

DOI:10.12154/j.qbzlgz.2021.06.003

数据驱动的社交网络舆情极化群体画像构建研究*

王晰巍^{1,2,3} 贾若男¹ 孙玉姣¹¹吉林大学管理学院 长春 130022; ²吉林大学大数据管理研究中心 长春 130022;³吉林大学国家发展与安全研究院网络空间治理研究中心 长春 130022)

摘要: [目的/意义] 尽管社交网络为用户提供了自由表达观点的平台,但在社交网络舆情形成和发展过程中,网民形成的群体极化也带来了严重的社会、经济和政治冲击,理解群体极化现象成为社交网络舆情发展中的新问题。[方法/过程] 从数据层、分析层和概念层三个层面构建数据驱动的社交网络舆情极化群体用户画像模型,采用网络爬虫收集新浪微博平台中的“离婚冷静期”舆情话题数据,并对极化群体演化及类型、静态特征、动态特征、各时期极化群体进行画像分析。[结果/结论] 数据分析结果表明,极化意见动态演变过程中,极化群体呈现由多极化向两极化转变的趋势。各个时期不同立场的极化群体画像呈现出复合性特征,既存在相同规律又存在明显差异。本研究丰富了群体极化和画像研究的理论和方法,并通过实证分析为网络舆情的干预和引导提供了实践指导。

关键词: 社交网络 舆情 群体极化 画像

Research on the Construction of Data-driven Social Network Public Opinion Polarized Group Portrait

Wang Xiwei^{1,2,3} Jia Ruonan¹ SunYujiao¹¹School of Management, Jilin University, Changchun, 130022;²Research Center for Big Data Management, Jilin University, Changchun, 130022;³Cyberspace Governance Research Center, National Academy of Development and Security, Jilin University, Changchun, 130022)

Abstract: [Purpose/significance] Although social networks provide a platform for users to freely express their opinions, the group polarization formed by Internet users during the formation and development of social network public opinion also brings serious social, economic and political impacts. The phenomenon of understanding group polarization becomes a new problem in the development of social network public opinion. [Method/process] A data-driven user portrait model of polarized groups in social networks is constructed at three levels: data level, analysis level and conceptual level. A web crawler is used to collect data on the topic of “divorce cooling-off period” in Sina Weibo platform the evolution and types of polarized groups, static characteristics, dynamic characteristics, and portrait analysis of polarized groups in each period are also analyzed. [Result/conclusion] The results of data analysis show that polarized groups show a trend from multi-polarization to bipolarization during the dynamic evolution of polarized opinions. The portraits of polarized groups with different positions in each period show composite characteristics, with both the same patterns and obvious differences. This study enriches the theories and methods of group polarization and portrait research, and provides practical guidance for the intervention and guidance of online public opinion through empirical analysis.

Keywords: social network public opinion group polarization portrait

*本文系国家社科基金重大项目“大数据驱动的社交网络舆情主题图谱构建及调控策略研究”(项目编号:18ZDA310)的研究成果。

1 引言

2021年2月中国互联网信息中心发布的第47次《中国互联网络发展状况统计报告》指出,截至2020年12月我国网民规模达到9.89亿^[1]。短视频、即时通讯和网络新闻等社交平台的用户规模不断增长,自由表达的平台环境也催生了网络舆论环境的多变性和复杂性。《2020年互联网舆情形势分析与展望》指出,随着国内疫情形势趋于稳定非疫情的社会热点事件正在逐渐升温,网络舆论呈现丰富多元的发展态势^[2]。

社交网络为公众提供了信息交流和互动的平台,近几年公众关心的问题在社交网络平台中出现了网络舆情群体极化的现象,并带来多重影响。网络舆情中群体极化的产生受到舆情受众、舆情环境、舆情事项和受众观点的影响,群体极化现象也反过来影响舆情事件的进程和走向^[3]。现有一些研究成果对社交网络群体极化的正面和负面影响进行了分析,包括极化对公众交流和社区形成的促进,以及极化带来的种族隔离、政治影响和社会僵局^[4]等。目前对于社交网络极化的研究多从宏观角度来分析舆情分化的规律和成因,较少从个体特征、互动方式和群体中的个体需要等微观层面进行分析^[5]。用户画像可以实现用户数据的提取和抽象,能够直观地展现不同类型群体用户的全貌和关键特征^[6]。目前用户画像方法已经被广泛应用在精准营销、个性化服务和网络舆情治理等方面^[7]。然而从现有研究成果来看,基于数据驱动方法进行社交网络舆情极化群体画像研究的成果则相对较少。

本文在研究中试图解决以下3个研究问题:(1)在社交网络舆情事件中,怎样构建数据驱动的极化群体画像?(2)在社交网络舆情事件中,不同阶段的极化群体画像及其关键属性特征有何异同?(3)如何借助极化群体画像为社交网络舆情的监管引导和用户群体定位提供对策与建议?本文在对相关文献进行梳理的基础上,基于数据分析方法对社交网络舆情事件中的群体极化阶段进行划分,将用户画像应用到社交网络的极化群体分析中,结合用户静态属性特征和动态属性特征进行极化群体画像的构建,并对极化群体进行抽象概括。结合“离婚冷静期”舆情事件进行实证分析,以了解社交网络舆情事件中极化群体的动态变化及其关键特征。本文可为社交网络舆情和用户画像研究提供新的研究思路和视角,同时为社交网络舆情事件的监管引导、用户群体定位提供理论依据和实践指导。

2 相关概念及文献综述

2.1 社交网络舆情传播及演化

社交网络舆情,指的是社交网络环境中围绕特定事件的产生和发展,传达出的民众、媒体和政府等多个主体的意见、态度和情感等信息的综合^[8]。近年来,国内外学者对社交网络舆情的研究聚焦在舆情主体、舆情信息传播和舆情响应及舆情引导上,主要以各类型突发事件、虚假信息以及谣言作为研究对象。在舆情主体的研究上,主要探究政府、社会组织、媒体、意见领袖和网民等对事件热度和信息传播效果的影响和作用^[9]。

现有研究中认为社交网络舆情的演化可划分为三个阶段、四阶段、五阶段和六阶段^[10],并认为每个阶段的细分各有差异,但总体来看发展阶段是社交网络舆情演化中的主体要素之一,能够从动态角度揭示舆情演化的形态和规律。除此之外,还有研究根据突发事件的特点将社交网络舆情阶段划分为“四点四阶段”^[11]等。在社交网络舆情极化群体的研究中,对于极化群体阶段的划分多以舆情发展阶段为基础,再结合具体事件特征进行划分。如在对网民发帖极化现象进行频次分析的基础上,将群体极化阶段划分为潜伏期、成长期、成熟期和消退期4个阶段^[12];根据舆情讨论热度的变化将群体极化划分为预热期、高潮期、延续期、突变期和衰退期5个阶段^[13]等。

2.2 极化群体的成因与影响

极化群体最早是由 Stoner James^[14]提出,认为在群体决策情景中个体的观点会受到群体间彼此讨论的影响,从而产生群体一致性的结果。研究普遍将极化群体按分化的集体数分为单极化、双极化与多极化。在社交网络中,群体极化形成的异质群体在一定程度上促进突发事件和社会热点问题的交流和讨论,从而有利于社会的发展和进步。群体极化也会产生一定的负面影响,由于群体极化存在极端情感,在舆情事件中往往衍生出谣言和网络暴力等负向社会影响,因而对构建和谐良好的网络舆论环境带来较大的风险和压力^[15]。

国内外学者对极化群体的研究主要集中在群体极化形成的原因和影响上,采用定性、定量和混合研究方法。关于群体极化成因的研究表明,群体极化的分化程度受到舆情受众关系强度的变化、社交媒体的引导、时间的推移、社会比较频率、群体的认同、选择性接触和群际情绪等方面的因素驱动^[16-17]。至于群体极化影响的研究,国外较为注重群体极化对公众政治观点的影响,国内研究则侧重通过分析群体极化的动力机制,发

现群体极化现象所产生的正面和负面两种社会效应^[18]。

2.3 数据驱动的用户画像

用户画像的概念最早由美国交互设计之父 Alan Cooper^[12]提出,是一种基于用户真实数据的虚拟代表。用户画像将用户的需求、个性化偏好和兴趣描述为一个概念化集合,在社交平台用户个性化服务、商业智能营销以及舆情监控预警等方面已有广泛应用。用户画像的构建方法可以通过决策树、因子分析、层次聚类、K均值聚类^[19]等定量方法,或定性与定量相结合的混合研究方法来实现。

国内外研究主要从基于用户行为、用户兴趣偏好、用户关注的主题和用户的情绪^[12]四个方面进行用户画像的构建。基于用户行为的画像构建主要是通过网络日志或用户行为日志实现用户画像的分析^[11];基于用户兴趣偏好画像的构建主要是利用数学模型对用户兴趣进行追踪,并建立用户兴趣库或数学模型;基于用户关注的主题主要是利用用户生成的文本信息,通过LDA主题模型和贝叶斯网络等方法进行主题挖掘从而建立用户画像^[20];基于用户情绪的方法主要是通过对用户情绪状态的挖掘从而进一步完善用户画像,并对用户行为和情感倾向做出预测^[21]。

2.4 社交网络舆情极化群体画像问题的提出

学术界关于社交网络舆情群体极化的研究主要基于传播学、社会学和心理学等相关理论对舆情群体极化的成因和影响进行分析。从现有研究来看,国内关于舆情群体极化的研究大多倾向于从情感极化的角度对舆情的群体进行分析,重点关注极化的情感测度、极化机理、极化效应和极化原因等^[3];国外舆情极化群体的研究内容则相对较为丰富,除了对极化群体的动因、表现和影响进行研究外,还倾向于将极化群体分为不同的类型,如区别于情感极化的种族、意识形态和政党极化等^[22]。可见,国内外关于舆情群体极化的研究多从宏观的角度来对舆情的分化规律和成因进行分析,但是从群体极化中的用户群体这一角度进行研究的成果却相对较少。

极化群体画像作为单用户画像研究的纵深,能够在区分用户群体的基础上更加全面地呈现群体的极化特征,具有重要的研究意义和应用价值^[23]。群体画像的研究主要围绕群体兴趣偏好、行为

特征、群体间信息传播模式等展开,构建方法以本体、聚类、复杂网络分析和深度学习等数据挖掘技术为主^[24]。因此,将社交网络舆情群体极化与群体画像相结合,进行社交网络舆情极化群体画像构建,有助于准确地对舆情中的极化群体特征进行分析,从而对舆情进行更好的监管和治理。

3 数据驱动的社交网络舆情极化群体画像模型

3.1 画像模型

数据驱动的社交网络舆情极化群体画像的构建,是一个将用户扁平化数据进行层次化和概念化的过程,即从最底层的数据经过各种不同的数据分析方法,再到最顶端的人物角色进行概念化,从而将扁平化的用户数据变为一个具有高度概括性和标签化的画像。基于此,本研究构建了一个数据驱动的社交网络舆情极化群体画像模型如图1所示。数据驱动的社交网络舆情极化群体画像模型共分成3层,即数据层、分析层和概念层。数据层是社交网络采集的各类用户数据;分析层是通过各种方法对数据进行操作将数据转换为画像标签的过程;概念层是数据驱动的极化群体画像构建的最高抽象级,即概念化所有标签后生成的极化群体画像。

3.2 数据层

在数据层通过网络爬虫从社交平台获取用户数据,主要包括用户的性别、会员等级、微博数、关注

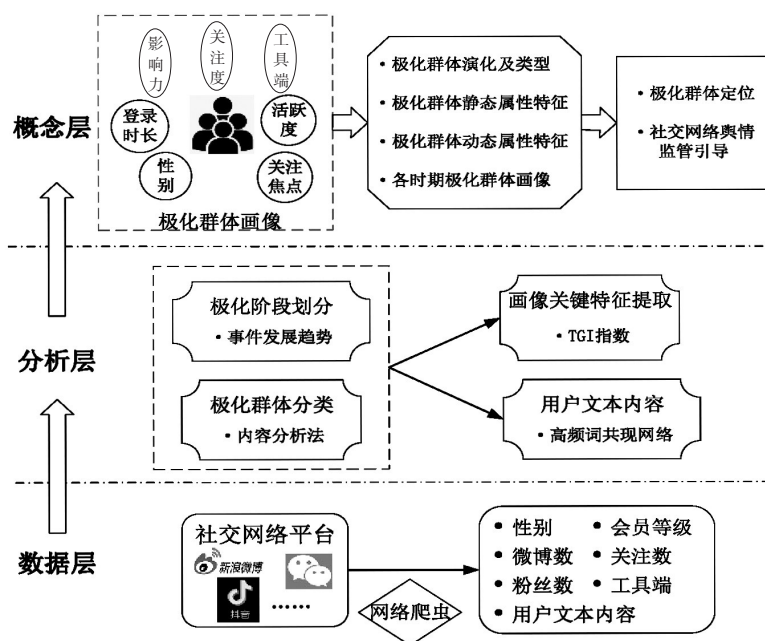


图1 数据驱动的社交网络舆情极化群体画像模型

数、粉丝数和工具端等用户基本信息标签,主要从用户的个人信息界面获取。工具端代表用户参与事件的途径,移动端指通过手机客户端来参与事件讨论,非移动端指通过网页或不同类型的浏览器进行互动。会员等级代表用户的连续登录时长,等级越高连续登录时长越长。微博数可以代表用户的活跃程度。关注数代表用户的信息关注度,粉丝数则在一定程度上代表用户的影响力。其中,性别和工具端属性为二分变量,即1为男性、0为女性;1为移动端、0为非移动端;微博会员等级、微博数、关注数和粉丝数为连续型变量。用户的文本内容是用户在对舆情事件进行转发、评论、发博等过程中产生的文本数据,可以体现用户对舆情事件的关注焦点。

3.3 分析层

本研究中分析层主要包括4个方面,即极化阶段划分、极化群体分类、用户文本内容分析和画像关键特征提取。

在极化阶段的划分方面,在社交网络舆情群体极化的研究中对于群体极化阶段的划分多以舆情发展阶段为基础,再结合具体事件特征进行划分。如在对网民发帖极化现象进行频次分析的基础上,将群体极化阶段划分为潜伏期、成长期、成熟期和消退期4个阶段^[25]。根据舆情讨论热度的变化将群体极化划分为预热期、高潮期、延续期、突变期和衰退期5个阶段^[13]等。基于此,本研究也将根据事件的发展趋势进行极化阶段的划分。

在极化群体的分类上,本研究主要通过内容分析法对极化群体的类型进行划分。内容分析法被证明可以应用于“任何形式的语言交流,以回答谁对谁说了什么、为什么以及产生了什么效果”等问题,非常适合于对社交网络通信文本进行编码分析^[26]。内容分析编码由两名从事社交网络舆情和用户信息行为研究的作者分别进行。

在用户文本内容分析方面,采用高频次共现网络。共现网络是从关键词之间的共现关系角度构建的单词网络,同时考虑到关键词之间的关联和词频。从单词的频率和关系两个角度对文本内容进行分析,将文本表示为单词共现网络。构建单词共现网络的优势在于能够直接和明确地表示文本的内容^[27],因此在对极化群体用户动态属性进行分析的基础上,本研究按照分词、去除停用词、提取高频词、提取行特征等步骤来构建高频词共现网络。对于共现网络的统计特征,本文采取网络分析的方式以关键词的共现关系为基础,对

网络中节点之间的关系、重要性以及网络结构进行直观的展示^[28]。其中PageRank算法是节点重要性评价的一个重要标准,已被证明在社交网络领域具有实用性,节点的PageRank值越高说明节点在网络中越重要^[29]。聚类系数能够反映网络的凝聚程度,值越大则网络的节点联系越紧密^[30]。

在画像关键特征提取方面,采用TGI指数(Target Group Index)。TGI指数是反映目标群体在规定研究范围内的吻合度指数,可以用来识别群体特征中的关键或突出特征,在营销、媒体调查和群体用户关键特征提取等方面已有应用^[31]。TGI指数的计算公式如公式(1)所示,TGI数值越大表明目标群体吻合度越强势。因此,本研究基于内容分析和网络爬虫得到每个极化类型下的用户静态属性数据后,分别计算不同属性分类的TGI值。最后,对各个属性特征的TGI值进行比较,选取TGI值最大的属性特征作为该群体用户的关键属性特征。

$$\text{TGI指数} = \frac{\text{目标群体中具有某一特征的群体所占的比例}}{\text{总体中具有相同特征的群体所占的比例}} \times \text{标准数} 100$$

公式(1)

3.4 概念层

概念层是数据驱动的社交网络舆情极化群体画像的最高级抽象,在这一层将生成具有静态和动态属性特征的极化群体画像。首先,是对不同时间阶段极化群体类型的呈现,能够展示极化群体在事件发展全过程的动态演变。其次,是极化群体画像的属性特征,从静态和动态属性特征两个角度对极化群体的基本信息和关注焦点进行全面、动态的展示。最后,是各时期整体的极化群体画像,通过各时期极化群体画像的生成与分析,找出不同时期各个极化类型群体的典型特征,从而进行用户群体的定位,为利用用户信息、制定舆情干预和引导决策提供指导和借鉴。

4 数据采集及数据结果

4.1 数据来源及处理

新浪微博作为中国影响最广、使用人数最多的社交媒体平台之一,日活跃用户超2亿,具有较高的交互可视性并与中国网民的公共生活深度融合^[32],为广大网民提供自由交流和表达态度的平台,成为反映国内舆论环境的代表性平台之一,因此,本文选择新浪微博作为数据来源平台。《中华人民共和国民法典》作为中国第一部以法典命名的法律,备受公众关注,其第一千零七十七条“离婚冷静期”的设立在微博等社交媒体平

台上引起巨大舆论反响,话题讨论数近15万,阅读数超7亿次^[33]。因此,本文选取#离婚冷静期#话题下的数据进行研究。本文利用新浪微博API,通过网络爬虫软件LocoySpider (V2010SP3),对2020年12月01日至2020年12月15日“离婚冷静期”话题下的数据进行爬取,主要包括该话题下的博文、转发和评论等文本内容,以及用户昵称、ID、发布时间、工具端、会员等级、粉丝数、关注数等字段。共收集到原始数据34942条,经过数据清洗,即删除空白、重复、http超链接等无效数据后共得到有效数据30437条。

4.2 极化群体的演化及类型

根据该舆情事件的发展趋势进行极化阶段的划分。由于每个时间段的话题讨论量存在较大差异,为了优化趋势图效果,使用对数函数对话题讨论量进行了处理,得到的话题舆论趋势如图2所示,可见2020年12月1~4日的话题讨论量虽有明显波动,但维持在较高水平,5日开始话题讨论量明显下降;8日以后话题讨论量虽有波动但基本在维持在相同水平。因此本文将极化群体阶段分为3个,即极化意见震荡期(2020/12/01-04)、成熟期(2020/12/05-07)、平静期(2020/12/08-15)。

在极化意见动态演变的3个时期中,极化群体类型呈现由多极化向两极化转变的趋势。在极化意见震荡期,有超过半数的极化群体持中立意见(61.6%),反对群体(36.%)次之,赞成群体(1.4%)最少。过渡到极化意见成熟期时,持中立意见(54.4%)与反对意见(44.6%)的极化群体基本各占半数,而持赞成意见(0.8%)的极化群体占比已下降到不足1%的水平。在极化意见平静期,中立意见群体(70.5%)再次占据主导地位,但赞成意见群体已经完全消失。其次,在极化意见动态变化的3个时期,中立意见群体始终在极化群体中占据主导地位;反对意见的极化群体呈现先增长后下降的趋势;持赞成意见的极化群体占比不断下降直至消失。

4.3 极化群体画像的静态特征

用户的静态属性特征主要通过性别、会员等级、微博数、关注数、粉丝数和工具端等自然数据获取。由于微博数、关注数和粉丝数与其他用户属性特征在量纲上存在较大差异,并且数据较为分散,所以本文根据顶

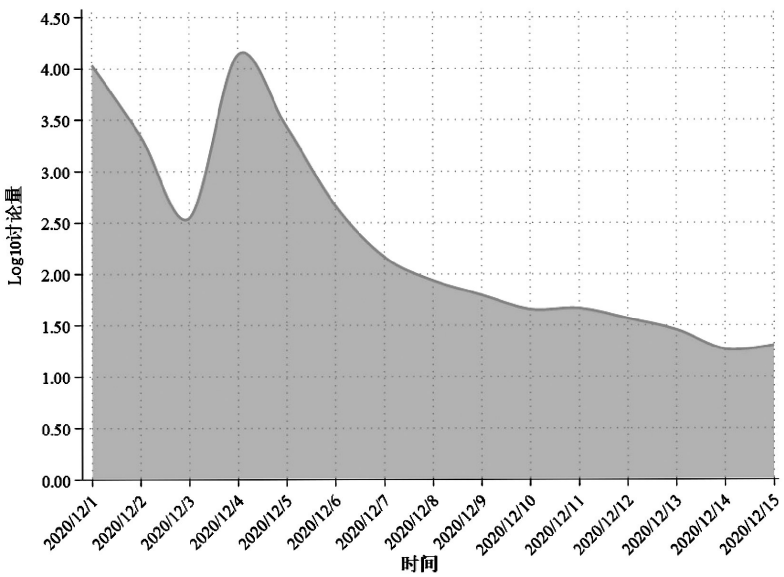


图2 “离婚冷静期”舆论趋势图

部四分位数、中间两个四分位数和较低四分位数对3种属性特征进行划分^[34],并将级别定义为高、中、低。其中,顶部四分位数对应级别为高,中间两个四分位数对应级别为中,较低四分位数对应级别为低。以极化意见震荡期用户的3种属性为例,表1显示了3个级别的值和范围。

表1 用户属性四分位划分结果(极化意见震荡期)			
用户属性	低	中	高
微博数	[0, 613]	[613, 6970]	(6970, 149776]
关注数	[0, 165]	[165, 585]	(585, 20000]
粉丝数	[0, 45]	[45, 317]	(317, 613153]

在计算TGI指数前,对静态特征数据进行了卡方检验,结果如表2所示,P值均小于0.05,说明静态特征间存在的差异有统计学意义。TGI指数能够反映目标群体在规定范围内的吻合度,其中TGI指数等于100表示平均水平,在此基础上的TGI指数越高则目标群体吻合度越强。下页表3是关键静态属性特征及其TGI指数,可见在极化的不同时期,持反对意见的极化群体多通过移动端发表意见,并且以女性为代表,会员等级中等偏上。从微博数和粉丝数来看,这类群体的活跃

表2 卡方检验结果

静态特征	卡方值	P
工具端	1712.813	.000
性别	955.315	.000
会员等级	3979.830	.000
微博数	381.130	.000
关注数	382.630	.000
粉丝数	385.645	.000

表3 各阶段极化群体画像关键静态属性特征及其TGI指数

极化阶段		工具端	性别	会员等级	微博数	关注数	粉丝数
1~4 日极化意见震荡期	反对	移动端 100.49	女 100.61	4级 129.62	低[0, 612.5] 106.14	低[0, 165] 107.52	低[0, 45] 104.02
	中立	移动端 100.59	男 101.72	3级 105.57	高[6970, 149776] 104.42	高[585, 20000] 104.42	高[317, 613153] 101.90
	赞成	非移动端 428.69	女 103.53	5级 123.22	低[0, 612.5] 194.80	高[585, 20000] 123.22	低[0, 45] 143.97
5~7 日极化意见成熟期	反对	移动端 112.55	女 105.66	7级 224.16	中[498, 7382] 110.07	高[591, 19943] 105.33	低[0, 46] 112.08
	中立	非移动端 123.63	男 108.05	5级 137.64	高[7382, 87981] 106.13	低[0, 131] 102.68	高[395, 3342506] 110.55
	赞成	非移动端 234.39	男 208.10	6级 253.03	低[1, 498] 150.45	中[131, 591] 133.33	高[395, 3342506] 134.14
8~15 日极化意见平静期	反对	非移动端 130.77	女 125.93	5级 340.00	低[2, 219] 188.89	中[165, 585] 140.00	低[0, 45] 151.11
	中立	移动端 107.94	男 141.67	4级 141.67	中[219, 3128] 116.67	高[585, 20000] 123.96	高[317, 613153] 123.96

度和影响力处在较低的水平;关注数所反映的信息关注度水平存在明显差异。持中立意见的极化群体以男性为主要代表,会员等级中等偏下,并且从微博数、关注数和粉丝数来看,这类群体不仅具有较高的活跃度和影响力,而且对信息的关注度水平也相对较高。持赞成意见的用户群体只出现在极化的前两个阶段,虽然会员等级相对较高,但是从微博数来看活跃度较低,信息关注度和影响力存在显著差异。

4.4 极化群体画像的动态特征

用户的动态属性特征主要是用户发布的文本内容,它们会随着事件或话题的发展而发生变化,“离婚冷静期”各阶段的动态属性特征如图3、图4和图5所示。图中的节点代表不同的高频词,节点间的边代表高频词之间的共现关系,边越粗说明共现次数越多。

结合表4和图3可见,在极化群体意见震荡期,中立和赞成意见极化群体的词共现网络结构更加聚集,关注和讨论的话题较为集中。持反对意见的极化群体主要观点是强调“婚姻自由”,认为“离婚冷静期”与先前法律相比,在某种意义上是一种“历史倒退”。持中立意见的群体则主要对离婚冷静期法条本身和涉及的其他事物进行解释和解读,如“孩子”“财产”和“家暴”,以及离婚申请的流程如“预约”“登记”和“申请”等,内容较为集中。至于持赞成意见的群体,内容相对其他两类群体来说较为分散,存在多组没有关系的词对。关注的内容主要是“离婚冷静期”的实施能够带来的预期后果,如“降低离婚”;虽然“增加法院工作量”,但给“律师”行业带来需求等。

在极化意见成熟期,由于持赞成意见的极化群体比例较少,生成的文本信息不足以生成高频关键词共现网络,所以仅分析了另外两种群体的动态属性特征。结合表5和图4可见,中立意见极化群体的词共现网络结构更为紧密,讨论的内容较为集中。在持反

表4 震荡期高频词共现网络指标

指标	反对		中立		赞成	
	关键词	PageRank 值	关键词	PageRank 值	关键词	PageRank 值
关键词重要性度量(前10位)	离婚	0.198	离婚	0.235	增加	0.058
	结婚	0.067	结婚	0.11	清楚	0.055
	一方	0.058	孩子	0.05	可怜	0.047
	成年人	0.057	登记	0.041	考试	0.044
	自由	0.057	预约	0.041	冷静期	0.043
	反思	0.049	申请	0.041	离婚	0.043
	倒退	0.044	一方	0.032	法院	0.041
	申请	0.041	同意	0.032	逃脱	0.036
	婚姻	0.04	上海	0.032	婚姻	0.036
	平均聚类系数	0.508	平均聚类系数	0.722	平均聚类系数	0.727

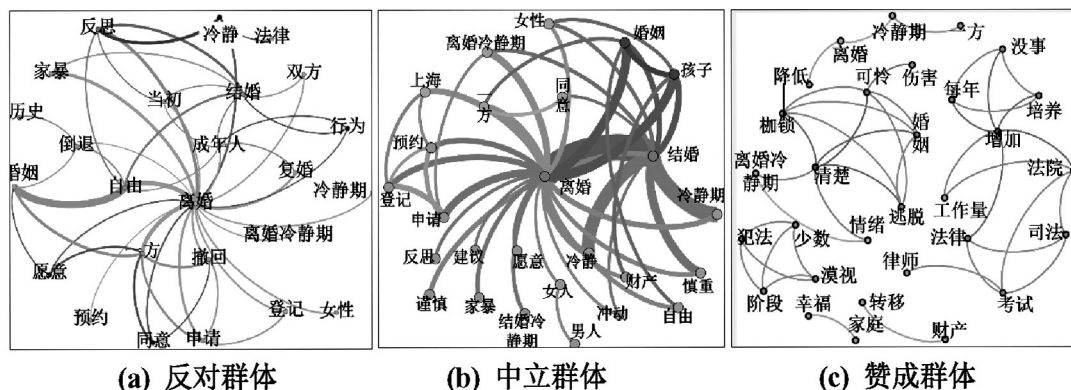


图3 震荡期各群体高频词共现网络

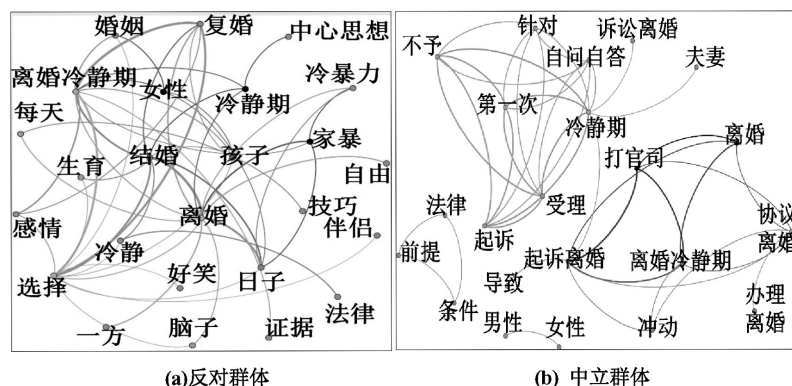


图4 成熟期各群体高频词共现网络

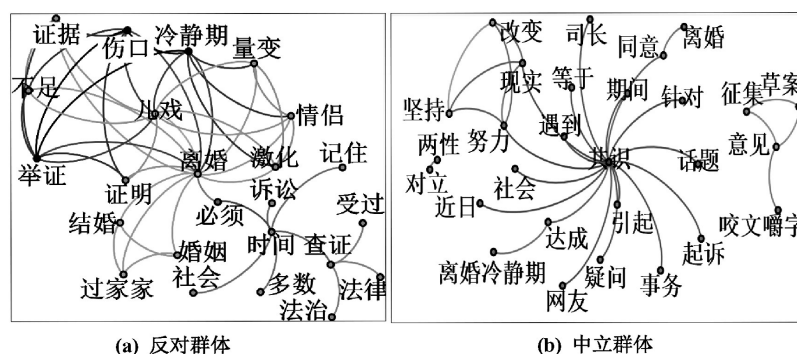


图5 平静期各群体高频词共现网络

表5 成熟期高频词共现网络指标

指标	反对		中立	
	关键词	PageRank 值	关键词	PageRank 值
关键词重要性度量(前10位)	选择	0.121	冷静期	0.084
	离婚	0.116	起诉离婚	0.076
	离婚冷静期	0.098	协议离婚	0.068
	结婚	0.08	离婚冷静期	0.055
	日子	0.054	针对	0.05
	复婚	0.053	第一次	0.05
	孩子	0.044	自问自答	0.05
	冷静期	0.042	受理	0.05
	家暴	0.038	男性	0.045
网络结构度量	平均聚类系数	0.559	平均聚类系数	0.883

对意见的极化群体中,一方面是对离婚冷静期实施后预期问题的关注,如“暴力”“冷暴力”“生育”和“孩子”等问题。另一方面是直接表达对于法条的不满,如“好笑”和“脑子”等。在持中立意见的群体中,网络结构明显存在几个不互联的分块,如图4(b)中的左上和右下。但总的来看,该类型群体还是主要关注对法条的解读,如离婚冷静期与“诉讼离婚”“协议离婚”的关系、办理离婚的方式,如“起诉离婚”“受理”“打官司”等。

结合表6和图5可见,在极化群体意见平静期,反

对和中立意见极化群体的词共现网络结构差异变小,两类群体主要关注的内容与上一成熟期相比趋于稳定。在持反对意见的群体中,主要关注的还是对法条实施后预期问题,但比成熟期内容有所扩展,如“举证不足”“诉讼时间”等。还有对法条直接表达的不满,如“离婚冷静期”的实施不是“儿戏”和“过家家”等,持中立意见的极化群体主要还是关注对法条的解读。

4.5 各时期极化群体画像

下页图6为各时期极化群体画像。在极化意见震荡期反对群体以女性为代表,主要通过移动端进行信息交互,具有较低水平的影响力、活跃度和信息关注度,会员等级居中,这些极化群体主要关注婚姻自由和离婚法律的前后对比。这一时期的中立群体以男性为代表,主要通过移动端进行信息交流,具有较高水平的影响力、活跃度和信息关注度,会员等级相对较低,这一极化群体主要关注法条的解读。赞成群体以女性为代表,主要通过非移动端进行信息交互,具有较高水平的信息关注度,但影响力和活跃度较低,会员等级则相对较高,这一极化群体主要关注法条实施带来的积极预期。

在极化意见成熟期反对群体以女性为代表,主要通过移动端进行信息交流,具有较低的活跃度,但信息关注度和活跃度分别处于较高和中等水平,会员等级相对较高,这一极化群体主要关注法条实施的消极预

表6 平静期高频词共现网络指标

指标	反对		中立	
关键词重要性度量(前10位)	关键词	PageRank 值	关键词	PageRank 值
	时间	0.11	共识	0.26
	离婚	0.082	意识	0.052
	查证	0.068	努力	0.05
	儿戏	0.049	现实	0.05
	伤口	0.049	改变	0.037
	举证	0.049	坚持	0.037
	冷静期	0.037	同意	0.036
	激化	0.037	达成	0.036
	量变	0.037	两性	0.036
网络结构度量	平均聚类系数	0.712	平均聚类系数	0.567

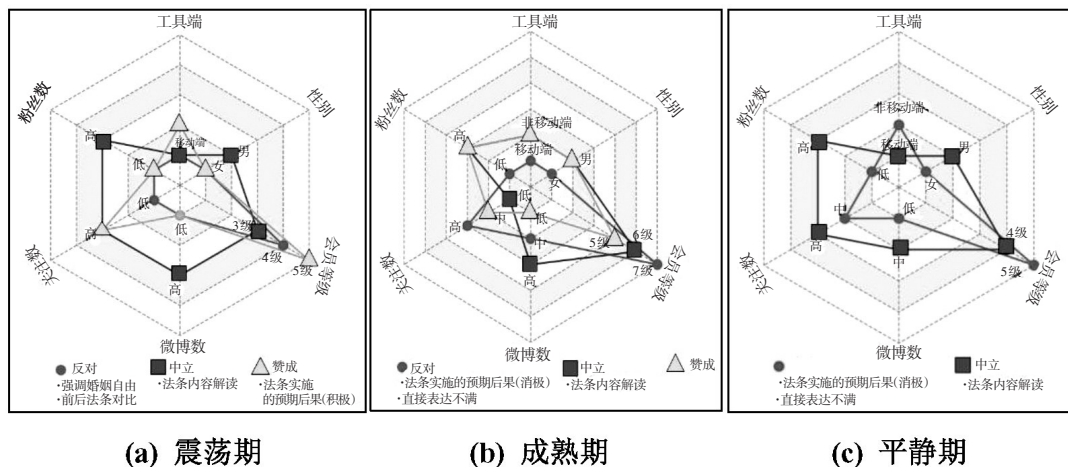


图6 各时期极化群体画像

期和直接表达不满。中立群体以男性为代表,主要通过非移动端参与这一事件,具有较高水平的影响力和活跃度,但信息关注度较低,会员等级相对较高,这一极化群体主要关注法条的解读。赞成意见的群体以男性为代表,主要通过非移动端进行信息交互,具有较高水平的影响力和中等水平的信息关注度,但活跃度较低,会员等级较高。

在极化意见平静期反对群体以女性为代表,主要通过非移动端进行信息交互,影响力水平和活跃度较低,但信息关注度水平居中,会员等级相对较高,这一极化群体主要关注法条实施的消极预期和直接表达不满。

5 讨论分析

5.1 极化群体类型的动态演化

在“离婚冷静期”舆情事件发展的整个过程中,极化群体存在着多极化向两极化动态演化的过程。在社会交往中个体往往会作为一类人的代表与其他个体交往,他们会努力去获得或维持社会认同,但当个体不满足于当前的社会认同时就会选择脱离该类型的群体^[6]。所以当任何热点事件发生后,大多数个体想要寻求社会认同而不想被孤立,就会选择在一个群体中寻求归属。因此,他们会根据自己最初的感觉和态度选择一个立场,这样就形成了初期的极化群体类型。但随着事件的不断发展,信息扩散程度不断增加以及信息完整度不断完善,个体在信息交互过程中与其他群体有了新的互动,对热点事件的理解和态度就产生了新的变化,不同类型的极化群体便会在极化过程中获得或失去一部分支持者,这时不同时期的极化群体类型在

意见交互中就产生了动态变化。

在社交网络的沟通交流中,多数影响可以使个体关注该群体并引导其向该群体的集体思维转变^[17]。在热点事件发生的初期,一部分人由于无法猜测事件背后的原因或对事件全貌缺乏整体的认知,总是会持观望态度并对事件持保留意见,这时的个体并不会贸然表达自己的立场,但是又不想让自己处于被孤立的状态^[35],因而会通过加入对热点事件本身的解读来获得社会认同和对事件的全面认知。随着不同极化群体之间的辩论和群体内部一致性的不断提高,持中立意见的群体数量不断扩大,对其他个体的影响和吸引也会不断增强。这也是“离婚冷静期”事件在3个极化时期,持中立意见的群体始终占据主导位置,而赞成意见的群体占比不断减少直至消失的原因。

5.2 极化群体的性别特征

在极化群体中持反对意见的极化群体均以女性为代表,而持中立意见的极化群体则以男性为代表,这在一定程度上表现出男女对立的局面。这与“离婚冷静期”事件性质有关,结合不同极化群体的动态属性特征来看,显然持反对意见的群体更加关注法律实施带来的预期消极后果,而中立意见的群体则更加关注对法律的解释与解读。由于传统社会观念上认为女性在社会地位和经济地位上的弱势,再加上法律条文的专业性较强,公众对法条会存在一定程度的理解偏差。因而大多数持反对立场的群体认为,在“冷静期”产生的家庭暴力问题、财产问题等对于女性都是极为不利的,并且对女性的不利影响要大于男性。另外,这种性别对立的观点以及直接对“离婚冷静期”表达不满的观

点,在极化群体中进一步发酵和强化也将会是引发次生舆情的重要潜在风险。

在此类事件的极化群体引导与调控上,可以从极化群体的信源和信道二元化疏导上展开。一方面,从信息源头入手。在获得政府有关部门许可的前提下,社交网络服务提供商可以采用关键词过滤、情感识别等技术,对能够引发次生舆情的敏感词条、语句和带有强烈负面情绪煽动色彩的用户进行识别和屏蔽,从源头上对舆论信息进行正面的过滤和管控。另一方面,从信息传播过程入手。可以通过扩大能够带来客观、积极影响的极化群体信息扩散渠道和范围,助力其成为极化过程中的多数影响群体。可以促进中立意见群体对法条解读信息的交互和扩散,扩大客观、正确、积极信息的传播和影响范围,使公众对法条实施过程、可能带来的预期结果等有更加完整和客观的认知,进而将这类具有负面影响的极化群体逐步转化成同质的公众性意见,以实现积极的社会认同,达到扭转极化负面影响的效果,同时也能够避免次生舆情的产生。

5.3 极化群体的复合性特征

研究发现,“离婚冷静期”舆情事件中各时期的极化群体画像呈现出复合性特征,即相同类型的极化群体画像并不是完全一致,他们既存在相同规律又存在明显差异。“离婚冷静期”舆情事件中各时期的中立群体均具有较高的影响力和活跃度,但信息关注度却分布参差不齐。相比之下,反对群体虽然信息关注度水平也参差不齐,但影响力和活跃度则均处在较低水平。因此,对极化群体的引导和调控可以采取复合性策略,根据需要扬长避短,并借助不同群体的优势来消解极化的负面影响。

对于粉丝数、微博数和会员等级中等偏高的中立群体来说,不仅立场稳定不存在引发次生舆情的风险,而且具有良好的粉丝基础和活跃度,具备强大的舆论把控潜力。因此,该类群体应该被视作舆情传播引导和极化干预的关键用户群体,社交网络服务提供商要维护好该类核心用户群体,给予足够的发展空间,对有积极影响的内容进行推广,从而让其有助于树立和传播良好和正确的舆情观念。对于信息关注度较低的群体来说,其信息来源较为单一,信息获得较为被动,可以有针对性地增加对这类极化群体的信息推送频率,全面、持续地提供个性化信息,增强他们对事件的客观和全面认知。同时,还能够增强用户正面体验和提高用户活跃度。除此之外,可以利用高影响力和高活跃度群体的领导优势,通过使用移动工具来提高信息曝

光的时效性,抢占信息曝光先机,从而提高正面和积极信息内容的曝光频率。在发挥和增强多数影响作用的同时,挤占负面和不利信息的曝光渠道,避免次生舆情的发生。

6 结语

在理论层面,本研究在确定群体极化阶段和极化类型的基础上,从数据层、分析层和概念层三个层面构建数据驱动的社交网络舆情极化群体用户画像模型,为社交网络极化群体用户画像的研究提供了新的理论研究框架。同时,分析了不同舆情阶段群体极化类型的演化,为社交网络舆情的极化群体用户画像的研究和舆情引导提供了新的理论依据。

在实践层面,选取典型舆情事件“离婚冷静期”作为实证分析对象,所得到的数据结果和研究结论对了解新法条的民意基础和民众关切内容具有现实意义。其次,本研究对于极化群体属性特征的分析结果能够为社交网络服务提供商和政府有关部门提供具体的目标用户群体定位,有利于网络舆情的管控和次生舆情的预防。

在研究过程中,由于社交网络中群体极化的复杂性本研究也存在一定的局限。本文对群体极化的分类和画像构建是基于特定事件的,在接下来的研究中我们将考虑扩大事件范围,将本文的研究方法与理论应用到其他类型的事件中,如突发公共卫生事件和网络谣言等。在未来的研究中,可以考虑采用人工智能或机器学习的方法,扩大研究样本的数量,从而加强研究结果的代表性和普适性。

参考文献

- [1] 中国互联网信息中心.第47次中国互联网络发展状况统计报告[EB/OL]. [2021-02-03].http://www.cac.gov.cn/2021-02/03/c_1613923423079314.htm.
- [2] 人民网舆情数据中心.2020年互联网舆情形式分析与展望[EB/OL]. [2020-09-17].<http://download.people.com.cn/yuqing/eleven16003359621.pdf>.
- [3] 高俊峰,黄 微.网络舆情信息受众情感极化的生发机理及干预措施分析[J]. 情报理论与实践, 2019, 42(5): 122-127.
- [4] KimYonghwan, KimYoungju. Incivility on Facebook and political polarization: the mediating role of seeking further comments and negative emotion[J]. Computers in Human Behavior, 2019, 99(OCT.):219-227.
- [5] Chen T, Li Q, Fu P, et al. Public opinion polarization by individual revenue from the social preference theory[J]. International Journal of Environmental Research and Public Health,

- 2020, 17(3): 946.
- [6] 林燕霞,谢湘生. 基于社会认同理论的微博群体用户画像[J]. 情报理论与实践, 2018, 41(3): 142-148.
- [7] 刘静, 安璐. 突发公共卫生事件中社交媒体用户应急信息搜寻行为画像研究[J]. 情报理论与实践, 2020, 43(11): 8-15.
- [8] 崔鹏, 张巍, 何毅, 等. 突发公共事件网络舆情演化及政府应对能力研究[J]. 现代情报, 2018, 38(2): 75-83.
- [9] Du L, Feng Y, Tang L Y, et al. Time dynamics of emergency response network for hazardous chemical accidents: a case study in China[J]. Journal of Cleaner Production, 2020, 248: 119239.
- [10] 马溯川, 张法明. 新媒体时期涉军网络舆情态势趋向及治理路径——以“老虎团涉嫌猥亵幼童”事件为例[J]. 情报杂志, 2019, 38(9): 99-104, 133.
- [11] 胡峰. 重大疫情网络舆情演变机理及跨界治理研究——基于“四点四阶段”演化模型[J]. 情报理论与实践, 2020, 43(6): 23-29+55.
- [12] Cooper A, Reimann R, Cronin D, et al. About Face: The Essentials of Interaction Design[M]. John Wiley & Sons, 2014.
- [13] 辛文娟, 赖涵. 群体极化视域下网络舆情的演化机制研究——以微博网民讨论“浙江温岭杀医案”为例[J]. 情报杂志, 2015, 34(2): 47-52.
- [14] Stoner J A F. A comparison of individual and group decisions involving risk[D]. Massachusetts Institute of Technology, 1961.
- [15] 卢智增, 高翔. 社交媒体平台群体极化的形成机理与引导策略研究[J/OL]. 情报理论与实践: 1-12 [2021-03-10]. <http://kns.cnki.net/kcms/detail/11.1762.g3.20210225.1827.004.html>.
- [16] 汪明艳, 余丽彬, 朱译冰. 舆论反转中群体极化效应的影响因素研究[J]. 情报杂志, 2018, 37(9): 106-112, 119.
- [17] Yuan Jiangjun, Shi Jiawen, Che Juyan, et al. Modeling and simulation analysis of public opinion polarization in a dynamic network environment[J]. Concurrency and Computation: Practice and Experience, 2020, 32(19): e5771.
- [18] Lee C, Shin J, Hong A. Does social media use really make people politically polarized? Direct and indirect effects of social media use on political polarization in South Korea[J]. Telecommunications and Informatics, 2018, 35(1): 245-254.
- [19] Jansen J, Jung S G, Salminen J O. Data-driven personas for enhanced user understanding: combining empathy with rationality for better insights to analytics[J]. Data and Information Management, 2020, 4(1): 1-17.
- [20] 叶光辉, 曾杰妍, 胡婧岚, 等. 城市画像视角下的社会公众情感演化研究[J]. 数据分析与知识发现, 2020, 4(4): 15-26.
- [21] 任中杰, 张鹏, 兰月新, 等. 面向突发事件的网络用户画像情感分析——以天津“8·12”事故为例[J]. 情报杂志, 2019, 38(11): 126-133.
- [22] Wilson A E, Parker V, Feinberg M. Polarization in the contemporary political and media landscape[J]. Current Opinion in Behavioral Sciences, 2020, 34: 223-228.
- [23] Salehi A, Ozer M, Davulcu H. Sentiment-driven community profiling and detection on social media[C]. Proceedings of the 29th ACM Conference on Hypertext and Social Media. New York: ACM, 2018: 229-237.
- [24] 刘蕾蕾, 王胜涛, 胡正银. 社区画像研究综述[J]. 图书情报工作, 2019, 63(23): 122-130.
- [25] 杨志强, 孙冉冉. 网络政治参与视域下青年网络群体极化形成机制研究[J]. 青年探索, 2014(6): 24-28.
- [26] Ngai C, Singh R G, Lu W, et al. Grappling with the COVID-19 health crisis: content analysis of communication strategies and their effects on public engagement on social media[J]. Journal of Medical Internet Research, 2020, 22(8): e21360.
- [27] Celardo L, Everett M G. Network text analysis: a two-way classification approach[J]. International Journal of Information Management, 2020, 51: 102009.
- [28] Camacho D, Panizo-Lledot Á, Bello-Ortiz G, et al. The four dimensions of social network analysis: an overview of research methods, applications, and software tools[J]. Information Fusion, 2020, 63: 88-120.
- [29] Cla B, Jb A, Lei Z B, et al. Opinion community detection and opinion leader detection based on text information and network topology in cloud environment - ScienceDirect[J]. Information Sciences, 2019, 504(Dec): 61-83.
- [30] 马超. 健康议题辟谣社群的类别构成与社群结构研究——基于多主体谣言协同治理的视角[J]. 情报杂志, 2019, 38(1): 96-105.
- [31] Albedah A M, Khalil M K, Khalil A A, et al. Use of the target group index survey to evaluate the cigarette smoking profile in Saudi Arabia[J]. Saudi Medical Journal, 2011, 32(10): 1055-1059.
- [32] 新浪微博. 2020年第三季度财务报告[EB/OL]. [2020-12-28]. <https://finance.sina.cn/usstock/mggd/2020-12-28/detail-iiznezxs9422515.d.html>.
- [33] 微热点. “离婚冷静期”能缓解爆发的离婚潮吗? [EB/OL]. [2020-12-29]. https://www.sohu.com/a/396002263_499808.
- [34] Fan C, Jiang Y, Yang Y, et al. Crowd or hubs: information diffusion patterns in online social networks in disasters[J]. International Journal of Disaster Risk Reduction, 2020, 46: 101498.
- [35] Wang Q, Yang X, Xi W. Effects of group arguments on rumor belief and transmission in online communities: an information cascade and group polarization perspective[J]. Information & Management, 2018, 55(4): 441-449.
- [作者简介] 王晰巍, 女, 1975年生, 教授, 博士生导师, 吉林大学大数据管理研究中心主任, 吉林大学国家发展与安全研究院网络空间治理研究中心主任。
- 贾若男, 女, 1995年生, 吉林大学管理学院博士研究生(通讯作者)。
- 孙玉姣, 女, 2000年生, 吉林大学管理学院本科生。
- 收稿日期: 2021-06-07