



开放科学
(资源服务)
标识码
(OSID)

社会安全事件网络舆情多属性演化分析与知识图谱构建

于凯¹ 杨富义²

1. 新疆财经大学公共管理学院 乌鲁木齐 830012;
2. 新疆财经大学信息管理学院 乌鲁木齐 830012

摘要: [目的/意义] 挖掘社会安全事件网络舆情的传播规律与多属性演化特征, 为网络舆情事件的应对和治理提供参考。[方法/过程] 以“货拉拉女乘客跳车事件”为例, 从舆情主题、参与主体和网民情绪三个角度进行社会安全事件网络舆情多属性演化分析, 探究舆情各阶段网民的关注焦点和情感动向; 其次基于事理逻辑构建舆情事件知识图谱, 以明确舆情事件间的关联关系和演化规律。[结果/结论] 不同阶段的舆情主题、关键主体与情感特征具有明显的区别。网络舆情多属性演化分析有助于舆情监管部门迅速把握事件的多属性演化动态, 而舆情事件知识图谱可以厘清事件传播路径、定位事件关键治理节点并还原舆情全貌。[局限] 后续研究将考虑细粒度情感分析, 同时结合事件及关系抽取模型构建舆情事件知识图谱, 扩展事件类型和数据规模。

关键词: 社会安全事件; 网络舆情; 知识图谱; 主题挖掘; 情感演化

中图分类号: G35; TP391

Multi-Attribute Evolution Analysis and Knowledge Graph Construction of Network Public Opinion on Social Security Events

YU Kai¹ YANG Fuyi²

1. School of Public Administration, Xinjiang University of Finance and Economics, Urumqi 830012, China;
2. School of Information Management, Xinjiang University of Finance and Economics, Urumqi 830012, China

Abstract: [Objective/Significance] Excavate the propagation law and multi-attribute evolution characteristics of network public opinion on social security events, so as to provide reference for the response and governance of network public opinion events. [Methods/Processes] Taking the “LaLaMove Female Passenger Jumping Incident” as an example, this paper analyzes the multi-attribute evolution of social security events’ network public opinion from the perspectives of public opinion theme, participants

基金项目 新疆维吾尔自治区自然科学基金项目“基于多层网络模型的信息传播源头定位研究”(2019D01A22); 新疆维吾尔自治区社科基金项目“智能时代新疆重大舆情和突发事件舆论治理与引导机制研究”(21BTQ162)。

作者简介 于凯(1974-), 副教授, 研究方向为数据挖掘与知识图谱、突发事件与网络舆情; 杨富义(2000-), 硕士研究生, 研究方向为网络舆情、知识图谱, E-mail: 1782369282@qq.com。

引用格式 于凯, 杨富义. 社会安全事件网络舆情多属性演化分析与知识图谱构建 [J]. 情报工程, 2022, 8(3): 14-30.

and netizens' emotions, and explores the focus and emotional trends of netizens at all stages of public opinion; Secondly, the knowledge graph of public opinion events is constructed based on the logic of affairs to clarify the correlation and evolution law between public opinion events. [Results/Conclusions] There are obvious differences in public opinion themes, key subjects and emotional characteristics in different stages. The multi-attribute evolution analysis of network public opinion help public opinion regulators quickly grasp the multi-attribute evolution dynamics of events, and the knowledge graph of public opinion events can clarify the propagation path of events, locate the key governance nodes of events, and restore the overall picture of public opinion. [Limitations] The follow-up research will consider fine-grained emotion analysis, and combine the event and relationship extraction model to build the knowledge graph of public opinion events, so as to expand the event type and data scale.

Keywords: Social security events; network public opinion; knowledge graph; topic mining; emotional evolution

引言

随着信息技术的进步和互联网的不断普及,以微博为代表的社交媒体平台逐渐成为舆论产生和传播的主战场,其中由社会安全事件引发的网络舆情在短时间内迅速发酵极易引发各种风险,给舆情监管工作带来巨大挑战。截至2021年6月,我国网约车用户规模达3.97亿,较2020年12月增长3123万,占网民整体的39.2%^[1]。一方面,网约车平台为用户出行提供了更为便捷的服务;另一方面,由于平台监管机制存在缺陷、安全保障体系不够完善,近年来网约车安全事故频发。2018年,“乐清顺风车乘客遇害事件”、“滴滴空姐遇害事件”相继发生,造成了极大的社会负面影响和舆论压力,导致滴滴平台被全国范围内叫停并严令整改。在此背景下,网约车行业为何再次出现“货拉拉女乘客跳车事件”?研究此类舆情事件的传播规律及多属性演化特征对于政府相关部门引导舆情健康发展并预防此类事件发生具有重要意义。

面对海量、碎片化、复杂多变的网络舆情数据,传统的舆情分析方法表现出明显的局限

性,必须开拓思路创新研究更为科学的知识组织技术和智能化的知识处理技术^[2]。知识图谱作为一种语义丰富的知识表示方式,在舆情热点话题发现、趋势预测等方面具有很大的优势。其中以事件为中心的知识图谱聚焦动态事件及其间的顺承、时序和因果关系,为挖掘动态事件信息及逻辑关系、揭示事件的传播演化规律提供有力支持。本文以“货拉拉女乘客跳车事件”为例,结合知识图谱相关方法探索知识图谱与网络舆情的融合应用,以期相关部门舆情治理工作提供借鉴。

1 相关研究

1.1 社会安全事件网络舆情

突发事件网络舆情是指网民围绕即将发生或已发生的突发事件,来呈现个人或群体的情绪、态度、意见等带有个人倾向性的信息^[3]。社会安全事件网络舆情则是存在于社会安全领域的网络舆情,目前国内针对此类问题的研究主要集中于舆情特征与演变机理^[4]、舆情传播与演化规律^[5]、舆情监测与预警机制^[6]、舆情应对与治理策略^[7]等方面,部分研究成果也涉

及公众情感分析^[8]、政策议程设置等内容。陈璟浩等^[9]运用生存分析方法对70起重大社会安全事件进行分析,通过研究各类社会安全事件网络舆情的生存周期,探索了作案人数、社会地位等因素对网络舆情生存时间的影响;陈贵梧等^[10]通过构建“两阶段多源流”理论框架,以滴滴顺风车安全事件为例分析了网络舆论与公共政策之间的关系。国外学者主要关注社交网络舆情中的信息传播机制、情绪识别与群体决策等问题,并采用仿真模拟、社会网络分析、传播动力学等方法探究舆情演变特征。Kim等^[11]对网络信息传播与公共舆论之间的关系进行了研究,发现微博舆情的极化和分散与政治意识形态密切相关;Uthirapathy等^[12]提出了一种基于舆论冲击矩阵的社交媒体网络突发公共事件实时舆论预测方法;Bandhakavi等^[13]从带标记的Twitter语料库中提取用于Twitter情感分析的词汇,建立情绪语料库用于情感分类和情感强度预测。

在网络舆情的演化方面,研究者主要聚焦舆情主体、主题与网民情感的演化分析,并采用网络舆情主题图谱、情感图谱、话题图谱等形式进行可视化表示。其中网络舆情主体图谱是由网络舆情主体之间的点赞、评论、转发等行为形成的关于特定舆情事件的主体关系网络,对于精准定位舆情治理关键主体具有重要价值。情感图谱是融合网络用户社交关系与情感态度倾向的分布图^[14],构建社交用户的情感图谱有利于掌握网民情绪动向,并快速锁定重要的情绪传播节点。如张柳等^[15]基于LDA主题模型构建了微博用户主题聚类图谱,研究了网络社群间的主题传播路径。安璐等^[16]结合利益相关者的社会网络情感网络图谱分析了公众情感变

化态势。刘雅姝等^[17]通过构建舆情话题图谱以识别话题讨论内容并对话题演化过程进行追踪。王丹等^[18]从主体、客体、全景生态三个维度构建了微博舆情图谱。周红磊等^[19]基于态势感知理论构建突发公共卫生事件话题-情感图谱,探究了新冠疫情话题背后的网民情感演变。

1.2 基于知识图谱的突发事件网络舆情研究

知识图谱是揭示实体之间关系的语义网络,2012年由谷歌公司首次提出,目前已被广泛应用于语义搜索、智能问答、个性化推荐等领域^[20]。构建以舆论研判分析需求为导向的舆论知识图谱,可以提高总体态势感知能力,及时掌握各种苗头性倾向性问题^[21]。国内外学者从多个角度进行了网络舆情与知识图谱融合研究的探索,如王晰巍等^[22]利用知识图谱方法对突发事件网络舆情领域相关文献进行梳理及可视化分析,指出了该领域的研究热点、现状及发展趋势;娄国哲等^[23]提出了基于知识图谱的网络舆情知识组织架构;袁国栋等^[24]提出了基于知识发现的突发事件舆情风险演化模型。有研究通过构建舆情事理图谱并对舆情进行推演及预测,如田依林等^[25]在事件泛化的基础上构建了事理图谱来剖析新冠相关舆情的发展趋势和演化路径;Hernes等^[26]提出了一种金融事件知识的语义表示方法,通过自动处理和分析金融事件的意义以辅助决策。其他研究成果则利用知识图谱辅助网络舆情监测、研判和治理,如马哲坤等^[27]提出了一种基于知识图谱的网络舆情监测方法;张思龙等^[28]对基于知识图谱的网络舆情研判系统的构建实践和应用进行了探讨;安宁等^[29]分别构建了微博和抖音平台的知识图谱并进行对

比分析,发现不同平台上舆情信息内容和知识分布存在差异。

综上所述,当前国内外学者主要采用案例分析、仿真模拟、生存分析、复杂网络分析等方法对网络舆情进行研究,而网络舆情与知识图谱的融合研究集中于舆情事件主题与情感演化、舆情趋势预测、舆情知识组织方法、舆情生态等方面,构建的图谱主要有主题图谱、情感图谱、多维图谱和事理图谱。针对某一特定类型突发事件网络舆情的知识图谱构建研究较少,所涉及的方法难以适用于其他类型的突发事件网络舆情研究。此外,使用事理图谱进行舆情推演与预测的研究多基于事件之间的顺承和因果关系,对于舆情事件时空关系和用户情感变化趋势不能很好的表达。

针对以上情况,本文首先结合舆情生命周期理论,从舆情主题挖掘、舆情主体关系分析、

网民情感演化三个角度进行社会安全事件网络舆情多属性演化分析,以探究网络舆情的传播规律与阶段性演化特征。其次,基于事件演化逻辑构建网络舆情事件知识图谱,以揭示事件发展脉络和演化路径,定位关键事件管控节点并还原舆情事件全貌。最后,以“货拉拉女乘客跳车事件”为例,采用LDA主题模型、社会网络分析、情感分析方法针对实际案例展开研究,在此基础上提出网络舆情治理和引导策略,以期舆情监管部门提供借鉴。

2 研究过程

2.1 研究框架

本文的研究过程由数据来源与处理、网络舆情事件多属性演化分析和舆情事件知识图谱构建三部分组成,具体研究框架如图1所示。

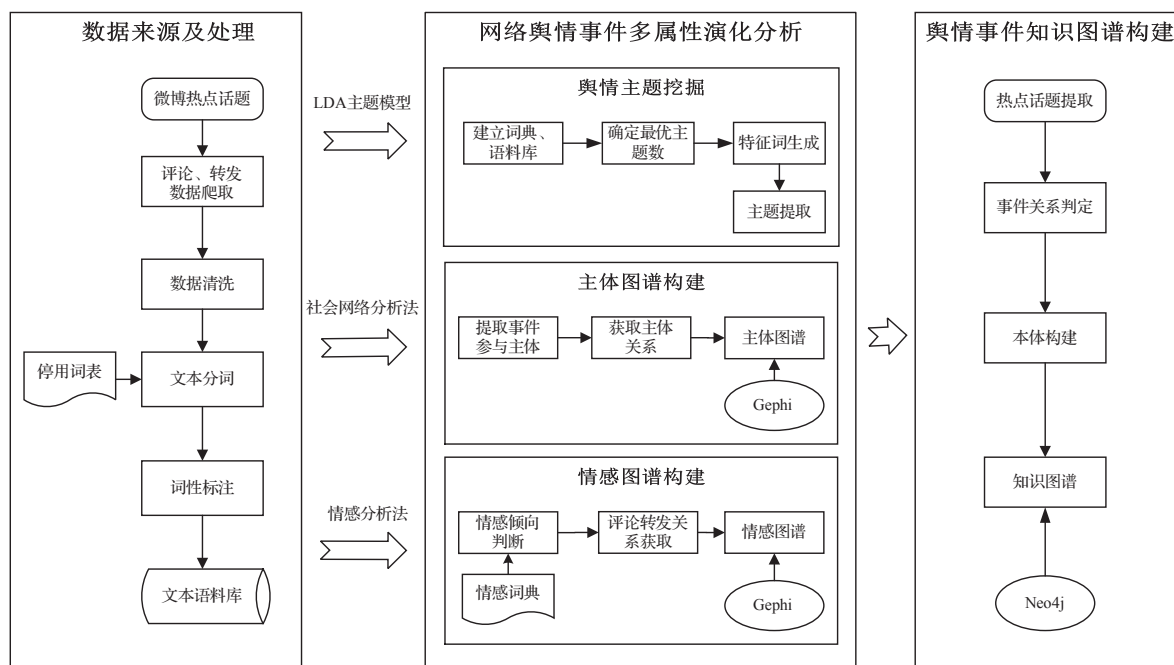


图1 研究框架

2.2 LDA主题模型构建

LDA 主题模型由 Blei 等于 2003 年提出, 其采用词袋方法, 将每篇文档视为一个词频向量, 考虑潜在的语义信息, 并以概率分布的形式给出每篇文档的主题^[30]。记每个舆情阶段的评论集为 $D=(d_1, d_2, \dots, d_m)$, 其中包含 m 篇博文评论数据, 评论集中共有 k 个主题, 记为 $Z_i, i=(1, 2, \dots, k)$, 每条评论由 n 个词语构成, 记为 $d(w_1, w_1, w_n)$, 主题分布和词分布都服从多项式分布, 分别记为:

$$Z | (\theta) = \text{Multinomial}(\theta) \quad (1)$$

$$W | Z, \phi = \text{Multinomial}(\phi) \quad (2)$$

则在评论 d_j 条件下生成词 w_i 的概率表示为:

$$P(w_i | d_j) = \sum_{s=1}^K P(w_i | z=s) * P(z=s | d_j) \quad (3)$$

其中, $P(w_i | z=s)$ 表示词 w_i 属于第 s 个主题的概率, $P(z=s | d_j)$ 表示第 s 个主题在评论 d_j 中的概率。

其次采用基于余弦相似度的自适应最优 LDA 模型选择方法来确定最优主题数, 具体步骤如下:

①取初始主题数为 k , 得到初始模型, 计算各主题之间的相似度, 计算公式如下:

$$\text{similarity} = \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}} \quad (4)$$

其中 A_i, B_i 分别代表向量 A 和 B 的各分量。

②增加或减少 k 值, 重新训练模型, 再次计算各主题之间的相似度。

③重复步骤②直到得到最优 k 值。

2.3 网络舆情事件知识图谱构建

事件是人类社会的核心概念之一, 人们的社会活动往往是事件驱动的, 网络舆情管控不仅需要掌握网络舆情的演化特征、演化规律及传播特点, 而且需要准确地分析舆情事件之间的关联关系, 把握舆情的演化路径^[31]。本文在构建网络舆情事件知识图谱时采用自顶向下的构建方式, 先为知识图谱定义好本体^[32]与数据模式, 再将实体加入到知识库中。具体步骤如下:

- (1) 构建网络舆情事件本体;
- (2) 判定事件中各热点话题之间的逻辑关系;
- (3) 提取舆情事件的所有参与主体;
- (4) 将热点话题间的逻辑关系与实体间的关系以三元组的形式进行整理, 并通过代码导入 Neo4j 图数据库中。

采用 protégé 软件构建网络舆情事件本体, 结果如图 2 所示, 该本体包括热点话题、媒体、一级评论者、二级评论者和转发用户 5 类实体, 如表 1 所示, 实体之间的关系主要有报道、评论、回复和转发。图 2 中事件实体即舆情事件相关的微博热点话题, 事件间的逻辑关系包括顺承、因果、上下位和转折关系, 如表 2 所示。

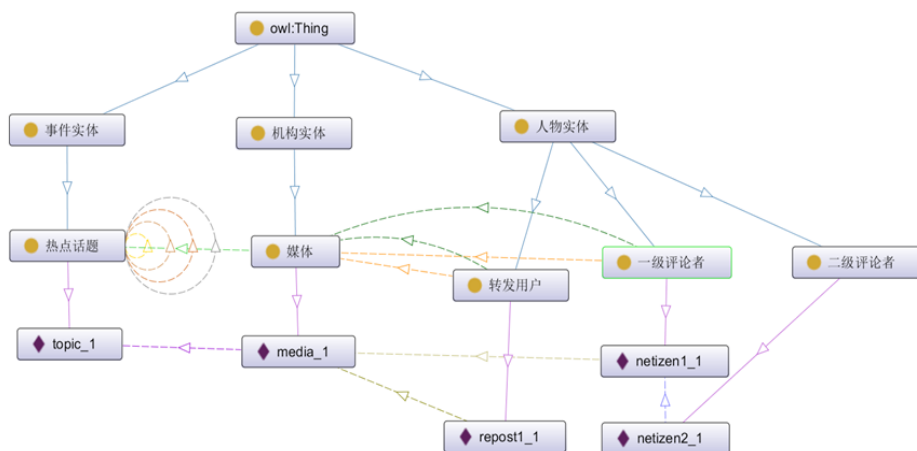


图2 网络舆情事件本体结构图

表1 网络舆情事件本体中的实体类别

实体类别	属性	实例
事件实体	话题ID 话题名称 阅读量 讨论量 阶段 主题	topic_1 #23岁女生在货拉拉车上跳窗身亡# 14.8亿 81000万 爆发期 女性安全
机构实体	机构ID 机构名称 粉丝量 关注数	media_1 头条新闻 1.04亿 1676
人物实体	用户ID 用户名称	netizen1_1 也桃c netizen2_1 juve吐司 repost1_1 月亮天天开心

表2 网络舆情事件本体中的关系类别

关系类型	描述	neo4j编码	实例
报道	媒体对热点话题的报道	report	media_1——report——>topic_1
评论	一级评论者对媒体发表博文的评论	comment	netizen1_1——comment——>media_1
回复	用户对一级评论者的回复	reply	netizen2_1——reply——>netizen1_1
转发	用户对博文的转发	repost	repost1_1——repost——>media_1
顺承	某事件紧接着另一事件的发生	顺承	topic_1——顺承——>topic_2
因果	某一事件导致另一事件的发生	因果	topic_1——因果——>topic_2
上下位	某事件是另一事件的上位或下位事件	上下位	topic_1——上下位——>topic_2
转折	某事件与另一事件形成对立	转折	topic_1——转折——>topic_2

3 实证分析

3.1 数据来源及处理

微博作为当前最热门和极具代表性的社交平台之一,在反映社情民意、传播网络舆情方面发挥着重要作用,较高的用户活跃度和广泛的人群覆盖面使得微博热议话题及其评论转发数据蕴含着巨大的研究价值。本文以新浪微博

为数据源,选取“货拉拉女乘客跳车事件”为研究对象,筛选该事件相关热点话题共 35 个,结合舆情生命周期理论将该事件划分为爆发期、蔓延期、二次爆发期和衰退期四个阶段,绘制舆情热度趋势如图 3 所示。在此基础上,采用八爪鱼数据采集软件爬取图 3 中话题下热门微博的所有评论及转发数据,经数据清洗后共保留评论 47195 条,转发数据 22110 条。

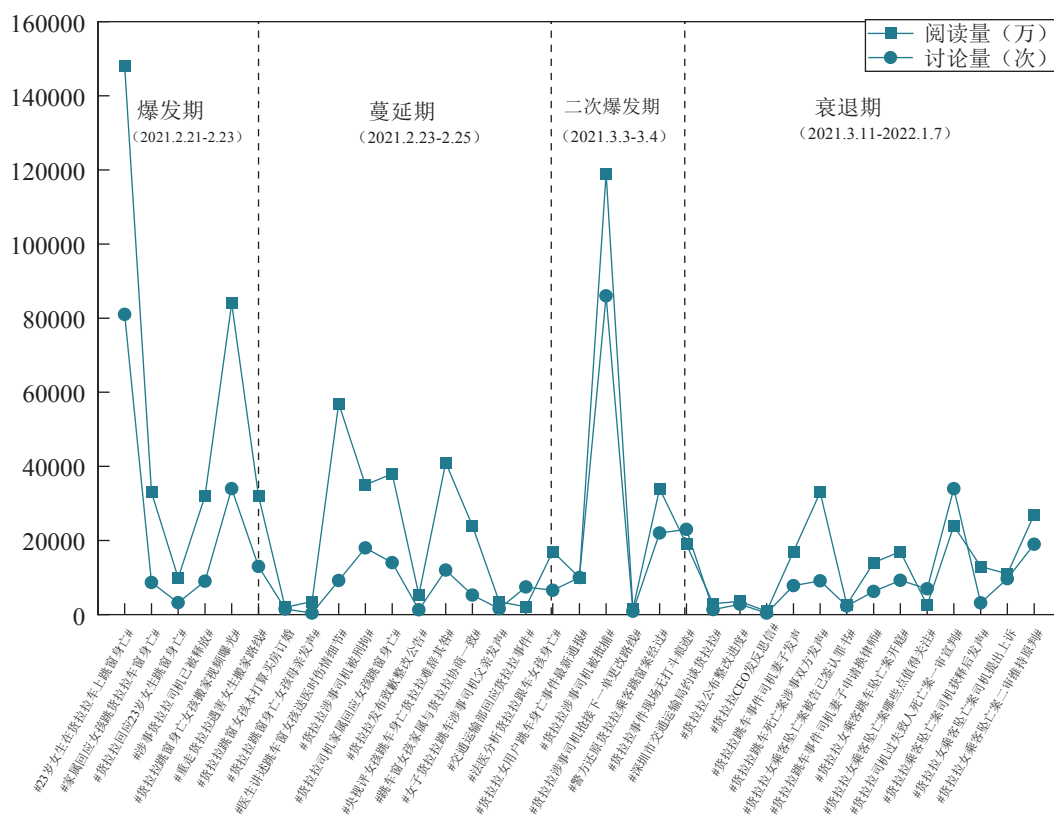


图3 “货拉拉女乘客跳车事件”舆情热度趋势图

3.2 “货拉拉女乘客跳车事件”多属性演化分析

3.2.1 各阶段舆情主题挖掘

为探究舆情演化过程中网民的关注焦点和心理动态,本文采用 LDA 模型挖掘舆情不同阶段的主题。根据 2.2 节构建的 LDA 模型,首

先将评论及转发数据进行去重合并、分词、词性标注并去除停用词,筛选部分情感词汇作为本事件专属情感词典,其次基于知网情感词典和自建词典将词语集划分为正向、中性和负向三类并分别建立对应的语料库,使用 Python 的 gensim 模块对每个舆情阶段正面、中性、负面

评论数据分别构建 LDA 主题模型，对每个主题生成 10 个最有可能出现的词语以及相应的概

率，根据特征词总结舆情各阶段网民评论关注点和情绪特征，结果如表 3 所示。

表 3 舆情各阶段主题分布

舆情阶段	特征词	正面主题	中性主题	负面主题
爆发期 (topic1-topic6)	女孩、跳车、司机、真相、搬家、自杀、希望、垃圾、危险、宠物、行车记录仪、跳窗、平台、警方、刹车、证据、猜测、侵害	逝者安息、期待真相大白、相信警方	平台整改、监控录音、女性安全	女孩跳窗自杀、司机三次偏航、货拉拉倒闭
蔓延期 (topic7-topic17)	老实人、希望、女孩、跳窗、法医、平台、赔偿、谅解、家属、公道、弟弟、监控、可怕、凶手、尸体、自杀、刹车、威胁	女孩家属发声、司机被刑拘、平台赔偿	女孩拒付运费、弟弟制造舆论、货拉拉致歉、	严惩凶手、犯罪嫌疑、司机言语威胁
二次爆发期 (topic18-topic22)	司机、平台、超时、赚钱、法律、搬家、尊重、过失、悲剧、态度、报警、沟通、跳车、理解、服务、拒绝	互相尊重、女孩月薪过万、人情世故	沟通重要性、超时搬运、司机不容易	态度恶劣、女孩过度恐慌、司机过失致死
衰退期 (topic23-topic35)	司机、舆论、支持、男女对立、羁押、无罪、跳车、判决、上诉、律师、倒霉、法院、依法治国、司法、长沙、正义	支持司机上诉、依法判决、程序合法	法院指定法律援助、案件未公开审理、平台整改措施	舆论定罪、男女对立、司机被超期羁押

结合图 3 与表 3，对舆情各阶段主题挖掘结果分析如下：

事件爆发期：“女孩跳窗坠亡”、“货拉拉司机被释放”等话题引发广泛讨论，该时期由于信息不完整，网民主要围绕事件真相进行猜测或推理。讨论关键词集中于“行车记录仪”、“性侵”、“真相”、“平台”等，表达对女孩遭遇的同情与惋惜、对货拉拉司机的愤怒和谴责、对平台审核机制、监管力度以及办事效率的质疑，体现出对网约车市场的不信任，同时也引发了对女性安全的担忧。此阶段官方仍未给出回复，舆情热度持续上升并衍生出多个热搜话题。

事件蔓延期：随事件的持续发酵，更多细节被曝光，长沙高新公安、货拉拉平台等有关部门纷纷回应并采取措施，司机再次被刑拘，公众情绪得到一定程度的稳定。评论特征词涉及“司机”、“平台”、“法医”、“涉事双方家属”等多个对象，讨论焦点主要有女孩家

属的安抚与赔偿、事件证据和真相的继续挖掘以及对平台的问责。

二次爆发期：在警方公布事件调查结果后，网民关注度再次上升，舆情二次爆发。此阶段高频关键词为“沟通”、“服务态度”、“过失致死”、“悲剧”等，即网民主要关注女孩和司机的职业和背景、二人沟通过程中存在的问题以及对事件真相的分析，与爆发期相比，网民的关注点更加宽泛，对事件的思考也更全面。

衰退期：随着信息的充分披露及监管部门的介入，货拉拉道歉并进行整改，司机入狱，事件整体关注度有所下降，但舆情仍在半年多的时间内存在反复趋势。2022 年 1 月 7 日，“货拉拉女乘客坠亡案二审维持原判”再次引发热议，网民对于审判结果的态度呈现出两极分化的态势。此阶段网民讨论焦点主要是司机是否应该上诉以及法律程序的公平问题，同时涉及一些男女对立话题的讨论，评论特征词表现为

“司法正义”、“支持上诉”、“羁押”、“舆论”等。

根据 LDA 主题分析结果可以看出,网民针对本次事件的关注范围在不断扩大,涉及主体在不断增多,争议的焦点也在不断深化。从最初的就事论事,关注司机、女孩、货拉拉平台等主要涉事主体,到质疑网约车行业准入门槛、抵制货拉拉;从试图还原真相,到深挖女孩职业背景及家庭关系,引发“女孩弟弟故意制造舆论”、女孩职业是“杀猪盘”、“女性出行安全”、“资本压榨底层劳动人民”、“性别对立”、“法律公正”等一系列讨论;从探究事件发生原因到质疑案件审理结果,不断发酵

出新的热议话题。

3.2.2 舆情事件主体关系分析

网络舆情主体即通过网络表达意见的个人或组织机构,主要包括普通网民、意见领袖、媒体、政府机构等^[33]。舆情主体推动舆情事件的传播和发展,并决定事件的发展方向,各主体之间的关注、点赞、评论、转发等行为形成特定舆情事件的主体关系网络。本文将处理后的评论转发数据导入 Gephi 软件绘制“货拉拉女乘客跳车事件”舆情主体图谱,其中节点包括主流媒体、一级评论者、二级评论者和转发者,节点之间的关系包括评论、回复和转发,结果如图 4 所示。

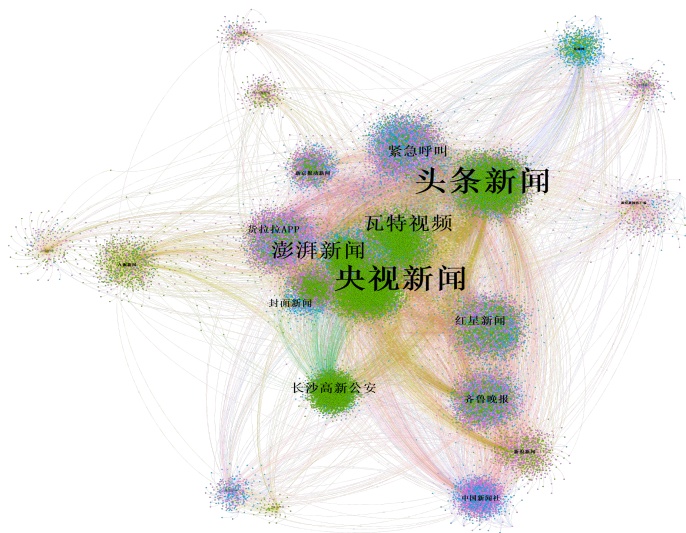


图 4 “货拉拉女乘客跳车事件”舆情主体图谱

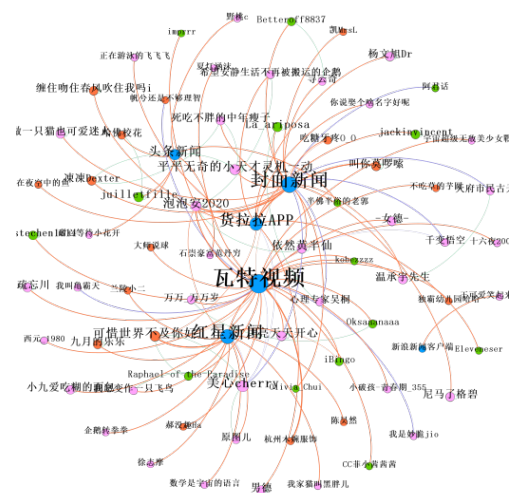
该主体图谱由 74213 个节点, 87313 条连边组成, 其中紫色连边表示评论关系, 橙色表示转发关系, 绿色表示回复关系。从结构上来看, 该网络平均度为 1.177, 网络直径为 6, 模块度为 0.809, 社团数量为 20。图中较大聚簇表示

在事件传播过程中影响力较大的媒体, 如“头条新闻”、“央视新闻”、“澎湃新闻”、“瓦特视频”、“长沙高新公安”等, 他们对事件的报道引起了大量网民的关注和讨论, 对事件舆论走势和网民情感态度的变化有着重要影响。

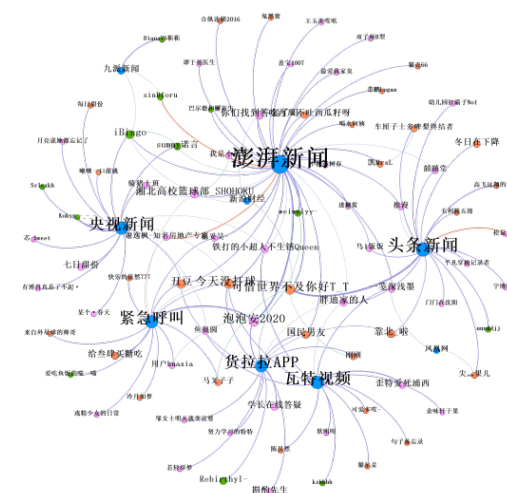
其次,“新京报动新闻”、“九派新闻”、“环球网”、“新浪财经”等媒体也是舆情传播过程中的重要节点。网络中度最大的节点为“头条新闻”,其出度为11165,此外大部分节点的度较小,表明大多数网民主要通过主流媒体的报道以了解事件动向,但对于事件的参与程度并不高,影响力也有限。

在微博舆情演化过程中,前排热门评论往往会为舆论后续的讨论奠定基调和方向,对于

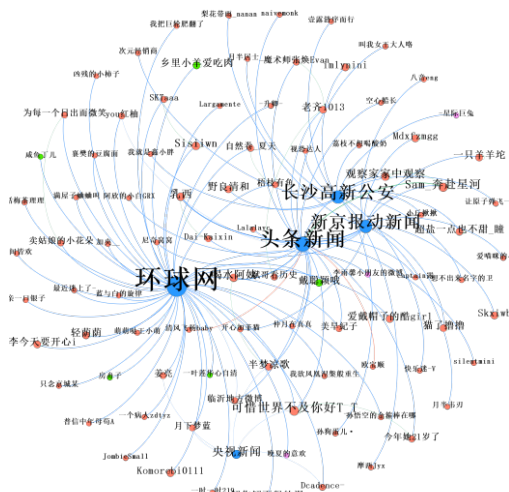
舆情趋势的判断具有参考价值。为进一步明确各阶段的关键主体及其相互关系,本文选取了度较大的前100个用户绘制不同阶段的关键主体图谱,结果如图5所示。由图5可知,不同阶段的关键主体差异较大,但均呈现以主流媒体为核心,普通网民交互影响的规律。事件爆发期和二次爆发期主体图谱的节点分布较为集中,而蔓延期和衰退期网络社团数量更多,网络结构更分散。



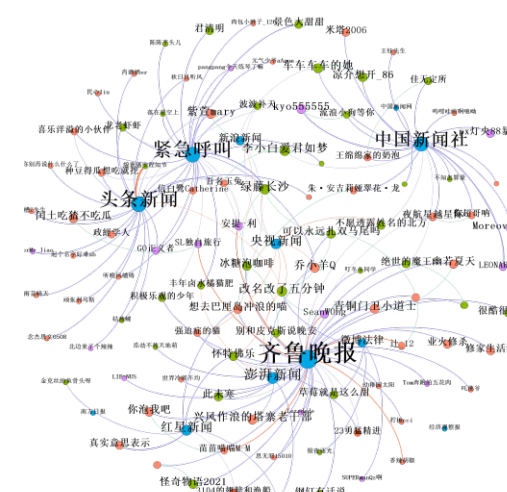
a.爆发期关键主体图谱



b.蔓延期关键主体图谱



c.二次爆发期关键主体图谱



d.衰退期关键主体图谱

图5 各阶段舆情关键主体图谱

3.2.3 网络舆情情感演化图谱分析

对网络舆情中的评论数据进情感分析,可以分析舆情发展过程中不同情绪类别的分布情况及其变化趋势,以追踪舆情发展动态,掌握网民信息诉求。在获取“货拉拉女乘客跳车事件”微博评论转发关系网络的基

础上,划分网民的情绪类型并嵌入社会关系网络中,以探究网民情绪的动态变化。采用基于词典的情感分类方法将评论和转发数据划分为正面、中性和负面三种类型,计算舆情各阶段中每种情感类型的占比,结果如图6所示。

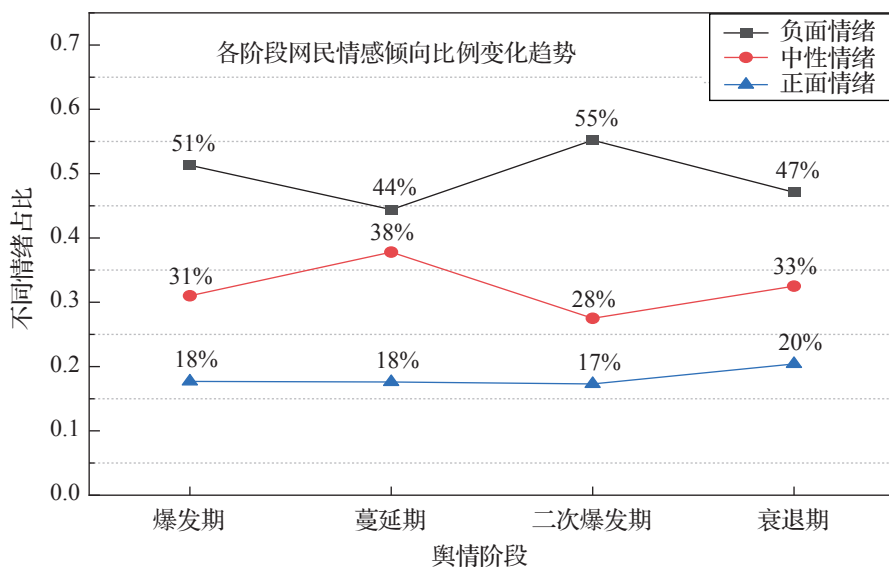


图6 各阶段网民情感倾向比例变化趋势

由图6可知,在整个舆情周期内,负面情绪所占比例最高,并在舆情二次爆发时达到最高点。正面情感所占比例最低且变化幅度很小,最大值仅为20%,表明较少数网民能理性对待本次事件并保持乐观态度。中性情感占比位于正负二者之间,其变化趋势与负向情感相反,在蔓延期达到峰值。总体来看,网民针对此次事件的情感整体较为消极,体现了“货拉拉事件”对公众情感和社会信任造成的不良影响。

为直观展示“货拉拉女乘客跳车事件”舆情情感演化动态,通过 Gephi 绘制基于舆情周期的情感演化图谱。如图7所示,图中节点表

示媒体及参与舆情事件微博用户,节点颜色表示该微博用户的情感类别,红色代表负向情感、绿色代表正向情感、黄色代表中性情感;节点间的连边表示微博用户的关联关系,紫色表示转发关系、绿色表示回复关系、橙色表示评论关系。

如图7a所示,爆发期情感图谱由15979个节点,17297条连边组成,网络平均度为1.802,社团数量为17,网络直径为3。网络整体呈现出以“瓦特视频”、“封面新闻”、“红星新闻”等网络媒体为核心节点高度聚集的态势,且红色节点居多,表明在爆发期少数热点话题引发高度关注,网民愤怒、惋惜等消极情绪集中爆发。

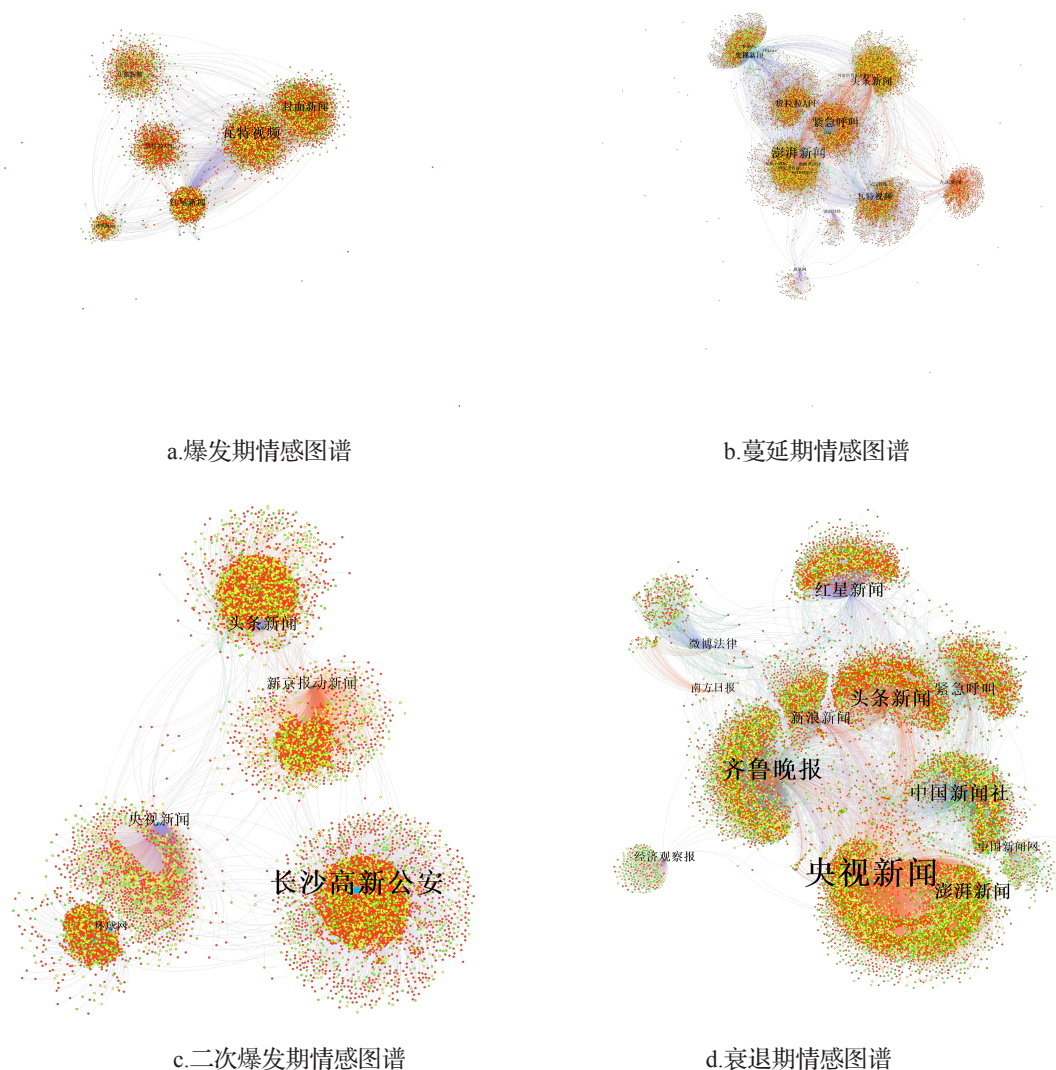


图7 各阶段舆情情感图谱

如图7b所示，蔓延期情感图谱由21245个节点，23435条连边组成，网络平均度为1.103，社团数量为41，网络直径为4。整体结构较爆发期更为发散，社团数量更多，表明舆情热度虽有所下降但仍在蔓延。除“澎湃新闻”、“紧急呼叫”等媒体之外，主流媒体“央视新闻”开始介入并表明立场，“货拉拉APP”官方发表致歉整改声明，此阶段网民负面情绪由51.3%

下降至44.6%，正面情绪占比则上升约7%。

如图7c所示，二次爆发期情感图谱由11299个节点，12010条连边组成，网络平均度为1.056，社团数量为56，网络直径为3。警方通报案件细节后，“长沙高新公安”、“央视新闻”、“头条新闻”等主流媒体成为新的核心节点，这类媒体公信力与影响力更强，其对舆论引导及事态发展起到关键作用。此阶段网民负面情

绪达到峰值，而正面情绪占比最低。

如图 7d 所示，舆情衰退期情感图谱由 23878 个节点，27120 条连边组成，网络平均度为 1.136，社团数量为 11，网络直径为 4。在衰退期舆情表现出发散趋势，公众情绪趋于平稳，热点话题数量较多但关注度整体不高。核心节点以主流媒体居多，此阶段网民消极情绪下降，正面和中性情绪占比有所上升。其中，以“中

国新闻网”、“中国新闻社”、“微博法律”为核心的团簇中绿色节点所占比例较高。

3.3 网络舆情事件知识图谱构建

基于上述多维分析结果，结合 2.3 节中的网络舆情事件知识图谱构建流程，使用 Neo4j 建立“货拉拉女乘客跳车事件”知识图谱。部分数据导入代码如表 4 所示。

表 4 Neo4j 中实体和关系构建代码（部分）

导入对象	实例	Cypher 代码
媒体节点表	media_1, 头条新闻, 媒体, 1.05 亿, 1678	LOAD CSV WITH HEADERS FROM "file:///media.csv" AS line Create(event:Media{media_id:line.media_id, name:line.name, label:line.label, 粉丝量:line.粉丝量, 关注数:line.关注数})
热点话题表	topic_1, #23 岁女生在货拉拉车上跳窗身亡#, 热点话题, 14.8 亿, 81000, 爆发期, 女性安全	LOAD CSV WITH HEADERS FROM "file:///topic.csv" AS line Create(event:Topic{topic_id:line.topic_id, name:line.name, label:line.label, 阅读量:line.阅读量, 讨论量:line.讨论量, 所属阶段:line.stage, 主题:line.topics})
报道关系表	media_1, topic_1, 报道, directed	LOAD CSV WITH HEADERS FROM "file:///media-topic.csv" AS line with line match(from:Media{media_id:line.start}) merge(to:Topic{topic_id:line.end}) merge(from)-[r:report{relation:line.relation, type:line.type}]->(to);

为获取每个事件子话题的关键主体，结合舆论传播的“前 10 效应”^[34]，选取每个热点话题下热门微博的前 10 个一级评论者和转发用户参

与事件知识图谱的构建。考虑到二级评论的重要性及数量通常低于一级评论，故选取一级评论下前 5 个二级评论者加入图谱中，结果如图 8 所示。

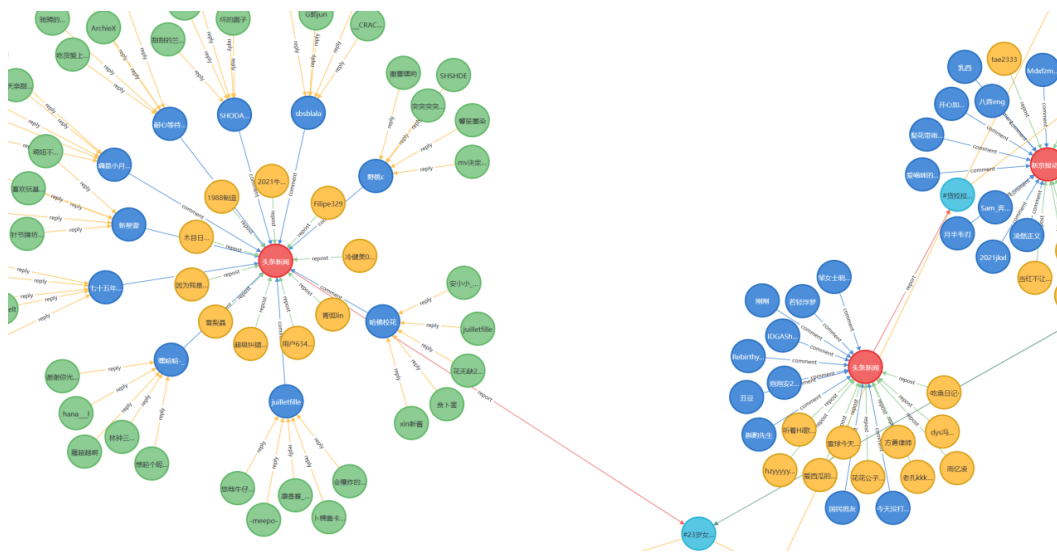


图 8 “货拉拉女乘客跳车事件”网络舆情事件知识图谱（部分）

如图8所示,“货拉拉女乘客跳车事件”知识图谱由2104个实体、2106条关系组成,图中红色节点代表媒体,浅蓝色节点代表热点话题,深蓝色、绿色和黄色节点分别代表一级评论、二级评论和转发用户。根据所构建的舆情事件

知识图谱,可明确事件的整体脉络、各话题的基本信息以及重要参与主体。使用Cypher语言可以查询与某一特定话题相关联的所有实体和关系,有利于事件复盘分析与关键用户的精准引导。本事件各子话题间的演化逻辑如表5所示。

表5 各子话题演化逻辑

start	end	relation
topic_1: #23岁女生在货拉拉车上跳窗身亡#	topic_3: #货拉拉回应23岁女生跳窗身亡#	顺承
	topic_4: #涉事货拉拉司机已被释放#	顺承
topic_3: #货拉拉回应23岁女生跳窗身亡#	topic_12: #货拉拉发布致歉整改公告#	顺承
	topic_2: #家属回应女孩跳货拉拉车窗身亡#	顺承
topic_4: #涉事货拉拉司机已被释放#	topic_8: #货拉拉跳窗身亡女孩母亲发声#	顺承
	topic_10: #货拉拉涉事司机被刑拘#	顺承
topic_5: #货拉拉跳窗身亡女孩搬家视频曝光#	topic_6: #重走货拉拉遇害女生搬家路线#	顺承
topic_6: #重走货拉拉遇害女生搬家路线#	topic_21: #警方还原货拉拉乘客跳窗案经过#	顺承
	topic_5: #货拉拉跳窗身亡女孩搬家视频曝光#	顺承
topic_8: #货拉拉跳窗身亡女孩母亲发声#	topic_7: #货拉拉跳窗女孩本打算买房订婚#	上下位
	topic_9: #医生讲述跳车窗女孩送医时伤情细节#	顺承
topic_9: #医生讲述跳车窗女孩送医时伤情细节#	topic_17: #法医分析货拉拉跟车女孩身亡#	顺承
topic_10: #货拉拉涉事司机被刑拘#	topic_11: #货拉拉司机家属回应女孩跳窗身亡#	因果
	topic_15: #女子货拉拉跳车涉事司机父亲发声#	顺承
topic_11: #货拉拉司机家属回应女孩跳窗身亡#	topic_14: #跳车窗女孩家属与货拉拉协商一致#	顺承
	topic_13: #央视评女孩跳车身亡货拉拉难辞其咎#	顺承
topic_12: #货拉拉发布致歉整改公告#	topic_24: #货拉拉公布整改进度#	顺承
topic_13: #央视评女孩跳车身亡货拉拉难辞其咎#	topic_16: #交通运输部回应货拉拉事件#	顺承
topic_16: #交通运输部回应货拉拉事件#	topic_23: #深圳市交通运输局约谈货拉拉#	顺承
topic_18: #货拉拉女用户跳车身亡事件最新通报#	topic_19: #货拉拉涉事司机被批捕#	因果
topic_19: #货拉拉涉事司机被批捕#	topic_26: #货拉拉跳车事件司机妻子发声#	因果
topic_20: #货拉拉涉事司机抢接下一单更改路线#	topic_1: #23岁女生在货拉拉车上跳窗身亡#	因果
	topic_20: #货拉拉涉事司机抢接下一单更改路线#	上下位
topic_21: #警方还原货拉拉乘客跳窗案经过#	topic_22: #货拉拉事件现场无打斗痕迹#	上下位
	topic_18: #货拉拉女用户跳车身亡事件最新通报#	顺承
topic_23: #深圳市交通运输局约谈货拉拉#	topic_24: #货拉拉公布整改进度#	因果
topic_24: #货拉拉公布整改进度#	topic_25: #货拉拉CEO发反思信#	顺承
topic_26: #货拉拉跳车事件司机妻子发声#	topic_27: #货拉拉跳车死亡案涉事双方发声#	顺承
topic_27: #货拉拉跳车死亡案涉事双方发声#	topic_30: #货拉拉女乘客跳车坠亡案开庭#	顺承
topic_28: #货拉拉女乘客坠亡案被告人已签认罪书#	topic_29: #货拉拉跳车事件司机妻子申请换律师#	转折
	topic_32: #货拉拉司机过失致人死亡案一审宣判#	顺承
	topic_28: #货拉拉女乘客坠亡案被告人已签认罪书#	顺承
topic_30: #货拉拉女乘客跳车坠亡案开庭#	topic_31: #货拉拉女乘客跳车坠亡案开庭哪些点值得关注#	上下位
	topic_32: #货拉拉司机过失致人死亡案一审宣判#	顺承
topic_32: #货拉拉司机过失致人死亡案一审宣判#	topic_33: #货拉拉乘客坠亡案司机获释后发声#	顺承
topic_33: #货拉拉乘客坠亡案司机获释后发声#	topic_34: #货拉拉女乘客坠亡案司机提出上诉#	转折
topic_34: #货拉拉女乘客坠亡案司机提出上诉#	topic_35: #货拉拉女乘客坠亡案二审维持原判#	顺承

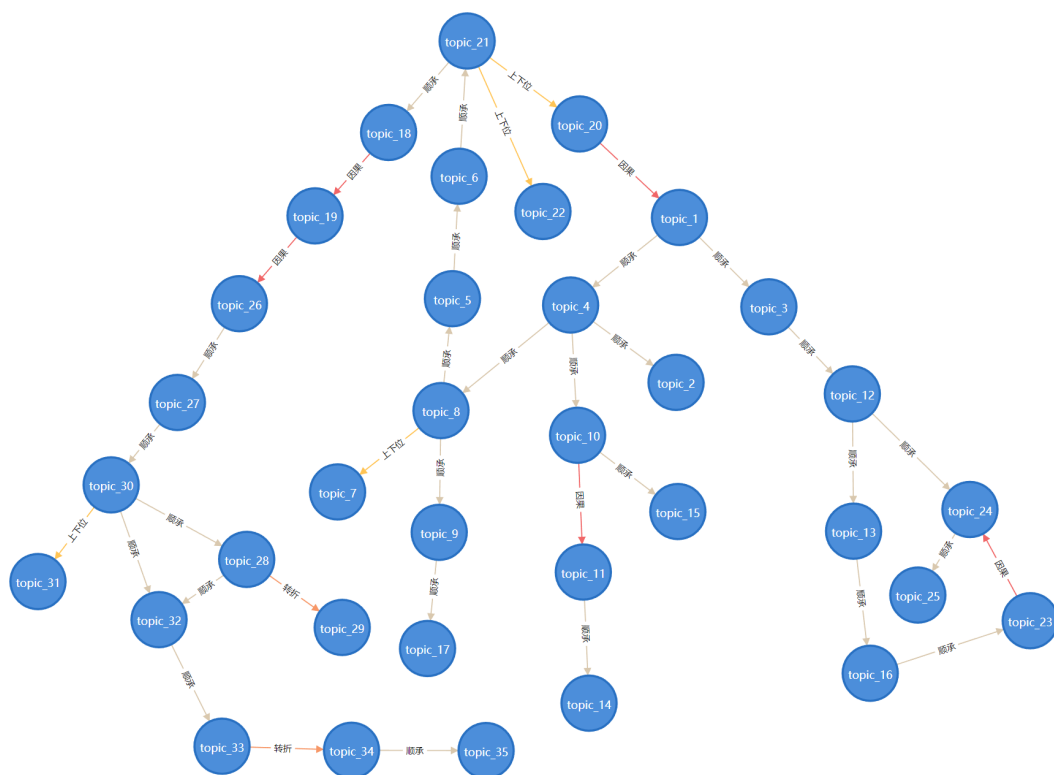


图9 “货拉拉女乘客跳车事件”各子话题演化路径

如图9所示,“货拉拉女乘客跳车事件”主要围绕警方、涉事双方父母以及货拉拉平台3个链条进行传播,事件之间的逻辑关系多为顺承关系,同时也涉及因果、转折和上下位关系。图中 topic_4:“# 涉事货拉拉司机已被释放#”、topic_8:“# 货拉拉跳窗身亡女孩母亲发声#”、topic_21:“# 警方还原货拉拉乘客跳窗案经过#”和 topic_30:“# 货拉拉女乘客跳车坠亡案开庭#”为事件传播演化的关键节点,在这些节点处,事件产生多个分支话题,极易引发衍生舆情,有关部门应予以重点关注并及时引导。

4 管理启示

根据以上多属性演化分析结果可以看出,不同舆情阶段的舆情主题、参与主体和网民情

感具有明显的差异。从舆情主题角度来看,各阶段舆情的主题均围绕涉事主体、货拉拉平台、警方和法院而产生,但随着舆情事件信息的不断披露和有关部门的相继介入,网民对事件的认知更加全面,关注主题更广泛,思考也逐渐理性而深入;从参与主体角度来看,舆情事件主体图谱整体呈现以主流媒体为传播中心,涉事平台、警方、有关部门、网络媒体等多方共同参与,意见领袖与普通网民交互影响的复杂态势。其中主流媒体和官方机构的影响力和公信力较强,而大部分普通网民的参与度和影响力有限;从网民情感角度来看,在整个舆情周期内网民负面情绪占主导地位,并在舆情爆发期和二次爆发期得到集中宣泄,而在舆情蔓延期和衰退期,网民情绪相对平缓,情绪传播网

络较为发散,网络社团数量增加,表明舆情热度有所下降,更多理性的言论和观点得到关注。

“货拉拉女乘客跳车事件”作为网约车市场的又一负面案例,暴露了网约车市场经营和监管方面存在的问题,导致了公众对各网约车平台的抵触和社会信任的缺失,让司法公信力也受到质疑,给社会带来较为严重的不良影响。纵观整个事件,愤怒、非理性甚至煽动性的言论比比皆是,如对司机与女孩的个人信息泄露与网络暴力、对货拉拉平台的大力抵制、对资本压榨底层民众的控诉以及对性别对立情绪的挑拨,这些网络舆论乱象进一步激化了社会矛盾并引发社会恐慌,背离了“清朗网络空间”的建设目标。对此类舆情事件的治理需要采取更加及时和精准的引导措施,从平台、媒体、社会公众、司机、女孩等方面进行反思并提出建设性意见。

5 结语

本文构建了社会安全事件网络舆情本体及知识图谱,结合实例从多个维度对舆情事件进行可视化分析,得出以下结论:(1)社会安全事件网络舆情的主题、参与主体及公众情感呈现出明显的阶段性特征,舆情事件的治理需要更加及时和精准的引导措施。(2)舆情热度、参与主体、主题特征与网民情感动态密切相关。舆情热度越高,越容易出现网民群体的聚集与情绪的集中爆发,而当舆情热度下降时,对应的社交网络结构更加发散、网民情绪波动幅度变小并产生多个衍生话题。(3)通过对事件演化逻辑的梳理及传播路径的可视化,可明晰舆情演化过程中的关键节点,对这些节点进行管

控有利于控制舆情后续走势。

研究结果为网络舆情的监管工作提供了参考:首先,在舆情的治理过程中,应把握事件发展的关键节点,注重主流媒体对网民情绪的引导,避免为了流量和热度挑起话题争议,引发次生舆情。其次,舆情参与主体的社交行为和情绪特征对舆情主题的演化及事件的发展具有主导作用,有关部门应实时监测网络舆情多属性演化动态,制定针对性和差异化的应对策略。具体来说可以采取以下措施:一是通过控制舆情主题引导网民关注焦点,如强化舆情各阶段正面主题的影响力,借助主流媒体对正面主题的引导来对冲负面主题,以削弱舆论的消极影响从而控制事态发展;二是通过控制各阶段关键用户的情感倾向从而影响网民整体的情绪分布;三是主动发起新的话题以转移网民的注意力,把握舆情引导主动权,持续跟进并及时回应社会关切;四是利用自然语言处理、社会网络分析和网络舆情知识图谱等方法 and 工具辅助舆情数据和处理和舆情态势的综合分析。

综上所述,结合主题、主体与情感的多属性演化分析有助于舆情监管部门迅速把握事件的多属性演化动态,而舆情事件知识图谱可以厘清事件传播路径、定位事件关键管控节点并还原舆情全貌,对有关部门的舆情分析与治理工作提供了新的思路。未来研究考虑对网民情感进行细粒度分析,同时结合事件及关系抽取模型构建网络舆情知识图谱,扩展事件类型和数据规模。

参考文献

- [1] 中国互联网络信息中心(CNNIC).第48次中国互联网络发展现状统计报告[R].中国互联网络信

- 息中心, 2021.
- [2] 王兰成, 娄国哲. 基于知识图谱的网络舆情管理方法与实践研究[J]. 情报理论与实践, 2020, 43(6):97-101.
- [3] 王睿, 张恩普, 李婷. 基于微博的突发事件网络舆情对策研究[J]. 情报科学, 2016, 34(4):94-98.
- [4] 付仁德. 社会安全事件网络舆情的演变机理及对策研究[D]. 北京: 中国人民公安大学, 2017.
- [5] 兰月新, 曾润喜. 突发事件网络舆情传播规律与预警阶段研究[J]. 情报杂志, 2013, 32(5):16-19.
- [6] 王雷. 社会安全事件智能监测与预警系统及综合平台[J]. 科学技术与工程, 2021, 21(24):10127-10133.
- [7] 王仕勇. 大数据时代的社会舆情治理: 何以可能与何以可为[J]. 重庆社会科学, 2021(12):84-95.
- [8] 习海旭, 蒋红芬, 程志凡, 等. 特定事件下网络舆情的情感分析与可视化方法[J]. 情报理论与实践, 2020, 43(9):132-136.
- [9] 陈璟浩, 李纲. 突发社会安全事件网络舆情演化的生存分析——基于 70 起重大社会安全事件的分析[J]. 情报杂志, 2016, 35(4):70-74.
- [10] 陈贵梧, 林晓虹. 网络舆论是如何形塑公共政策的? 一个“两阶段多源流”理论框架——以顺风车安全管理政策为例[J]. 公共管理学报, 2021, 18(2):58-69.
- [11] Kim K, Baek Y M, Kim N. Online news diffusion dynamics and public opinion formation: a case study of the controversy over judges' personal opinion expression on SNS in Korea[J]. Social Science Journal, 2015, 52(2):205-216.
- [12] Uthirapathy S E, Sandanam D. Real-Time opinion prediction method for emergency public events in social media networks using opinion hit matrix[J]. Revue d' Intelligence Artificielle, 2020, 34(4):507-514.
- [13] Bandhakavi A, Wiratunga N, Massie S, et al. Emotion-aware polarity lexicons for Twitter sentiment analysis[J]. Expert Systems, 2018, 38(7):1-14.
- [14] 戴杏云, 张柳, 戴伟辉, 等. 社交网络的情感图谱研究[J]. 管理评论, 2016, 28(8):79-86.
- [15] 张柳, 王晰巍, 黄博, 等. 基于 LDA 模型的新冠肺炎疫情微博用户主题聚类图谱及主题传播路径研究[J]. 情报学报, 2021, 40(3):234-244.
- [16] 安璐, 欧孟花. 突发公共卫生事件利益相关者的社会网络情感图谱研究[J]. 图书情报工作, 2017, 61(20): 120-130.
- [17] 刘雅姝, 张海涛, 徐海玲, 等. 多维特征融合的网络舆情突发事件演化话题图谱研究[J]. 情报学报, 2019, 38(8):798-806.
- [18] 王丹, 张海涛, 刘嫣, 等. 全景生态视角的微博舆情多维图谱构建研究[J]. 情报学报, 2019, 38(12):1275-1285.
- [19] 周红磊, 张海涛, 张鑫蕊, 等. 话题-情感图谱: 突发公共卫生事件舆情引导的切入点[J]. 情报科学, 2020, 38(7):15-21.
- [20] 肖仰华. 知识图谱: 概念与技术(第1版)[M]. 北京: 电子工业出版社, 2019: 9-10.
- [21] 程新斌. 对重大舆情与突发事件舆论引导研究的分析与对策[J]. 西南民族大学学报(人文社会科学版), 2022, 43(2):235-240.
- [22] 王晰巍, 韦雅楠, 邢云菲, 等. 社交网络舆情知识图谱发展动态及趋势研究[J]. 情报学报, 2019, 38(12): 1329-1338.
- [23] 娄国哲, 王兰成. 基于知识图谱的网络舆情知识组织方法研究[J]. 情报理论与实践, 2019, 42(1):58-64.
- [24] 袁国栋. 基于知识发现的突发事件舆情风险演化模型研究[J]. 情报科学, 2021, 39(3):32-36.
- [25] 田依林, 李星. 基于事理图谱的新冠肺炎疫情网络舆情演化路径分析[J]. 情报理论与实践, 2021, 44(3):76-83.
- [26] Hernes M, Bytniewski A. Knowledge representation of cognitive agents processing the economy events[C]. Asian Conference on Intelligent Information and Database Systems. 2018: 392-401.
- [27] 马哲坤, 涂艳. 基于知识图谱的网络舆情突发话题内容监测研究[J]. 情报科学, 2019, 37(2): 33-39.
- [28] 张思龙, 王兰成, 娄国哲. 基于知识图谱的网络舆情研判系统研究[J]. 现代情报, 2021, 41(4): 10-16.
- [29] 安宁, 安璐. 跨平台网络舆情知识图谱构建及对比分析[J]. 情报科学, 2022, 40(3):159-165.
- [30] Blei D M, Lafferty J D. Dynamic topic models[C]. Proceedings of the 23rd International Conference on Machine Learning. New York: ACM Press. 2006: 113-120.
- [31] 王兰成, 张思龙, 许和旭. 网络舆情事理图谱构建及应用[J]. 中华医学图书情报杂志, 2021, 30(5):17-23.
- [32] Studer R, Benjamins V R, Fensel D. Knowledge engineering: principles and methods[J]. Data & Knowledge Engineering, 1998, 25(1-2):161-197.
- [33] 徐蓓蓓, 杨子江, 朱世伟, 等. 基于本体的突发事件舆情知识库建设研究[J]. 情报杂志, 2019, 38(4):132-137.
- [34] 李良荣, 于帆. 网络舆论中的“前 10 效应”——对网络舆论成因的一种解读[J]. 新闻记者, 2013(2):50-53.