

基于LDA模型的中国大学生心理韧性研究主题识别与演进分析

黄 靖, 康文馨

江汉大学教育学院, 湖北 武汉

收稿日期: 2026年7月13日; 录用日期: 2026年8月12日; 发布日期: 2026年8月24日

摘 要

该研究基于潜在狄利克雷分配(LDA)主题模型, 对中国知网(CNKI)数据库中618篇关于中国大学生心理韧性的期刊文献进行了计量分析。研究旨在系统梳理该领域的核心研究主题与演化路径。通过使用Python进行文本预处理和主题挖掘, 研究识别出八个核心主题, 包括群体心理健康测评、压力挫折与韧性培育、主观幸福感等。研究发现, 该领域的研究范式经历了从“量表测量”到“机制挖掘”再到“实践应用”的转型。主题强度分析显示, “压力挫折与心理韧性培育”是当前热点, 而“留守大学生”相关研究热度持续下降。主题演化分析揭示了“社会支持-心理韧性-主观幸福感”这一经典作用链条的形成与巩固。该研究通过数据驱动的方法, 为理解大学生心理韧性研究的全貌提供了可量化的知识图谱, 并为未来的科研选题和高校心理健康教育提供参考。

关键词

大学生, 心理韧性, LDA主题模型, 研究趋势, 主题演化

Research on Psychological Resilience of Chinese College Students Based on LDA Model: Topic Identification and Evolution Analysis

Jing Huang, Wenxin Kang

College of Education, Jiangnan University, Wuhan Hubei

Received: July 13, 2026; accepted: August 12, 2026; published: August 24, 2026

Abstract

This study employs a Latent Dirichlet Allocation (LDA) topic model to conduct a bibliometric analysis of 618 journal articles on Chinese college students' psychological resilience, retrieved from the China National Knowledge Infrastructure (CNKI) database. The research aims to systematically delineate the core research themes and evolutionary trajectories within this field. Through text pre-processing and topic mining using Python, eight focal themes were identified, including group-level mental health assessment, stress-adversity and resilience cultivation, and subjective well-being. Findings indicate that the research paradigm in this domain has undergone a transition from "scale-based measurement" to "mechanism exploration" and subsequently to "practical application." Topic intensity analysis reveals that "stress-adversity and resilience cultivation" constitutes the current research hotspot, whereas studies pertaining to "left-behind college students" have exhibited a persistent decline in popularity. The thematic evolution analysis further uncovers the formation and consolidation of the classic mediational chain—"social support → psychological resilience → subjective well-being." By adopting a data-driven approach, this study furnishes a quantifiable knowledge map for comprehending the holistic landscape of psychological resilience research among college students, and offers insights for future research topic selection as well as university-based mental health education interventions.

Keywords

College Students, Psychological Resilience, LDA Topic Model, Research Trends, Topic Evolution

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

心理韧性(Resilience)又称心理弹性、复原力,是指个体面对逆境、创伤、威胁或其他重大压力时能迅速恢复和成功应对的良好适应过程,强调个体在压力情境中的适应性与恢复力(胡月琴, 甘怡群, 2008)。对于大学生来说,他们正处于身心发展的关键阶段,既要应对课程考核、升学考研等学业压力,又要处理人际、亲密关系等社交矛盾,同时还要直面求职就业、网络沉迷、信息过载等数字化时代特有困境,多重压力叠加下,抑郁、焦虑、情绪失调等心理适应问题日益突出。大学生的心理健康问题日益严峻(赵玉芳, 左有霞, 2024)。心理韧性的补偿模型认为,心理韧性作为一种心理健康的保护性因素,能够缓冲威胁性因素给个体带来的负面影响(史璐等, 2023)。心理韧性水平不仅关乎个体的心理健康与成长,也影响着高等教育人才培养的质量与效能。一项横断历史 meta 分析研究显示,我国大学生心理韧性水平整体呈下降趋势(史璐等, 2023),这反映了当代大学生心理健康干预的紧迫需求,也凸显了深化心理韧性研究的必要性。

关于中国大学生心理韧性研究已有的综述工作,大多采用传统文献梳理和 CiteSpace 可视化分析等方法,从内涵结构、影响因素、作用机制等横向维度进行归纳总结,这为理解该领域的研究格局提供了重要的基础性贡献(张妍妍, 赵帅, 2024)。如今大数据时代也为社会科学研究带来了范式层面的拓展机遇。潜在狄利克雷分配模型(Latent Dirichlet Allocation, LDA)作为一种三层贝叶斯概率主题模型,由 Blei 等人于 2003 年提出,通过构建“文档-主题-词汇”的生成式结构,能够在不预设主题的前提下自动发现大

规模文本语料中的隐藏主题分布(Blei et al., 2003)。该方法已在图书情报、计算机科学、管理科学等领域得到广泛应用。基于此, 本研究以近二十年中国大学生心理韧性领域的中文期刊文献为语料, 运用 LDA 主题模型对该领域的研究主题进行系统识别与演化分析, 通过数据驱动的方法, 弥补传统综述在主题演化追踪上的局限, 为该领域的后续研究提供可量化、可验证的知识图谱, 为大学生心理健康教育工作提供更多思路。

2. 研究设计

2.1. 数据来源

数据收集来源于中国知网(CNKI)数据库, 以“大学生 AND 心理韧性”为主题词进行检索, 检索范围为期刊文献, 以 Excel 格式自定义导出每篇文献的题目、年份、关键词、摘要等信息, 对导出结果进行人工筛选, 剔除与大学生、心理韧性主题无关、摘要残缺、征稿通知、书评、会议转载等非有效研究文献, 最终得到符合条件的 618 篇实证与理论类期刊文献。按年份统计每年的文献发文量, 得到大学生心理韧性的发文数量趋势, 如图 1 所示。

根据文献的发文量数据, 将大学生心理韧性研究划分为三个阶段。2008~2015 年年度文献量从个位数增长至 20 余篇, 年均增速平稳, 属于领域“从无到有”的初步探索期; 2016 年发文量首次突破 40 篇, 此后八年稳定在 30~60 篇之间, 研究体系日趋成熟, 2016~2023 年属于快速发展期; 2024 年起发文量突破 70 篇, 2025 年达到 149 篇的历史峰值, 研究进入爆发式增长阶段, 属于深化应用期。

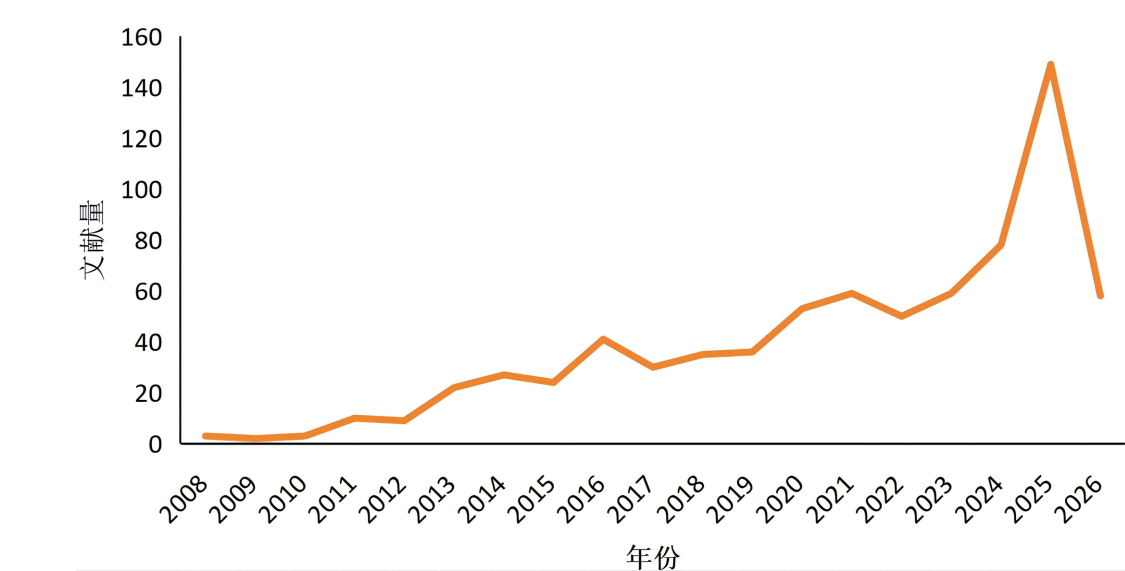


Figure 1. Publication volume of research on psychological resilience in college students
图 1. 大学生心理韧性发文量

2.2. 研究方法

采用潜在狄利克雷分配(LDA)主题模型分析法对我国大学生心理韧性研究文献进行热点主题识别与演进分析, 首先使用 Python 语言对收集到的期刊文献进行文本预处理, 包括中文分词、停用词过滤以及同义归一等操作; 其次构建 LDA 主题模型, 通过计算困惑度和主题一致性指标确定最优主题数量, 归纳出该领域的核心研究主题; 然后通过计算各主题在不同年份的强度值来识别热点主题; 最后绘制词云图和桑基图来展示我国大学生心理韧性研究主题的发展脉络与演进路径, 研究框架如下(图 2)。

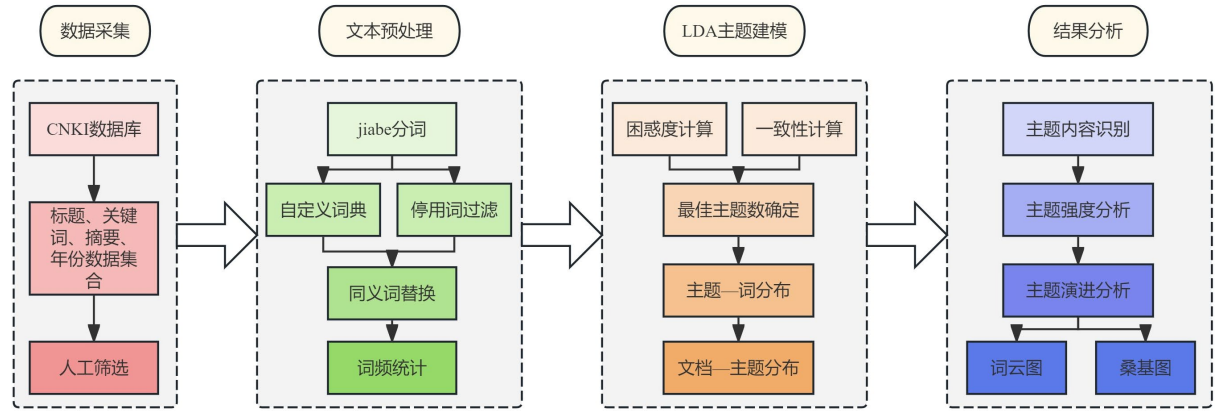


Figure 2. Research framework diagram
图 2. 研究框架图

2.2.1. 文本预处理

采用 Python 中的 jieba 分词工具对每篇文献的题目和摘要文本进行分词处理。为提升分词准确性，加载自定义词典，包含“心理韧性”“社会支持”“应对方式”“中介效应”“自我效能感”等专业术语，同时屏蔽某些无意义的单字或符号。停用词表采用哈尔滨工业大学和百度中文停用词表的结合，并依据本研究语料特点进行人工增补，移除诸如“我们”“他们”“这个”“那个”以及“研究”“发现”“表明”“本文”等与研究主题无直接关联的高频功能词。将同义但书写形式不同的词汇进行归一化处理，例如将“中介效应”“中介作用”“中介机制”统一为“中介效应”，将“复原力”“心理弹性”“抗逆力”统一为“心理韧性”。剔除词频低于 5 次且文档频率低于 3 篇的罕见词，以降低模型噪声。

2.2.2. LDA 主题模型构建

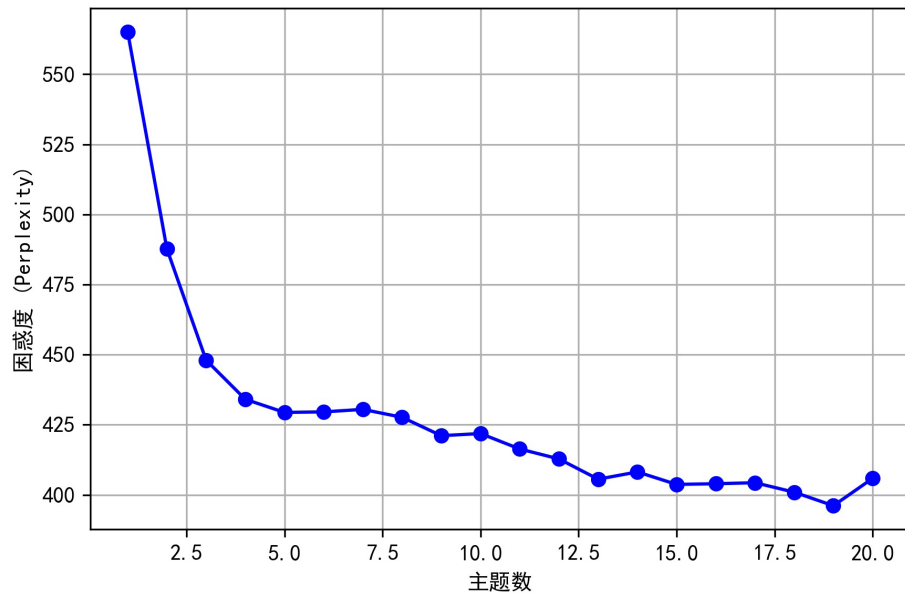


Figure 3. Perplexity results
图 3. 困惑度计算结果

LDA 模型通过概率推断实现对文本语料潜在主题结构的识别与抽取。调用 Python 3.2 中 sklearn 库进

行 LDA 主题模型构建, 设置参数为默认参数 $\alpha = 0.1$, $\beta = 0.01$, 迭代次数为 50 (邱均平, 沈超, 2021)。设定主题数量 K 的取值范围为 1 至 20, 步长为 1, 分别计算主题困惑度和主题一致性, 并结合可视化图形来确定最佳主题数量。困惑度是衡量模型对未见文档的预测能力, 困惑度越低表示模型的泛化性能越好; 主题一致性是衡量同一主题下高概率词项之间的语义一致程度, 使用 C_v 指标, 数值越高表明主题越清晰(吴炯炯, 2023)。

如图 3 所示, 当主题数量为 1~5 时, 困惑度大幅下降, 当主题数达到 8 之后, 曲线下降斜率显著放缓, 趋于平缓; 如图 4 所示, 当主题数为 8 时, 一致性曲线出现一处明显的局部峰值, 相比邻近 $K = 7$ 、 $K = 9$ 的一致性数值更高; 结合 LDAvis 可视化图形, 8 个主题的空间分布清晰, 重叠区域极少, 各主题圈层边界分明, 不存在大面积融合、语义高度重合的主题。最终确定最优主题数为 8, 针对最优的主题数量, 调整 α 和 β 的取值, 进一步提升模型性能。

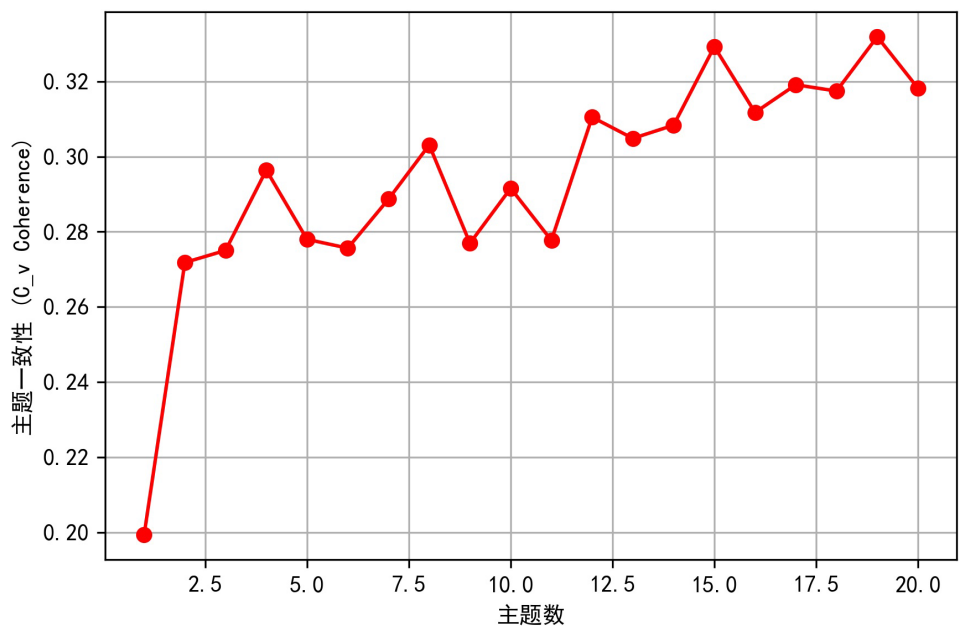


Figure 4. Topic coherence results
图 4. 主题一致性计算结果

3. 结果与分析

3.1. 主题内容分析

根据 LDA 主题建模结果, 识别出 8 个研究主题, 为缓解 LDA 主题人工命名过程中存在的主观性偏差, 本研究采取多重校验策略: 首先依托模型输出的主题 - 词汇概率分布提取各主题前 10 个高频特征词; 其次结合各主题下代表性文献摘要语义进行交叉比对; 同时参照领域已有相关文献对同类主题的命名范式综合归纳, 最终确定各主题名称, 主题 - 特征词分布如表 1 所示。

结合可视化图(图 5)整体来看, 主观幸福感(Topic 4)为独有高权重变量, 区别于其他主题侧重压力、抑郁、干预的视角, 以积极情绪、幸福感为核心结果变量, 形成差异化研究脉络; Topic 1 侧重群体心理健康测评, Topic 3 侧重行为、情绪预测, 都落在大学生基础心理现状测量, 研究范式高度相近, 因此文本语义重叠度最高; Topic 6 和 Topic 8 均涉及特定细分人群与心理干预, 二者空间位置紧邻; Topic 2、Topic 5 和 Topic 7 语义相近, 均围绕压力、负性情绪、外部支持开展机制分析。

Table 1. Topic-term distribution of the literature
表 1. 文献主题 - 特征词分布

主题编号	主题标识	前 10 个高频特征词
Topic 1	群体心理健康测评	心理健康(0.085)、自我(0.079)、量表(0.029)、幸福感(0.024)、医学生(0.023)、高职(0.021)、评价(0.019)、能力(0.019)、效能(0.018)、管理(0.017)
Topic 2	压力挫折与心理韧性培育	心理健康(0.039)、压力(0.019)、培育(0.017)、能力(0.016)、策略(0.012)、挫折(0.011)、支持(0.011)、协同(0.011)、家庭(0.011)、工作(0.011)
Topic 3	运动行为与负性情绪预测	量表(0.054)、体育锻炼(0.044)、压力(0.043)、依恋(0.024)、预测(0.022)、抑郁(0.022)、生活(0.022)、事件(0.020)、情绪(0.019)、手机成瘾(0.018)
Topic 4	主观幸福感的影响	主观幸福感(0.092)、情绪(0.025)、能力(0.024)、症状(0.024)、家庭(0.022)、依赖(0.020)、人际交往(0.019)、手机(0.017)、积极情绪(0.017)、网络(0.015)
Topic 5	社会支持的保护机制	社会支持(0.097)、抑郁(0.062)、量表(0.056)、自我效能感(0.032)、预测(0.029)、自杀意念(0.027)、正念(0.026)、学业(0.022)、人际关系(0.020)、正向(0.017)
Topic 6	压力应对方式	量表(0.056)、焦虑(0.052)、应对方式(0.044)、领悟社会支持(0.037)、护生(0.032)、自尊(0.027)、情绪智力(0.026)、积极应对(0.023)、方式(0.021)、问卷(0.021)
Topic 7	留守大学生心理特征差异	留守经历(0.058)、得分(0.048)、维度(0.040)、因子(0.036)、量表(0.035)、家庭(0.031)、女生(0.020)、积极认知(0.020)、预测(0.020)、男生(0.019)
Topic 8	积极心理学干预研究	对照组(0.035)、创业(0.031)、就业(0.026)、心理资本(0.020)、心理健康(0.018)、情绪(0.017)、实验组(0.017)、效果(0.015)、积极心理学(0.013)、团体(0.013)

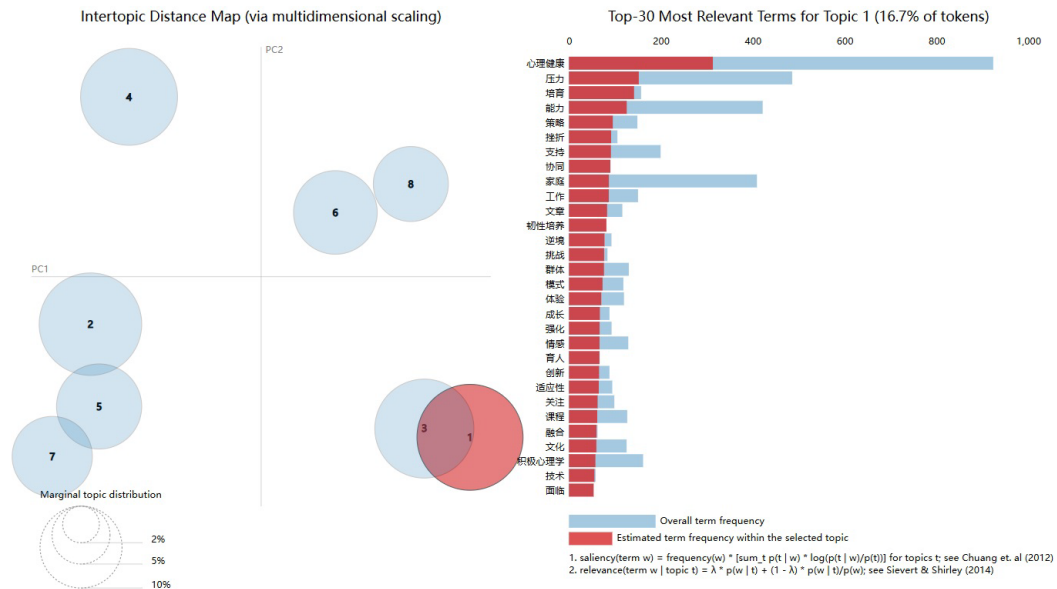


Figure 5. LDA model visualization
图 5. LDA 模型可视化

3.2. 主题强度分析

主题强度表示各主题在各年份的相对重要性。如图 6 所示, Topic 2 (压力挫折与心理韧性培育)是全图增长最显著的主题,是第三阶段强度最高的主题,说明面向大学生压力挫折的韧性培育是当下领域的研究热点,学界研究资源持续向实践培育类方向倾斜。Topic 3 (运动行为与负性情绪预测)和 Topic8 (积极心理学干预研究)的主题强度总体来说保持较高的趋势,比较平稳,波动较小,代表体育锻炼、压力、抑郁情绪、积极心理学、团体干预研究的机制研究是贯穿近 20 年的稳定基础主线,是领域长期重要的干预研究分支。Topic 4 (主观幸福感的影响)三阶段持续缓慢增长,前期热度最低,近年稳步回升,积极情绪、主观幸福感的保护机制研究长期稳步发展,在深化应用期热度逐步回暖。Topic 7 (留守大学生心理特征差异)呈现出持续衰退的状态,说明留守大学生专项横断面调查是早期本土热点,但近十年研究热度持续衰退,一方面,随着国家精准扶贫战略与乡村振兴的全面推进,绝对意义上的留守儿童大幅减少,因此留守大学生这个群体规模随之减少(安俊, 王志敏, 2018);另一方面,可能是国内大学生心理健康研究正从聚焦特定边缘群体的病理性,向关注全体学生积极心理品质发展发生转变(马颖哲等, 2016)。研究者不再单独聚焦留守群体,仅将留守背景作为调节变量融入其他模型。

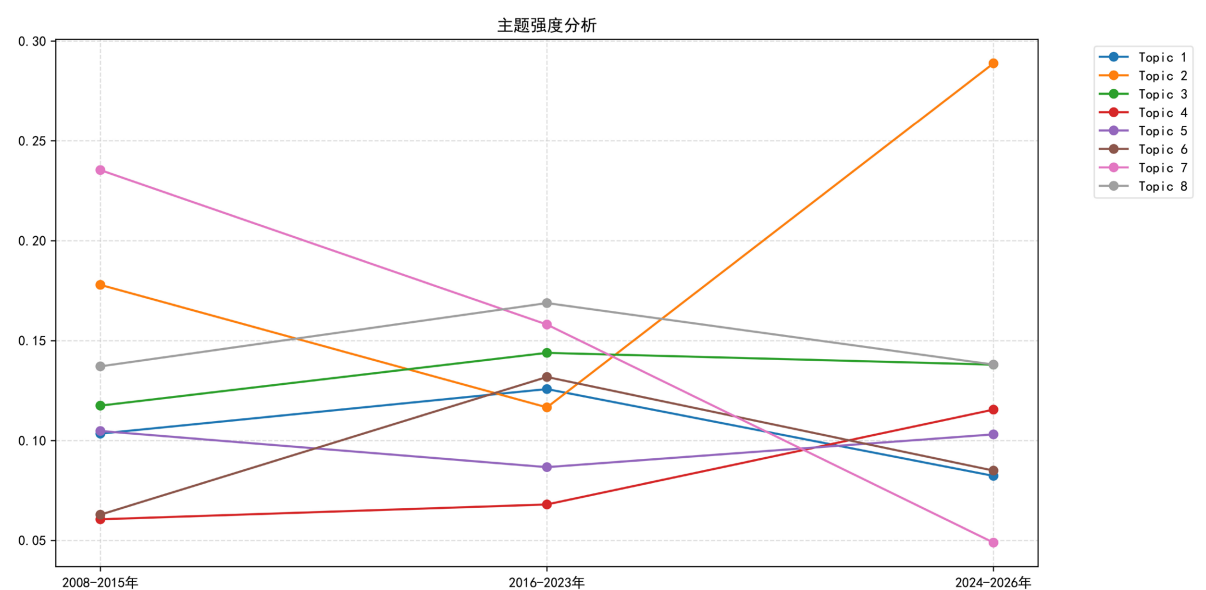


Figure 6. Topic strength analysis of the literature
图 6. 文献主题强度分析

3.3. 主题演进分析

词云图是将文本高频词汇可视化的图形工具,通过词语字体大小区分出现频率,字体越大代表该词汇在对应阶段文献中出现次数越多,能够快速反映不同阶段的研究核心与热点方向(包琛, 汪云海, 2021)。

初步探索期(2008~2015)词云结果显示(见图 7),“心理韧性”“社会支持”“量表”“维度”为领域核心词汇,“留守经历”“医学生”“高职”等本土特殊群体词汇占比突出。该阶段研究以引进修订韧性量表、开展横断面问卷调查为主要范式,侧重检验社会支持、家庭教养、压力等变量与大学生心理韧性的简单相关关系,关注留守、医学高职等群体韧性得分差异,但缺乏复杂作用机制分析与干预实践探索。

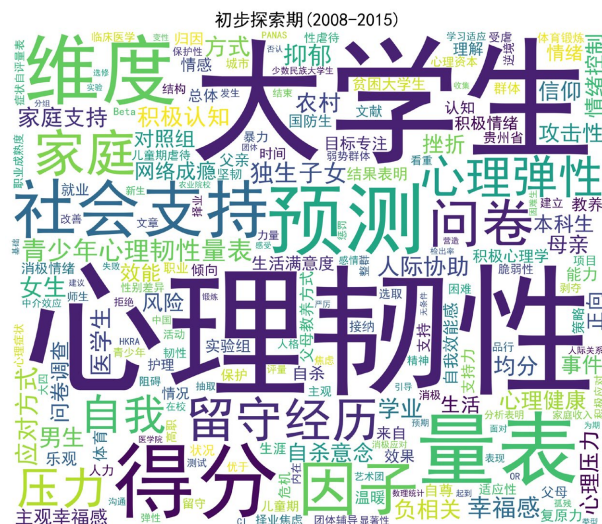


Figure 7. Word cloud of the literature in the initial exploration period
图 7. 初步探索期的文献主题词云图

快速发展期(2016~2023)词云结果显示(见图 8), “心理韧性” “量表” “情绪” “心理健康” “压力” 为高频词汇, “中介效应” “Bootstrap” “护生” “手机成瘾” “主观幸福感” “PTSD” “实验组” 等新词大量涌现。相较于初步探索期, 该阶段研究范式实现明显升级, 中介、链式中介、调节等复杂作用机制成为主流分析思路; 研究群体进一步细分, 护生作为高压专业群体成为重点研究对象, 留守经历仍持续受到关注; 同时叠加手机网络成瘾等数字化、公共卫生时代特有压力源, 积极心理学视角快速兴起。此外“实验组” “对照组” “团体辅导” 等词汇表明, 准实验类干预实证研究开始批量出现, 领域从单纯现状测量转向机制挖掘与初步干预探索并行。



Figure 8. Word cloud of the literature in the rapid development period
图 8. 快速发展期的文献主题词云图

深化应用期(2024~2026)词云结果显示(见图 9), “心理健康” “焦虑” “抑郁” “社会支持” 仍保持核心地位, “协同” “育人” “培育” “文化” “体育锻炼” “就业创业” “链式中介” 为阶段独有高频

5 (社会支持的保护机制)、Topic 6 (压力应对方式)、Topic 7 (留守大学生心理特征差异)及 Topic 8 (积极心理学干预研究)的大量流入。这一变化标志着研究视角发生从“问题导向”到“积极取向”的范式转型。在阶段 3, Topic 4 保持较强的自我循环能力,同时向 Topic 2 输出显著流量,这表明,主观幸福感研究在经历快速扩张后,逐渐与韧性培育研究形成有机整合。

Topic 5 (社会支持的保护机制),在阶段 1 仅做简单相关分析的社会支持研究,而阶段 2 重新形成的 Topic 5,生成了宽厚流线并完整流向阶段 3 的 Topic 5,与 Topic 2 (压力挫折与韧性培育)、Topic 4 (主观幸福感的影响)之间形成稳定的三角循环,验证了“社会支持→心理韧性→主观幸福感”这一经典保护性因素链(程子良等, 2026),说明当前研究关注外部支持系统与个体韧性的交互效应。

Topic 6 (压力应对方式)主要流向 Topic 2 (压力挫折与韧性培育),部分流向 Topic 8 (积极心理学干预),说明应对方式研究正与积极心理干预技术结合。Topic 8 (积极心理学干预研究),向 Topic 2 (韧性培育)、Topic 3 (运动行为)、Topic 7 (留守群体)均有输出,显示积极心理学干预作为一种通用技术框架被广泛应用于各细分领域。Topic 1 (群体心理健康测评),已从“量表开发与测评”转向“机制探究与干预实践”,测量学主题逐渐融入具体研究而非独立存在。

Topic 3 (运动行为与负性情绪预测)总体保持稳定,且自我循环效应较强,接收多个方面的主题,说明运动干预作为具身化的韧性提升路径,既不被其他主题淹没,也未出现爆发式增长,形成了一条平行于认知干预、社会支持干预的行为干预研究脉络。

Topic 7 (留守群体)主题流线持续收缩衰退,学界首先关注最显性的脆弱群体,留守大学生因其特殊的成长经历成为心理韧性研究的切入点。后期不再将留守背景作为独立核心研究对象,仅作为调节变量嵌入其他机制模型。

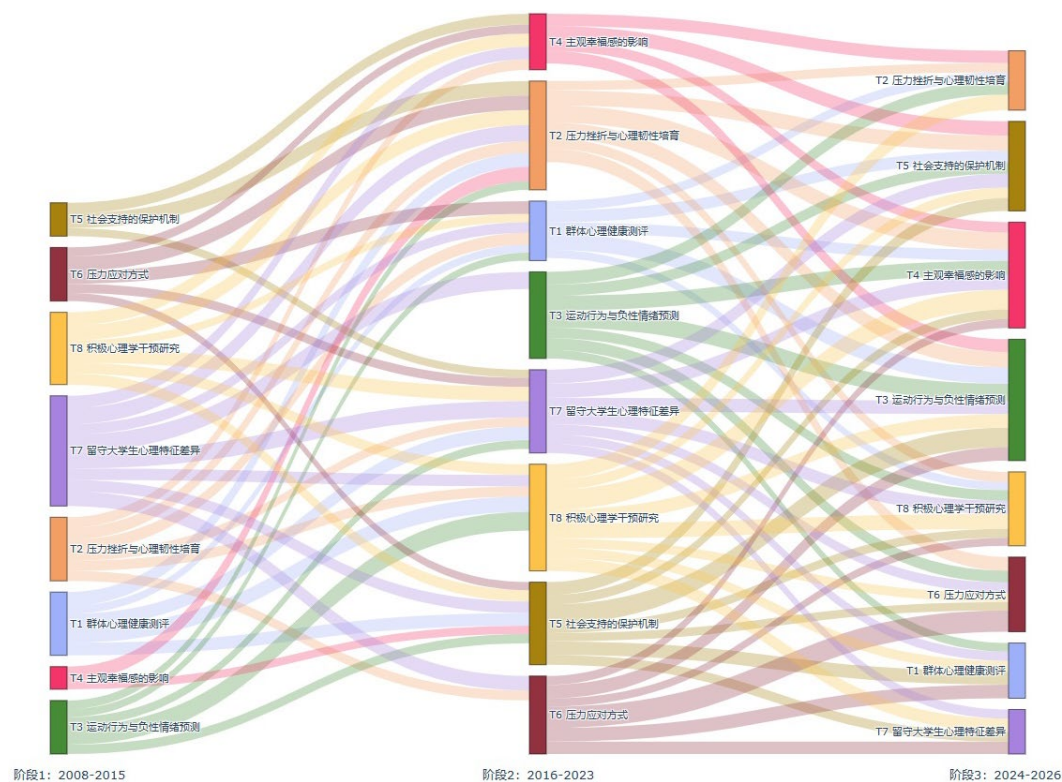


Figure 10. Sankey diagram of topic evolution
图 10. 主题演进桑基图

4. 结论

本研究通过 LDA 模型识别出国内大学生心理韧性领域八大核心主题, 并揭示出研究范式从“量表测量→机制挖掘→实践应用”的清晰转型轨迹。主题强度与演化分析表明, 压力挫折与韧性培育(Topic 2)已成为当前最热方向, “社会支持-心理韧性-主观幸福感”(Topic 5→Topic 2→Topic 4)的经典作用链条稳固形成, 积极心理学干预(Topic 8)与运动行为预测(Topic 3)作为稳定分支贯穿始终, 而留守群体专项研究(Topic 7)则持续衰退, 整体呈现“早期全面铺开、中期多元分化、后期向应用汇聚”的演进特征。

上述结果不仅勾勒出领域知识图谱, 也为未来研究指明了可突破的方向。其一, 鉴于 Topic 2 (压力挫折与韧性培育)与 Topic 8 (积极心理学干预)已成为热点, 而现有干预研究多依赖现实人际支持(Topic 5), 在生成式人工智能深刻改变青年社交生态的当下, 未来应突破传统社会支持框架, 重点探索大语言模型(LLMs)、AI 心理陪伴智能体等在大学生压力缓冲与韧性即时干预中的作用机制, 推动“AI+ 心理”数字化韧性赋能的新范式。其二, 针对 Topic 1 (群体心理健康测评)与 Topic 3 (运动行为与负性情绪预测)长期依赖横截面量表与静态中介模型的局限, 未来应结合可穿戴设备与数字表型技术, 捕捉大学生在学业挫败、就业受挫等真实困境下的心理韧性动态波动轨迹, 借助生态瞬时评估与机器学习算法构建个体化预警模型(徐冲等, 2024; 杨丽等, 2023), 从而将静态机制研究升级为动态精准干预。

此外, 本研究仍存在若干局限性, 需在后续研究中加以改进。首先本研究仅选取知网期刊文献, 数据来源较为单一, 且未对核心期刊与普通期刊开展分组对比分析, 可能导致研究主题的覆盖范围受限。其次, LDA 主题模型本身蕴含多重主观决策环节, 例如最优主题数 K 的确定虽综合了困惑度、主题一致性指标与可视化校验, 但不同指标给出的最优值并非完全一致, 最终选择 8 个主题仍依赖研究者的判断; 主题命名过程虽借助特征词排序与代表性文献摘要交叉验证, 但命名标签的凝练仍不可避免地带有主观解读成分等。为增强结论的稳健性, 后续研究可扩充多源数据, 包括国内外期刊, 进行跨库比较; 引入动态主题演化模型, 捕捉更细粒度的主题演变轨迹, 并减少对预设参数的人工依赖; 采用多研究者独立命名与一致性检验, 或利用外部专家评审提升主题标签的效度, 为大学生心理韧性领域的量化综述提供更可靠的方法论基础。

参考文献

- 安俊, 王志敏(2018). 新时代农村留守儿童关爱与教育扶贫三维体系构建. *新课程(中)*, (4), 26-27.
- 包琛, 汪云海(2021). 词云可视化综述. *计算机辅助设计与图形学学报*, 33(4), 532-544.
- 曹婧, 曹雯, 梁利苹(2025). 家校协同对大学生心理韧性的影响研究. *科教导刊*, (24), 101-104.
- 程子良, 廖传景, 胡嘉欢, 徐敏, 翁添富(2026). 领悟社会支持对贫困大学生主观幸福感的影响: 自尊与心理韧性的链式中介作用. *黑龙江科学*, 17(1), 63-67.
- 胡月琴, 甘怡群(2008). 青少年心理韧性量表的编制和效度验证. *心理学报*, (8), 902-912.
- 马颖哲, 靳小川, 叶静(2016). 积极心理学视域下大学生心理健康教育模式的构建. *教育教学论坛*, (49), 45-46.
- 邱均平, 沈超(2021). 基于 LDA 模型的国内大数据研究热点主题分析. *现代情报*, 41(9), 22-31.
- 史璐, 焦姣, 牡丹(2023). 中国大学生心理韧性水平变迁的横断历史研究. *中国心理卫生杂志*, 37(10), 894-898.
- 吴帼帼(2023). 近五年信息素养主题演化及影响力分析. *情报工程*, 9(1), 92-101.
- 吴婷婷, 黄雅杰, 张哲(2026). 大学生心理韧性: 理论脉络、整合性概念框架与育人启示研究. *社会与公益*, (8), 273-275.
- 徐冲, 张尚, 赵丹, 杨丽, 白苏好(2024). 基于生态瞬时评估法探索自杀意念对情绪的影响. *中国临床心理学杂志*, 32(4), 837-842.
- 杨丽, 张强, 白苏好, 周玥, 安莉(2023). 使用机器学习算法预测大学生自杀尝试风险. *中国临床心理学杂志*, 31(3), 525-529+634.

张妍妍, 赵帅(2024). 大学生心理韧性研究综述及展望. *高教论坛*, (2), 120-124.

赵玉芳, 左有霞(2024). 新时代大学生心理健康监测体系构建与实践进路. *中国高等教育*, (23), 60-64.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3, 993-1022.