

2021-2022学年第二学期

《新媒体时代的公共关系》课程期末论文

此处为论文中文题目，要求居中填写

主标题不超过24个汉字；可加副标题（副标题前加破折号），副标题与主标题间空一行的位置

主标题：黑体，小二，居中

副标题：楷体\_GB2312，四号，居中

阅后删除此文本框。

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

**内容提要：**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、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 三号黑体字

1.1 四号黑体字

1.1.1 小四号黑体字

正文用小四号宋体字，固定行间距20磅。

表格或图表要有表（图）头：如表1 物流的概念和范围。资料来源注于表的下方。表格用五号字。表格或word图表要居中。要有脚注[[1]](#footnote-0)。要有参考文献，用小四号字。要有页码，页码居中。

表1 示例

|  |  |
| --- | --- |
| 本 质 | 过 程 |
| 途径或方法 | 规划、实施、控制 |
| 目标 | 效率、成本效益 |
| 活动或作业 | 流动与储存 |
| 处理对象 | 原材料、在制品、产成品、相关信息 |
| 范围 | 从原点（供应商）到终点（最终顾客） |
| 目的或目标 | 适应顾客的需求（产品、功能、数量、质量、时间、价格） |

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。研究现状！！！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

情感分析，或称意见挖掘，是人们的观点，情绪，评估对诸如产品，服务，组织等实体的态度。

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学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 聚类分析

聚类分析的目标是在相似的基础上收集数据来分类。聚类技术被用作描述数据，衡量不同数据源间的相似性，以及把数据源分类到不同的簇中。

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2682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。

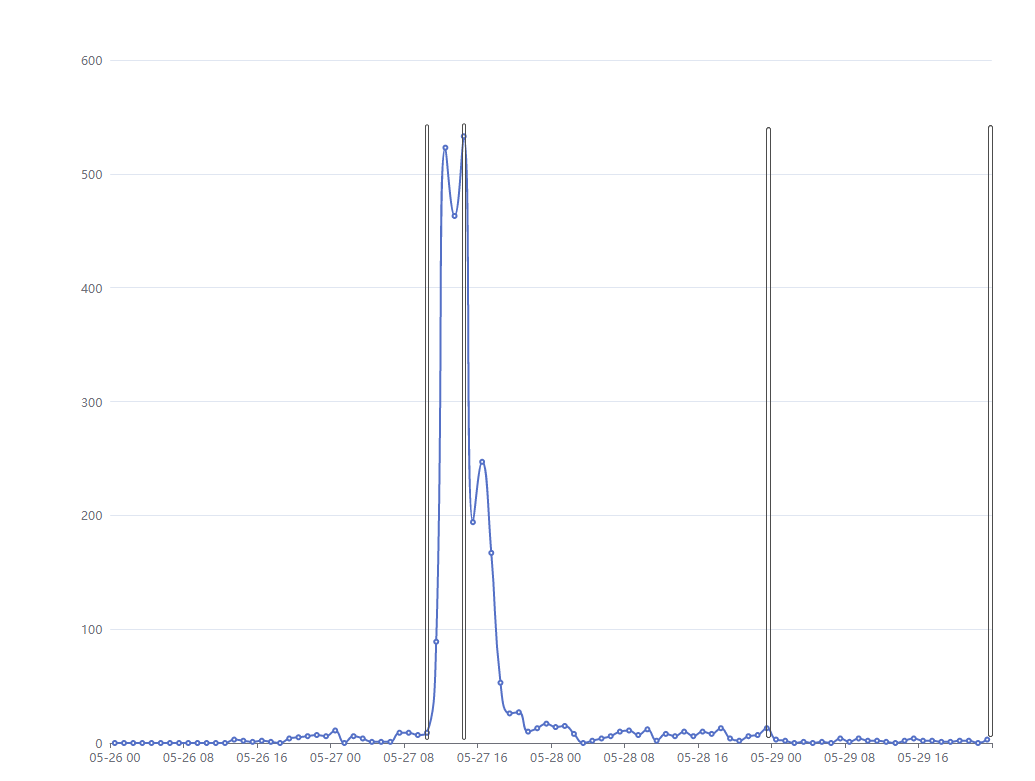


图 1

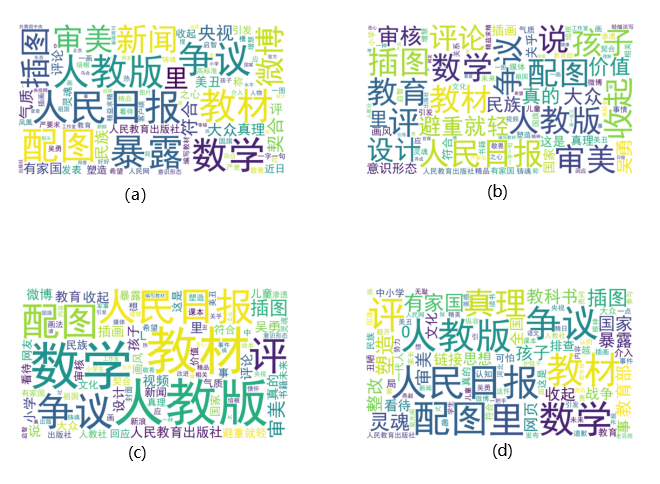


图 2

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。

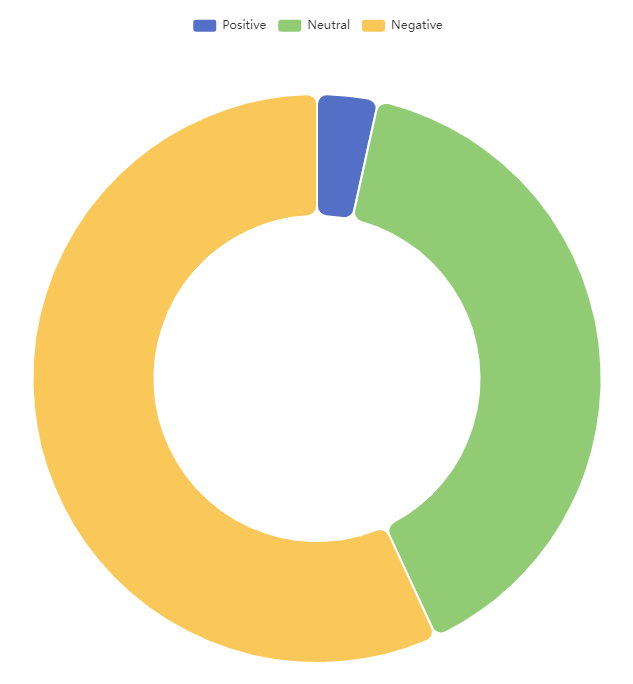


图 3

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。

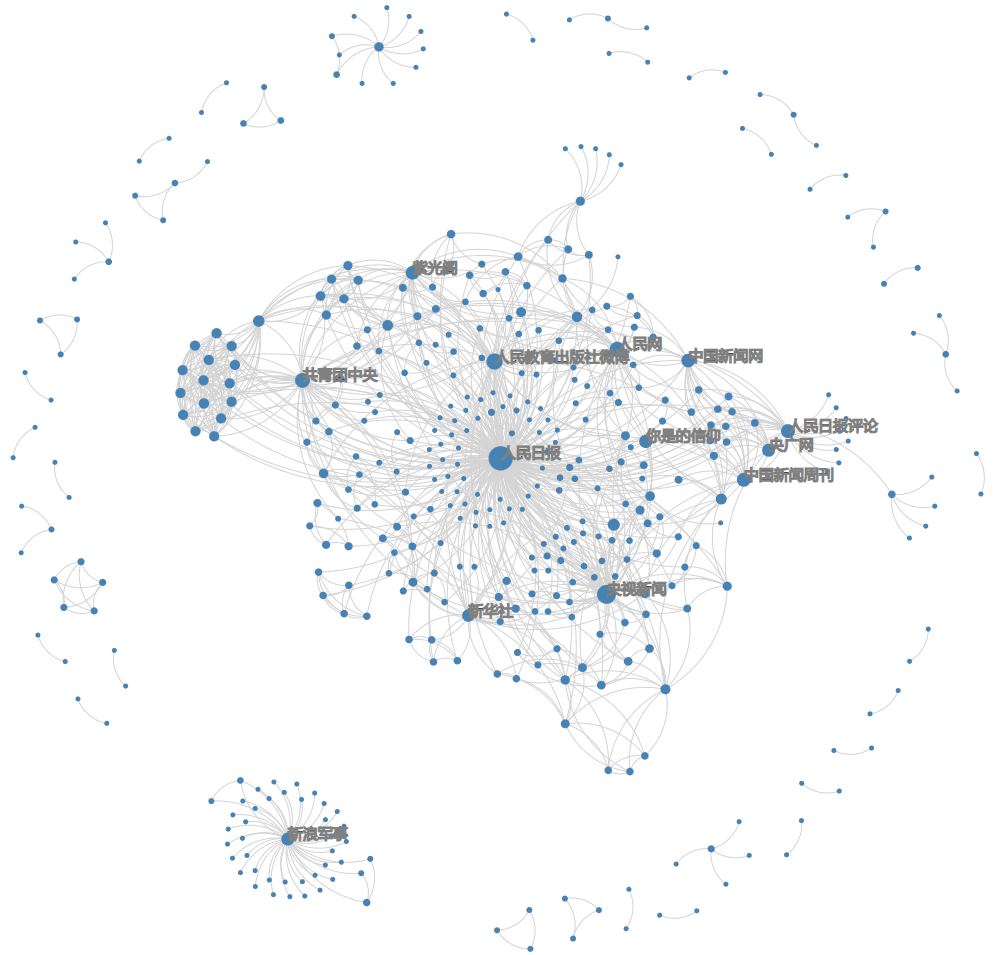


图 4

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）。

4.2.2 舆情网络的聚类分析

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构成，如“人民日报”，“人民教育出版社微博”，“央视新闻”等，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”外，还有大V和小有名气的普通用户，可以看到这些用户之间的关系极其紧密（见图6）。

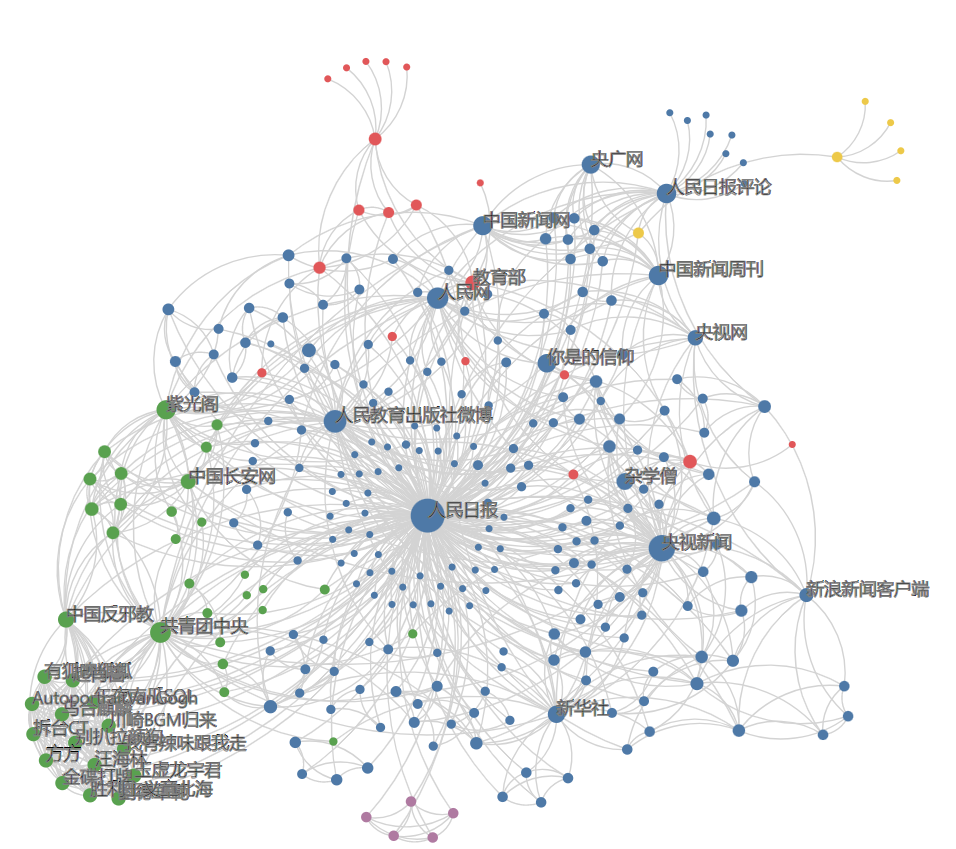


图 5

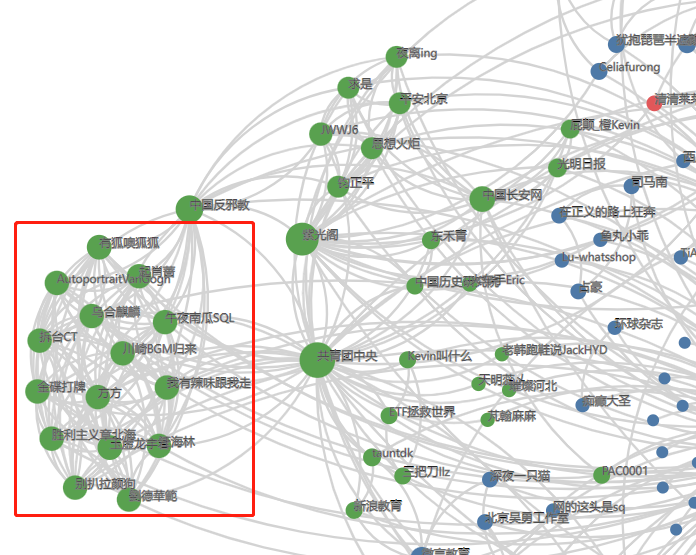


图 6

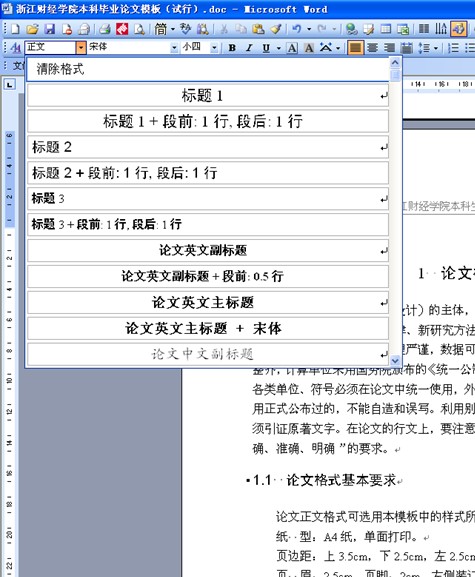


图1 样式

参考文献

[1] 蒋有绪, 郭泉水, 马娟等. 中国森林群落分类及其群落学特征[M]. 北京: 科学出版社, 1998.

[4] 张志祥. 间断动力系统的随机扰动及其在守恒律方程中的应用[D]. 北京: 北京大学数学学院, 1998.

[5] 辛希孟. 信息技术与信息服务国际研讨会论文集: A集[C]. 北京: 科学出版社, 1994.

[6] 白书农. 植物开花研究[M]//李承森. 植物科学进展. 北京: 高等教育出版社, 1998: 146-163.

[7] 李晓东, 张庆红, 叶瑾琳. 气候学研究的若干理论问题[J]. 北京大学学报:自然科学版, 1999, 35(1): 101-106.

[8] Jack, W., & Suri, T. (2014). Risk sharing and transactions costs: evidence from kenya's mobile money revolution. **American Economic Review**,104(1),183-223.

1. 参见邱风：制度创新与技术选择——中国农业经济发展的路径研究，中国财政经济出版社，2004年版，第33-35页。 [↑](#footnote-ref-0)