

DOI: 10.13718/j.cnki.xdzk.2023.08.005

于凯, 王云乐. 互动仪式链视角下隐私泄露事件网络舆情分析 [J]. 西南大学学报(自然科学版), 2023, 45(8): 48-60.

互动仪式链视角下隐私泄露事件网络舆情分析

于凯^{1,2}, 王云乐²

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

摘要: 对社交媒体中隐私泄露事件所产生的网络舆情进行分析, 在互动仪式链视角下关注不同群体间主题及情感的变化, 帮助政府部门完善隐私政策的制定并提高公众的隐私保护意识。对豆瓣小组中隐私泄露问题的相关讨论进行数据分析, 主要利用 BERTopic 主题模型抽取各类小组关注的核心问题, 借助百度 AI 进行情感分类并对比分析。研究表明, 互动仪式中相互关注的焦点使得各小组内对于隐私泄露问题有不同的侧重, 科技类小组注重数据安全性, 而生活类小组认为容忍一定的隐私泄露可以换取更加便利的生活; 其情感共享结果表明, 约 75% 的用户对于隐私泄露持正面情绪。根据研究结果提出对策建议, 促进用户隐私与网络发展的有效协同, 构建更加完善的隐私保护体系, 为数字化可持续发展提供有效支撑。

关键词: 隐私泄露; 主题挖掘; 互动仪式链; 舆情分析

中图分类号: TP391; G206

文献标志码: A

文章编号: 1673-9868(2023)08-0048-13

开放科学(资源服务)标识码(OSID):



Analysis of Online Public Opinion on Privacy Breaches from the Perspective of Interactive Ritual Chains

YU Kai^{1,2}, WANG Yunle¹

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

2. School of Public Administration, Xinjiang University of Finance and Economics, Urumqi 830012, China

Abstract: The analysis of online public opinion generated by privacy breaches in social networks and the changes in themes and sentiments among different groups in the context of interactive ritual chains helps government departments to improve privacy policies and raise public awareness of privacy precautions. The data was processed and keywords co-occurred in the discussion of privacy leaks in Douban groups, the core concerns of each group were extracted with the help of BERTopic themes, the emotion was classified and comparative analyzed with the help of Baidu AI. The study shows that the focus of the interactive ritu-

收稿日期: 2023-06-08

基金项目: 国家自然科学基金项目(72264036, 72164034); 新疆维吾尔自治区社科基金项目(21BTQ162); 新疆维吾尔自治区自然科学基金项目(2019D01A22); 中国科学院西部青年学者项目(2020-XBQNXZ-020); 新疆财经大学研究生科研创新项目(XJUF2023K034)。

作者简介: 于凯, 博士, 教授, 主要从事突发事件应急管理、信息传播与舆情治理等研究。

als has led to a different emphasis on the privacy issue within each group. The technology group focused on data security and the lifestyle group believed that a certain amount of tolerance can be exchanged for greater convenience. Its sentiment sharing results indicate that about 75% of users had positive feelings about privacy breaches. Based on the findings of the study, recommendations are made to promote effective synergy between user privacy and online development to build a better privacy protection system and provide effective support for sustainable digital development.

Key words: privacy breach; topic mining; interactive ritual chains; opinion analysis

社交媒体行业随着移动通讯技术的发展而不断迅速演进,公众隐私边界在频发的信息泄露问题中被重新界定。隐私泄露是公众在网络安全中最容易受到影响的部分,在社交媒体中该现象更为明显。部分平台所获取的权限远超过其需要使用的数量,社交媒体中频繁发生的隐私泄露事件严重影响了用户隐私披露的意愿^[1]。2021年9月《中华人民共和国数据安全法》正式施行,在网络安全问题中,公众更加关注个人信息泄露问题,隐私主体被恶意骚扰甚至引发网暴事件被广泛讨论,面对数字化转型中的挑战,隐私泄露成为亟待解决的问题之一。

本研究从互动仪式链视角出发,从主题和情感两个维度进行分析,探究不同用户群体对于隐私泄露事件的态度。结合互动仪式链中相互关注的焦点及情感的共享等基本要素,分析隐私泄露中的痛点问题,挖掘不同群体对于隐私泄露事件产生容忍度差异的原因,对社交媒体上该现象的网络舆情展开讨论,为相关政策的制定提供有效依据。进而帮助有关部门和平台明确披露个人信息的边界,促进隐私服务的完善,在保证用户隐私披露安全性、主动性及满意度的同时维系平台的蓬勃发展。基于以上目的,本文重点回答以下问题:(1)不同群体对于隐私泄露事件的关注点是否有差异?(2)各群体对于隐私泄露事件的侧重点是否与其群体属性有关?(3)互动仪式对隐私泄露事件的情感倾向是否有一定的影响?

1 相关研究

互动仪式链理论提出,群体活动的参与者利用相互关注的焦点及共享的情感状态^[2]在互动仪式中获得群体归属感和身份认同感^[3]。其基本要素包含身体在场、共同的情感状态、共同关注的焦点及对局外人设限4个方面。该理论的核心是情感能量的传递,个体情感能量受到群体的影响,早期用于研究主体间的互动关系^[4]或理论修正^[5]。由于网络的发展,身体在场这一要素逐渐演化为网络空间的虚拟在场^[6],所以也有学者运用这一理论对网络舆情进行分析,探究视频弹幕网站中的情感演化规律^[7]或是特殊群体的情绪转化过程^[8]。

将互动仪式链理论与网络舆情分析相结合丰富了研究视角,在对隐私泄露问题的探讨中引入该理论可以构建多视野分析框架。隐私泄露问题指未经授权公开隐私主体个人信息的行为,在社交媒体不断发展的今天是一种普遍现象。目前的研究主要包括对隐私泄露原因的探究^[9];保护对策的有效性分析^[10];影响隐私泄露的关键因素讨论^[11]。有学者结合具体案例分析隐私泄露的原因^[12];以图书馆为平台找寻隐私泄露的途径^[13]。由于个体差异存在容忍度的区分^[14],有学者在医疗数据泄露^[15]和应用程序泄露^[16]等方面建立了容忍度测量方法。在隐私泄露行为发生后,应完善举报机制以减少信息扩散^[17],加强社交媒体管理减少再次发生的概率^[18]。面对突发的新冠疫情,借助大数据实现“数据防疫”取得了较好的成果,但这也使公众重新思考隐私披露的安全性问题,在信息收集过程中加大了隐私泄露的风险,容易使公众的负面情绪转移为网络暴力行为,产生较为严重的后果^[19]。

社交媒体中的文本数据较为杂乱,分析过程中需要对数据进行整理,对信息进行挖掘。对隐私泄露事件进行主题挖掘及情感分析是研究该类事件网络舆情的两个重要维度。目前主题识别的研究方法已较为成熟,有学者利用LDA模型对在线健康社区中用户关注的主题词进行提取^[20];利用LDA2vec模型进一步提

高政策文本语义内涵挖掘的完整性^[21]；针对隐私泄露中的围观者进行主题抽取^[22]。也有学者借助 LSTM 算法辅助主题分类研究，提出按照词组分词抽取主题以提高准确度^[23]。在网络舆情的研究中，引入情感分析可以拓展研究角度。目前使用较多的情感分析方法主要为 3 类，即利用情感词典进行情感分类^[24]、借助卷积神经网络建立情感分类模型^[25]及基于深度学习对情感极性进行分析^[26]。现有改进分析方法多是基于这 3 类发展出的更加精确的情感分类模型。

相关研究已从多方面关注到隐私泄露的风险性，但多从用户容忍度及政策完善等方面对该现象进行讨论，少有研究关注到隐私泄露的主题集中性及各用户群体关注中心的差异性。故本文探究了隐私泄露事件在互动仪式链视角下的痛点问题，借助主题挖掘与情感倾向分析，对不同群体间主题差异性 & 情感倾向分布进行挖掘，针对社交媒体中该现象的网络舆情展开讨论，为相关政策的制定提供有效依据，促进隐私服务的完善。

2 研究方法

豆瓣小组实行进组审核制，满足对局外人设限这一基本要素，同时组内成员存在共同兴趣或身份属性，具有相似的视角。故以豆瓣小组作为社交媒体数据来源，爬取各小组内对于“隐私泄露”事件的讨论，获取发帖时间、讨论详情及小组类别等数据，借助哈工大停用词表去除停用词、jieba 分词工具进行分词及人工去除无意义字符等预处理后，利用 BERTopic 模型和百度 AI 情感分析模块从主题挖掘与情感分析两个维度对数据进行分析与整合。对应研究框架见图 1。

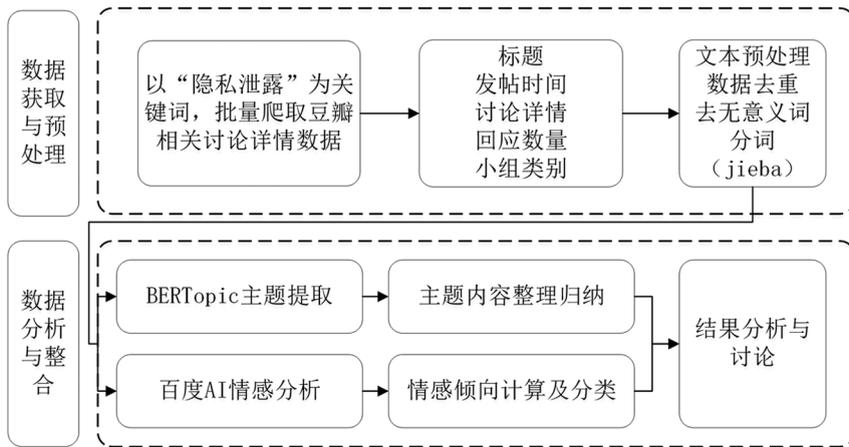


图 1 研究框架图

2.1 互动仪式链

互动仪式链理论包含 4 个基本要素，即身体在场、焦点共同、情感共享及对外设限 4 个组成要素，基于该理论构建豆瓣小组关于隐私泄露事件主体关系图，如图 2 所示。

随着社交媒体的发展，身体在场这一要素不再局限于实体，豆瓣小组将组内成员聚集起来，实现网络空间的虚拟在场。同时组内成员关注领域相同，在隐私泄露事件中有共同的焦点，互相分享传播所关注的内容。当事件发生引发较多话题时，人们倾向于靠近观点相似或是情感倾向相似的讨论，在不断表达观点的过程中产生情感共享进而积累情感能量。同时对外设限这一要素能使用户产生身份认同感，进一步参与讨论。基于互动仪式链理论，进行主题挖掘以探究各群体关注焦点，计算情感倾向以分析用户间情感共享状态，利用主题相似度判断不同小组之间是否具有差异。

互动仪式会产生一定的结果。组内成员不断参与小组讨论，通过发帖或评论推动互动仪式的持续进行，进而衍生出组内专有词汇或暗号，加强成员对小组的认同感，产生群体之间的团结。不断积累的情感能量及对焦点的认同或反对短期内表现为观点表达或情绪流露，到达一定阈值后各用户之间产生情感联结，推动用户

参与下一个互动仪式. 线上积累的情感在线下接触同类型事件时会产生一定的行为映射, 推动支持或反对某些措施的线下行为. 同时由于互动仪式而产生的群体团结, 组内成员对于小组内部的观点有较强的认可, 进而形成维护群体道德标准, 对不符合组规或者破坏群体团结的人表示反击或不满.

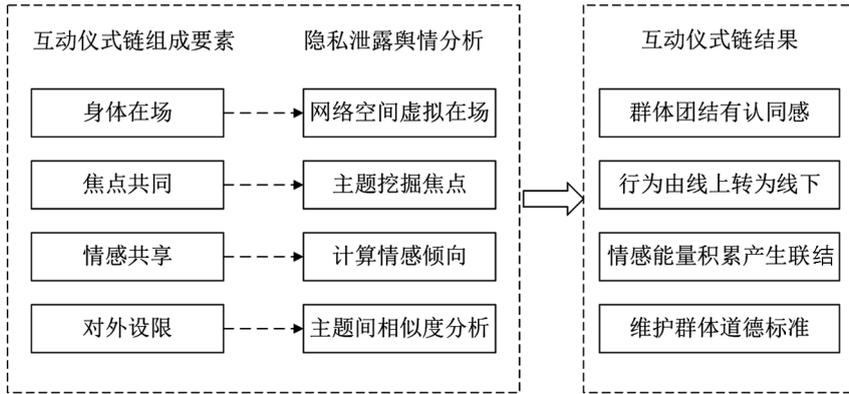


图 2 主体关系图

2.2 BERTopic 主题挖掘

BERTopic 是基于深度学习的一种主题建模方法^[27], 利用 BERT 嵌入和基于类的 TF-IDF 创建密集集群, 可以对大量的短文本数据进行高效的建模, 非常适合用于对豆瓣讨论等文本进行主题提取的任务. 该模型的原理如图 3 所示, 在 BERTopic 的每一阶段都可以选择更加匹配数据的处理办法. 本研究根据所选文本特征, 首先利用 SpaCy 文本预处理库对数据进行处理, 再用 UMAP 降维算法(Uniform Manifold Approximation and Projection for Dimension Reduction)对词嵌入向量进行降维处理^[28], 通过 HDBSCAN 聚类算法(Hierarchical Density-Based Spatial Clustering of Applications with Noise)对降维后的向量进行聚类, 最后提取与主题最为相关的候选词.

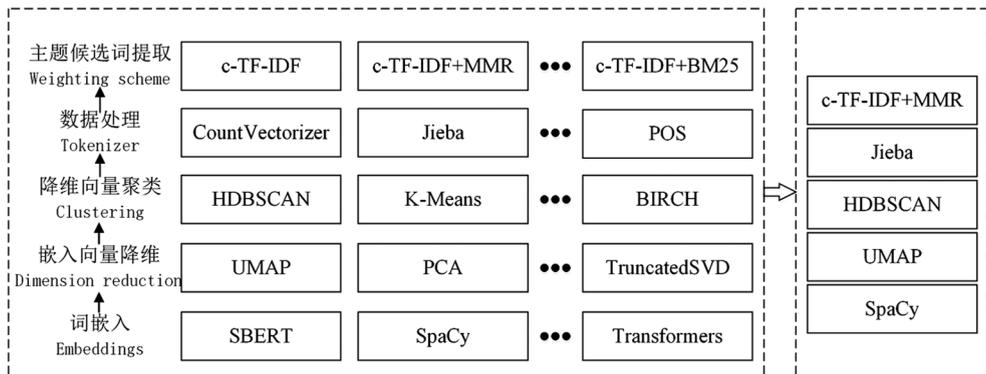


图 3 BERTopic 模型原理

根据公式(1)计算 BERTopic 模型中所得主题内的词重要性, 以 TF-IDF 为基础的改进 c-TF-IDF, 挖掘出每个主题中的表义词. $W_{x,c}$ 表示词 x 在语义簇 c 中的重要性得分, $tf_{x,c}$ 表示词 x 在语义簇 c 中出现的频率, f_x 表示所有语义簇中词 x 出现的频率, Ave 表示每个语义簇中出现词汇数的均值.

$$W_{x,c} = tf_{x,c} \cdot \log\left(1 + \frac{Ave}{f_x}\right) \quad (1)$$

对所获取的文本进行基本处理之后, 利用 BERTopic 进行主题挖掘及主题词提取, 获取某一讨论的主题概率分布图, 如图 4 所示. 由该图可以看出, 该讨论内容来自某一个主题的概率比其他主题的高, 其主题分布较为准确. 将提取出的主题进行相似度对比, 结果如图 5 所示, 各主题间相似程度较低, 利用 BERTopic 对豆瓣小组的文本进行主题抽取能够取得较好的效果.

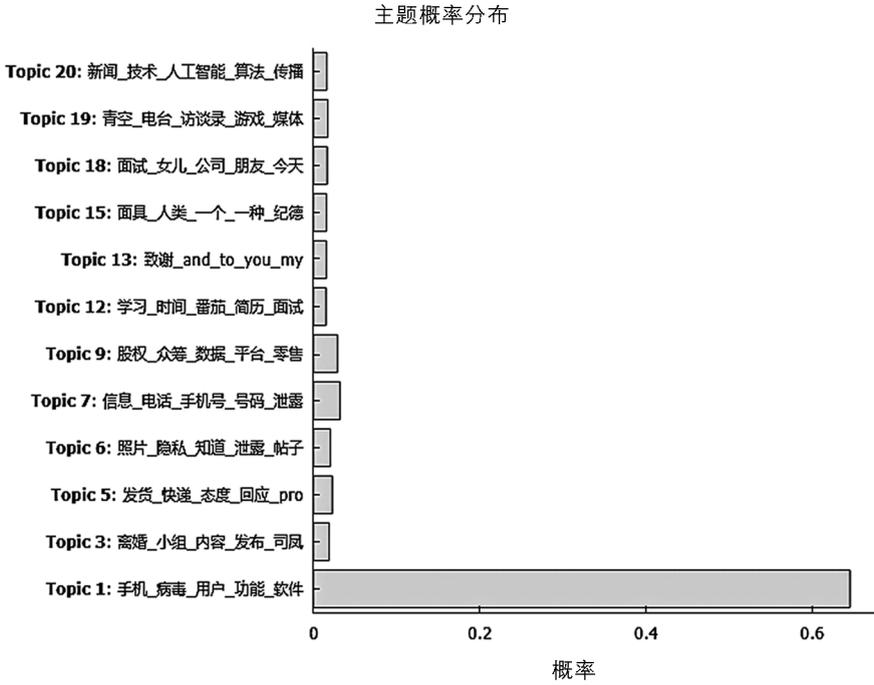


图 4 主题概率分布图

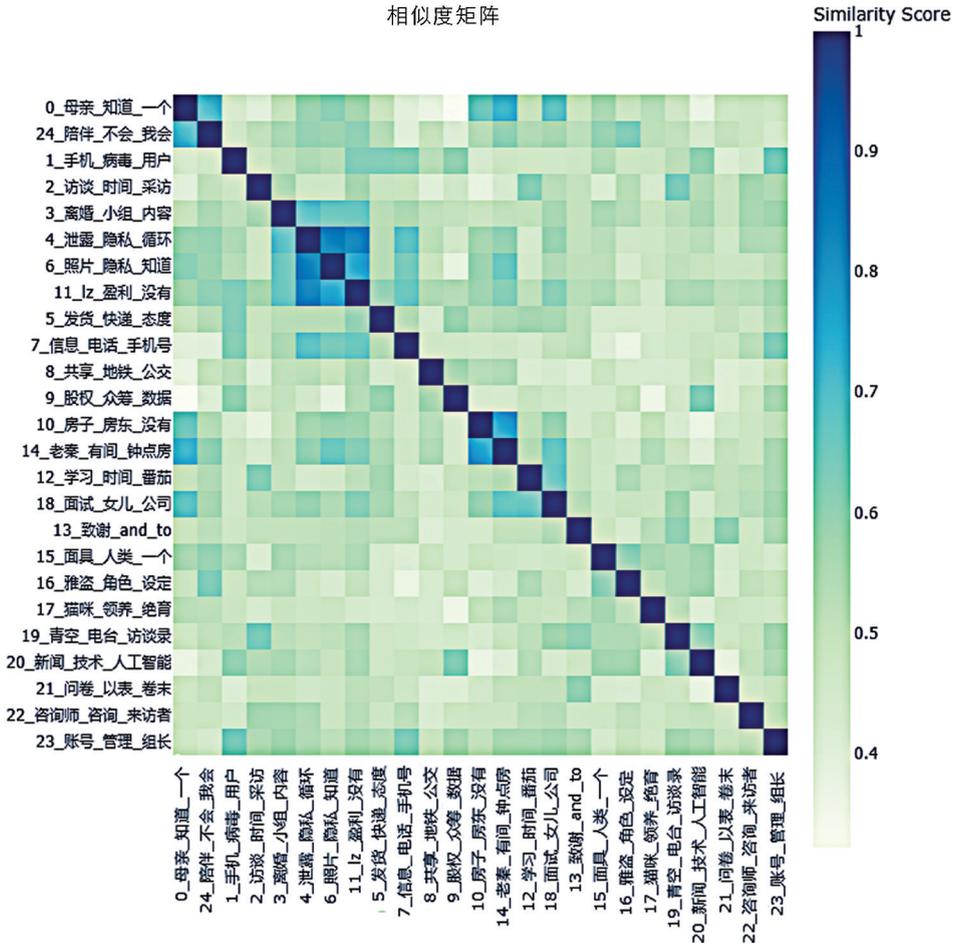


图 5 主题相似度热力图

利用余弦相似度计算不同类别间的主题相似程度, 对文本内容进行分词后向量化, 将两个向量的空间夹角的余弦值作为两个文本的相似度, 值越大相似度越高, 利用公式(2)判断两个向量夹角余弦, 其中 \mathbf{A} 、 \mathbf{B} 为两个 n 维向量.

$$\cos\theta = \frac{\sum_{i=1}^n (\mathbf{A}_i \times \mathbf{B}_i)}{\sqrt{\sum_{i=1}^n (\mathbf{A}_i)^2} \times \sqrt{\sum_{i=1}^n (\mathbf{B}_i)^2}} \quad (2)$$

2.3 情感倾向分析

情感分析与主题挖掘相结合能对网络舆情中的观点态度有更多维度的识别. 在现有的情感分析方式中, 百度情感倾向分析结合了情感词典和机器学习两种情感分析方法, 相较于单一情感分析, 具有更高的效率及准确度. 其情感倾向词库中含有来自各个领域的正、负向语料库, 在这些语料库的基础上训练模型使得通用准确率达到 92%^[29], 普适性以及准确性都较高, 能够为本研究提供较为准确的实验结果.

百度情感倾向分析通过对文本信息的情感分类(积极、消极)和对应的可信度来实现情感评价. 将数据通过百度 AI 的接口输入, 得到情感分类结果及可信度. 主要步骤为: ① 通过 http 请求向百度 AI 接口提交需要处理的文本数据. ② 百度 AI 处理后返回情感极性分类结果, 正面情感与负面情感的概率, 方便之后对情感结果展开进一步分析研究.

除此之外, 在将数据输入百度情感分析模型前, 需要人工清洗文本内容, 将表情符号及相关网络用语转换为常用语. 部分用户习惯于使用 emoji 或颜文字表达情感, 这部分内容在处理时需选择与其意思相近的词语做替换, 以得到更准确的情感分类结果.

3 研究过程

社交媒体的发展使互动仪式链组成要素中的身体在场能够以虚拟身份表达, 对外设限这一要素也由于各网络平台的特性得以成立. 同时豆瓣小组具有的泛连接用户关系、审核申请制的特性也决定了组内成员有着共同的关注点及共享的情感状态. 互动仪式链的 4 个基本要素在豆瓣小组这一平台上都被满足.

3.1 数据采集与预处理

选取豆瓣为数据来源, 借助八爪鱼采集器以“隐私泄露”为关键词爬取评论数据共 1 539 条, 以哈工大停用词表为基础, 结合本文数据采集过程中出现的无意义字符与数值等, 对文本中的无效内容进行剔除. 借助 jieba 分词工具进行分词, 再通过关键词共现展示词组两两出现在一篇文档中的次数, 揭示词组之间的逻辑关系进而分析它们所代表文档的结构变化. 经过数据去重后得到 1 102 条数据, 时间跨度为 2006 年 11 月至 2023 年 3 月, 各年份分布图如图 6 所示.

自 2006 年至今, 豆瓣中关于“隐私泄露”的讨论在 2009 年、2014 年、2021 年出现过 3 次起伏. 2009 年豆瓣开始实行组长实名制, 用户需提供身份信息等内容; 2014 年奇虎 360 与腾讯公司关于隐私侵占垄断的诉讼以 360 败诉收尾, 该类事件被推向公众视野; 2021 年央视 315 曝光了三起规模较大的隐私泄露事件, 同时健康宝通行码等疫情相关小程序引发较多讨论.

根据豆瓣官方对小组的类别标签进行简单分类, 涉及“隐私泄露”话题的小组类别共 20 个, 各小组类别占比如图 7 所示. 生活类小组占比最多达到 27.86%, 科技、情感、校园、职场类小组关于“隐私泄露”有较多讨论, 其讨论主题各有侧重.

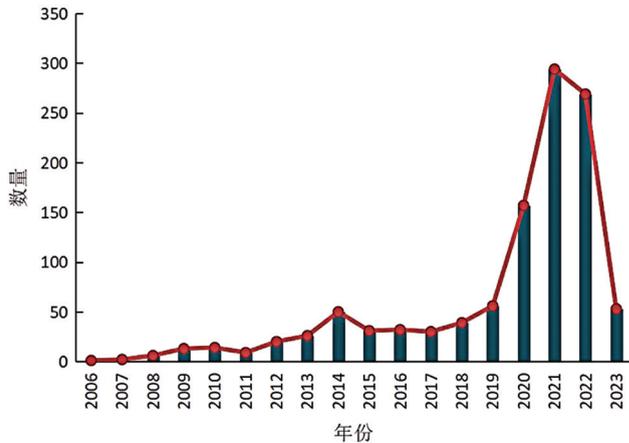


图 6 隐私泄露讨论热度

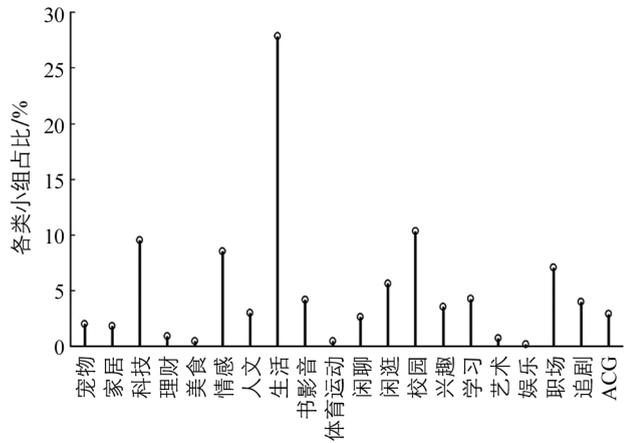


图 7 隐私泄露话题各类小组讨论占比

在对隐私泄露问题的讨论中,整体关注点多集中在社交媒体使用过程中是否存在泄露隐私、避免隐私泄露的技术及身边隐私泄露相关事件.除了用户自身行为导致的隐私泄露,一些被动的泄露行为也会引起广泛讨论,如公司或学校无法避免的信息填写等.在关键词提取后构建共现矩阵,选取其中共现次数较高的关键词,如表 1 所示,其共现次数越多说明二者关系越密切.其中信息、手机及用户等词都有较高频次的共现,说明手机使用过程中产生的信息泄露问题是公众关注的重点.除此之外,泄露方式及渠道也较多被讨论.

表 1 共现矩阵

	信息	手机	用户	问题	数据	网络	平台	个人信息	方式
信息	0	100	123	148	74	97	96	86	95
手机	100	0	79	93	40	39	36	29	48
用户	123	79	0	95	73	57	63	49	67
问题	148	93	95	0	76	68	66	50	92
数据	74	40	73	76	0	49	40	34	56
网络	97	39	57	68	49	0	48	51	55
平台	96	36	63	66	40	48	0	50	48
个人信息	86	29	49	50	34	51	50	0	42
方式	95	48	67	92	56	55	48	42	0

3.2 焦点共同——主题识别

借助 BERtopic 主题挖掘工具对讨论文本进行主题分析,利用 LDA 主题提取进行辅助分析,随机选取 4 个类别的某条讨论绘制文档—主题分布图,如图 8.从图中可以看出所选文档的某 1 个或 2 个主题概率相比其他主题概率高,即文档均存在主要主题,较好地实现了文本主题的划分.

在互动仪式链中,组内成员有着共同的关注焦点,柯林斯认为个体在参与到集体的互动中时有较强的身份认同感,周围群体的讨论易引起用户积极参与.对各类小组所关注的主题进行挖掘,其中科技、生活、校园及职场 4 个小组的主题数量及其概率词如表 2—表 5 所示.科技类小组主要关注隐私泄露的具体类别,如后台程序或者照片定位等,同时关注苹果等手机品牌在数据安全方面的优劣性.在 2010 年腾讯与奇虎 360 之争中发布讨论的也多为科技类小组,进一步说明该类小组讨论侧重于产品或软件而非用户行为.

生活组对于隐私泄露问题的讨论多集中在身边事,相较于科技组侧重从数据角度避免隐私泄露,生活组更注重隐私泄露后的补救措施及情绪抚慰.同时生活组的讨论主题还包括购物软件、快递和外卖等涉及私人信息的隐私保护问题,以及提出切实有效的隐私保护措施,如涂黑快递信息等.相较于男性,女性更加担心隐私泄露问题,Topic1、Topic6 中都提及到女性隐私泄露问题.

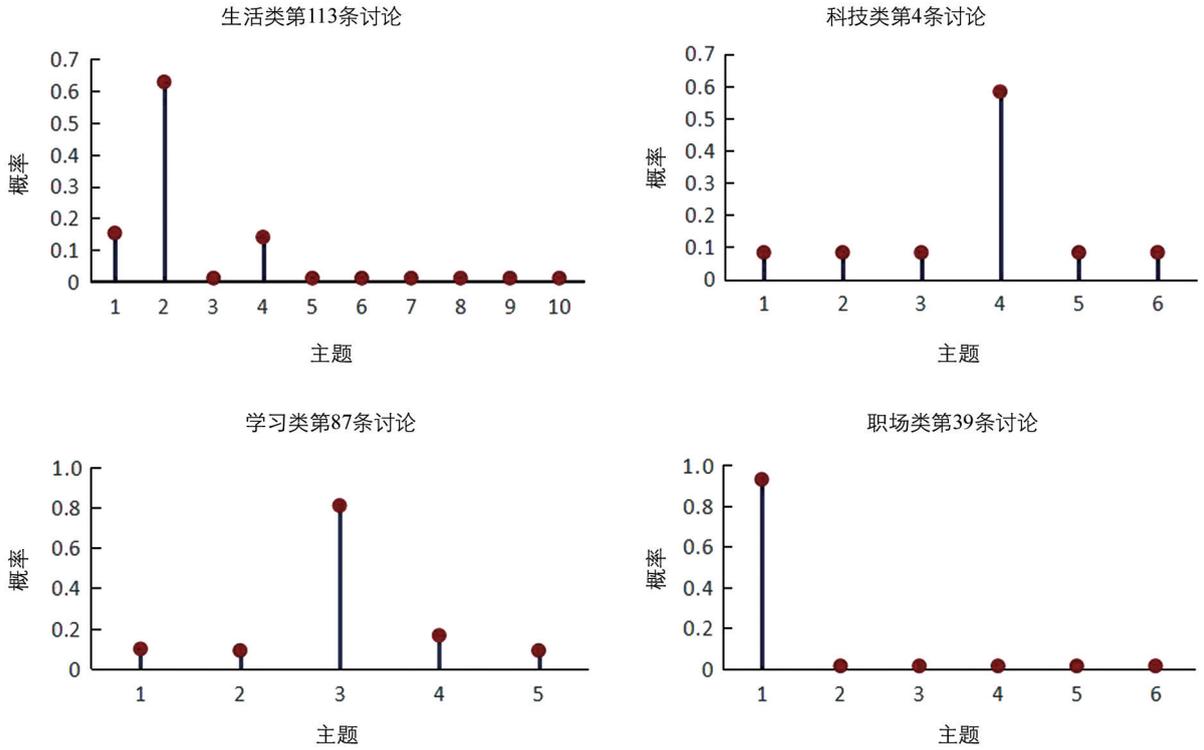


图 8 文档-主题分布

表 2 科技类小组隐私泄露讨论主题分类

主题编码	主题简介	主题概率词									
Topic1	软件后台运行泄露信息	测试	软件	垃圾	后台	工作	版本	权限	用户	功能	设置
Topic2	系统获取较多用户权限	音乐	用户	漏洞	信号	人员	系统	密码	版本	信息	三星
Topic3	聊天定位泄露地址	朋友	定位	时间	地图	好友	体验	信息	功能	聊天	版本
Topic4	照片备份不安全	文件	设备	应用	程序	智能	管理	照片	备份	工具	自动 传输
Topic5	互联网数据信息安全问题	数据	技术	发展	互联网	信息	安全	风险	系统	信息	公司
Topic6	腾讯与 360 管家之争	系统	企业	云盘	电脑	腾讯	软件	服务	管家	模式	记录
Topic7	适老化设备隐私问题	智能	老人	设备	谷歌	功能	产品	平台	工作	无线	摄像头
Topic8	购物平台泄露信息	广告	照片	免费	视频	微信	购物	用户	游戏	利用	录音
Topic9	苹果数据安全性	苹果	数据	屏幕	功能	平板	团队	服务	设计	三星	云盘
Topic10	信息泄露导致短信骚扰	病毒	百度	移动	电信	浏览器	短信	骚扰	程序	传播	监听

表 3 生活类小组隐私泄露讨论主题分类

主题编码	主题简介	主题概率词									
Topic1	朋友遇到隐私泄露	关系	女生	影响	专业	情绪	联系	建议	微信	内容	朋友
Topic2	职业泄露	职业	正确	问题	信息	话题	原因	微信	情绪	交流	专业
Topic3	看见相关发帖的讨论	内容	帖子	理由	帮助	话题	建议	故事	问题	帮忙	大量
Topic4	隐私泄露导致的害怕情绪	朋友	照片	害怕	感觉	过程	事情	联系	手机	名字	经历
Topic5	购物平台泄露隐私	店家	地址	姐妹	建议	记录	无法	威胁	记录	店铺	信息
Topic6	生活中避免隐私泄露	家庭	影响	外卖	电话	打火机	措施	快递	难受	帮助	生气

表 4 校园类小组隐私泄露讨论主题分类

主题编码	主题简介	主题概率词
Topic1	论文是否存在信息泄露	论文 工作 朋友 学校 专业 手机 过程 研究 同学 信息
Topic2	问卷调查泄露隐私	视频 问卷 内容 毕业 论文 文字 时间 信息 语音 信息 微信
Topic3	社交媒体泄露隐私	媒体 研究 个人 对象 用户 信息 数据 过程 手机 现象
Topic4	论文查重平台是否泄露	研究生 学校 查重 形式 时间 对象 主题 私信 老师 可以
Topic5	信息采集泄漏隐私风险	车票 学校 身份证 情况 微信 模式 共享 无法 感兴趣 风险

表 5 职场类小组隐私泄露讨论主题分类

主题编码	主题简介	主题概率词
Topic1	服务业摄像头数量的讨论	行业 服务 公司 问题 摄像头 老板 联网 影响 过程 目的
Topic2	招聘平台泄露简历	项目 个人 信息 简历 招聘 网络 行业 情况 朋友 经历 辞职
Topic3	隐私泄露问题寻求法律帮助	法律 工作 身份证 劳动 单位 建议 专业 经验 问题 目的
Topic4	公司或许存在隐私泄露行为	公司 证据 工作 方式 法律 网络 联系 辞职 拷贝 事情
Topic5	联网项目的隐私泄露风险	电脑 网络 经验 联网 无法 项目 姓名 信息 建议 禁止
Topic6	隐私泄露案件的法律后果	法律 后果 起诉 方式 对方 结果 影响 工作 事情 时间

职业属性集中的校园类小组和职场类小组在关于隐私泄露问题的讨论中侧重点不同。校园类小组更多关注论文中隐私泄露的问题。如 Topic1 所示,包括需要完成的问卷及论文的查重是否造成隐私泄露等。由于学生较多受学校管束,所以部分用户对学校的信息采集等必要措施也进行了讨论,他们认为共享文档等内容存在泄露隐私的隐患。我国在校大学生寒暑假期间往往有跨区域流动,Topic5 说明他们出行过程中更加担心车票等是否会泄露信息,共享单车等出行软件是否有良好的隐私保护性。与其他群体相比,学生群体愿意花费更多的时间在社交媒体上,接触不同类的社交媒体也使得该类小组对于社交媒体泄露隐私的问题更加关注,其中就有抖音等平台随手拍行为是否注意到了路人打码等问题的讨论。

与校园类小组不同的是,职场类小组更加侧重对隐私泄露行为可能会带来的后果及风险的分析,更加注重从法律角度对隐私泄露问题进行解读,在严重的隐私泄露事件中,个人力量在法律上能否与企业所抗衡。Topic1 显示,部分公司摄像头数量较多,公司内部网络泄露员工隐私等现象都引起了较多讨论。Topic2 说明由于个人简历往往包含较完整的个人信息,有部分用户担心在求职过程中被招聘软件非法收集简历造成信息泄露。

3.3 情感共享——情感倾向分析

用户在发布讨论时会带有情感词,这些词主要反应了用户在对共同关注的焦点讨论后的情感状态。对用户情感倾向进行分类,选取其中 10 类小组的情感分布,如图 9 所示。在同一小组中,用户情感状态容易受到组内其他讨论的影响,相互关注的焦点也影响着用户的情感,随着互动的推进,用户情感不断波动而产生情感能量的积累。在隐私泄露事件的讨论中,负面情感在职场、闲聊、校园、美食、理财、家居及 ACG 类小组中占比较高。其中校园类小组关于隐私泄露事件的讨论最为负面,其原因可能是学生群体在情感上容易冲动,情感起伏较大,而娱乐类小组正面情感最多,该小组属性偏向乐观,组内整体情感较为积极,所以对隐私泄露事件的讨论也偏向正面。而情感类、生活类等小组由于其组内情感趋于中性,日常讨论内容较为平和,故对于隐私泄露事件的情感并未出现较明显的波动。

总体而言,用户对于隐私泄露的容忍度维持在一定阈值范围内,在该范围内情感倾向偏向积极。如 3.2 节中所提,校园类 Topic5 及职场类 Topic3 都涉及到基本身份信息的泄露,该些讨论使得整体情感倾向偏向消极。而在娱乐类小组中情感倾向较为积极,原因在于组内讨论多为各类防止隐私泄露的措施,及

接到部分骚扰电话的积极应对行为。部分用户认为,大数据带来的便利在某种程度上是以隐私泄露为代价的,故对于使用这类软件造成的隐私泄露保持积极情绪。除此之外,对于政府部门进行防疫等必要举措时产生的隐私披露,公众大都表示可以理解并接受。

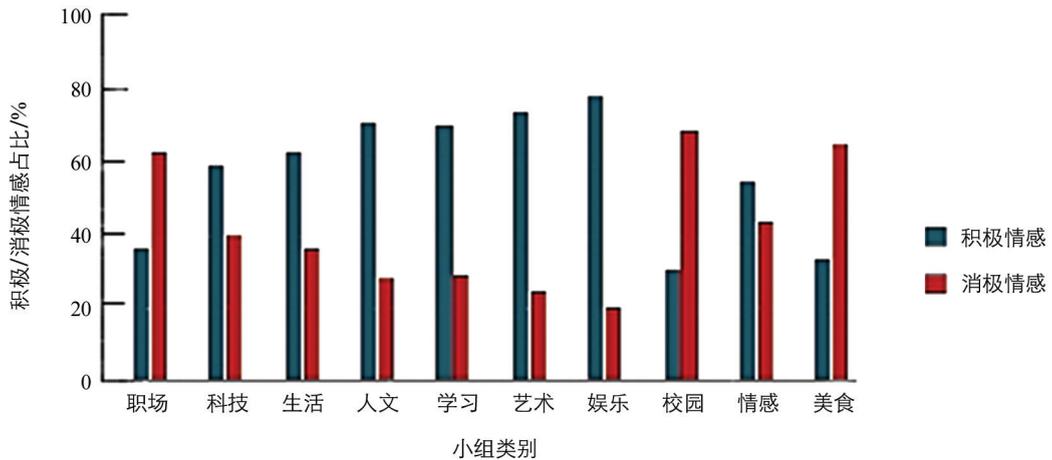


图9 各类小组情感倾向

3.4 对外设限——主题相似度计算

互动仪式链中的一个基本要素是对局外人设限,而豆瓣小组的进组申请审核制满足这一要素,组外成员的讨论受到一定的限制。利用余弦相似度计算各类小组在讨论隐私泄露事件时所涉及的主题是否相似,并从中选取10个结果,如表6所示。大部分小组间所讨论的主题相似度都较低,但生活类小组因为其内容涉及面较为广泛,与各个小组都有交集,所以与其他小组间主题相似度略高。

表6 部分类别小组间主题相似度

	ACG	宠物	家居	科技	理财	美食	情感	人文	生活
ACG	1.00	0.12	0.29	0.30	0.22	0.17	0.40	0.30	0.44
宠物	0.12	1.00	0.17	0.13	0.13	0.07	0.22	0.16	0.22
家居	0.29	0.17	1.00	0.37	0.29	0.18	0.52	0.36	0.51
科技	0.30	0.13	0.37	1.00	0.29	0.19	0.42	0.52	0.70
理财	0.22	0.13	0.29	0.29	1.00	0.16	0.40	0.37	0.46
美食	0.17	0.07	0.18	0.19	0.16	1.00	0.29	0.16	0.30
情感	0.40	0.22	0.52	0.42	0.40	0.29	1.00	0.45	0.72
人文	0.30	0.16	0.36	0.52	0.37	0.16	0.45	1.00	0.63
生活	0.44	0.22	0.51	0.70	0.46	0.30	0.72	0.63	1.00

4 讨论与分析

互动仪式不断进行构成互动仪式链,用户在互动仪式链中不断参与讨论、分享观点、表达情感,感受到所在群体的成员身份,产生归属感,进而对破坏群体团结的行为产生愤怒。共同关注的焦点与互相影响的情感使群体中产生了专属符号,在对隐私泄露事件的态度表达及预防措施讨论中,对用户线下行为的实践产生一定的影响。互动仪式链产生群体团结、情感联结与线下行为3个结果,将互动仪式结果与隐私泄露相关舆情结合起来提出措施建议,帮助社交媒体平台与相关部门进一步完善隐私保护措施。

4.1 借助群体团结增强隐私意识

隐私泄露具有较强的危害性与破坏性,而我国公民尚未形成较强的个人隐私保护意识,不能充分认识到身份证等重要信息在网络传播可能造成的危害。互联网的不断发展使得隐私泄露事件不断发生,电信诈

骗等利用已泄露的信息实施的违法行为,在中老年群体中较为常见,给社会稳定造成了一定影响。互动仪式产生的群体团结使得豆瓣各小组内部对于隐私泄露问题持不同看法,科技方面关注较多的群体对隐私泄露更加敏感,如 2022 年豆瓣平台展示用户 IP 地址引发该类小组较为激烈的讨论,而生活类小组却较多保持平和看待该事件。

由于群体团结这一互动仪式的结果,各小组内部关注点较为统一。相关部门可利用群体团结这一特性,对于各小组可能产生的衍生主题提出预见性的举措,提前对关键问题进行应对,借助对违反群体团结的愤怒情绪推动政策落地。同时为了避免严重的隐私泄露,在数据发展与保护信息之间达到平衡,还需要多方面举措并行。各类 APP 对于隐私政策需要更加审慎,做到公开透明。在隐私政策制定中体现出尊重用户的更能获得用户信任^[30],也可以避免用户的“被自愿”行为。除职场类小组外,其他类小组对于隐私泄露可能产生的后果没有明确的认知,各社交媒体在日常推送中需增加相关内容的表述,以提升用户隐私保护意识,并及时监测各群体焦点的主题演化过程,采取有效策略减缓可能的负面影响。

在隐私泄露问题上维持动态的平衡点,利用小组管理员的举报反馈机制来调整组内相关内容的比例,达到公众知情权与隐私权的互相完善。比如,豆瓣和其他社交媒体平台在疫情流调等情况下可能需要公布用户个人信息,这时可以根据各小组的特点进行正向引导,发布脱敏的讨论内容,以减少用户的抵触情绪。同时,还应该聚焦各群体的痛点问题,例如校园组的“论文泄露”、ACG 组的“账户被盗”以及生活组的“地址威胁”,集中处理以减少由隐私问题引发的负面舆情。

4.2 基于情感联接进行疏导

通过不断进行互动仪式,参与者逐渐获得情感能量,并从早期不稳定的情感状态逐渐转为稳定。参与者的情感会受到组内情感的影响,当组内情感能量积极时,参与者的情感也偏向正面。利用这一仪式结果,可以疏导隐私泄露引发负面影响的舆情,避免舆情态势恶化。

根据对各类小组内隐私泄露问题的情感倾向进行分析,发现校园类小组和职场类小组的情感倾向较为负面,涉及外卖泄露地址和理财产品信息不安全等讨论时会呈现负面情感。研究表明,这些小组中的隐私泄露问题多为被动、无法避免的情况,需要用户自己承担后果,因此组内的情感低落。由于互动仪式中的回音室效应,信息在封闭的空间内不断重复,导致情感的积累和群体极化现象。通过情感联结可以对舆情产生反作用,将情感能量转化为促进正面舆情的枢纽,减少负面反馈,从而疏导消极情感。

当严重的隐私泄露事件发生时,通常会引发大量负面舆情。因此,平台监管部门应提升监管能力,建立情感词语料库,结合关键词及时捕捉负面情感信息,避免隐私泄露范围的进一步扩大。同时,结合各群体内隐私泄露问题产生的情感演化过程,制定全面的预案以应对波动趋势。由于疫情防控等措施的推进,用户对于隐私泄露的容忍度有所改变,合理的隐私披露在一定程度上可以被部分用户接受,情感也更偏向正面。官方媒体在公布个人信息时需对其内容用途进行说明公示,并禁止隐私信息的转发扩散,从而增加公众的信任度。此外,公布隐私事件的处理结果,打断负面情绪的恶性循环。

4.3 利用行为结果完善隐私保护

相互关注的焦点及情感共享的状态使得互动仪式参与者之间形成较为紧密的社会关系和较强的信任度,用户对破坏群体团结的行为产生愤怒情绪,进一步发表维护群体的言论或线下实践。当用户对讨论的内容产生认同,并且情感共鸣较高时,其所认定的内容对线下行为产生一定的影响。利用这一特点提出相关建议,促使隐私被更好的保护,在网络发展与信息泄露之间达到平衡。

隐私泄露问题中的行为结果包含两方面,一方面是对隐私保护措施的讨论,例如涂掉快递信息、设置外卖虚拟电话号码以及关闭不必要的软件权限获取等;另一方面是对隐私泄露事件的扩大行为,转发和点赞等行为,这会增加隐私泄露的关注度,同时部分谴责的讨论也会进一步传播被泄露的信息。同时警惕这种扩大行为容易引发的次生舆情,尤其是在社交媒体平台上,微信、抖音等平台加速了信息泄露。因此,需

建立隐私保护优化模型,并善用举报和投诉机制来应对这一问题。但生活中的隐私泄露问题往往是无意识的,为此,各社交媒体平台在用户隐私政策中应特别强调相关内容,以提醒用户注意隐私保护。

信息化建设的不断深化,推出的线上办公小程序等内容,在便利生活的同时都面临着隐私问题。为了应对隐私问题,政府部门应建立多种信息的反馈渠道,以便了解存在的隐私风险,根据已经发生的隐私泄露事件,完善管理体系,并检查信息安全漏洞。同时,对隐私泄露的惩处措施予以公示,强调其法律后果,避免再次发生类似事件。为了提高个人隐私保护意识,应将线上的情感传递与线下的行为结果进行结合,遵循平台隐私管理条例,保持良好的社交环境。相关企业完善信息安全保护机制,采用加密算法并对其安全性定期评测,以完善隐私保护体系。

5 结语

本文围绕隐私泄露这一话题,从互动仪式链视角出发,分析了豆瓣各类小组的互动机制。结合 BER-Topic 主题挖掘及情感倾向性分析方法,对这类事件中的焦点问题及产生的情感进行了分析。研究结果显示,不同群体由于职业、性别和兴趣的不同,在隐私泄露问题上持有不同的观点。科技类小组更关注技术问题及数据安全性,职场类小组更关注发生隐私泄露问题的法律后果和公司合法性,生活类小组中表现出女性较男性更为注重隐私保护这一特点。处于互动仪式链中的成员在共同关注的焦点下有着共享的情感状态,其组内讨论影响着个人的情感能量。具有职业属性的校园类和职场类小组情感较为负面,而娱乐类和生活类小组整体情感较积极,对隐私泄露容忍度较高。对于隐私泄露问题,可以通过引导互动仪式链产生的群体团结、情感联接和相应行为结果,提高公众的隐私素养,并完善隐私保护体系。

本研究还存在一定的局限性。首先,以豆瓣小组为研究平台,缺少多平台案例的对比分析。其次,在主题挖掘阶段未对讨论中的图片及表情包做处理,情感分析阶段对于多模态的内容未进行更进一步的研究。在后续研究中考虑融合影响较大的隐私泄露事件,多维度对该类事件的演化规律进行讨论,借助深度学习等多模态的分析方法,更准确地挖掘公众对隐私泄露现象的看法,并逐步关注不同类别事件隐私泄露的演化过程,以进一步加强隐私保护措施。

参考文献:

- [1] 池毛毛,王俊晶,王伟军. 突发重大疫情下基层政府信任对公民信息不披露意愿的影响机制研究 [J]. 情报学报, 2021, 40(6): 666-678.
- [2] 兰德尔·柯林斯. 互动仪式链 [M]. 林聚任,王鹏,宋丽君,译. 北京:商务印书馆,2017.
- [3] 谷学强. 互动仪式链视角下网络表情包的情感动员——以“帝吧出征 FB”为例 [J]. 新闻与传播评论, 2018, 71(5): 27-39.
- [4] 沈霄,王国华. 网络直播+政务与用户的互动研究——基于互动仪式链的视角 [J]. 情报杂志, 2018, 37(5): 111-116, 175.
- [5] 李钧鹏,茹文俊. 论虚拟社区中的互动仪式链 [J]. 广东社会科学, 2020(4): 201-211.
- [6] 赵雪芳. 互动仪式链视角下网络舆情的演化及连带效应研究——以“错换人生 28 年”事件为例 [D]. 天津:天津师范大学, 2022.
- [7] 魏来,王伟洁. 基于互动仪式链理论的视频网站弹幕信息情感分析——以 Bilibili 健康科普类视频为例 [J]. 情报理论与实践, 2022, 45(9): 119-126.
- [8] CLARKE J M, WARING J. The Transformative Role of Interaction Rituals within Therapeutic Communities [J]. Sociology of Health & Illness, 2018, 40(8): 1277-1293.
- [9] 吴娜,李文立,吕欣,等. 泄露他人隐私行为意向的影响要素研究 [J]. 科研管理, 2015, 36(11): 139-147.
- [10] 陆雪梅,古春生. 大数据环境下用户信息隐私泄露成因分析和保护对策 [J]. 现代情报, 2016, 36(11): 66-70.
- [11] HALLAM C, ZANELLA G. Online Self-Disclosure: The Privacy Paradox Explained as a Temporally Discounted Balance between Concerns and Rewards [J]. Computers in Human Behavior, 2017, 68: 217-227.

- [12] 金元浦. 大数据时代个人隐私数据泄露的调研与分析报告 [J]. 清华大学学报(哲学社会科学版), 2021, 36(1): 191-201, 206.
- [13] 易红, 任竞. 图书馆大数据服务环境下用户隐私泄露容忍度的实证研究 [J]. 图书馆论坛, 2016, 36(4): 57-64.
- [14] BAZAROVA N N, MASUR P K. Towards an Integration of Individualistic, Networked, and Institutional Approaches to Online Disclosure and Privacy in a Networked Ecology [J]. Current Opinion in Psychology, 2020, 36: 118-123.
- [15] 臧国全, 张凯亮. 医疗数据隐私泄露容忍度的计量分析 [J]. 信息资源管理学报, 2020, 10(4): 70-78, 108.
- [16] 李水莲, 张陈杰. 不合规 App 隐私政策文本结构与情感倾向研究 [J]. 青年记者, 2022(18): 57-59.
- [17] 王忠. 大数据时代个人数据隐私泄露举报机制研究 [J]. 情报杂志, 2016, 35(3): 165-168, 79.
- [18] MASUR P K, SCHARKOW M. Disclosure Management on Social Network Sites: Individual Privacy Perceptions and User-Directed Privacy Strategies [J]. Social Media + Society, 2016, 2(1): 205630511663436.
- [19] 王晰巍, 刘宇桐, 李玥琪. 突发公共卫生事件中公民隐私泄露舆情的情感演化图谱研究 [J]. 情报理论与实践, 2022, 45(3): 19-27.
- [20] 尹德虎. 在线健康社区中基于 LDA 模型的话题热度动态演化趋势研究 [D]. 昆明: 昆明理工大学, 2019.
- [21] 胡吉明, 钱玮, 李雨薇, 等. 基于 LDA2Vec 的政策文本主题挖掘与结构化解析框架研究 [J]. 情报科学, 2021, 39(10): 11-17.
- [22] 谭春辉, 陈晓琪, 梁远亮, 等. 隐私泄露事件中社交媒体围观者情感分析 [J]. 情报科学, 2023, 41(3): 8-18.
- [23] 霍朝光, 霍帆帆, 董克. 基于 LSTM 神经网络的学科主题热度预测模型 [J]. 图书情报知识, 2021(2): 25-34.
- [24] 杜毅贤, 徐家鹏, 钟琳颖, 等. 网络舆情态势及情感多维特征分析与可视化——以 COVID-19 疫情为例 [J]. 地球信息科学学报, 2021, 23(2): 318-330.
- [25] 张柳, 王晰巍, 黄博, 等. 基于字词向量的多尺度卷积神经网络微博评论的情感分类模型及实验研究 [J]. 图书情报工作, 2019, 63(18): 99-108.
- [26] 常城扬, 王晓东, 张胜磊. 基于深度学习方法对特定群体推特的动态政治情感极性分析 [J]. 数据分析与知识发现, 2021, 5(3): 121-131.
- [27] GROOTENDORST M. BERTopic: Neural Topic Modeling with a Class-Based TF-IDF Procedure [EB/OL]. (2022-03-11) [2023-05-21] arXiv: 2203. 05794. <https://arxiv.org/abs/2203.05794>.
- [28] MCINNES L, HEALY J, SAUL N, et al. UMAP: Uniform Manifold Approximation and Projection [J]. Journal of Open Source Software, 2018, 3(29): 861.
- [29] 赵凯, 王鸿源. LDA 最优主题数选取方法研究: 以 CNKI 文献为例 [J]. 统计与决策, 2020, 36(16): 175-179.
- [30] 刘百灵, 夏惠敏, 李延晖. 移动购物用户信息披露意愿影响因素的实证研究——基于公平理论和理性行为理论视角 [J]. 情报理论与实践, 2017, 40(5): 87-93.

责任编辑 孙文静