创业管理

网络圈层化视域下大学生社交特征识别及演化分析

——基于BERTopic主题建模

尹慧林，郑毅恒，周 欢

（湖南工业大学经济与管理学院，株洲 412007）

摘要：随着网络化时代的推进，大学生社会交往深受互联网影响，“大学生圈群”成为高校普遍存在的社交现象。使用Python采集知乎和百度贴吧平台上以大学生社交为主题的文本为数据源，采用BERTopic主题建模，以识别大学生社交的主题分布及提取其主题特征，分析不同时间序列上的主题内容演化趋势。研究发现，大学生社交话题呈现出显著的多元化与圈层化，核心主题可归纳为4个社交领域：人际交往社交、职业发展社交、学习合作社交和现实焦虑社交。演化分析表明，大学生的核心诉求与心态变化在不同阶段发生着规律性演变。将BERTopic模型应用于大学生社交领域的动态主题演化分析，有效捕捉了圈层语境下话题的细粒度特征与时序性规律。未来研究可进一步结合多模态数据探索主题演化与特定社会事件的关联性。

关键词：大学生社交；圈层化；BERTopic模型；主题特征；主题演化

中图分类号：G250 **文献标志码：**A **文章编号：**1671－1807(2026)09－0297－07

随着互联网技术的深度发展与社交模式的持续演进，大学生互联网使用率也急剧上升，催生出“圈群”这一社交形式。各类在线社交平台通过算法推荐和兴趣社群，将用户划分为众多内部认同度高、外部差异性强的圈群，深刻重塑了个体信息获取与社会交往的方式[1]。大学生作为数字原住民，其社交行为高度依赖于微博、知乎、贴吧等社交媒体。这些平台不仅构成了大学生日常互动的主要场域，更是其塑造身份认同、获取社会资本、感知外部世界的关键渠道[2]。

2017年国家颁布的《中长期青年发展规划(2016—2025年)》特别强调青年应主动、自信地参与群体性交往去适应社会、融入社会。在网络化背景下，青年群体基于趣缘纽带进行互动、交往等社交行为，形成了大小不同、文化不一的社交圈群，在同一社交圈群之下又会分化出许多层级，这种基于趣缘而结成的有着同质文化属性的聚合体叫作网络圈层[3]。然而当前，在网络“圈层化”的现实背景下，大学生面临新的社交挑战。一方面，信息茧房与回音壁效应可能导致视野窄化、观点极化，加剧群体间的认知隔阂[4-5]；另一方面，圈群内部的话语模式、情感结构与议题焦点往往隐蔽而流动，难以被外部观察和准确把握[6]。既有研究多从宏观理论或问卷调查入手，对圈群内部海量、动态、真实的文本数据进行细粒度挖掘尚显不足，致使对大学生社交的主题偏好及演化规律的认识仍较为模糊[7]。

因此，精准识别圈层环境下大学生社交的核心主题，并动态追踪其演化路径，对于理解青年群体、促进其健康发展具有重要的理论与现实意义。研究聚焦于知乎与百度贴吧两大典型圈层化平台，通过Python爬虫构建语料库，采用主题的双向编码器表示(bidirectional encoder representations for topic，BERTopic)主题模型对评论文本进行主题挖掘与演化分析，旨在揭示大学生社交话题的分布规律与动态变迁，为理解当代大学生的社交需求、情感动向与关注焦点提供依据。

收稿日期：2025-12-23

基金项目：2024年度湖南省大学生创新训练计划一般项目(S202411535097)；湖南省教育科学“十四五”规划2025年度一般项目(XJK25BGD009)；2025年湖南工业大学学位与研究生教育教改研究重点项目(JGZD25003)

作者简介：尹慧林(2001—)，女，湖南永州人，硕士研究生，研究方向为数据挖掘与分析、知识管理；郑毅恒(2003—)，男，山东邹城人，研究方向为科技情报服务；通信作者周欢(1982—)，女，湖南益阳人，博士，教授，研究方向为情报学理论方法、管理决策。

1 文献综述

1.1 大学生网络社交“圈层化”研究现状

网络“圈层化”是算法推荐与用户偏好共同作用下的虚拟社群聚类现象[8]。研究普遍认为,大学生群体的高度同质性与移动互联网的高渗透率是圈层形成的温床。例如,邱月[1]指出,大学生通过加入各种兴趣社群,寻求身份认同和情感归属,呈现出独特的社交行为特征,但群体跨圈层交流存在一定障碍。其次,在社交影响方面,研究多聚焦于圈层化对大学生心理与社会化的双重效应。有研究表明,圈层为大学生提供了安全表达和自我实现的空间,有助于社会资本积累[9]。但也有研究表明,圈层内的回音壁效应易诱发观点极化、信息茧房及群体孤独[10]。还有学者从不同角度看待大学生网络社交。例如,Hui 等[11]发现通过一般亲社会行为和学校志愿服务能对青少年生活满意度和自尊产生积极影响;Okray 等[12]认为习惯性使用 SNS 会使得大学生受虚拟世界影响容易产生逃避行为从而影响正常生活。尽管这些研究提供了有价值的见解,但现有研究多依托宏观理论或问卷调查,对圈层内部文本数据进行定量化挖掘的探索尚显不足,限制了对大学生网络社交行为的理解。

1.2 主题挖掘研究现状

主题挖掘是自动识别海量非结构化文本潜在语义结构的核心技术[13]。以线性判别分析(linear discriminant analysis,LDA)为代表的统计概率模型曾是该领域的基准算法[14],被广泛应用于社会科学、新闻分析和生物信息学等领域。例如,曹树金和曹茹烨[15]结合 LDA 主题模型与支持向量机(support vector machine,SVM)分类算法对情报学领域中文社会科学引文索引(Chinese Social Sciences Citation Index,CSSCI)文献摘要进行潜在主题识别;陈淑一和朱华桂[16]基于 LDA 主题模型构建高校新闻网文本主题挖掘框架,识别高校管理的9个关键锚点并分析其演化趋势;Dai 等[17]利用 LDA-SCGB(LncRNA-disease associations based on condensed gradient boosting)模型融合奇异值分解与凝聚梯度提升算法,预测长链非编码 RNA(long non-coding RNA,lncRNA)与疾病的潜在关联;孟薇和王晟聪[18]对兰德公司的 229 份涉华研究进行词频分析和关键词语义网络分析,探讨其注意力分布与阶段性特征。后来研究者提出了专门针对短文本的模型,此外,Word2Vec、GloVe 等词嵌入技术进一步深化了主题模型的应用。近年来,传统主题模型因运行速度慢、生成能力不够强而面临挑战,学者们提出了 BERTopic 主题建模技术,通过预训练语言模型生成上下文感知的词向量。

1.3 BERTopic 主题建模研究现状

BERTopic 是由 Grootendorst[19]提出的一种融合预训练语言模型和层次聚类技术的主题识别方法。该模型利用 Sentence-BER 生成高质量的文档嵌入向量,通过降维和密度聚类,自动识别主题并有效过滤异常点,再基于词频和语义信息的基于类别的词频-逆文档频率(class-based term frequency-inverse document frequency,c-TF-IDF)算法提取主题关键词,生成可解释性强、质量高的主题表征。因此,BERTopic 模型考虑了文本数据中词汇的语义、结构及序列,使获取的主题在一定程度上可实现词义消歧,生成的主题质量显著优于 LDA、非负矩阵分解(non-negative matrix factorization,NMF)等传统模型[20]。近年来,BERTopic 的效能在多个领域得到验证。例如,王益成等[21]运用 BERTopic 对生物技术科技报告进行主题挖掘,识别出该领域未来的研究方向;苗宏慧等[22]借助该模型对东北三省科技人才政策文本进行主题识别与量化分析,揭示区域政策的结构特征与差异化导向;毛太田和郑锐[23]通过构建 BERTopic 模型探究人形机器人专利技术主题与演化路径;李豪等[24]利用 BERTopic 模型挖掘计算机科学领域论文主题以识别学者研究兴趣特征,通过构建 Prompt 指令引导大语言模型自动生成学者研究兴趣的动态描述文本,验证了其在长文本上的适用性。在社交媒体短文本场景下,该模型也广泛用于用户评论挖掘和舆情主题追踪。

然而,现有研究多聚焦于生物医学、政策或专利等规范文本,针对社交媒体中“圈层化”特征鲜明、语境隐蔽的大学生社交文本关注不足。因此,采用 BERTopic 模型对大学生社交圈群进行主题挖掘与演化分析,旨在精准洞察青年群体的社交动态。

2 数据来源与研究方法

2.1 数据来源

选取知乎与百度贴吧作为语料来源。知乎凭借高质量问答属性集聚了大量活跃大学生群体,百度贴吧则通过兴趣驱动的“吧”呈现出典型圈层化特征,均具有较强代表性。研究利用 Python 爬虫,于 2025 年 8 月采集帖子标题、正文及评论等原始数据17 359条,经人工去重、剔除内容缺失及无关文本等清洗预处理,最终保留 2018—2025 年的有效文本数据 12 253 条。

2.2 研究方法

通过 Sentence-BERT 级数据分析将文本数据以每句话为单位划分为不同的句子,借助BERTopic模型用于文本聚类和主题建模。BERTopic 模型利用了预训练的 BERT 模型,通过词向量的聚合和聚类,将文本数据转化为主题集合,主题建模过程如图 1 所示。首先,载入 jieba、pandas、BERTopic、基于层次密度的噪声应用空间聚类(hierarchical density-based spatial clustering of applications with noise,HDBSCAN)及均匀流形逼近与投影(uniform manifold approximation and projection,UMAP)等多个 Python 数据库,获取包含百度贴吧和知乎两个社交平台评论文本,使用 jieba 分词工具对评论文本进行数据预处理,在此期间读取哈工大停用词表,通过 Sentence-Transformers 将 content 文本"切割"为句子级列表形式数据。随后,在嵌入阶段,使用 Sentence-BERT 生成文档向量,选取功能最强支持多语言的"pertrain \ paraphrase-multilingual-MiniLM-L12-v2"词向量训练模型进行文本嵌入;在聚类阶段,分别采用 UMAP 降维方法和 HDBSCAN 聚类方法聚类相似文档,在对比多次建模分析结果和参考已有研究的基础上,最终确定 UMAP 邻近样本点数量参数设置为 15(n_neighbors),嵌入数据的降维空间设置为 5(n_components),HDBSCAN 生成的最小聚类规模设置为 10(min_cluster_size),该数值越大,主题挖掘数量越少;在主题表示阶段,选择 sklearn 机器学习工具库中的 CountVectorizer 特征数值化方法。最后,利用 c-TF-IDF 算法进行加权,识别文本数据中的主题分布,对模型结果可视化。

3 分析与讨论

3.1 大学生社交圈群主题分布识别

运用 BERTopic 模型对预处理后的12 253条有效文本数据进行主题建模,最终识别出 9 个稳定的社交主题簇,并生成大学生社交主题间距图与主图相似度热力图(图 2)。图 2(a)显示,各主题在多维向量空间内分布相对均衡且边界清晰,表明生成的主题在语义上具有较好的区分度,模型成功捕捉到了不同社交话题间的差异性;图 2(b)显示,除中轴线自相关区域外,其余绝大部分单元格色调普遍偏浅,表明所得主题间相关程度较低,主题分布结果相对合理。

3.2 大学生社交圈群主题特征分析

运用 BERTopic 模型提取前 8 个显著主题,其 Top5 关键词及权重分布(图 3)揭示了大学生社交关注点从现实生存到价值认同的多元映射,各主题的关键词分布特征如下。

从权重分布看,主题 3 以"论文"(0.23)为核心,协同"期刊"与"发表",映射出极高浓度的学术研究诉求。主题 0 与主题 1 呈现明显的学业周期特征,前者侧重"专业"(0.13)与"大学",聚焦入校适应与学科选择;后者则聚焦于"大学"与"毕业",捕捉大学生面对学业终点时的规划与考研心态。主题 4 突出"宿舍"(0.16)及室友关系,凸显住宿环境在校园人际互动中的枢纽地位;主题 2 以"兼职"(0.31)为高权词,反映出学生通过社会实践实现经济自主的强烈倾向;主题 5 聚焦"老师"与"辅导员",体现大学生对学校管理层级的关注;主题 6 则直指"社交"与朋友交往的心理需求;而主题 7 以"志愿"(0.32)和"高考"为核心,揭示出学生在升学衔接阶段的求助与咨询热点。

可视化分析显示,不同主题呈现出显著的词权差异。主题 2、3、7 的核心词权重远高于其他词汇,表明讨论焦点在"兼职、论文、志愿"等特定领域高度集中;而主题 1、5、6 的词权分布相对均衡,反映了话题内容的多元化。此外,模型识别出同一高频词"大学"在不同语境下的权重偏移:在主题 0 中指向专业认知,在主题 1 中指向学业规划,在主题 6 中则指向社交环境。这种基于上下文语义的细粒度区分,验证了 BERTopic 在识别圈层语境差异方面的有效性。

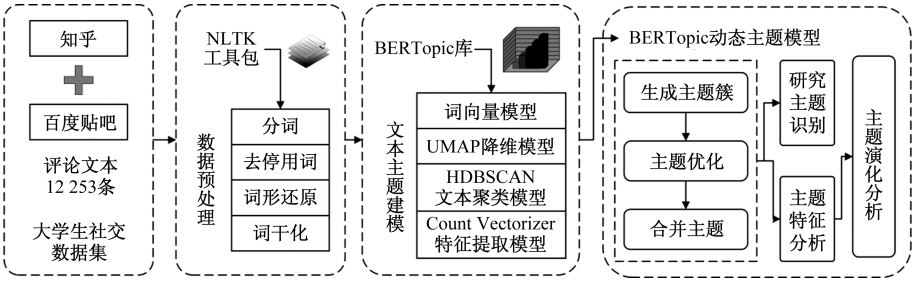


图 1 BERTopic 主题建模过程

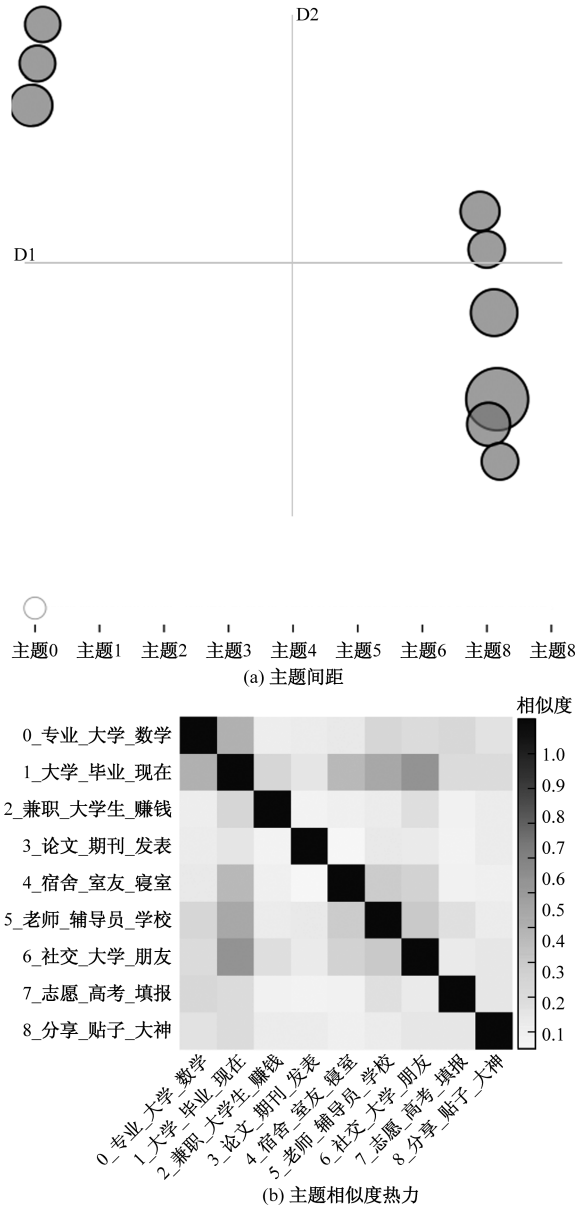


图 2　主题间距图和主题间热力图

3.3　大学生社交圈群层次聚类及层次划分

HDBSCAN 是由 Campello 等[25]提出的一种结合密度聚类与层次聚类的算法，该算法通过集成不同 ε 参数下的基于密度的噪声应用空间聚类（density-based spatial clustering of applications with noise，DBSCAN）聚类结果，识别不同密度的簇类，对参数选择具有较强的鲁棒性。采用 HDBSCAN 算法对 BERTopic 提取的主题及其关键词进行层次聚类，挖掘大学生社交圈群中隐含的主题结构及其层次关系（图 4）。在聚类初始阶段，“宿舍_室友_寝室”与“老师_辅导员_学校”两个主题首先实现合并，表明校园生活环境与师生互动话题在大学生社交讨论中具有紧密的内在联系。随着聚类过程的

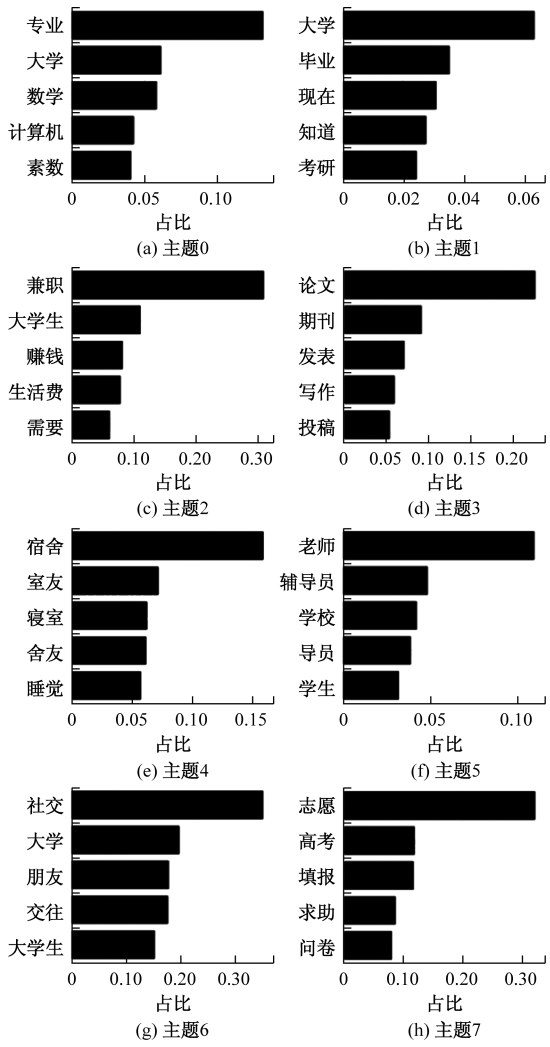


图 3　大学生社交主题内容及主题词占比

推进，“大学_毕业_现在”与“社交_大学_朋友”两个主题聚为一类。在更高的相似度层次上，“志愿_高考_填报”与“专业_大学_数学”形成新的簇群，显示出入学前的院校志愿抉择到入学后专业学科探索的学业生命周期延续性。“分享_帖子_大神”与“论文_期刊_发表”以及“兼职_大学生_赚钱”这 3 个主题最终聚为一个大类，共同构成了大学生以工具性需求为导向的现实生存话语簇群，集中体现其在学术资本积累、经济自立尝试及数字化资源互助层面的务实表征。

基于层次聚类分析结果，结合各主题的关键词语义内涵，将大学生的社交话题进一步整合为 4 个核心领域，即人际交往社交、职业发展社交、学习合作社交和现实焦虑社交。

3.4　大学生社交圈群主题演化趋势

在 BERTopic 主题识别的基础上，结合帖子发布时间戳，对 2018—2025 年不同圈层下大学生社交

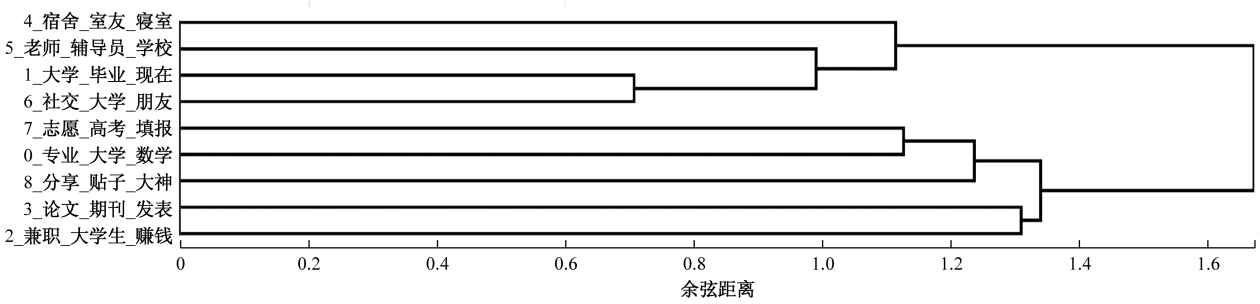


图 4　大学生社交主题层次聚类分析

主题的频率进行时序演化分析(图 5)。研究发现,社交主题的演化并非孤立的文本波动,而是大学生群体对宏观社会变迁、教育政策调整及现实生存压力的动态映射。

在 2018—2022 年的早期阶段,社交话题呈现出由常态校园生活向危机应激模式转化的过程。2019 年前,以主题 4 为代表的线下校园生活话题占据主流,反映了物理空间社交的稳定性。然而,2020 年重大公共卫生事件的爆发打破了既有秩序,社交行为大规模向线上迁移。毕业生面对招聘停滞与“云毕业”困境产生的剧烈心理震荡,直接驱动主题 1 在 2020 年的阶段性爆发,揭示出突发事件可通过改变物理隔离状态重塑大学生的社交话语权。

2023—2024 年,大学生社交主题呈现出结构性爆发态势,这种井喷现象实则是多年政策累积与竞争压力叠加的产物。在教育政策层面,随着教育部《本科毕业论文(设计)抽检办法(试行)》的深入实施以及学术评价体系的持续下沉,学术诚信与科研产出的门槛被提高。这一政策转向直接驱动了主题 3 关于论文发表与期刊研究的热度阶梯式攀升,说明在学历贬值背景下科研资本已演变为大学生应对保研及就业竞争的刚性防御机制。同时,国家对基础学科建设与“强基计划”的持续推进,促使学生社交的核心动力回归至主题 0 所代表的学科专业知识建构,专业身份认同成为数字化圈层互动中的核心逻辑。

与此同时,宏观经济增速放缓与毕业生规模屡创新高的现实矛盾,深刻重构了大学生的实用主义社交导向。主题 2 和主题 8 所代表的兼职赚钱与资源分享话题的高频出现,凸显出经济周期波动对个体生存策略的影响。在生活成本上升与就业预期不稳定的双重挤压下,大学生的社交动机从“经验积累”转向“经济补缺”。网络社交平台已不再仅仅是单纯的信息交互场所,而成为大学生通过“大神

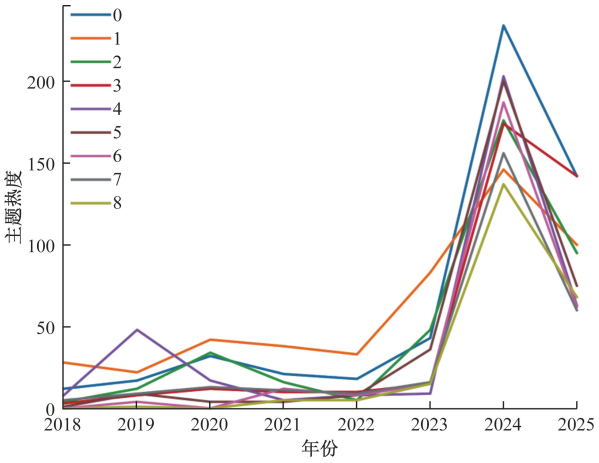


图 5　2018—2025 年不同圈层下大学生网络社交主题演化趋势

分享”与“兼职实践”来实现自我救赎与资源协同的生存网络,印证了外生性环境压力对学生群体行为逻辑的深度塑造。2025 年,各主题频率的高位回落并不代表关注度的退潮,而是预示着大学生已完成从应激性爆发向常态化数字化生存的平稳过渡。

综上所述,大学生社交话题的演化并非均质线性发展,而是内源性需求与外生性环境压力共同作用的结果。职业发展与现实焦虑类话题呈现出外部压力响应,紧随社会经济周期与政策波动而起伏;而学术合作与人际交往类话题则表现为“内源性渐进演化”,受教育制度变迁与个体成长周期的深层影响。

4　结论与启示

4.1　研究结论

基于知乎与百度贴吧平台的12 253条大学生社交文本数据,运用 BERTopic 主题模型、层次聚类与演化分析,对大学生社交圈群的主题特征与动态规律进行探究,得出如下结论。

(1)大学生社交话题呈现圈层化结构与多元化的层次谱系。通过 BERTopic 模型识别出的 9 个稳

定主题，并非孤立存在，而是依据其语义内涵，通过层次聚类自然聚拢为4大核心社交领域，包括职业发展社交、学习合作社交、人际交往社交与现实焦虑社交。该结构清晰地映射了大学生作为学习者与准社会人的双重身份。其中，职业发展与学习合作社交构成了其面向未来的前瞻性社交维度，而人际交往与现实焦虑则更多地反映了其应对当下学习生活的现实性社交维度。

(2)社交话题的演化规律呈现内生长周期与外生环境压力的双重驱动特征。时序演化分析表明，不同主题的演化路径存在显著差异。职业发展类话题与宏观经济社会环境紧密关联，表现出“应激响应”模式，热度随外部压力变化而显著波动。学术合作类话题则与学生的内在成长周期同步，呈现“渐进演化”特征，热度随年级提升与学术任务加重而稳步攀升。而现实焦虑类话题的波动则直接反映学生在经济与网络资源获取上的迫切需求。研究发现揭示了大学生的社交讨论是感知、回应并适应内外压力的重要窗口。

4.2 研究启示

基于上述结论，为有效引导大学生社交圈群文化，提升其社交质量与心理健康水平，从4个核心社交领域出发，提出如下治理启示。

(1)建立职业发展社交的动态预警与逆向干预机制。研究表明该类话题具有显著的应激响应特征，因此当“兼职、赚钱、裁员、学历贬值”等词云在校园圈层中异常波动时，高校不应仅发布常规通知，而应实施算法逆向引导，建立基于主题词频变化的焦虑指数监测预警系统，向话题高频圈层定向推送大厂校招复盘、基层就业政策、技能补丁课程等干预性内容，利用信息对冲打破焦虑自我强化的信息茧房，将应激状态转化为理性的求职准备。

(2)打造学习合作社交的去中心化学术共同体。针对学术压力下移导致的学术社交需求激增现象，学校应顺应圈层化特征。不应止步于入驻社群，而应通过学术领袖培育计划，选拔专业能力强、圈层影响力大的青年学生或辅导员担任学术合伙人，在知乎、贴吧等非正式社交场域领办微型学术社区。同时，建立线上协作学分认证体系，将网络平台自发形成的优质学术讨论、代码开源协作、笔记共享等非正式学习行为纳入素质学分考核，通过制度激励引导圈层社交向高阶学术治理转型。

(3)实施人际交往社交的“O2O场景修复”工程。针对宿舍关系、师生互动等恒常性社交需求，应利用圈层画像强化线下支持。通过大数据识别出长期沉溺于孤岛型社交圈层或表现出退缩社交特征的宿舍单元。在线下，通过重塑宿舍公共空间，将线上识别出的共同兴趣转化为线下的微型垂直社团。重点关注圈层讨论中暴露出的“寝室冷暴力”“师生权力不对等”等隐蔽问题，建立辅导员与学生圈层的数字热线，实现线上识别矛盾、线下化解纠纷的闭环。

(4)构建现实焦虑社交的数字资源补偿与就业防线。针对学生对低价流量、兼职补缺的刚性需求，高校应提供实物性支持。可与电信运营商合作提供学习专用免流包，并整合校内数字图书馆资源，开发一键触达的学术资源离线包，从根源上缓解因数字鸿沟产生的焦虑。针对兼职诈骗高发的痛点，由校内勤工俭学部门建立校内实名兼职任务流转平台，通过审核企业资质，将校内外可靠的勤工俭学岗位进行圈层化分发，直接推送至话题最活跃的兼职群组，用官方信誉的“硬链接”替换社交平台中来源不明的“软广告”，保障学生的经济安全与职业尊严。

参考文献

[1] 邱月. 圈层化网络社交生态下社交媒体用户话语建构及媒介审视[J]. 郑州大学学报(哲学社会科学版)，2025，58(1)：126-133，141.

[2] 杨华，赵景阳. 大学生网络社交圈层化的成因、隐忧与破解[J]. 学校党建与思想教育，2024(18)：79-81.

[3] 朱博，朱翠明. 透视青年“圈层化”文化现象[J]. 人民论坛，2023(19)：97-99.

[4] 邓春林，樊彦伶，邹凯，等. 圈层舆论极化下社交媒体用户信息偏食行为生成机理研究[J]. 情报理论与实践，2025，48(10)：63-72，100.

[5] 董晶，白芳睿，吴丹. 网络圈层化背景下社交媒体用户信息茧房滞留行为研究[J]. 情报理论与实践，2025，48(1)：140-149，167.

[6] YANG Y，ZHANG Y，JIANG L，et al. Importance of epistemic emotions and collaborative interactions in collective knowledge building：evidence from high-and low-performance groups[J]. Interactive Learning Environments，2025，33(7)：4309-4333.

[7] 郑苗，张家靖，刘向阳. 圈层化背景下高校大学生获得感的追寻、困境及引导[J]. 高教探索，2025(1)：123-128.

[8] 晏齐宏. 信息圈层化及其形成机制：基于ERGM的分析[J]. 湖南师范大学社会科学学报，2024，53(1)：111-124.

[9] 罗琳. 青年网络“圈层化”的时代特征、生成机制与风险防控[J]. 中国青年社会科学，2022，41(3)：75-83.

[10] CINELLI M，DE FRANCISCI MORALES G，GALEA-

ZZI A, et al. The echo chamber effect on social media [J]. Proceedings of the National Academy of Sciences, 2021, 118(9): e2023301118.

[11] HUI B P H, AU A K Y, NG J C K, et al. From social networking site use to subjective well-being: the interpersonal and intrapersonal mediating pathways of prosocial behavior among vocational college students in China [J]. International Journal of Environmental Research and Public Health, 2022, 20(1): 100.

[12] OKRAY Z, DIREKTÖR C, SIMSEK A H. Social network site usage among adults: social networking status and virtual loneliness[J]. European Journal of Psychology and Educational Research, 2018, 1(1): 11-18.

[13] 黄佳佳, 李鹏伟, 彭敏, 等. 基于深度学习的主题模型研究[J]. 计算机学报, 2020, 43(5): 827-855.

[14] CHAGNON E, PANDOLFI R, DONATELLI J, et al. Benchmarking topic models on scientific articles using BERTeley[J]. Natural Language Processing Journal, 2024, 6: 100044.

[15] 曹树金, 曹茹烨. 情报学论文创新性评价研究: LDA 和 SVM 融合方法的应用[J]. 图书情报知识, 2022, 39(4): 56-67.

[16] 陈淑一, 朱华桂. 突发事件处置背景下的高校管理锚点探析: 基于疫情防控形势下的 N 大学新闻网 LDA 主题挖掘[J]. 江苏高教, 2023(5): 67-71.

[17] DAI C, WANG L, DENG Y, et al. LDA-SCGB: inferring lncRNA-disease associations based on condensed gradient boosting[J]. BMC Bioinformatics, 2025, 26(1): 1-27.

[18] 孟薇, 王晟聪. 拜登政府时期美国智库涉华研究的注意力变化及其逻辑研究: 基于兰德公司 229 份报告的文本分析[J]. 赤峰学院学报(汉文哲学社会科学版), 2025, 46(3): 40-45.

[19] GROOTENDORST M. BERTopic: neural topic modeling with a class-based TF-IDF procedure[J]. arXiv preprint arXiv: 2203. 05794, 2022.

[20] 冯立杰, 秦浩, 王金凤, 等. 融合专利数据与社交媒体数据的潜在颠覆性技术识别: 基于深度学习模型[J]. 情报学报, 2024, 43(2): 181-197.

[21] 王益成, 蒋星宇, 郑彦宁. 基于 BERTopic 模型的科技报告主题挖掘与演化分析: 以生物技术领域为例[J]. 情报科学, 2024, 42(9): 51-60.

[22] 苗宏慧, 全情爽, 舒心. 基于 BERTopic 的科技人才政策文本主题识别与量化分析: 以东北三省为例[J]. 现代情报, 2025, 45(4): 110-121.

[23] 毛太田, 郑锐. 基于 BERTopic 的中国人形机器人专利主题识别与演化分析[J]. 科技和产业, 2025, 25(23): 131-138.

[24] 李豪, 张柏苑, 邵蝶语, 等. 融合 BERTopic 和 Prompt 的学者研究兴趣生成模型: 以计算机科学领域为例[J]. 情报科学, 2025, 43(1): 127-136, 160.

[25] CAMPELLO R J G B, MOULAVI D, SANDER J. Density-based clustering based on hierarchical density estimates[C]//Advances in Knowledge Discovery and Data Mining: 17th Pacific-Asia Conference, PAKDD 2013. Heidelberg: Springer Berlin Heidelberg, 2013: 160-172.

Analysis of Classification and Evolution of College Students' Social Interaction from Perspective of Online Network Layering: Modeling Based on BERTopic Topic

YIN Huilin, ZHENG Yiheng, ZHOU Huan

(School of Economics and Management, Hunan University of Technology, Zhuzhou 412007, China)

Abstract: With the advancement of the networked era, college students' social interactions are deeply influenced by the internet, and 'college student circles' have become a common social phenomenon on campuses. Python was used to collect texts on the topics of college student socialization from the platforms Zhihu and Baidu Tieba as data sources, and BERTopic topic modeling was applied to identify the thematic distribution of college student social interactions and extract their topic features, and trends of theme evolution over different time series were analyzed. It is found that college students' social topics exhibit significant diversification and stratification, with core themes categorized into four social areas: interpersonal socialization, career development socialization, collaborative learning socialization, and reality anxiety socialization. Based on the evolutionary analysis, college students' core needs and mindset changes follow regular patterns at different stages. The BERTopic model was applied to the dynamic topic evolution analysis in the field of college student socialization, effectively capturing the fine-grained features and temporal patterns of topics within contextual circles. The suggestion is put forward to further explore the relationship between topic evolution and specific social events by integrating multimodal data.

Keywords: college student social interactions; stratification; BERTopic model; topic features; topic evolution