

一种基于共词网络社区的科研主题演化分析框架*

程齐凯 王晓光

[摘要] 共词网络在一定程度上可以表示特定学科领域的知识结构。为分析主题演化过程,将网络社区的演化分为 6 种类型,分别为产生、消亡、分裂、合并、扩张与收缩。在此基础上,利用 Z-value 算法和社区相似度算法,构建一个科研主题演化分析模型。与传统的基于词频的分析思路相比,所提出的基于共词网络社区演化分析的框架不强调词频的变化,而是强调词间关系的变化,试图通过中观层面的网络社区的演化分析揭示科研主题发展规律。

[关键词] 共词网络 网络社区 社区发现 网络演化 新兴趋势探测

[分类号] G202

1 引言

自 20 世纪 50 年代以来,科研产出增长迅速,新兴学科不断产生,学科体系结构日趋复杂,有效获取科研信息的难度逐渐加大。科技管理和科技信息服务机构纷纷加大了情报分析研究的力度,力求从科学文献中尽早发现新兴主题,并评估其发展趋势,以辅助科研决策。由此促使新兴趋势探测(emerging trend detection, ETD)研究成为情报学前沿课题。

随着各种数字图书馆的建设,从海量科学文献中识别科学发展脉络和学科发展趋势的数据基础已经具备。TOAS、ThemeRiver、TimeMine、CiteSpace II 等专门用于探测科研发展趋势的软件系统也随之产生,但是囿于对新兴科研主题的浮现机理的认识不足,多数新兴趋势探测系统的效果并不理想。为了开发更有效的探测方法和更高效的探测系统,必须加强科研主题演化规律的基础性研究,只有摸清规律,才能实现探测方法与手段的创新。为了分析科研主题的演化过程,本文提出一种新的基于共词网络的主题演化分析框架。

2 相关研究综述

2.1 新兴趋势探测

A. Kontostathis 在 2003 年提出了新兴趋势(emerging trend)的概念,即随着时间推移逐渐引起人们兴趣并被越来越多的学者讨论的主题领域^[1]。新

兴趋势探测有两个角度:一个是引文分析,另一个是主题词汇分析^[2]。引文分析存在时间滞后性问题,在新兴主题的揭示上不如词汇直观,因此研究者更倾向于使用时效性比较强的主题词汇分析^[3-4]。

Le. Minh-Hoang 将 ETD 过程分为三个阶段:主题表示(representation)、主题识别(identification)和主题判定(verification)^[5]。主题表示阶段主要利用大规模文献数据集进行主题抽取。这一阶段的关键是从语义上对代表同一概念的词汇进行归一化合并处理,避免因词形异化导致词频下降而被噪音淹没。在主题识别阶段,研究者常借助相似性计算和层次聚类来发现科研主题。随着数据分析规模扩大和复杂网络思想的引入,部分研究者开始使用动态网络可视化技术进行趋势分析^[6],并利用社区识别方法取代缺陷明显的层次聚类法进行主题识别,推动知识网络可视化分析成为趋势探测的前沿方法^[7]。2009 年, M. L. Wallace 等人利用复杂网络层次社区快速展开算法研究了共被引网络内的聚类问题,结果表明社区发现算法在科研主题识别上具有天然的优势^[8]。

判定科研主题的状态与趋势是 ETD 的一个难点。除了传统的词频统计法以外,研究者开始利用背景知识或使用更复杂的方法来进行新兴主题的判定。2006 年, Le. Minh-Hoang^[5]提出根据科研主题的 6 个属性值来衡量主题的受关注程度及有用性,在此基础上殷蜀梅于 2008 年提出了一套更全面的用于新兴趋势评估

* 本文系国家自然科学基金项目“基于语义共词网络演化的学科新兴趋势浮现机理与探测研究”(项目编号:71003078)和中央高校基本科研业务费专项资金资助项目(项目编号:2012104010204)研究成果之一。

[作者简介] 程齐凯,武汉大学信息管理学院博士研究生;王晓光,武汉大学信息管理学院副教授,博士,通讯作者,E-mail: whu_wxg@126.com。

收稿日期:2012-11-27 修回日期:2013-02-20 本文起止页码:91-96 本文责任编辑:易飞

的指标体系^[9]。2012年,Tu Yining和Seng Jialang又提出了新颖指数(NI)和已发表量指数(PVI)两个新型指标,以此来判断新兴主题^[10]。

综合国内外研究进展来看,随着知识网络和科学图谱研究的发展,从网络演化计量角度揭示科研主题的发展状态正在成为一种新颖的研究思路。基于以往的科研经验可知,判断一个主题是新兴主题、热门主题、衰退主题还是死亡主题,都必须考虑时间维度,即在考虑主题自身科学价值的同时,也必须考虑该主题及其相关主题在以往的表现情况,即是否出现过、何时出现以及近几年的发展情况。此外,识别一个学科领域的新兴主题,还必须考虑主题之间的关系,从主题演化过程中进行推断。任何科研主题都不是凭空出现的,都与以往曾经出现过的其他主题存在或多或少的联系。事实上,很多主题都是在以往的科研基础上逐渐孕育而生的,因此揭示科研主题的演化过程、规律和态势对于新兴主题的探测具有至关重要的意义。

2.2 共词网络与网络社区

从知识网络的形成原理和功能作用来看,知识网络的复杂结构与学科领域结构之间存在天然的对立关系。对知识网络的挖掘不仅可以实现学科潜在知识的发现,还能用于学科前沿的识别和趋势探测。共词网络是以文章关键词这种特殊的知识单元及其共现关系为基础构建的一类知识网络,它与引文网络以及共被引网络一样具有重要的方法论价值^[11-12]。

社区(community)是社会网络中的常见现象,它由一群高度聚集、联系紧密的节点聚集组成。社区结构是一种介于宏观和微观之间的网络特征。对于真实网络,同属于一个社区的顶点更有可能具有相似的性质或相近的功能。例如在WWW网络中,同一个社区的页面通常表达相近的主题;在神经网络中,一个社区通常对应一个功能组;在科研合作网络中,同一个社区内的学者有着相近的研究兴趣。网络中的社区结构有助于人们更深刻地理解网络结构和功能之间的关系。

随着网络社区研究的兴起,各种知识网络中的社区现象引起了学者的关注。M. E. J. Girvan和M. E. Newman的研究表明引文网络与合著网络内社区现象十分明显^[13]。K. W. Boyack等人利用期刊引文关系绘制的科学全局地图上也显示出明显的聚集现象^[14]。知识网络中社区的普遍存在意味着这些社区不是可有可无的。R. Lambiotte和P. Panzarasa认为知识网络中的社区与学科关系密切,可以被视为一种知识地图领域的划分机制和学科前沿方向标^[15]。

2.3 网络社区识别

在信息科学领域,社区识别存在两个方向:一种是基于拓扑关系的社区识别,另外一种是基于主题的社区识别^[16]。基于拓扑关系的社区识别主要依赖图论方法。这种识别方法主要依赖于网络拓扑关系,而不考虑网络节点和网络关系的性质,因而适合于任何复杂网络。基于主题的社区识别主要针对那些网络节点是一个或多个文本集合的网络,如博客网络和大学网络。这种识别主要依赖两个节点间拥有的主题相似性,在层次聚类基础上形成一个树状结构,以此显示哪些节点属于一个社区。相对而言,基于拓扑关系的社区识别方法比层次聚类法更有优势,因为它不需要预先设定聚类数目和确定树状图的层次切割点。

目前,理论物理学界和计算机学界已经基于图论思想提出了众多社区识别算法,最有代表性的一类方法是基于优化网络模块度(modularity)的方法。模块度是由M. E. J. Newman提出的衡量网络划分好坏的一种指标。模块度值,也叫Q值,其计算方法为:

$$Q = \sum_i (e_{ij} - a_i^2) \quad (1)$$

其中 e_{ij} 表示社区*i*和社区*j*之间的边数占总边数的比率; $a_i = \sum_j e_{ij}$ 表示有一个端点在社区*i*中的边占总数的比率。从本质上来说,基于模块度的算法是根据边的中介性和模块度的变化来进行社区识别的。提出模块度方法之初,该方法只能适用于无权网络。2004年,M. E. J. Newman又提出了一种新算法,将模块度算法扩展到了加权网络上^[17]。

2008年,K. W. McCain利用引文网络和社区发现算法进行了科研主题的识别研究^[18]。随后,M. L. Wallace、Y. Gingras和R. Duhon又利用两个案例研究证明了将社区发现方法用于研究方向的识别不仅是可行的,更是一种非常理想的思路,它能比传统的共被引分析揭示更多的知识领域的结构细节^[18]。

3 共词网络中的社区及其对应主题的表达

以往的研究表明共词网络内也存在社区现象,这些网络社区与学科体系存在一定的对应关系。不同层次的词汇社区代表了特定的学科、专业以及研究方向,因此共词网络中社区的演化在一定程度上揭示了科研主题的发展过程^[11-12]。

在确认了共词网络内的社区具有特定的指示性意义后,下一步就必须确定这些社区代表的主题。由于共词网络中的节点就是文章关键词,确定社区代表主

题的过程也就转化为寻找社区核心节点的过程,少数核心节点代表了社区对应的科研主题。

在复杂网络中,节点的重要性指标有很多,除了传统的中心度、声望等指标外,还有 Pagerank 值。这些指标都从网络全局层面进行考虑,计算每一个节点在整个网络中的边数、中介性以及与其他节点的连接情况,进而判断出全局层面的核心节点。这些指标虽然能揭示出每个节点在全局范围的地位,但无法揭示一个节点在一个特定社区内的重要性。

为了寻找社区内的代表性节点,笔者使用了 Z-Value 值。该指标由 R. Guimerà 等人提出^[19],它可以衡量网络节点与其他节点联系的紧密性,是一个在地区层面而非全局层面揭示节点重要性的指标。

Z-Value 定义如下:

$$z_i = \frac{k_{s_i}^i - \langle k_{s_i}^j \rangle_{j \in s_i}}{\sqrt{\langle (k_{s_i}^j)^2 \rangle_{j \in s_i} - \langle k_{s_i}^j \rangle_{j \in s_i}^2}} \quad (2)$$

其中 $k_{s_i}^i$ 表示节点 i 到社区 s_i 中其他节点的连接数, s_i 表示 i 所在的社区, $\langle \dots \rangle$ 表示平均数。Z-value 值越高,表明节点与其所在社区内其他节点联系越紧密。在使用 Z 值后,根据经验,共词网络中每个社区的代表节点,即对应的主题,就可以由一个或多个 Z-value ≥ 2.5 的节点表示。

4 共词网络中的社区演化

4.1 社区演化过程和社区演化状态

从知识社会学角度来看,知识的创造具有连续性,旧知识常常是新知识产生的基础。在特定的学科领域内,公认的基础理论、广泛传播的研究范式、薪火相传的研究团体等多种社会性因素都会导致知识创造与发展带有遗传式的性状,形成了诸如思想、流派、学说这种表现现象。部分情报学者曾使用“知识基因”、“情报基因”这种类比式的概念,解释这种现象存在的潜在逻辑。然而随着社会环境的变化、科研人员的替代、新物质和新规律的发现、观察实验和经验总结,科研主题在持续发展的同时,也会发生一定的突变,新主题不断产生的同时,旧主题也在不断消亡。新旧主题之间常常关系密切,新主题常在旧主题的消亡过程中孕育产生,这就形成了科研领域常见的研究主题演化现象。

在共词网络中,网络社区不是一成不变的。在不同的时间段内,网络社区的数量、大小、密度、结构等属性并不一致,网络社区的演化既包括社区自身内部节点、关系和结构的变化,也包括社区间关系和位置的变化。参考 G. Palla, A. L. Barabási 和 T. Vicsek 的做

法^[20] 笔者将网络社区的演化过程定义为 6 种形式:

定义 1: 产生: 指 t 时间段不存在的社区,在 $t+1$ 时间段产生;

定义 2: 消亡: 前 t 时间段存在的社区,在 $t+1$ 时间段没有存在;

定义 3: 分裂: 前 t 时间段的社区,在 $t+1$ 时间段分化成为两个或多个新的社区;

定义 4: 合并: 前 t 时间段的两个或者多个社区,在 $t+1$ 时间段合成一个新的社区;

定义 5: 扩张: 前 t 时间段存在的社区,在 $t+1$ 时间段继续存在,但规模扩大;

定义 6: 收缩: 前 t 时间段存在的社区,在 $t+1$ 时间段继续存在,但规模缩小。

演化过程是在连续时间段的社区关系的基础上定义的,仅仅观察一个时段的社区,则需要对社区在演化中的状态进行定义。受 G. Palla 等^[20] 和钱铁云等^[21] 的启发,笔者在上述演化过程定义的基础上进一步定义了社区演化的以下 6 种状态:

定义 7: 生成态: t 时段的社区内部形成了两个或者多个相对独立的潜在子社区,并在 $t+1$ 时段分裂形成多个社区,则该社区处于生成态。如果一个社区处于生成态,去除社区中少数节点,会形成两个或者多个内部密度较大的独立社区。

定义 8: 吸收态: t 时段的社区同其他社区产生较大的关联,并在 $t+1$ 时间段吸收其他社区的节点形成新的社区,如果该社区在新形成的社区内有着显著性的体现,则认为该社区处于吸收态。

一个社区可以同时处于生成态和吸收态,社区在分裂的同时,包含着对其他社区节点显著性的吸收,则该社区同时处于生成态和吸收态。吸收是方向性的,两个社区合并形成新社区的过程可能是以一方为主体的吸收行为,也可能是双方相互吸收融合的结果。

定义 9: 扩张态: 社区在后续时间段得以保持,但后继社区规模出现显著扩张。

定义 10: 浸润态: 社区在后续时间段得以保持,其后继社区规模无显著变化。

定义 11: 收缩态: 社区在后续时间段得以保持,其后继社区规模出现显著收缩。

定义 12: 消亡态: 社区在后续时间段不再存在,既没有分裂形成多个新社区,也没有吸收其他社区节点形成新的社区,则认为该社区处于消亡态。

演化过程和演化状态是对社区演化动态特征和静态特征的分别描述,两者存在着一定的对应关系。演

化状态反映了社区在演化过程中的可能性,并在后续时段通过一定的过程实现演化,状态的判定可以用于预测后续演化行为。

无论是演化过程的判定还是演化状态的确定,均涉及对 t 和 $t+1$ 两个连续时间段的社区关系的分析,而这可以进一步简化为为 t 时段的所有网络社区寻找前驱和后继。

4.2 社区前驱后继的发现方法

从“情报基因”的视角来看,社区的前驱后继发现问题,就是寻找相似性状的问题,即寻找 t 时间段和 $t+1$ 时间段内具有相似性的社区的问题。网络相似度计算是一个现实的难题,它既需要考虑节点相似性,也需要考虑关系的相似性。常见计算方法有三种:①点相似度,即前后两个连续时刻的社区之间点的重合度达到某个阈值,就认为这两个社区之间存在演化关系;②关系相似度,即前后两个连续时刻的社区之间关系的重合度达到某个阈值;③综合计算点相似度和关系相似度。第一种方法较为方便快捷;第二种方法对于加权网络来说,如何将权重放入重合度计算公式内是个难题;第三种方法也面临同样的难题。

寻找一个网络社区前驱后继的常用方法有三种,分别为分类方法、语言模型方法、相似度模型方法。这里首先给出一些概念的符号表示: G_t 表示为时间段 t 的网络, M_{ii} 表示 G_t 中编号为 i 的社区, G_{t+1} 是 G_t 对应的下一个时间段的共词网络, $Des(M_{ii})$ 表示社区 M_{ii} 的后继社区, $Pre(M_{ii})$ 表示社区 M_{ii} 的前驱社区。

分类方法将前一时间段的社区作为分类的训练集,前一时间段的社区作为分类类目,利用社区内节点的特征训练出分类器,将后续时间段的节点作为分类数据予以分类,从而建立后续时间段社区同前一时间段社区的关联。分类模型可以采用朴素贝叶斯、SVM 等。

语言模型方法利用网络社区的节点构建语言模型,从而将社区相似度的计算转化为语言模型的相似度计算。具体而言,对 $M_{ii} \in G_t$, 建立语言模型 LM_{ii} , 同样,为时间段 $t+1$ 的网络 G_{t+1} 的各个社区 $M_{(t+1)j}$ 建立语言模型 $LM_{(t+1)j}$, 定义 $M_{(t+1)j}$ 的前驱为:

$$Pre(M_{(t+1)j}) = (M_{ii} | M_{ii} \in G_t, d(LM_{ii}, LM_{(t+1)j}) < \delta) \cup \operatorname{argmin}_{M_{ii} \in G_t} (d(LM_{ii}, LM_{(t+1)j})) \quad (3)$$

其中 δ 是一个可调节的阈值; $\operatorname{argMin}(\cdot)$ 表示与 $G_{(t+1)}$ 中与 $M_{(t+1)j}$ 距离最小的社区,做析取运算可以保证 $G_{(t+1)}$ 中任一社区都能找到一个前驱; $d(LM_i, LM_j)$ 为语言模型 LM_i 和语言模型 LM_j 的距离函数。一般

地, d 使用 KL 距离:

$$KL(P || Q) = \sum_{x \in X} P(x) \log \frac{P(x)}{Q(x)} \quad (4)$$

相似度模型方法主要计算两个社区节点集合的相似性,如果前后两个连续时间段中的社区相似度超过一定阈值,则认为两个社区存在前驱后继关系,定义社区 $M_{(t+1)j}$ 的前驱为:

$$Pre(M_{(t+1)m}) = (M_{ii} | M_{ii} \in G_t, d(M_{ii}, M_{(t+1)j}) < \delta) \cup \operatorname{argmax}_{M_{ii} \in G_t} (d(LM_{ii}, LM_{(t+1)j})) \quad (5)$$

其中 δ 是可调节的阈值, d 是任意的相似度计算公式,如余弦相似度计算公式、Jaccard 距离、欧几里德距离等。为了确保每个社区都能找到前驱,需要将公式左部(以析取符号隔开)和右部做析取运算。

通过以上三种方法和一定的阈值,可以为任意时间段上的任意网络社区找到前驱与后继,从而揭示其演化路径,确定社区的演化状态。

4.3 社区状态的确定

给出社区演化各种状态的明确指标是非常困难的,但是,可以利用社区的前驱后继性质确定社区所处的状态。

社区生成态意味着社区内部形成了两个或者多个相对独立的潜在子社区,但这些潜在子社区在该时间段却并没有各自独立开来,而在后续时间段分裂。社区 M_t 处于生成态的界定指标为:

$$|Des(M_t)| > 1 \quad (6)$$

社区处于吸收态意味着社区将通过吸收其他社区子节点形成新的社区,且原社区在新社区中有着显著性的体现,社区 M_t 处于吸收态的界定指标为:

$$\exists M_{(t+1)} \in Des(M_t), |Pre(M_{(t+1)})| = 1 \text{ and } Sig(M_t, M_{(t+1)}) > \delta \quad (7)$$

其中 δ 是一个阈值参数。 $Sig(A, B)$ 表示社区 A 在社区 B 中的显著性度量,如果社区 A 的元素显著地表现在社区 B 中,则 $Sig(A, B)$ 值较大。 $Sig(A, B)$ 可以使用 Jaccard 系数简单表示。

社区 M_t 在后续时间段继续存在,没有出现分裂、融合或者消亡现象,即 $|Des(M_t)| = 1$, 则 M_t 可能处在扩张态、浸润态、吸收态三种状态之一。用社区规模函数 $g(M_t)$ 来表示 M_t 的规模,简单起见,可以令 $g(M_t) = |M_t|$, $|M_t|$ 表示社区的节点数量。

社区处在扩张态的界定指标为:

$$\begin{cases} |Des(M_t)| = 1 \\ g(M_{t+1}) - g(M_t) > \xi \end{cases} \quad (8)$$

其中 M_{t+1} 是 M_t 的后继, ξ 是一个阈值参数且 $\xi >$

0 表示社区在后续时间段的扩张幅度。

如果社区处于浸润态,则社区在后续时间段得以保持,且后继社区规模(结构)无显著变化。定义社区浸润态的界定指标为:

$$\begin{cases} |Des(M_t)| = 1 \\ |g(M_{t+1}) - g(M_t)| \leq v \end{cases} \quad (9)$$

其中 v 为阈值参数且 $v \geq 0$ 。

如果社区处在收缩态,社区在后续时间段得以保持,其后继社区出现显著性的收缩。社区收缩态的界定指标为:

$$\begin{cases} |Des(M_t)| = 1 \\ |g(M_t) - g(M_{t+1})| \geq v \end{cases} \quad (10)$$

其中 v 是一个阈值参数且 $v > 0$ 。

社区处于消亡态,意味着社区不存在后继。显然,社区处于消亡态的界定指标可以定义为:

$$|Des(M_t)| = 0 \quad (11)$$

5 结 语

为了分析科研主题的演化规律,本文提出了一种新的基于共词网络的科研主题演化分析框架,在此框架下给出了网络社区对应主题表示算法,定义了社区演化事件类型、社区演化状态以及社区演化关系的判断算法。与以往的科研主题演化分析思路不同,本文不强调词频的变化,而是将复杂网络和网络演化思想引入情报分析过程,强调词间关系的变化,试图在网络视角下,通过中观层面的网络社区的演化分析发现科研主题发展规律。

无论从情报学角度还是从复杂网络的角度来看,判断两个科研主题是否存在演化关系以及存在何种演化关系都是一个难题。首先从情报学角度来看,对于科研主题之间关系的判断需要结合特定的学科背景,深入分析这两个主题的研究目标、发展历史、研究人员等信息;另一方面,从复杂网络角度来看,对网络社区演化的判断需要详细分析两个社区的节点、边、结构等信息,即使两个社区拥有相似的节点、边和结构,判定两个社区之间的关系类型也不容易。在没有公认的判定标准的条件下,以相似度作为社区演化关系的判定标准是最为可行的方法。但在此过程中,如何设定阈值又是一个难题。阈值的大小将直接决定社区演化关系的判定,对于不同的网络是否设定阈值、设定什么样的阈值,还需在实践中摸索确定。

参考文献:

[1] Kontostathis A, Galitsky L M, Pottenger W M, et al. A survey of

emerging trend detection in textual data mining [C]//Berry M. A Comprehensive Survey of Text Mining. Heidelberg: Springer-Verlag, 2003: 185-224.

- [2] 陈仕吉. 科学研究前沿探测方法综述[J]. 现代图书情报技术, 2009(9): 28-33.
- [3] Pottenger W M, Yang Tinghao. Detecting emerging concepts in textual data mining [J]. Computational Information Retrieval, 2001, 100(1): 89-105.
- [4] Roy S, Gevry D, Pottenger W M. Methodologies for trend detection in textual data mining [C/OL]. [2012-09-06]. <http://www.cse.lehigh.edu/~billp/pubs/ETDMethodologies.pdf>.
- [5] Le Minh-Hoang, Ho Tu-Bao, Nakamori Y. Detecting emerging trends from scientific corpora [J]. International Journal of Knowledge and Systems Sciences, 2005, 2(2): 63-69.
- [6] Chen Chaomei. CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature [J]. Journal of the American Society for Information Science and Technology, 2006, 57(3): 359-377.
- [7] Börner K, Chen Chaomei, Boyack K W. Visualizing knowledge domains [J]. Annual Review of Information Science & Technology, 2003, 37(1): 179-255.
- [8] Wallace M L, Gingras Y, Duhon R. A new approach for detecting scientific specialties from raw cocitation networks [J]. Journal of the American Society for Information Science and Technology, 2009, 60(2): 240-246.
- [9] 殷蜀梅. 判断新兴研究趋势的技术框架研究[J]. 图书情报知识 2008(5): 76-80.
- [10] Tu Yining, Seng Jialang. Indices of novelty for emerging topic detection [J]. Journal of Information Processing and Management, 2012, 48(2): 303-325.
- [11] 王晓光. 科学知识网络的结构与演化(I): 共词网络方法的提出[J]. 情报学报, 2009, 28(4): 599-605.
- [12] 王晓光. 科学知识网络的结构与演化(II): 共词网络可视化与增长动力学[J]. 情报学报, 2010, 29(2): 314-322.
- [13] Newman M E J, Girvan M. Finding and evaluating community structure in networks [J]. Physical Review E, 2004, 69(2): 026113.
- [14] Boyack K W, Klavans R, Börner K. Mapping the backbone of science [J]. Scientometrics, 2005, 64(3): 351-374.
- [15] Lambiotte R, Panzarasa P. Communities, knowledge creation, and information diffusion [J]. Journal of Informetrics, 2009, 3(3): 180-190.
- [16] Ding Ying. Community detection: Topological vs. Topical [J]. Journal of Informetrics, 2011, 5(4): 498-514.
- [17] Newman M. Analysis of weighted networks [J]. The American Physical Society 2004, 70(5): 56-64.
- [18] McCain K W. Assessing an author's influence using time series historic graphic mapping: The oeuvre of Conrad Hal Waddington

- (1905 - 1975) [J]. Journal of the American Society for Information Science and Technology, 2008, 59(4): 510 - 525.
- [19] Guimer'a R, Sales-Pardo M, Amaral L A. Classes of complex networks defined by role-to-role connectivity profiles [J]. Nature Physics, 2007, 3(1): 63 - 69.
- [20] Palla G, Barabási A, Vicsek T. Quantifying social group evolution [J]. Nature, 2007, 446(7136): 664 - 667.
- [21] 钱铁云, 李青, 许承瑜. 面向科技主题发展分段的社区核心圈技术 [J]. 计算机科学与探索, 2010(2): 170 - 179.

A New Research Frame for Analyzing the Evolution of Research Topics Based on Co-word Network Communities

Cheng Qikai Wang Xiaoguang

School of Information Management, Wuhan University, Wuhan 430072

[Abstract] The knowledge structure of a discipline can be expressed by a co-word network. Six types of evolution of community are recognized, which are birth, death, splitting, merging, growth, and contraction. A topic evolution analysis framework is created based on Z-value mechanism and community match mechanism. Comparing with the traditional methods, the method proposed in this paper focus on the links between words rather than the frequency of words, which is a innovation in the field of ETD from the mesoscopic network view.

[Keywords] Co-word Network Network Communities Community Detection Network Evolution Emerging Trend Detection

(上接第 77 页)

- [11] 刘启柏. 古籍防霉 [J]. 四川图书馆学报, 1979(3): 93 - 102.
- [12] 周崇润. 图书纸张挥发物对其耐久性的影响 [C] // 中国国家图书馆. 中文善本古籍保存保护国际研讨会论文集. 北京: 北京图书馆出版社, 2002: 273 - 281.
- [13] 邱晓刚. 对使用“羧甲基纤维素”作为胶粘剂修补古籍的分析 [J]. 广东图书馆学报, 1990(3): 91 - 90.
- [14] 周崇润, 李景仁. 修裱防腐剂硫酸铜对文献纸张耐久性的影响 [C] // 中国国家图书馆. 中文善本古籍保存保护国际研讨会论文集. 北京: 北京图书馆出版社, 2002: 265 - 272.
- [15] 吴余佩娴, 林焯荣. 香港中文大学图书馆中文古籍管理与保存: 不同阶段的发展和挑战 [C] // 中国国家图书馆. 中文善本古籍保存保护国际研讨会论文集. 北京: 北京图书馆出版社, 2002: 45 - 52.
- [16] 王国强, 孟祥凤. 中国古代文献保护方法的现实价值 [J]. 图书情报工作, 2012, 56(3): 104 - 108.
- [17] 亚尔地斯. 书的护持和糟踏 [M] // 书的礼赞. 叶灵凤, 译. 北京: 三联书店, 1998: 69 - 75.

Study on the Basic Path of Protection Technology System Construction of Chinese Ancient Books

Wang Guoqiang

Department of Information Management, Zhengzhou University, Zhengzhou 450001

[Abstract] The mixed phenomenon of old and new protection technologies of contemporary Chinese ancient books shows the disorder both in path and practice. The basic path of protection technology system construction of Chinese ancient books is the fundamental and overall problem about protection technology system construction and practice orientation. And it is mainly based on Chinese traditional books protection technology, and supplemented by contemporary physical technology and chemical technology. The path conforms to the ancient books protection principles, adapts to the physical characteristics of protection objects and traditional protection technology, and is verified in Chinese ancient books protection practice.

[Keywords] ancient book ancient books protection ancient books protection technology