

意见领袖的跨圈层传播与“回音室效应”^{*}

——基于深度学习文本分类及社会网络分析的方法

■ 汤景泰 陈秋怡

【内容摘要】“回音室效应”使得社交媒体的信息传播质量受到质疑,但其成因与作用机制并未形成共识。圈层的同质化程度可能与成员活跃度有关,成员数多并且互动活跃的子群,子群的用户身份类型及其参与的内容主题呈现多样性的趋势,并且存在着跨子群的互动关系,活跃子群并没有形成明显的回音室效应。另一方面,意见领袖参与内容主题数增多,其跨圈层的传播影响力也具有增大的趋势,跨圈层意见领袖一定程度上有助于破除“回音室效应”。

【关键词】意见领袖;回音室效应;跨圈层传播;文本分类;社会网络分析

一、引言

随着数字媒体的涌现和信源的剧增,注意力成为稀缺的资源,很少有议题、人物或群体能够长期吸引公众的注意力。在注意力稀缺的形势下,媒体对用户注意力竞争而形成的“注意力经济”,使得吸引用户成为实现经济、社会或政治目标的重要条件。^①而选择性接触理论认为,用户对自己的信息消费做决策时,更喜欢符合自身意识形态或强化自身观点的内容,回避挑战自身意见的内容。^②这样在注意力竞争以及认同驱动选择的框架下,媒体与内容产品的分众逻辑与趋势逻辑愈加显著。^③

但随着英国脱欧、剑桥数据分析丑闻等事件的发生,这种逻辑的弊端也引发了越来越多的讨论。注意力经济驱使下的媒体竞争、基于内容或用户的推荐算法,被认为强化了选择性接触(selective exposure)^④、加速了受众碎片化(audience fragmentation),形成相对封闭的“回音室”或“过滤气泡”,使得公共讨论的质量与多样性下降,甚至导致意见的极化(polarization)以及虚假信息的泛滥。^⑤另外有研究发现,社交媒体在使人快速地寻求具有相似观点的用户与社区的同时,也可能导致公共领域的分裂,形成人口与行为特征同质化的联系网络、缺少互动关系的圈层^⑥。这种假设中人们进入了只由有相似观点的人以及同质化的信息所构成的圈层的概念,即“过滤气泡”(filter bubble)^⑦或“回音室”(echo chamber)^⑧。目前,这一现象已经引起了较大的社会忧虑。在国内,近年来异军突起的新闻类

App 今日头条因其推荐算法也饱受非议。因此,在这一背景下,讨论回音室效应的形成及其破除机制,具有较高的理论意义和现实价值。

二、文献综述

回音室效应是国内外众多学者关注的热点话题,过滤气泡或者回音室效应的存在及其形成的原因存在着许多争议。例如,一项随机对照实验的研究发现,内组群的信息流入一定程度上抑制了向不同社会群体传播的信息,促进了对信息交换的正面反应,却减少了对交换信息的负面评价,呈现了信息回音形成的过程。^⑨对荷兰 Twitter(推特)平台的研究则发现,用户有意识地选择与自己社区主题与情感对应的信息,共同构筑了“回音室”的信息循环。^⑩而另一项针对英国脱欧的脸书帖子的大规模分析发现,选择性接触的信息消费模式导致了两个界限分明的新闻媒体社区的产生,两者的话题与情感存在显著差异,具有明显的回音室效应特点,呈现两极分化格局。^⑪在特定的问题上寻求同质的群体可能表明对于一个特定问题的自我涉入程度(ego-involvement),并且参与这个群体的时长可能会增加自我涉入程度,从而将圈层中的参与者推向极端观点。^⑫

但另外一些研究则发现,不应夸大过滤气泡或者回音室效应。一项对内容个性化算法效果的研究认为,现有经验证据无法支撑对过滤气泡效应的过分担忧。^⑬另一项研究则发现,顺应推荐系统结果的用户相比不使用推荐系统的用户反而获得了更多样化的信息。^⑭若整体

^{*} 本文系国家自然科学基金重大项目“互联网群体传播的特点、机制与理论研究”(项目编号:15ZDB142)的研究成果。

网络呈现小世界特点,那么连接的多样性可能促进特定个体网络内信息共享的多样性。特别是弱连接能够扩展既定的熟人网络,为社交媒体环境带来更多样的信息。^⑮有研究表明,在社交媒体上花费的时间似乎增加了网络的异质性,积极地使用社交媒体会带来更多样化的信息体验;^⑯而社交性的指标(组成员与朋友的数量)是预测过滤气泡效应的重要因素。^⑰因此,网络的异质性可能与个体使用社交媒体的活跃度有关。

可见,回音室效应的形成有一定条件。另外相关研究也发现,除了社交媒体使用活跃度,不同媒体平台以及传播结构之间的回音室效应存在差异。弱连接能够拓展信息的多样性,但如果网络结构呈现具有较强信息控制力的星形网络或者具有较长社会距离的线形网络,则可能具有较为显著的回音室效应。^⑱因此,有研究认为单一的媒体研究和使用的定义和测量的研究是有缺陷的,需要在一个多媒体环境的现实背景下测试“过滤气泡”或“回音室”的理论,而媒体“食谱”多样的用户会主动回避“回音室”。^⑲但也有研究表明,弱关系能够增强网络结构的多样性,网络信息流的多样性能够减少“过滤气泡”的影响。^⑳可见,网络结构是影响信息“回音室效应”的一个重要因素,信息传播的网络结构与信息内容异质性的关系是一个值得深入探索的领域。

不仅如此,尽管社交媒体由弱连接组成的网络可能由异质成员构成,能够提供立场不同、多元的信息,但也不意味着这些信息能被理解与接受。涉及心理学角度的研究表明,人们可能会看到异质内容,但不一定会点击对立观点的内容。^㉑在某一特定问题上有强烈立场的高度自我参与性的个体,通常会认为媒体内容是有偏见的,这种效应被称为敌意媒体现象(hostile media phenomenon)。^㉒同时,受众容易高估自己观点的公共支持度,这种认知偏差被称为虚假共识效应(FCE)。一项研究发现,网络的同质性较强,则虚假共识效应显著较高。^㉓

综上所述,关于网络平台上的同质化圈层(“过滤气泡”与“回音室”)以及信息传播多样性的问题存在种种争论,但一定程度上表明了社交媒体中互动的同质化与多样化之间存在动态而复杂的关系,在不同的条件下社交网络的结构与信息异质度可能是不断变化的。分散的社交媒体用户在兴趣、身份认同、政治立场等因素的驱使下形成多个交流相对紧密的圈层,如果圈层的排斥性与防御性较强、他者信息流通不畅则可能形成“回音室”,导致意见极化^㉔;但如果圈层参与者具有较高的信息搜集的主动性与积极性,则可能突破圈层的限制,使得其信息流通具有多样化、跨圈层传

播的特点。因此,传播圈层中的意见领袖与是否会形成回音室效应关系密切。

这就涉及到意见领袖的识别及其作用机制的研究。目前识别社交媒体意见领袖的方法主要有两种思路:一是构建指标体系进行用户影响力的排序,借助层次分析法或者熵权法赋予指标权重,研究的差异只是在于指标体系考虑的因素不同,如采用社交媒体用户自身的特征(是否认证、发博活跃度等)以及内容本身的传播影响力(转发、点赞、评论等)^㉕;或者是加入了社会网络分析的指标(如中介中心度、出入度、page-rank等)^㉖。二是对用户进行聚类,首先挖掘的是群体,然后进一步对群体内的成员影响力进行评估。为了提升识别的准确性和全面性,有的研究还考虑了更多维的特征数据^㉗,加入了文本挖掘的方法进行主题聚类与情感分析^㉘,进而构建内容质量层面的指标,甚至结合时序数据进行超网研究^㉙。

基于已有的研究可以看出,引入社会网络分析以及文本挖掘的方法对识别意见领袖具有较好的应用价值,能够给对社交媒体用户行为的研究带来更深入的认识。但是在机器人“水军”虚假数据涌现的背景下,缺乏对网络结构以及内容属性等因素考量的指标设计,有可能忽略了用户实际的传播的质量,高估了用户和内容同质化圈层中“意见领袖”的传播力。因此,本研究在前人对社交媒体回音室效应及过滤气泡效应问题研究的基础上,通过深度学习文本分类以及社会网络分析的方法,对微博平台上意见领袖的圈层化以及跨圈层传播的问题进行研究,探讨意见领袖对回音室效应的影响。

本研究假设在分散的用户圈层中,存在联系不同圈层且具有较大影响力的用户主体,即跨圈层的意见领袖——他们是沟通不同圈层的关键节点,并使得多样化的信息得以传播。本研究以新浪微博平台数据为例,对以下问题进行分析:

Q1:用户的身份类型与其参与的内容主题倾向是什么关系?

Q2:不同的用户圈层参与的内容主题呈现什么样的分布?

Q3:用户发布的内容异质度与其跨圈层的传播力之间是什么关系?

Q4:跨圈层的意见领袖对于不同圈层传播的内容异质度有何影响?

三、研究设计

(一)研究对象

本研究以新浪微博平台的头部用户及其被转发的

微博为研究对象,时间范围是2019年1月1日到6月30日。首先,通过大规模微博数据抓取、按用户被转发数总和聚合计算,得到原发博用户共8577962位,作为微博用户样本U。^③用户样本U的被转发总和呈现幂律分布的特点,并且前0.05%的用户量产生了80%的被转发量,而30%左右的用户转发数不超过1(如图1所示),可见微博用户的转发量集中度非常高。

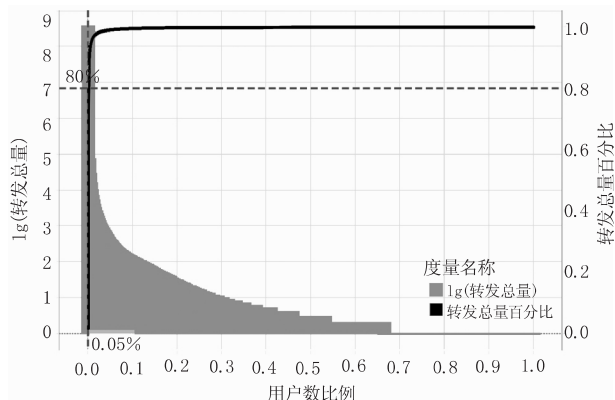


图1 用户被转发数总和分布帕累托图

意见领袖是在二级传播中担任关键节点的传播主体,很大程度影响了信息的扩散。因此本次研究关注基于微博的转发关系计算的头部用户,通过微博转发数观察用户的信息扩散能力,并且取前0.05%的头部用户($n=4288$)作为本次研究的用户样本U1。进一步地选取用户U1样本中转发数大于1万的微博作为分析的微博样本W1($n=11285$),并对其进行文本分类。

(二) 研究方法

1. BERT 多类别文本分类

2018年,谷歌AI团队发布的BERT模型在机器阅读理解水平测试SQuAD1.1中表现出色,被认为是自然语言处理的集大成者。BERT模型沿袭了利用深度神经网络预训练(pre-training)模型的方法思路,即基于海量的语料数据,依据词汇上下文的关系将每一个词汇标注为数值向量,以达到准确表征自然语言词汇语义的目的。^④这种非监督的机器学习方法,能够为训练数据标注节省大量的人工成本,也为下游任务节省了从零开始训练模型的成本。而BERT的新进展在于采取深度双向(deep bidirectional)的机制对词汇进行表征,即代替传统的仅用左侧(或右侧)的词汇预测目标词汇的单向方法或使用单独的左侧(或右侧)上下文预测目标词汇的浅层双向方法,使用全向的上下文预测被遮罩词汇,生成理想的通用语言表征模型。^⑤

BERT另一重要的机制是微调(fine tuning),即根据

特定任务的数据对模型进行调整——具体在文本分类的任务上,则是使用带有标签的数据集进行有监督的机器学习,对模型进行训练。由于BERT的预训练采用的模型架构基于多层双向Transformer编码器(multi-layer bidirectional transformer),能够适应采取其他模型(如LSTM、GRU)构建小规模分类任务的需求。

因此,基于深度双向编码机制以及半监督预训练模型的思想,在神经网络架构优化、数据规模效应的影响以及强大计算能力支持下,BERT具有较强的文本特征学习能力,也有较强的迁移学习能力。本研究使用BERT的目的在于构建文本分类器,对微博W1的内容主题以及意见领袖U1的身份信息进行多类别分类(multiclass classification)。

首先,分别从新浪微博“热门微博分类”与“名人堂”获取具有类别标签的微博数据($n=28961$)、包含认证信息与简介的用户数据($n=65045$)。在标签数据预处理的过程中,对相似的类别进行了整合(如“电影”和“电视剧”的类别整合为“影视”,“媒体”“媒体精英”和“时评”整合为“新闻与传媒”等),以提高模型特征识别的准确度。接下来,利用两个带类别标签的数据集分别对BERT模型进行微调训练与测试。为防止过拟合,模型在验证集准确度下降前中断训练。最后的模型测试精确率、召回率、F1值都在80%左右(如表1所示)。

表1 BERT模型微调结果

参数	微博主题分类模型	用户身份分类模型
准确度	0.9010	0.8256
损失函数值	0.3375	0.6895
测试精确率	0.8138	0.8099
测试召回率	0.8134	0.8031
F1值	0.8114	0.8026

2. 社会网络分析法

社会网络分析理论在复杂网络理论的影响下不断发展,关系如何影响行为、个体行为如何影响网络结构、群体行为如何凝聚成场力等问题是社会网络分析关注的最基本的研究问题。本研究采取社会网络分析的方法,对微博用户现实的转发行为及其网络结构进行研究。首先,提取用户U1之间的转发关系,以转发微博为边,原发用户作为源节点,转发用户作为目标节点,组成有向网络。然后,采取模块度(modularity)聚类的方法划分分子群。模块度旨在衡量网络划分的质量:在模块度较高的网络中,每个子群中各节点间有密集的连线,而处在不同子群的节点间的连线则很少。^⑥因此,在同一组群中的节点有着更为紧密的联系或传播关系。克劳赛特(Clauset)、纽曼(Newman)和

莫尔(Moore)对这一算法进行了调整,使其能够适应规模更大、分布更为离散的大型网络。^④因此,本研究利用 No-dexl 内置的 Clauset - Newman - Moore 算法对网络依据模块度进行子群的聚类。基于网络结构的子群聚类方法能够弥补简单依据用户身份或者传播内容类型划分组群的局限性,有助于观察社交媒体用户行为本身的特点。

另一方面,依据简单的行为数据指标统计(如转发

量、评论量等)衡量意见领袖的传播影响力,容易忽视对网络结构的考量,高估单一中心控制的星形网络中意见领袖的传播力——在此类型的网络中容易形成组员与内容同质化的回音室效应。^⑤因此本研究重点关注跨圈层的传播行为,利用跨子群以及跨层级的边排序方法衡量节点的跨圈层传播力,以识别跨圈层的意见领袖。

整体研究设计的路径如图2所示。

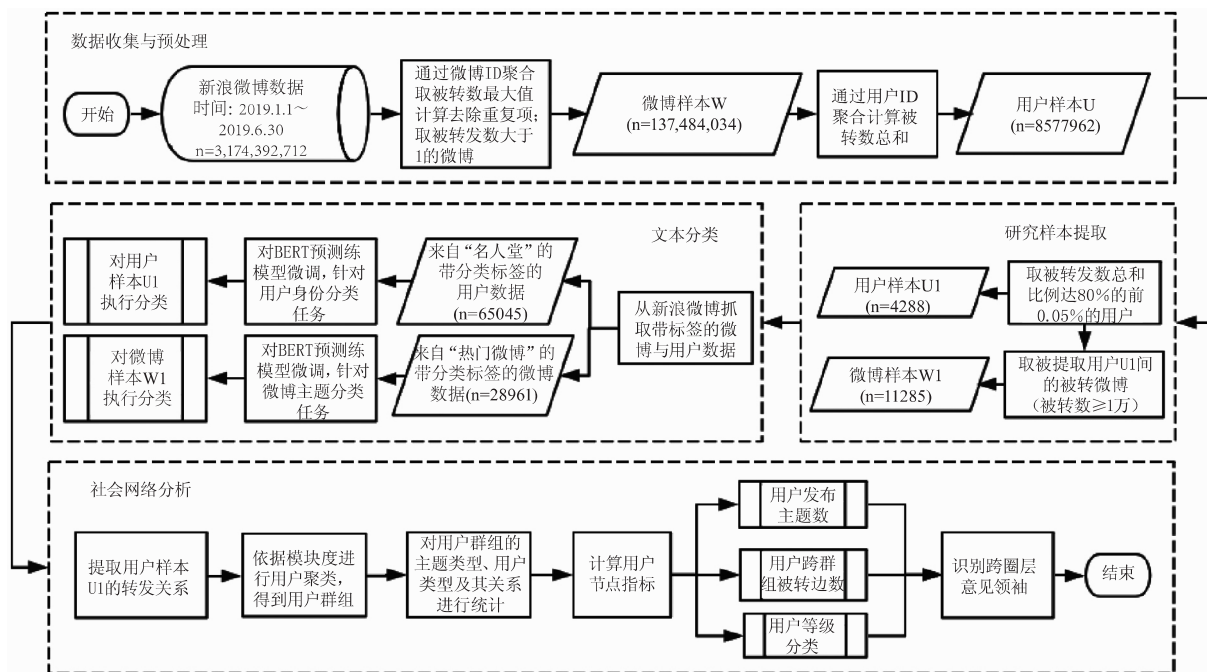


图2 研究设计路径

四、分析结果

(一) 用户的身份及其主题参与倾向

利用 BERT 模型构建分类器的分类结果显示,微博头部用户的身份以及参与的微博主题类型都有较高的集中度。“明星娱乐”“时尚与美妆”“搞笑”“音乐”“影视”等文娱、消费生活领域的用户占比较高,其次才是“新闻与传媒类”的用户,而“教育”“政务”“健康与医疗”“公益”等涉及公共生活领域的用户占比则较低(如图3所示)。

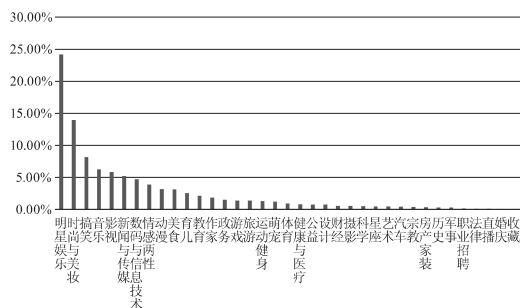


图3 各类型用户数比例分布

而在微博主题方面,“情感”“音乐”“影视”“综艺”类的内容占比很高,其次才是“社会”类的新闻内容(如图4所示)。可见微博整体具有明显的娱乐与消费生活的属性倾向。

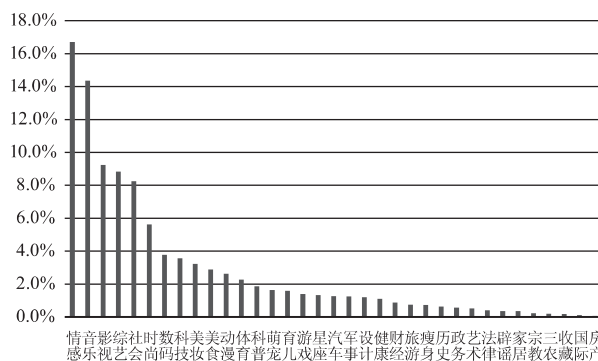


图4 各主题类型微博数比例分布

那特定身份类型的头部用户参与的内容主题倾向又如何呢?对用户身份类型与参与的微博主题类型(包括发布与转发)进行交叉统计显示(如图5所示):身份

类型与主题类型之间具有一定对应关系。例如,明星娱乐类的用户主要参与的内容主题类型是“音乐”(17%)、“情感”(14%)、“影视”(12%)、“综艺”(12%)等;新闻与传播类用户则是“社会”(21%)等。但也有与用户类型不完全相符的内容主题参与倾向的存在,如育儿类头部用户参与比例更高的主题类型是“综艺”(16%)、“时尚”(20%)、美食类头部用户为“情感”(18%)等。而且

从整体看,各类别用户参与的内容主题数平均超过10,参与的内容主题类型是较广泛的。因此用户的身份类型与主题参与倾向之间并非简单的线性关系,不同类型的用户对特定主题也具有注意力的交叠关系。在这种情况下,只依据用户身份类型或者内容主题划分群组就具有很大的局限性,需要引入社会网络关系分析在子群中进一步观察其实际的传播行为。

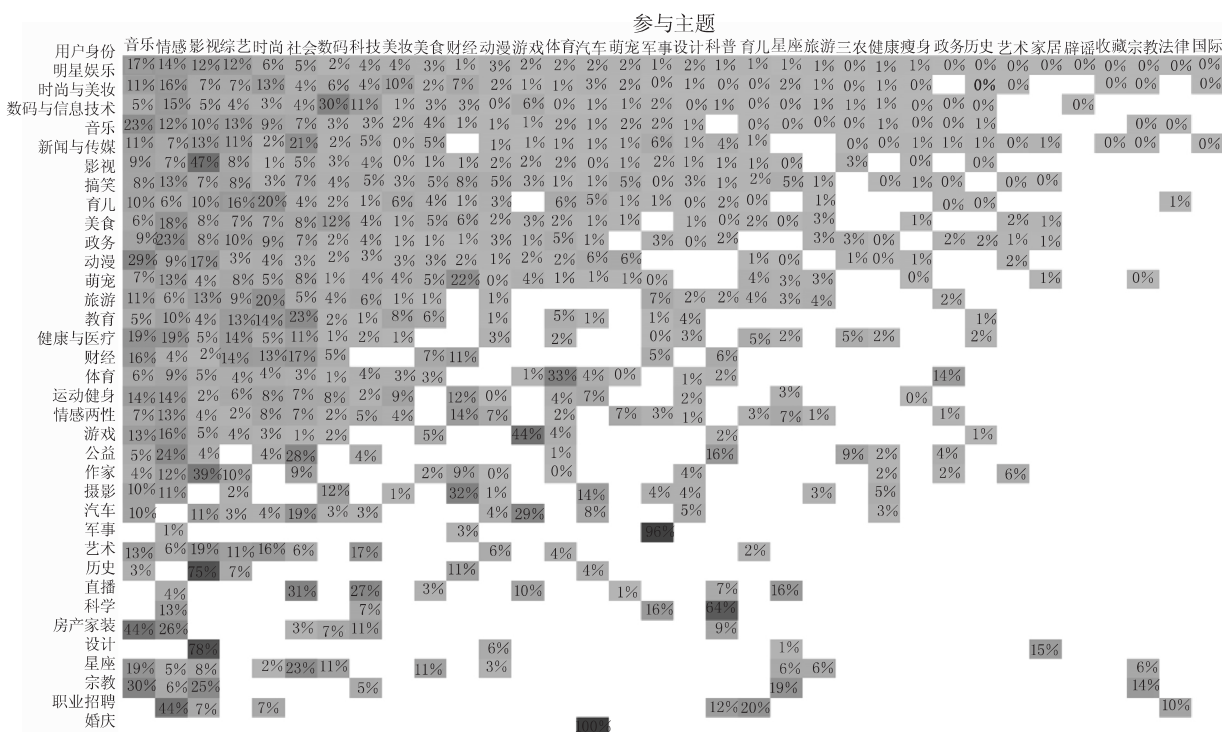


图5 不同身份类型的用户参与的内容主题倾向

(二) 子群参与的内容倾向与主题圈层化

子群聚类的结果显示,子群之间的成员的联系活跃度具有明显的差距(如图6所示)。

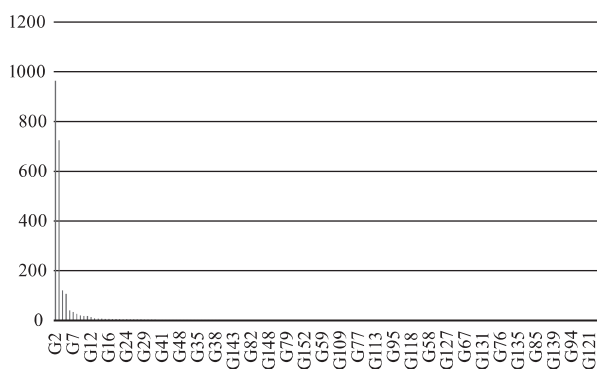


图6 各组别用户数分布

79%的组别只有一位成员,表明其与其他头部用户缺少互动关系,处于相对孤立的状态,用户类型单

一,更有可能形成单一中心控制的星形网络。而成员数量更多的子群中(如子群G1-G10),用户类型与主题分布则具有多样性,头部用户之间的互动关系也更为活跃(如图7所示)。

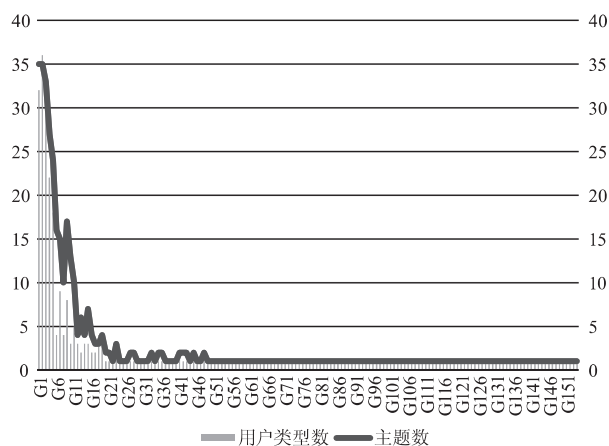


图7 各组别用户类型数与主题数分布

那不同子群参与的主题又呈现着怎样的分布情况呢?图8呈现了前十个用户互动活跃子群的参与主题成分。可以发现,只有“情感”“音乐”“影视”“综艺”等少数几个主题在多个子群中都具有高占比水平,而大部分主题只在特定子群中呈现高占比水平。例如,

G2、G3的“社会”、G1、G4的“时尚”、G5的“数码”等。这表明特定几类主题“情感”“音乐”“影视”“综艺”)在不同的活跃子群中都具有广泛的传播,而大部分主题则具有圈层化的特点,只在特定的几个子群中具有较高的集中度(如表2所示)。

	组别									
	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10
情感	12.63%	21.94%	13.78%	12.00%	15.64%	27.06%	19.05%	24.39%	11.11%	30.00%
音乐	19.26%	9.08%	13.33%	11.11%	11.85%		19.05%	39.02%	16.67%	2.50%
影视	8.62%	8.77%	15.32%	12.89%	11.37%	1.18%	11.11%	7.32%	3.70%	2.50%
综艺	15.91%	3.44%	6.54%	8.00%	5.21%		9.52%			
社会	6.21%	10.71%	9.42%	5.78%	4.74%	1.18%	3.17%			5.00%
时尚	9.79%	2.37%	3.65%	11.56%	4.27%	4.71%	3.17%		5.56%	
科技	3.58%	3.43%	3.14%	4.44%	2.37%	3.53%	11.11%	7.32%		17.50%
数码	2.82%	3.55%	3.33%	5.33%	12.80%		1.59%	2.44%		
动漫	1.68%	3.68%	2.56%	1.33%	0.95%	3.53%	3.17%	4.88%	1.85%	2.50%
美食	2.07%	2.85%	2.95%	4.44%	2.84%	3.53%		7.32%	7.41%	
美妆	3.27%	1.75%	4.17%	6.67%	0.47%	5.88%			1.85%	2.50%
科普	0.88%	3.25%	2.37%	1.78%	2.84%	3.53%	1.59%		1.85%	2.50%
萌宠	0.91%	3.08%	1.47%	1.33%	0.47%			2.44%	7.41%	
体育	2.79%	1.46%	2.56%	1.78%	7.11%					
星座	0.51%	2.27%	0.83%	0.44%	2.37%	10.59%		2.44%	3.70%	5.00%
育儿	0.82%	2.19%	1.67%	0.44%	1.42%	4.71%	1.59%		7.41%	2.50%
军事	0.85%	1.76%	2.63%	1.33%					3.70%	
游戏	0.88%	1.31%	1.73%		1.90%	4.71%	3.17%	2.44%		5.00%
财经	0.43%	1.39%	0.71%	1.78%	0.95%	4.71%				10.00%
设计	1.09%	1.52%	0.71%	0.44%	2.84%	1.18%	3.17%		1.85%	
汽车	0.99%	0.76%	1.99%	1.33%	2.84%	18.82%	3.17%			
健康	0.52%	1.90%	0.77%	1.33%	0.47%				11.11%	
瘦身	0.34%	1.28%	0.64%						7.41%	
历史	0.29%	1.27%	0.64%	1.33%	2.84%					
政务	0.59%	0.78%	0.96%							
旅游	0.56%	0.91%	0.71%	0.44%	0.95%	1.18%				
法律	0.23%	0.70%	0.19%	0.44%						
艺术	0.15%	0.93%	0.26%	0.44%						12.50%
辟谣	0.39%	0.44%	0.13%	1.33%						
家居	0.31%	0.44%	0.26%	0.44%					3.70%	
三农	0.42%	0.01%	0.26%							
国际	0.05%	0.25%	0.06%		0.47%					
宗教	0.08%	0.18%	0.26%				6.35%			
收藏	0.09%	0.21%							3.70%	
房产	0.02%	0.16%								

图8 子群组别与参与内容主题的倾向

(三) 子群活跃度与子群异质性的关系

但另一方面,主题的圈层化只出现在子群的局部网络。在整体网络上,前十个活跃子群参与的主题数平均大于10,子群用户身份也具有多样性(如图7所示)。由于数据分布较为分散,不具备正态分布

性,使用斯皮尔曼相关性分析研究发现:子群成员数与子群主题数、用户类型数都具有显著的正相关关系(Sig均小于0.05)(如表3所示),这说明若子群成员数量增加,子群的主题数以及用户类型数也具有增加的趋势。成员数多而互动活跃的子群,其参

与的主题与用户类型更有可能呈现多样化的特点,没有形成明显的信息回音室效应。但在成员数少、与其他群组的头部用户缺乏互动关系的子群,则呈现参与主题以及用户类型单一的趋势,更有可能形成回音室效应。

表 2 具有圈层化特点的主题及其所在子群

子群	主题及其在子群中的占比
G1	时尚(9.79%)、社会(6.21%)
G2	社会(10.71%)
G3	社会(9.42%)
G4	时尚(11.56%)、美妆(6.67%)
G5	数码(12.80%)、体育(7.11%)
G6	汽车(18.82%)、星座(10.59%)
G7	科技(11.11%)
G8	科技(7.32%)、美食(7.32%)
G9	健康(11.11%)
G10	科技(17.5%)、艺术(12.50%)

观察前十个活跃子群的聚类结果(如图9所示)可发现,这些具备参与主题与用户类型多样性的子群之间存在着许多跨子群的传播关系,整体上的信息流动并不是封闭的,很有可能存在积极联系不同子群的跨圈层传播的意见领袖。那么,如何对跨圈层的意见领袖进行识别呢?

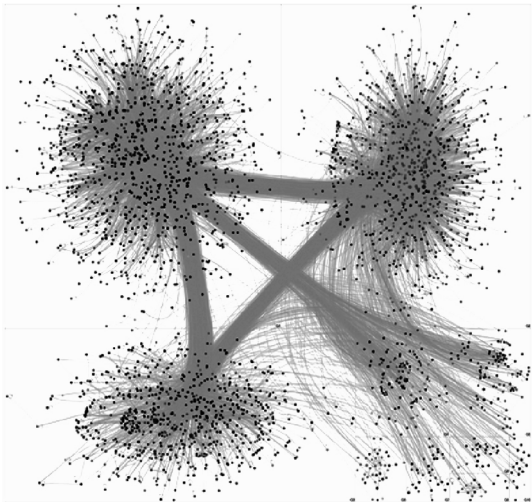


图 9 前十个子群聚类结果

表 3 子群成员数、主题数与用户类型数之间的相关性分析

		子群用户		
		子群成员数	子群主题数	类型数
斯皮尔曼 Rho	子群成员数 相关系数	1.000	0.825 **	0.723 **
	子群主题数 相关系数	*	1.000	0.852 **
	子群用户类型数 相关系数			1.000

注: ** . 在 0.01 级别(双尾) 相关性显著。

(四) 跨圈层意见领袖的作用

“圈层”包含圈子关系以及层级关系,子群聚类呈现的是圈子关系。为了对用户跨层级的传播关系进行观察,本研究通过计算头部用户总转发量的四分位数,将用户分为信息扩散能力递减的四个等级: A 级、B 级、C 级以及 D 级(如图 10 所示),并对存在跨子群关系的头部用户进行计算(如图 11 所示)。

图 11 呈现跨圈层传播关系,左侧表示微博原发用户的等级及其所处的子群,可以发现原发微博 65.75% 都来自于 A 级用户,而 A 级用户主要分布在 G1、G2、G3 这三个子群;右侧则表示转发微博的用户的等级及其所处的子群,转发微博的用户中占比更高的是 D 级用户(33.07%),且主要分布在 G1、G2、G3 这三个子群;线条则表示微博从发布到被转发的流向过程,线条的粗细代表微博条数的多少,可以发现 A、B、C 等级的用户都存在积极转发下级用户微博的行为。通过

统计总跨子群与层级的边数并进行排序,识别出的前十位跨圈层意见领袖如表 4 所示。这些跨圈层的意见领袖参与的内容主题都具有多样化的特点。

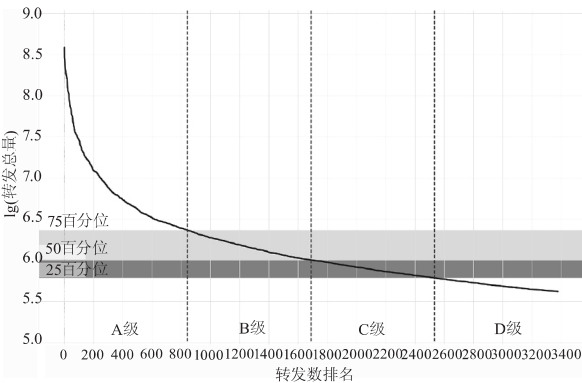


图 10 用户层级划分结果

进一步对用户跨圈层的被转发边数以及主题数进行相关性分析,结果显示:跨圈层的被转发边数与

主题数之间存在显著的正相关关系(如表5所示)。这表明,在很大程度上,意见领袖参与内容主题数增多时,跨圈层的传播影响力也具有增大的趋势。跨圈层意见领袖因而成为联系不同子群的关键节点,影响着子群之间的信息流通以及圈层内部的信息多样性。

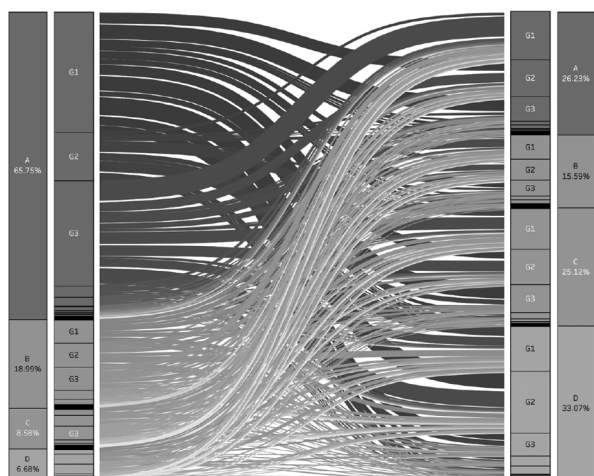


图 11 意见领袖跨圈层传播桑基图

表 4 前十位跨圈层意见领袖

用户名	参与主题数	跨圈层边数
人民日报	25	186
央视新闻	25	132
新浪娱乐	18	70
亚尼大帝	18	66
ECHO 骑着小毛驴去拉萨	29	65
锦衣夜行焱十一	27	52
头条新闻	24	48
让红包飞	17	43
共青团中央	15	42
新浪电影	14	40

表 5 发布主题数与跨圈层被转边数之间相关性分析结果

相关性		
		跨圈层被转发边数
斯皮尔曼 Rho ^①	子群主题数	0.758 **
Sig. (双尾)	子群主题数	0.000

注: ** . 在 0.01 级别(双尾) 相关性显著。

注释:

① Webster James G., Ksiazek Thomas B. *The Dynamics of Audience Fragmentation: Public Attention in an Age of Digital Media*. Journal of Communication, vol. 62, no. 1, 2012, p. 41.

五、结语

本文采取机器学习文本分类以及社会网络分析的方法,以新浪微博平台为例,探讨了意见领袖的跨圈层传播与回音室效应的问题。研究发现,微博的转发量非常集中,0.05%的头部用户创造了80%的微博转发量,呈现明显的幂律分布特点,并且头部用户类型与微博主题类型分布较为集中。但这并不能说明新浪微博平台具有典型的“回音室效应”。通过对用户身份以及参与内容主题倾向的分析发现,不同类型用户参与的内容主题实际上相对广泛,身份与内容参与之间并非简单的线性关系。

而对子群的分析则发现,确实存在与其他头部用户缺少互动关系而表现的相对孤立的组别,这些组别更有可能形成用户以及内容同质化的圈层,出现回音室效应;但在成员多而互动活跃的子群,其用户身份类型以及参与的内容主题具有多样性的趋势,并且存在着跨子群的互动关系。这些活跃子群并没有形成明显的“回音室效应”。

最后,本研究对用户分级并进行跨圈层边数排序,识别出跨圈层意见领袖,发现意见领袖的跨圈层被转发边数与参与主题数之间存在显著的正相关关系;表明意见领袖参与内容主题数增多时,跨圈层的传播影响力也有增大趋势,因而成为联系不同子群的关键节点,影响着子群之间的信息流通以及圈层内部的信息多样性,这对于避免回音室效应具有积极作用。

但是由于本文的数据采集局限在2019年上半年,历史数据的时间范围影响了研究结果,并且本次研究关注的是头部用户之间的转发关系,呈现的是局部网络的结果。未来的研究需要观察更广泛历史范围的用户行为表现,并且尽可能分析更大规模的网络结构,以提高研究的普适性。其次,研究并未进行文本的情感分析,意见领袖被转关系的正负面属性并未纳入考量,若要对意见领袖的影响进一步研究,对文本属性还需更细致的考察。另一方面,本文研究的是以新浪微博为代表的单一平台的网络结构。由于不同社交媒体平台存在着不同的规则设置与传播模式,在未来的研究中也需要对网络结构的多平台对比予以关注。最后,本研究采取的是横截面数据的统计,验证的是相关关系,未来应该引入对时间序列的分析,进一步对各因素之间的因果关系进行深入的研究,解释网络动态演变的机制。

- ② Garrett R. Kelly. *Echo Chambers Online: Politically Motivated Selective Exposure Among Internet News Users*. Journal of Computer – Mediated Communication ,vol. 14 ,no. 2 2009. p. 267.
- ③ Dijk José van ,Thomas Poell. *Understanding Social Media Logic*. Media and Communication ,vol. 1 ,no. 1 2013. p. 6.
- ④ Garrett R. Kelly ,Dustin Carnahan ,Emily K. Lynch. *A Turn Toward Avoidance? Selective Exposure to Online Political Information 2004—2008*. Political Behavior ,vol. 35 ,no. 1 2013. p. 131.
- ⑤ Allcott Hunt ,Matthew Gentzkow. *Social Media and Fake News in the 2016 Election*. Journal of Economic Perspectives ,vol. 31 ,no. 2 2017. pp. 223—224.
- ⑥ McPherson Miller ,Lynn Smith – Lovin ,James M Cook. *Birds of a Feather: Homophily in Social Networks*. Annual Review of Sociology ,vol. 27 ,no. 1 2001. p. 416; Furman Ivo ,Asli Tun. *The End of the Habermasian Ideal? Political Communication on Twitter During the 2017 Turkish Constitutional Referendum*. Policy & Internet <https://onlinelibrary.wiley.com/doi/abs/10.1002/poi3>. 2019. p. 13
- ⑦ Pariser E. *The Filter Bubble: What the Internet Is Hiding from You*. New York: Penguin. 2011. p. 10.
- ⑧ Jamieson K ,Cappella J. *Echo Chamber: Rush Limbaugh and the Conservative Media Establishment*. London: Oxford University Press. 2008. pp. 75—76.
- ⑨ Kim Soojong. *Directionality of Information Flow and Echoes Without Chambers*. PLOS ONE ,vol. 14 ,no. 5 2019. p. e0215949.
- ⑩ Wieringa Maranke ,Daniela van Geenen ,Mirko Tobias Schäfer ,Ludo Gorzeman. *Political Topic – Communities and Their Framing Practices in the Dutch Twittersphere*. Internet Policy Review ,vol. 7 ,no. 2 2018. p. 2.
- ⑪ Del Vicario Michela ,Fabiana Zollo ,Guido Caldarelli ,Antonio Scala ,Walter Quattrociocchi. *Mapping Social Dynamics on Facebook: The Brexit Debate*. Social Networks ,vol. 50 2017. p. 12.
- ⑫ McEwan Bree ,Christopher J. Carpenter ,Jill E. Hopke. *Mediated Skewed Diffusion of Issues Information: A Theory*. Social Media + Society ,vol. 4 ,no. 3 2018. p. 638.
- ⑬ Borgesius Frederik J. Zuiderveen ,Damian Trilling ,Judith Moeller ,Balázs Bodó ,Claes H. de Vreese ,Natali Helberger. *Should We Worry about Filter Bubbles*. Internet Policy Review ,vol. 5 ,no. 1 2016. p. 10.
- ⑭ Nguyen Tien T ,Pik – Mai Hui ,F. Maxwell Harper ,Loren Terveen ,Joseph A. Konstan. *Exploring the Filter Bubble: The Effect of Using Recommender Systems on Content Diversity*. Proceedings of the 23rd International Conference on World Wide Web ,ACM 2014. p. 683.
- ⑮②③ Sanz – Cruzado Javier ,Pablo Castells. *Enhancing Structural Diversity in Social Networks by Recommending Weak Ties*. Proceedings of the 12th ACM Conference on Recommender Systems ,ACM 2018. p. 235.
- ⑯ Groshek Jacob ,Karolina Koc – Michalska. *Helping Populism Win? Social Media Use ,Filter Bubbles and Support for Populist Presidential Candidates in the 2016 US Election Campaign*. Information ,Communication & Society ,vol. 20 ,no. 9 2017. pp. 1399—1400.
- ⑰ Bechmann Anja ,Kristoffer L. Nielbo. *Are We Exposed to the Same “News” in the News Feed*. Digital Journalism ,vol. 6 ,no. 8 2018. pp. 997—998.
- ⑱③⑤ 李卫东、彭静《社交网络平台信息传播的回声室效应仿真实验分析》,《现代传播》2019年第4期,第148页。
- ⑲ Dubois Elizabeth ,Grant Blank. *The Echo Chamber Is Overstated: The Moderating Effect of Political Interest and Diverse Media*. Information ,Communication & Society ,vol. 21 ,no. 5 2018. p. 730.
- ⑳ Bakshy Eytan ,Solomon Messing ,Lada A. Adamic. *Exposure to Ideologically Diverse News and Opinion on Facebook*. Science ,vol. 348 ,no. 6239 ,2015. p. 1131.
- ㉑ Vallone R. P ,Ross L ,Lepper M. R. *The Hostile Media Phenomenon: Biased Perception and Perceptions of Media Bias in Coverage of the Beirut Massacre*. Journal of Personality and Social Psychology ,vol. 49 ,no. 3 ,1985. p. 581.
- ㉒ Luzsa Robert ,Susanne Mayr. *Links Between Users’ Online Social Network Homogeneity ,Ambiguity Tolerance ,and Estimated Public Support for Own Opinions*. Cyberpsychology ,Behavior ,and Social Networking ,vol. 22 ,no. 5 2019. p. 325.
- ㉓ DiFonzo Nicholas ,Jerry Suls ,Jason W. Beckstead ,Martin J. Bourgeois ,Christopher M. Homan ,Samuel Brougner ,Andrew J. Younge ,Nicholas Terpstra – Schwab. *Network Structure Moderates Intergroup Differentiation of Stereotyped Rumors*. Social Cognition ,vol. 32 ,no. 5 2014. p. 410.
- ㉔ 王佳敏《突发事件中意见领袖的识别和影响力实证研究》,《情报学报》2016年第2期,第171页。
- ㉕ 彭丽微《基于SNA与模糊TOPSIS的网络舆情关键节点识别分类模型研究》,《现代情报》2017年第8期,第19页。
- ㉖ 曹玖新《基于多维特征分析的社交网络意见领袖挖掘》,《电子学报》2016年第4期,第899页。
- ㉗ Yu Xiao ,Xu Wei ,Xia Lin. *Algorithms of BBS Opinion Leader Mining Based on Sentiment Analysis*. Proceedings of the 2010 international conference on Web information systems and mining 2010. p. 363.
- ㉘ 张磊《语义社会网络的超网络模型构建及关键节点自动化识别方法研究》,《现代图书情报技术》2016年第3期,第9页; Ma Ning ,Yijun Liu. *SupernodeRank Algorithm and Its Application in Identifying Opinion Leader of Online Public Opinion Supernetwork*. Expert Systems with Applications ,vol. 41 ,no. 4 2014. p. 1366.
- ㉙ 数据来源由暨南大学新闻与传播学院计算新闻传播研究中心提供支持。
- ㉚ Bengio Yoshua ,Réjean Ducharme ,Pascal Vincent ,Christian Janvin. *A Neural Probabilistic Language Model*. J. Mach. Learn. Res. ,vol. 3 2003. p. 1142.
- ㉛ Devlin Jacob ,Ming – Wei Chang ,Kenton Lee ,Kristina Toutanova. *BERT: Pre – Training of Deep Bidirectional Transformers for Language Understanding*. ArXiv: 1810.04805 <http://arxiv.org/abs/1810.04805>. 2019.
- ㉜ Newman M. E. J. *Modularity and Community Structure in Networks*. Proceedings of the National Academy of Sciences ,vol. 103 ,no. 23 2006. p. 8577.
- ㉝ Clauset A ,Newman M. E. J ,Moore C. *Finding Community Structure in very Large Networks*. Physical Review E ,vol. 70 ,no. 6 2004. p. 1.

(作者汤景泰系暨南大学新闻与传播学院教授、博士生导师;陈秋怡系暨南大学新闻与传播学院硕士研究生)

【责任编辑:张毓强】