如图 3.2 所示。

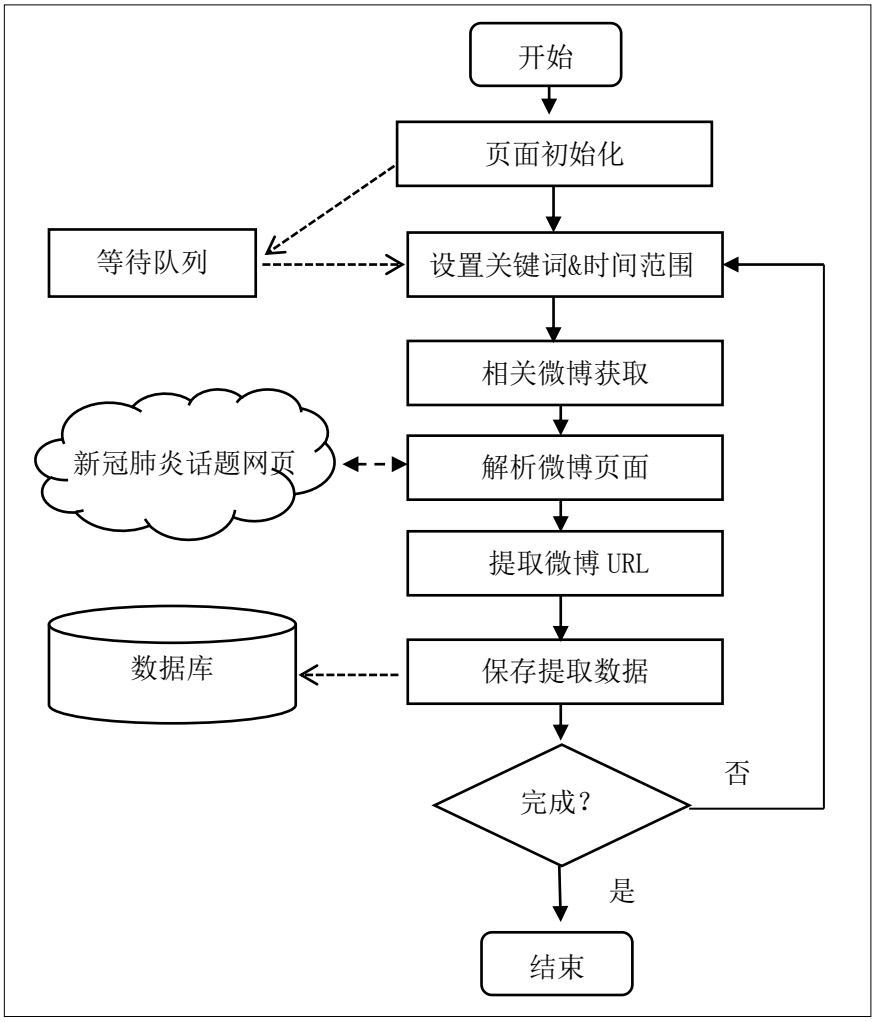


图 3.2　微博数据网络爬虫工作流程

3.1.3 数据采集结果

基于网络爬虫数据采集方法对新冠肺炎微博舆情数据进行采集，以“新冠肺炎疫情通报”为关键词，时间范围为 2020 年 1 月 1 日到 2020 年 4 月 9 日，采集的第一张表的字段包括微博发布时间、微博内容 ID、微博内容、评论数、微博

转发数和点赞数，采集了 100 天的微博数据，第二张表的字段包括微博内容 ID，评论博主 ID、评论内容、评论时间、评论博主性别和所在城市，总共采集了 95437 条数据，采集的样本数据如图 3.3 和图 3.4 所示。

微博内容ID	日期	微博内容	转发数	评论数	点赞数
4464171633076141	1月24日	#全国确诊新型肺炎病例#【#29省累计新冠肺炎确诊病例830例#】1月23日0-24时，27个省（区、市）报告新增确诊病例259例，新增死亡患者8例。新增治愈出院6例。19个省（区、市）报告新增疑似病例680例。	2967	10010	188333
4464534334753932	1月25日	#全国确诊新型肺炎病例#【#全国新增444例新型肺炎确诊病例# 累计确诊1287例】国家卫健委通报，1月24日0-24时，29个省（区、市）报告新增确诊病例444例，新增死亡患者16例。新增治愈出院3例。19个省（区、市）报告新增疑似病例1118例。全	15258	74833	1839022

图 3.3　“新冠肺炎”微博内容数据格式

微博内容ID	评论博主ID	评论内容	评论时间	性别	城市
4464255397344620	6243761175	我希望武汉好一点，湖北好一点，中国好一点。	1月23日	女	河南
4468307178385440	6276537427	钟老先生和所有医护人员都太辛苦了，也一定要保护好自己！	2月18日	男	北京
4488634033755310	2004130393	中国尽全力守好国门吧。严防境外输入病例引发疫情反弹，将是对中国的严峻考验	3月9日	男	安徽
4482830899823490	5839604250	严防境外输入！不能让之前努力白费！	3月17日	女	辽宁
4491033490706180	6795862063	必须加强外籍人员管理!	4月7日	男	湖北

图 3.4“新冠肺炎”微博评论数据格式

3.2 数据预处理

网络爬虫采集到的数据格式并不是理想的，存在不完整、缺失及冗余等问题，如果直接用爬取到的数据进行文本分析，会使模型效果不佳。因此，在进行文本分析之前应该进行数据预处理，对数据进行清洗、整理，使数据更整体化一，提高分析结果的准确性。解决数据确实无效等问题的基础上，通过文本清洗、中文分词、删除停用词与特殊字符等一系列操作，将上述爬取到的微博评论转化为可进行文本分析的格式。操作过程均在 Python 中实现，下面将数据预处理过程进行展示。

3.2.1 数据清洗

数据清洗是为了解决数据缺失、重复和错误数据等一系列问题，使数据具有较高的真实性和有效性，避免脏数据影响模型结果。为了更好的清洗数据，首先引入必要的第三方库并导入数据，具体代码如表 3.1 所示。

表 3.1 数据清洗代码内容

```
import pandas as pd
import numpy as np
import re
import jieba
from senta import Senta
from wordcloud import WordCloud
from matplotlib import pyplot as plt
from pyecharts import Map
import seaborn as sns
from province_city import province_city
plt.rcParams['font.sans-serif']=['Hiragino Sans GB']
plt.rcParams['axes.unicode_minus']=False
raw_data = pd.read_excel('/Users/J1739/Desktop/毕业论文/微博数据分析/微博评论.xlsx')
```

根据计算发现评论博主 ID 与评论博主昵称分别有 81862 和 81868 条，即有 6 个博主在数据采集期间更换过一次昵称，并且平均每个博主在收集的数据集内评论 1.2 条。基于此可以认为每个博主在数据采集期间都只发了一条评论，因此，无需考虑由于一人多评而导致情感分析的偏差。而对于对于时间，只需要具体到哪一天即可；对于地点，也只需要精确到省份即可，具体代码如表 3.2 所示。

表 3.2 数据清洗代码内容

```
def transform_date(s):
    nums = re.compile(r'\d+').findall(s.split(' ')[0])
    date = ''
    date += nums[0] if len(nums[0]) == 2 else '0' + nums[0]
    date += nums[1] if len(nums[1]) == 2 else '0' + nums[1]
    return date

data.评论时间 = raw_data.评论时间.apply(transform_date)
  ##城市只需要关注省份即可
data.博主城市 = raw_data.博主城市.apply(lambda x : str(x).split(' ')[0])
```

经过以上一系列操作，本文得到了初步数据集。接下来主要对评论文本数据进行规范化处理。

3.2.2 文本规范化处理

（1）文本清洗

微博评论信息是本文进行文本分析的重点，但是微博评论中包含了部分无关以及不必要的标识和字符，需要对其进行删除。通过观察文本内容，发现有很多评论里面包含了类似 xa0，u200b 等字符，这些是由于爬取的时候编码问题导致的。其次，很多评论文本内容开头是‘回复’、‘的表态’这样的词语。因此需要将上述提到的字符删除。进一步观察筛选出的数据集，发现评论文本中存在英文字母缩写的现象。因此，统计出所有连续英文词汇出现的次数，逐一观察，把频率高的一部分替换成相应的中文字符，具体代码如表 3.3 所示。

表 3.3 文本清洗代码内容

```
##统计所有出现的英文词汇
eng = []
for i in data.评论内容:
    eng += [x for x in re.compile('[a-zA-Z]*').findall(i) if x!='']

##查看出现频次前 20 的字母组合并适当替换成中文词汇
pd.value_counts(eng)[:20]
```

（2）中文文本分词

在微博评论中，词是可以反映情感状态最小的独立的文本单元。因此，在进行中文分词之前首先需要分词，将词切分出来作为特征词。英文分词相对于中文分词要简单许多，中文词语之间没有明确的标识符号，并且一词多义，在不同的语境中意义不同。现如今，中文分词工具层出不穷，如 jieba、 LTP、 Stanford 汉语分词和 SnowNLP 等[40]。Jieba 是基于 Python 第三方词库的中文分词，是目前分词效果较好且使用最为广泛的工具之一，Jieba 分词基于前缀词典进行高效词图扫描，生成句子中所有可能分词结果的有向无环图（DAG），采用动态规则的方法查找出最大概率的路径，从而找出基于词频的最大切分组合，对于未登录词，使用基于汉字成词的 HMM 模型（隐马尔科夫模型），采用 Viterbi 算法[43]。。

基于 Jieba 分词可支持多种分词模型（全模式、精确模式和搜索引擎模式），并且支持自定义词典，对 Python2 和 Python3 均兼容，支持多种编程语言（Java、C++、Rust、PHP、R、Node.js 等）。

（3）删除停用词与特殊字符

删除停用词与特殊字符也是文本规范化处理的重要一部分，停用词是指文本最常用的词，但可能对表达文本意义及情感没有太多价值和含义。例如文本中的“的”属于结构助词，通常用来表达两者之间的关联，本身没有实际意义，因此需要在文本分析前将其删除，保留更能体现主题和情感的词语[45]。目前还没有普遍或已穷尽的停用词表。因此，本文将 github 上包括哈工大停用词、百度停用词、四川大学机器智能实验室停用词、中文停用以及最全中文停用词表进行合并、去重，得到新的中文停用词表，总共 2462 个。在文本分析中，特殊符号没有有用的价值，需要将特殊字符加入停用词表，使用停用词表对分词后的数据进行过滤，得到删除停用词与特殊字符之后的分词结果。

4 网络舆情传播主题演化与情感时序变化分析

发布、转发、评论和点赞等是微博用户主要的行为表现。当前，微博用户行为逐渐用于研究突发公共卫生事件背景下的网络舆情中，安璐等人揭示了突发公共卫生事件下的微博话题在各阶段与微博用户之间存在正相关关系[15]；孟吉杰以关于“H7N9”事件的政务媒体微博为研究对象，研究结果显示，微博用户在传播事件紧张与预防知识等信息方面表现很积极，参与意识很强[41]。由此可见，通过分析微博话题下用户的转发、评论和点赞等行为，可以充分说明舆情演化规律。因此本文主要分析转发、评论和点赞这三种微博用户行为之间的相关性，分析这三者对与网络舆情演化的驱动作用，进而选择其中一个指标依托生命周期理论，对新冠肺炎疫情舆情演化周期划分。其次进行新冠肺炎疫情舆情传播周期各阶段主题挖掘与主题演化分析，结合传播周期各阶段情感倾向变化对舆情生命周期中不同时期窗口下研究新冠肺炎疫情舆情传播过程中共转变原因进行剖析，即新冠肺炎疫情舆情传播变化的原因分析。

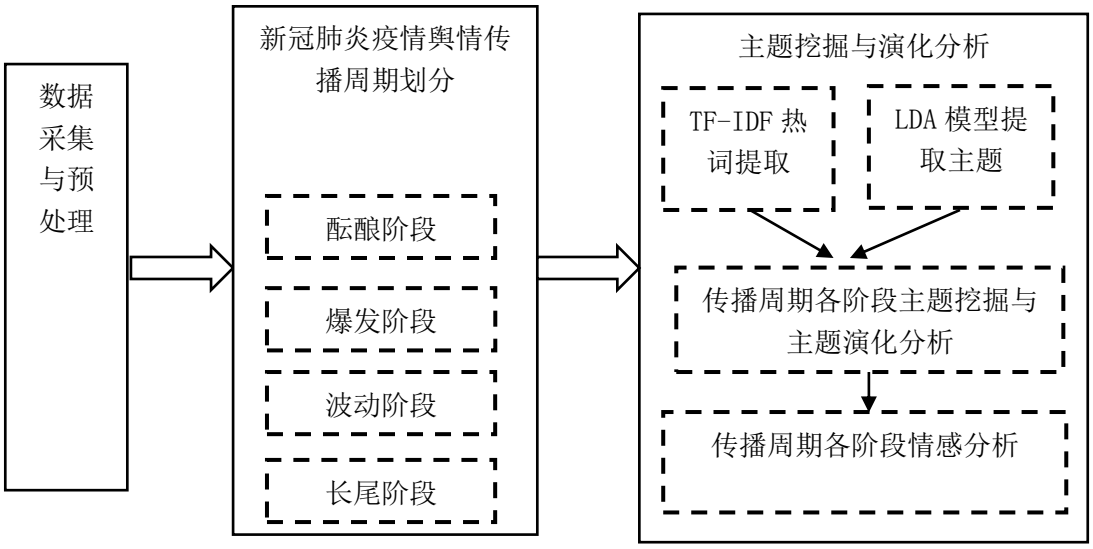


图 4.1 新冠肺炎疫情网络舆情演化分析框架

4.1 舆情传播阶段划分

为了剖析新冠肺炎疫情网络舆情的发展趋势与传播周期演化过程，本文依托上述有关网络舆情演化周期划分的理论基础与新冠肺炎疫情的发展历程，通过分析微博博文的转发、评论和点赞数之间的相关性，进而从其中选用一个变量对舆

情演化周期进行划分。

4.1.1 转发量、评论数及点赞数的相关性分析

转发、评论和点赞等用户行为代表用户对微博话题的整体态度，是促使网络舆情发展的基础。转发加快了微博信息的传播速度，是用户自主将感兴趣或者有价值的微博内容附上自己的见解进行转发，进而扩散了该条微博内容，因此，转发数可以测定微博内容被传播的频率和范围大小。评论是用户可以在别人所发布的微博内容下方，表达自己的看法和意见，能够反映用户对该微博内容所表达的情感倾向和态度，同时能够展示用户的参与程度，因此，评论文本中含有丰富的研究价值。点赞分为两种情况，一种是亲友之间的互动，一种是表达用户对博主发布的微博内容价值认可，表示喜爱，后者是本文所研究的。上述三种行为在一定程度上能够推动网络舆情的发展，尤其转发数更能说明网络舆情的传播过程。通过分析爬取的微博数据中转发数、评论数与点赞数之间的相关关系，判断“新冠肺炎疫情事件”中这三种用户行为两两之间是否存在相关关系，进而决定是否选用其中的一个指标来分析新冠肺炎微博舆情传播的变化趋势。使用 SPSS21.0 软件计算，结果如表 4.1 所示。

表 4.1 三项指标的相关分析

		转发量	评论数	点赞数
转发量	Pearson 相关性	1	.797	.714
	显著性（双侧）		.000	.000
	N	121	121	121
评论数	Pearson 相关性	.797	1	.893
	显著性（双侧）	.000		.000
	N	121	121	121
点赞数	Pearson 相关性	.714	.893	1
	显著性（双侧）	.000	.000	
	N	121	121	121

借助 Person 简单相关系数分析三种用户行为之间的相关性，具体分析情况有三种：当两两之间的相关系数大于零时，代表某两种用户行为之间存在正线性相关关系；当相关系数小于零时，表示某两种用户行为为负线性相关关系；当相

关系数等于零时，表示某两种用户行为之间无线性相关关系。从表 3.1 中可以清晰地看出，转发数、评论数与点赞数这三者之间的相关系数都大于 0.7，点赞数与评论数之间的相关关系高达 0.9，说明这三种用户行为之间存在很强的线性相关关系，三者对网络舆情的推动作用趋同，而转发数更能说明网络舆情的传播频率和范围。因此，接下来本文只选用转发数作为舆情传播阶段划分的主要变量。

4.1.2 新冠肺炎疫情网络舆情传播阶段划分

通过分析从 1 月 1 日到 4 月 9 日之间有关“新冠肺炎”或“肺炎”的微博转发量，由于疫情期间转发量、评论量和点赞数变化幅度较大，因此为了便于观察，图 4.2 中纵轴采用对数坐标轴，然后绘制趋势图，如图 4.2 所示：

根据图 4.2 中数据收集期间微博转发量的变化情况，可以发现目标话题的传播趋势为峰值分布无规律的多峰特征，且波动较剧烈。为了准确表示舆情发展规律，结合舆情发展反复性的特点与突发公共卫生事件舆情传播的生命周期理论，将舆情阶段划分为酝酿阶段(2020 年 1 月 1 至 2020 年 1 月 19 日)、爆发阶段(2020 年 1 月 20 日至 2020 年 1 月 31 日)、波动阶段(2020 年 2 月 1 至 2020 年 2 月 29 日)和长尾消散阶段(2020 年 3 月 1 日至 2020 年 4 月 9 日)，这个阶段为关于国内疫情的舆情基本处于稳定阶段，但是由于境外疫情逐渐加重与境外输入逐渐增多。因此，又引发了关于境外新冠肺炎疫情的新一轮舆情爆发期。

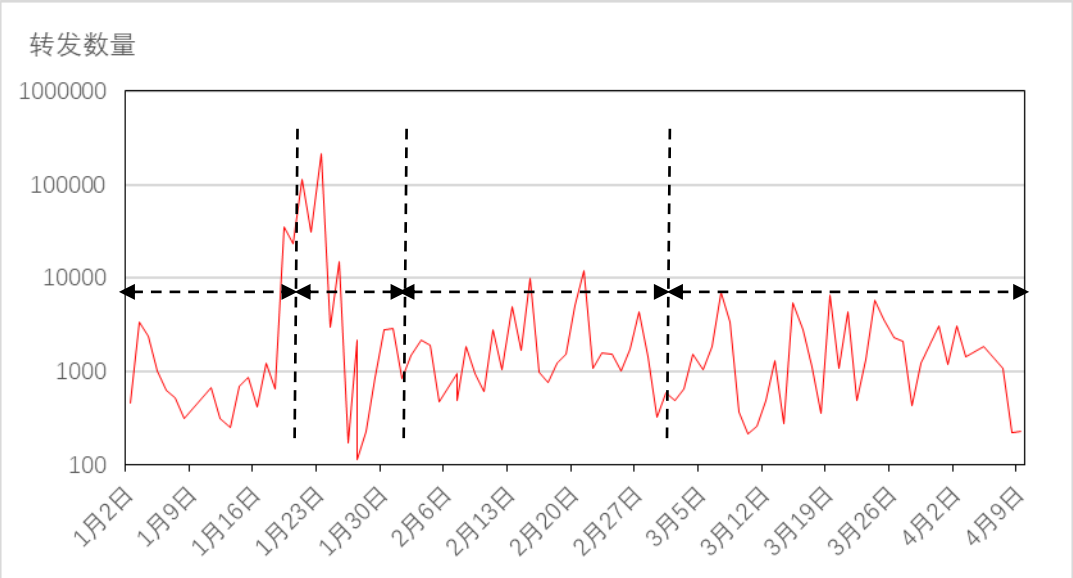


图 4.2 新冠肺炎网络舆情播的变化趋势

为了剖析新冠肺炎疫情网络舆情传播周期内出现转折点的内部原因，下文结合疫情事件本身做一些细致分析。结合新冠肺炎事件本身进行分析（见图 4.3），将此次重大突发公共卫生事件舆情传播阶段划分为四个阶段，具体分析如下：

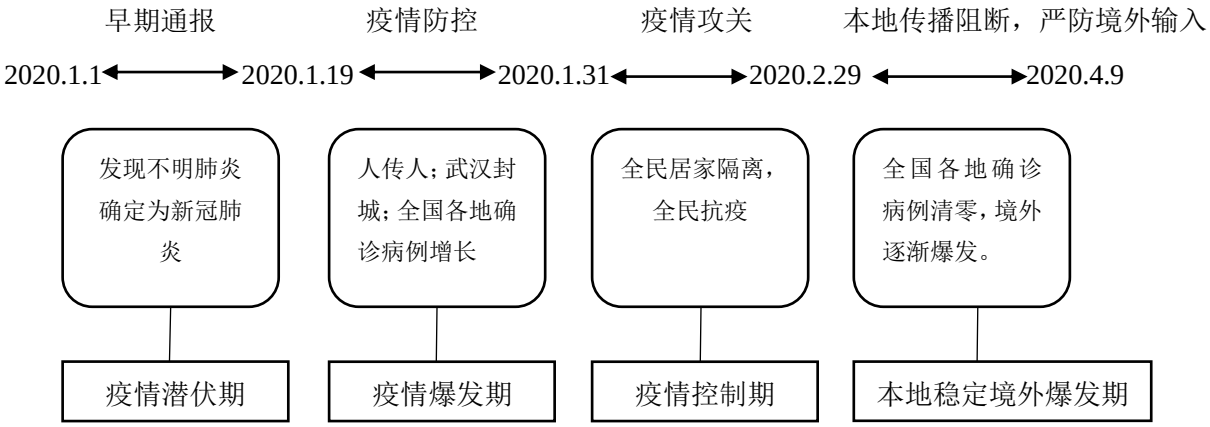


图 4.3 新冠肺炎疫情事件发展时间主线

（1）酝酿阶段。酝酿阶段是指突发公共卫生事件发生之后，此时的信源渠道多种多样，信息内容五花八门，大部分信息真实性有待商榷，官方媒体报道不确定性强，大部分网民没有意识到事件的严重性，只能猜测评论，经过网民用户的转发、评论及点赞，事件在小范围内形成部分舆情，但传播内容模糊、渠道狭窄且可信度低。2019 年 12 月 30 日，武汉市卫健委发布关于做好不明原因肺炎防空的紧急通知，并且回应了不明肺炎的有关情况；2020 年 1 月 1 日到 2020 年 1 月 19 日之间，国家卫健委宣布武汉市不明肺炎为新型冠状病毒肺炎。在此阶段，微博仅有小部分群体参与讨论，并没有引起大多数网民的关注与热议。因此本文将 2020 年 1 月 1 日至 2020 年 1 月 19 日划分为疫情舆情传播酝酿阶段。

（2）爆发阶段。爆发阶段是指关于事件的关注和热议铺天盖地，包括微博平台在内的其他多方媒体及传统媒体及都参与共同宣传报道，最后演变成全民参与讨论的舆情爆发阶段。舆情的爆发往往需要短时间内能够引发网民讨论的积极性或者刺激情绪的导火索。在本次网络舆情事件中，1 月 20 日，钟南山院士确定了新冠肺炎存在“人传人”现象，瞬间引发全民热议，带着疑惑和探究的心理开始对突如其来的疾病展开讨论，这期间会产生大量的谣言和引导性强的言论，1 月 23 日武汉市全面封城，全国各地相继启动一级响应政策，全国各级政府及民众才逐渐意识到新冠肺炎疫情的严重性，针对新冠肺炎疫情的狙击战开始全面打

响。因此，本文将 1 月 20 日至 1 月 31 日划分为舆情爆发阶段。

（3）波动阶段。波动阶段是指舆情爆发之后，民众接受事实后舆情呈现下降趋势，但是期间会出现小幅度波动舆情，有出现反弹的迹象，这个阶段并非每种事件都能引起，往往出现在既复杂又周期长的突发公共卫生事件中，如果事件的发展过程呈现不够明晰，且对公众的舆论置之不理，不但不能推动事件的进展，而且容易引发“二次舆情”。在本次网络舆情事件中，1 月 31 日到 2 月 29 日左右，全国开始全面控制疫情，形成全面居家隔离，全民抗疫的局势，通过相关部门出台的应急措施及医疗手段的提升，疫情由不可控到可控可防，但是期间全国各地确诊病例时有新增，使得舆情热度反复升降。因此，本文将 2 月 1 日至 2 月 29 日的疫情舆情传播阶段划分为波动阶段。

（4）长尾阶段。长尾阶段是指事件得到较为妥善处理，公众讨论热度开始消退的阶段。在新冠肺炎舆情事件中，由于涉及范围广，传播速度快，随着我国疫情的稳定，境外疫情却逐渐加重，境外输入病例逐渐增多，因此该事件的舆情并未消退，而是出现了长尾现象。3 月 1 日到 4 月 9 日之间，国内本土疫情基本阻断，确诊病例逐渐治愈，全国各地病例清零，各地区疫情控制均呈明显向好趋势，但是境外疫情又逐渐爆发，境外输入病例逐渐增多，对国内疫情防控又开始严峻的挑战。因此，3 月 1 日至 4 月 9 日处于疫情舆情传播周期的长尾阶段。

结合新冠肺炎疫情事件发展时间主线和其引起的网络舆情传播变化趋势分析可以看出，有关事件的微博传播变化趋势与疫情事件本身的发展主线具有密切的关系，因此，针对一件正在发展雏形阶段的事件，及时分析其微博传播状况，可以预见性的进行积极引导，避免事件持续发酵。

4.2 舆情传播主题演化分析

为了挖掘出新冠肺炎疫情网络舆情传播周期中各阶段的主题，勾勒舆情事件主题演化规律和情感倾向的时序发展趋势，为舆情决策与分析提供科学依据。上述结合生命周期理论已将新冠肺炎疫情的网络舆情划分为四个阶段，接下来结合 TF-IDF 模型、LDA 主题模型及情感分析模型，将时间维度融入微博文本分析中，主要进行了包括时序主题挖掘，挖掘新冠肺炎疫情事件发展过程中其舆情主题信息与演化规律，为政府及相关部门针对舆情事件实施有效监控、制定预警决策提

供必要的理论支撑。为了更好的识别网络舆情演化，进行舆情监控和情感预警，在此基础上引入情感分析，通过分析网络舆情传播各阶段中的情感变化，深层次的探究新冠肺炎疫情的网络舆情传播周期规律，把握公众的情感倾向。

首先基于 TF-IDF 算法对舆情传播周期各阶段的文档，提取能够表达文本主旨的关键词，得出传播周期各阶段的热词排序，并绘制词云图加以直观展示。其次采用 LDA 主题模型挖掘新冠肺炎疫情舆情的主题特征，并结合 TF-IDF 提取出的热词对主题下的特征词进行筛选，以辅助识别传播周期各阶段热点主题信息，最后揭示在生命周期各阶段微博舆情的热点主题分布及演化规律。

运用 TF-IDF 算法提取出文本数据中传播周期各阶段对应的热词，绘制出词云图，见图 4.4（a）-（b），并从其中选取各周期排名前十的词语，见表 4.2 所示。

（a）酝酿阶段

（b）爆发阶段

（c）波动阶段

（d）长尾阶段

图 4.4 传播周期各阶段热词词云图

表 4.2 TF-IDF 结果

起始阶段		爆发阶段		波动阶段		长尾阶段	
高频词	词频	高频词	词频	高频词	词频	高频词	词频
肺炎	1410	武汉	1465	疫情	2308	中国	3962
不明原因	1403	疫情	1167	确诊	1936	隔离	3895
武汉市	1322	确诊	1021	治愈	1803	境外	2587
病毒	1062	肺炎	1014	医疗队	1783	输入病例	2065

诊断	973	一级相应	938	隔离	1688	疫情	1860
感染	942	宣传	891	中国	1634	入境	1803
市场	821	封城	819	防控	1523	严防	1726
海鲜	889	口罩	792	肺炎	1400	物资	1703
通报	879	防控	725	医院	1184	检测	1675
查明	796	治愈	666	口罩	1181	治疗	1657

从词云图（图 4.4）及提取的热词（表 4.2）可以看出，在酝酿阶段，权重较大的热词有“肺炎”、“不明原因、“武汉市”、“病毒”、“诊断”、“感染“等词，可以看出该阶段网友主要关注的是关于新冠肺炎的致病原因及是否传染等问题；在爆发阶段，权重较大的热词有“武汉”、“疫情”、“确诊”、“肺炎”、“一级响应”、“封城”等，此时网民的关注点转向了疫情本身的发展情况及全国各地如何纷纷采取防疫措施；在波动阶段，权重较大的热词有“疫情”、“确诊”、“治愈”、“医疗队”、“隔离”、“防控”等，可见波动阶段的热词有一部分连续爆发阶段，此外出现了“治愈”、“医院”等词，可见网民对疫情的发展抱有较高的期望；在长尾阶段，除了承接上一阶段的热词之外，出现了“境外”、“输入病例”、“入境”等权重较大的词有，此时国内疫情基本稳定，但是境外输入逐渐增多，网民的关注点主要转向了境外输入话题。

可以发现，以上通过 TF-IDF 模型分析只能提取出传播周期各阶段的热词，反映出舆情事件各阶段有出现频率高的热词，观察发现，各阶段权重较高的词语在各阶段基本涵盖，但是不能概括出主题。而 LDA 主题模型抽象层级更高，能够更好地概括出传播周期各阶段的主题分布，并进行对比分析及主题演化分析。

采用 LDA 主题模型对预处理后的文本数据进行层次主题聚类，本文依据网络舆情传播周期划分出的阶段，根据采集数据时间形成各个阶段的文本集，接着将文本集引入 LDA 主题模型，先通过计算困惑度得到最优主题集合 k{topic1,topic2 ,…, topick}。本文选取了排序前 10 的主题词作为该主题的特征词，并对各个主题进行了概括，结果如表 4.3 所示。

表 4.3 新冠肺炎疫情舆情传播周期各阶段主题分布

阶段	编号	主题词	主题概括
起始	topic1	武汉、不明、肺炎、原因、病毒、措施、调查、通报、发现、海鲜	疫情通报及溯源病源
	topic2	发热、咳嗽、病毒、感冒、发烧、流感、呼吸、传染、症状、观察	新冠肺炎症状

阶段	topic3	治疗、治愈、患者、救治、出院、医学、医学、卫生、观察、药物	前期治疗措施
爆发阶段	topic1	武汉、确诊、病例、疫情、通报、新增、中国、湖北、新增、数字	疫情发展情况
	topic2	人传人、钟南山、传染、感染、病毒、肺炎、食物、调查、途径、野味	传染途径及方式
	topic3	封城、口罩、隔离、控制、接触、封路、防控、返程、聚集、实名制	全国防控措施
	topic4	红十字会、贪污、严惩、寒心、捐赠、物资、唐志红、政府、官僚、湖北	红十字会及官员不当行为
	topic5	医生、医护人员、一线、呼吁、措施、救治、职业、支援、救治、科研	对医护人员的关注
波动阶段	topic1	确诊、新增、数据、疑似、感染、病例、现存、检测、拐点、出院	疫情发展情况
	topic2	隔离、居家、抗疫、报备、控制、活动轨迹、管制、警惕、口罩、防范	疫情防控防范措施
	topic3	全球、新冠、疫情、韩国、意大利、爆发、日本、防控、救援、中国	新冠肺炎全球流行
	topic4	开学、上学、高三、毕业、复工、经济、延迟、上班、休学、恢复	复工复学意愿
	topic5	领导、好干部、严惩、党员、停职、责任、免职、隐瞒、失职、严查	官员不当行为
	topic6	李文亮、医生、医护人员、钟南山、英雄、天使、医护人员、一线、支援、致敬	对医护人员的肯定与支持
	topic7	隐瞒、造谣、辟谣、自私、谣言、拒绝、警惕、泛滥、严惩、添乱、	谣言情况
长尾阶段	topic1	入境、严防、输入、驱逐出境、风险、韩国、蔓延、封锁、美国、病毒	对境外输入的态度
	topic2	自费、隔离、治疗、检测、隐瞒、严惩、核酸、添乱、感染、控制	境外输入应该采取的防护措施
	topic3	社会主义、骄傲、大国担当、中国、共产党、制度，榜样、国家、遵守、春天	对战疫成果的肯定及对祖国的赞扬
	topic4	隔离、口罩、严防、居家、潜伏期、严查、机场、社区、观察、广东	国内持续防控措施
	topic5	数字、清零、新增、本土、医务人员、无症状、李文亮、致敬、英雄、清空	疫情发展情况
	topic6	上学、复工、孩子、工作、旅游、电影、快递、初三、开学、高三	复工复学意愿

由表 4.3 可以看出，新冠肺炎疫情背景下的网络舆情传播各个阶段的主题内容侧重点存在一定的差异，但是具有话题继承性和连续性。舆情酝酿阶段共得到三个主题，分别是武汉市不明原因肺炎的报道与致病原因、新冠肺炎症状及前期的治疗措施，是对舆情事件较为宏观且不够全面的探讨。在该阶段，相应的管理部门应该第一时间公开透明有关突发公共卫生事件的初步核实情况，动态发布事件进展，及时告知公众相关防护措施，快速并有针对性地回应公众疑问，避免舆情持续扩大，引发不必要的衍生舆情。

爆发阶段共得到五个主题，早期网民多关注新冠肺炎事件的消息报道及具体伤害性，随官方发布内容增多，话题逐渐演变为对疫情发展动态、新冠肺炎传播途径与方式、应急防控措施及对一线工作医护人员的关注，除此之外，红十字会及政府行为成为网民热议话题，网民指责红十字会物资使用不力及武汉政府治理

不力。爆发阶段含有大量的舆情信息，是网民对舆情事件态度和情绪的集中表达，该阶段相关部门的管控重点应在于舆情引导和控制，网络意见领袖在舆情演化和分裂中有重要的地位，因此，应以权威评论意见积极引导网民情绪，传达主流价值观，弘扬正能量。

波动阶段共得到七个主题，分别为疫情发展情况、疫情防控防范措施、新冠肺炎开始全球流行、复工复学意愿、官员不当行为整治、对医护人员的肯定与支持以及谣言散布情况。疫情发展情况、疫情防控防范措施、政府官员在疫情期间的不作为及对一线医护人员的赞扬与肯定为继承话题，是由上一阶段舆情持续发酵而形成的话题，说明舆情具有连续性与扩散性。此外，事件谈论内容进一步演变为网民所处情境的相关衍生话题，分别为复工复学的意愿、新冠肺炎全球大流行及谣言散布等内容。该阶段是舆情管控的关键时期，相关部门通过获取实时情报，对事件舆论演变进行动态分析，针对重点问题及时做出应对干预，及时回应公众诉求及告知应对措施，提高政府公信力，引导舆情导向持续向好发展。

长尾阶段共得到六个主题，分别为对境外输入的态度、境外输入应该采取的防护措施、对战疫成果的肯定及对祖国的赞扬、国内持续防控措施、疫情发展情况及复工复学意愿。随着国内疫情逐渐好转稳定，国内持续防控措施、疫情发展情况及复工复学意愿受到持续关注，但关注度有所下降，人们逐渐开始关注境外输入和疫情国际发展情况，此外，讨论话题又衍生出在隔离期间的生活类话题，包括工作，旅游等积极情感表达类继承话题及致敬祖国，感谢人民付出的情感类话题。值得注意的是，疫情发展情况、疫情防控措施在后三个阶段中被持续关注。该阶段相关部门应重点对新一轮舆情进行监督和干预，快速并有针对性地发布权威解释，对即将消退的舆情，如国内开学、就业以及复工复产等问题，出台相关措施并及时发布，做好善后恢复。此外，通过舆情评估，对整个应对过程进行经验总结，发现类似舆情事件传播应对规律，在未来类似突发公共卫生事件发生时，可将已有知识经验快速向情报转化[17]。

通过以上分析可以看出网络舆情传播周期各阶段主题转变迅速的同时具有关联性和继承性。疫情下的网络舆情具有很强的聚焦性，持续被公众关注，一直到事情演化结束，较高和持续的关注度容易诱发类似网络舆情，进而使网络舆情传播内容出现具有关联性。此外，突发公共卫生事件中，公众对事件的进展异常

敏感，每当有新的进展，公众讨论度极速上升，产生新的关注点，引发新的舆情主题，呈现舆情主题易变的特征。总的来说，此次新冠肺炎疫情舆情事件主题演化规律呈现舆情主题关联、舆情主题易变的传播规律。

4.3 舆情传播情感倾向时序演化分析

探究新冠肺炎疫情网络舆情主题演化背后的网民情感倾向时序演变情况，有助于政府实施舆情良性引导和舆情及时管控。前两部分已经将新冠肺炎疫情舆情传播周期划分为四个阶段，并且挖掘了舆情传播周期各阶段的主题演化规律。在此基础上，这部分主要从动态角度深入研究网民情感倾向时序变化情况。

值得一提的是，Senta 本身只支持二分类，但是其实是输出一个正面和负面的概率，然后直接把较大概率对应的感情作为结果，对于一些看似中性的句子，往往会给出 0.5 的情感得分。基于此，本文在 Senta 相关源码的基础上，将输出的概率值进行三分类：大于 0.6 的划分为正向，小于 0.4 的则为负向，介于二者之间的为中性。

由于本文主要探究在舆情主题演化背景下的情感倾向变化，因此，需要展示情感倾向演化的时序图。对于情感值的时序变化，本文直接计算出每天的情感平均得分，然后绘其中两条虚线分别代表着正负倾向的分界线。绘制出情感倾向时序变化图，得到的时序图如图 4.5 所示。

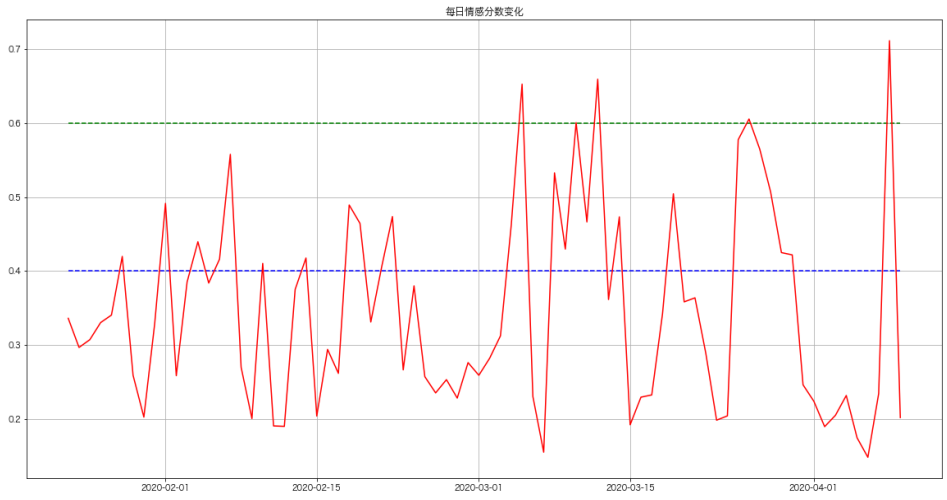


图 4.5 情感倾向时序变化图

按照舆情划分的阶段，第一阶段是舆情酝酿期，该阶段的情感波动较小，情感处于紧张焦虑期。由图 4.6 中可以看出这个阶段的情感值处于逐渐上升趋势，由一开始的负面情绪转向中性情绪，由于发布的关于武汉不明原因肺炎通告，公众面对突如其来的未知疾病充满猜测、怀疑，表现出担忧、恐惧等负面情绪。随着主题演化，调查表明，目前肺炎感染人数较少，且未发现存在人传人现象，因此，大多数网民紧张焦虑的负面情绪逐渐得到缓解，情感值开始逐渐上升，慢慢由负面情绪转向中性情绪，表现出相信政府、等待官方回应等情绪。总体来看，该阶段的情感波动较小，由一开始的负面情绪转向中性情绪。因此，该阶段相关部门应该及时公布关于新冠肺炎疫情的最新调查，及时满足公众对于不明原因疫情的恐慌，起到稳住舆情极速恶化的势头的作用。

第二阶段是舆情爆发期，该阶段的情感值整体呈负面情绪，由前一阶段的中性情绪快速下降至负面情绪，情感处于悲观恐慌期。从图 4.6 中可以看出情感值骤然下降，结合舆情主题挖掘来看，阶段的舆情主题主要为疫情发展情况、新冠肺炎传播途径与方法及防护措施。1 月 20 日，钟南山院士公布由新冠病毒引起的肺炎存在人传人的现象，引起了全民恐慌，紧接着，全国各地出现新冠肺炎病例，武汉市紧急封城，疫情的爆发与蔓延形势的突发性与防控任务的迫切性引起了全国民众的恐慌与悲观，其中，还主要包括对部分政府措施及患者就医困难等问题的不满，新冠肺炎疫情的舆情高峰形成后，政府行为成为情的核心部分，政府每一次行动都可能导致舆情转换方向。总的来看，舆情爆发阶段情感值总体呈负向情绪，该阶段相关部门应该及时响应政策，出台强有力的科学防控措施，同时要保证公开透明疫情数据，引导民众情绪积极向上，避免引发而持续二次舆情。

第三阶段是舆情波动期，该阶段的情感值由低到高，又逐渐下降，整体呈波动状态，情感基本处于自信振作期。由图 4.6 可以看出，情感值一开始逐渐性正面情绪转化，转至中性情绪时又下降至负面情绪，随后情感值一直处于上下波动状态。第三阶段前期我国已经全面进入抗疫阶段，各地采取联防联控和内防外控的应急机制，全面了落实隔离、追踪等防疫措施，民众看到政府如此强有力的抗疫措施和国内疫情逐渐趋于稳定的局势，负面情绪明显减少，逐渐衍生出的关于复产复学意愿主题。但是情感值并没有持续向好，负面情感占比仍然较大，并有明显的下降趋势，在第三阶段后期由于疫情逐渐开始全球扩散，网民的焦点逐渐

转向国外并且混杂一部分谣言。其中，负面情感主要包括对国外疫情的担忧、华人同胞受到歧视、外国媒体对我国的负面报道等。总的来看，舆情波动阶段情感值总体呈波动不定，大部分呈现负面情绪，该阶段相关部门应该及时回顾前期舆情演化规律，及时分析舆情演化，预测后期舆情演化，针对未解决的重点问题做出应对干预，加强舆情预测预警，引导舆情导向持续向好发展。

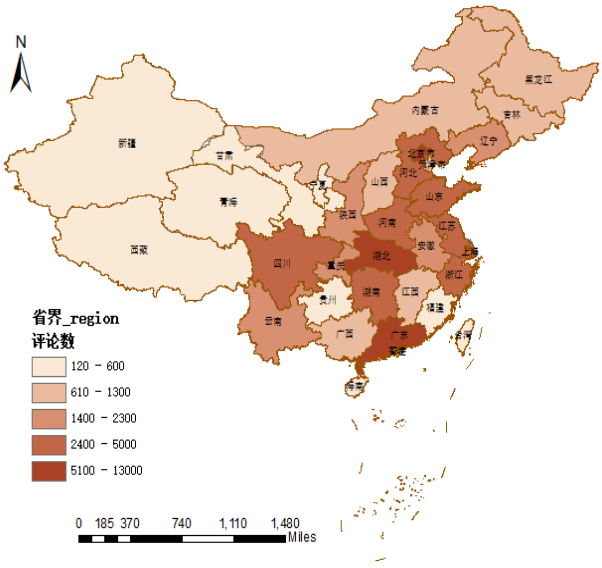
第四阶段是舆情长尾期，该阶段的情感值由低到高，期间伴有小频率波动，整体呈正面中性情绪状态，基本处于平稳恢复期。由图 4.6 可以看出，情感值由负面情绪逐渐转向正面情感，并且连续好几天呈现中性情绪，期间也存在小波动。根据主题演化分析，在第四阶段前期，我国疫情防控持续向好，多地确诊病例清零，复工复产实现一定规模，对比国外疫情的严峻性，逐渐演化出赞扬祖国伟大、中国社会主义制度的优越性及对灾后生活的向往的主题，情感值持续向好。但是全球疫情持续加重，国内疫情稳定的局势受到了境外输入病例的影响，国内防疫工作中心也开始注重外防，民众开始担忧境外输入病例会对国内战役成果产生威胁，存在国内疫情二次爆发的隐患。因此，情感值开始下降，涌现出关于对境外输入应该采取的防护措施等主题，情感值跌到了最低点。随着国内对境外输入病例的妥善处理，民众再次看到党和政府的实力，情感值又开始回温。在该阶段相关部门应该及时回顾新冠肺炎疫情舆情应对过程，及时发现应对过程中存在的纰漏与短板，重点分析舆情主题发生演变的拐点和情感倾向发生较大变化的节点，进行经验总结，梳理出突发公共卫生事件网络舆情演变规律和情感时序特征，为未来应对类似事件舆情传播提供参考价值，并且形成针对突发公共卫生事件舆情管理系统，从而降低社会治理成本。

5 网络舆情空间分析

把握网络舆情演化规律及映射出的情感倾向是政府及相关部门识别舆情传播阶段，实施舆情积极引导和舆情管控的重要依据，网络舆情传播主题演化及情感倾向演化会呈现不同时间阶段特征之外，在地理位置的制约下也存在空间上的分布差异。此次新冠肺炎疫情传播范围广，危害性强，全国各个省份均不同程度受到波及。因此，伴随疫情而生的网络舆情在不同区域会表现出不同的强度，微博用户地理位置和嵌入文本中的地理位置信息是研究网络舆情空间分布特征的富有价值的指标。为了充分挖掘微博数据蕴含的位置信息与情感信息，本文借用空间可视化方法，对比分析统计出的不同区域微博评论用户和疫情本身确诊病例数的空间分布情况，同时量化不同区域网络用户在疫情期间表现出的情感平均值和被提及次数，在此基础上进行舆情社区发现分析，刻画出空间层面上舆情网络结构。挖掘出不同区域网络舆情的空间差异性，对于政府及相关部门制定差异化防疫对策及措施具有重要的现实意义。

5.1 评论用户数量分布

为了使新冠肺炎疫情背景下的网络舆情势态直观可见，利用微博用户地理位置，统计出不同区域微博用户发评数量，用其表示不同区域对此次新冠肺炎疫情的讨论积极性和对疫情发展态势的关注度。本文首先统计出爬取的有关新冠肺炎微博评论样本数据中全国 34 个省份的网络用户发表的评论数量，其次，整理出 2020 年 1 月 1 日至 2020 年 4 月 9 日中国 34 个省（市、区）卫健委官方网站公布的新冠肺炎累计确诊人数，利用 Aicgis10.4 软件进行空间可视化展示，绘制出微博评论人数空间分布图（图 5.1）和疫情确诊人数空间分布图（图 5.2）。



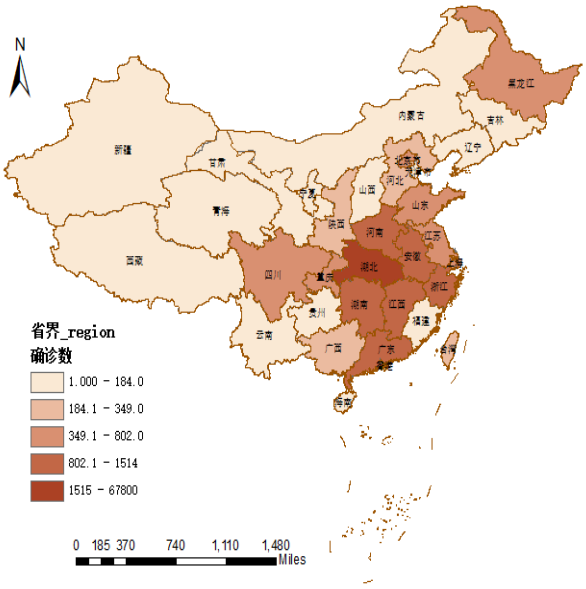


图 5.1 微博评论人数空间分布　　**图 5.2　疫情确诊人数空间分布**

从微博参与评论人数空间分布图来看，有关新冠肺炎疫情的话题讨论覆盖全国各个省份，即由疫情引起的网络舆情具有全域覆盖性和全民参与性。其中，所在地是广东、北京和湖北的网络用户人数和评论参与人数明显远多于其他省份，参与率（参与评论人数占样本评论总人数的百分比）分别高达 14.6%、13.1%和11.9%，此外，广东、北京和湖北三个省份的疫情确诊人数也明显高于其他省份，是防控压力和抗疫难度较高的区域，因此该区域的网络用户对疫情发展过程的讨论积极性和关注度最高；参与率大于 4%的省份依次为河南、上海、山东、四川、江苏、湖南和浙江，这七个省份均为此次疫情扩散较为严重的区域，疫情爆发前期这些区域确诊人数持增加，随着全球疫情的爆发，后期境外输入病例逐渐增多；西北地区，东北地区除辽宁和西南地区除四川包括福建、海南和内蒙古在内的多个省份疫情发展程度较低，这与疫情本身不太严重以及后期采取较为严格及时的防控机制有很大的关系。因此，这些区域的微博评论人数相对较少，对疫情发展参与积极性和关注度较低。

从疫情确诊人数空间分布图来看，全国各省新冠肺炎累计确诊病例数存在较强的空间相关性，确诊病例人数高、低分布具有明显的空间聚集性，高确诊人数的地区主要集中在与湖北省相邻近的区域或周边区域，并且这种聚集特征随时间推移逐渐趋于稳定的态势。各省份新冠肺炎累计确诊病例数的分布具有明显集聚

性，虽然在不同时间点存在较小差异，但整体的空间分布特征并没有发生变化，即新冠肺炎高确诊人数的地区主要集中在湖北省以及与之相邻或周边的省份，如河南省、湖南省、浙江省、广东省、安徽省和江西省等。从地理位置和经济发展程度上看，安徽省和河南省距离湖北省武汉市最近，其交通最为便利；而浙江省紧邻安徽省与江西省，经济较为发达，贸易往来以及人员流动较为频繁等等。另外，通过探索性空间数据分析发现新冠疫情确诊病例数的分布与中国人口密度分布图极为相似，其中高确诊人数主要集中在中国东南部各省份，在中国人口密度分布图中与之对应的是“胡焕庸线”[①]的右侧，即高人口密度的中国东南半壁。

总的来说，全国各省新冠肺炎累计确诊人数存在较强的空间相关性，确诊人数高、低分布具有明显的空间聚集性，高确诊人数的地区主要集中在与湖北省相邻近的区域或周边区域，而由疫情引起的网络舆情具有全域覆盖性和全民参与性。此外，还可以发现，由新冠肺炎疫情引起的网络舆情空间分布与新冠肺炎疫情确诊病例数据的空间分布特性整体具有相似一致性，但也存在局部差异性。相似性一致性体现在网络舆情关注度与疫情确诊人数空间分布之间存在正相关关系，疫情越严重的区域，其网络舆情关注度也较高，不过根据距离疫情爆发地的远近呈现一定的距离衰减规律。局部差异性体现在某些区域疫情本身较为严重，但是其有严格的防疫措施，使得确诊病例人数在短时间内不再增加，这就使得网民的关注度逐渐降低，产生网络舆情空间分布与疫情确诊病例数据的空间分布特性的局部差异性。有关新冠肺炎疫情微博参与评论人数与累计确诊人数之间有密切的关系，地方疫情越严重的地方以及越靠近湖北地区的地方，其地区网络民众参与疫情网络舆情演化的讨论热度和对疫情发展形势的关注度越高。这为下一部分分析不同省份对新冠肺炎疫情发展过程所呈现出来的情感值奠定了基础。

[①] “胡焕庸线”是由中国地理学家胡焕庸(1901-1998)在 1935 年提出的划分我国人口密度的对比线，也称黑河（爱辉）—腾冲线，首次揭示了中国人口分布规律。即自黑龙江瑷珲至云南腾冲画一条直线（约为 45°），线东南半壁 36%的土地供养了全国 96%的人口；西北半壁 64%的土地仅供养 4%的人口。 二者平均人口密度比为 42.6∶1。

5.2 空间情感状态分析

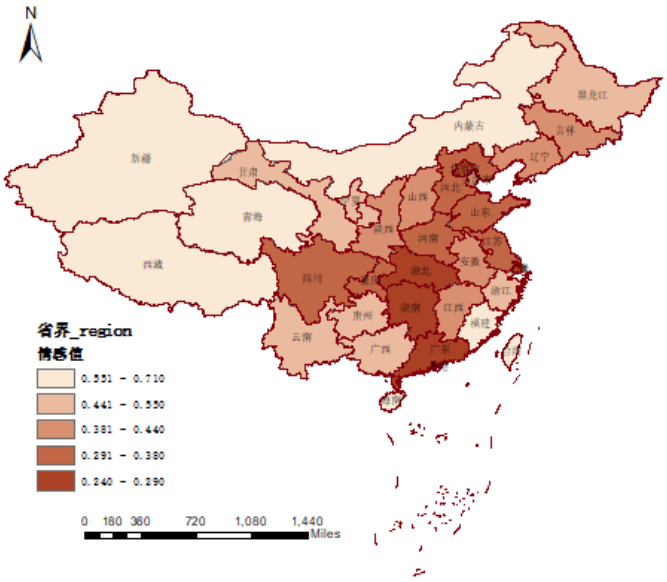


图 5.3 全国各省份网民情感值分布图

为了研究不同区域的情感倾向特征，依据微博用户地理位置信息将不同省份的文本数据归类，计算文本情感值的均值作为表示不同区域网络用户的情感倾向空间特征的评价指标，首先利用 Python 计算出不同区域的情感均值，接着通过 Aicgis10.4 软件生成情感值空间分布图，更加直观地反映情感倾向空间特征，绘制出全国各省份网民情感值空间分布图，见图 5.3。如图 5.3 所示，根据情感值数高低可以把 34 个省份划分 5 个大区域。区域 1 包含的省份有青海、海南、福建、新疆、澳门、西藏、内蒙古和台湾，这一区域情感值总体呈正向情绪，情感值在 0.55 至 0.7 之间，由前一部分分析可以看出，该区域的新冠肺炎疫情严重程度较低且受境外输入病例的影响较小，微博评论人数较少，对疫情发展的关注度较低，折射出来的情感倾向总体趋于积极。区域 2 包含江西、天津、浙江、黑龙江、贵州、甘肃、广西、宁夏和云南等省份，这一区域情感值总体呈中性情绪，情感值在 0.44 至 0.55 之间，其中浙江地区疫情较为严重，但是浙江省在疫情防控和复工复产等方面做出了积极贡献，被网友称为“满分答卷”，呼吁其他省份积极“抄作业”，成为了全国疫情防控及复工复产工作的榜样省份，因此，即使浙江省确诊人数并不少，但是其表现出的情感值较积极；宁夏和甘肃等省份

虽然确诊人数与评论人数较少，但后期承接了不少境外输入病例，有受境外输入影响，其情感值相比其他确诊人数少的省份稍微较低；其余几个省份疫情严重程度一般，且评论人数较少，用户情感倾向总体较为乐观。区域 3、区域 4 包含四川、香港、河南、山东、江苏、重庆、河北、辽宁、安徽、陕西、山西及吉林等省份，这一区域情感值在 0.31 至 0.44 之间，整体呈负向情绪，该区域中部分省份外出务工的人数较多，春节前返乡人数多，由于新冠肺炎病毒具有较长的潜伏性，潜在患病人数较多，又因为该区域前期疫情扩散较为严重，后期受海外疫情及境外输入性病例的影响，导致网络用户较为担忧，大多数表现出消极情绪。区域 5 包含的省份有广东湖北、北京、湖南和上海，这一区域的情感值最低，情感值在 0.29 之下，该区域社会经济发达，人口流动性强，疫情肆虐程度较高，本土确诊病例长期未清零，境外输入病例持续增加，多重因素致使该区域网络用户情绪普遍低迷，长期心情愉悦度低，情感倾向为负向。

总的来看，由新冠肺炎疫情引起的网络舆情事件，其区域情感倾向特征与疫情事件严重程度有很强的相关性，但是是负向相关性，具体表现为疫情越严重的省份和越靠近疫情爆发地的省份，其区域公众对疫情网络舆情的讨论积极性和疫情发展趋势的关注度越高，但是呈现出的情感值就越低，相反区域疫情较轻的区域，其呈现出的情感值就越高。因此，情感值较低的区域，与新冠肺炎疫情最严重的区域存在较高的吻合度。

5.3 舆情网络社团挖掘

前面利用微博用户地理位置信息，对比分析微博评论人数和疫情确诊病例的空间分布情况，量化不同区域情感倾向特征，发现由疫情引起的网络舆情在空间分布中存在特有的特征，但是并没有深层次剖析文本数据中蕴含的地理位置信息。因此，进行舆情网络社区挖掘，统计出在疫情期间各个省份被提及的次数及构建出各个省份的舆情网络结构，深层次的分析网络舆情在空间中特有的属性。

本文需要构造的舆情网络社团是一组由省级行政区点位位与有向共现词链接组成的集合，省级行政区域间联系紧密且存在社团化或群组化的结构[53]。依据第二章有关舆情网络社团构建的理论知识，舆情网络社团结构发现过程是根据共现省级行政区域词链，将省级行政区节点一个个划分到不同社团中的过程，社团

内部节点存在某种特质[53]。

本文首先从爬取的微博评论文本中检索出含省级行政单位或下辖城市地名信息的微博评论者，将下辖城市扩展到相应的省级行政区，根据每个省市所发出的评论中提及其他省市的次数来构建一个邻接矩阵。这里引入了一个外部文件，该文件保存着中国各个省市的信息，通过该文件将评论者所在地及评论文本中提及的省市统一归到34个省份中，进而构建出34个省市在被提及次数的邻接矩阵，在此基础上统计出每个省份被提及的次数，见表5.1所示。

表5.1 省市之间被提及次数

省份	被提及次数	省份	被提及次数
湖北	3413	山西	165
广东	1592	福建	141
北京	1523	贵州	123
山东	994	台湾	118
河北	674	陕西	104
云南	564	江苏	81
浙江	517	海南	77
四川	458	天津	75
西藏	380	新疆	70
河南	356	宁夏	58
上海	336	甘肃	49
内蒙古	289	黑龙江	46
香港	275	江西	44
辽宁	241	重庆	44
吉林	214	广西	25
安徽	210	青海	11
湖南	203	澳门	7

通过统计每个省份在新冠肺炎疫情事件评论文本中被提及的次数，从表5.1中可以看出，其中湖北被提及的次数最多，出现了3413次，广东与北京被提及的次数也较多，分别为1592和1523。其次山东、河北、云南与浙江被提及的次数在500以上。这与本章第一部分地方得出的结论相吻合，疫情越严重的地方以及越靠近湖北地区的地方被提及次数越多，说明上述地区被网友广泛讨论，是舆情管理的重中之重的区域，这些区域的疫情发展情况及防疫管理措施备受网友关注。因此，及时分析出网友重点关注的区域，在全国统一防疫政策与制度下各区

域政府应该差异化出台疫情防控制度，及时跟进舆情管控，防止区域舆情崩塌，进而影响区域政府公信力。

之后将得出的各省市被提及次数的邻接矩阵导入 Gephi 软件，生成舆情网络社团图。在模块化中设定好参数即可进行 Louvain 社区检测算法，将各个省份聚类到不同社团，以不同的颜色标出，同一个社团内的省份之间的互动比在不同社团的省份要更频繁[6]。不同的边的粗细由他的度数决定，点的大小则由他的出度决定。最后得到的舆情社交网络结构图 5.4 所示。

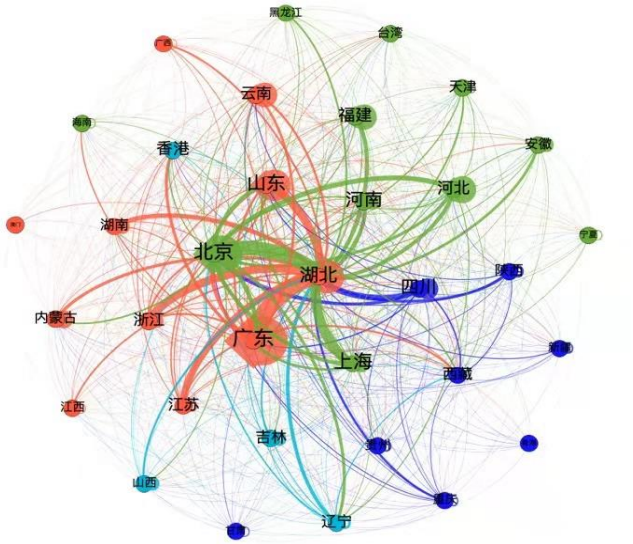


图 5. 4 基于“被提及人数”的社团结构

如图 5.4 所示，利用 Gephi 软件 Modularity 模块，进行社团结构检测，将 34 个省份划分为 4 个社团，模块度计算结果为 0.32，说明省份之间具有较好的社团结构特征，划分的 4 个社团中，同一社团内省份之间联系比不在同一个社团的省份联系更紧密。节点大小代表各省份的加权度，即考虑到不同个省份用户发微博的数量，又考虑到不同省份在评论文本中被提及的次数，反映节点在舆情社团结构中的层次。节点间粗细不同的边代表被提及次数合并的权重，边的连接次数数量代表节点之间的互评次数，二者共同反映节点在舆情网络中被关注程度高低[22]。由于生成的是有向图，顺时针代表图的连接方向，从图 5.4 可以看出，如广东湖北之间较粗的一条代表广东用户提及湖北，较细的一条代表湖北用户讨论广东。湖北被其他各个省份广泛提及，其中，广东、北京、山东、上海、浙江及江苏的网友提及湖北次数明显多于其他省份，这与当地疫情发展情况、经济发展状

况和人口数量有着较大的相关性。将节点的入度与出度统一纳入考虑，节点活跃度较高的省份有广东、北京、湖北、山东、河南、上海及四川，基本都是前期疫情较为严重或者后期境外输入病例颇多的省份，这与前几部分得出的结论相吻合。同一个社团省份具有一定的地理相关性，本文的社团一用橘红色表示，社团二用绿色表示，社团三用深蓝色表示，社团四用浅蓝色表示。通过观察这四个社团可以明显看出聚类的四个社团内部具有很强的空间相关性。

通过以上分析可以发现，突发公共卫生事件背景下的网络舆情具有很强的空间特性，不仅和疫情的严重程度由密切的联系，还和社会经济发达程度，人口数量多少及交通运输设施情况具有较强的联系。疫情最严重的区域是全国各个省份广泛关注的焦点，同时，临近最严重区域的省份及周边省份也是被关注的重点。此外，经济发达，人口密度大及交通便利的省份也被广泛关注，其关注度仅次于疫情最严重的区域。可见，相关公共部门在制定防控管理政策及措施的时候应该分区域差异性对待处理，不能“一视同仁”，“一篮子”装。同时相关媒体在公布相关信息的时候应该根据区域差异性发布地域倾向性新闻，进而引导各个区域网民情感势头向好，避免出现部分区域网民情绪极端，出现“地域黑”等现象。

6 结论与启示

6.1 研究结论

6.1.1 网络舆情传播主题演化与情感时序变化特征

对新冠肺炎疫情事件网络舆情传播主题演化与情感时序变化特征进行分析，可以发现舆情传播过程中的主题演化与情感变化特征具有一定的规律性。

首先，新冠肺炎疫情网络舆情演化周期可以划分酝酿期、爆发期、波动期和长尾期这四个阶段，该舆情事件演化周期中酝酿期大致有 17 天，爆发期有 14 天，波动期有 29 天，长尾期有 40 天，新冠肺炎疫情舆情事件与一般舆情事件存在明显差异，具有较长时间的波动期和长尾期，网络舆情并没有经过一段时间彻底消亡，演化过程中出现阶段反复及波动频率较高的规律。

其次，根据划分的舆情传播周期，将评论数据划分到不同阶段挖掘舆情主题，进行舆情主题演化分析，发现舆情传播各个阶段的主题内容侧重点存在一定的差异，但是具有话题继承性和连续性且不同阶段的话题关联强度有所差异。随着疫情事件演化，不同阶段主题侧重点不断发生改变。如酝酿阶段的主题主要为武汉市不明原因肺炎的报道、新冠肺炎症状及前期的治疗措施；爆发阶段的主题逐渐演变为对疫情发展动态、新冠肺炎传播途径与方式、应急防控措施及对一线工作医护人员的关注等；波动阶段的主题演化为疫情发展情况、疫情防控防范措施、新冠肺炎开始全球流行、复工复学意愿、官员不当行为整治、对医护人员的肯定与支持以及谣言散布情况等话题。其中疫情发展情况、疫情防控防范措施、政府官员在疫情期间的不作为及对一线医护人员的赞扬与肯定为继承话题，是由上一阶段舆情持续发酵而形成的话题；长尾阶段的主题分别为对境外输入的态度、境外输入应该采取的防护措施、对战疫成果的肯定及对祖国的赞扬、国内持续防控措施、疫情发展情况及复工复学意愿。值得注意的是，疫情发展情况、疫情防控措施在后三个阶段中被持续关注。此外，可以发现，舆情传播前两个阶段，网民关注热点多为疫情本身及防控措施等，随着疫情好转，疫情相关衍生话题超越有关疫情事件本身话题，值得注意的是，舆情传播后两个阶段主题数量增多，主题演化逻辑趋于复杂，因此，突发公共卫生事件后期的舆情管控及引导极为重要，

需要格外重视舆情传播过程中的主题和网民的情感波动，及时加强人文关怀，宣传正能量，弘扬主旋律。突发公共卫生事件本身已经给公众带来生活及工作等多方面的消极影响，因此，相关部门需要因时因地及时调整应对举措，将各种政策有效落到实处，实现更好的舆情引导和管理。

最后，分析随着时间变化，不同阶段主题演化背景下的情感倾向变化特征，发现随着舆情传播周期演化，主题背景下的情感变化浮动较大，总的负面情绪和中性情绪居多。舆情酝酿阶段的情感波动较小，情感处于紧张焦虑期，情感值由一开始的负面情绪转向中性情绪；舆情爆发阶段的情感值整体呈负面情绪，由前一阶段的中性情绪快速下降至负面情绪，情感处于悲观恐慌期；舆情波动阶段的情感值由低到高，又逐渐下降，整体呈波动状态，情感基本处于自信振作期；舆情长尾期阶段的情感值由低到高，期间伴有小频率波动，整体呈正面中性情绪状态，基本处于平稳恢复期。疫情事件随着时间推移，经网络不断传播、情绪持续发酵，往往会引起较为极端的情绪，会产生难以预想的后果。因此，掌握不同阶段的舆情演化走势和情感倾向特征在突发公共卫生事件网络舆情引导和社会综合管理等方面具有颇高的现实价值。

6.1.2 网络舆情的空间区域特性

通过对比分析在新冠肺炎疫情期间不同区域微博评论用户和疫情本身确诊病例数的空间分布情况，同时量化不同区域网络用户表现出的情感平均值，在此基础上进行舆情社区发现分析，统计出在疫情期间各个省份被提及的次数，进而刻画出空间层面上舆情网络结构，深层次的挖掘由具有地理特性的突发公共卫生事件引起的网络舆情在不同区域之间网络舆情的影响情况。发现新冠肺炎疫情引起的网络舆情具有很强的空间特性。

首先，新冠肺炎疫情背景下的网络舆情具有全域覆盖性和全民参与性。网络舆情的传播和新冠肺炎疫情本身的发展在空间范围内存在不完全一致的现象，在疫情发展初期，确诊病例只出现在武汉市及周边的几个省份，疫情本身的发展只在小范围内扩散，但是由疫情引发的网络舆情传播范围则扩散至全国各个省份，引起全民广泛关注，上述说明新冠肺炎疫情本身的发展情况与经济发展情况、人口密度及地理空间距离相关，具有较强的空间集聚性，但是由其引发的网络舆情

在空间范围内的传播并没有完全受地理区域限制，可以实现短时间内跨区域传播影响。

其次，新冠肺炎疫情背景下的网络舆情空间分布特性与疫情确诊人数空间分布情况整体呈现相似一致性，但是也存在局部差异性。相似一致性体现在各区域网络舆情的关注度（微博评论人数空间分布与各省份被提及次数）与疫情确诊人数空间分布情况之间存在正相关关系，即某一区域疫情越严重，该区域的网络舆情无论是评论人数还是被提及次数，都远高于其他区域，这些区域节点活跃度较高。如湖北疫情最严重，其评论人数与被提及次数相应也就更高，自然其网络舆情关注度就越高，对广东与北京而言，本土确诊病例数相对较多，疫情较为严重，而且后期境外输入病例也多，因此也具有较高的舆情关注度，社团节点活跃度较高，在舆情管控措施方面可以相互参照；而疫情相对较轻的省份，网络舆情关注度较小，但是相似一致性根据距离远近存在距离衰减规律，越靠近疫情严重地方的省份，其网络舆情越高，离疫情严重地方越远的省份，其评论人数和被提及次数也较少，即网络舆情关注度越小。如湖北相邻的省份河南、湖南等的网络关注度高于距离较远的西北几个省份。局部差异性体现在网络舆情覆盖范围在少数区域存在差异性，如在疫情初期，距离湖北较远的广东、北京和浙江等省份对疫情事件关注度持续保持高关注度，特别是广东和北京，这说明网络舆情所反应的地理区域特性具有局部差异性。这与区域经济、人口、交通、政治文化等因素相关，北京和广东网络普及度高，经济发达，交通便利，人口流量大。互联网资源供给本身倾向于经济活动聚集的区域，在空间分布上往往是非均衡[47]。因此，由新冠肺炎疫情引发的网络舆情在空间分布上必然与疫情本身之间存在相似一致性的同时存在局部差异性。

最后，新冠肺炎疫情网络舆情的各区域情感倾向性与各区域疫情严重程度有强相关性，但是是负向相关性。从全国范围来看，网络舆情的空间情感状态受疫情事件本身影响，发现疫情越严重的省份，其网络舆情空间情感状态越倾向于负面情绪，疫情较轻的省份，其网络舆情空间情感状态倾向于中性及正面情绪，因此，网络舆情空间情感状态与疫情事件本身之间呈现负相关。如湖北的确诊病例数最多，其情感值较低，同时像广东与北京，确诊病例数也相对较多，而且后期境外输入病例逐渐增多，因此这两个省份也具有很低的情感值。

6.1.3 网络舆情普适性结论

综合上述得出的新冠肺炎疫情网络舆情传播主题演化与情感时序变化特征和网络舆情的空间区域特这两部分结果，可以类推分析得出较为普适性的结论：

（1）舆情周期演化：舆情周期演化一般可以分为四个阶段，分别为酝酿期、爆发期、波动期及长尾期或消退期。其中每个阶段的时间长短分别与事件本身有关，有部分舆情事件不存在波动期这一阶段。

（2）舆情主题演化：舆情主题演化呈现出一定规律，总的来说呈现现舆情主题关联、舆情主题易变的规律。起始阶段舆情事件相关主题数量较少且受关注程度低，未形成规模化；在爆发期和波动期用户参与量大幅度提升，舆情主题数增多；长尾期主题数量逐渐减少，有可能出现新的关注点。政府及相关管理部门在处理类似突发公共卫生事件时，将舆情传播数据波动幅度与主题强度演化规律进行对比，及时识别舆情演化的周期阶段。另一方面，在舆情演化周期研判的基础上，实现舆情预警、引导和控制的管理流程。在舆情传播的起始阶段，进行及时研判并预警，推进舆情良性发展；在爆发阶段和波动阶段，注重舆情引导；在舆情长尾阶段，注重舆情评估，探索舆情传播应对规律，形成预案以维护社会稳定。

（3）舆情情感倾向变化：突发公共卫生事件网络舆情情感变化主要可以分为两个阶段，前两个阶段的情感主要为紧张、慌恐、焦虑及消极的负面情绪，其中不乏小频率的波动，后两个阶段的情感主要为振作、自信、平复及积极的中性及正面情绪，其中也存在负面情绪波动。政府及相关部门应该在不同阶段及时以权威评论意见积极引导网民情绪，传达主流价值观，防止出现极端情绪，引发不必要的社会动荡。

（4）舆情空间分布情况：具有地理特性的突发公共卫生事件引起的网络舆情不仅具有全域覆盖性和全民参与性，而且与疫情本身空间分布特性整体存在一致性的同时存在局部差异性，相似一致性根据距离远近还存在距离衰减规律。网络舆情区域空间情感倾向性与疫情本身严重程度呈负相关性。因此，结合突发公共卫生事件本身的空间分布差异特点，挖掘区域粒度上的公众对网络舆情的关注度和情感特性，对于政府及相关部门制定差异化防疫对策及措施具有重要的科学意义。

6.2 研究启示

习近平总书记在研究应对新冠肺炎疫情期间曾指出，做好宣传教育和舆论指导工作至关重要，需要强信心，暖人心，聚民心，维护社会大局稳定[49]。结合本文对新冠肺炎疫情网络舆情的实证研究成果，总结出有关突发公共卫生事件网络舆情传播周期主题演化及情感倾向时空变化特征的普适性结论。为了政府及相关部门在未来能够更好的应对突发公共卫生事件，减少社会治理成本，本文提出如下几点建议：

一、完善信息发布机制，全面应对网络舆情

突发公共卫生事件和其他事件不同，其话题敏感度强，传播周期长，涉及群体广泛，疫情与舆情互相交织的复杂背景下，如果在事件爆发之后，政府及相关部门报道不及时或者不实，极容易引发公众负面情绪激增。因此，在突发公共卫生事件网络舆情应对方面，需要尽快完善信息发布机制，形成专职信息发布部门主导的联合应对方式，从公众切身利益出发，及时报道事件的来龙去脉，不能单一公布事件结果，要多层次多维度报道事件的危害程度、致病原因、防护措施和医疗物资等方面的新闻，并且应当及时处理对突发公共卫生事件防护措施实施不力的领导及当事人，提振民心民意，避免给公众留下政府不作为或者包庇不作为行为的刻板印象，减少负面舆情传播。

二、提升科学研判能力，有效管控网络舆情

突发公共卫生事件背景下的网络舆情演化迅速，影响范围深远，在疫情初期，网络舆情处于潜伏状态，政府及相关部门存在被假象蒙蔽，掩盖事实，堵塞言论的现象，容易错失遏制疫情及舆情发展的最佳时期。因此，政府应利用人工智能技术及时收集舆情言论，分析舆情演化趋势和重点区域加强对网络舆情的科学研判能力，加大人员、资金、设施等方面的保障力度，从科学视角出发做决策，针对舆情及时积极正确引导，并采取相应的措施，使舆情降温。

三、提升官媒公信力，正向引导网络舆情

网络舆情传播的便捷性和广泛性对政府公信力提出了较大的考验，在互联网时代，爆发突发公共卫生事件期间，提升和维护政府公信力的主要依托为官方媒体，官方媒体在网络舆情管控中担任着重要角色，是公众获取信息的重要来源之一，是舆论导向的引导者之一。因此，如何提升官媒公信力，正向引导网络舆情

是政府应对网络舆情时的重要任务。提升官方媒体公信力，主要可以从两个方面着手：一是树立“以人为本”的服务理念，坚持将人民利益置于首位，全面实现信息真实报道，及时科普宣传。信息公开透明在实现公众对政府有效监督的同时，是判断官方媒体公信力的重要标准。突发公共卫生事件爆发之后，网络舆情随之而来，官方媒体在真实报道事件的同时，需要加大宣传事件发展过程中出现的正能量事件和人物，树立正确的舆论导向，营造良好的舆论环境；二是尽快出台健全网络舆情方面的相关法律文件和政策，由于网络的开放性等特点，突发公共卫生事件爆发之后，存在部分不法分子造谣生事，借助网络平台肆意妄为，这就需要官方媒深入宣传中央重大文件和决策部署，及时澄清事实，还原真相，把握舆情主导方向，引导公众稳定情绪，让不良舆论无生存环境。

四、致力打造专业团队，健全网络舆情监测体系

随着互联网科技及人工智能越来越强大，网络舆情监测体系还需要进一步完善，需要搭建更为准确网络舆情监测和研究平台，打造更为专业的网络舆情研究人才，高校应该注重培养网路舆情分析方面的人才，同时提供长期开展舆情监测的平台。同时尽快出台关于网络舆情管理的规范化政策，充分利用数据挖掘、自然语言处理及机器学习等技术，及时准确的掌握各种新兴媒体诸如微博、微信、论坛等平台的信息和舆情动向，实现舆情预警全网覆盖，及时采集挖掘舆情数据，构建专门舆情预测模型，实施全方位动态监测，预测舆情演化趋势，切实了解掌握公众情感倾向，对监测过程中发现的异常因素及时展开分析，及时通报预测结果和信息。

参考文献

[1]2019-2020 年中国移动社交行业年度研究报告[R].艾媒网,2020.

[2]Blei, D. M., Ng, ,A.Y. and Jordan, MI. Latent Dirichlet allocation[J]. Journal of Machine Learning Research, 2003,3:993-1022.

[3]CHOI S,BAE B. The real-time monitoring system of social big data for disaster management[M].Berlin, Germany: Springer, 2015.

[4]Dey L, Khurdiya A, Mahajan D. Topical evolution and regional affinity of tweets [C].Proceedings of 2013 International Symposium on Computational and Business Intelligence, New Delhi, India, August 24-26. Los Alamitos, CA: IEEE, 2013: 297-300.

[5]Esuli A, Sebastiani F. Determining the semantic orientation of terms through gloss classification ACM, 2005:617-624.

[6]Fink S. Crisis Management: Planning for the Inevitable[M].New York: American Management Association,1986: 20.

[7]Gomide J, Veloso A, Meira Jr W, etal. Dengue Surveillance Based on a Computational Model of spatio-temporal Locality of Twitter[C]//Proceeding soft the 3rd International Web Science Conference. ACM, 2011:1-8.

[8]GRIFFITHS T L, STEYVERS M. Finding scientific of topics[J]. Proceedings of the National Academy of Sciences,2004,101(suppl1):5228-5235.

[9]Mark G, Bagdouri M, Palen L, et al. Blogs as a collective wat diary[C]. Proceedings of the ACM 2011 Conference on Compurer Supported Cooperative Work, Seattle, Washington, USA February 11-15. New York: NY: ACM, 2012: 3716.

[10]Neppalli V K, Caragea C, Squicciarini A, et al. Sentiment analysis during Hurricane Sandy in emergency response[J]. International Journal of Disaster Risk Reduction, 2017, 21:213-222.

[11]Sun Y, Wang S, Li Y, et al. ERNIE: Enhanced Representation through Knowledge Integration.2019.

[12]Wang J., Guo M.Y., Zhang L., et al. Research on Dissemination Rule of Public

Opinion from SNA Perspective:Taking the Vaccine Safety Event as an Example[J]. Studies in Media and Communication,2017(5): 42-49.

[13]Wiebe J, Ril off E. unannotated texts[J]. Creating subjective and objective sentence classifiers from Computational Linguistics and Intelligent TextProcessing, 2005:486-497.

[14]Zhao L, Wang J , Huang R , et al. Sentiment contagion in complex networks[J]. Physica A Statistical Mechanics & Its Applications, 2014,394(2):-1723

[15]安璐,杜廷尧,余传明,周利琴,李纲.突发公共卫生事件的微博主题演化模式和时序趋势—以 Twitter 和 Weibo 的埃博拉微博为例[J].情报资料工作,2016(05):-4452

[16]曹树金.我国网络舆情研究现状及其知识增长趋势分析[J].情报资料工作,2016(06):17-22.

[17]曾子明,黄城莺.面向疫情管控的公共卫生突发事件情报体系研究[J].情报杂志,2017,36(10):79-84.

[18]曾子明,万品玉.融合演化特征的公共安全事件微博情感分析[J].情报科学,2018,36(12):3-8+51.

[19] 蒋明敏,王雪芬,刘玥.基于 LDA 模型的网络舆情研究进展与演化分析[J].泰山学院学报,2020,42(02):116-124.

[20]陈福集,陈婷.舆情突发事件演化探析——基于意见领袖引导作用视角[J].情报资料工作,2015(02):23-28.

[21]陈兴蜀,常天祐,王海舟,赵志龙,张杰.基于微博数据的“新冠肺炎疫情”舆情演化时空分析[J].四川大学学报(自然科学版),2020,57(02):409-416.

[22]陈映雪,甄峰,王波,邹伟.基于微博平台的中国城市网络信息不对称关系研究[J].地球科学进展,2012,27(12):1353-1362.

[23]董悦,王梦.基于情感分析与 LDA 模型的网络舆情案例研究[J].价值工程,2019,38(34):169-172.

[24]董云虎.舆论引导工作要把握好“时、度、效”[J].求是,2013(20):40-41.

[25]高洁,杨宝龙,赖思宇,武虹,赵立新,杨逸萌.基于微博数据的“新冠肺炎”互联网舆情分析[J].今日科苑,2020(02):57-64.

[26]高榕.“新冠”疫情防控期间的网络舆情分析[J].今传媒,2020,28(04):37-41.

[27]黄卫东,陈凌云,吴美蓉.网络舆情话题情感演化研究[J].情报杂志,2014,33(01) : 102-107.

[28]蒋知义,马王荣,邹凯,李黎.基于情感倾向性分析的网络舆情情感演化特征研究[J].现代情报,2018,38(04):50-57.

[29]康伟.突发事件舆情传播的社会网络结构测度与分析——基于“11·16 校车事故”的实证研究[J].中国软科学,2012(07):169-178.

[30]兰月新,邓新元.突发事件网络舆情演进规律模型研究[J].情报杂志,2011,30(08):47-50.

[31]雷蒙德,荆静. 特定事件下网络舆情的情感分析与可视化方法[J/OL].情报理论与实践:1-8[2020-06-12].

[32]李博诚.基于微博的突发公共卫生事件网络舆情演化研究[D].吉林大学,2020.

[33]刘雅囡.网络舆情生成与演化机制研究[D].南京邮电大学,2015.

[34]李杰,陈思宇,张静文,徐培罡.基于大数据的疫情地理传播与网络舆情时空关系研究[J].地理信息世界,2020,27(03):31-34+41.

[35]李杰,赵阳.基于 WebGIS 的突发事件网络舆情可视化设计与实现[J].测绘地理信息,2014,39(04):38-41.

[36]李沐南.Louvain 算法在社区挖掘中的研究与实现[D].北京:中国石油大学(北京),2016.

[37]梁冠华,鞠玉梅.基于舆情演化生命周期的突发事件网络舆情风险评估分析[J].情报科学,2018,36(10):48-53.

[38]王曰芬,王一山.传播阶段中不同传播者的舆情主题发现与对比分析[J].现代情报,2018,38(09):28-35+144. [39]刘铭,王晓龙,刘远超.基于词汇链的关键短语抽取方法的研究[J].计算机学报,2010,33(07):1246-1255.

[40]刘志明,刘鲁.基于机器学习的中文微博情感分类实证研究[J].计算机工程与应用,2012,48(01):1-4.

[41]孟吉杰.突发事件政务微博发布的实证研究—以“上海发布”典型案例为例[D].上海:上海交通大学,2014.

[42]任凯,吴冬芹,郭黎黎.基于生命周期理论的公共危机舆情事件研究[J].现代信

息科技,2019,3(24):1-4.

[43]石凤贵.基于 jieba 中文分词的中文文本语料预处理模块实现[J].电脑知识与技术,2020,16(14):248-251+257.

[44]宋海龙,巨乃岐,张备,濮小金.突发事件网络舆情的形成、演化与控制[J].河南工程学院学报(社会科学版),2010,25(04):12-16.

[45]孙艳,周学广,付伟.基于主题情感混合模型的无监督文本情感分析[J].北京大学学报(自然科学版),2013,49(01):102-108.

[46]汪玉叶.突发公共卫生事件的舆情管理探析[J].西部学刊,2020(07):130-134.

[47]王波,甄峰,席广亮,钱前,吴乘月,张浩.基于微博用户关系的网络信息地理研究——以新浪微博为例[J].地理研究,2013,32(02):380-391.

[48]王秀娟.新冠肺炎疫情的网络舆情趋势与治理[J].新闻传播,2021(01):24-26.

[49]习近平.在中央政治局常委会会议研究应对新型冠状病毒肺炎疫情工作时的讲话[J].求是，2020,(4):1-5.

[50]徐迪.基于空间可视化的大数据舆情研判体系建构研究[J].情报科学,2019,37(03):22-26.

[51]许鑫,章成志,李雯静.国内网络舆情研究的回顾与展望[J].情报理论与实践,2009,32(03):115-120.

[52]易承志.群体性突发事件网络舆情的演变机制分析[J].情报杂志,2011,30(12):6-12.

[53]张岩,李英冰,郑翔.基于微博数据的台风“山竹”舆情演化时空分析[J].山东大学学报:工学版.(2020-03-11)

[54]赵岩,王利明,杨菁.公共危机事件网络舆情生命周期特征分析及对策研究[J].经济研究参考,2015(16):57-69.

[55]中国互联网络信息中心.中国互联网络发展状况统计报告[R].北京:中国互联网络信息中心，2020.

[56]周宏仁,唐铁汉.网络舆情电子政务知识读本[M].北京:国家行政学院出版社,2002.

[57]张鹏,崔彦琛,兰月新,吴立志.基于扎根理论与词典构建的微博突发事件情感分析与舆情引导策略[J].现代情报,2019,39(03):122-131+143.

致谢

时光如匆匆流水，转眼三年的硕士研究生学习生涯即将结束。在毕业论文完稿之际意味着离毕业之时日趋渐进，回顾过去，往事历历在目，感慨良多，心情久久不能平复，三年的时光给我的人生长河中增添了不一样的色彩。这篇在导师及其他老师同学们的帮助下完成的毕业论文是我的研究生生涯画上句号的关键一笔。从论文的选题构思到总体框架，从内容设计到模型选择，从结果调优到定稿，几经修改，倾注了大量的心血，回顾整个撰写过程，要感谢的人实在太多。

首先，我要衷心的感谢我的研究生导师刘明老师，刘老师是我最尊敬、最崇拜的老师。从论文的选题、开题、撰写、初稿、预答辩到修改、最终定稿，整个过程中刘老师不断地给予我宝贵的建议和帮助，令我受益匪浅。除此之外，刘老师严谨细致的教学精神和渊博深邃的学识对我往后的工作及生活带来了很多积极的引导。在这里再次向刘老师致以真诚的祝福和感谢。

其次，感谢兰财统计学院的各位老师们，从本科阶段开始，是统计学院的各位老师们为我打开了一扇统计学的大门，让我感受到了统计学的魅力，是老师们传授的专业知识和研究理论让我有能力、有信心坚定的选择未来继续从事有关统计学的工作。毕业在即，想对老师们表达深深的感谢，因为有你们，才有我充实的七年时光和值得期许的未来。

然后，感谢相互交流学习的同学们，感谢一路相伴的大学及研究生舍友，是你们一次次将我从迷茫与困惑中解救出来。在论文撰写初期，遇到了很大的挑战，大学及研究生舍友给予了很大的鼓励和帮助，帮我在自我否定之后重新燃起希望。感谢你们在我最美好的年华里带来的温暖和美好回忆。

最后，感谢我最爱的父母长期以来对我学业的支持，没有你们辛勤的付出，就没有我顺利的求学之路，你们是我未来继续努力的动力，一路以来你们是我坚强的后盾，希望我早日能成为你们最坚实的后盾！