

引用格式:杨妹,王妍,王传彪,李昊展,刘铭.突发事件网络的传播规律及特征——以知网事件为例[J].中国传媒大学学报(自然科学版),2022,29(06):58-67.

文章编号:1673-4793(2022)06-0058-10

突发事件网络的传播规律及特征 ——以知网事件为例

杨妹^{1,2},王妍^{1,2*},王传彪^{1,2},李昊展^{1,2},刘铭^{1,2}

(1.中国传媒大学数据科学与智能媒体学院,北京100024;2.中国传媒大学融合与传播国家重点实验室,北京100024)

摘要:以“知网事件”为例,构建微博转发网络,运用社会网络分析方法,从网络全局特征、网络结构特征、节点影响力三个角度,分析该事件传播规律,识别核心节点;采集意见领袖发布的文本信息,运用LDA主题提取以及SnowNLP情感分析,分析意见领袖的观点倾向。实例分析表明:此次事件参与传播的人数多、范围广,但个体化强,互相联系松散,意见领袖在转发网络中具有较大的出度、度中心性和接近中心性,以及较小的限制度,属于4-shell层。意见领袖发布内容均持中立偏积极的态度,主题为事件阐述和经营状况。

关键词:突发事件;意见领袖;社会网络分析;LDA主题提取;情感分析

中图分类号:G206;G250 **文献标识码:**A

Network propagation law and characteristics of emergencies ——take the CNKI incident as an example

Yang Mo^{1,2}, Wang Yan^{1,2*}, Wang Chuanbiao^{1,2}, Li Haozhan^{1,2}, Liu Ming^{1,2}

(1.School of Data Science and Media Intelligence, Communication University of China, Beijing 100024, China; 2.State Key Laboratory of Media Convergence and Communication, Communication University of China, Beijing 100024, China)

Abstract: Taking "CNKI" as an example, constructing a Weibo forwarding network, using social network analysis methods, from three perspectives of network global characteristics, network structure characteristics, and node influence, we analyzed the propagation law of the event, and identified the core Nodes. Then we collected text information released by opinion leaders, and used LDA topic extraction and SnowNLP sentiment analysis to analyze opinion leaders' opinion tendencies. The case analysis shows that the number of people involved in the communication of this event is large and the scope is wide, but the individuality is strong and the relationship is loose. The opinion leaders have a large out-degree, degree centrality and closeness centrality in the forwarding network, and belong to the 4-shell layer. The content released by opinion leaders has a neutral and positive attitude, and the themes are event descriptions and business conditions.

Key words: emergency; opinion leaders; social network analysis; LDA topic extraction; sentiment analysis

基金项目:北京市自然科学基金(9202018);中国传媒大学中央高校基本科研业务费专项资金资助(CUC21GZ005);中国广播电视协会学术研究项目(2020ZGLH015);国家自然科学基金(11901542)

作者简介(*为通讯作者):杨妹(1999-),女,硕士研究生,主要从事复杂网络上的传播动力学研究。Email:2364419515@qq.com;王妍(1979-),女,博士后,教授,研究方向为复杂网络与信息计算,媒体大数据分析,舆情动力学建模与仿真。Email:wy@cuc.edu.cn;王传彪(1984-),男,博士,副教授,主要从事动力系统研究。Email:cbwang@cuc.edu.cn;李昊展(1998-),男,硕士研究生,主要从事复杂网络、数据挖掘研究。Email:404839496@qq.com;刘铭(1999-),女,硕士研究生,主要从事复杂网络上的传播动力学研究。Email:876692648@qq.com

1 引言

突发公共事件^[1]是指突然发生的,可能会造成社会危害的事件,其中往往存在复杂的、富有争议的因素,容易引起网民的关注和讨论。互联网的实时性和去中心化提升了信息的扩散速度和影响范围,自媒体等意见领袖的加入使传统的议程设置发生变化,传播主体因其账号影响力的高低不同,信息获取的实效性不同,对舆情发展的推动力也不尽相同。

当舆情发生时,网络信息传播的过程是官方媒体首先发布事件报道形成信息源,而后意见领袖介入,对信息进行筛选、加工,并掺入自己的观点^[2]。网络意见领袖不能直接控制网民“怎么想”,但他们可以利用信息权力,以迎合网民的方式解读舆情,并将特定的价值理念编辑进“意见”,潜移默化地引导网民“想什么”^[3]。网民会同时接收信息源和再加工的信息,并选择是否加入到信息传播的过程中。

因此,在突发性公共舆情事件中,回答如下三方面问题对于厘清舆情传播脉络,掌握舆情规律进而正确引导和控制舆情具有重要作用。

(1) 知网事件舆论走势存在什么规律,有哪些特征?

(2) 社交网络信息传播呈现网络传播特征,每个人即是信息接收者,也是传播者,舆情事件意见领袖和把关人对信息传播的走势起重要的主导作用,在知网事件中信息传播的网络结构特征和拓扑结构中的“意见领袖”等重要节点有哪些特征?

(3) 关键节点(即意见领袖)起到什么作用,有怎样的观点输出及情感倾向?

本文运用社会网络分析方法探究参与舆情传播的各个节点之间的拓扑关系,以及重要节点的作用,分析“知网被市场监管局立案调查”舆情事件,在此基础上使用LDA主题提取和情感识别,探究重要节点的舆论引导倾向。

2 相关研究

对于重大突发事件引发的网络舆情,有多种研究角度。根据拉斯韦尔5W模式,传播过程中存在五种基本要素:传播者、信息内容、媒介、接收者和传播效果,每种要素均可作为研究的切入点。刘波维^[4]等人对舆情研究的视角进行分析,将不同视角的网络舆情研究成果进行编码和分析,得出舆情研究的几大范畴,包括主体、客体、引发者、舆情传播理论、舆情功能、传播内容等。本文从中选取用户互动和信息内容作为研究角度。

基于用户互动的角度,众多用户在社交网络上的互动行为构成了巨大的复杂网络,此前有众多研究人员利用社会网络分析方法探究传播结构及重要节点的识别。基于信息内容的角度,有研究者运用话题识别和情感分析捕捉热点话题,判断舆论导向。

Huang^[5]等认为可以将复杂网络中度最大的节点看作“意见领袖”,通过影响这些节点来控制信息在网络中传播,这一方法被应用于多种复杂网络的意见领袖挖掘。Kitsak^[6]等人指出,网络中节点的传播能力与节点所处的位置有重要关系,处于网络核心位置的节点有较高的影响力。赵蓉英^[7]等采用社会网络分析方法,进行突发事件网络舆情关键节点识别研究,通过网络社群图分析、接近中心性分析和K-丛分析等,挖掘出舆情传播的内在结构以及演化规律。刘小平^[8]等通过测度部分媒体微博的网络结构特征,结合中心度与LeaderRank算法衡量微博节点的影响力。孙羽^[9]等基于网络分析和文本挖掘划分意见领袖社团,分析“特朗普美国大选游行事件”中各类意见领袖的特点。安璐^[10]等从意见领袖的评论内容及用户间的交互入手识别网络意见领袖。刘嘉琪^[11]等基于社会网络分析方法,分析意见领袖、普通节点与在线群体影响力间的关系。

吴晓娟^[12]利用LDA模型分析“蓝色钱江纵火案”在不同舆情时期的微博话题演化过程。Yuan^[13]等利用LDA提取飓风期间的Twitter数据中公众关注的问题。Wang^[14]等利用LDA模型识别新冠肺炎疫情期间用户最常发布的主题,并结合点赞量、评论量和转发量等数据对不同话题进行了用户行为分析,研究了用户关注度的变化和话题参与度差异。黄仕靖^[15]等对数据文本进行情感分析及LDA主题模型建模,对不同阶段及不同群体的舆情时空演化及差异进行内容分析。刘雅姝^[16]等在突发事件网民评论的话题图谱上,利用LDA模型提取了话题图谱实体中的话题和时间属性,从多维特征融合角度追踪突发事件舆情的演化过程。肖倩^[17]等将卷积神经网络与LDA模型相结合,利用了舆情热度与其传播过程间的潜在关联,设计了热点舆情识别方法。邓春林^[18]对突发公共安全事件的微博评论进行聚类,用LDA主题模型对评论文本进行关键要素提取,得出归因维度,进而通过情感分析软件对各维度进行情感倾向度分析。

已有研究提出构建社交网络,利用网络结构特征、节点位置、节点中心性特征、社群分析等多种方法研究意见领袖的识别,利用LDA主题提取模型研究舆情事件演化过程,对文本数据进行情感倾向分析。上述成果对突发事件的舆情研究具有重要的借鉴意义。

但已有研究仅聚焦用户互动或传播内容进行单一角度的信息传播特征研究,本文将以上研究综合起来,以更全面、更完整、更系统的视角分析突发性舆情事件。

因此,在以上学者的研究基础上,本文采用社会网络分析方法以突发事件“知网被市场监管局立案调查”为例,先探究转发网络特征,识别网络中的意见领袖,再进一步提取意见领袖的观点输出及情感倾向。

3 研究设计

3.1 研究对象

本文选取微博和百度作为主要研究平台。研究对象是5月13日知网事件爆发期间参与信息传播的博主,包括官媒用户、自媒体用户、普通用户,共计37794名,以及部分博主发布的文章内容。

3.2 数据获取

本文使用Python语言爬虫程序爬取微博数据,获得热搜话题转、赞、评情况。通过UID跟踪微博主体转发关系,经过清洗后获得知网报道转发网络。使用八爪鱼采集器爬取百家号数据,得到百度移动端和PC端的用户搜索指数,以及媒体发布的文章内容。

3.3 研究方法

3.3.1 网络拓扑结构特征分析

当人们通过社会网络交换信息的同时,也建立起了一个庞大的传播信息网络。个体间的相互作用对信息传播有着至关重要的影响。本文建构了一个经典的无权有向转发网络,使用Python中的Networkx包,网络分析工具Gephi和Pajek,对转发网络进行定量的整体分析,对个体间关系和相互影响进行精细测量。

研究流程如图1所示。



图1 转发网络研究指标

各指标对应的表达式如表1所示。

表1 表达式

指标	表达式
平均度	$K = \frac{\sum_{i=1}^N k_i}{N}$
平均聚类系数	$C = \frac{\sum_{i=1}^N \frac{2E_i}{k_i(k_i - 1)}}{N}$
平均路径长度	$L = \frac{1}{2N(N-1)} \sum_{i \neq j} d_{ij}$
度中心性	$DC(i) = \frac{k_i}{N-1}$
接近中心性	$CC(i) = \frac{1}{d_i} = \frac{N-1}{\sum_{i \neq j} d_{ij}}$
特征向量中心性	$EC(i) = c \sum_{j=1}^N A_{ij} x_j$
限制度	$C(i) = (p_{ij} + \sum_q p_{iq})^2$

其中, k_i 表示与节点*i*直接相连的节点数; N 表示网络中节点总数; E_i 表示节点*i*与相邻点间直接连接的边数; d_{ij} 表示目标节点*i*到节点*j*的距离; A_{ij} 是邻接矩阵中的元素; c 表示矩阵*A*中特征值最大值的倒数; p_{ij} 代表行动者*i*投入到相连点*j*的关系所占比例; q 是除*i*或*j*之外的每个第三者, p_{iq} 代表第三者*q*投入到相连点*j*的关系所占比例。

3.3.2 意见领袖发表内容分析

(1) LDA主题提取

LDA(隐含狄利克雷分布)^[19]是一种文档主题生成模型,也称为一个三层贝叶斯概率模型,包含词、主题和文档三层结构。一篇文章的每个词都是通过“以一定概率选择了某个主题,并从这个主题中以一定概率选择某个词语”这样一个过程得到。文档到主题服从多项式分布,主题到词服从多项式分布,本文需要提取出意见领袖输出的核心报道主题。先进行分词处理、分词向量化,随后使用Python中的库pyLDAvis,指定单词个数上限和主题个数等参数,将数据填进模型,抽取主题,最后用pyLDAvis可视化。

(2)情感分析

情感分析是判定文本中对某一话题所表现出的态度或情绪倾向性的过程、技术和方法,本文需要判断意见领袖输出的观点携带的情感倾向。采用Python中的库SnowNLP和Pandas,直接对文件中的文本进行情感值打分。SnowNLP^[20]可以进行文本预处理,例如分词、词性标注提取关键词等,同时该程序包也可以进行情感分析,使用的基本模型是贝叶斯模型,即对于有两个类别*c*₁和*c*₂的分类问题来说,其特征为*w*₁, ..., *w*_{*n*},特征之间是

相互独立的,属于类别 c_1 的贝叶斯模型的基本过程为:

$$P(c_1|w_1, \dots, w_n) = \frac{P(w_1, \dots, w_n|c_1) P(c_1)}{P(w_1, \dots, w_n)}, \quad \text{其中}$$

$P(w_1, \dots, w_n) = P(w_1, \dots, w_n|c_1) \cdot P(c_1) + P$. 最终得到的取值为 $[0, 1]$, 0表示极端负面, 1表示极端正面。

研究流程如图2所示。

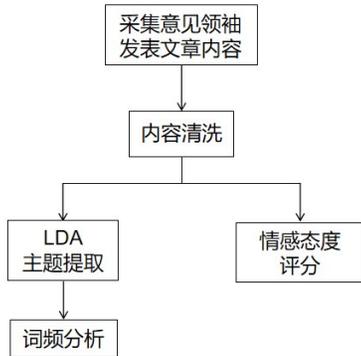


图2 文章内容研究指标

4 知网事件基本统计分析

4.1 事件时间线梳理

知网在论文平台中拥有极高的市场占有率,近年因连年涨价的高昂服务费被多所高校抵制。如图3所示,从去年赵德馨教授的论文被下架,国家市场监督管理总局就表示将核实研究知网是否涉嫌垄断;到今年中科院停用知网,国家市场监督管理总局在网上留言回应,表示正在依法开展相关工作。2022年5月13日,市场监管总局对知网涉嫌垄断行为立案调查。

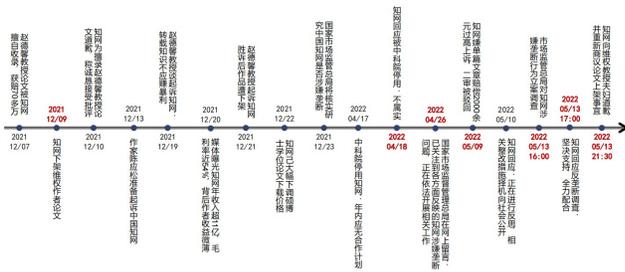


图3 知网事件时间轴

4.2 知网搜索热度情况

图4搜索指数源于百度移动端和PC端的用户搜索情况。图中标出五个较高极值点,对应上图时间轴中红色的时间节点。

此次事件前期潜伏期长,从去年开始关于知网的

议题就开始频繁出现在大众视野中,基本都是负面议题,议程设置的功能得到充分发挥。事件爆发之前,知网的负面形象已经在人群中广泛传播。

2021年12月中旬,因赵教授系列事件,知网短暂维持在较高搜索热度,之后热度回落。从今年二月底开始,知网一直处于较高搜索热度,与毕业生准备毕业论文时间重合。4月18日知网搜索热度飙升,当天中科院宣布年内与知网无合作计划。4月26日,国家市场监督管理总局在网上留言表示已关注到各方面反映的知网涉嫌垄断问题,正在依法开展相关工作。此前赵教授的妻子周秀鸾也进行维权,5月9日,知网就赔偿金额过高提出上诉,二审被驳回。5月13日搜索热度再次飙升,当天下午4点,由中国之声首发报道,市场监管总局宣布对知网涉嫌垄断行为立案调查,此次舆情爆发。

此次事件爆发期短,因为这是一个收尾事件,前期知网的所作所为现在国家已经开始调查,大众得到了想看到的结果。

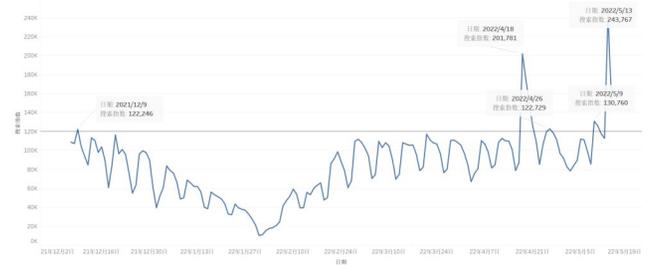


图4 知网搜索指数

4.3 话题参与情况

图5为话题#知网涉嫌垄断被立案调查#词条中参与讨论用户的转、赞、评情况。筛选出点赞量大于90的博主列入下图,按点赞数由大到小排序,其中媒体官号与自媒体号均为13个,各占50%。发挥议程设置的功能,这些媒体和自媒体通过提供信息来吸引受众关注到知网事件中。受众不是被动的信息接收者,而是积极的网络传播参与者,通过转、评、赞的方式给媒体予以反馈。其它媒体的信息均从信息源“中国之声”获得,它们承担了意见领袖的作用。而“中国之声”既是信息源,又是意见领袖,它作为信息源起到传递信息的功能,作为意见领袖发挥的作用是在他的追随者中构成信息和影响的重要来源,并左右追随者的态度。它的特殊地位给予它比其它媒体高出数十倍的数据量。

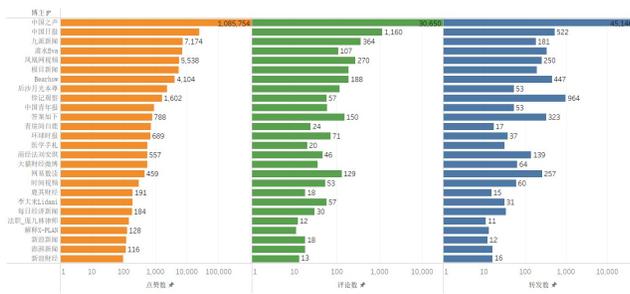


图5 话题#知网涉嫌垄断被立案调查#转赞评情况

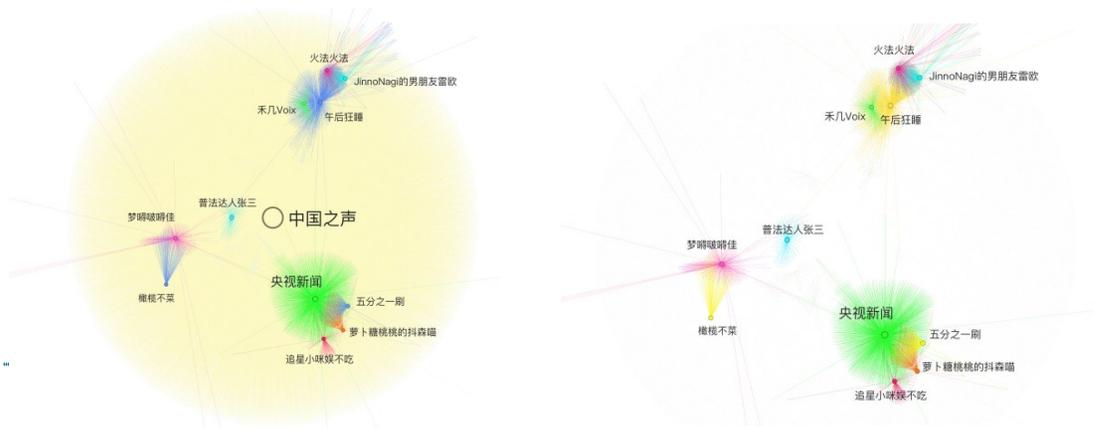


图6 转发网络可视化(右无中国之声)

将采集到的新浪微博转发和被转发用户数据导入Gephi,绘制转发网络云图,如图6所示。运用Python计算得到转发网络的基本指标,如表2所示。根据运算结果可以看出,总计37794名用户参与到知网话题的信息传播,他们之间发生转发的总次数为42482次,整体上节点用户数量规模较大,相互之间传播互动数量较少。参与传播的人数多、范围广,原因可能是潜伏期已经历几次传播,了解知网事件的受众规模大。

知网信息转发网络的平均度值为2.248,表示每个节点用户之间平均相互发生联系的用户数量为2.248。此转发网络的直径仅为2,表示网络中任意两点之间的最长距离较小。普通用户直接与传播网络中的意见领袖建立转发联系,中介节点非常少。网络密度接近0,为稀疏网络。

表2 转发网络基本指标

指标	数值
节点(用户)数	37794
边数	42482
平均度	2.248
平均聚类系数	0.042
平均路径长度	1.002
网络直径	2

5 “转发”网络的拓扑结构特征分析

转发是舆情信息扩散的方法和主要途径,而意见领袖在其中处于关键位置,对于信息扩散起到主导作用,最大化地促进信息传播。故本节构建转发网络,先分析网络全局特征,再综合多种指标识别网络中的意见领袖节点。

5.1 网络全局特征:庞大而松散的网络

5.2 网络结构特征:多层次阶梯式扩散传播的无标度网络

5.2.1 连接状况严重不均的异质网络

运用Gephi统计分析功能得出知网信息转发网络直径为2,即网络中任意两点间距离的最大值仅为2。平均聚类系数为0.042,该值体现用户之间的聚集程度,即网络中两个用户之间存在联系的概率为0.042,用户之间联系不紧密。平均路径长度为1.002,即用户平均通过1.002个节点用户就可以与其他节点用户接触。该网络聚类系数小,平均长度远小于6,不符合基于六度分隔理论的小世界网络。

使用Gephi统计分析功能得出图7度分布图,为使图中度分布显著,则排除度最大节点中国之声,得出知网事件转发网络中的结点的度符合幂律分布,故该网络为节点度数严重不均匀的无标度网络。

网络中转发链路的长度短、宽度窄,传播速度快,传播效率高。用户之间的接触和信息交流严重不均匀,大部分普通用户为底层边缘节点,传播路径到他们即终止,只有少数关键节点在信息传播中起主导作用。关键节点的识别将在下节详细讨论。

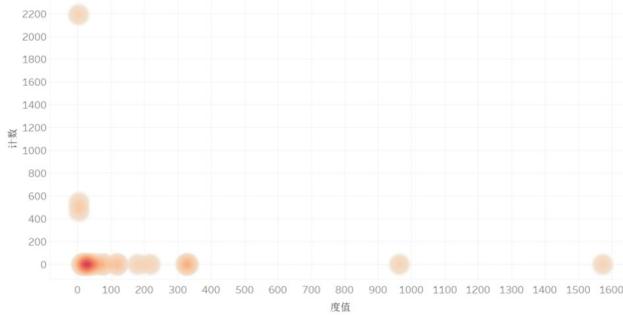


图7 网络节点度分布(除中国之声)情况

5.2.2 社群内部层级扩散

运用Gephi的模块化功能社区探测算法分解网络。解析度越小,社区越多;解析度越大,社区越小。采用适中的解析度3,得出4个子网络,如图8所示。采用Modularity Class进行颜色渲染,直观地得出四个凝聚

子群,其规模差异大,集群现象明显,如图9所示。

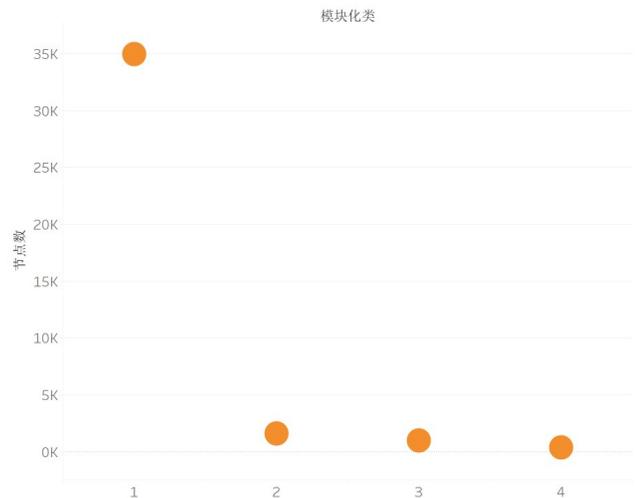


图8 模块化社区探测结果

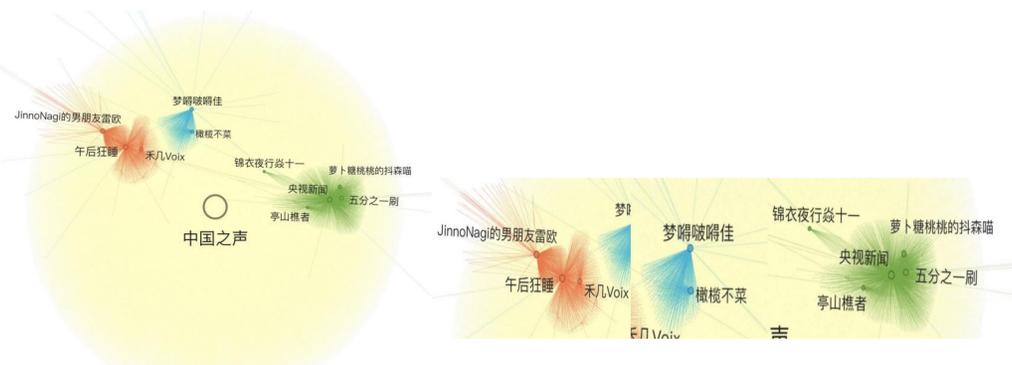


图9 modularity class 颜色渲染图

规模最大的子网络是由信息源“中国之声”主导的,拥有整体网络中92.42%的节点,数量超过三万。它与其他三个子网络均有紧密联系,三个子网络各自无过多联系。

排名第二的子网络以官媒“央视新闻”主导,另外四个大度节点“亭山樵者”、“五分之一刷”、“锦衣夜行焱十一”、“萝卜糖桃桃的抖森喵”均转发了央视新闻的信息;排名第三的子网络以自媒体博主“午后狂睡”主导,另外两个大度节点“禾几Voix”、“JinnoNagi的男朋友雷欧”均转发了“午后狂睡”的信息;排名第四的子网络以自媒体博主“梦啾啾啾佳”主导,大度节点“橄榄不菜”转发了“梦啾啾啾加”的信息。而三个小规模子网络的主导节点均转发了信息源“中国之声”的信息,传播过程呈多阶梯型扩散体系,从信息源到受众中间出现多个层次的意见领袖,形成多级传播,如图10所示。

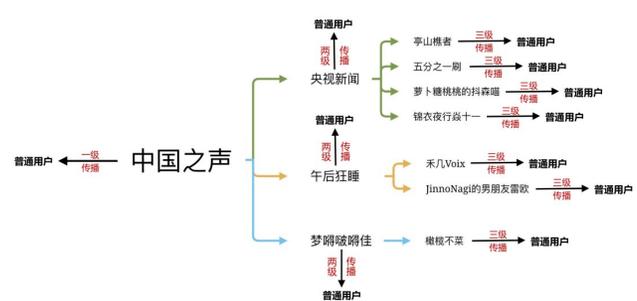


图10 多级传播示意图

虽然信息传播的参与人数多,但是舆论主要由一方主导,不存在舆论对立和舆论失焦的现象。在整个网络传播过程中,从一级传播到多级传播均有涉及。

5.3 关键节点的特征:传播过程中的核心领导力

5.3.1 节点指标测度分析

(1) K-Shell分解

K壳分解可用于区分网络中节点的重要程度,识

别网络中的关键节点,分解过程是逐渐剥除当下度为1的节点,也就是最不重要的节点,重复循环该操作直至网络中没有度为1的节点。K值由小到大,重要程度递增,4壳节点是网络中的关键节点,中国之声、央视新闻、午后狂睡等核心领导力节点均位于4-shell中。各类节点数量如图11所示。K核子图如图12所示,K值相同节点存在小范围聚集情况,但各个子群联系不紧密。

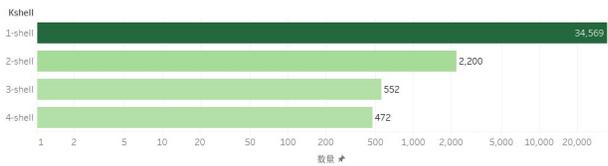


图11 知网信息转发网络K-Shell数量

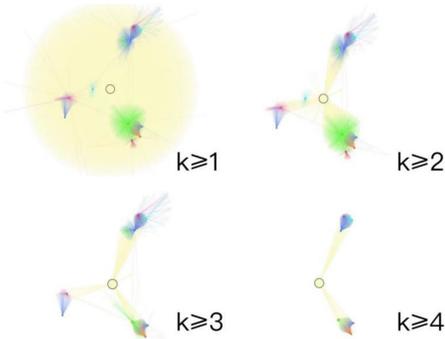


图12 知网信息转发网络K-Shell子图

(2)度中心性

该转发网络排名前10的节点用户度中心性结果,如图13所示。通过对节点用户度中心性分析可以看出,形成转发网络中心度排名前三的节点分别是初始源节点中国之声、央视新闻、午后狂睡。前两者是媒体官号,后者是影评博主,粉丝量都达到百万千万的级别,发挥着意见领袖的作用。节点用户的度中心性越大,他们的节点圈越大,从图6也可以看出这三个节点用户的圈较大。

出度表示被转发次数,入度表示转发次数。出体量远大于入度,信息流动呈单向化。用户普遍仅转发一次,最多转发四次,如图14所示。说明受众对知网事件的参与度不深入,倾向于单一地转发有影响力的信息源。

其中中国之声作为信息源头,出度达到37729,相当于被转发37729次。作为央广旗下的媒体,同时具备信息源和意见领袖的作用,其消息来源真实度高,获取信息早,社会地位高,更容易受到受众的追捧,故

度中心性处于垄断地位。



图13 知网信息转发网络度中心性和出度情况

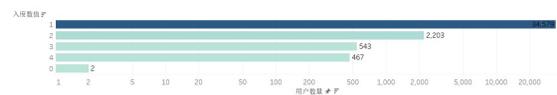


图14 知网信息转发网络入度情况

(3)接近中心性

图15列出接近中心性0.5以上的用户,这些用户到其他用户的最短距离很小,接近几何上的中心位置,更容易参与到信息传播过程中。相对地独立性也较强,不易受到其他节点的影响。有些节点同时拥有较大的度中心性和接近中心性,如中国之声、央视新闻、午后狂睡、火法火法等。他们是知网信息传播过程中的核心领导力,在形成的转发网络中发挥着重要的作用。

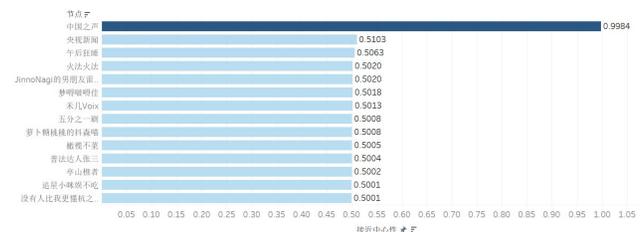


图15 知网信息转发网络接近中心性

(4)结构洞

结构洞^[21]是指两个关系人之间的非重复关系。如果两者之间缺少直接的联系,而必须通过第三者才能形成联系,那么行动的第三者就在关系网络中占据了一个结构洞,结构洞是针对第三者而言的。节点占据的结构洞越多,它的地位就越重要,通常用限制度来测量。

将数据导入Pajek中,算出限制度,如图16,取数值小于0.24的节点,与对应节点的度中心性比对,基本走势成反比。限制度指节点间存在阻碍的程度,值越小的节点,越不易受到控制,拥有更多的资源,具有

更多的结构洞,他们是传播网络中的信息中枢,这些节点的度中心性也处于领先地位。

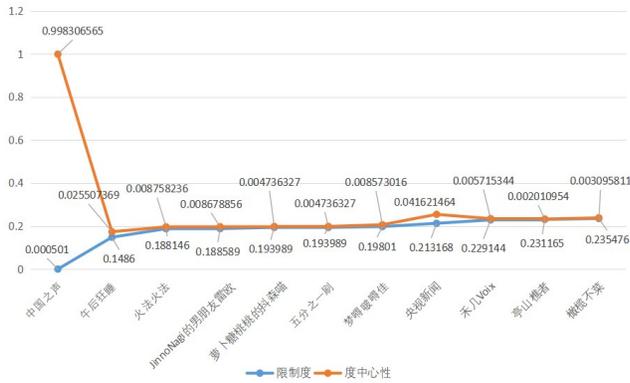


图16 知网信息转发网络限制度

(5)特征向量中心性

特征向量中心性强调节点邻居的数量和质量,特征向量中心性高的节点连接了很多其他重要节点。图17显示的是用户数量与特征向量中心性数值的关系。特征向量为1的节点均是通过和少量高分值节点相连来提高自身的重要性,他们是积极参与到传播网络中的普通用户,多次转发意见领袖的信息。但大部分普通用户没有深入参与到传播过程中,对此事件的关注浅尝辄止,值为0.0158的用户高达91.42%。

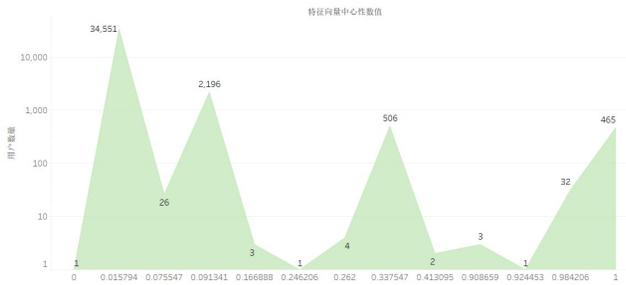


图17 知网信息转发网络特征向量中心性

5.3.3 意见领袖特征

意见领袖是关键节点,是信息中枢,担任传播中的核心领导力,在其中起到信息传递和意见分发的重要作用。综上所述,他们的特征即是拥有较大的出度、k值、度中心性和接近中心性,以及较小的限制度。

通过对拥有这些特征的节点取交集,我们可以识别出网络中的意见领袖,包括官媒“中国之声”和“央视新闻”,影评人“午后狂睡”,动漫博主“火法火法”和“JinnoNagi的男朋友雷欧”,娱乐博主“梦啾啾啾佳”和“橄榄不菜”,军事博主“禾几Voix”,饭圈大粉“萝卜糖桃桃的抖森喵”和“五分之一刷”,法律博主“普法达人张三”,

科技博主“亭山樵者”。他们分属不同领域,影响不同圈层,再次体现了此次事件的影响范围之广。其中信息源“中国之声”处于转发网络中的垄断地位,其出度值、度中心性和接近中心性以及首发微博的转赞评都排在首位,并且数据量比其他意见领袖高数十倍。

6 意见领袖输出的观点及态度分析

上一节讨论了意见领袖节点在转发网络中的特征和作用,即传播主体之间的互动。本节针对意见领袖发布的文章内容进行讨论,也就是传播内容。

基于主流媒体及自媒体发声,利用文本分析方法提取其语义信息,观察意见领袖对舆情事件的加工解读,绘制词云如图18所示。“中国”、“市场”、“垄断”、“调查”、“学术”、“论文”等词汇被提及频率排行前列,媒体大多从客观事实角度出发陈述事件,整体对该事件持中立态度。



图18 媒体发文词云图

6.1 报道主题:事件阐述和经营状况

利用LDA主题提取方法得到主题数困惑度情况如图19所示,选择较低困惑度对应的主题数,得到“事件阐述”、“经营状况”两个主题。

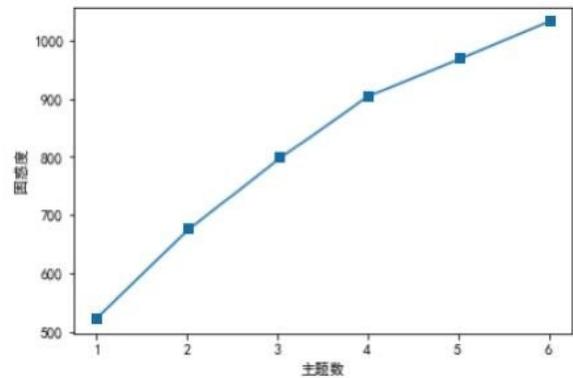


图19 主题数困惑度情况

两个主题各自包含的排名前二十的单词词频排序分别如图20、图21所示。“事件阐述”主题中,“知网”、“中国”、“论文”三个单词P值大于0.95,与主题相关程度最

高,该主题与事件发展始末的叙述高度相关。“经营状况”主题中,前20个单词与主题相关程度差别不大,大多数单词用来描述知网当前经营情况。“事件阐述”内最低词频高于“经营状况”内最高词频,说明媒体把主要关注点放在事件本身,衍生话题关注度较低。

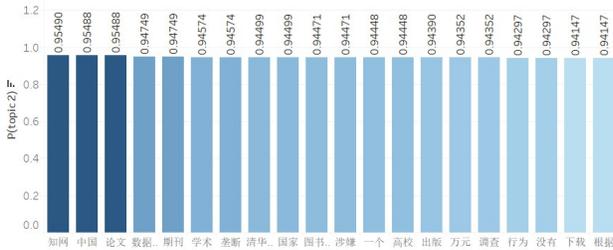


图20 “事件阐述”单词词频排序

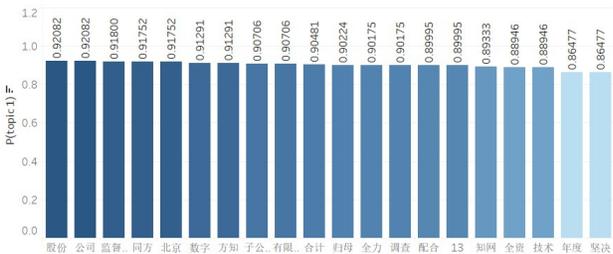


图21 “经营状况”单词词频排序

综上,媒体报道主要从此次立案调查事件的发展始末阐述以及知网经营状况两个角度展开。这两个角度也是在解答受众关心的问题,即知网为什么会调查,以及知网在本行业的垄断地位从何而来。

6.2 报道态度:迎合受众,持中立偏积极

使用SnowNLP计算百家号正文情感得分,其分布情况如图22所示。其中,得分大于0.55表示积极,0.45-0.55表示中立,小于0.45表示消极。

过半数文章对知网事件进行积极阐述,即对知网被市场监管局调查持支持态度。其他文章均采用中立情绪叙述,没有消极情绪的文章出现。同时,通过观察原始得分数据,所有文章得分均在0.48-0.62之间,说明媒体大多持中立偏积极的态度对事件进行报道。

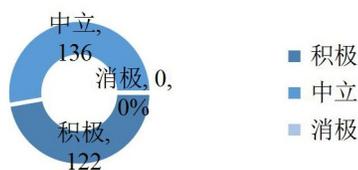


图22 内容情感态度

之前对知网的报道大部分为负面报道,与其在业内处于垄断地位的经营状况和对学者的不公平待遇有关,已形成“学界苦知网久矣”的舆论倾向。此次知网被调查,媒体持的态度符合先前的议程走向,与先前知网系列报道舆情导向一致。

7 结论

本研究基于5月13日的知网事件从传播网络和发布内容两个层面分析传播规律和舆论倾向。得出如下结论:

(1)此次事件前期潜伏期长,爆发期短,参与传播的人数多、范围广,转发网络符合无标度网络特征。舆论主要由一方主导,不存在舆论对立和舆论失焦的现象。传播呈单向流动,过程呈多阶梯型扩散体系,从信息源到受众中间出现多个层次的意见领袖,形成多级传播。

(2)用户个性化强,互动性低,互相联系松散,普通用户主要与传播网络中的意见领袖直接建立转发关系,中介节点非常少。仅有少量普通用户多次转发关键节点的信息,大部分普通用户没有深入参与到传播过程中。意见领袖拥有较大的出度、k值、度中心性和接近中心性,以及较小的限制度。识别出的意见领袖分属不同领域,影响不同圈层。其中信息源“中国之声”处于转发网络中的垄断地位,其各个特征均排在首位。

(3)事件的议程设置与先前知网系列报道舆情导向一致。报道内容均持中立偏积极的态度,无消极内容,报道主题以事件阐述和经营状况为主。

本次突发性舆情事件的传播特征分析可以为舆情预测和控制提供如下参考:(1)对于舆情预测,如果事件存在潜伏期,可以通过提取报道主题和情感倾向,观察潜伏期的议程设置与此次事件是否一致,来预测用户反应和舆论走向。若一致,则舆论由一方主导的几率较大;若不一致,则需要警惕次生舆情引起的讨论。(2)对于信息控制,可以用以上指标识别其中的关键节点,有针对地持续监测,以减少工作量。通过控制这些节点,提取他们输出信息内容的主题,判断与目标舆情倾向是否相符,来决定扩散还是抑制这些节点的传播,让他们充当把关者。

参考文献(References):

[1] 刘高勇,黄靖钊,艾丹祥.融合立场检测和主题挖掘的突发公共事件网络舆情演化研究[J].广东工业大学学报,2022,39(03):32-40+48.

- [2] 熊尧,李弼程,王子玥.基于两级传播理论的舆论超网络传播分析[J].智能系统学报,2020,15(05):870-879.
- [3] 杨慧民,陈锦萍.网络意见领袖建构网络意识形态的逻辑理路及其应用[J].理论导刊,2022(04):53-58+78.
- [4] 刘波维,曾润喜.网络舆情研究视角分析[J].情报杂志,2017,36(02):91-96.
- [5] Huang X, Vodenska I, Wang F, et al. Identifying influential directors in the United States corporate governance network [J]. Physical Review E, 2011:84(4):1-8.
- [6] Kitsak M, Gallos L K, Havlin S, et al. Identification of influential spreaders in complex networks [J]. Nature Physics, 2010, 6(11):888-893.
- [7] 赵蓉英,王旭.突发事件网络舆情关键节点识别及导控对策研究——以“大贤村遭洪灾事件”为例[J].现代情报,2018,38(01):19-24+30.
- [8] 刘小平,田晓颖.媒体微博的社会网络结构及其影响力分析[J].情报科学,2018,36(01):96-101+123.
- [9] 孙羽,裘江南.基于网络分析和文本挖掘的意见领袖影响力研究[J].数据分析与知识发现,2022,6(01):69-79.
- [10] 安璐,胡俊阳,李纲.基于主题一致性和情感支持的评论意见领袖识别方法研究[J].管理科学,2019,32(01):3-13.
- [11] 刘嘉琪,齐佳音,陈曼仪.基于社会网络分析的意见领袖与在线群体影响力关系研究[J].情报科学,2018,36(11):138-145.
- [12] 吴晓娟.基于微博文本的网络舆情主题演化分析[D].南京:南京大学,2018.
- [13] Yuan FX, Li M, Liu R. [J]. International Journal of Disaster Risk Reduction, 2020, 51:101798.
- [14] Wang J Z, Zhou Y, Zhang W, et al. Concerns expressed by Chinese social media users during the COVID-19 pandemic: content analysis of Sina Weibo microblogging data [J]. Journal of Medical Internet Research, 2020, 22(11): e22152.
- [15] 黄仕靖,吴川徽,袁勤俭,夏镜然.基于情感分析的突发公共卫生事件舆情时空演化差异研究[J].情报科学,2022,40(06):149-159.
- [16] 刘雅姝,张海涛,徐海玲,魏萍.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(08):798-806.
- [17] 肖倩,谢海涛,刘平平.一种融合LDA与CNN的社交媒体中热点舆情识别方法[J].情报科学,2019,37(11):27-33.
- [18] 邓春林,周舒阳,杨柳.大数据环境下公共安全突发事件微博用户评论的归因分析[J].情报科学,2021,39(01):48-55+80.
- [19] 张伦.计算传播学导论[M].北京:北京师范大学出版社,2018:75.
- [20] 张伦.计算传播学导论[M].北京:北京师范大学出版社,2018:51-52.
- [21] Burt R S. Structural holes and good ideas [J]. American Journal of Sociology, 2004, 110(2):349-399.

编辑:王谦

(上接第57页)

- [17] Dai J, Qi H, Xiong Y, et al. Deformable convolutional networks [C]//IEEE International Conference on Computer Vision, 2017.
- [18] Zhou X, Wang D, Krähenbühl P. Objects as points [DB/OL]. arXiv:190407850, 2019.
- [19] Geng Z, Sun K, Xiao B, et al. Bottom-up human pose estimation via disentangled keypoint regression [C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021.
- [20] Jaderberg M, Simonyan K, Zisserman A, et al. Spatial transformer networks [C]// NIPS'15: Proceedings of the 28th International Conference on Neural Information Processing Systems, 2015, 2: 2017-2025.
- [21] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition [DB/OL]. arXiv:14091556, 2014.
- [22] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition [C]//IEEE Conference on Computer Vision and Pattern Recognition, 2016.
- [23] Ren S, He K, Girshick R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 39(6): 1137-1149.
- [24] Zhu X, Hu H, Lin S, et al. Deformable convnets v2: more deformable, better results [C]//IEEE Conference on Computer Vision and Pattern Recognition, 2019.
- [25] Newell A, Huang Z, Deng J. Associative embedding: End-to-end learning for joint detection and grouping [C]// 31st Annual Conference on Neural Information Processing Systems, 2017: 2278-2288.
- [26] Sun K, Geng Z, Meng D, et al. Bottom-up human pose estimation by ranking heatmap-guided adaptive keypoint estimates [DB/OL]. arXiv:200615480, 2020.
- [27] Lin T-Y, Maire M, Belongie S, et al. Microsoft coco: Common objects in context [C]//European Conference on Computer Vision, 2014: 740 - 755.
- [28] Ronchi M R, Perona P. Benchmarking and error diagnosis in multi-instance pose estimation [C]//IEEE International Conference on Computer Vision, 2017.

编辑:王谦