

# 基于 BERTopic 模型和 Louvain 算法的颠覆性技术主题演化研究——以新能源汽车产业领域为例<sup>1</sup>

吕鲲<sup>1\*</sup> 傅元美<sup>1</sup> 鲍泽孟鸿<sup>2</sup>

1. 宁波大学商学院 宁波 315211

2. 上海大学文化遗产与信息管理学院 上海 200444

**摘要：**颠覆性技术是关乎国家竞争力和国际地位的重大课题，精准识别出颠覆性技术主题并明确其发展脉络，能够为推动颠覆性技术发展的超前部署提供坚实的理论基础，有利于快速抢占技术发展先机。以技术、市场、政策为数据源，获取专利、市场报告、公开政策数据，根据技术生命周期理论划分时间窗口；基于 BERTopic 主题模型并引入聚类评价指标和 K-means 算法，优化聚类效果，结合颠覆性技术特征指标，识别颠覆性技术主题；通过构建主题共现网络展开主题关联横向演化，再借助 Louvain 社区发现算法实现多源主题融合，并构建多源融合主题演化网络展开主题社区纵向演化，全面分析颠覆性技术主题演化过程。以新能源汽车产业领域为例进行实证分析，结合政策解读验证了该方法的可行性。

**关键词：**BERTopic；颠覆性技术识别；技术主题演化；新能源汽车产业；多源数据融合

**分类号：**U491.8; G353.1

## Study on the Evolution of Disruptive Technology Themes Based on BERTopic Model and Louvain Algorithm: A Case Study in the New Energy Vehicle Industry

Kun Lv<sup>1\*</sup>, Yuanmei Fu<sup>1</sup>, Zemenghong Bao<sup>2</sup>

1. Business School, Ningbo University, 315211, China

2. School of Cultural Heritage and Information Management, Shanghai  
University, Shanghai 200444, China

**Abstract:** Disruptive technology is a critical subject related to national competitiveness and international standing. Accurately identifying disruptive technology themes and clarifying their developmental trajectories can provide a robust theoretical foundation for the proactive advancement of these technologies, facilitating the swift capture of technological opportunities. This paper utilizes data from technology, market, and policy sources to gather patents, market reports,

<sup>1</sup> 本文系国家社会科学基金青年项目“‘双碳’目标下‘技术-经济-区域’信息融合的创新生态系统构建及其协同演化研究”（编号：22CTQ028）研究成果。

\* 通讯作者邮箱：lvkun@nbu.edu.cn

and public policy data, dividing time windows according to the technology lifecycle theory. Based on the BERTopic topic model, the clustering evaluation index and K-means algorithm are introduced to optimize the clustering effect, and the disruptive technology theme is identified by combining the disruptive technology characteristic index. We construct a thematic co-occurrence network to explore horizontal thematic evolution, and by leveraging the Louvain community detection algorithm, we achieve multi-source theme integration, constructing a multi-source integrated theme evolution network for a comprehensive analysis of the evolution process of disruptive technology themes. An empirical analysis is conducted in the new energy vehicle industry, and the feasibility of the method is verified through policy interpretation.

**Keywords:** BERTopic; disruptive technology identification; technological theme evolution; new energy vehicle industry; multi-source data integration

## 1 引言

近年来,科技革命和产业变革的浪潮不断涌现,重塑着全球科技竞争的新格局。在此过程中,颠覆性技术创新逐渐占据国家竞争力和国际地位提升的重要战略地位。党的十九大强调要建设世界科技强国,就一定要锻造以自主创新为利刃的颠覆性技术创新之剑。在此背景下,颠覆性技术成为科技创新领域研究的前沿热点。

然而,当前我国在颠覆性技术发展上面临着诸多挑战,包括主题的精准识别、发展动态的有效追踪以及未来趋势的科学预测,这些问题直接关联产业的未来竞争力与国际地位。尤其在全球环境、社会和治理(ESG)标准日益严格的当下,新能源汽车产业作为绿色转型的先锋,其颠覆性技术发展中的难题,例如电池能量密度与安全性同步提升、推动氢燃料电池商业化等,更是引起了广泛关注。因此,精准识别关键产业领域的颠覆性技术主题并分析其发展动态和未来趋势,对于科技前沿的突破至关重要。

为了解决这些问题,前人在颠覆性技术主题识别和演化方法的研究上有了极大推进,但受限于单一数据源信息,难以全方位呈现颠覆性技术的动态发展轨迹。鉴于此,整合多源数据进行深度挖掘与分析,准确识别颠覆性技术主题,并探究其发展动态,是科研人员全面把握颠覆性技术发展脉络和未来趋势的关键所在,更是精准定位技术突破口、加速技术创新的重要手段,能够为国家科技创新的战略布局提供坚实的理论基础与实证依据。

## 2 文献综述

颠覆性技术概念最早是由美国哈佛商学院教授 Christensen[1]提出的,他认为颠覆性技术是一种新兴技术,起初在性能上不如现有技术,但却能通过降低成本、提高便利性等方式,逐渐获得市场认可,最终取代现有技术。随着科技的快速发展,颠覆性技术逐渐成为学术界关注的焦点。学者们纷纷对颠覆性技术展开研究,现有研究方向包括颠覆性技术概念辨析、

特征分析、识别方法、演化路径、预测分析等。

## 2.1 颠覆性技术识别方法

目前对于颠覆性技术的识别方法主要分为两类，即定性分析法和定量分析法。定性分析法，如德尔菲法[2]、技术路线图法[3]、情景分析法[4]等，依赖专家知识和经验，能够输出可理解性较高的分析结论，在早期研究中广泛应用，但存在主观性强、过程周期长、投入成本高、适用程度低等局限性[5]。定量分析法，主要采用模型分析法、文献计量法、指标识别法等，追求量化分析的客观性，在数据分析中展现出强大的效能，其识别结果也更加科学、精确，是当前颠覆性技术识别的主流方法。

模型分析法是指通过运用数理统计原理构建颠覆性技术量化模型的方法[6]。在已有研究中，于光辉[7]和武建龙[8]等将专利挖掘分别与 Bass 模型和 Gompertz 模型相结合，从“技术—市场”特征维度综合识别颠覆性技术；熊焰[9]等则构建了“突变—演化”模型，从颠覆性技术的突变特征与演化路径出发，有效提高了颠覆性技术识别的精度。

文献计量法主要指基于专利或论文数据，运用路径分析、文献知识关联、社会网络分析等方法实现颠覆性技术识别。例如，Momeni[10]等依托专利引文分析追踪技术发展路径，并结合 k-core 中心性分析以及技术进步主题模型识别出潜在的颠覆性技术；苑朋彬[11]等利用引文与施引专利映射投入产出关系，通过分析专利知识流动，识别具有颠覆性潜力的技术方向；纪亚琨[12]等借助 IPC 分类号构建领域知识网络，通过网络指标监控新颖性知识的动态扩散，从而在专利层面发掘潜在的颠覆性技术。

指标识别法是根据颠覆性技术特征或某一理论基础构建相应评价指标体系以识别和预见颠覆性技术的方法。Dahlin[13]等认为一个成功的颠覆性发明应该是新颖的、独特的、对未来技术有影响的；王康[14]等通过构建专利融合的影响力、相似性和新颖性指标，鉴定了技术融合中的颠覆性；赵玉桐[15]等构建多维度指标，结合新颖性、融合性与影响力，实现了跨领域颠覆性技术的识别；陈育新[16]等基于核心—边缘理论，从技术基础性、影响力、突变性三个维度来评估颠覆性潜力。综上，颠覆性技术核心特征包含新颖性、突变性、影响力、融合性等。

## 2.2 技术主题识别方法

技术主题包括具有根本性创新和突破性的颠覆性技术，也包括那些在现有技术基础上进行改进或优化、不会导致市场根本性变革的技术。技术主题识别是颠覆性技术识别的基础[17]，多采用构建主题模型的方法。这种方法主要利用文本数据中丰富的语义信息，通过向量表征识别技术术语并聚类成技术主题，进而从海量专利、文献中挖掘快速成长的技术主题。其中最具代表性的模型之一就是 LDA 模型[18]，这是一种旨在揭示文档集合中潜在主题结构的概率生成模型，能够深入挖掘技术领域的隐含知识，具有较好的技术主题识别效果。但该模型也存在忽略词汇、上下文之间语义关系以及主题聚类过程中容易造成混淆等局限性[19]，为此需要构建其他模型或改进该模型。随着大数据和人工智能的兴起，机器学习技术

迈向新高度,推动主题识别模型进一步完善,不断从 LDA 等传统模型中拓展与创新。例如,将 LDA 与 Word2Vec 结合的 LDA2Vec 模型[20]可同时学习主题和词向量表示,弥补了 LDA 模型忽略词汇之间语义关系的不足;将 BERT 与 LDA 结合的 BERT-LDA 模型[21]可充分考虑文本的语义信息,弥补了 LDA 模型缺乏上下文之间语义关系的缺陷。尽管 BERT-LDA 模型增强了文本的语义表示,但是在主题聚类及解读等方面仍有待改进[22]。为了解决这一问题,Grootendorst M[23]提出 BERTopic 模型,该模型使用 BERT 预训练模型和降维算法极大地保留文本语义信息,并引入 c-TF-IDF 算法进行主题标识,具有明显的模块化特征,可以同时结合其他模型或算法进行主题微调 and 聚类优化[24],增强了主题表达能力,进一步提升了技术主题识别的准确性和可解释性。

## 2.3 技术主题演化方法

技术主题演化是揭示技术演化路径的重要方法,它是指技术主题在不同的、连续的、滑动的时间窗口之间表现出来的动态变化,旨在描述技术从萌芽、发展、成熟到衰退的发展脉络。DOSI[25]于 1982 年首次提出技术演进轨迹理论,认为技术创新路径有其特定规律,为后续技术演化路径的研究奠定了理论基础。范维熙[26]等提出了技术发展阶段和专利引文网络构建技术演进路径的思路,但由于其侧重关注高显示度指标,存在一定的时滞性弊端。而后,Yoon[27]根据生长曲线和生命周期划分时间窗口,利用专利分类号来指导技术演化分析,一定程度上克服了时滞性问题,但未能深入挖掘专利文本信息,在演化描述上较为粗糙。在此基础上,有学者[28][29]将技术生命周期理论与文本挖掘方法相结合,基于主题模型,从大量非结构化文本内容中提取有价值的信息进行技术演化分析,提高了技术主题演化分析的动态性和准确性。为了实现演化结果的可视化,知识网络分析成为技术主题演化分析的重要手段,它可以通过网络节点的度、中介性、中心性等网络指标找出重要节点,然后根据这些重要节点在时间序列上的变化探究主题的演化情况[30]。例如,王曰芬[31]等通过构建主题共现网络,利用中心性指标找出不同时期的关键主题,清晰展示知识领域的演进路径;邢晓昭[32]等则结合文本挖掘与社会网络分析,进一步细化了技术主题的演变细节;许海云[33]等人的研究更具创新性,构建了跨科学、技术和产业的三层知识网络,并应用网络社区相似度算法,深度剖析了颠覆性技术在不同知识领域间的互动与演化过程。

综上所述,目前关于颠覆性技术的识别与技术主题演化分析的研究已取得显著进展。这些研究共同推进了对颠覆性技术发展全貌的深入理解,但仍然存在一些不足之处:①在数据源选择上,多数研究聚焦于专利或文献等单一数据源,未能融合包括市场报告、政策文本等在内的多源数据对颠覆性技术进行更全面的度量;②在主题聚类上,现有主题模型在聚类过程中容易出现主题重叠、缺乏细粒度和可解释性等问题;③在主题演化上,目前学者大多纵向分析主题演变路径,在横向分析主题关联关系的演化过程方面尚有欠缺。因此,本文基于多源数据,依据技术生命周期理论划分时间窗口,利用语义信息保留度更高的 BERTopic 模型进行文本挖掘,并在该模型的基础上引入聚类评价指标和 K-means 算法二次优化主题聚类效果,提升主题聚类准确性、细粒度和可解释性,再结合颠覆性技术指标识别该技术主题,

同时通过 Louvain 算法对不同数据源的主题进行社区划分, 以实现多源主题融合。在主题演化分析过程中, 本文借助社会网络分析工具, 不仅横向分析了不同数据源中的主题关联关系演化过程, 还纵向剖析了多源融合主题社区的演变路径, 更全面地展示了技术主题的发展趋势, 为有效把握相关领域颠覆性技术的发展提供理论支撑。

### 3 研究设计

本文构建了一种基于 BERTopic 模型的颠覆性技术主题识别方法, 并引入聚类评价指标和 K-means 算法优化聚类效果, 通过 Louvain 社区发现算法实现多源主题融合, 利用社会网络分析法, 综合分析技术主题演化过程, 具体包括以下四个步骤: ①数据获取与预处理: 确定数据来源包括专利、市场报告和公开政策并收集相关数据, 对数据集进行文本清洗、文本分词、去除停用词。②多源数据时序划分: 对各数据源数据量和总数据集数据量进行技术 S 曲线拟合, 并结合技术生命周期理论划分各阶段的时间窗口。③颠覆性技术主题识别: 利用 BERTopic 模型对数据集进行主题挖掘和表示, 结合主题新颖性、主题突变性和技术影响力三大颠覆性技术指标识别出其中的颠覆性技术主题, 构建得到颠覆性技术主题矩阵。④颠覆性技术主题演化: 在颠覆性技术主题识别结果的基础上分别进行主题关联横向演化分析和主题社区纵向演化分析。具体研究思路如图 1 所示:

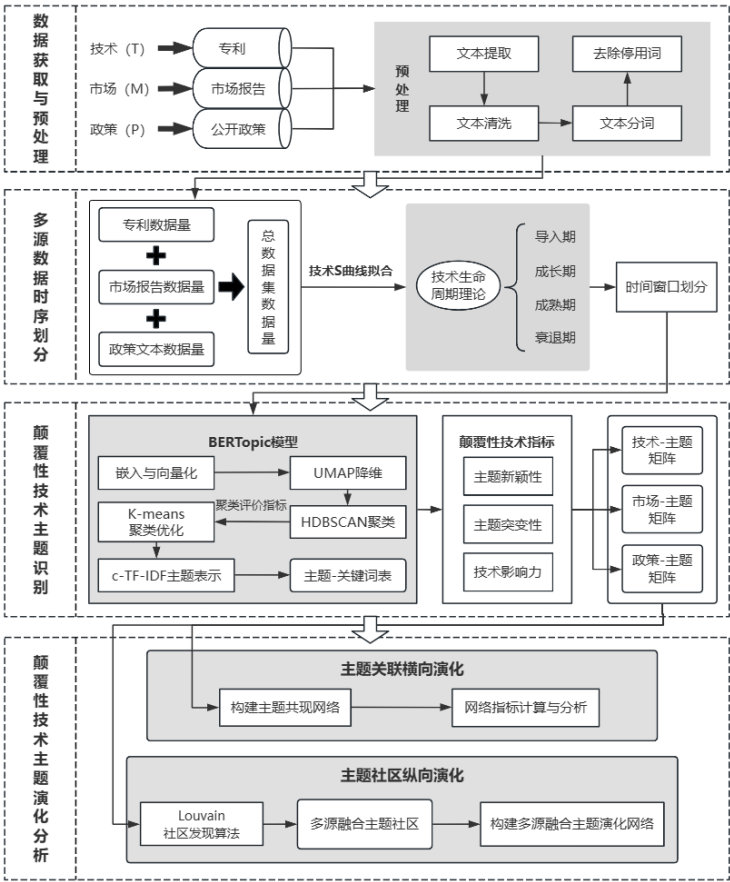


图 1 研究思路

3.1 数据获取与预处理

多源数据的获取对全面融合信息、把握技术发展态势具有重要作用。专利代表现有的技术研发成果，从技术角度反映了技术发展的情况。但对于一些具有颠覆性特征的技术来说，由于它们没有路径依赖性，所以仅从技术角度对其进行识别存在一定的局限性。而在市场报告和公开政策中，包含着相关颠覆性技术的市场情况与政策引导方向，且客观性与真实性较强，因此将市场报告、公开政策与专利数据进行融合，能够从不同侧面发现技术重要研究内容，实现尽可能全面、准确的分析。此外，尽管学术论文作为重要且常用的数据来源之一，但由于其发表周期较长，可能无法提供实时的市场和技术信息，无法及时反映相关产业领域的最新发展，因此本研究在数据选取上未包括学术论文。

本文以专利、市场报告、公开政策作为数据来源，收集文本标题、时间、摘要以及正文字段组信息，建立初始数据集，并进行文本清洗、文本分词、去除停用词等预处理。具体而言，先将无意义文本和重复文本筛除，然后使用 jieba 分词库对文本进行分词。接着进行停用词处理，去除句子中没有实际意义的词汇或标点符号等，如“啊”“吧”等。这些无意义的词汇对文本分析没有帮助，去除后文本将更加清晰、易于理解和分析，从而可以减少后续处理的工作量。此外，针对不同的文本类型和领域，需要进行相应的预处理。例如，对于技术领域的文本，需要了解相关的专业术语和概念，以便更好地理解和分析文本内容。

3.2 多源数据时序划分

一项技术在整个发展过程中会经历 4 个阶段：导入期、成长期、成熟期、衰退期。相应的技术行为数据随时间变化呈现 S 形轨迹。通过各类数据源每年的数据量随时间的变化，分别绘制技术发展曲线，并与技术生命周期 S 曲线进行拟合，可以综合判断该技术经历了哪些发展阶段或正处于哪个发展阶段，从而为数据集划分时间窗口提供依据。因此，笔者对技术、市场、政策三个数据源每年的数据量进行技术 S 曲线拟合，并结合技术生命周期理论，将数据量 S 曲线拟合图与技术生命周期曲线图进行发展阶段的综合对比，从而确定时间窗口节点。

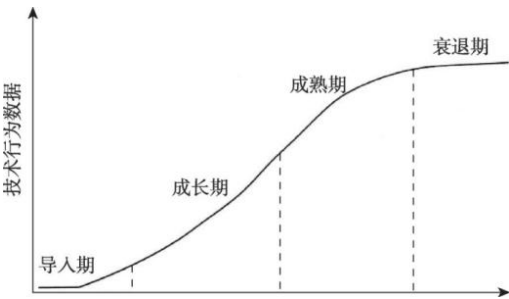


图 2 技术生命周期曲线图



### 3.3 颠覆性技术主题识别

对专利文本摘要以及市场报告、公开政策全文本运用 BERTopic 模型，结合颠覆性技术特征指标，分别识别出技术、市场、政策三个数据源的颠覆性技术主题。

#### 3.3.1 BERTopic 模型

主题建模是一种常用的文本挖掘技术，能够有效地从大量文本数据中提取出主题信息。近年来，随着深度学习技术的发展，主题建模方法也得到了进一步的改进。其中，BERTopic 是一种在 BERT 基础上进行主题建模的技术，通过默认的 UMAP 降维、HDBSCAN 文本聚类 and c-TF-IDF 提取主题词，能够充分考虑文本语义信息，并在提取技术主题时保留重要的词汇。由于 BERTopic 模型流程分模块进行，因此除了默认算法外，每个模块都可以根据研究的数据集特点选择更合适的算法。

本文基于选取的专利、市场报告和政策数据集，构建细粒度的 BERTopic 模型，具体步骤如下：

（1）文档嵌入与向量化：基于 Transformer 语言模型进行预训练，将文档转换为嵌入表示，然后基于 Sentence-BERT 框架，将文档进行向量化。

（2）降维：由于文档向量维度很高，在嵌入空间中非常稀疏，不容易进行聚类，所以需要先进行降维，使用全局和局部意义保留性较高的 UMAP 算法将文档向量降到二维空间中。

（3）文档聚类：默认采用 HDBSCAN 算法，该算法可以无监督自动推荐最优的簇类结果，允许将噪声建模为离群值，防止将不相关的文档分配给任何集群。

为了提高聚类结果细粒度，本研究加入人工监督，综合轮廓系数（Silhouette Coefficient Index）、CH 指数（Calinski-Harabasz Index）、DB 指数（Davies-Bouldin Index）三个聚类评价指标结果，对簇的数量进行二次优化，并输入 K-means 算法中再次聚类。一般情况下，SC 值越高，CH 值越高，DB 值越低，聚类效果越好。

（4）主题表示：默认采用 c-TF-IDF 算法，这是一种基于 TF-IDF 的改进方法，不仅考虑了特征词在当前文档中的频率，还考虑了特征词在其他同一类别文档中的频率，从而更好地体现了特征词在当前类别下的重要性。具体如公式（1）所示，其中， $t$  表示单词（term）， $c$  代表 class， $A$  表示每个 class 的平均单词数量， $tf_{t,c}$  表示  $t$  在  $c$  中的频率， $tf_t$  表示  $t$  在所有类别中的频率。

$$W_{t,c} = tf_{t,c} \cdot \log\left(1 + \frac{A}{tf_t}\right) \quad (1)$$

与传统的主题建模方法相比，BERTopic 在主题建模过程中引入了深度语义向量的概念，使得模型能够更好地捕捉文本中的语义信息，从而提高了主题建模的准确性和有效性。此外，BERTopic 模型还能够有效地处理稀疏性问题，使得模型在处理大规模文本数据时具有更好的性能。

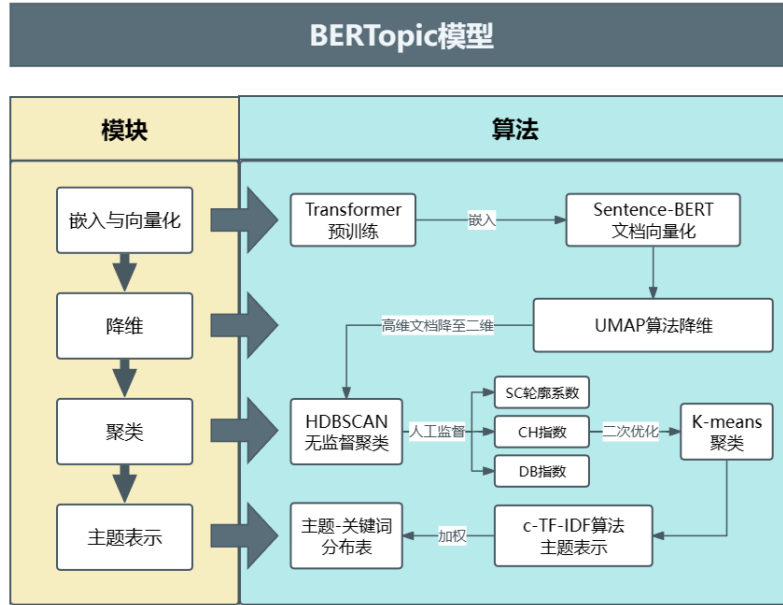


图 3 BERTopic 模型思路

### 3.3.2 颠覆性技术指标

通过主题建模能够得到多个技术主题，但并非每个技术主题都具有颠覆性，因此需要结合颠覆性技术特征构建相应指标，对以上的技术主题进行测度，从而得到颠覆性技术主题。结合已有的关于颠覆性技术特征的研究[34]，本文构建主题新颖性[35]、主题突变性[36]和技术影响力[37]三个测度指标，进一步筛选出颠覆性技术主题。具体指标内容如下：

#### 1. 主题新颖性

颠覆性技术早期具有一定的边缘性和创新型，可以从技术主题的新颖度来反映，即某个技术主题是否在时间窗内首次出现。将主题所对应的主题文档按照时间切片排序，并以出现次数不为 0 的这一年为开始年，主题  $z$  在  $t$  年的新颖度计算如公式（2）所示，其中  $t$  代表当前年份， $FY$  代表开始年份， $NI_t^z$  代表  $t$  年主题  $z$  的新颖度。

$$NI_t^z = \frac{1}{t - FY + 1} \quad (2)$$

#### 2. 主题突变性

突变性作为颠覆性技术的重要特性之一，是指颠覆性技术既需要经历技术发展生命周期的一般阶段，又在演化中呈现不连续性和阶段性爆发的特征，具有时效性及突发性。相关学者们一般通过分析词频数量特征、技术主题内容突变等来测度。基于此，本文借鉴词突发检测算法，测度不同时段节点的突发性，从而测度技术突变可能性。通常认为技术突变性越大，越有可能成为潜在颠覆性技术，具体算法如公式（3）所示，其中  $TF_{v_i^s}$  表示节点  $v_i^s$  在某一时段内出现的频次， $\overline{TF_{v_i^s}}$  表示节点  $v_i^s$  在研究区间内出现频次的平均值， $TFS_{v_i^s}$  表示节点  $v_i^s$  在研究区间内出现频次的标准差。



$$\text{Mut}(v_i^s) = \frac{\text{TF}_{v_i^s}}{\text{TF}_{v_i^s} + 1 + 2\text{TFS}_{v_i^s}} \quad (3)$$

### 3. 技术影响力

麦肯锡认为颠覆性技术影响范围广泛，能影响其他多个技术的发展应用情况，影响力越大，其技术价值与创新能力越大。接近中心性衡量一个节点到所有其他节点的路径长度，表征了一个节点在网络中最大组成部分的参与程度，可直接或间接地接触到网络中的其他节点，知识节点接近中心性越高，在网络中具有越高的信息传播效率，说明其影响力越强。计算如公式（4）所示，其中 $v_i^s$ 和 $v_j^s$ 为知识网络中的不同节点， $|V^s|$ 代表知识网络中的关键词节点总数， $d_{v_i^s, v_j^s}$ 代表 $v_i^s$ 和 $v_j^s$ 之间的最短路径长度（含权重）。

$$\text{CC}(v_i^s) = \frac{|V^s| - 1}{\sum_{j=1}^{|V^s|} d_{v_i^s, v_j^s}} \quad (4)$$

## 3.4 颠覆性技术主题演化

### 3.4.1 基于社会网络分析法的主题关联横向演化

主题关联横向演化能够从不同主题之间的关联关系角度揭示主题的分布情况，例如各个主题在网络中处于何种位置，这些主题之间的距离如何，以及某些主题之间存在一定的关联度。这些信息作为定性判断的依据，为后续的主题发现算法提供更加精确的输入，以便更好地识别领域主题社区。

本文运用社会网络分析法来揭示技术主题之间的关联关系。具体包括以下四个步骤：

- （1）从聚类后的文本分布表中筛选出识别为颠覆性主题的聚类文本，并将筛选后聚类内的文本向量转化为主题向量。
- （2）根据主题向量计算各个主题之间的相似度，并构建相似度矩阵。
- （3）构建主题共现网络图，网络中的节点代表选取的主题词，两个节点之间的边代表它们的相似度。
- （4）计算中间中心性、接近中心性、特征向量中心性、聚集系数、权重和标准化点度中心性等网络指标，用以具体分析主题之间的网络关系。

中间中心性表示一个节点在网络中的“中介”地位，即通过该节点的最短路径数量。接近中心性表示一个节点到网络中所有其他节点的平均距离，该中心性高的节点在网络中具有较高的可达性，能够快速地与网络中的其他节点进行通信。特征向量中心性考虑了节点的连接质量，即一个节点的重要性不仅取决于它连接的节点数量，还取决于它所连接的节点的重要性。聚集系数衡量一个节点的邻居节点之间的连接程度，该系数高的节点意味着其邻居节点之间形成了紧密的子群。权重表示节点之间连接的强度或频率。标准化点度中心性衡量一个节点的直接连接数，该中心性高的节点在网络中连接广泛，可能涉及多个不同的领域或社

区。

### 3.4.2 基于 Louvain 算法的主题社区纵向演化

模块度（Modularity）是一个衡量网络社区结构强度的指标，它表示社区内部连接的紧密程度与随机网络中期望的连接紧密程度的差异。模块度越高，说明社区内部的连接越紧密，社区结构越明显。

Louvain 算法是一种基于模块度的社区发现算法，它通过迭代的方式优化网络的模块度来识别社区结构。相比于传统算法，其时间复杂度较低，且划分效果更好，是目前最主流的社区发现算法之一。具体工作原理如下：

（1）初始化：将每个节点视为一个独立的社区。

（2）局部优化：对于每个节点，尝试将其移动至其邻居节点所在的社区，并计算模块度的增益。如果移动后模块度增益为正，则进行移动。模块度（Modularity）计算如公式（5）

所示，其中  $A_{ij}$  是节点  $i$  和节点  $j$  之间的边的权重， $k_i$  是节点  $i$  的度（无权图中的边数，或加权图中的权重和）， $m$  是网络中所有边的权重之和， $c_i$  和  $c_j$  分别是节点  $i$  和节点  $j$  所属的社区， $\delta(c_i, c_j)$  是克罗内克函数，当  $c_i = c_j$  时为 1，否则为 0。

$$Q = \frac{1}{2m} \sum_{ij} [A_{ij} - \frac{k_i k_j}{2m}] \delta(c_i, c_j) \quad (5)$$

（3）全局优化：重复步骤（2），直到网络中的模块度不再增加。

（4）社区合并：将每个社区视为一个新的节点，社区间的边权重为原社区间所有边的权重之和。构建新的网络，并返回步骤（2）。

（5）终止条件：当模块度不再增加时，算法停止。

### 3.5 颠覆性技术研究方法运用的综合优势

为了在技术浪潮中占据前沿阵地，及时捕捉并引领颠覆性技术的兴起，对颠覆性技术的探索不仅需关注技术成果的影响力及其核心属性，还需深入挖掘其随时间演进的知识积累轨迹。基于此，在技术识别结果的基础上，进一步构建技术演化脉络，明确技术发展趋势，成为精确预判技术未来走向的必经之路。针对这些挑战，本研究创新性地融合 BERTopic 模型和 Louvain 算法，针对技术、市场、政策三重数据源，构建了横纵向双维度的颠覆性技术演化分析框架。

在主题识别方法中，BERTopic 模型通过深度学习驱动，不仅能够深度挖掘语义特征，还保持了重要词汇的识别能力；通过 UMAP 降维，能够有效解决高维空间处理的难题，保留文档的全局和局部意义，提高了后续聚类分析的效率与准确性；通过 HDBSCAN 算法的无监督特性及其对异常值的灵活处理，加之结合聚类评估指标和 K-means 算法的聚类优化策略，提升了聚类结果的准确性和细粒度；通过 c-TF-IDF 算法，考虑词汇在类别内部及跨

类别的重要性，进一步优化了主题表达的精确性和可解释性。

在多源主题融合中，Louvain 算法以模块度为核心评估标准，有效衡量网络中社区结构的紧凑性，通过局部优化与全局优化循环迭代，实现社区划分的持续优化，不仅大幅降低了计算复杂度，相较于其他传统方法，还显著提高了社群划分的质量与效率。Louvain 算法的迭代策略与模块度最大化目标相结合，确保了即便在大规模网络中也能迅速收敛至高质量的社区结构解决方案，因而在应对多数据源主题社区划分的情况下依然能保持较高的处理效能。

上述优化后的 BERTopic 模型与 Louvain 算法的有机结合，可以在深度挖掘语义特征、保留文本全局与局部意义、构建细粒度主题聚类的基础上，进行颠覆性技术主题识别，降低了单独使用 HDBSCAN 算法聚类的粗糙性，并极大程度上保持了多源主题融合过程中的数据处理质量与效率。在此过程中，同时利用社会网络分析法构建主题共现网络和多源融合主题网络，描绘技术主题的横向与纵向演化轨迹，为深入理解颠覆性技术的发展趋势提供了新视角。因此本文充分利用多源数据中的语义特征和文本意义，通过聚类识别与指标分析确定技术的颠覆性，深入剖析技术发展脉络，为颠覆性技术识别与演化路径分析提供创新的研究方法。

## 4 实证分析

新能源汽车产业作为国家战略性新兴产业，承载着产业升级、技术创新和绿色发展的重任。在政策扶持和市场需求的双重驱动下，该产业迅速崛起，具有极大的颠覆性潜力。新能源汽车产业领域的研究覆盖了电池技术、驱动系统、充电基础设施以及车辆智能化等多个关键方面，这些技术对于促进能源结构优化、产业经济升级具有重要作用。因此，本文以新能源汽车产业领域为例开展实证研究，旨在验证本文对颠覆性技术主题识别和演化过程分析的有效性，并为该产业的技术发展和战略规划提供参考。

### 4.1 数据获取与预处理

CNKI 国家标准全文数据库是中国专利文献的全面资源库，提供详实的技术描述。中国知网重要报纸数据库是国内领先的新闻资讯和市场报告资源平台，覆盖经济、产业、科技领域，提供快速更新的市场动态。北大法宝是国内权威的法律信息平台，及时汇集国家和地方法政策法规。因此，本文选择 CNKI 国家标准全文数据库、中国知网重要报纸数据库和北大法宝分别作为专利、市场报告和公开政策的数据来源，以确保研究的全面性、时效性和权威性。

以“新能源汽车”or“电动汽车”or“纯电动汽车”or“混合动力汽车”or“燃料电池汽车”作为关键词，对新能源汽车领域的专利、市场报告和公开政策三类数据进行检索。考虑到产业前期三类数据数量较少，本文设置时间跨度为 2009-2023 年，检索时间为 2024 年 1 月 20 日。经过筛选和去重处理后，总计获得 85047 条信息，其中包括 71500 项专利、13273 篇市场报告和 274 条政策文本。

分别对原始数据集进行预处理，包括文本清洗、文本分词、去除停用词等操作，以保证后续分析的准确性。通过 Python 中的 Jieba 库进行分词，并加载自定义字典防止部分词汇被错误切割，例如“混合动力汽车”“燃料电池汽车”“电池组”“热管理系统”等词汇。本研究的停用词表在哈工大停用词表的基础上加入了部分人工筛选停用词，例如“第一条”“本发明”“设有”等。

4.2 多源数据时序划分

笔者依据技术生命周期理论对数据集进行时序划分，运用 LogletLab 软件对上述三个数据源每年的数据量进行技术 S 曲线拟合。根据  $R^2$  值可以判断出总数据集的曲线拟合效果优良。经过与 3.2 中图 2 技术生命周期曲线图各发展阶段的曲线趋势比对后，笔者将专利、市场、政策以及总数据集的技术 S 曲线拟合图分别划分出导入期、成长期、成熟期和衰退期，如图 4、图 5 所示。综合比对时间节点后，认为将目前收集到的新能源汽车产业数据集划分为 T1、T2、T3 这三个时间窗口能够与技术生命周期曲线图的发展阶段大致重合，如表 2 所示。

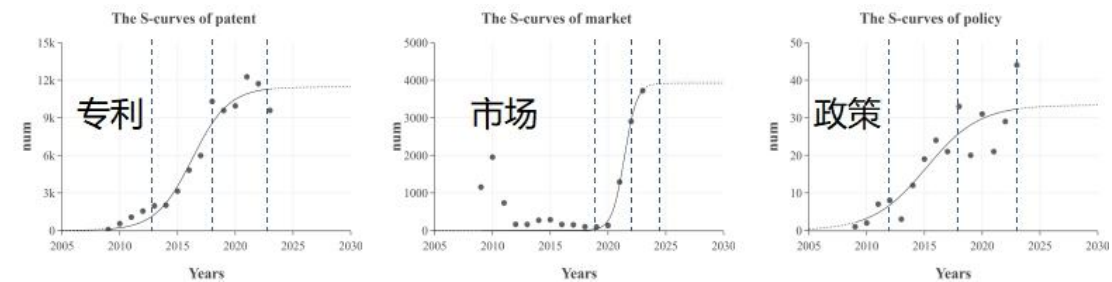


图 4 专利、市场、政策数据技术 S 曲线拟合结果

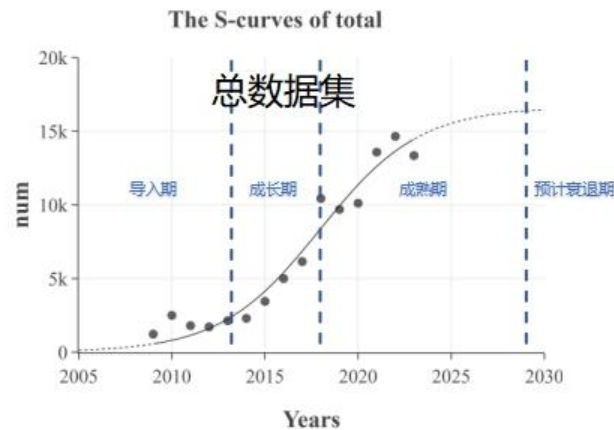


图 5 总数据集技术 S 曲线拟合图

表 1 总数据集技术生命曲线拟合效果指标

Logistic		d	K	a	tm	r	SSE	RMS	MAD	MAPE	SE	ln[MLE]	AICc	R <sup>2</sup>	p	1%	10%	50%	90%	99%
The S-curves of total	Phase 1	0.000	21312	13.9	2020	0.317	11696870	806	551	0.310	883	-146	300	0.946	1.46e-11	2005	2012	2020	2027	2034

表 2 技术生命周期时间窗口划分

导入期（T1）	成长期（T2）	成熟期（T3）	预计衰退期
---------	---------	---------	-------

### 4.3 颠覆性技术主题识别

#### 4.3.1 主题识别

本研究采用 BERTopic 主题建模方法，分别基于上述三个数据源的文档，针对 T1、T2、T3 时间窗口的文本数据提取主题。

首先采用 paraphrase-multilingual-MiniLM-L12-v2 句向量模型嵌入预处理后的文本。该模型用 SBERT 进行预训练，支持中英文等多语言，在训练速度和性能上都表现优越。然后对得到的向量化文本进行 UMAP 降维，使高维度的文本向量降至二维空间。为了验证文本语义的一致性，笔者计算了降维前后文本向量的余弦相似度，结果平均值为 0.793，表明降维后的文本向量在很大程度上保留了原始高维空间中的语义结构。

在文本聚类时，先采用 HDBSCAN 算法自动聚类，观察到该算法聚类质量较好，各个聚类之间重叠度较低，但缺乏细粒度，簇数多为 5 或 10 以内，不利于更加精准地识别新能源汽车产业领域的颠覆性技术主题。例如图 6 展示了 HDBSCAN 算法下的 T3 时期政策数据聚类结果数量为 3，而本文后续通过图 12 展示的实际聚类数量为 9，细粒度更高且保持重叠度较低，聚类效果更好。因此，为进一步细分聚类，笔者结合 SC 轮廓系数、CH 指数、DB 指数三个聚类评价指标结果来确定聚类数，并通过 K-means 算法再次聚类。一般情况下，SC 值越高，CH 值越高，DB 值越低，聚类效果越好。综合如图 7、图 8 和图 9 所示的聚类评价指标结果，考虑聚类的质量与解释性，最终确定各个数据源在不同时间窗口下总共划分出 108 个聚类，如表 3 所示。

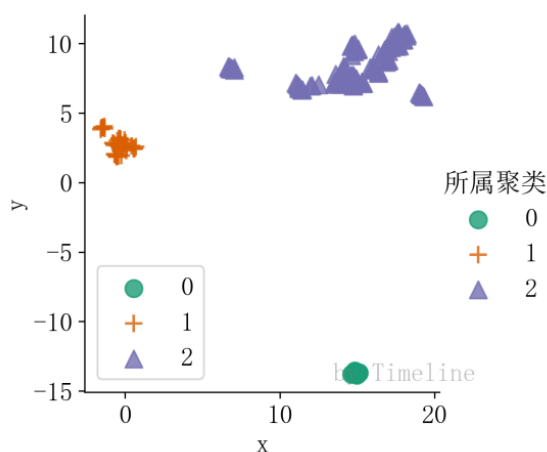


图 6 HDBSCAN 算法聚类结果（T3 时期政策数据）

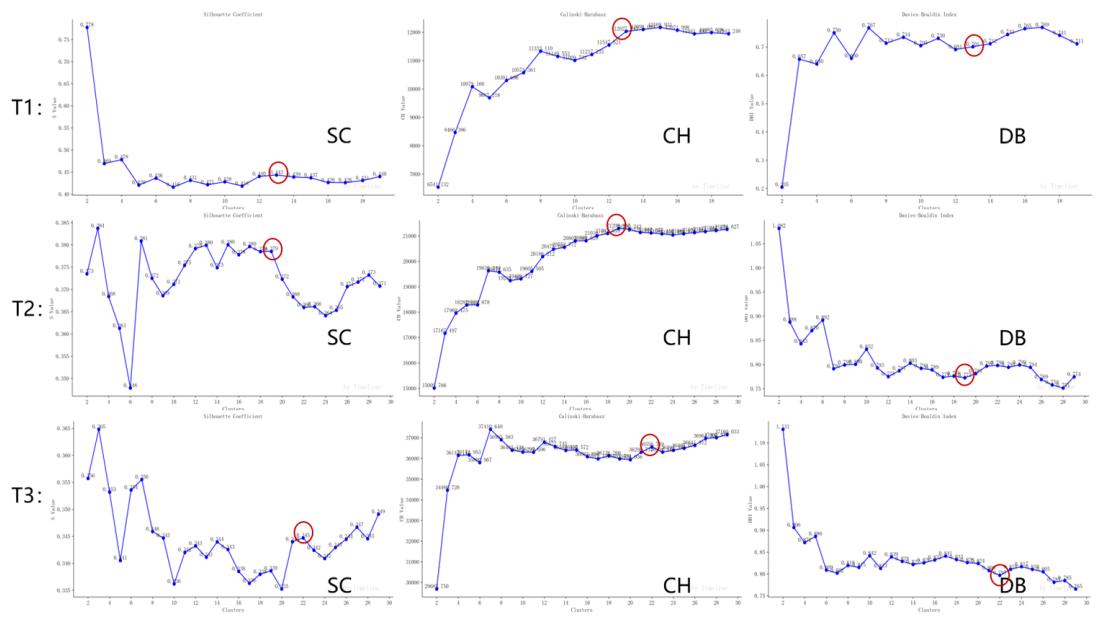


图 7 专利数据 (T1-T3) 聚类评价指标

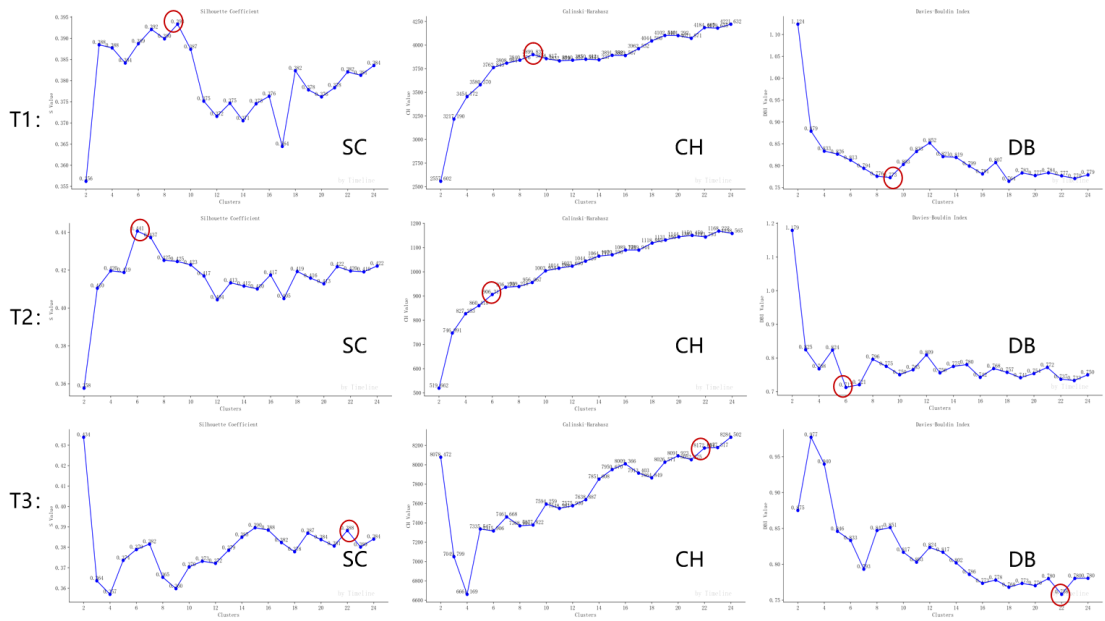


图 8 市场数据 (T1-T3) 聚类评价指标



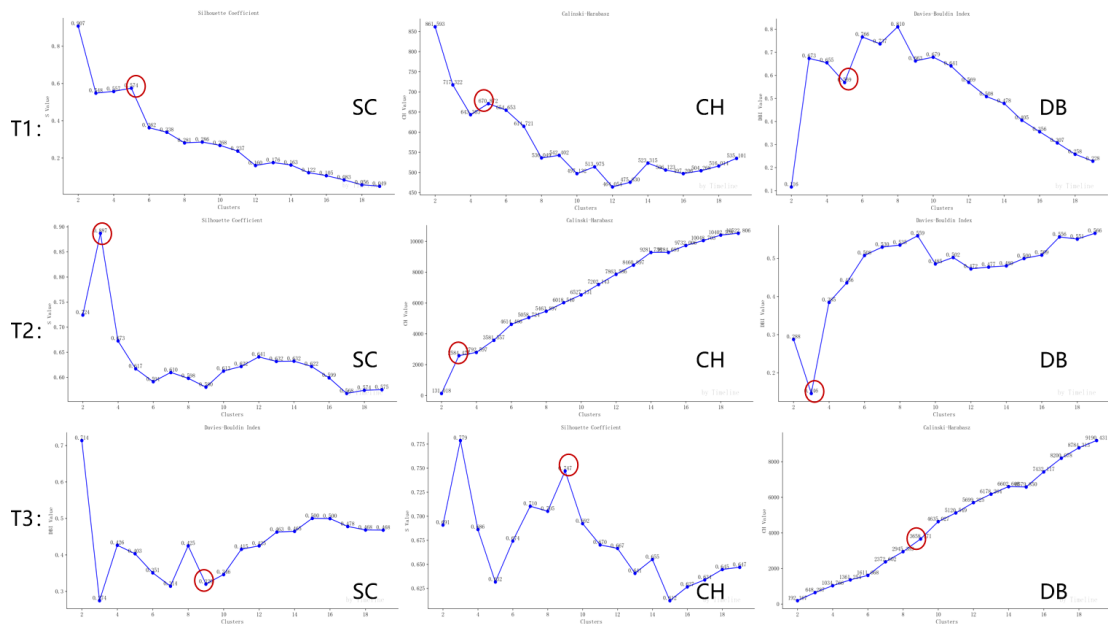


图 9 政策数据（T1-T3）聚类评价指标

表 3 各个数据源在不同时间窗口下的聚类数量

数据源 \ 时间窗口	T1	T2	T3
专利	13	19	22
市场报告	9	6	22
公开政策	5	3	9

此外，通过绘制各个聚类的二维坐标图，更直观地展现优化后的聚类结果之间的重叠度较低且细粒度较高，如图 11、图 12 和图 12 所示。最后采用 c-TF-IDF 算法提取各个聚类的主题特征词，为之后识别出来的颠覆性技术主题命名标签奠定基础。

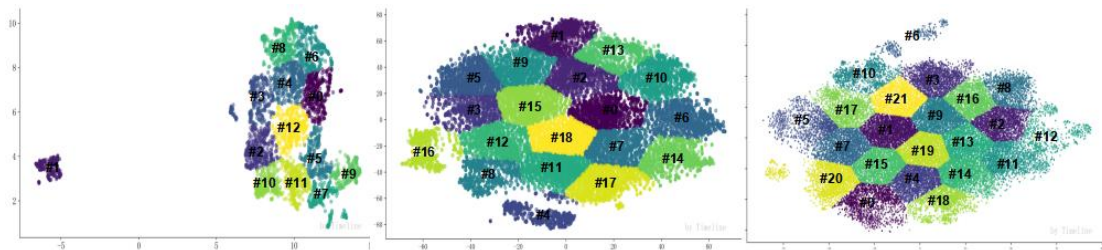


图 10 专利聚类二维坐标图

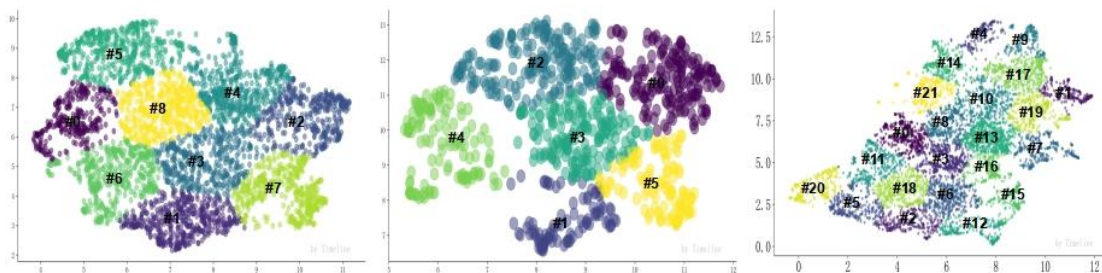


图 11 市场聚类二维坐标图



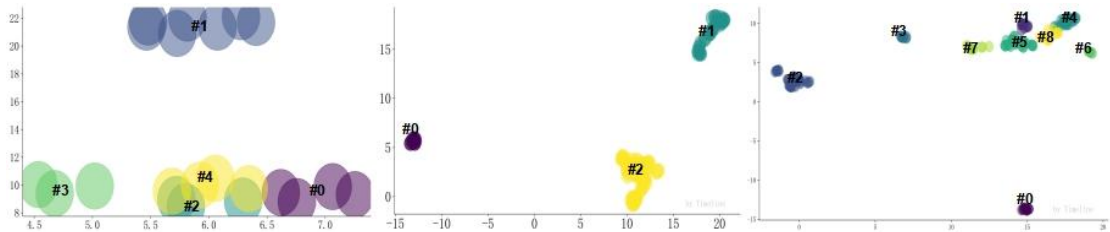


图 12 政策聚类二维坐标图

4.3.2 颠覆性技术指标测量

根据 3.3.2 可知，颠覆性技术主题通常表现出一定的主题新颖性、突破性和较大的技术影响力。因此本文综合考量上述三项指标，计算出各指标得分并通过熵权法赋权求和，得到颠覆性技术主题综合得分。考虑到不同时间窗口下技术背景不同，本文分别对不同时间窗口的各个数据源进行评分和赋权，并根据是否超过组别平均分来判定选择，例如表 4 展示了 T1 时期的专利主题颠覆性评分结果。最终在三个时间阶段中共识别出 49 个颠覆性技术主题（包括各时间段中重复相似的主题），并根据主题关键词权重分布对其进行命名表示，具体如表 5、表 6、

时间段	市场主题	主题关键词
2009 - 2013 (T1)	Ma1 产业合作与投资建设	有限公司;本报讯;公司;项目;合作;签署;基地;投资;协议;市长
	Ma2 政策补贴与市场推广	补贴;试点;购买;财政部;新能源汽车;推广;政策;示范;出台;私人
	Ma3 产品展示与应用示范	电动车;世博会;世博;纯电动汽车;混合;上海;动力;世博园;亮相;绿色
	Ma4 产业规划与核心技术	规划;电池;我国;产业化;锂电池;扶持;动力电池;混合动力汽车;电机;板块
	Ma7 充电基础设施建设	电动汽车;充电;充电站;建设;电动;充电桩;公交车;记者;国家电网;我市
2014 - 2018 (T2)	Mb0 政策驱动与基础设施建设	新能源汽车;补贴;充电;推广应用;政策;新能源车;推广;意见;设施;财政部
	Mb2 领先企业与技术革新	特斯拉;电池;电动车;电动汽车;汽车;动力电池;公司;巨头;锂电池;锂电
	Mb4 市场动态与增长趋势	同比;增长;万辆;市场;中国汽车工业协会;增速;产销;数据;产销量;显示
2019 - 2023 (T3)	Mc0 市场动态与销售趋势	中国汽车工业协会;中汽协;环比;排队;零售;小时;秘书长;本报讯;国庆...
	Mc2 生产能力与技术进步	有限公司;车间;生产;项目;科技;基地;位于;研发;零部件;半导体
	Mc4 消费市场与政策支持	消费;恢复;工作;消费市场;消费品;优化;出台;重点;商务部;零售总额
	Mc5 产业发展与国家战略	发展;推进;制造;打造;创新;习近平;总书记;构建;战略;高端
	Mc11 绿色发展与低碳战略	绿色;提出;低碳;能源;体系;碳达峰;双碳;十四五;目标;碳中
	Mc12 企业业绩与资本动态	公司;业绩;亿元;净利润;上市公司;上市;股份;A 股;披露;板块
	Mc15 成本上涨与价格波动	价格;动力电池;电池;上涨;材料;锂电池;锂电;回收;资源;宁德
	Mc16 国际市场与供应链挑战	特斯拉;涨价;原材料;车型;影响;欧盟;上调;欧洲;芯片;短缺
	Mc20 疫情防控与政策措施	措施;疫情;国务院;部署;政策措施;发放;稳住;介绍;扎实;一揽子

表 7 所示。

表 4 T1 时期专利主题颠覆性评分结果

聚类	新颖度	突变性	影响力	综合得分
----	-----	-----	-----	------

指标权重	0.004	0.316	0.68	
<i>C0</i>	0.063	0.800	0.177	<i>0.373</i>
<i>C6</i>	0.063	0.842	0.157	<i>0.373</i>
<i>C4</i>	0.063	0.825	0.151	<i>0.364</i>
<i>C3</i>	0.063	0.784	0.124	<i>0.332</i>
<i>C11</i>	0.063	0.862	0.077	<i>0.325</i>
<i>C8</i>	0.063	0.852	0.077	<i>0.322</i>
<i>C9</i>	0.063	0.706	0.124	<i>0.308</i>
C10	0.063	0.586	0.158	0.293
C2	0.063	0.582	0.158	0.292
C1	0.063	0.660	0.077	0.261
C12	0.063	0.570	0.077	0.232
C7	0.067	0.551	0.077	0.226
C5	0.067	0.381	0.134	0.212
组别平均分				<b>0.301</b>

注：表中黑色加粗、斜体表示颠覆性技术主题筛选结果

表 5 专利主题矩阵

时间段	专利主题	主题关键词
2009 — 2013 (T1)	Ta0 动力电池管理系统	充电; 电池组; 供电; 放电; 电量; 串联; 动力电池; 单体; 并联...
	Ta3 车辆动力与控制系统	控制; 信号; 控制器; 方法; 制动; 传感器; 踏板; 转速; 转矩; 控制系统...
	Ta4 电气系统与安全保护	电路; 输出; 电压; 输入; 检测; 开关; 电流; 电源; 继电器; 电阻...
	Ta6 智能通信与电网管理系统	信息; 数据; 系统; 管理; 通信; 监控; 电网; 充电桩; 智能; 终端...
	Ta8 车辆集成与数据处理	模块; 采集; 总线; 实时; 显示; 整车; 处理; 通讯; 状态; 接口...
	Ta9 热管理系统	散热; 冷却; 空调; 温度; 水泵; 制冷; 散热器; 热量; 压缩机; 制热...
2014 — 2018 (T2)	Ta11 车辆结构与设计	固定; 结构; 安装; 实用新型; 底盘; 总成; 车身; 支架; 强度; 车架...
	Tb1 智能电网与充电管理	方法; 系统; 信息; 用户; 步骤; 电网; 确定; 模型; 终端; 目标...
	Tb2 控制系统与通信模块	模块; 信号; 控制器; 采集; 控制系统; 通信; 总线; 故障; 通讯; 测试...
	Tb3 无线供电与导轨技术	接收; 供电; 无线; 发射; 线圈; 传输; 高频; 无线电; 端子; 道路...
	Tb5 电力电子与电路设计	电路; 输出; 电压; 输入; 电源; 开关; 电流; 直流; 电阻; 高压...
	Tb6 电机驱动与动力系统	电机; 驱动; 齿轮; 离合器; 动力; 传动; 发动机; 机构; 变速器; 车轮...
2019 —	Tb8 充电设施设计与布局	充电桩; 设置; 新能源汽车; 端面; 连接; 底部; 顶部; 配合; 左右; 滑动...
	Tb9 电池管理系统与检测技术	检测; 管理系统; 单元; 发送; 监测; 放电; 识别; 单体; 均衡; 人机交互...
	Tb13 车辆数据采集与分析	获取; 计算; 数据; 车辆; 行驶; 实时; 监控; 电量; 参数; 车载...
	Tb15 可再生能源与储能技术	太阳能; 电能; 发电机; 蓄电池; 发电; 电池组; 动力电池; 电池板; 续航...
2019 —	Tc2 动力电池状态和健康管理	温度; 里程; 能耗; 动力电池; 介质; 平均; 续航; 环境温度; 出水口...
	Tc3 油门控制和驾驶踏板技术	扭矩; 控制; 转速; 踏板; 信号; 控制器; 开度; 油门; 指令...

2023	Tc5 车辆状态监控与同步技术	判断;车辆;状态;采集;是否;阈值;剩余;实际;获得;系数...
(T3)	Tc8 负荷调度与优化建模	方法;模型;获取;目标;计算;优化;建立;负荷;调度;构建...
	Tc11 热管理系统	系统;换热器;冷却;回路;空调;压缩机;冷却液;制冷;热泵;热管理系统...
	Tc13 电池安全与防护技术	转动;壳体;缓冲;电池组;减震;作用;产生;电芯;模组;灭火...
	Tc16 通信平台与故障处理	模块;发送;接收;通信;平台;终端;故障;识别;单元;预约...
	Tc17 电机驱动与整车控制系统	电机;输出;驱动;离合器;整车;制动;模式;齿轮;驾驶;行星...

表 6 市场主题矩阵

时间段	市场主题	主题关键词
2009 - 2013 (T1)	Ma1 产业合作与投资建设	有限公司;本报讯;公司;项目;合作;签署;基地;投资;协议;市长
	Ma2 政策补贴与市场推广	补贴;试点;购买;财政部;新能源汽车;推广;政策;示范;出台;私人
	Ma3 产品展示与应用示范	电动车;世博会;世博;纯电动汽车;混合;上海;动力;世博园;亮相;绿色
	Ma4 产业规划与核心技术	规划;电池;我国;产业化;锂电池;扶持;动力电池;混合动力汽车;电机;板块
	Ma7 充电基础设施建设	电动汽车;充电;充电站;建设;电动;充电桩;公交车;记者;国家电网;我市
2014 - 2018 (T2)	Mb0 政策驱动与基础设施建设	新能源汽车;补贴;充电;推广应用;政策;新能源车;推广;意见;设施;财政部
	Mb2 领先企业与技术革新	特斯拉;电池;电动车;电动汽车;汽车;动力电池;公司;巨头;锂电池;锂电
	Mb4 市场动态与增长趋势	同比;增长;万辆;市场;中国汽车工业协会;增速;产销;数据;产销量;显示
2019 - 2023 (T3)	Mc0 市场动态与销售趋势	中国汽车工业协会;中汽协;环比;排队;零售;小时;秘书长;本报讯;国庆...
	Mc2 生产能力与技术进步	有限公司;车间;生产;项目;科技;基地;位于;研发;零部件;半导体
	Mc4 消费市场与政策支持	消费;恢复;工作;消费市场;消费品;优化;出台;重点;商务部;零售总额
	Mc5 产业发展与国家战略	发展;推进;制造;打造;创新;习近平;总书记;构建;战略;高端
	Mc11 绿色发展与低碳战略	绿色;提出;低碳;能源;体系;碳达峰;双碳;十四五;目标;碳中
	Mc12 企业业绩与资本动态	公司;业绩;亿元;净利润;上市公司;上市;股份;A股;披露;板块
	Mc15 成本上涨与价格波动	价格;动力电池;电池;上涨;材料;锂电池;锂电;回收;资源;宁德
	Mc16 国际市场与供应链挑战	特斯拉;涨价;原材料;车型;影响;欧盟;上调;欧洲;芯片;短缺
	Mc20 疫情防控与政策措施	措施;疫情;国务院;部署;政策措施;发放;稳住;介绍;扎实;一揽子

表 7 政策主题矩阵

时间段	市场主题	主题关键词
2009 - 2013 (T1)	Pa2 政策导向与技术研发支持	科技部;财政部;工作;组织;部门;材料;评估;发展;制造;科技厅...
2014 - 2018 (T2)	Pa3 电力系统与标准制定	传导;直流;装置;电动汽车;交流;质量;认证;业务;类别;提交...
	Pb0 证券市场与投资管理	证券;份额;误导;投资人;准予;会计核算;开放式;契约型;披露;买卖...
	Pb1 税收优惠与车辆购置	国家税务总局;免征;购置税;申请材料;转型;文明;财政部;车型;审查...
2019 - 2023	Pc0 购置税免征政策	购置税;免征;财政部;部门;中华人民共和国;审查;车型;提交;企业...
	Pc3 节能与新能源汽车优惠政策	享受;减免;节约能源;节能;优惠政策;交通运输;新能源;通知;申请...
2023	Pc6 行政许可与机动车辆生产	确需;行政许可;许可法;许可;推广应用;国务院;审批;机动车辆;行政...

4.3.3 主题识别效果检验

为了检验本文主题模型的识别效果，笔者利用社会网络分析法，构建不同数据源中各个时间窗口下的主题共现网络图，以展示各数据源主题的内部关系，如图 13、图 14、图 15 所示。可见各数据源在不同时间窗口下的主题之间的互联较为紧密且主题分布稀疏度高，再次印证了本文的主题聚类效果表现良好。

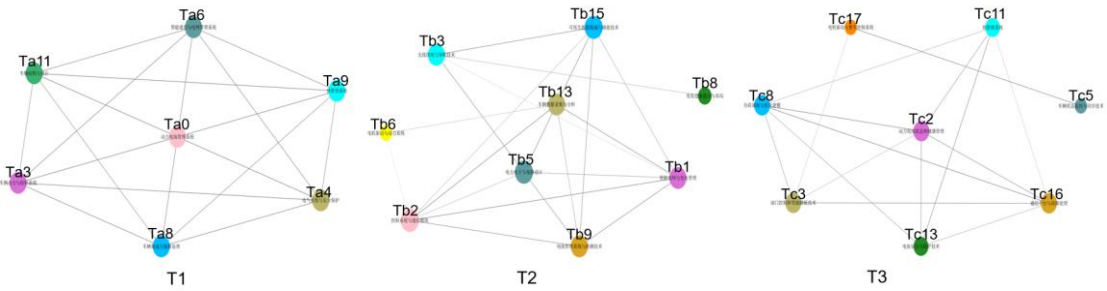


图 13 专利颠覆性技术主题网络图

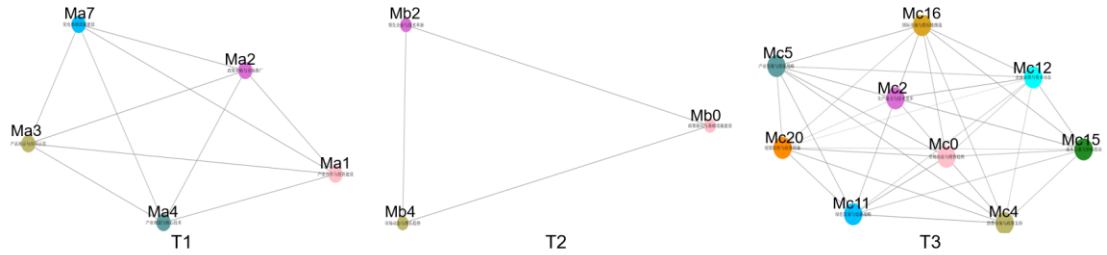


图 14 市场颠覆性技术主题网络图

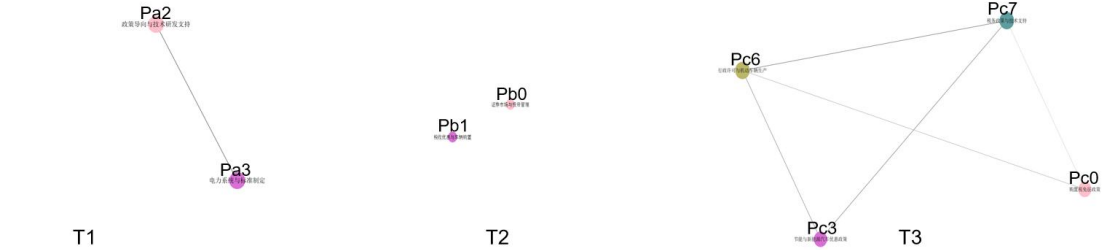


图 15 政策颠覆性技术主题网络图

为了验证本文主题模型的优越性，笔者将本文改进后的 BERTopic 模型与传统 BERTopic、Word2Vec、LDA 三种主题模型进行对比实验，采用十折交叉验证法对数据集进行训练和测试，得到本文主题模型的准确率、召回率、F1 值都比上述三种对比模型高，如表 8 所示，表明本文改进后的 BERTopic 模型在主题识别效果上具有显著优势。

表 8 主题模型识别效果对比结果

主题模型	准确率/%	召回率/%	F1 值/%
改进的 BERTopic (本文模型)	79.1%	85.3%	81.8%
传统 BERTopic	74.2%	80.9%	78.5%
Word2Vec	70.6%	77.4%	73.7%
LDA	65.7%	71.2%	67.9%

此外，通过对比专利、市场、政策三个来源识别出的颠覆性技术主题，可以发现多源数据能够识别出单一数据源所没有包括的主题，从而提高主题的覆盖率。市场来源包含“产业合作与投资建设”、“产业规划与核心技术”等专利来源所缺失的主题，这些主题对于理解市场环境对颠覆性技术的应用具有重要意义。政策来源包含“政策导向与技术研发支持”、“电力系统与标准制定”等专利来源所缺失的主题，这些主题对于理解政策环境对颠覆性技术发展的支持机制至关重要。因此，相较于单一专利来源的识别结果，将这些专利、市场和政策的多源主题进一步融合分析，不仅可以涵盖技术细节，还涉及市场动态和政策支持等多个方面，能够更全面地识别和理解颠覆性技术主题并把握技术发展趋势。

#### 4.4 颠覆性技术主题演化

##### 4.4.1 主题关联横向演化分析

在识别出的颠覆性技术主题的基础上，分析不同时间窗口下所有主题之间的关联关系变化。首先，分别从技术、市场和政策三个来源的聚类后文本分布表中筛选出识别为颠覆性主题的聚类文本。其次，分别在 T1、T2 和 T3 时间窗口下，将这三个来源的颠覆性聚类文本向量进行合并，计算合并后文本向量的平均值，从而转化为“技术-市场-政策”主题向量。然后，根据主题向量计算各个主题之间的余弦相似度，构建得到主题相似度矩阵。接着，根据相似度矩阵，分别在 T1、T2 和 T3 时间窗口下，通过 NetworkX 库构建“技术-市场-政策”主题共现网络图，网络中的节点代表选取的主题词，两个节点之间的边代表它们的余弦相似度，如图 16 所示。最后，通过 NetworkX 库计算相关网络指标，对比三个时期的网络指标变化，详细分析主题关联的横向演化，具体指标如表 9、表 10、表 11 所示。

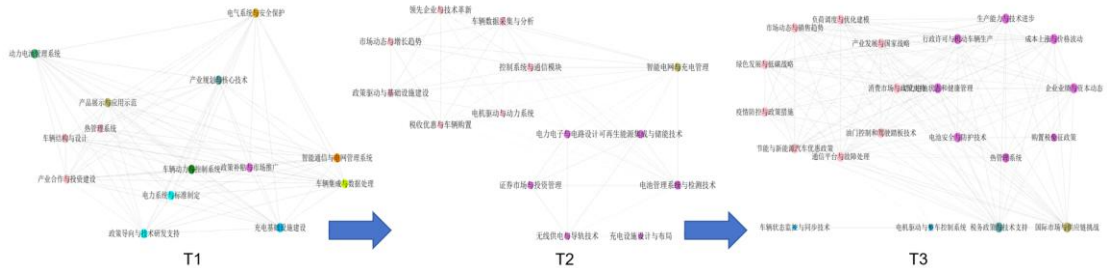


图 16 “技术-市场-政策”主题共现网络图

##### 1. T1 时期（2009-2013 年）

- (1) 中心性指标：所有主题的中间中心性为 0，接近中心性为 1，特征向量中心性为 0.267，表明网络中节点之间的距离较短，节点的重要性相对均衡。

- (2) 聚集系数：所有主题的聚集系数为 1，表明每个主题的邻居节点之间都完全连接。
- (3) 权重和标准化点度中心性：所有主题的权重和标准化点度中心性相同，表明网络中节点的连接强度和连接数相对均衡。

反映了在这一时期,新能源汽车产业正处于起步阶段,技术、市场和政策均在探索之中。

2. T2 时期（2014-2018 年）

- (1) 中心性指标：一些技术主题的中间中心性和接近中心性提高，如“控制系统与通信模块”、“车辆数据采集与分析”，表明这些主题在网络中的重要性增加。
- (2) 聚集系数：除了“电机驱动与动力系统”和“充电设施设计与布局”，其他主题的聚集系数略有下降，表明这些主题的社区结构变得不那么紧密。
- (3) 权重和标准化点度中心性：一些技术主题在网络中的权重和标准化点度中心性明显较高，如“控制系统与通信模块”、“车辆数据采集与分析”、“电池管理系统与检测技术”以及“智能电网与充电管理”，表明这些主题在网络中的重要性更高。

反映了在这一时期,新能源汽车产业快速发展,技术进步显著,充电基础设施建设加速。

3. T3 时期（2019-2023 年）

- (1) 中心性指标：技术和市场主题的中间中心性和接近中心性进一步提高，特别是“油门控制和驾驶踏板技术”、“消费市场与政策支持”等主题，表明这些主题在网络中的桥梁作用和可达性显著提升。
- (2) 聚集系数：聚集系数进一步下降，特别是技术主题，表明这些主题的社区结构更加松散。
- (3) 权重和标准化点度中心性：市场和政策主题的权重和标准化点度中心性显著提升，表明这些主题的重要性显著增加。

反映了在这一时期,新能源汽车产业进入成熟期,市场对新能源汽车广泛接受,政策持续推动,技术创新和市场需求推动产业向智能化、网联化发展。

表 9 T1 时期“技术-市场-政策”主题共现网络指标

主题来源	标签	中间中心性	接近中心性	特征向量中心性	聚集系数	权重	标准化点度中心性
市场	产业合作与投资建设	0	1	0.267	1	15	1.154
技术	车辆结构与设计	0	1	0.267	1	15	1.154
技术	热管理系统	0	1	0.267	1	15	1.154
市场	政策补贴与市场推广	0	1	0.267	1	15	1.154
市场	产品展示与应用示范	0	1	0.267	1	15	1.154
市场	产业规划与核心技术	0	1	0.267	1	15	1.154
市场	充电基础设施建设	0	1	0.267	1	15	1.154
政策	政策导向与技术研发支持	0	1	0.267	1	15	1.154
政策	电力系统与标准制定	0	1	0.267	1	15	1.154
技术	动力电池管理系统	0	1	0.267	1	15	1.154

技术	车辆动力与控制系统	0	1	0.267	1	15	1.154
技术	电气系统与安全保护	0	1	0.267	1	15	1.154
技术	智能通信与电网管理系统	0	1	0.267	1	15	1.154
技术	车辆集成与数据处理	0	1	0.267	1	15	1.154

表 10 T2 时期“技术-市场-政策”主题共现网络指标

主题来源	标签	中间中心性	接近中心性	特征向量中心性	聚集系数	权重	标准化点度中心性
技术	控制系统与通信模块	0.075	0.929	0.336	0.712	14	1.077
技术	车辆数据采集与分析	0.044	0.867	0.317	0.764	13	1
市场	领先企业与技术革新	0.024	0.813	0.296	0.822	12	0.923
市场	市场动态与增长趋势	0.024	0.813	0.296	0.822	12	0.923
市场	政策驱动与基础设施建设	0.004	0.684	0.242	0.929	10	0.769
政策	税收优惠与车辆购置	0.004	0.684	0.242	0.929	10	0.769
技术	电机驱动与动力系统	0	0.619	0.181	1	8	0.615
技术	电池管理系统与检测技术	0.040	0.867	0.318	0.764	13	1
技术	可再生能源集成与储能技术	0.051	0.813	0.282	0.756	12	0.923
技术	电力电子与电路设计	0.051	0.813	0.282	0.756	12	0.923
政策	证券市场与投资管理	0.024	0.722	0.226	0.821	10	0.769
技术	无线供电与导轨技术	0.015	0.684	0.196	0.857	9	0.692
技术	充电设施设计与布局	0	0.520	0.104	1	6	0.462
技术	智能电网与充电管理	0.040	0.867	0.318	0.764	13	1

表 11 T3 时期“技术-市场-政策”主题共现网络指标

主题来源	标签	中间中心性	接近中心性	特征向量中心性	聚集系数	权重	标准化点度中心性
市场	消费市场与政策支持	0.035	0.909	0.233	0.882	20	1
市场	产业发展与国家战略	0.006	0.909	0.239	0.935	20	1
技术	负荷调度与优化建模	0.006	0.909	0.239	0.935	20	1
市场	市场动态与销售趋势	0.002	0.870	0.231	0.978	19	0.95
市场	绿色发展与低碳战略	0.002	0.870	0.231	0.978	19	0.95
市场	疫情防控与政策措施	0.028	0.870	0.221	0.890	19	0.95
政策	节能与新能源汽车优惠政策	0.002	0.870	0.231	0.978	19	0.95
技术	通信平台与故障处理	0.002	0.870	0.231	0.978	19	0.95
技术	油门控制和驾驶踏板技术	0.113	0.870	0.210	0.794	19	0.95
市场	企业业绩与资本动态	0.006	0.909	0.239	0.935	20	1
市场	成本上涨与价格波动	0.006	0.909	0.239	0.935	20	1
市场	生产能力与技术进步	0.006	0.909	0.239	0.935	20	1



政策	行政许可与机动车辆生产	0.006	0.909	0.239	0.935	20	1
技术	动力电池状态和健康管理	0.006	0.909	0.239	0.935	20	1
技术	电池安全与防护技术	0.004	0.833	0.228	0.949	19	0.95
技术	热管理系统	0.003	0.800	0.216	0.958	18	0.9
政策	购置税免征政策	0	0.645	0.152	1	13	0.65
市场	国际市场与供应链挑战	0.006	0.909	0.239	0.935	20	1
政策	税务政策与技术支持	0.006	0.909	0.239	0.935	20	1
技术	电机驱动与整车控制系统	0.006	0.541	0.040	0.667	6	0.3
技术	车辆状态监控与同步技术	0	0.488	0.015	1	4	0.2

#### 4.4.2 多源主题融合

在各个时间窗口下的主题相似性矩阵的基础上，通过 Community 库，运用 Louvain 社区发现算法识别主题社区，将相似的主题分组在一起，以融合技术、市场、政策三个数据源的颠覆性技术主题。为保证合理的社区规模，笔者将社区内部主题节点最小值设置为 2，最终得到 3 个关系较为紧密的主题社区（模块度  $Q>0.47$ ），如图 17 所示。

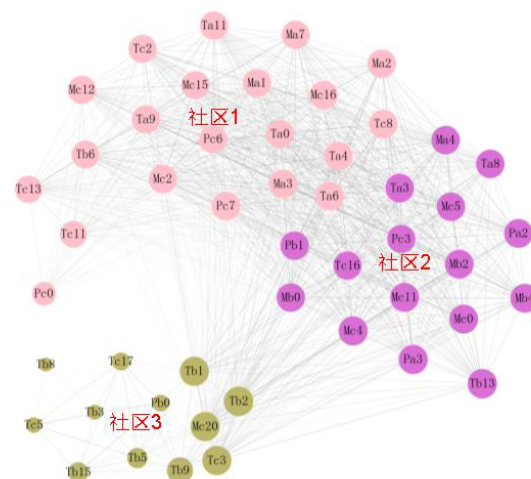


图 17 技术、市场、政策在所有时间窗口下的主题融合

根据社区中包含的主题词，对这 3 个主题社区进行命名表示。社区 1 为“新能源汽车全产业链创新与生态构建”，该社区包含了从产业合作、政策支持、产品应用到动力电池及电气系统等多个方面的内容。社区 2 为“新能源汽车技术进步与市场发展驱动”，该社区聚焦于产业规划、政策导向、市场动态、技术革新以及绿色发展战略。社区 3 为“新能源汽车智能化与能源管理技术革新”，该社区突出了智能电网、可再生能源集成、无线供电、电力电子以及车辆状态监控等智能化与能源管理方面的颠覆性技术。

#### 4.4.3 主题社区纵向演化分析

融合后的主题社区成为新聚类，其中的特征词为原主题词。根据原主题所在的时间窗口以及主题之间的相似度，借助 VOSviewer 软件绘制出多源信息融合主题下的网络演化图，

如图 18 所示。

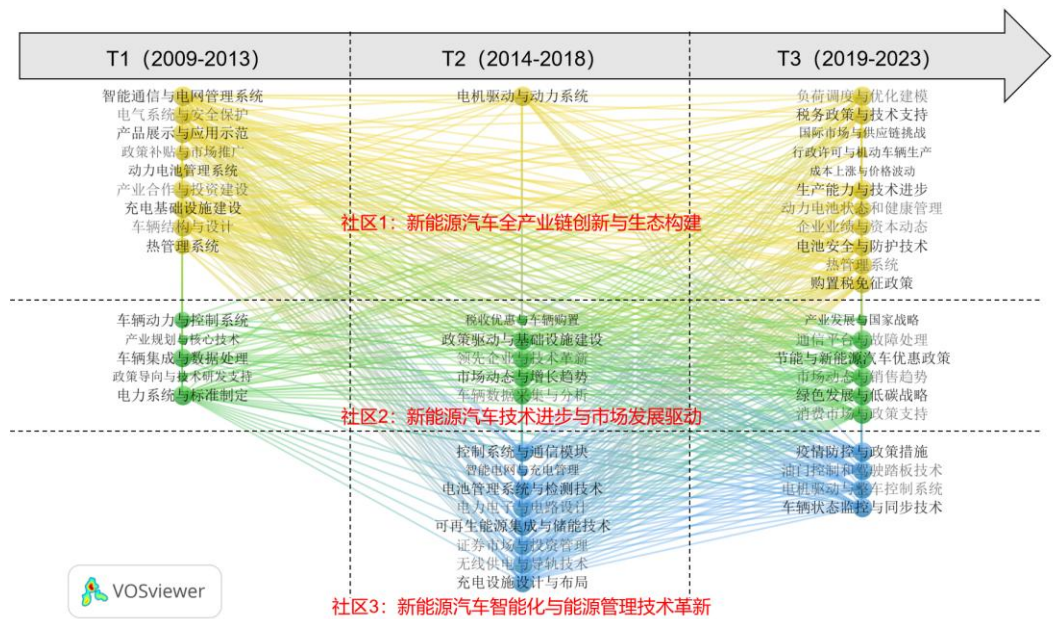


图 18 多源主题融合演化网络

## 1. 新能源汽车全产业链创新与生态构建

### (1) 基础设施建设与政策引导阶段（2009-2013 年）

此阶段的总链接强度普遍较高，表明早期新能源汽车产业的发展重点在于基础设施建设和政策推广，如“产业合作与投资建设”、“政策补贴与市场推广”、“充电基础设施建设”等。

### (2) 核心技术突破阶段（2014-2018 年）

这一阶段的重心转向了“电机驱动与动力系统”，其链接强度虽相对单一，但重要性显著，说明行业开始深入核心技术的研发，以提升新能源汽车的动力性能和效率，标志着从基础设施建设转向了技术深化。

### (3) 市场成熟与政策优化阶段（2019-2023 年）

最后阶段，主题变得更加多元化且深入，涉及“企业业绩与资本动态”、“成本上涨与价格波动”、“国际市场与供应链挑战”等市场运营的多个方面，显示行业正经历从政策驱动到市场驱动的转变。同时，“生产能力与技术进步”、“税务政策与技术支持”、“电池安全与防护技术”、“动力电池状态和健康的管理”等技术主题与政策主题的加强，说明产业开始关注生产效率、产品质量、安全性能及政策环境的持续优化。表明产业正在向更全面供应链管理方向发展，同时也在积极构建一个更加完善和高效的产业生态系统。

## 2. 新能源汽车技术进步与市场发展驱动

### (1) 技术与规划奠基阶段（2009-2013 年）

这一阶段的主题重点落在了“产业规划与核心技术”、“政策导向与技术研发支持”以及“电力系统与标准制定”等方面，权重较高，表明早期行业发展的重点在于构建技术基础

和规划框架。同时“车辆动力与控制系统”和“车辆集成与数据处理”的重要性也凸显出来，意味着技术进步与车辆智能化开始起步。

## (2) 政策激励与市场响应阶段（2014-2018 年）

这一时期，政策继续扮演关键角色，但市场因素开始显现其影响力。“政策驱动与基础设施建设”仍保持高权重，表明政府在推动基础设施建设的同时，也开始通过“税收优惠与车辆购置”等措施直接刺激市场需求。“市场动态与增长趋势”成为新的关注点，反映出行业正逐渐由政策推动向市场需求拉动转变。此外，“领先企业与技术革新”、“车辆数据采集与分析”等主题的出现，意味着技术创新和数据分析对于企业竞争力的重要性日益增加。

## (3) 市场成熟与国家战略融合阶段（2019-2023 年）

最后阶段，市场发展更加成熟，国家战略层面的考量成为主导。“市场动态与销售趋势”、“绿色发展与低碳战略”以及“消费市场与政策支持”成为核心主题，显示出市场机制的完善、环保理念的普及以及政策对消费市场的直接作用。同时，“产业发展与国家战略”、“节能与新能源汽车优惠政策”和“通信平台与故障处理”等主题的强化，表明行业已进入全面发展期，不仅关注经济效益，更注重与国家绿色发展战略的对接，以及智能化水平的提升和用户体验的优化。

# 3. 新能源汽车智能化与能源管理技术革新

## (1) 初始孕育期（2009-2013 年）

此阶段没有明显的颠覆性技术主题，但根据新能源汽车领域的一般发展规律，这一时期可以理解为新能源汽车智能化与能源管理技术的初始孕育阶段。在这个阶段，行业的各类基础技术开始积累，但尚未形成明显的焦点话题或大规模的应用趋势。

## (2) 技术萌芽与初步发展期（2014-2018 年）

此阶段“控制系统与通信模块”的权重最高，反映智能化控制和车辆间通信技术的兴起，这是新能源汽车迈向智能化的关键步骤。“智能电网与充电管理”和“电池管理系统与检测技术”的高权重，表明此时期行业内开始重视能源管理和优化电池性能，通过智能化手段提高充电效率和电池寿命。“可再生能源集成与储能技术”和“电力电子与电路设计”的出现，预示着行业对绿色能源利用和高效能源转换的关注增加。

## (3) 技术深化与新挑战应对期（2019-2023 年）

此阶段，“疫情防控与政策措施”的权重最高，显示出这一阶段行业需应对全球公共卫生事件带来的新挑战，如何确保生产、销售及服务流程在疫情期间的连续性和安全性成为关注点。“电机驱动与整车控制系统”和“油门控制和驾驶踏板技术”则表明在车辆核心驱动系统和驾驶操控技术上的持续优化和创新。“车辆状态监控与同步技术”的出现，意味着智能化管理与维护技术的提升，通过实时监控车辆状态来保障运行安全和效率。

# 4.5 有效性验证

综上所述，通过多源融合主题演化分析主要获得三个新能源汽车产业领域颠覆性技术主题的信息，由于目前没有权威机构公开的该领域颠覆性技术演化路径数据集，难以利用召回率、准确率、精确率等定量的指标对分析结果进行检验，因此本文根据国务院印发的《新能

源汽车产业发展规划（2012-2020 年）》、《新能源汽车产业发展规划（2021-2035 年）》等国家规划和政策，对我国新能源汽车产业的关键核心方向进行了深度研判，以确保识别结果符合国家发展战略的颠覆性技术。在此基础上，笔者邀请了新能源汽车产业领域的权威专家对拟定的演化路径进行评审。专家团队包括新能源汽车领域的知名教授、设计师、以及拥有丰富产业经验的企业高管。经过一个月的评审与讨论，专家团队一致认可了本文的演化分析结果，认为其准确反映了行业技术发展的历史脉络与未来趋势。专家们特别强调了几个关键点：①从智能电网与充电管理到电池管理系统，再到车辆智能化与状态监控技术的发展，这些技术趋势与国家规划中的关键技术突破方向高度契合，显示出良好的前瞻性和可行性。②技术演进路径与国家政策导向之间紧密联系，尤其是在《新能源汽车产业发展规划》等政策框架下，技术进步与政策支持相辅相成，共同推动了行业的快速发展。③技术演进与市场需求的变化也紧密相连，如疫情防控期间对政策措施的灵活调整，以及对电机驱动与整车控制系统的重视，都体现了行业对市场反馈的快速响应。

## 5 结论

颠覆性技术作为引领科技革命和产业重构的核心力量，在我国实施创新驱动发展战略的过程中发挥着不可替代的作用，因此构建更加全面、深层次的颠覆性技术研究体系至关重要。本文以技术、市场、政策为多重数据源，根据技术生命周期理论划分时间窗口；基于 BERTopic 主题模型结合颠覆性技术特征指标，识别颠覆性技术主题；通过构建主题共现网络展开主题关联横向演化，再借助 Louvain 社区发现算法实现多源主题融合，并构建多源融合主题演化网络展开主题社区纵向演化，全面分析颠覆性技术主题演化过程。本文以新能源汽车产业领域为例进行实证分析，结合政策解读验证了该方法的可行性。实验结果表明，优化后的 BERTopic 模型通过结合聚类评估指标和 K-means 算法，能够识别出比单独使用 HDBSCAN 算法自动聚类簇类更多且重叠度低的聚类，提升了聚类结果的准确性和细粒度。同时 Louvain 算法在应对多数据源主题社区划分的情况下依然能保持较高的处理效能。

本文的研究方法具备以下优势：①通过整合多重数据资源，打破了传统单一数据源分析的局限，有助于促进颠覆性技术的全面识别、演化和预测；②利用语义信息保留度更高的 BERTopic 模型进行文本挖掘，并在该模型的基础上引入聚类评价指标和 K-means 算法二次优化主题聚类效果，提升主题聚类准确性、细粒度和可解释性；③在演化分析中，不仅横向探索了主题之间的相互关系，还纵向剖析了多源融合主题社区的形成与演变机制，为深入理解颠覆性技术的演化路径提供了新视角。

然而，受限于笔者的研究能力，本文尚有若干不足之处：①尽管本文已涵盖主要数据类型，但仍存在未纳入的潜在重要数据源，如社交媒体、学术文献等，可能影响分析的全面性。②随着技术快速迭代，现有的 BERTopic 模型可能需要不断优化调整，以适应新出现的语言表达和专业术语变化。③在主题提取和颠覆性技术主题识别方面，本文还存在一定的主观性和不确定性。

## 参考文献

- [1] Christensen C M. The Innovator's Dilemma: When New Technologies Cause Great Firms to Fail [M]. Boston, MA: Harvard Business School Press, 1997: 9-10.
- [2] 李晓龙, 鲁平, 李存斌. 基于 Delphi 和 DEMATEL 法影响国网的颠覆性创新技术影响因素综合排序分析[J]. 科技管理研究, 2017, 37(06): 127-133.
- [3] 卢光松, 卢平. 技术路线图与颠覆性技术创新[J]. 科技进步与对策, 2011, 28(03): 11-15.
- [4] 王知津, 周鹏, 韩正彪. 基于情景分析法的技术预测研究[J]. 图书情报知识, 2013, (05): 115-122. DOI: 10.13366/j.dik.2013.05.001.
- [5] 许佳琪, 汪雪锋, 陈虹枢, 等. 跨领域颠覆性技术主题识别研究——以脑科学技术为例[J]. 图书情报工作, 2024, 68(15): 44-57. DOI: 10.13266/j.issn.0252-3116.2024.15.004.
- [6] Firat A K, Woon W L, Madnick S. Technological forecasting-a review[R/OL]. Cambridge: Massachusetts Institute of Technology, (2008-09-01). <http://web.mit.edu/smadnick/www/wp/2008-15.pdf>.
- [7] 于光辉, 宁钟, 李昊夫. 基于专利和 Bass 模型的颠覆性技术识别方法研究[J]. 科学学研究, 2021, 39(08): 1467-1473+1536.
- [8] 武建龙, 刘禹彤, 陈劲, 等. 基于专利挖掘和 Gompertz 模型的颠覆性技术识别方法研究[J]. 科研管理, 2024, 45(04): 62-72.
- [9] 熊焰, 张凌恺, 陈旭, 等. 基于“突变—演化”模型的颠覆性技术识别方法及应用[J]. 情报杂志, 2023, 42(12): 119-126+152.
- [10] Momeni A, Rost K. Identification and monitoring of possible disruptive technologies by patent-development paths and topic modeling[J]. Technological Forecasting and Social Change, 2016, 104: 16-29.
- [11] 苑朋彬, 邢晓昭. 专利知识流动视角下的颠覆性技术方向识别研究——以 6G 太赫兹通信技术领域为例[J]. 情报杂志, 2023, 42(11): 142-146.
- [12] 纪亚琨, 余翔, 张奔, 等. 专利网络视角下的潜在颠覆性技术识别——以自动驾驶领域为例[J]. 情报杂志, 2022, 41(12): 46-50+139.
- [13] Dahlin B K, Behrens M D. When is an invention really radical?[J]. Research Policy, 2005, 34(5).
- [14] 王康, 陈悦. 技术融合视角下基于专利的颠覆性技术识别研究[J]. 情报杂志, 2022, 41(04): 29-36+134.
- [15] 赵玉桐, 杨建林. 基于跨领域专利的颠覆性技术识别研究——以人工智能领域为例[J]. 情报理论与实践, 2023, 46(03): 174-182.
- [16] 陈育新, 李健, 韩毅. 核心—边缘理论视角下的颠覆性技术识别研究[J]. 情报理论与实践, 2022, 45(08): 121-129.
- [17] 单晓红, 韩晟熙, 刘晓燕. 基于技术主题演化的颠覆性技术识别研究[J]. 情报理论与实践, 2023, 46(08): 113-123.
- [18] BLEI D M, NG A Y, JORDAN M I. Latent dirichlet allocation[J]. Journal of machine Learning research, 2003, 3: 993-1022.
- [19] WANG X, MCCALLUM A. Topics over time: a non-markov continuous-time model of topical trends[C]//Proceedings of the 12th ACM SIGKDD international

- conference on knowledge discovery and data mining.New York:ACM,2006:424-433.
- [20] MOODY C.Mixing dirichlet topic models and word embeddings to make LDA2Vec[J].ar Xiv preprint,ar Xiv:1605.02019,2016.
- [21] 王秀红,高敏.基于 BERT-LDA 的关键技术识别方法及其实证研究——以农业机器人为例[J].图书情报工作,2021,65(22):114-125.
- [22] 曹琨,吴新年,白光祖,等.基于“科学-技术”复杂网络的关键核心技术识别研究——以数控机床领域为例[J/OL].数据分析与知识发现,1-18[2024-08-28].<http://kns.cnki.net/kcms/detail/10.1478.G2.20240506.1342.002.html>.
- [23] GROOTENDORST M. BERTopic: neural topic modeling with a class-based tf-idf procedure [J]. arXiv preprint arXiv:2203.05794.
- [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/OL].情报科学,1-21[2024-08-28].<http://kns.cnki.net/kcms/detail/22.1264.G2.20240726.1726.010.html>.
- [25] DOSI G.Technological paradigms and technological trajectories:asuggested interpretation of the determinants and directions of technical change[J].Research Policy,1982,11:147-162.
- [26] 范维熙,费钟琳.基于德温特专利引文网络的技术演进路径研究——以太阳能电池技术为例[J].情报杂志,2014,33(11):62-66.
- [27] YOON J,JEONG B,LEE W H,et al.Tracing the evolving trends in electronic skin(e-skin)technology using growthcurve and technology position-based patent bibliometrics[J].IEEE Access,2018,6:26530-26542.
- [28] 李雪,王婧朱,刘心蕊,等.基于时序关联和价值评估的技术主题演化研究——以生物制药产业为例[J].数字图书馆论坛,2023,19(06):22-30.
- [29] 冉从敬,田文芳,贾志轩.技术生命周期视域下专利技术主题演化分析方法研究——以视频图像处理技术领域为例[J/OL].情报理论与实践,1-15[2024-08-28].<http://kns.cnki.net/kcms/detail/11.1762.G3.20240327.1033.004.html>.
- [30] 李雪思,张智雄,王宇飞,等.领域知识演化分析方法综述[J].数据分析与知识发现,2024,8(01):1-15.
- [31] 王曰芬,王金树,关鹏.主题-主题关联的学科知识网络构建与演化分析[J].情报科学,2018,36(09):9-15+102.
- [32] 邢晓昭,任亮,雷孝平,等.基于专利主题演化的颠覆性技术识别研究——以类脑智能领域为例[J].情报科学,2023,41(03):81-88.
- [33] 许海云,王超,陈亮,等.颠覆性技术的科学-技术-产业互动模式识别与分析[J].情报学报,2023,42(07):816-831.
- [34] 黄鲁成,蒋林杉,吴菲菲.萌芽期颠覆性技术识别研究[J].科技进步与对策,2019,36(1):10-17.
- [35] 王海军,于佳文.基于专利发展路径的颠覆性技术识别:以智能语音领域为例[J].科技管理研究,2022,42(6):170-181.
- [36] 谭晓,西桂权,苏娜,等.科学-技术-项目联动视角下颠覆性技术识别研究[J].情报杂志,2023,42(2):82-91.
- [37] 冯倩倩,张光宇,戴海闻,等.颠覆性技术遴选的指标体系与流程设计——基于扎根理论的多案例研究[J].科技管理研究,2021,41(24):50-59.

## 作者简介

吕鲲，副教授，博士，硕士生导师，E-mail: lvkun@nbu.edu.cn；傅元美，在读本科生；  
鲍泽孟鸿，硕士研究生。