

---

# 基于共词分析法的学科主题演化研究方法体系构建

唐果媛

中国科学院文献情报中心 北京 100190

摘要：[目的/意义] 学科主题演化研究作为学科情报研究的一个方向，分析学科主题演化趋势，可以有效地揭示学科知识发展变化及其相互作用的特征和规律，有助于科研人员判断学科变化趋势以及某一领域的研究热点，调整科研方向，亦有利于决策层制定发展战略、规划研究布局、支撑重点领域发展等。本文在支撑科技发展和科研管理方面都具有现实意义。[方法/过程] 本文构建了基于共词分析法的学科主题演化研究方法体系，包括4个模块，分别是：数据准备、演化阶段划分、主题识别和主题演化分析。[结果/结论] 在主题识别阶段改进了词频 g 指数来选取共词分析的对象；在主题演化分析模块，提出从静态和动态两个角度来分析学科主题的演化情况，构建三维战略坐标来进行静态分析，并构建学科主题演化现象识别模型来进行动态分析。

关键词：共词分析法 学科主题演化 词频 g 指数 三维战略坐标

分类号：G250

Building the method system of subject theme evolution based on co-word analysis method

Tang Guoyuan

National Science Library, Chinese Academy of Sciences, Beijing 100190

Abstract:[Purpose/significance] Research of subject theme evolution is one of the aspects of subject intelligence research. It can reveal the development and changes of subject knowledge and the characteristics and laws of their interactions effectively to analyze the trend of subject theme evolution. It can also help the researchers determine trends of the subject's changes and the hotspots of the particular subject areas, and adjust research direction. At the same time, it can help decision-makers formulate development strategies, plan the layout, and support the development of key areas. It has practical significance for supporting the development of science & technology and the management of scientific research. [Method/process] This paper builds the method system of subject theme evolution based on co-word analysis method. It includes four modules: data preparation, evolution phasing, theme identification, and analysis of the theme evolution. [Result/conclusion] This paper

---

uses the improved word frequency g index to select the object of co-word analysis in the module of theme identification. This paper analyzes the subject theme evolution from both static and dynamic perspective. This paper presents a three-dimensional strategic diagram for static analysis, and builds the recognition model of subject theme evolutionary phenomenon for dynamic analysis.

**Keywords:** co-word analysis, subject theme evolution, word frequency g index, three-dimensional strategic diagram

---

## 1 引言

邓小平在 1988 年提出“科学技术是第一生产力”。习近平总书记在 2014 年也提出“科技是国家强盛之基，创新是民族进步之魂”。在当今中国，创新驱动发展已经上升为国家战略，科技创新促进发展也日益受到国家的重视和支持。科学研究、技术开发是科技创新的重要基础，支撑科学研究、技术开发的重要组成部分之一是学科情报研究。在学科情报研究中选择“学科主题演化”研究是具有积极意义和实用价值的。该研究有助于科研人员判断学科变化趋势以及某一领域的研究热点，调整科研方向，亦有利于决策层制定发展战略、规划研究布局、支撑重点领域发展等。总之，本文在支撑科技发展和科研管理方面都具有现实意义。

同时，随着计算机技术和网络技术的发展，文献的数字化不断推进，网络环境下的文献数据资源几乎成指数增长，面对如此海量的数据，如何有效地对学科领域的主题进行分析对于研究者进行思路整理及发现研究主题非常重要。

目前对学科主题演化研究的科学计量方法主要有共引分析法、词频分析法和共词分析法，这三种方法都具有一定的局限性。共引分析法的局限性在于需要庞大的引文索引作为基础，且以文献作为分析对象，而文献包含的主题可能不唯一；词频分析法的局限性在于只关注单个关键词的词频；共词分析法的局限性在于选择关键词具有一定的主观性。虽然这三种方法都存在一定的局限性，但通过对比发现共词分析法具有明显的优势。相比于词频分析法，共词分析法不仅能分析高频词，而且更关注这些词之间的联系，从而反映出概念之间的关系。相比于共引分析法，共词分析法的优势在于能深入文献内部，以文献内部的关键词作为分析对象，从更微观的角度去揭示学科主题演化规律。

虽然目前已有研究人员利用共词分析法来研究学科主题演化，但仍存在一些不足之处。首先是共词分析法本身的缺陷，表现在选取高频关键词时，大多数研究人员都根据经验从关键词词频、关键词共现词频、关键词词频和共现词频相结合等角度来选取一定数量的关键词来进行共词分析，主观性较强，缺乏理论的指导，虽然有部分研究人员采用一些计算公式来选取高频词作为共词分析的对象，但这些公式存在局限性，如高频低频词界分公式的普适性还有待验证，词频  $g$  指数没有考虑到关键词的词频相同的情况。其次在可视化展示方面，战略坐标在展示学科领域在某一阶段的研究结构时，用向心度和密度呈现了学科领域主题内部联系情况和主题间相互作用情况，但没有对同一阶段不同主题的热度加以区分，没有体现出研究人员对这些主题感兴趣的程度。

本文的创新性在于：在构建的基于共词分析法的学科主题演化研究方法体系

---

中，通过改进的词频 g 指数突破了词频 g 指数未考虑相同词频关键词的局限性，在战略坐标的基础上提出了三维战略坐标，以弥补战略坐标不能体现主题的研究热度的局限性，并且构建出学科主题演化现象识别模型，使学科主题演化的研究得到了深化，实用性更强，为科学研究和技术开发提供了支撑。

## 2 总体方法体系构建

本文构建了基于共词分析法的学科主题演化研究方法体系（见图 1），主要用来解决学科主题演化研究的三个问题（when-what-how）：即演化阶段如何划分、各演化阶段都包含了哪些主题、这些主题随着时间的发展都产生了哪些变化。本文构建的方法体系主要包括 4 个模块：数据准备、演化阶段划分（when）、主题识别（what）、主题演化分析（how）。其中，数据准备模块，是根据所要分析的学科领域选择合适的文献数据库，并基于特定的研究目的设计检索策略；演化阶段划分模块，是为了解决学科主题演化研究中“when”这个问题，即如何划分学科领域的演化阶段，主要方法<sup>[1]</sup>有：等距离定长法、科技文献增长规律法、学科发展历程法、试错法（通过不断试错的方法来划分演化阶段，使得每个阶段包含的文献数量大致相同<sup>[2]</sup>）；主题识别模块，是为了解决学科主题演化研究中“what”这个问题，即每个演化阶段都包含了哪些主题，本文构建的方法体系采用的是基于关键词的共词分析法，其流程包括获取关键词、构建关键词共词矩阵、主题聚类，在获取关键词的步骤中，本文对词频 g 指数进行了一定的改进；主题演化分析模块解决了学科主题演化研究中“how”这个问题，即每个演化阶段的主题随着时间的发展产生了哪些变化，本文从静态和动态两个角度出发，通过构建三维战略坐标来进行静态分析，三维战略坐标的三个维度分别为密度、向心度和强度，基于相似指数构建了一个主题演化现象识别模型来进行动态分析。

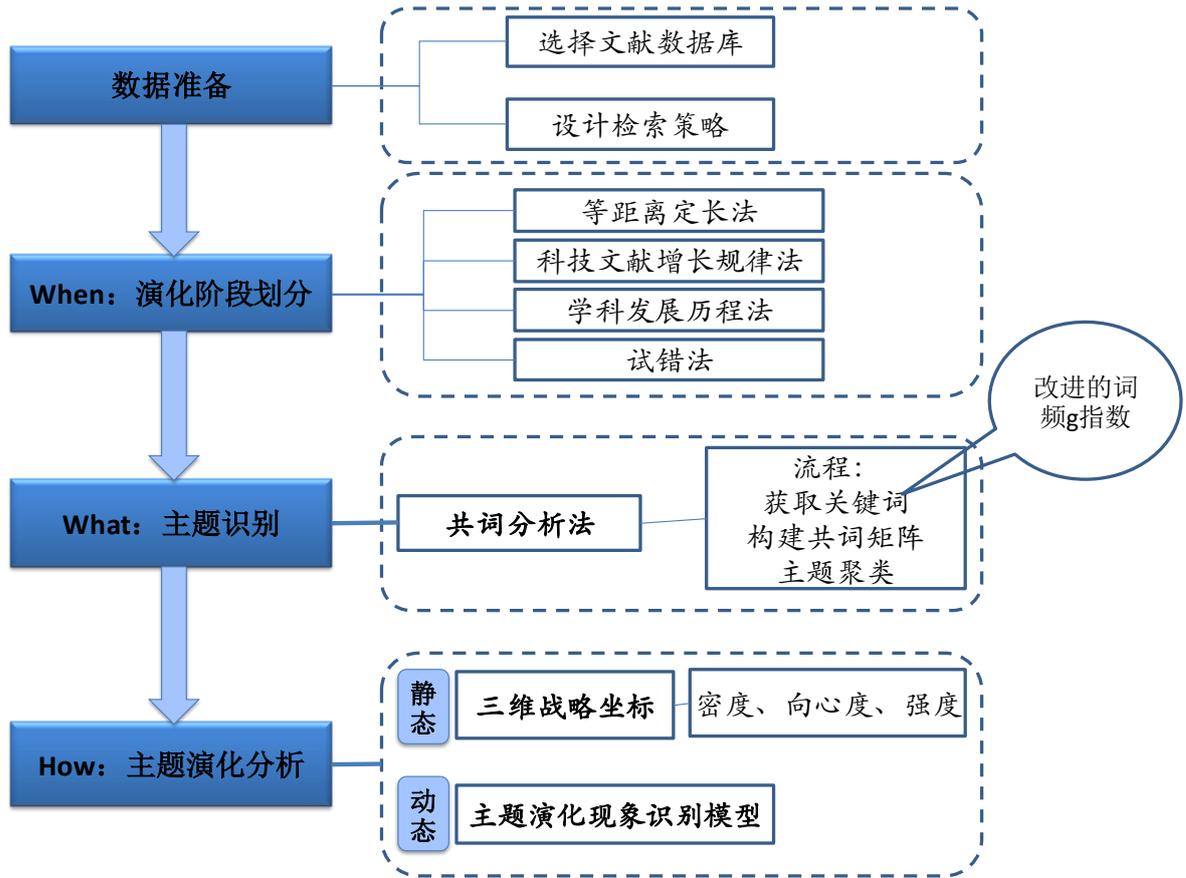


图 1 基于共词分析法的学科主题演化研究方法体系

笔者在《基于共词分析法的学科主题演化研究进展与分析》<sup>[1]</sup>一文中已对方法体系中的数据准备模块、演化阶段划分模块以及主题识别模块进行了较为详细的阐述，在此不再赘述。本文将重点分析在主题识别模块中提出的改进词频  $g$  指数，以及在主题演化分析模块中构建的三维战略坐标和主题演化现象识别模型。

### 3 改进词频 $g$ 指数的提出过程

#### 3.1 选取共词分析对象的主要方法

学科主题识别模块的分析流程如图 2 所示：首先从选取的学科领域的数据集选择 and 提取共词分析法的具体对象（关键词），然后利用选取的关键词构建共词矩阵并进行归一化，最后是采用合适的聚类方法识别每个演化阶段包含的主题。



图 2 基于共词分析法的学科主题识别的分析流程

研究人员选取共词分析对象——高频词的方法主要有 3 种，按照出现的时间顺序依次为：经验判定法、高频低频词界分公式和词频 g 指数。

研究人员采用经验判定法一般从关键词词频<sup>[3]</sup>、关键词共现词频<sup>[4]</sup>、关键词词频和共现词频相结合<sup>[5]</sup>等三个角度来选取关键词，此方法仅依据研究人员的经验，主观性较强，缺乏理论的指导。

高频低频词界分公式<sup>[6]</sup>是 Donohue 于 1973 年根据齐普夫第二定律提出来的，公式为： $T = \frac{-1 + \sqrt{1 + 8 \times I_1}}{2}$ ，其中  $I_1$  是出现一次的词的个数，T 为高频低频词词频临界值。高频低频词界分公式是从关键词词频的角度出发来选取共词分析对象，具有一定的理论指导，但普适性还有待进一步验证。

词频 g 指数是杨爱青等<sup>[7]</sup>根据 g 指数提出用于选取高频词的方法。g 指数的定义为：一个科学家的分值为 g，当且仅当在他发表的 N 篇论文中有 g 篇论文总共获得了不少于  $g^2$  次的引文总数，而  $(g+1)$  篇论文总共获得了少于  $(g+1)^2$  次的引文总数<sup>[8]</sup>。词频 g 指数的计算方法与 g 指数类似，词频 g 指数的定义是：在研究领域中，一个关键词的数量分值为 g，当且仅当此研究领域的关键词总量 N 中，有 g 个关键词其累计出现频次不少于  $g^2$  次，而  $g+1$  个关键词其累计出现频次少于  $(g+1)^2$  次<sup>[7]</sup>。

词频 g 指数的计算过程见表 1，其中 i 是关键词按词频降序排列后的序号，N 为关键词总数， $F_i$  是序号为 i 的关键词的频次，且满足  $F_{i-1} \geq F_i$ 。

表 1 词频 g 指数的计算过程

关键词序号 i	关键词词频 $F_i$	约束条件
1	$F_1$	$F_1 \geq 1^2$
2	$F_2$	$F_1 + F_2 \geq 2^2$
.....	.....	.....
g-1	$F_{g-1}$	$\sum_{i=1}^{g-1} F_i \geq (g-1)^2$
g	$F_g$	$\sum_{i=1}^g F_i \geq g^2$
g+1	$F_{g+1}$	$\sum_{i=1}^{g+1} F_i < (g+1)^2$
.....	.....	.....
N	$F_N$	$\sum_{i=1}^N F_i < N^2$

相比于从关键词词频、关键词共现频次等角度根据研究人员的经验选出一定数量的关键词的方法，利用高频低频词界分公式和词频 g 指数选出的关键词更为

自然和客观。

### 3.2 改进的词频 $g$ 指数

词频  $g$  指数没有考虑关键词词频相同的情况,还存在改进的地方。当  $F_g=F_{g+1}$  的时候,即第  $g$  个关键词和第  $g+1$  个关键词的词频相同,若按照词频  $g$  指数计算出的关键词个数来获取关键词,则可能出现多种选词结果。以杨爱青等<sup>[7]</sup>在提出词频  $g$  指数时用的实证数据为例,如表 2 所示,计算出词频  $g$  指数为 19,序号为 19 的关键词的词频为 12,而词频为 12 的关键词共有 5 个,这 5 个关键词中只选择了 2 个,那么选择哪两个关键词就成为了一个问题。

表 2 2006 年图情领域前 30 个关键词及对应序号<sup>[7]</sup>

序号 ( $g$ )	序号平方	2006 年的关键词字段	出现频次	累计频次
1	1	图书馆	90	90
2	4	数字图书馆	51	141
3	9	高校图书馆	50	191
4	16	本体	27	218
5	25	知识管理	27	245
.....	.....	.....	.....	.....
18	324	图书馆管理	12	363
19	361	信息素质	12	375
20	400	图书馆学	12	387
21	441	图书馆员	12	399
22	484	电子资源	12	411
23	529	知识产权	11	422
.....	.....	.....	.....	.....
29	841	数字化	10	456
30	900	电子政务	9	465

为了解决这个问题,本文对词频  $g$  指数进行了一定的改进,在考虑到关键词词频相同的情况下,在计算出词频  $g$  指数之后,选取词频  $\geq F_g$  的关键词作为高频词。那么在表 2 中列出的 30 个关键词中,选取词频  $\geq 12$  的 22 个关键词作为高频词。利用改进后的词频  $g$  指数来选取共词分析对象,获取的关键词具有唯一性。

## 4 学科主题演化分析

### 4.1 静态分析——三维战略坐标

战略坐标图是在建立关键词的共词矩阵和聚类分析的基础上,用可视化的形

式来展现学科领域的研究结构,关注的是学科领域中主题内部联系情况和主题间相互作用情况。

战略坐标是 Law 等<sup>[9]</sup>于 1988 年提出用来描述某一学科领域内容的内部联系情况和学科领域间内容相互影响情况的一种可视化方法。战略坐标是一个二维坐标, X 轴为向心度, Y 轴为密度, 坐标原点是两条轴的中位数或者平均数。其中, 密度用来度量主题内关键词之间的紧密联系程度, 它表示该主题在未来维持和发展自身的能力。某一主题密度的计算方式有多种, 如该主题中关键词共现频次的平均值、中位数或者平方和。向心度是用来量度主题内关键词与其他主题内关键词之间的紧密程度, 表明一个主题领域和其他主题领域的相互影响程度。一个主题与其他主题联系越紧密, 这个主题在该学科领域中的地位就越趋于中心位置。向心度可以通过该主题中所有关键词与其他主题中的关键词之间的连接强度进行计算。如该主题中关键词与其他主题中关键词共现频次的总和、平方和的开平方或者前 6 个共现频次的总和。

战略坐标把空间划分为四个象限, 见图 3。在第一象限中, 主题内部联系紧密, 研究趋于成熟, 并在学科领域中处于中心地位; 在第二象限中, 主题内部联系紧密, 研究已形成一定规模, 但在整个学科领域中处于边缘位置; 在第三象限中, 主题内部联系比较松散, 研究还不成熟, 且处于学科领域的边缘位置; 在第四象限中, 主题内部联系比较松散, 研究虽不成熟, 但与其他主题的联系比较紧密, 因此该主题有进一步发展的空间, 是潜在的发展主题。

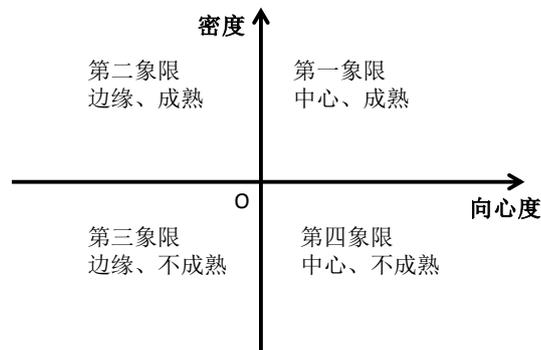


图 3 战略坐标

战略坐标的优点是能通过计算学科领域内热点主题的向心度和密度, 并以坐标图来可视化学科领域主题内部联系情况和主题间相互作用情况, 展现了学科领域的研究结构, 既有量化指标又能进行可视化展示, 同时, 还能通过对比不同时期的战略坐标图展示热点主题演变的过程。在对学科领域内的高频关键词进行战略坐标分析时, 得到的主题应是学科领域内的研究热点, 这些研究热点在同一阶段的热度是不同的, 研究人员对这些研究热点感兴趣的程度也不同, 不同热度的研究热点其生命力也不同。战略坐标只展示了主题的向心度和密度而没有对研

究主题的热度加以区分。

因此，本文在战略坐标图的基础上，增加一个维度——强度，即构建一个三维战略坐标图，三个维度分别为强度、密度、向心度。强度是用来体现主题在当前演化阶段被研究的热度，一个主题的程度越大说明研究人员对该主题的研究兴趣越浓厚，研究人员继续研究该主题的可能性就越大，那么该主题表现出的生命力就越强。密度是用来体现主题内部发展情况，密度越大，则说明主题内部联系越紧密，结构越稳定，学科主题具有生命周期的基本特征，即学科主题在发展过程中会经历产生、发展、成熟和消亡等四个阶段。学科主题在不同的发展阶段其表现出的生命力是不同的，以成熟阶段作为分界线，在主题发展成熟之前主题的生命力较强，在发展成熟之后，若主题融入了新的知识点，那么主题依然有较强的生命力，若没有，则主题发展遇到瓶颈开始消亡，生命力减弱。向心度是用来体现主题之间相互作用的情况，若一个主题与其他主题联系紧密，即处于学科领域的核心地位，那么该主题吸收新知识点的可能性就越大，若一个主题相对较独立，那么其吸收新知识点的可能性就越小。

图 4 展示的是三维战略坐标的示意图，横坐标为向心度，纵坐标为密度，原点为主题向心度和密度的均值，圆圈面积代表主题的程度，圆圈面积越大，主题的程度越大，在当前阶段产生的文献数量越多，强度大于等于均值的主题用橙色圆圈表示，强度小于均值的主题用蓝色圆圈表示。

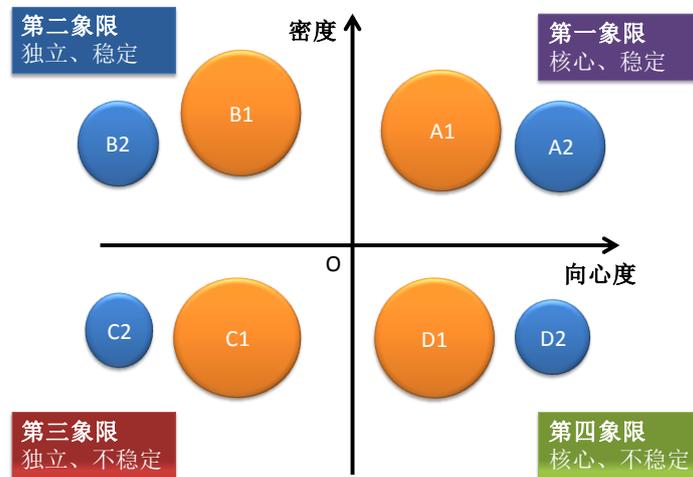


图 4 三维战略坐标

根据主题在三维战略坐标图中的位置以及强度的大小，本文将学科主题划分为 8 种类型，分别为核心稳定热门主题、核心稳定常规主题、独立稳定热门主题、独立稳定常规主题、独立不稳定热门主题、独立不稳定潜在主题、核心不稳定热门主题、核心不稳定潜在主题，其判断标准见表 3。

表 3 主题类型划分标准

主题类型	举例	判断标准
核心稳定热门主题	A1	主题位于第一象限，且强度大于等于均值。
核心稳定常规主题	A2	主题位于第一象限，且强度小于均值。
独立稳定热门主题	B1	主题位于第二象限，且强度大于等于均值。
独立稳定常规主题	B2	主题位于第二象限，且强度小于均值。
独立不稳定热门主题	C1	主题位于第三象限，且强度大于等于均值。
独立不稳定潜在主题	C2	主题位于第三象限，且强度小于均值。
核心不稳定热门主题	D1	主题位于第四象限，且强度大于等于均值。
核心不稳定潜在主题	D2	主题位于第四象限，且强度小于均值。

三维战略坐标图中密度、向心度和强度的计算公式如下，其中向心度和密度的计算公式是 Mark William Neff 等<sup>[10]</sup>于 2009 年提出的，相比于直接用关键词共现频次计算向心度（如：该主题中关键词与其他主题中关键词共现频次的总和、平方和的开平方或者前 6 个共现频次的总和）和密度（如：关键词共现频次的平均值、中位数或者平方和）的方法，这两个公式的优势是，在关键词共现频次归一化的基础上进行计算，消除了绝对词频带来的影响，并且计算的是关键词之间所有的共现频次而不仅仅是排名靠前的共现频次。

$$\text{向心度: } C_m = \frac{\sum_{i,j} O_{i,w}}{O_{i,i}}; \quad [n(N-n)]$$

其中， $C_m$  为主题  $m$  的向心度； $O$  为 Ochiai 系数； $i$  为主题  $m$  内部的关键词； $w$  为不属于主题  $m$  的关键词； $N$  为总词数； $n$  为主题  $m$  内关键词的数量。

$$\text{密度: } D_m = \frac{\sum_{i,j} O_{i,j}}{O_{i,i}}; \quad n!/[2!(n-2)!]$$

其中， $D_m$  为主题  $m$  的密度； $O$  为 Ochiai 系数； $i$  和  $j$  都为主题  $m$  内部的关键词； $n$  为主题  $m$  内关键词的数量。

$$\text{强度: } S_m = \frac{P_m}{P_N};$$

其中， $S_m$  为主题  $m$  的密度； $P_m$  为主题  $m$  的文献数量； $P_N$  为当前演化阶段总的文献数量。

强度的公式是本文提出来的，用主题内文献的数量强度来体现该主题在当前演化阶段的热度，该主题的文献数量占当前演化阶段所有文献的比例越大，则强度越大，说明研究人员在这个阶段对该主题的研究兴趣越大。

## 4.2 动态分析——学科主题演化现象识别模型

在一个演化阶段中，学科领域包含了若干个学科主题，学科主题是由若干个联系紧密的关键词组成。学科领域中的知识在不断发展演变的过程中具有遗传变异的特性，那么学科主题作为学科知识的一种表现形式，也具有遗传变异的特性。不同演化阶段之间学科主题的演化情况的判定是基于主题之间的关系，本文将主题之间的关系定义为两种：相似主题和不相似主题。相似主题是指相似程度较高的主题，本文是基于两个主题之间拥有的相同关键词来计算它们之间的相似程度。同理，不相似主题是指相似程度较低的主题。

目前基于共词分析的学科主题演化相似度测度的算法主要有三种：非相似指数、影响指数和出处指数、相似指数（见表 4），它们是以关键词的频次或共现频次作为主题演化分析的量化标准。

表 4 测度相似度的三种指数

名称	计算公式	指标含义
非相似指数 <sup>[11]</sup>	$T = \frac{C_i + C_j}{C_{ij}}$	$C_i$ 和 $C_j$ 分别表示主题 $i$ 和主题 $j$ 中包含的关键词数量， $C_{ij}$ 表示主题 $i$ 和主题 $j$ 中相同关键词的数量。
影响指数（I）和 出处指数（P） <sup>[12]</sup>	$I_{ij} = \frac{2M_{ij} + Ln_{ij}}{2N_i};$ $P_{ij} = \frac{2m_{ij} + ln_{ij}}{N_j}$	$m_{ij}$ 表示主题 $i$ 和主题 $j$ 共有的关键词数量； $ln_{ij}$ 在影响指数中表示出现在主题 $i$ 和前一阶段的主题 $j$ 中，但不属于前一阶段其他主题的共有关键词数量； $ln_{ij}$ 在出处指数中表示出现在后一阶段的主题 $j$ 和前一阶段主题 $i$ 中，但不属于前一阶段其他主题的共有关键词数量； $N_i$ 表示主题 $i$ 中关键词的数量， $N_j$ 表示主题 $j$ 中关键词的数量。
相似指数 <sup>[13]</sup>	$SI = \frac{2W_{ij}}{W_i + W_j}$	$W_{ij}$ 表示主题 $i$ 和主题 $j$ 中共有的主题词数量， $W_i$ 和 $W_j$ 分别表示主题 $i$ 和主题 $j$ 中包含的主题词数量；2 倍是为了提高相似度的显示度，便于观察和分析数值结果的差异情况。

由于本文构建的方法体系是基于共词分析法对学科主题进行聚类，不存在同一个关键词在同一演化阶段中的多个主题中存在的现象，因此影响指数和出处指数不适合在本文中用来测度相邻演化阶段主题的相似度，而非相似指数和相似指数的计算原理相似。本文采用相似指数对相邻演化阶段中主题的相似程度进行计

算，同时会对相邻演化阶段的两个主题的相似指数设置一定的阈值，并规定，相似指数大于或等于该阈值的两个主题是相似主题，相似指数小于该阈值的两个主题是不相似主题。

本文借鉴 G. Palla<sup>[14]</sup>对社区网络演化过程的划分，将学科主题的演化过程也划分为 6 种：产生、扩张、合并、分裂、收缩、消失（见图 5），其中实线圆圈代表一个主题，虚线圆圈代表主题不存在，圆圈的大小代表主题包含的关键词数量，圆圈越大表明主题中的关键词越多，箭头表明两个主题之间的相似指数大于等于阈值，说明这两个主题之间存在演化关系，箭头越粗表明两个主题之间的相似程度越大。

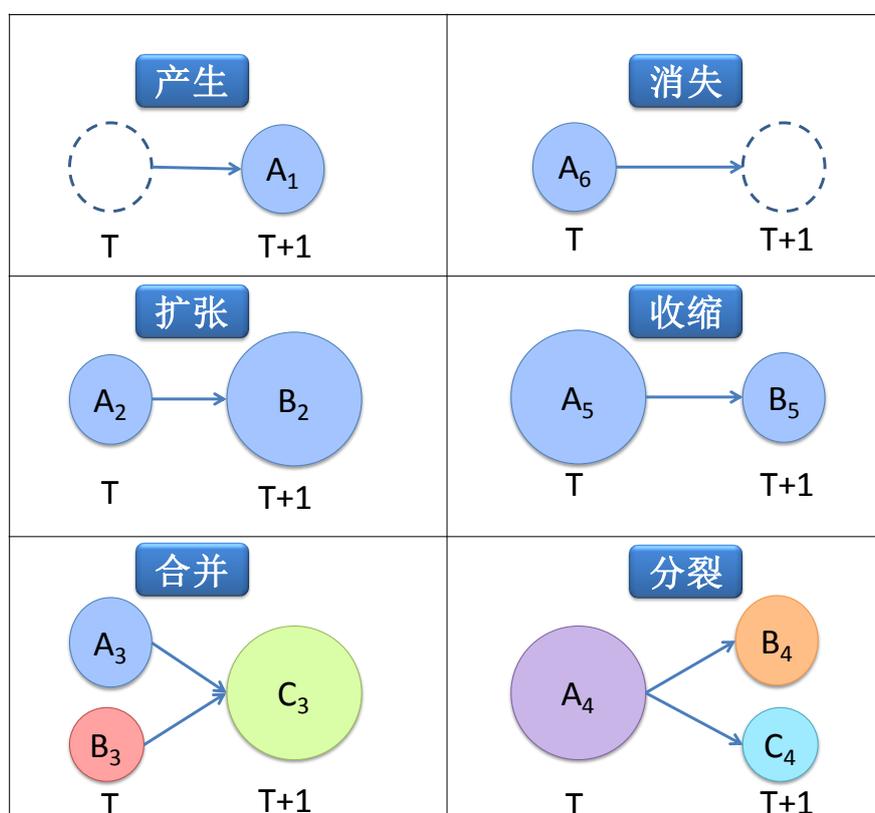


图 5 学科主题演化的 6 种现象

由于本文是对每个演化阶段中的高频关键词进行共词分析，揭示的是每个演化阶段中学科领域的热点主题。因此，以上 6 种学科主题演化现象仅仅分析热点主题的演变，本文给出了 6 种现象应满足的条件，如表 5 所示。

表 5 六种主题演化现象的定义及满足的条件

演化现象	定义	满足的条件
产生	指在 T 时间段不存在或还没成为热点的主题，在 T+1 时间段成为了热点主题。	T+1 时间段的主题 A1 与 T 时间段的所有主题的相似指数均小于阈值。
扩张	指前 T 时间段存在的热点主	T+1 时间段的主题 B <sub>2</sub> 与 T 时间段的主

	题,在 T+1 时间段继续存在,但包含的关键词数量增多。	题 A <sub>2</sub> 的相似指数大于等于阈值,且主题 B <sub>2</sub> 中的关键词数量大于 A <sub>2</sub> 。
合并	指前 T 时间段的两个或多个热点主题,在 T+1 时间段合并成一个新的热点主题。	T+1 时间段的主题 C <sub>3</sub> 与 T 时间段的主题 A <sub>3</sub> 和 B <sub>3</sub> 的相似指数均大于等于阈值。
分裂	指前 T 时间段的热点主题,在 T+1 时间段分裂成两个或多个新的热点主题。	T+1 时间段的主题 B <sub>4</sub> 和 C <sub>4</sub> 与 T 时间段的主题 A <sub>4</sub> 的相似指数均大于等于阈值。
收缩	指前 T 时间段存在的热点主题,在 T+1 时间段继续存在,但包含的关键词数量减少。	T+1 时间段的主题 B <sub>5</sub> 与 T 时间段的主题 A <sub>5</sub> 的相似指数大于等于阈值,且主题 B <sub>5</sub> 中关键词的数量少于主题 A <sub>5</sub> 。
消亡	指前 T 时间段存在的热点主题,在 T+1 时间段消失或不再成为热点主题。	T 时间段的主题 A <sub>6</sub> 与 T+1 时间段所有主题的相似指数均小于阈值。

## 5 小结

本文设计出了基于共词分析法的学科主题演化研究的方法体系,包括 4 个模块,分别是:数据准备、演化阶段划分、主题识别和主题演化分析。本文在词频 g 指数的基础上提出了改进的词频 g 指数;在学科主题演化分析模块,从静态和动态两个角度来分析学科主题的演化情况,本文在战略坐标的基础上提出了三维战略坐标来进行静态分析,三个维度分别为向心度、密度、强度,依据主题在三维战略坐标中的位置以及强度的大小将主题分为了 8 种类型,然后利用相似指数构建了学科主题演化现象识别模型来进行动态分析。在接下来的研究工作中,将选取某一研究领域来对本文设计出的基于共词分析法的学科主题演化研究方法体系进行实证分析,以进一步验证该方法体系的有效性。

[1]唐果媛,张薇. 基于共词分析法的学科主题演化研究进展与分析[J]. 图书情报工作,2015,59(5):128-136.

[2]Cambrosio A, Limoges C, Courtial J P, et al. Historical scientometrics? Mapping over 70 years of biological safety research with cword analysis[J]. Scientometrics, 1993, 27(2): 119-143.

[3]肖明,杨楠,李国俊. 基于共词分析的我国用户信息行为研究结构探讨[J]. 情报杂志,2010,29(S2):12-15,26.

- 
- [4]周玉芳. 知识图谱视野下科技查新研究的发展分析[J]. 现代情报,2012,32(6):25-28,32.
- [5]姜霖,王子朴,王晓虹. 基于 CSSCI 的体育人文社会学论文关键词分析[J]. 西南民族大学学报(人文社会科学版),2014(1):229-238.
- [6]魏瑞斌. 基于关键词的情报学研究主题分析[J]. 情报科学, 2006, 24(9):1400-1404.
- [7]杨爱青, 马秀峰, 张风燕等. g 指数在共词分析主题词选取中的应用研究[J]. 情报杂志, 2012, 31(2):52-55.
- [8]袁军鹏. 科学计量学高级教程[M]. 科学技术文献出版社, 2010:195-196.
- [9] Law J, Bauin S, Courtial J P, et al. Policy and the mapping of scientific change: A co-word analysis of research into environmental acidification[J]. *Scientometrics*, 1988, 14 (3) : 251-264.
- [10]Neff M W, Corley E A. 35 years and 160,000 articles: A bibliometric exploration of the evolution of ecology[J]. *Scientometrics*, 2009, 80(3): 657-682.
- [11]Callon M, Courtial J P, Laville F. Co-word analysis as a tool for describing the network of interactions between basic and technological research: The case of polymer chemistry[J]. *Scientometrics*, 1991, 22 (1) : 155-205.
- [12]Law J, Whittaker J. Mapping acidification research: A test of the co-word method[J]. *Scientometrics*, 1992, 23 (3) : 417-461.
- [13]Coulter N, Monarch I, Konda S. Software engineering as seen through its research literature: A study in co - word analysis[J]. *Journal of the American Society for Information Science*, 1998, 49 (13) : 1206-1223.
- [14] G P, AL B, T. V. Quantifying social group evolution[J]. *Nature*, 2007, 446(7136):664-667.