



浙江财经大学

**2021-2022 学年第二学期**  
**《新媒体时代的公共关系》课程期末论文**

题目：微博话题舆情网络分析  
——以“人民日报评人教版数学  
教材配图争议”为例

学生姓名 李松岳

学 号 190110910212

班 级 19 软工 2 班

2022 年 6 月

## 微博话题舆情网络分析

### ——以“人民日报评人教版数学教材配图争议”为例

**摘要：**结合社会网络分析与数据挖掘技术，为挖掘舆情传播网络中的信息提供新方案。梳理与舆情网络相关的研究现状，以情感分析、社会网络分析、谱聚类分析为主要方法，并以最新热点话题为例进行研究及可视化展示。研究的结果验证社会网络分析与数据挖掘技术相结合，能够充分分析舆情传播网络。

**关键词：**微博；舆情传播网络；可视分析；情感分析；社会网络分析

Weibo topic public opinion network analysis

——Take the "People's Daily Criticize People's  
education edition of mathematics textbooks  
Illustration Controversy" as an example

**Abstract:** Combining social network analysis and data mining technology, it provides a new solution for mining information in the public opinion dissemination network. Sort out the research status related to public opinion networks, take sentiment analysis, social network analysis, and cluster analysis as the main methods, and take the latest hot topics as examples for research and visual display. The results of the study verify that the combination of social network analysis and data mining technology can fully analyze the public opinion dissemination network.

**Keywords:** Weibo; Public opinion dissemination network; Visual analysis; Sentiment analysis; Social network analysis

# 1 引言

互联网的发展影响着社会舆情的传播方式，改变了人们对于信息获取，传播的方式，越来越多的舆情信息通过互联网进行扩散传播，而微博作为主流的社交网络交互平台，成为网络舆情热点事件传播的重要途径。

近年来，社会网络分析和数据挖掘在舆情传播网络中逐渐广泛应用。黄宪成基于社会网络分析方法对自闭症议题的传播现状，揭示新媒体对弱势群体的赋权特征<sup>[1]</sup>。邱泽国等人提出 PCA-Spectral-LDA 方法，为舆情文本分析提供了新的视角<sup>[2]</sup>。牟冬梅等人结合社会网络分析方法与情报学方法，多维度分析了微博舆情事件，证明该研究视角能够充分解读舆情事件不同社会属性的外化表现<sup>[3]</sup>。

从传播学视角出发，利用社会网络分析方法和数据挖掘方法，将舆情网络的内容与传播结构相结合，为网络舆情的监管及分析舆情背后所反映的社会问题提供了新的研究视角。基于可视化以及交互的形式，可以直观展示微博舆情网络的关键信息，为理解公众反应，舆情传播提供有力支持。

## 2 研究方法

### 2.1 情感分析

文本情感分析也称为意见挖掘，是指利用自然语言处理、文本挖掘及计算机技术等方法对带有情感色彩的文本进行提取、分析和处理的过程，从而帮助用户获取有效信息。通过情感分析能否了解到网民对某个事件的看法，识别出网民发布内容的情感趋势。文本情感分析主要有基于词典和机器学习两种方法。本文是基于词典的情感分析，运用 SnowNLP 提供的语料库分析。

### 2.2 社会网络分析

社会网络分析是由社会学家根据数学方法、图论等发展起来的定量分析方法，近年来该方法在各个领域广泛应用，并发挥了重要作用。社会网络分析是社会领域比较成熟的分析方法，社会学家们利用它可以比较得心应手地解释一些社会学问题。

### 2.3 谱聚类分析

谱聚类是一个被广泛应用的聚类算法，它对数据分布的适应性相较于其他聚类算法而言更强，聚类效果更优秀，因其实现简单最近几年越来越受学者们的青睐。谱聚类具有的优点如下：谱聚类算法只需要数据间的相似度矩阵，对于稀疏数据的聚类效果非常好；相较于传统聚类算法，由于谱聚类算法使用了降维，因此在处理高维数据聚类时的复杂度更低；谱聚类算法基于谱图理论，它能够在任意形状的样本空间上进行聚类并收敛于全局最优解。

## 3 数据采集与数据预处理

### 3.1 数据采集

本文对微博话题“人民日报评人教版数学教材配图争议”下所有微博进行爬虫，时间区间选择5月26日0:00分至5月29日23:59分，爬取微博评论的发布时间，用户名，微博评论等主要信息。总共爬取有效微博数量为2681条。

### 3.2 数据预处理

本文主要关注微博之间的转发关系，转发关系在话题中具有动态，交互的特性，适合反映出舆论的传播。故根据转发关系构建出舆论传播网络。网络的点由该话题下所有发布微博的用户组成，每一个点保存着“id”编号和“name”用户名的信息。根据微博评论中的“@xxx@yyy”关键词，构建出用户之间的转发连接。剔除离群点之后，总共筛选出符合要求的用户节点数为446个，因而构建出446\*446的有向邻接矩阵，该矩阵为后续进行可视化以及聚类分析提供数据来源。

## 4 研究发现

### 4.1 舆情网络的内容分析

#### 4.1.1 舆情内容的演进分析

借鉴现有研究成果，将整个舆情事件的生命周期划分为潜伏期、爆发期、蔓延期和衰退期。以时间为横轴，微博发文量为纵轴，绘制出微博发文量随时间变化的趋势图。事件从发酵到衰退的时间较短，但仍能将整个生命周期划分为潜伏期（5月26日0时至5月27日10时），爆发期（5月27日10时至5月27日14时），蔓延期（5月27日14时至5月29日0时），衰退期（5月29日0时至5月29日23时）四个阶段，如图1所示。通过对舆情事件的阶段划分，能够对舆情事件整体传播趋势的不同时间阶段进行有效合理的划分。按照划分出的四个阶段，分别对微博舆情内容进行jieba分词，提取关键词，计算词频，绘制词云图。四个阶段中，“人民日报”，“人教版”，“争议”，“配图”，“数学”作为词频最高的关键词在四个阶段中都有出现。而不同时期的其余核心关键词有所不同，如图2的(a)所示，潜伏期以“审美”，“暴露”为主要关键词，表明网民大多在讨论人民日报评教材问题是审美问题；如图2的(b)所示，爆发期以“孩子”，“教育”，“审核”，“避重就轻”为主要关键词，公众的注意力转移到小学时期的教育上，并质疑教材审核是否达标，且有一大部分网民批评人民日报的评论在“避重就轻”；如图2的(d)所示，“衰退期”以“整改”，“国家”为主要关键词，即公众更加关注国家，政府能够加强管控，整改该教材。

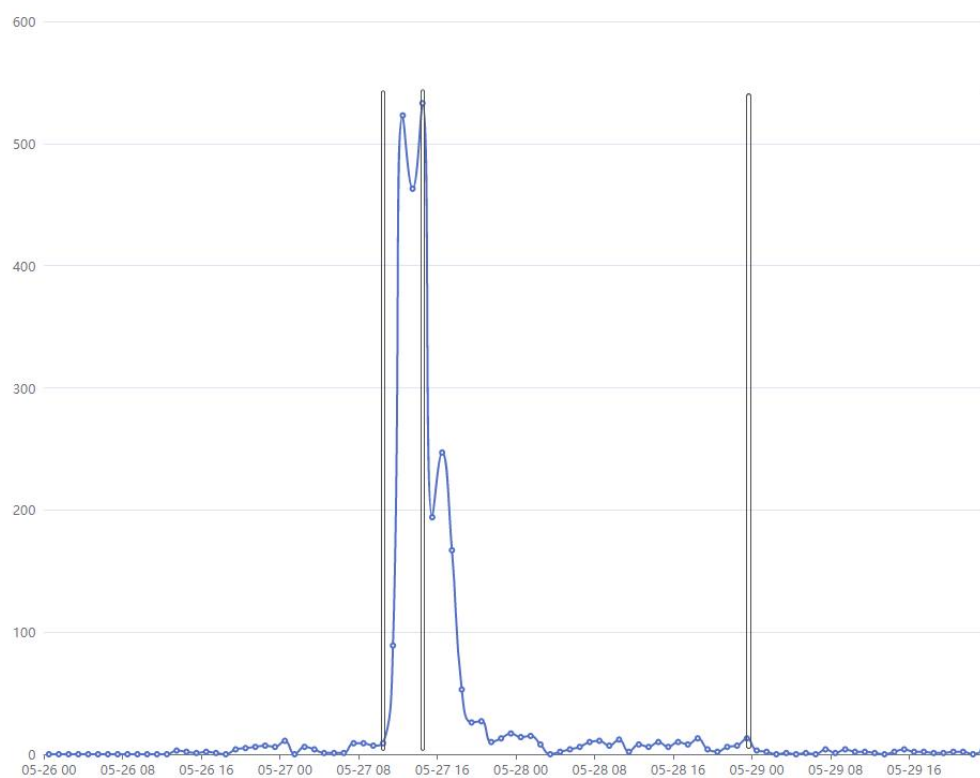


图 2

### 4.1.2 舆情内容的情感分析

舆情传播的演进过程中，公众的情感极大影响着舆情事件的导向，传播速度，传播范围等。通过分析公众的情感，能够反映出公众对于教育问题的态度，针对消极评论能够及时发现并进行引导。本文使用的是 Python 第三方包 SnowNLP，相较于其他工具，其在中文自然语言处理中表现更为出众。情感评分取值从 0 到 1，一条微博评论得分大于等于 0.6 分会被认为是积极的，处于 0.5 至 0.6 分之间为中性，低于 0.5 分则被认为是消极评论。在 2681 条微博中，有 93 条评论被认为是积极的，有 1063 条被认为是中性态度的，有 1525 条被认为是消极态度的。如图 3 所示，消极态度占比最高，积极态度占比极小。

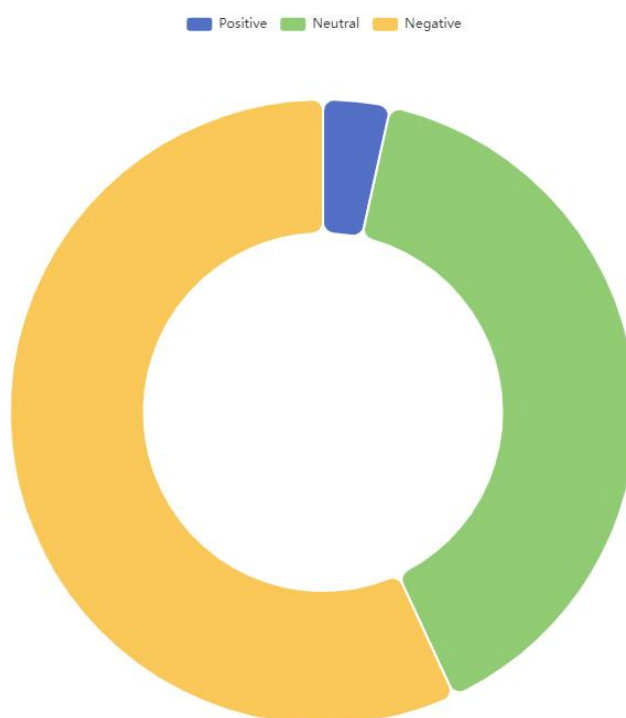


图 3

## 4.2 舆情网络的结构分析

本文将预处理后的数据构建有向邻接矩阵，通过 D3 可视化工具，利用 Force-Direct 算法生成该矩阵的关系网络（见图 4）。可以看得到该网络中的成员关系极其复杂。

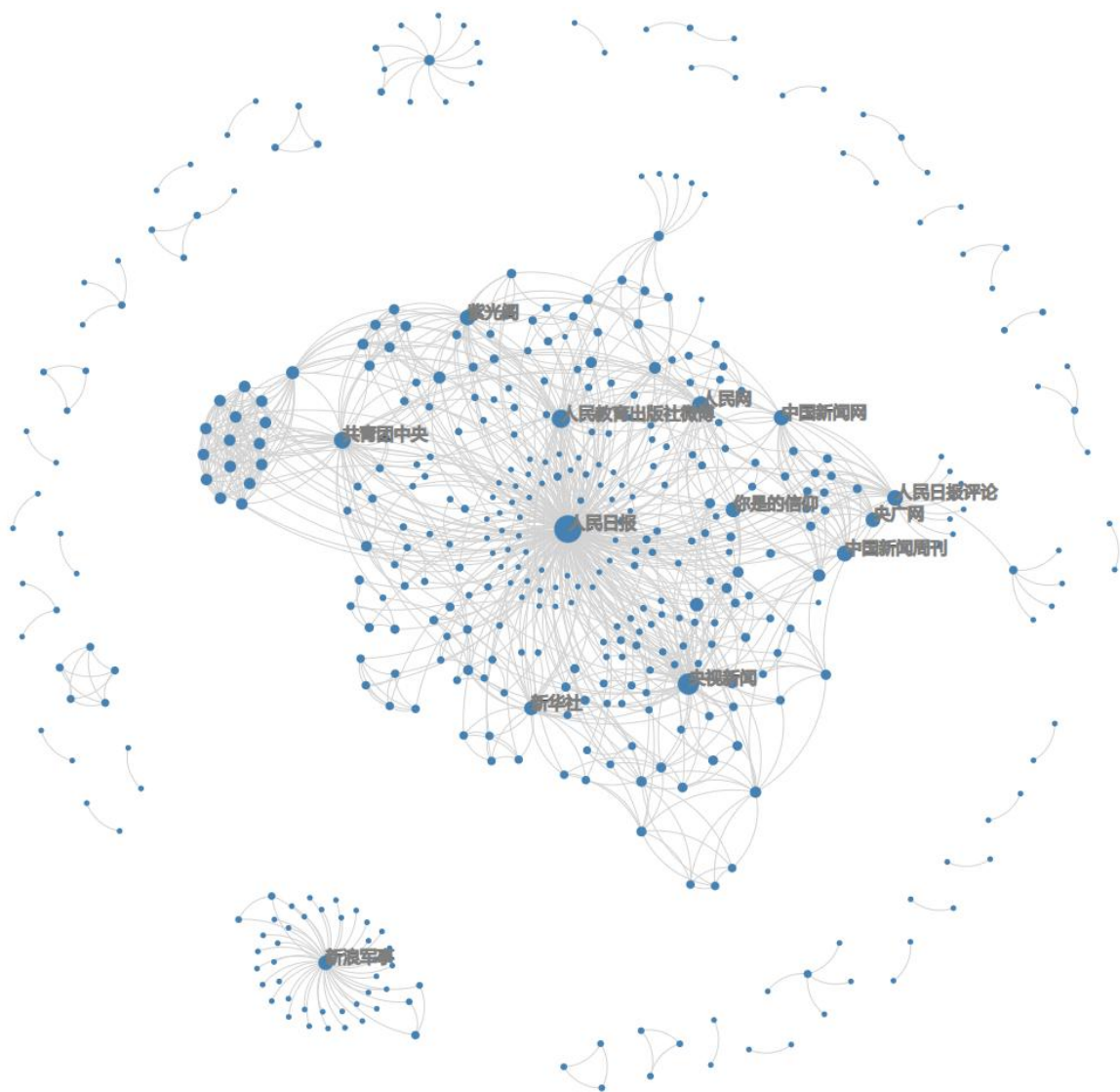


图 4

#### 4.2.1 舆情网络的参数分析

通过 Python 的第三方包 Networkx 计算，网络中存在 446 个节点，1073 条边。计算网络密度得 0.0108，说明网络中的成员之间总体联系较为松散，互动性不强。

点度中心性分析，因传播网络是有向网络，故节点有入度和出度，某节点的入度表示被转发关系，出度表示转发关系。利用 in\_degree 和 out\_degree 命令分别计算出所有节点的入度和出度，并进行排序，取入度，出度数值前十的节点。入度排名前 10 的节点（见表 1），出度排名前 10 的节点（见表 2）。

ID	Name	Indegree	Outdegree
1	人民日报	194	88
2	新浪军事	42	1

3	人民教育出版社微博	41	6
4	央视新闻	37	33
5	共青团中央	32	33
6	人民网	27	4
7	人民日报评论	25	1
8	紫光阁	22	9
9	中国新闻周刊	19	1
10	新华社	17	13

表 1

ID	Name	Indegree	Outdegree
1	人民日报	194	88
2	央视新闻	37	33
3	中国反邪教	3	23
4	共青团中央	32	22
5	玉虚龙宇君	0	17
6	拆台 CT	1	16
7	我有辣味跟我走	3	14
8	新华社	17	13
9	别扒拉颜狗	4	12
10	方方	5	12

表 2

借助点度中心性的测量和描述，可以得出以下结论：

1. 点入度分析：“人民日报”是最大的被转发节点，“人民日报”作为主流媒体和话题中的用户，被转发次数远远高出其它用户。其余 9 个（新浪军事，央视新闻，共青团中央等）被转发量较高的用户都是官方媒体，可以得出在此次舆情事件中，官媒重视该事件并被公众广泛传播。
2. 点出度分析：转发次数最多的依旧是“人民日报”，“人民日报”在该舆情话题中处于绝对中心的地位，对舆论传播起到决定性的作用。除了一些官方媒体同样积极转发外。也有大 V 或拥有一定量的粉丝的用户积极转发。（“方方”拥有 487w 粉丝，“拆台 CT”拥有 44W 粉丝，“玉虚龙宇君”，“我有辣味跟我走”，



“别扒拉颜狗”等粉丝数相对较少，在 2w 到 10w 不等），这些用户积极转发，并构建起一个关系极其紧密的传播网络。

#### 4.2.2 輿情网络的聚类分析

网络聚类分析可以根据用户之间的网络关系进行归类，可以发现在传播网络中的社群。谱聚类是从图论中演化出来的算法，在聚类中得到广泛的应用，对数据分布的适应性更强，聚类效果也很优秀。本文保留该传播网络中最大的连通分量，使用 Python 的第三方包 sklearn 中的 SpectralClustering 进行谱聚类，设置聚类数量为 5，根据节点所属不同的类别，使用 5 种不同颜色进行映射（见图 5）。蓝色类是较大的传播网，主要由一些官方媒体（“人民日报”，“人民教育出版社微博”，“央视新闻”等）以及传播能力不强的普通用户组成，这是该传播网络中的主干网络。绿色类中的成员，除了包含具有广泛传播力的“共青团中央”和“紫光阁”两个官媒外，还有大 V 和小有名气的普通用户，可以观察到这些用户之间的转发关系极其紧密（见图 6）。

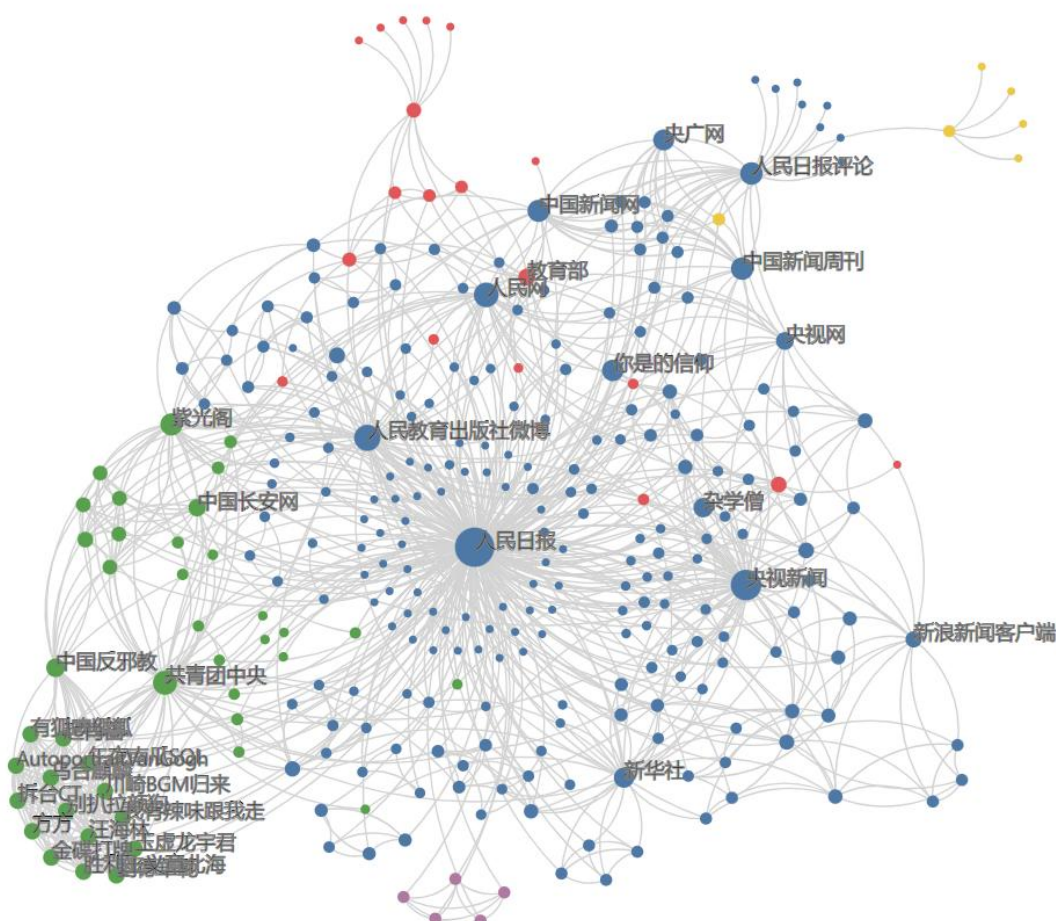


图 5

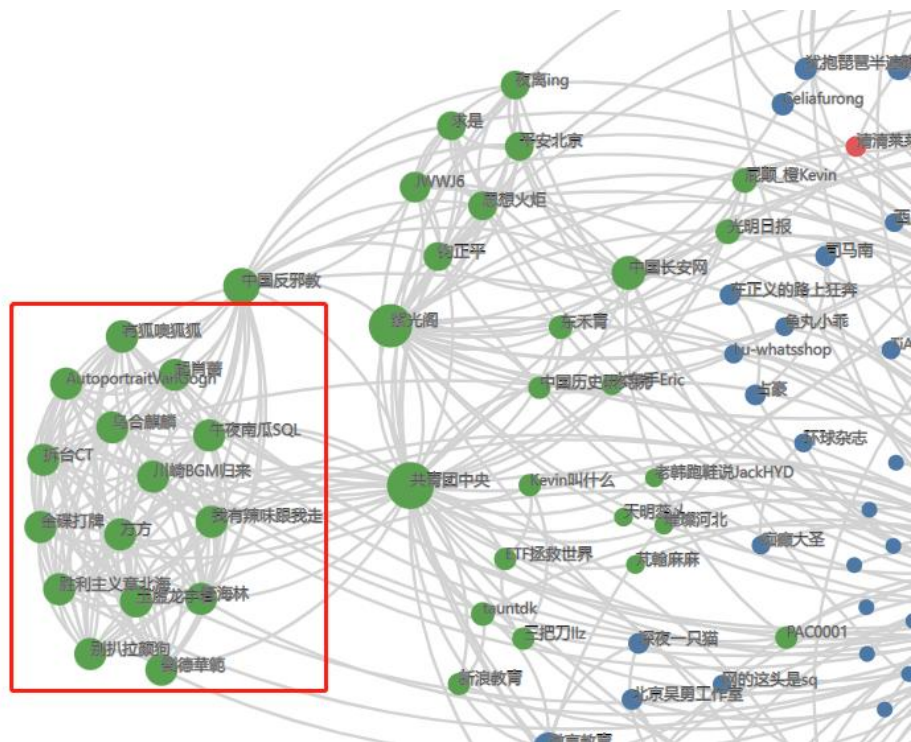


图 6

## 5 结论与展望

本文基于网络爬虫技术和网络文本挖掘方法，爬取微博文本信息，再利用 Python 对数据进行清洗和预处理操作的基础上，结合社会网络分析与数据挖掘技术，进行文本情感分析，利用 D3 可视化工具，对舆情传播网络进行可视化建模，辅助发现网络中的关键传播节点和传播子群。通过舆情演化分析发现：舆情事件的四个阶段呈现出不同的主题关键词。通过文本情感分析发现：该事件中公众的情绪多为消极态度，主要原因公众认为教材问题不仅仅是审美问题，而人民日报的回应并没有直击要害。通过舆情网络的结果分析发现：官媒在此次事件中传播范围较广，是网络的主要节点。而部分大 V 和有一定量粉丝的用户也积极在此次事件中发声，共同构建起了一个关系及其紧密的传播网络。

本文还存在一些局限之处。第一，由于微博信息噪音过大，再采集过程中出现脏数据，以及采集之后出现评论的缺失，会导致传播网络的构建不完整。第二，进行聚类分析时，由于网络并不是连通的，于是只保留了最大的连通分量，这会导致舆情网络关键节点的缺失。

## 参考文献

- [1] 黄月琴,黄宪成.“转发”行为的扩散与新媒体赋权——基于微博自闭症议题的社会网络分析[J].新闻记者,2021(05):36-47.
- [2] 邱泽国,贺百艳.基于 PCA-Spectral-LDA 的网络舆情聚类 and 情感演进分析:一个微博文本挖掘研究[J].系统科学与数学,2021,41(10):2906-2918.
- [3] 牟冬梅,邵琦,韩楠楠,王萍,金姗,靳春妍.微博舆情多维度社会属性分析与可视化研究——以某疫苗事件为例[J].图书情报工作,2020,64(03):111-118.
- [4] 赵金楼,成俊会.基于 SNA 的突发事件微博舆情传播网络结构分析——以“4.20 四川雅安地震”为例[J].管理评论,2015,27(01):148-157.
- [5] 王晰巍,邢云菲,赵丹,李嘉兴.基于社会网络分析的移动环境下网络舆情信息传播研究——以新浪微博“雾霾”话题为例[J].图书情报工作,2015,59(07):14-22.
- [6] 陈骁,黄曙光,秦李.基于微博转发的社交网络模型[J].计算机应用,2015,35(03):638-642.
- [7] 何建民,李雪.面向微博舆情演化分析的隐马尔科夫模型研究[J].情报科学,2016,34(04):7-12.
- [8] 高芳芳,林心婕.围绕在线问诊的舆论焦点与网络情绪研究——基于微博舆情的分析[J].未来传播,2022,29(02):32-40+120-121.
- [9] 姚乐野,吴茜,李明.新冠肺炎疫情微博舆情传播的网络结构分析[J].图书情报工作,2020,64(15):123-130.
- [10] 饶浩,陈海媚.主成分分析与 BP 神经网络在微博舆情预判中的应用[J].现代情报,2016,36(07):58-62+70.
- [11] Shmargad, Y. (2018). Twitter influencers in the 2016 US congressional races. *Journal of Political Marketing*, 1-18.
- [12] Li, Y., Gao, X., Du, M., He, R., Yang, S., & Xiong, J. (2020). What causes different sentiment classification on social network services? evidence from weibo with genetically modified food in China. *Sustainability*, 12(4), 1345.

- [13] Li, T., Yu, G., & Yu, Y. (2019, January). Research on Influence Evaluation and Application of University Weibo Users. In 2nd International Conference on Social Science, Public Health and Education (SSPHE 2018) (pp. 72-76). Atlantis Press.
- [14] Chen, J., & She, J. (2012, June). An analysis of verifications in microblogging social networks--Sina Weibo. In 2012 32nd International Conference on Distributed Computing Systems Workshops (pp. 147-154). IEEE.
- [15] Huang, R., & Sun, X. (2014). Weibo network, information diffusion and implications for collective action in China. *Information, Communication & Society*, 17(1), 86-104.
- [16] Lei, K., Zhang, K., & Xu, K. (2013, December). Understanding Sina Weibo online social network: A community approach. In 2013 IEEE Global Communications Conference (GLOBECOM) (pp. 3114-3119). IEEE.
- [17] Lei, K., Liu, Y., Zhong, S., Liu, Y., Xu, K., Shen, Y., & Yang, M. (2018). Understanding user behavior in Sina Weibo online social network: A community approach. *IEEE Access*, 6, 13302-13316.
- [18] Xing, S., & Fan, Z. (2020, January). A Method for LDA-based Sina Weibo Recommendation. In Proceedings of the 2020 2nd International Conference on Big Data Engineering and Technology (pp. 54-57).
- [19] Yang, X., Xu, S., Wu, H., & Bie, R. (2019). Sentiment analysis of Weibo comment texts based on extended vocabulary and convolutional neural network. *Procedia computer science*, 147, 361-368.
- [20] Li, Y., Zhou, X., Sun, Y., & Zhang, H. (2016). Design and implementation of Weibo sentiment analysis based on LDA and dependency parsing. *China Communications*, 13(11), 91-105.