

## 热度演化视角下新兴主题识别分析研究 \*

■ 万校基<sup>1,2</sup> 李海林<sup>1,3</sup> 何雨晴<sup>1</sup> 杨润奇<sup>1</sup> 林海龙<sup>1</sup><sup>1</sup> 华侨大学工商管理学院 泉州 362021<sup>2</sup> 华侨大学东方企业管理研究中心 泉州 362021<sup>3</sup> 华侨大学现代统计与大数据研究中心 厦门 361021

**摘要:** [目的/意义] 识别领域新兴主题, 不仅可为科研人员的科学选题提供意见指导, 亦可为战略性新兴产业发展规划的精准制定提供技术支持。[方法/过程] 针对新兴主题概念界定不统一、识别时高维数据复杂难处理等问题, 提出一种基于主题热度演化的新兴主题识别方法。结合关键词重要性、平均相似性系数和近邻传播聚类算法自适应识别领域研究主题。以主题新颖度、主题影响力和主题潜力表征新兴主题, 借助天际线算法和主成分分析法获取主题热度排名时间序列, 通过主题热度排名的时间演化趋势识别新兴主题。[结果/结论] 通过对中国知网 2012—2021 年供应链领域期刊论文进行数据处理和挖掘, 本研究表明, 所提出的方法可以有效地识别该领域新兴主题, 这为相关科研人员和政府部门的选题和规划制定提供了决策支持。

**关键词:** 新兴主题 AP 聚类 天际线算法 主题热度 热度演化**分类号:** G251**CSTR:** 32305.14.CN11-1541.2024.22.012**DOI:** 10.13266/j.issn.0252-3116.2024.22.012

**引用本文:** 万校基, 李海林, 何雨晴, 等. 热度演化视角下新兴主题识别分析研究 [J]. 图书情报工作, 2024, 68(22): 126-138. (Citation: Wan Xiaoji, Li Hailin, He Yuqing, et al. Research on the Identification and Analysis of Emerging Topics from the Perspective of Popularity Evolution[J]. Library and Information Service, 2024, 68(22): 126-138.)

## 1 引言 /Introduction

大数据时代, 海量数据为科学研究和国家发展带来了重大机遇, 然而, 面对海量科技文献数据, 科研人员难以及时有效地获取领域研究前沿和最新研究进展。与此同时, 国家及地方政府在制定战略性新兴产业发展规划时, 也面临如何甄选符合时代和国家需要的新兴产业困境。为此, 迫切地需要对相关领域的新兴主题进行识别与分析。

目前学术界对新兴主题的概念缺乏统一的认识, 各项研究关注的视角和特性也不同。此外, 部分研究在识别新兴主题时存在数据高维复杂难处理、识别粒度不足和受限于专家个人知识积累和偏好等问题。本文将主题新颖度、主题影响力和主题潜力表征领域新兴主题, 并结合文献计量、近邻传播聚类算法和天际线算法对其进行识别。首先, 通过词频分析法、关键词权重公式和 Ochiai 系数构建加权高频关键词相似性矩阵; 其次, 借助近邻传播 (affinity

propagation, AP) 聚类算法和平均相似性系数细粒度识别领域研究主题; 然后, 以细粒度主题的新颖度、影响力和潜力为基础, 通过天际线算法和主成分分析法获取主题热度排名时间序列; 最后, 通过分析主题热度的演化趋势识别领域新兴主题。

本文的主要贡献体现在: ①充分考虑新兴主题概念、内涵、特征, 从主题新颖性、影响力和潜力三方面来表征新兴主题; ②度量不同关键词在文献中的重要性, 并结合 Ochiai 系数、平均相似性系数和近邻传播聚类算法细粒度识别领域研究主题; ③结合领域研究主题特征和天际线算法构建多维度主题天际线集合, 并通过分析主题热度排名演化趋势精准识别新兴主题。

## 2 文献综述 /Literature review

### 2.1 新兴主题定义

新兴主题一般也被称为新兴研究话题和新兴研究趋势, 与前沿主题不同的是, 该类型主题在近期出

\* 本文系福建省社会科学规划项目“基于大数据分析的国家自然科学基金结题项目产出绩效评估研究”(项目编号: FJ2023B109)研究成果之一。

**作者简介:** 万校基, 副教授, 博士, 硕士生导师; 李海林, 教授, 博士, 博士生导师, 通信作者, E-mail: hailin@hqu.edu.cn; 何雨晴, 硕士研究生; 杨润奇, 本科生; 林海龙, 本科生。

**收稿日期:** 2024-02-28 **修回日期:** 2024-06-14 **本文起止页码:** 126-138

版权所有 ©《图书情报工作》杂志社有限公司, 未经许可不得转载 (Copyrights © LIS Press Co., Ltd. Reproduction is prohibited without permission)

现但并没有得到广泛关注,具有较大发展潜力,未来有可能成为前沿主题,但也有一些主题可能演化为非前沿性热点主题或衰老主题<sup>[1]</sup>。现有研究对新兴主题的概念并没有达成共识,不同研究者有不同的认识,如新兴主题概念的最早提出者 N. Matsumura 等<sup>[2]</sup>认为具有未来发展潜力的主题是新兴主题,而 D. Rotolo 等<sup>[3]</sup>认为新兴主题是一种全新的、发展速度相对较快的主题,具有新颖性、增长性和持续性等特征。国内学者方面,李静等<sup>[4]</sup>认为基金项目的新兴主题具备潜在阶段和突破阶段两个主题周期阶段点,并将潜在阶段低于平均水平但之后超过平均水平的主题定义为新兴主题。黄鲁成等<sup>[5]</sup>认为新兴主题具有高关注度、高成长潜力度和高关联度特征,而杨金庆和张力<sup>[6]</sup>认为学科交叉融合是新兴主题形成的内在驱动力之一。

## 2.2 新兴主题识别方法

在新兴主题识别方法方面,现有研究主要有基于定性分析的识别、基于文献计量的识别和基于文本挖掘的识别。

### 2.2.1 基于定性分析的识别

技术路线图法<sup>[7]</sup>和专家意见法<sup>[8]</sup>是常见识别新兴主题的定性方法,其中技术路线图可以描绘某一技术领域内发展轨迹、预期目标以及实现目标所需步骤和资源,通过深入分析这些内容可以进一步识别那些正在形成或即将形成的新兴主题,该方法主要关注技术层面发展,缺乏对经济、社会、政策等方面因素考虑。专家意见法主要是通过专家经验和知识来对领域新兴主题进行识别,主要有两种应用模式,一种是先借助相关指标对主题进行初步筛选,再由专家进行最终判定<sup>[9]</sup>;另外一种是先让专家组确定候选主题,再依据相关指标识别新兴主题<sup>[10]</sup>。该方法专家依赖性较为明显,耗时较长,并且通用性和可复制性较弱<sup>[11]</sup>。

### 2.2.2 基于文献计量的识别

文献计量分析法是一种基于数学和统计学的定量分析方法,主要是通过对论文、专利等文献数据的数量、质量、引用、机构和作者等指标进行评估和分析,进而揭示相关学科领域研究现状和发展趋势。现有研究者主要采用共现分析法、引文分析法、共被引分析法、文献耦合分析法等识别领域新兴主题。

共现分析一般是通过分析文献中不同词汇或实体共同出现的频率,进而识别相关领域热点、研究趋势或学科交叉点。如冯佳和张云秋<sup>[12]</sup>借助 g 指数确定高频词汇,并对呈增长趋势的高频词汇进行共现分析,最后通过 VOSviewer 聚类识别了数据挖掘领域新

兴主题。引文分析主要是通过对期刊、专利等文献数据之间的引用关系进行分析,进而评估文献学术影响力,揭示知识流动规律。如 H. Sasaki 等<sup>[13]</sup>借助引文网络和主题模型识别了纳米碳领域新兴主题。共被引分析是通过研究两篇或多篇文献被其他文献共同引用的次数,以揭示它们之间知识关联。如 C. Chen<sup>[14]</sup>在新开发的 CiteSpace 软件中整合了突变检测算法和共被引分析法,并对由共被引和施引文献构建的网络进行了深入分析,进而探测了科学文献中新兴趋势和瞬态模式。文献耦合分析主要关注施引文献,通过比较不同文献所引用的共同参考文献的数量和类型,可以分析这些文献之间关联程度和相似性。如 W. Glanzel 和 B. Thijs<sup>[15]</sup>借助文献耦合分析法和文本相似性算法获取了文档集合中核心主题,并将文献集的异常增长、新生和研究内容的转移视为了新兴主题 3 种范式。此外,部分研究者还融合了引文指标和 alt-metrics 指标<sup>[16]</sup>、战略分析图<sup>[17]</sup>完成了对新兴主题的识别。

当前基于文献计量的新兴主题识别方法不需要构建复杂模型,操作较为简单,然而该类方法对文档语义和语境关注不足,难以充分解释文档与主题、主题与词汇间的隐含关系和语义联系<sup>[1]</sup>。

### 2.2.3 基于文本挖掘的识别

文本挖掘是一种能够对海量文本数据进行信息或知识抽取的方法,通过该方法可以自动挖掘非结构化文本中隐含的语义信息,从而最终实现主题识别<sup>[1]</sup>。目前该类方法在识别新兴主题方面得到了较为广泛的应用<sup>[18]</sup>。现有研究主要是通过概率突发和关联规则、SAO (subject-action-object) 结构、LDA (latent Dirichlet allocation) 及其改进模型 PLDA (parallel latent Dirichlet allocation) 等识别新兴主题。

概率突发通常是指某个事件或话题在短时间内发生的频率超过它的平均水平,有助于识别出突然增加的事件或话题,而关联规则是数据挖掘中一种常用的技术,通过计算支持度和置信度可以找到与新兴主题相关话题、词汇或事件。如 M. Xu 等<sup>[19]</sup>以图书情报学为例,首先利用 LDA 模型生成主题并计算概率,然后通过检测连续时间跨度中概率突发来发现新出现的主题,最后结合关联规则和词汇相似度来识别新兴主题。SAO 结构一般描述文本中三元组关系,主要包括 subject (主语)、action (动作) 和 object (宾语) 三元素,通过识别和分析文本中 SAO 结构,不仅有助于从文本中提取结构化信息,亦可深入地理解文本中表达的语义。如黄鲁成等<sup>[20]</sup>通过 CiteSpace 和

二次检索确定突现文献施引文献,借助 AlchemyAPI 自然语言处理工具提取对应文献摘要 SAO 结构,最后结合语义相似度计算程序和多维尺度分析法识别新兴主题。LDA 是一种基于贝叶斯学习的话题模型,它将每一篇文档视为一个词频向量,能够识别大规模文档集或语料库中隐藏的主题信息。如白敬毅等<sup>[21]</sup>借助 LDA 主题模型和多维尺度分析法探测了燃料电池领域新兴主题,并通过曲线拟合分析法预测了新兴主题未来发展趋势。作为 LDA 改进模型,PLDA 模型在大型文档集合应用中具有良好的可扩展性,能够更好地揭示词语的语义关联关系,提升主题识别效率和精确性<sup>[22]</sup>。如张家惠和丁敬达<sup>[8]</sup>以文化遗产领域中的期刊论文、学位论文、会议文献、基金项目等多源数据为研究对象,利用 PLDA 识别这些数据源文献主题,并结合 VSM (vector space model) 模型和主题新颖度、增长率、关注度等文献计量指标筛选该领域新兴主题。

现有基于文本挖掘的新兴主题识别方法能够对以全局视角对大规模文本进行深层次分析,可以较为客观地探测出文本隐含的新兴主题。然而,多数主题模型仍然存在事先需要确定主题数量和数据高维复杂难处理等问题。

### 3 理论基础 /Theory foundation

#### 3.1 近邻传播聚类算法

近邻传播聚类 (affinity propagation, AP) 算法是一种考虑了观测值之间的“信息传递”的聚类算法。该算法主要思想是将全部的数据点都看作是潜在的聚类中心 (exemplar), 然后通过数据点之间的相似性进行聚类, 不断迭代更新数据点之间的吸引度 (availability) 和归属度 (responsibility) 信息, 直至迭代达到收敛, 最优类代表集合和聚类中心形成。在该算法中, 代表程度  $r(i, k)$  和合适程度  $a(i, k)$  的具体计算公式为:

$$r(i, k) = s(i, k) - \max_{k' \neq k} \{s(i, k') + s(i, k')\} \quad \text{公式 (1)}$$

$$a(i, k) = \min \left\{ 0, r(k, k) + \sum_{i' \in \{i, k\}} \max \{0, r(i', k)\} \right\} \quad \text{公式 (2)}$$

$$a(k, k) = \sum_{i' \neq k} \max \{0, r(i', k)\} \quad \text{公式 (3)}$$

其中, 代表程度  $r(i, k)$  表示在考虑其他潜在聚类中心点之后,  $k$  点作为数据对象  $i$  所在簇的中心的程度; 合适程度  $a(i, k)$  表示在考虑了其他点对  $k$  点成为聚类中心的支持后, 数据对象  $i$  选择点  $k$  作为聚类中心的程度。

在信息更新过程中, 为减少振荡所带来的影响,

需引入衰减系数  $\lambda (0 < \lambda < 1)$ 。每条信息被设置为前一次迭代信息值的  $\lambda$  倍与本次信息更新值的  $(1-\lambda)$  倍的累加, 具体表达公式为:

$$r_{s+1}(i, k) = (1-\lambda) * r_{s+1}(i, k) + \lambda * r_s(i, k) \quad \text{公式 (4)}$$

$$a_{s+1}(i, k) = (1-\lambda) * a_{s+1}(i, k) + \lambda * a_s(i, k) \quad \text{公式 (5)}$$

鉴于 AP 聚类算法无需事先指定最终聚类个数, 可以自适应实现聚类, 本文将借助该算法细粒度识别领域研究主题, 该算法的具体流程如表 1 所示:

表 1 AP 聚类算法的流程  
Table 1 Procedure of AP clustering algorithm

AP 聚类算法
输入: 初始相似度矩阵 $S$ , 预设迭代次数 $M$ 以及衰减系数 $\lambda$
输出: 聚类集合 $A$
Step1: 初始化代表程度和合适程度为 $r(i, k) = a(i, k) = 0$ ;
Step2: 根据公式 (1) 至公式 (3), 计算样本点间代表程度 $r$ 和合适程度 $a$ ;
Step3: 通过公式 (4) 和公式 (5) 消除更新信息过程中的振荡;
Step4: 重复 Step2 和 Step3, 直到实际迭代次数大于预设迭代次数 $M$ 或聚类中心在迭代过程中连续不改变时终止计算;
Step5: 最优类代表集合 $A$ 和聚类中心形成。

#### 3.2 Skyline 算法

天际线 (skyline) 常被应用在多目标决策、数据挖掘和数据库可视化中, 早在 1975 年, H. Kung<sup>[23]</sup> 以最大向量问题呈现了该算法。实际上, skyline 是指从给定的一组多维空间数据集  $U$  中选择一个子集, 该子集中任意一点都不能被  $U$  中的其他点所支配, 这些不被支配的点被称为 skyline 点。所谓的支配关系<sup>[24]</sup>是指: 当且仅当  $P$  至少在其中一个维度优于  $Q$ , 而在其他维度不差于  $Q$ , 那么  $P$  支配  $Q$ 。因此, skyline 算法最终返回的是一系列互不支配, 平行的解。

鉴于本文以主题新颖度、主题影响力和主题潜力表征新兴主题, 为科学获取主题热度排名时间序列, 需构建三维天际线集合, 此时, 对应的天际线算法流程如表 2 所示:

表 2 skyline 算法的流程  
Table 2 Procedure of skyline algorithm

skyline 算法
输入: 点集 $V(x_1, x_2, x_3)$
输出: 各个天际线点集 $U_1, U_2, U_3, \dots, U_n$
Step1: 排列 $V$ , 使得 $x_1(v_1) > x_1(v_2) > \dots > x_1(v_n)$ ;
Step2: 初始化 $\max Y \leftarrow 0$ , $\max Z \leftarrow 0$ , 点集 $Sky$ 和点集 $Out$ 置为 $\emptyset$ ;
Step3: 扫描 $V$ ,
(1) 若 $x_2(v_i) > \max Y$ 且 $x_3(v_i) > \max Z$ , 则 $\max Y \leftarrow x_2(v_i)$ , $\max Z \leftarrow x_3(v_i)$ 并且跳转到 $V_{i+1}$ ;
(2) 若 $x_2(v_i) > \max Y$ , 则 $\max Y \leftarrow x_2(v_i)$ ; $Sky \leftarrow Sky \cup V_i$ 且跳转到 $V_{i+1}$ ;
(3) 若 $x_3(v_i) > \max Z$ , 则 $\max Z \leftarrow x_3(v_i)$ ; $Sky \leftarrow Sky \cup V_i$ 且跳转到 $V_{i+1}$ ;
(4) $Out \leftarrow Out \cup V_i$ ;
Step4: 输出 $Sky$ , $V \leftarrow Out$ ;
Step5: 重复 Step2~Step4, 直到 $V = \emptyset$ 。



## 4 研究思路和方法 /Research ideas and methods

### 4.1 研究思路

本文整体研究共分为两部分,第一部分为细粒度主题识别。通过词频分析、关键词权重和 Ochiia 系数构建加权高频关键词相似性矩阵,利用 AP 聚类算

法自适应获取初始主题簇,针对低于平均相似性的主题簇进行二次自适应 AP 聚类,进而得到细粒度主题。第二部分为新兴主题识别。提取细粒度主题新颖度、影响力和潜力等特征,借助天际线算法和主成分分析法构建主题热度排名时间序列,分析主题热度演化趋势以识别新兴主题。本文研究思路见图 1,下面将对这两部分内容进行具体介绍。

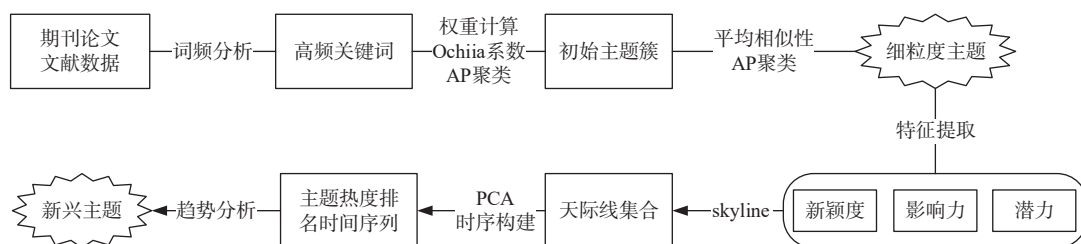


图 1 本文研究思路

Figure 1 Research ideas of this paper

### 4.2 细粒度主题识别

#### 4.2.1 初始主题识别

关键词作为描述一个文献主题的重要线索,是主题识别中不可缺少的部分。然而,传统的主题识别通常是以 Ochiia 系数进行关键词间的相似性度量,忽略了不同关键词在文献中的重要性。本文将充分考虑关键词在不同文献的重要性程度,通过文献计量和自适应 AP 聚类算法来实现初始主题识别,具体识别过程如下:

步骤 1: 确定高频关键词。借鉴万校基等<sup>[25]</sup>做法,选取频次超过 10 的关键词为高频关键词。

步骤 2: 构建加权高频关键词相似性矩阵。借鉴李海林等<sup>[26]</sup>做法,根据高频关键词在文献中出现的先后顺序来计算其重要性程度,并在 Ochiia 系数基础上构建加权高频关键词相似性矩阵,该过程涉及的公式为:

$$wKey_{pk} = \frac{K_p - K + 1}{\sum_{K=1}^{K_p} K} \quad \text{公式 (6)}$$

$$S_w(e, f) = \frac{\sum_{e, f \in Key_p} wKey_{pe} \times wKey_{pf}}{\sqrt{\sum_{e \in Key_p} wKey_{pe}^2} \sqrt{\sum_{f \in Key_p} wKey_{pf}^2}} \quad \text{公式 (7)}$$

$$S_w = \begin{bmatrix} Sw(1,1) & Sw(1,2) & \cdots & Sw(1,m) \\ Sw(2,1) & Sw(2,2) & \cdots & Sw(2,m) \\ \vdots & \vdots & \vdots & \vdots \\ Sw(m,1) & Sw(m,2) & \cdots & Sw(m,m) \end{bmatrix} \quad \text{公式 (8)}$$

其中,  $K_p$  为文献  $p$  中的关键词个数;  $wKey_{pk}$  表示文献  $p$  中第  $K$  个关键词对应的权重;  $S_w(e, f)$  为关键词

$e$  和关键词  $f$  的相似性;  $wKey_{pe}$  和  $wKey_{pf}$  分别为关键词  $e$  和关键词  $f$  共同出现在第  $p$  篇文献中对应的关键词权重;  $Key_p$  为关键词集合;  $G$  为文献数量。

步骤 3: 获取初始研究主题。以步骤 2 构建的加权高频关键词相似性矩阵为基础,借助如表 1 所示的 AP 聚类算法开展自适应聚类,以聚类簇的中心代表关键词表示所在簇的初始研究主题。

#### 4.2.2 细粒度主题识别

在首次聚类得到的部分主题簇中,其内部关键词成员之间的相似性偏低,以这些簇的中心代表关键词代表所在簇的研究主题将存在一定偏差,有必要对相似性偏低的主题簇进行二次聚类,具体操作过程如下:

步骤 1: 计算初始主题簇相似性及其平均相似性。通过初始主题簇中关键词成员之间相似性计算所在簇相似性,并以此为基础获取所有主题簇平均相似性,涉及的公式为:

$$AGVS_l = \frac{\sum_{e=1}^{n_l-1} \sum_{f>e}^{n_l} Sw(e, f)}{[n_l \times (n_l - 1)] / 2} \quad \text{公式 (9)}$$

$$AVGS = \frac{\sum_{l=1}^L AGVS_l}{L} \quad \text{公式 (10)}$$

其中,  $AGVS_l$  为第  $l$  个主题簇的相似性;  $n_l$  为第  $l$  个主题簇的关键词个数;  $L$  为初始主题簇个数;  $AVGS$  为所有主题簇的平均相似性。

步骤 2: 识别细粒度主题。针对相似性低于平均相似性的主题簇开展 AP 自适应聚类,以聚类后的簇中心代表关键词代表所在簇的研究主题,将其

与首次 AP 聚类中相似性较高的主题合并为领域细粒度主题。

### 4.3 新兴主题识别

#### 4.3.1 概念界定

通过文献梳理笔者发现,新兴是一个从形成到变得重要的过程<sup>[27]</sup>,新兴主题是具备成为热门主题潜力但尚未被广泛开展研究的主题,其出现时间较短,能够逐步引起关注并产生影响力,具有较大研究潜力<sup>[28]</sup>。为此,本文认为新兴主题具有明显新颖性、影响力和潜力等特征。为便于后续客观识别新兴主题,有必要对主题新颖度、主题影响力和主题潜力进行度量。

(1) 主题新颖度 (novelty index, NI)<sup>[29]</sup> 反映了主题的新颖性程度,当某主题刚出现时,其新颖度较高,随着时间变化,其新颖度将逐渐下降。本文将通过主题已出现时间的长短来计算主题新颖度<sup>[30-32]</sup>,即首先统计主题在各年份出现次数,然后以出现次数不为 0 的年份作为主题开始年份,最后通过如下公式计算主题新颖度:

$$NI_t^z = \frac{1}{t - FY + 1} \quad \text{公式 (11)}$$

其中,  $t$  代表当前年份;  $FY$  代表主题  $z$  开始出现的年份;  $NI_t^z$  代表  $t$  年主题  $z$  的新颖度。

(2) 主题影响力 (influence index, II) 反映了主题的影响能力,当某主题处于初期时,其影响力较弱,随着该主题不断受到关注,其影响力将逐渐增强<sup>[33]</sup>。本文将借鉴 Q. Wang 等<sup>[34]</sup>的做法,以主题相关论文的被引频次来度量主题影响力。

(3) 主题潜力 (potentiality index, PI) 反映了主题未来发展能力,一个主题是否具有发展成热点主题的潜力,很大程度取决于该主题在文本中的出现次数,反映了该主题的布局空间、研究热度和关注程度。本文将借鉴殷蜀梅<sup>[35]</sup>的做法,利用主题在文献中出现的次数来对主题潜力进行度量。

#### 4.3.2 识别步骤

本文将以细粒度主题的新颖度、影响力和潜力为基础,借助天际线算法、主成分分析法和主题热度演化趋势识别新兴主题,具体操作步骤如下:

步骤 1: 获取主题天际线集合。提取并计算每一年细粒度主题的新颖度、影响力和潜力,以其为基础,根据表 2 中的 skyline 算法获取每一年主题天际线集合。

步骤 2: 构建主题热度排名时间序列。逐年对主

题天际线集合进行主成分分析,从而得到同一天际线下不同主题热度排名  $TP$ , 汇总所有年份各主题热度排名信息,进而构建出所有主题热度排名时间序列。

步骤 3: 计算平均主题热度。对  $t$  年各主题热度排名  $TP$  进行平均处理,最终可得到如下平均主题热度:

$$ATP_t = \frac{\sum_{z=1}^N TP_t^z}{N} \quad \text{公式 (12)}$$

其中,  $ATP_t$  表示  $t$  年所有主题的平均热度排名,其综合反映了该年所有主题热度的平均水平;  $TP_t^z$  表示  $t$  年主题  $z$  的热度排名,综合反映了该主题在  $t$  年呈现出的新颖度、影响力和潜力等特征;  $N$  表示  $t$  年主题总数量。

步骤 4: 计算主题热度比。定义  $t$  年主题  $z$  的热度比为:

$$TPP_t^z = \frac{TP_t^z}{ATP_t} \quad \text{公式 (13)}$$

其中,  $TP_t^z$  和  $ATP_t$  同公式 (12) 定义。

显然,当  $TP_t^z < ATP_t$  时,  $TPP_t^z < 1$ , 说明主题  $z$  的热度低于平均水平;反之,则说明该主题热度高于平均水平。

步骤 5: 识别新兴主题。通过公式 (13) 分析所有主题热度比演化趋势,同时借鉴杨金庆等<sup>[36]</sup>做法,筛选出从  $TPP_t^z < 1$  到  $TPP_t^z > 1$  的主题,进而完成对新兴主题识别。

## 5 实验分析 /Experimental analysis

### 5.1 数据来源和处理

供应链在经济全球化背景下扮演着至关重要的角色,对于企业的运作效率和市场竞争力具有巨大的影响。因此,针对供应链领域开展相关研究具有重要的现实意义。鉴于海量期刊文献中隐含着目标领域研究现状和发展趋势,本文选取中国知网 (China National Knowledge Infrastructure, CNKI) 中收录于 SCI (Science Citation Index)、EI (Engineering Index)、CSSCI (Chinese Social Sciences Citation Index)、CSCD (Chinese Science Citation Database) 与供应链相关的期刊文献为实验对象,检索的主题词为“供应链”,来源时间为 2012—2021 年,数据收集时间为 2022 年 1 月 2 日。为排除不相关文献的影响,保证研究的可信度,剔除通知、征稿文件、专访和会议等文献,最终获得 9 638 篇有效文献。图 2 为供应链相关期刊文献的每年发表情况:

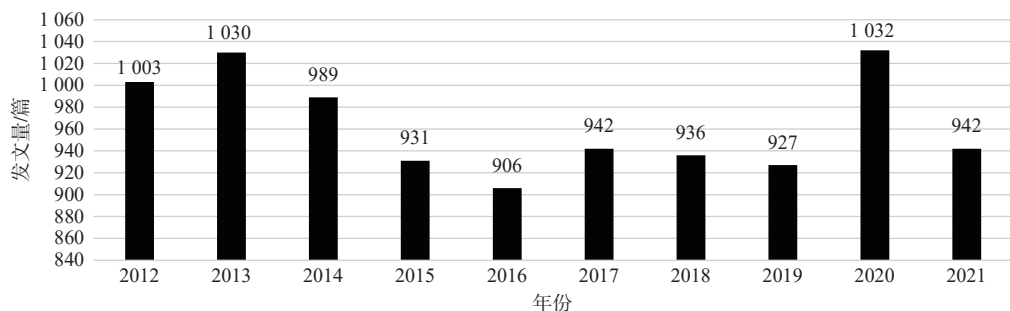


图 2 供应链相关期刊文献的每年发表情况

Figure 2 Annual publication volume of supply chain-related papers

由图 2 可知, 2012—2021 年, 供应链领域发表的论文数量整体近似呈 U 形变化, 说明学术界对供应链领域相关研究的热度近似呈先下降后上升趋势。此外, 2013 年和 2020 年的发文量略高于其他年份, 特别地, 发文量在 2020 年达到最高 (1 032 篇), 究其原因, 可能是 2019 年 12 月底新冠疫情暴发带来整个学术界对新冠肺炎疫情、应急物流、逆全球化和抗风险等相关研究的关注。

## 5.2 细粒度主题识别

### 5.2.1 初始主题识别

关键词是一篇论文主要内容和核心观点的精

炼, 能较好地概括文章主题。鉴于关键词具有较强的主题代表能力, 本文将用其来表示期刊文献主题。为便于后续初始主题识别, 借鉴万校基等<sup>[25]</sup>的做法, 选取频率超过 10 的 519 个高频关键词来展开分析。

由公式 (6) — 公式 (8) 可得到 519 阶加权关键词相似性矩阵, 通过 R 语言工具对其进行 AP 自适应聚类, 再借助 Gephi 软件对聚类结果进行可视化, 最终可得到如图 3 所示的 15 个初始主题簇:

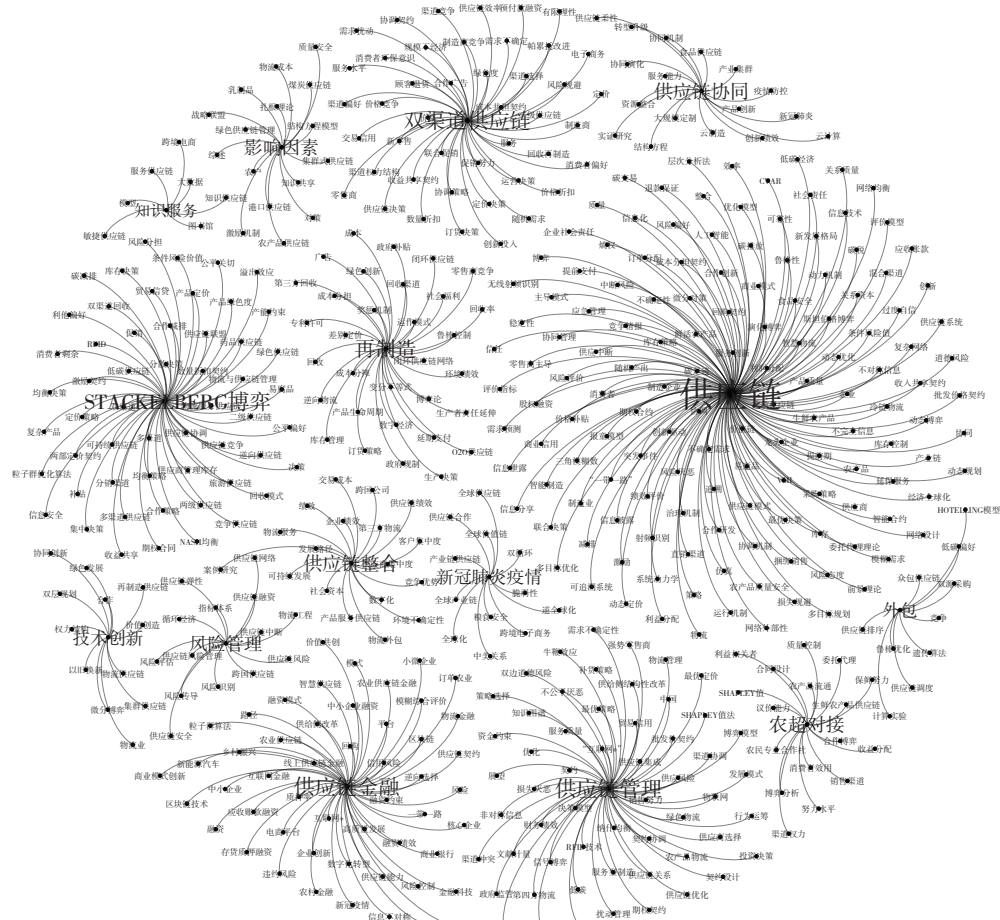


图 3 初次主题聚类结果

Figure 3 Initial topic clustering results



由图3可知,与同一个关键词相连的关键词成员同属于一个主题簇,每个簇中间的关键词与近邻的关键词联系密切,说明其是该主题簇的核心主题。通过AP聚类获得的15个初始核心主题分别为“供应链”“供应链管理”“供应链金融”“STACKELBERG博弈”“双渠道供应链”“再制造”“新冠肺炎疫情”“供应链整合”“影响因素”“风险管理”“供应链协同”“农超对接”“技术创新”“外包”和“知识服务”。

以初始核心主题“再制造”为例,其包含的关键词成员有“逆向物流”“闭环供应链网络”和“生产者责任延伸”等,从概念上可知这些关键词之间联系紧密。刘军军等<sup>[37]</sup>认为“逆向供应链”包括末端产品获取、逆向物流和再制造等环节,可持续运营管理的实践和研究对逆向供应链的关注在绿色供应链管理概念提出之后有所上升,其清晰揭示了“闭环供应链”“逆向物流”“绿色创新”和“再制造”这些关键词的密切关系。对于“供应链”为核心的主题簇,其簇内含有“不对称信息”“报童模型”和“质量管理”等关键词,从概念上来看联系并不太紧密,但王谦等<sup>[38]</sup>利用报童模型研究了非对称信息下“供应商—零售商”两级供应链的质量激励问题,刚好说明新问题的提出有可能将看似不紧密的主题联系起来。

为衡量首次AP聚类效果,通过公式(9)计算15个初始主题簇的平均相似性,最终可得到如表3所示结果:

表3 初始主题簇的平均相似性  
Table 3 Average similarity of initial topic clusters

主题簇序号	初始核心主题	AVGS <sub>i</sub>
1	供应链	0.016
2	供应链管理	0.037
3	供应链金融	0.043
4	STACKELBERG 博弈	0.039
5	双渠道供应链	0.049
6	再制造	0.068
7	新冠肺炎疫情	0.156
8	供应链整合	0.109
9	影响因素	0.122
10	风险管理	0.136
11	供应链协同	0.112
12	农超对接	0.114
13	技术创新	0.150
14	外包	0.179
15	知识服务	0.247

结合表3和公式(10)可知,所有初始主题簇的平均相似性为 $AVGS=0.105$ 。显然,针对相似性较低的主题簇,其簇内关键词成员之间相似性不高,即该

主题簇的中心代表关键词未能更好地反映成员关键词所蕴含的主题,因此,有必要对其进行二次聚类。

### 5.2.2 细粒度主题识别

由表3可知,初始核心主题“供应链”“供应链管理”“供应链金融”“STACKELBERG 博弈”“双渠道供应链”和“再制造”所在主题簇的平均相似性均低于平均相似性 $AVGS$ ,因此需要对其进行二次AP聚类。

类似首次AP聚类过程,对以上6个相似性较低的主题簇进行二次AP自适应聚类,最终可得90个主题簇。为便于后续主题特征提取,将其与首次聚类相似性较高的9个主题簇进行合并,最终可得到如图4所示的99个核心主题簇。

由图4可知,“风险管理”“农超对接”“知识服务”等中心代表关键词被分在了不同的核心主题簇中,而在同一个主题簇中的关键词往往具有较高的相关性,如在以“应急管理”为核心的主题簇中,含有“突发事件”“不对称信息”“鲜活农产品”等3个关键词成员,这些关键词有共同出现在吴忠和等<sup>[39]</sup>的研究中,该研究分析了鲜活农产品供应链在时间约束和不对称信息情景下应对突发事件的协调问题。

此外,笔者也发现,二次聚类后的大部分主题都由3个或4个关键词所刻画,说明在海量文献数据中,往往特定几个关键词之间的联系比较紧密,由此可以选取每个主题簇的中心代表关键词作为该主题簇的代表。以每个主题簇的中心代表关键词代表所在簇的主题,最终可得到99个细粒度主题。

## 5.3 新兴主题识别

### 5.3.1 主题热度排序

由于聚类后各个主题簇中关键词成员数量不一致,有可能导致含有较多成员的主题簇的整体新颖度、影响力和潜力偏高。为消除此影响,本文将以主题簇中心代表关键词的新颖度、影响力和潜力来表征所在簇的新兴主题。此外,由天际线算法得到的主题天际线集合,只是找出了在新颖度、影响力和潜力3个维度下的各主题集合热度,为消除多维数据特征影响并客观揭示各主题热度排名情况,还需要通过主成分分析法进行进一步降维。

以2021年数据为例,使用Python工具对表2中算法进行实现,一共可得到17条天际线集合。针对每个天际线集合,再次通过Python工具进行主成分降维,最终可得到2021年各主题热度排序结果,排名前10的主题及其对应的天际线集合见表4。

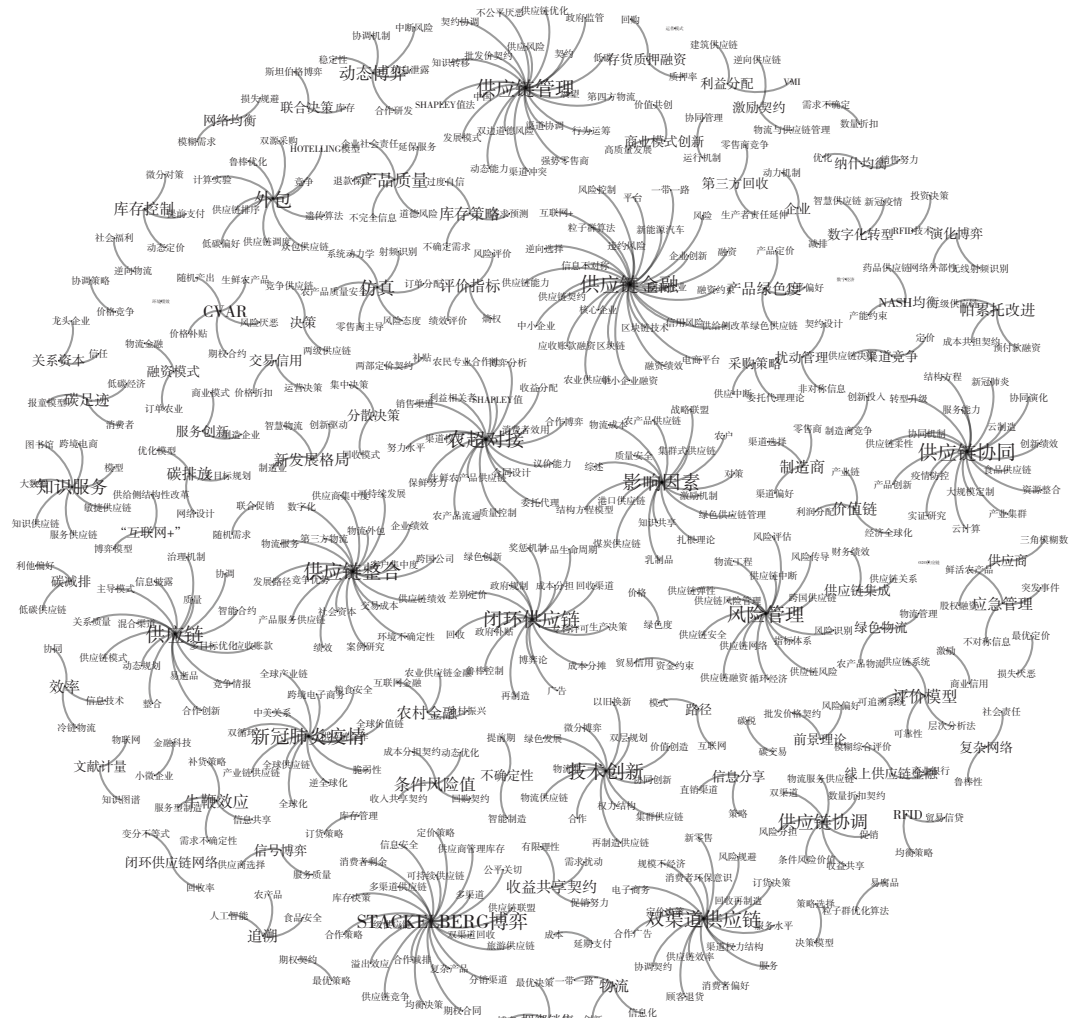


图 4 合并后的 99 个核心主题簇  
Figure 4 99 core topic clusters after merging

表 4 2021 年排名前 10 的主题及其对应天际线集合  
Table 4 Top 10 topics and their corresponding skyline collections in 2021

主题天际线集合	主题热度排序结果
(1) 新冠肺炎疫情、新发展格局、供应链金融、“互联网+”、O2O 供应链、供应链	1- 新冠肺炎疫情; 2- 供应链; 3- 新发展格局; 4- “互联网+”; 5- 供应链金融; 6-O2O 供应链
(2) 线上供应链金融、双渠道供应链、两部定价契约、闭环供应链	7- 线上供应链金融; 8- 两部定价契约; 9- 双渠道供应链; 10- 闭环供应链

5.3.2 主题热度排名时间序列构建

类似上文 2021 年主题热度排序方法，针对每一年各主题的热度进行排序，最终可得到如表 5 所示的

主题热度排名时间序列。显然，该时间序列一定程度上反映了对应主题热度在近 10 年的动态演化过程，有助于后续新兴主题的识别。

表 5 主题热度排名时间序列  
Table 5 Topic popularity time series

主题	年份									
	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
Topic 1	-	-	-	-	-	-	-	-	6	4
Topic 2	39	16	20	18	18	21	15	19	25	19
Topic 3	64	41	41	27	38	37	43	37	53	57
Topic 4	59	86	32	26	34	40	33	30	37	35
Topic 5	25	20	25	29	28	27	23	25	31	27
...	...	...	...	...	...	...	...	...	...	...
Topic 99	90	84	49	80	77	84	75	66	77	69



5.3.3 新兴主题及其热度比获取

结合表 5、公式 (12) 和 (13) 可得出各主题热度比  $TPP_t^z$ ，借鉴杨金庆等<sup>[36]</sup>做法，分析各主题热度

比演化趋势，并筛选出从  $TPP_t^z < 1$  到  $TPP_t^z > 1$  的主题，最终可获得如表 6 所示新兴主题及其热度比：

表 6 新兴主题及热度比  
Table 6 Emerging topics and their popularity ratio

主题	年份									
	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
风险管理	0.82	0.28	1.36	1.48	1.32	1.20	1.34	1.40	1.26	1.30
技术创新	0.88	0.38	0.38	0.20	0.22	0.34	0.34	0.90	1.12	1.08
碳税	0.34	1.44	1.14	1.12	1.02	1.40	1.44	1.44	1.28	1.34
商业信用	0.16	0.10	0.04	0.18	0.18	0.90	0.86	0.80	0.50	1.06
资金约束	0.64	0.52	0.7	1.04	1.48	1.60	1.58	1.54	1.34	1.50
供应链集成	0.68	0.98	1.44	1.22	1.12	1.18	1.36	1.22	1.32	1.28
激励契约	0.52	1.30	1.10	1.06	1.16	1.22	1.30	1.32	1.16	1.22
产品绿色度	0.92	1.14	1.38	1.34	1.40	1.52	1.40	1.38	1.46	1.38
渠道竞争	0.78	1.12	1.40	1.26	1.10	1.42	1.46	1.20	1.24	1.32

注：表中加黑表示  $TPP_t^z$  高于 1

5.3.4 结果分析

由表 6 可知，供应链领域存在风险管理、技术创新、碳税、商业信用、资金约束、供应链集成、

激励契约、产品绿色度和渠道竞争等 9 个新兴主题，其在 2012—2021 年主题热度比发展趋势如图 5 所示。下面将对图 5 中各新兴主题进行具体分析。

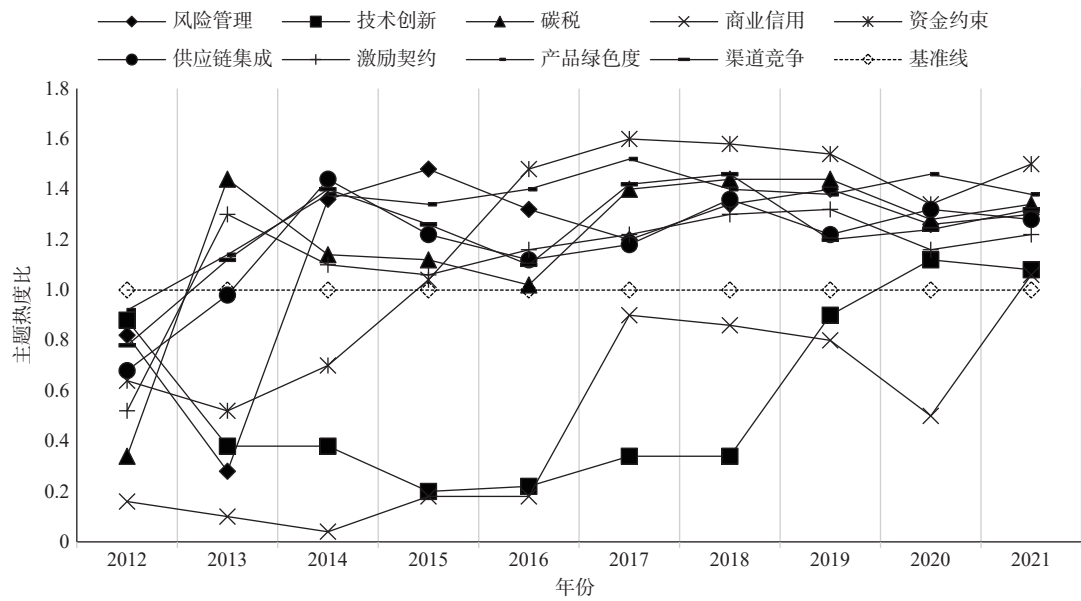


图 5 新兴主题热度比的演化趋势  
Figure 5 Evolution trend of popularity ratio of emerging topics

(1) 风险管理：对供应链风险进行科学管理有助于企业降本增效，现有研究主要聚焦于供应链风险事后控制、信息共享、行业范围拓展及动态预警等方面。该主题的整体热度比水平较高，从 2014 年开始就一直超过基准线，说明该主题从 2014 年开始可能已成为供应链领域中一个热点。现实中，风险管理和供应链之间联系非常紧密，企业应该重视风险管理和供应链管理的整合，加强各个环节沟通和协作，进而提升风险管理效率和供应链弹性。

(2) 技术创新：技术创新发展对供应链信息链有效运转意义重大。该主题主要关注供应链数据传递效率及信息透明度，期望利用区块链等技术帮助从事供应链金融的机构提升对优质企业的识别效率，使全链条实现信用共享，降低金融风险<sup>[40]</sup>。该主题热度比在 2020 年之前低于基准线，而后高于基准线，且呈现先下降后上升趋势，属于再次兴起的主题。当前，区块链、人工智能、数字孪生等一系列创新技术已开始

(3) 碳税: 环境保护意识日益增强, 碳税政策愈受关注, 其能有效减少碳排放并增加零售商的利润, 但却会使制造商利益受损。当前, 碳税环境下政府、制造商及零售商等三方博弈问题日益受到学者关注。该主题热度比从 2013 年开始就一直高于基准线, 属于近年来的研究热点。实际上, 碳税是一种重要的碳减排政策工具, 已在诸多发达国家和地区得到充分实践, 并取得良好成效, 将碳税引入我国碳减排政策体系中已成为理论界和实务界持续关注的重要课题。

(4) 商业信用: 商业信用的达成是供需双方均衡博弈的结果<sup>[41]</sup>。之前有关企业商业信用影响因素的研究多集中在外部市场环境和内部信息质量等方面。随着数字化技术的发展, 现有研究正转向数字化转型与商业信用供给关系等方面。该主题热度比从 2021 年开始才高于基准线, 具备明显的新颖性特征, 属于后发的新兴主题。供应链是实现商业信用价值最大化的场景, 新型供应链平台和商业银行的供应链金融产品等的介入, 不仅可以解决传统交易中存在的大部分问题, 亦可实现商业信用价值最大化。

(5) 资金约束: 随着供应链交易机制的健全和市场信誉体系的建立, 基于供应链上下游商业信用的融资逐渐成为缓解资金约束的重要手段。当前学者主要以双边资金约束的供应链策略和双渠道供应链信息为主要研究方向。该主题热度比保持较快增长, 从 2015 年开始实现高于基准线的转变。资金约束是供应链管理中不可忽视的重要因素, 资金链不稳定将直接影响供应链的正常运转, 企业需要通过有效的供应链融资和运营决策来应对资金约束带来的挑战。

(6) 供应链集成: 供应链集成在于通过信息集成与共享减少牛鞭效应影响。由于实践需要, 当前研究大多为集成系统优化设计及其与企业绩效、数字化转型和营运资金的相互关系<sup>[42]</sup>。该主题热度比波动比较稳定, 从 2014 年开始高于基准线。现实中, 供应链集成可以增强供应链响应速度、降低企业成本和提高企业战略优势等。可以预见, 该主题未来一段时间仍将受到学术界和企业界关注和重视。

(7) 激励契约: 激励契约是有效规避委托代理道德风险的重要手段, 当前研究主要关注各激励契约的实际效用, 例如激励契约对供应链碳减排<sup>[43]</sup>和定价决策的影响等。该主题热度比整体保持较高水平, 从 2013 年开始就高于基准线。在契约激励下, 可以确保订单准时交付, 保证产品质量, 提高用户满意度, 降低供应链成本, 提高整条供应链绩效和每一个成员

企业绩效, 为满足实践需要, 不同场景中的激励契约将不断被提出。

(8) 产品绿色度: 随着经济发展, 供应链能力增强, 产品供给能力提高, 单位产品的环境负荷增加, 产品供应各个阶段对环境的友好程度也愈发受到关注。现有研究主要分析双渠道绿色供应链<sup>[44]</sup>和闭环供应链<sup>[45]</sup>对产品全生命周期中能源资源输入等问题。该主题热度比从 2013 年开始就高于基准线, 受到较高关注。实际上, 产品绿色度和供应链之间存在相互依存和相互促进关系, 提升产品绿色度可以增强供应链整体竞争力, 而有效的供应链管理也可以推动产品绿色度的提升。

(9) 渠道竞争: 渠道竞争与电商平台、定价策略、消费者行为息息相关<sup>[46]</sup>。随着零售实践的不断深入, 厂商将面临渠道的多样化、冲突与竞争等一系列挑战。当前该主题主要围绕零售商模式、信息共享策略和定价策略等问题展开研究。该主题热度比整体保持较高水平, 从 2013 年开始就高于基准线, 具有较大的发展空间。现实中, 渠道竞争的压力会传导到供应链上, 要求供应链具备更高的效率和灵活性, 而供应链的优化和创新又会支持渠道在竞争中表现。

综上所述, 以上新兴主题不仅符合杨金庆等<sup>[36]</sup>的评判标准, 而且也都具有比较明显的新颖性、影响力和发展潜力。预计这些主题将继续获得学术界和企业界关注和重视, 并有望在不同应用场景中得到发展。

## 6 结论 / Conclusion

本文从主题热度的演化视角对领域新兴主题进行了科学识别。首先, 基于词频分析法、关键词重要性和近邻传播聚类自适应识别了初始研究主题; 其次, 结合平均相似性系数和 AP 聚类算法识别了细粒度研究主题; 然后, 以细粒度主题的新颖度、影响力和潜力等表征新兴主题, 并借助天际线算法和 PCA 降维构建主题热度排名时间序列; 最后, 利用主题热度排名的时间演化趋势精准识别了领域新兴主题。本文创新点主要体现在: ①充分考虑了关键词在不同文献中的重要性程度和各类聚类簇平均相似性, 并通过多次自适应 AP 聚类细粒度识别了领域研究主题, 解决了传统主题识别不精细问题; ②以主题新颖度、主题影响力和主题潜力表征新兴主题, 通过构建并分析主题热度比以识别新兴主题, 一定程度克服了传统新兴主题识别维度过高、数据复杂难处理和方法普适性偏弱等问题; ③动态展示了领域新兴主题热度比演

化趋势,有助于科研人员科学选题和政府部门甄选战略性新兴产业。

本文在获取主题热度时,仅选取了簇内中心代表关键词的新颖度、影响力和潜力等指标,忽略了其他特征变量可能带来的影响。同时,在细粒度识别领域研究主题时,仅选择知网期刊论文文献作为数据源。随着 Altmetrics 等相关学术社交媒体的兴起,未来新兴主题的发现可从多数据源和多特征变量中进行挖掘与分析,从而实现对新主题精准识别和及时跟踪。

### 参考文献/References:

- [1] 柴文越,刘小平,梁爽.新兴主题识别方法研究综述[J].现代情报,2023,43(12): 164-177. (CHAI W Y, LIU X P, LIANG S. Review of emerging topic identification methods[J]. Journal of modern information, 2023, 43(12): 164-177.)
- [2] MATSUMURA N, OHSAWA Y, ISHIZUKA M. Discovery of emerging topics between communities on WWW[C]//Asia-Pacific conference on Web intelligence: research and development, lecture notes in computer science. Berlin: Springer, 2001: 473-482.
- [3] ROTOLO D, HICKS D, MARTIN B R. What is an emerging technology?[J]. Research policy, 2015, 44(10): 1827-1843.
- [4] 李静,徐路路,赵素君.基于时间序列分析和SVM模型的基金项目新兴主题趋势预测与可视化研究[J].情报理论与实践,2019(1): 118-123. (LI J, XU L L, ZHAO S J. Prediction and visualization of emerging topics of fund sponsored projects based on time series analysis and SVM model[J]. Information studies: theory & application, 2019(1): 118-123.)
- [5] 黄鲁成,唐月强,吴菲菲,等.基于文献多属性测度的新兴主题识别方法研究[J].科学学与科学技术管理,2015(2): 34-43. (HUANG L C, TANG Y Q, WU F F, et al. Research on identification of emerging topics based on multi-attribute measurement of literature[J]. Science of science and management of S. & T., 2015(2): 34-43.)
- [6] 杨金庆,张力.学科交叉视角下新兴主题识别特征分析——以医学信息学为例[J].情报工程,2021,7(4): 3-12. (YANG J Q, ZHANG L. The characteristics analysis of emerging topic identification from interdisciplinary perspective: a case study of medical informatics[J]. Technology intelligence engineering, 2021, 7(4): 3-12.)
- [7] 谈毅,黄燕丽.基于过程的新兴技术规划与选择模型研究[J].科技管理研究,2007,27(8): 5-8. (TAN Y, HUANG Y L. Study on the new emerging technology planning and choosing model[J]. Science and technology management research, 2007, 27(8): 5-8.)
- [8] 张家惠,丁敬达.多源数据融合的新兴主题探测研究——以文化遗产领域为例[J].图书情报工作,2023,67(9): 23-31. (ZHANG J H, DING J D. Research on emerging topic detection of multi-source data fusion: taking cultural heritage as an example[J]. Library and information service, 2023, 67(9): 23-31.)
- [9] 任智军,乔晓东,张江涛.新兴技术发现模型研究[J].现代图书情报知识,2016(7): 60-69. (REN Z J, QIAO X D, ZHANG J T. Discover emerging technologies with LDA model[J]. New technology of library and information service, 2016(7): 60-69.)
- [10] JANG W, PARK Y, SEOL H. Identifying emerging technologies using expert opinions on the future: a topic modeling and fuzzy clustering approach[J]. Scientometrics, 2021, 126(8): 6505-6532.
- [11] 钟辉新.新兴趋势探测研究综述[J].现代情报,2017(12): 164-169. (ZHONG H X. Review on emerging trend detection[J]. Journal of modern information, 2017(12): 164-169.)
- [12] 冯佳,张云秋.基于新生词与知识图谱的新兴趋势探测研究[J].情报理论与实践,2015,38(11): 87-90, 138. (FENG J, ZHANG Y Q. Research on emerging trends detection based on new words and knowledge graph[J]. Information studies: theory & application, 2015, 38(11): 87-90, 138.)
- [13] SASAKI H, FUGETSU B, SAKATA I. Emerging scientific field detection using citation networks and topic models: a case study of the Nanocarbon field[J]. Applied system innovation, 2020, 3(3): 40.
- [14] CHEN C M. CiteSpace II: detecting and visualizing emerging trends and transients patterns in scientific literature[J]. Journal of the American Society for Information Science and Technology, 2006, 57(3): 359-377.
- [15] GLANZEL W, THIJSS B. Using 'core documents' for detecting and labelling new emerging topics[J]. Scientometrics, 2012, 91(2): 399-416.
- [16] 段庆峰,潘小换.利用社交媒体识别学科新兴主题研究[J].情报学报,2017,36(12): 1216-1223. (DUAN Q F, PAN X H. Identification of emerging topics in science using social media[J]. Journal of the China Society for Scientific and Technical Information, 2017, 36(12): 1216-1223.)
- [17] 陈稳,陈伟.基于计量指标多变量LSTM模型的新兴主题热度预测研究[J].数据分析与知识发现,2022,6(10): 35-45. (CHEN W, CHEN W. Predicting popularity of emerging topics with multivariable LSTM and bibliometric indicators[J]. Data analysis and knowledge discovery, 2022, 6(10): 35-45.)
- [18] XU S, HAO L Y, AN X, et al. Review on emerging research topics with key-route main path analysis[J]. Scientometrics, 2020, 122(1): 607-624.
- [19] XU M, LI G J, WANG X D. Detecting emerging topics by exploiting probability burst and association rule mining: A case study of library and information science[J]. Malaysian journal of library and information science, 2020, 25(1): 47-66.
- [20] 黄鲁成,张璐,吴菲菲,等.基于突现文献和SAO相似度的新兴主题识别研究[J].科学学研究,2016,34(6): 814-821. (HUANG L C, ZHANG L, WU F F, et al. The identification research of emerging topics based on burst articles and similarity



- of SAO[J]. Studies in science of science, 2016, 34(6): 814-821.)
- [21] 白敬毅, 颜端武, 陈琼. 基于主题模型和曲线拟合的新兴主题趋势预测研究[J]. 情报理论与实践, 2020, 43(7): 130-136. (BAI J Y, YAN D W, CHEN Q. Trend prediction of emerging topics based on topic model and curve fitting[J]. Information studies: theory & application, 2020, 43(7): 130-136.)
- [22] 高阳, 严建峰, 刘晓升. 朴素并行 LDA[J]. 计算机科学, 2015, 42(6): 243-246. (GAO Y, YAN J F, LIU X S. Naive parallel LDA[J]. Computer science, 2015, 42(6): 243-246.)
- [23] KUNG H T, LUCCIO F, PREPARATA F P. On finding the maxima of a set of vectors[J]. Journal of the ACM, 1975, 22(4): 469-476.
- [24] 魏小娟, 杨婧, 李翠平, 等. Skyline 查询处理[J]. 软件学报, 2008, 19(6): 1386-1400. (WEI X J, YANG J, LI C P, et al. Skyline query processing[J]. Journal of software, 2008, 19(6): 1386-1400.)
- [25] 万校基, 李海林, 龚燕燕, 等. 基于天际线算法的主题排序方法研究[J]. 情报学报, 2022, 41(4): 388-400. (WAN X J, LI H L, GONG Y Y, et al. Ranking method of topic popularity based on skyline algorithm[J]. Journal of the China Society for Scientific and Technical Information, 2022, 41(4): 388-400.)
- [26] 李海林, 万校基, 林春培. 基于关键词重要性和近邻传播聚类主题分析研究[J]. 情报学报, 2018, 37(5): 533-542. (LI H L, WAN X J, LIN C P. Theme analysis based on keyword importance and affinity propagation clustering[J]. Journal of the China Society for Scientific and Technical Information, 2018, 37(5): 533-542.)
- [27] FIELD J A, JOHNSON C A, ROSE J B. What is “emerging”?[J]. Environmental science and technology, 2006, 40(23): 7105.
- [28] 徐路路, 靳杨. 基于 FSD 模型的政府资助项目新兴主题探测与分析[J]. 科学学与科学技术管理, 2019, 40(2): 40-54. (XU L, JIN Y. Emerging topics detection and analysis of government funded projects based on FSD model[J]. Science of science and management of S. & T., 2019, 40(2): 40-54.)
- [29] 安璐, 余传明, 董丽, 等. 科研机构对新兴主题的贡献度可视化研究——以中美图情科研机构为例[J]. 图书情报工作, 2014, 58(13): 68-74. (AN L, YU C M, DONG L, et al. A visual study on the contribution of research institutions to the emerging themes: taking Chinese and American LIS institutions as examples[J]. Library and information service, 2014, 58(13): 68-74.)
- [30] TU Y N, SENG J L. Indices of novelty for emerging topic detection[J]. Information processing and management, 2012, 48(2): 303-325.
- [31] 范云满, 马建霞. 基于 LDA 与新兴主题特征分析的新兴主题探测研究[J]. 情报学报, 2014(33): 698-711. (FAN Y M, MA J X. Detection of emerging topics based on LDA and feature analysis of emerging topics[J]. Journal of the China Society for Scientific and Technical Information, 2014(33): 698-711.)
- [32] 叶光辉, 王灿灿, 李松桦. 基于 SciTS 会议文本的跨学科科研协作新兴主题识别及预测[J]. 情报科学, 2022(7): 126-135. (YE G H, WANG C C, LI S Y. Recognition and prediction of emerging topics in interdisciplinary scientific research collaboration based on SciTS conference text[J]. Information science, 2022(7): 126-135.)
- [33] HALL D, JURAFSKY D, MANNING C D. Studying the history of ideas using topic models[C]//Proceedings of the conference on empirical methods in natural language processing. Stroudsburg: Association for Computational Linguistics, 2008: 363-371.
- [34] WANG Q. A bibliometric model for identifying emerging research topics[J]. Journal of the Association for Information Science and Technology, 2018, 69(2): 290-304.
- [35] 殷蜀梅. 判断新兴研究趋势的技术方法分析[J]. 情报科学, 2008(4): 536-540. (YIN S M. Analysis of the methods for detecting emerging trend[J]. Information science, 2008(4): 536-540.)
- [36] 杨金庆, 肖兵, 程秀峰, 等. 基于 HDP 过程模型与学术会议的学科新兴主题发现研究——以“人工智能”领域为例[J]. 情报理论与实践, 2019, 42(4): 117-122. (YANG J Q, XIAO B, CHENG X F, et al. Detecting emerging topics of subjects based on HDP model and academic conferences: a case study in the field of artificial intelligence[J]. Information studies: theory & application, 2019, 42(4): 117-122.)
- [37] 刘军军, 冯云婷, 朱庆华. 可持续运营管理研究趋势和展望[J]. 系统工程理论与实践, 2020, 40(8): 1996-2007. (LIU J J, FENG Y T, ZHU Q H. The research trend and future directions on sustainable operations management[J]. Systems engineering theory and practice, 2020, 40(8): 1996-2007.)
- [38] 王谦, 谢春雨, 崔怡雯. 非对称信息下报童问题的质量激励研究[J]. 管理评论, 2018(6): 227-237. (WANG Q, XIE C Y, CUI Y W. Quality incentives based on newsvendor model under asymmetric information[J]. Management review, 2018(6): 227-237.)
- [39] 吴忠和, 陈宏, 梁翠莲, 等. 时间约束下不对称信息鲜活农产品供应链应对突发事件协调模型[J]. 中国管理科学, 2015, 23(6): 126-134. (WU Z H, CHEN H, LIANG C L, et al. Supply chain disruptions coordination model of fresh agricultural products under time constraints with asymmetric information[J]. Chinese journal of management science, 2015, 23(6): 126-134.)
- [40] 代业明, 于双. 碳税监管下考虑零售商双重行为偏好的再制造闭环供应链决策[J/OL]. 中国管理科学 [2024-06-13]. <https://doi.org/10.16381/j.cnki.issn1003-207x.2022.1639>. (DAI Y M, YU S. Remanufacturing closed-loop supply chain decision considering retailer's dual behavior preference under carbon tax regulation[J/OL]. Chinese journal of management science [2024-06-13]. <https://doi.org/10.16381/j.cnki.issn1003-207x.2022.1639>.)
- [41] 贡菲菲, 王铁昕. 国内大循环下区块链技术在供应链金融中的应用研究——基于沿海内贸场景的分析[J]. 价格理论与实践, 2023, 463(1): 117-121. (YUAN F F, WANG Y X. Research on the application of blockchain technology in supply

- chain finance under the domestic big cycle: analysis based on the coastal domestic trade scenario[J]. Price: theory & practice, 2023, 463(1): 117-121.)
- [42] 李琦, 刘力钢, 邵剑兵. 数字化转型、供应链集成与企业绩效——企业家精神的调节效应[J]. 经济管理, 2021, 43(10): 5-23. (LI Q, LIU L G, SHAO J B. The effects of digital transformation and supply chain integration on firm performance: the moderating role of entrepreneurship[J]. Business management journal, 2021, 43(10): 5-23.)
- [43] 范如国, 李玉龙. 信息不对称下企业低碳发展的激励契约设计[J]. 中国科技论坛, 2016, 247(11): 62-69. (FAN R G, LI Y L. Incentive contract design for enterprise's low-carbon development under asymmetric information[J]. Forum on science and technology in China, 2016, 247(11): 62-69.)
- [44] 余娜娜, 王道平, 赵超. 考虑产品绿色度的双渠道供应链协调研究[J]. 运筹与管理, 2022, 31(4): 75-81. (YU N N, WANG D P, ZHAO C. Study on coordination of dual-channel supply chain considering product green degree[J]. Operations research and management science, 2022, 31(4): 75-81.)
- [45] 吕宝龙, 张桂涛, 刘阳, 等. 考虑碳税和产品绿色度的闭环供应链网络 Nash 博弈均衡模型[J]. 中国人口·资源与环境, 2019, 29(1): 59-69. (LÜ B L, ZHANG G T, LIU Y, et al. Nash game equilibrium model of the closed-loop supply chain network considering carbon tax mechanism and product green degree[J]. China population, resources and environment, 2019, 29(1): 59-69.)
- [46] 刘丁瑞, 李登峰, 郑小雪. 跨境电商环境下考虑渠道服务水平的产品定价研究[J]. 计算机集成制造系统, 2020, 26(8): 2278-2287. (LIU D R, LI D F, ZHENG X X. Product pricing considering channel service level in cross-border e-commerce environment[J]. Computer integrated manufacturing system, 2020, 26(8): 2278-2287.)

#### 作者贡献说明 / Author contributions:

万校基: 提出论文思路与设计框架, 主导实验开展, 论文修改;

李海林: 论文思路, 论文修改与校对;

何雨晴: 论文撰写, 数据处理;

杨润奇: 论文撰写, 论文修改;

林海龙: 论文修改, 数据处理。

#### Research on the Identification and Analysis of Emerging Topics from the Perspective of Popularity Evolution\*

Wan Xiaoji<sup>1,2</sup> Li Hailin<sup>1,3</sup> He Yuqing<sup>1</sup> Yang Runqi<sup>1</sup> Lin Hailong<sup>1</sup>

<sup>1</sup>College of Business Administration, Huaqiao University, Quanzhou 362021

<sup>2</sup>Oriental Enterprise Management Research Center, Huaqiao University, Quanzhou 362021

<sup>3</sup>Modern Statistics and Big Data Research Center, Huaqiao University, Xiamen 361021

**Abstract: [Purpose/Significance]** Identifying emerging topics in disciplinary fields can not only guide researchers in selecting scientific topics, but also provide the technical support for the precise formulation of strategic emerging industry development plans. **[Method/Process]** Due to the lack of a unified definition for “emerging topics” and the challenges associated with processing complex, high-dimensional data in topic identification, this study proposed an emerging topic detecting method based on topic popularity evolution. It combined the keyword importance, average similarity coefficient and affinity propagation clustering algorithm to adaptively identify research topics in a field. The topic popularity was characterized by the novelty, influence and potentiality of topics. Time series of the topic popularity were obtained by skyline algorithm and principal component analysis method, and the emerging topics were accurately identified by the time evolution of the initial topic popularity. **[Result/Conclusion]** By data processing and mining journal papers in supply chain from 2012 to 2021 in the China National Knowledge Infrastructure (CNKI) database, this study demonstrates that the proposed method can effectively identify emerging topics in the field. It provides decision-making support for relevant researchers and government departments in selecting topics and planning.

**Keywords:** emerging topics affinity propagation skyline algorithm topic popularity popularity evolution

\*This work is supported by the Fujian Provincial Social Science Planning Project titled “Research on the Output Performance Evaluation of National Natural Science Foundation Closing Projects Based on Big Data Analysis” (Grant No. FJ2023B109).

**Author(s):** Wan Xiaoji, associate professor, PhD, graduate supervisor. Li Hailin, professor, PhD, doctoral supervisor, corresponding author. E-mail: hailin@hqu.edu.cn. He Yuqing, master candidate. Yang Runqi, undergraduate, Lin Hailong, undergraduate.

Received: 2024-02-28 Revised: 2024-06-14 Pages: 126-138