

doi :10.3772/j.issn.1000-0135.2009.03.012

## CiteSpace II 科学文献中新趋势与新动态的识别与可视化

陈超美(著)

(Drexel 大学信息科学与技术学院, 美国费城)

陈悦<sup>1,2</sup> 侯剑华<sup>1</sup> 梁永霞(译)

(1.大连理工大学 21 世纪发展研究中心 WISE 实验室, 大连 116024; 2.浙江大学公共管理学院, 杭州 310027)

[编者按] 此文是国际著名信息可视化专家陈超美博士的代表作,该文系统阐述了他所研制的用于文献引文网络分析的 CiteSpace II 可视化软件的原理、特点及应用案例。它属于多元、分时、动态的第二代信息可视化技术。其独到的创新之处,在于用 CiteSpace II 绘制的一幅科学知识图谱上,能够显示一个学科或知识域在一定时期发展的趋势与动向,形成若干研究前沿领域的演进历程。如今,知识图谱和信息可视化在各国蓬勃兴起,但陈超美博士开发的历时性动态可视化技术迄今尚十分鲜见,被公认于居国际领先水平。目前已升级到可自动聚类与术语标识的 CiteSpace2-2 最新版本,但本文仍是这一技术的基础性文献。为在我国推广和普及作为科学计量学前沿的科学知识图谱与知识可视化方法,征得作者同意,特在本期全文发表此文的中译本。

**摘要** 本文介绍了在科学文献中识别并显示科学发展新趋势和新动态的一种通用方法的最新研究进展。这项研究在理论和方法上极大地促进了知识领域可视化研究。研究领域 (specialty) 的概念和可视化基于信息科学中的两个概念——“研究前沿”和“知识基础”间的时变对偶 (time-variant duality)。研究前沿 (research front) 被定义为一组突现的动态概念和潜在的研究问题。研究前沿的知识基础 (intellecture base) 是它在科学文献中 (即由引用研究前沿术语的科学文献所形成的演化网络) 的引文和共引轨迹。Kleinberg 设计的跳跃检测算法 (burst detection algorithm) 适用于辨认新兴研究前沿专业术语概念<sup>[1]</sup>。Freeman 提出的中间中心性测度可以用来使潜在范式变化的关键点凸显出来<sup>[2]</sup>。我们设计并实现了两个互补的视图: 聚类视图 (cluster views) 和时区视图 (time-zone views)。这种方法的贡献在于: ①通过对研究前沿术语的算法运算,在动态中认识知识基础的本质; ②用研究前沿专业术语概念明确标出共引聚类的确切含义; ③直观地和靠算法识别的关键点的一致性大大简化了可视化的复杂性。CiteSpace II 应用 Java 程序实现了大规模生物集群灭绝 (mass extinction) (1981~2004 年) 和恐怖主义 (terrorism) (1990~2003 年) 两个研究领域的建模和可视化过程。可视化网络中的突出的趋势和关键点的作用经各自领域专家直接验证,这些专家本身就是关键点文章的作者。本文讨论了这项研究的实际意义,并明确了今后研究工作中存在的一系列挑战和机会。

**关键词** CiteSpace 信息可视化 科学前沿图谱 知识传播

### CiteSpace II : Detecting and Visualizing Emerging Trends and Transient Patterns in Scientific Literature

Chen Chaomei

(College of Information Science and Technology, Drexel University, Philadelphia, USA)

收稿日期: 2007 年 8 月 24 日

作者简介: 陈超美, 男, 1960 年生, 美国德雷赛尔大学信息科学与技术学院副教授 (终身教职), 1983 年获南开大学理论数学学士, 1991 年获牛津大学计算硕士, 1995 年获利物浦大学计算机学博士, 主要学科领域与研究方向: 信息科学、信息可视化和科学计量学, 包括信息与知识的可视化技术、复杂网络分析、科学前沿图谱绘制、科学技术发展新态势探测等方向。E-mail: chaomei.chen@cis.drexel.edu

Chen Yue<sup>1,2</sup> Hou Jianhua<sup>1</sup> and Liang Yongxia<sup>1</sup>

(1. Center for Development Studies on 21st Century, WISE Lab, Dalian University of Technology, Dalian 116024;

2. National Institute for Innovation Management, Zhejiang University, Hangzhou 310027)

**Abstract** This article describes the latest development of a generic approach to detecting and visualizing emerging trends and transient patterns in scientific literature. The work makes substantial theoretical and methodological contributions to progressive knowledge domain visualization. A specialty is conceptualized and visualized as a timevariant duality between two fundamental concepts in information science: research fronts and intellectual bases. A research front is defined as an emergent and transient grouping of concepts and underlying research issues. The intellectual base of a research front is its citation and co-citation footprint in scientific literature— an evolving network of scientific publications cited by research-front concepts. Kleinberg's (2002) burst-detection algorithm is adapted to identify emergent research-front concepts. Freeman's (1979) betweenness centrality metric is used to highlight potential pivotal points of paradigm shift over time. Two complementary visualization views are designed and implemented: cluster views and time-zone views. The contributions of the approach are that ① the nature of an intellectual base is algorithmically and temporally identified by emergent research-front terms, ② the value of a co-citation cluster is explicitly interpreted in terms of research-front concepts, and ③ visually prominent and algorithmically detected pivotal points substantially reduce the complexity of a visualized network. The modeling and visualization process is implemented in CiteSpace II, a Java application, and applied to the analysis of two research fields: mass extinction (1981 ~ 2004) and terrorism (1990 ~ 2003). Prominent trends and pivotal points in visualized networks were verified in collaboration with domain experts, who are the authors of pivotal-point articles. Practical implications of the work are discussed. A number of challenges and opportunities for future studies are identified.

**Keywords** CiteSpace, information visualization, science frontiers mapping, knowledge communication

## 1 引 言

根据引文半衰期(citation half-lives)的明显不同,科学文献可分为持续高被引的经典文献(classic articles)和在短暂时间内达到被引峰值的过渡文献(transient articles)<sup>[3]</sup>。过渡文献比经典文献更为普遍<sup>[4]</sup>。一篇研究文献被持续引用的平均时间长度与潜在研究领域的发展速度密切相关<sup>[5]</sup>。了解过渡文献转变为科学领域的动力机制,对于各学科领域的科学家都具有重要的现实意义。

科学文献的新趋势(Emergent trends)和突变(abrupt changes)是有内部和外部原因的。典型的内因包括新发现和科学重大突破,例如,在大规模生物集群灭绝(mass-extinction)中陨石坑的发现和天文学中超大规模黑洞的发现。外因是指那些可能引发科学家从新角度研究问题的事件。例如,9.11恐怖主义袭击引发了在国家安全、国家卫生、创伤后应激障碍(post-traumatic stress disorder, PTSD)等领域一系列新课题的提出。识别和理解由这些事件引发的学科发展新趋势和突变能极大的提高科学家适时应对这些变化的能力。需要指出的是复杂系统中的大规模变化在没有明显引发事件时可能表现出自组织临界性(Self-organized criticality, SOC)<sup>[6]</sup>。尽管也有人认

为科学文献的增长同自组织临界性相关<sup>[4]</sup>,但本文主要探讨与重大事件相关的变化。

普赖斯最早提出“研究前沿”的概念,用它来描述研究领域的动态本质。普赖斯观察到科学家似乎倾向于引用最新发表的文章,并将其称为即时因子(immediacy factor)<sup>[3]</sup>。某个领域的研究前沿是由科学家积极引用的文章所体现的。普赖斯认为,某个研究前沿大概由40~50篇最近发表的文章组成。迄今为止,对研究前沿至少有3种不同的认识形态:①共被引文献聚类<sup>[7,8]</sup>,②共被引文献聚类和所有引用这个聚类的文章<sup>[9,10]</sup>,③引用共群文章的文献聚类<sup>[11]</sup>。

知识基础是一个有利于进一步明晰研究前沿本质的概念<sup>[12]</sup>。如果我们把研究前沿定义为一个研究领域的发展状况(如研究思路),那么研究前沿的引文就形成了相应的知识基础。一个研究领域可以被概念化成一个从研究前沿 $\Psi(t)$ 到知识基础 $\Omega(t)$ 的时间映射 $\Phi(t)$ ,

$$\text{即 } \Phi(t): \Psi(t) \rightarrow \Omega(t)$$

我们研究的目的是要开发出一种能够识别和显示 $\Phi(t)$ 随时间发展的新趋势和突变的普适方法。在这里, $\Psi(t)$ 是一组在 $t$ 时刻与新趋势和突变密切相关的词和短语(如专业术语),这些专业术语被称作研究前沿术语。 $\Omega(t)$ 是由出现前沿术语的文章引

用的大量文章组成,它们之间的关系如下:

$$\Phi(t): \Psi(t) \rightarrow \Omega(t)$$

$$\Psi(t) = \{ \text{term} \mid \text{term} \in S_{\text{title}} \cup S_{\text{abstract}} \cup S_{\text{descriptor}} \cup S_{\text{identifier}} \wedge \text{IsHofTopic}(\text{term}, t) \}$$

$$\Omega(t) = \{ \text{article} \mid \text{term} \in \Psi(t) \wedge \text{term} \in \text{article}_0 \wedge \text{article}_0 \rightarrow \text{article} \}$$

$S_{\text{title}}$  表示一系列标题专业术语,  $\text{IsHofTopic}(\text{term}, t)$  表示布尔函数,  $\text{article}_0 \rightarrow \text{article}$  表示  $\text{article}_0$  引用  $\text{article}$ 。本文介绍两个实例应用,说明内在因素和外在因素对研究领域发展的影响。

本文逻辑结构如下:首先,对检测和显示动态网络中新趋势和突变的研究,以及对研究前沿进行定量研究的相关文献进行了概述。然后,介绍了一种用于识别和显示科学文献中发展趋势与突变(尤指由引发事件而产生的变化)的新方法。这项研究能使知识领域分析专家和科学家清晰地辨别和认识科学发展的结构和态势。我们用大规模生物集群灭绝研究(1981~2003年)和恐怖主义研究(1990~2003年)两个实例来阐明这种新方法的应用,最终的研究前沿视图经该领域专家鉴定而被证实,分析结果也经大量相关文献内容的查阅而得到考证。

## 2 识别和跟踪研究领域的演变

研究者们对用于辨别和跟踪学科发展前沿的定量方法已经做了大量的研究。研究前沿的动态本质,使得科学家、科研政策制定者和许多追赶科学快速发展的人们面临着巨大的挑战。理解研究前沿演变的动力机制对科学家、分析者和决策者辨认科学发展的新趋势和突变是非常关键的。

### 2.1 研究前沿和知识基础

研究前沿代表了一个研究领域的思想现状。普赖斯<sup>[3]</sup>注意到一个非常有趣的现象,即在科学引文网络中,越是频繁被引用的文章越可能是新近发表的文章。这种即时效果很好地解释了一些文章在发表很多年后将会被完全遗忘的现象。普赖斯将由 Burton 和 Kebler 早在 1960 提出的猜想表达为期刊文献可能是由两种有着不同半衰期的文献组成——经典文献和过渡文献,过渡文献实质上是对应于研究前沿<sup>[13]</sup>。普赖斯认为,“研究前沿是基于新近研究成果的,网络也变得越来越紧密”,他估计一篇文章之前有 30~40 篇相关文章发表,由此构成研究前沿。普赖斯将 200 篇有关  $N$ -rays 主题的文章按年代

顺序进行排列,通过引文矩阵(列引用行)来预测学科的研究前沿,研究前沿的包容界限是一篇引文发表前的大约 50 篇文章。很明显,通过研究相对小的文献网络的变化是有助于追踪不可计数的大量文献的发展轨迹。

与研究前沿相关的主要问题有:它是如何开始的?现状如何?其演化中的重要路径是什么?为了回答这样的问题,我们需要识别和分析新趋势的出现和与研究前沿密切相关的突变。我们还需要确定在某个特定时间上,基于相关知识基础的研究前沿的关注焦点,进而揭示导致研究前沿演变的重要知识转折点(intellectual turning points),并发掘不同研究前沿间的联系。

很多研究者从不同的角度研究学科领域的演变。Small 和 Griffith 认为,共被引文章聚类标志着当前活跃的研究领域<sup>[8]</sup>。共引聚类的本质决定于选择的词集(word profiles)。假设一个聚类中的文章  $D_i$  被  $K$  篇文章  $D_{i1}, D_{i2}, \dots, D_{ik}$  引用,选出在  $D_{ij}$  文章标题中最常出现的四个词形成文章  $D_i$  的词集。集合一个聚类中所有词集,选出出现频次最高的  $N$  个词,就形成了  $N$ -词聚类库( $N$ -word cluster-profile)。这种词库的最大优势就是简单,然而,初始词集的选择因受到  $K$  篇引用文献的标题限制,可能不足以揭示潜在主题领域的动态变化。

Braam 等将一个研究领域定义为“一群科学研究者关注的一系列相关问题和概念”<sup>[14]</sup>。他们根据共引聚类间的相似性来分析一个领域发展的连续性和稳定性。两个共引聚类的相似性依据于聚类词集的比较,尽管他们还没有提供图论或是其它的可视图形来说明研究领域从一个阶段到另一个阶段的变化,但他们辨别出同一个学科在不同发展阶段的一系列相似共引聚类。

Persson 基于 1986~1990 年间在《美国信息科学学会会刊》上发表的一篇关于研究前沿和知识基础的文章中认为研究前沿和知识基础的区别在于:从文献计量学来看,引文形成了研究前沿,被引文献组成了知识基础。<sup>[12]</sup>知识基础由共引聚类来表示,通过逐渐降低共引的阈值,他提出了所谓的知识基础的逐级扩大(stepwise enlargement),即随着共引阈值的降低,共引聚类中的文章越来越多,从而知识基础也逐渐增大。他还发现知识基础在相当长的一段时间内非常稳定。他注意到研究前沿形成的时候,如果被引文献的活跃度没有严格的限制,那么研究前沿聚类和知识基础聚类之间的对应关系也不是很

明显。在本文中,我们将研究领域定义研究前沿和知识基础两者之间的时变映射。

Morris 等绘制了研究前沿时间线索( time-line )可视图<sup>[11]</sup>,将研究前沿定义为持续被一组固定的、与时间无关的基本文章引用的大量文章。研究前沿的文章是基于文献耦合聚类而成的<sup>[15]</sup>。

美国科学信息所的 Henry Small 组织开发了一个基于网络的服务,即专题系列( The Special Topics series <http://www.esi-topics.com> )这个系列依据引文库记录了科学研究领域中的重大进步,Small 为之探索了 10 余年<sup>[8,16-18]</sup>。共被引文献聚类是从每年的引文数据中提取出来的。将相邻几年的聚类进行比较而辨别出的新要素作为新趋势的标识。相邻几年的共引文献聚类间相互重叠的程度为识别研究热点的转移提供了一种途径。表 1 总结了有关研究前沿和知识基础的不同界定。

在本文中,我们没有将引文局限在固定的、与时间无关的知识基础上,而是期望一个研究领域的知识基础将会同其研究前沿一起随着时间的变化而演变。这是因为我们定义的研究前沿是强调新趋势和突变的特征。研究前沿是时间变量映射的定义域,其知识基础是映射的值域( co-domain )。我们将研究前沿的知识基础重新简单定义为研究前沿在文献中的引用轨迹。

我们要弄清楚研究前沿的变动在多大程度上同其知识基础的稳定性相耦合。外在的事件会对研究前沿和知识基础间的内在联系产生怎样的影响也是我们关注的问题。例如,研究前沿的进程是如何通

过令人信服地发现而改变的?

## 2.2 识别新趋势

识别科学研究的新趋势和跟踪研究热点已越来越引起研究者的兴趣。尽管通过统计方法来确定发展趋势的做法仍很普遍<sup>[19,20]</sup>,但是已经有一些研究者开始进行动态可视化研究并用以辅助趋势探测<sup>[21]</sup>。

有关主题识别和跟踪的研究主要集中在以下五个问题:①叙述分段( story segmentation ),②主题探测( topic detection ),③主题跟踪( topic tracking ),④新闻的首次报道( first-story detection )和⑤叙述链接识别( story-link detection )。Roy 等总结了目前已有的趋势探测方法<sup>[20]</sup>。Kontostathis ,Galitsky ,Pottenger ,Roy 和 Phelps 全面总结了在文本挖掘中进行趋势探测的研究<sup>[22]</sup>。

Allan ,Papka 和 Lavrenko 介绍了一种用于鉴别新闻( new stories )的一次扫描算法( single pass algorithm )<sup>[23]</sup>。如果在现有文件中没有与某个新接收到的信息相类似的事件,那么这个新接收到的信息就被定义为一条新闻。这个算法能够很好的从早些的世界贸易中心爆炸事件中区分出俄克拉荷马州城市爆炸事件,但是还无法从其他法院的案例中确认出 O.J. Simpson 案。

Swan 和 Allan 建立了一个基于统计模型的二项式分布,以确定检索信息的显著性<sup>[24]</sup>。他们通过 CNN 广播新闻和路透社新闻专线(这是用来进行主题识别和跟踪研究的语料库)来检测他们的方法。

表 1 研究前沿和知识基础的定义

作者	年份	研究前沿	知识基础	聚类	标签
Price	1965	对于一篇指定引文,是由被频繁引用的近期文章( 30 ~ 50 篇)所组成的动态聚类	未定义	引文的最近行为( recentness )	
Small & Griffith	1974	共引聚类	未定义	共引	根据从引文中提取出的词集标注的被引文献
Braam et al.	1991	集中关注的一系列相关问题和概念	共引	词集	
Garfield	1991	共引聚类与引文的总和	共引		
Persson	1994	引用相同文献的文章	研究前沿的引文映像	共引	由标题词标注的文章
Morris et al.	2003	经常被一组固定的、与时间无关的基本文章引用的一组文章	固定的、与时间无关的文章群	文献耦合	由人工提取标题词标注的聚类
CiteSpace II ( this article )		正在兴起的理论趋势和新主题的涌现	共引网络	共被引文章和引用这些文章的术语的复合网络	从题目、摘要中提取的专业术语和出现频率突然增加的专业术语

文本分析有助于我们鉴别出某一段时期内文章中最热门的词语,通过这种热门词语的变化可以捕捉到潜在的主题变化。ThemeRiver 是一个可视化的系统,它用河流做隐喻,来描述文章主题随时间的变化<sup>[25]</sup>。主题的变化随着外部事件的时间线索而显示出来。主题河是由术语的频次支流组成,支流的宽度依据术语在不同时间段上出现频次的不同而发生变化。某个外部事件的出现可能会伴随着主题强度的突然变化。

Erten 等使用他们开发的 TGRIP 系统针对 ACM 会议论文集中的论文绘制了时序可视图(a temporal graph visualization)以寻找计算领域的研究热点主题、逐渐过时的主题和快速发展的主题<sup>[21]</sup>。他们的类图(category graph)由多个时间分区组成。不同权重(按变化率设定)的边表示出相邻时间分区的变化。每一年主题变化趋势是由使用比例最高的前五个标题词来识别。像“设计”(design)、“系统”(system)和“模拟”(simulation)这些词始终位于前列,而“自动辅助设计”(ADA)、“数据库”(database)和“并行(运算)”(parallel)这些词在列表中只是出现了几年,随后消失。他们的研究没有使用引文数据。

在 CiteSeer 文件数据库中,有用聚类方法来鉴别动态发展趋势的记录<sup>[19]</sup>。首先,基于引文挑选出有影响的文章以生成聚类,其次,由与聚类中的种子文章共被引的文章再组成一个新的聚类。两个聚类的相似性取决于它们共有的文章。各聚类的内涵是由各自包含文章中高频使用的标题词来标识的。

### 2.3 研究领域的识别和可视化

大多数传统的共引分析都集中在共被引文章的单个聚类,除了 Griffith 等,Small 是分析学科关系的,其余很少看到侧重于学科聚类间相互联系的研究<sup>[26,27]</sup>。Griffith 等发现聚类间共引链接强度要弱于聚类内的共引链接<sup>[26]</sup>。

在社会网络中,一个成员的影响可能依赖于他与网络中其他成员的关系。Granovetter 强调了弱连接(weak ties)的作用<sup>[28]</sup>。Burt 提出了结构洞的概念<sup>[29]</sup>。聚类内的连接往往要强于聚类间的连接。传统的引文网络分析集中在强连接(strong link)(如聚类内连接)。如果要想了解不同学科和不同主题间如何相互作用的,那么对其长期发展的本质,聚类间联接的研究以及搞清不同学科文献相互联系的原因是非常必要的。

点的中心性是一个用以量化点在网络中地位重  
万方数据

要性的图论概念。中间中心性是常用来进行中心性测度的指标<sup>[2]</sup>,它是指网络中经过某点并连接这两点的最短路径占这两点之间的最短路径线总数之比。中间中心性高的点往往位于连接两个不同聚类的路径上,用来确定群体的算法(community-finding algorithms)就是利用这个特点来区分网络中的聚类<sup>[30]</sup>。

中心性测量为发现不同学科的连接点或进化网络中的支点(tipping points)提供了一种计算方法。这种图论方法的优势在于,因为它独立于任何知识领域,所以其应用范围就极其广泛。而且这种方法只研究网络中少量的连接点,而不是整个网络,这样就有了很大的实际应用价值。首先,如果这些连接点显示出某个领域的结构和动态本质,那么就可以大大减少用户在理解上的负担。其次,接下来的自动文本概要的生成(automatic text summarization)和自然语言处理算法(natural language processing algorithms)处理过程,可以很有效地将研究重点集中在为数不多的连接点上。第三,连接点的鉴别是基于整体图论特征,而词集是基于初次出现频次,因而关键点法(a pivotal-point-based approach)扩展了词集法(a word-profile-based approach)。

在信息科学和其他学科中对引文网络和共引网络的研究已经有了长足的发展。加菲尔德的历史图表是描述引文链接的论文有向图,HisCite 是依此思路最近开发的工具<sup>[31]</sup>。相对而言,Small 等研究的共引网络是一个无向图<sup>[8,32]</sup>。SCI-Map 是由美国科学信息所早期开发的共引网络可视化系统,它没有使用通常的多维尺度分析,而是使用了基于地理三角剖分处理(geometric triangulation process)的二维显示算法。三角剖分处理首先将一篇人为选择的文章映射到坐标系的原点,然后将与第一篇文章共被引关系最强的文章置于恰当的位置,第三篇文章的位置是使用前两个位置距离三角剖析出来的。这种方法体现出最强链接的优势。最终,网络由中心向外逐渐扩展。Small 认为,处理过程依赖于研究对象的排序是这种方法的缺陷<sup>[17]</sup>。SCI-Map 以引文次数的立方根为半径,这样球体积恰好为引文次数。

有选择地控制连接的密度,也就是网络的精简,在诸如引文网络这样的复杂网络可视化研究中是一个有挑战性的实际问题。现实中的共引网络往往有大量而错综复杂的立连接。适当减少连接数量能使网络的主要结构更清晰更明显。

多维尺度的空间构型无法清晰表示连接。而最

小生成树(minimal spanning trees)和路径网络简化(pathfinder network scaling)选择了一系列最初连接而使显示出的网络图过于简单<sup>[33]</sup>。Pathfinder 网络最初应用于作者共引分析,而后扩展到一般的共引分析<sup>[34, 35]</sup>。路径网络简化是依据一个三角不等式检验以决定是保留还是删除某个连接,其标准是一个单连接路径的长度不能超过其他多个连接的路径的长度。Pathfinder 方法优于多维尺度方法之处在于它可以在共引网络中保持同步的发展模式<sup>[7]</sup>。改进的路径网络简化是为知识领域动态演进可视化研究而最近开发的<sup>[36]</sup>。

在许多研究中都探讨了文献网络变化可视分析的技术手段,如 Small 等分析了 AIDS 的研究进展<sup>[37]</sup>。Sandia 国家实验室开发了 VxInsight 系统,将科学文献表示为三维景图<sup>[38]</sup>,用户通过控制时间能够看到科学随时间的发展趋势。借鉴这一方法,我们设计了一种三维共引图景进化的时间模式<sup>[39]</sup>。

CiteSpace 是用来分析和可视共引网络的 Java 应用程序<sup>[36]</sup>。主要是用来帮助分析知识领域中的新趋势。它使用户可以将某个领域顺时进行“抓拍”,然后将这些抓拍的图片连接起来。我们已经使用初级版发掘出物理学中超理论革命中的转折点,但当时还有许多问题没能解决。为了区分起见,我们将初级版命名为 CiteSpace I,而将经过极大改进的新版本命名为 CiteSpace II,表 2 列出主要改进之处。新版最突出的特点是关键点的计算测量与可视属性的合并,这样做可以极大地减少用户在找寻知识结构中关键点时的负担。

表 2 CiteSpace I 与 CiteSpace II 的比较

	CiteSpace I	CiteSpace II
时区分割	√	√
共引网络/知识基础	√	√
基于时区分割的阈值	√	√
Pathfinder 精简	√	√
进化网络修补	√	√
时间编码网络可视化	√	√
n-gram 专业术语提取	—	√
突变检测	—	√
快速增长的主题词/研究前沿	—	√
中间中心性	—	√
引文年环可视化	—	√
异质网络	—	√
Pub Med MeSH 题头网络	—	√
时区视图	—	√
并行 Pathfinder	—	√

CiteSpace II 的新特点与三个中心概念相关:突变探测,中间中心性和异质网络。CiteSpace II 把研究领域概念化成研究前沿和知识基础间的映射函数,在这个映射函数概念框架下创建的这三个概念对解决以下三个问题非常重要:①识别研究前沿的本质,②标注研究领域,③及时识别新趋势和突变。

### 3 CiteSpace II

下面我们用 3 个实例来说明 CiteSpace II 的应用,并简要介绍如何应用这种新方法来分析激发事件是如何影响一个研究领域的演变。CiteSpace II 的概念模型见表 1。

#### 3.1 标注共引聚类

标注聚类涉及共引聚类的划界和解释。标准的方法是依据从引用共被引文献聚类的文章中提取出的词集。由于这些词是从形成初始聚类的文章中提取出来的,因而可以假定词集可以用来表示共引聚类的本质特征。词集的方法也有缺陷。首先,词集可能无法转化成一个焦点信息,用户和分析者可能需要花费大量的精力用于综合不同的词集。其次,基于集合词集的聚类标注可能趋于太泛而缺乏实际意义。实际上,许多研究者不仅仅关注最常用的专业术语,还对那些能导致重大变化的术语感兴趣。这些与新趋势相关的专业术语往往隐含在那些更宽泛的更持久的主题中。例如,对与生化武器威胁相关的卫生保健(health care)研究的兴起可能很容易就被掩盖在诸如 biological weapons (生化武器)这样更主导的专业术语中。

理想的标注方法应该能够从不引人注目的较为持久的主题中识别出明显的新趋势和快速变化。Small 在解释交叉学科共引连接时处理过相似的情形<sup>[27]</sup>。他提出有用的解释应该是能够区分有共引关系的两篇文章间异同之处。

Kleinberg 的突变检测算法可以用于检测一个学科内研究兴趣的突然增长<sup>[1]</sup>。尽管这个算法原本是用来检测单个词的突然出现,但也适于时间序列的多词专业术语和文章的引文分析。在 CiteSpace II 中,研究前沿是基于从题目、摘要、系索词(descriptors, 指标引文献主题的单元词或词组)和文献记录的标识符中提取出的突变专业术语(burst terms)而确定的,这些术语随后被用做专业术语和文章异质网络中的聚类标注。

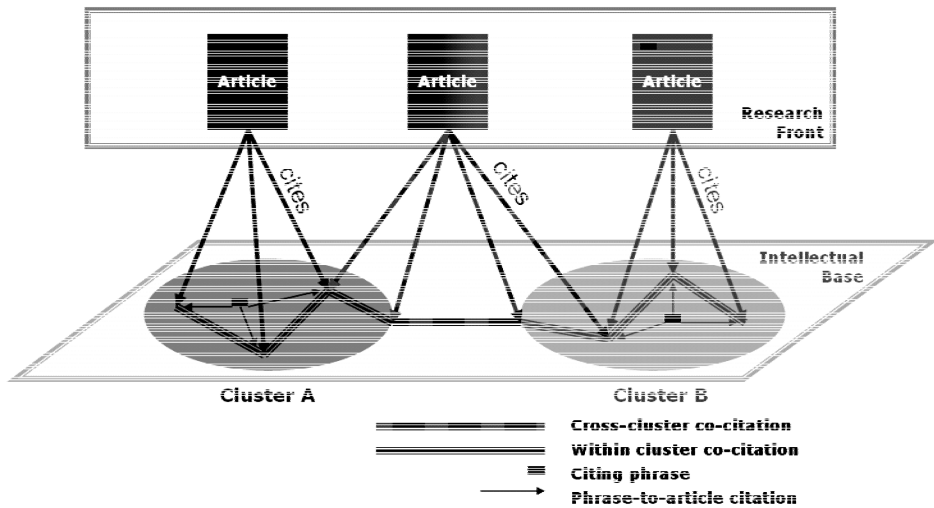


图1 CiteSpace II 的概念模型。用时间切片抓拍 (Time-sliced snapshot) 来显示研究领域的演变

### 3.2 提高可视化共引网络的时效 (timeliness)

运用引文专业术语来表征研究前沿也许会促进我们对其知识基础的理解。传统的共引分析集中在共被引文章的同类型网络研究。新近发表的文章由于没有足够的引文而不能充分体现在这种网络中,另外在共引网络中也没有涉及到词集。而突变检测算法,不管文章被引用多少次,都能够从中识别出突然涌现出的专业术语。因此,一个新的研究前沿,即使还没有吸引足够的引文,也能在大图中能被显示出来。

除了标准的有向图外,CiteSpace II 可以生成强调研究前沿和其知识基础间的顺时模式时区视图。时区视图是由一系列表示时区的条形区域组成,时区按时间顺序从左向右排列,因而研究前沿指向知识基础。这种显示算法是将弹簧模型算法 (spring embedder algorithm) 进行了改进,以使一个条目 (an item) 的横向变化被严格限制在所属时区内,但它的纵向变化完全取决于它同其它时区中条目的联系,这样更容易识别研究领域。时区视图是模仿整体显示时间线索可视 (overall layout of time-line visualization) 方法而设计的<sup>[11]</sup>,不同的是,为了突出研究前沿和其知识基础间之间的关系,CiteSpace 能够同时显示被引文献和引用的专业术语。

### 3.3 计算确定关键点

我们认为通过引用和共引关系可以追踪研究前沿的动态转变。假设  $\Psi_\alpha$  和  $\Psi_\beta$  分别是  $t$  和  $t + \Delta t$  时刻产生于  $\Omega_\alpha = \Phi(\Psi_\alpha)$  和  $\Omega_\beta = \Phi(\Psi_\beta)$  知识基础上

的,以文章  $\alpha$  和文章  $\beta$  为标识的主要研究前沿。突出出分别以文章  $\alpha$  和文章  $\beta$  为中心的两个共被引文章聚类。连接这两个聚类路径上的文章  $\{p^{(i)}\}$  描绘了从  $\Psi_\alpha$  向  $\Psi_\beta$  转变的特征。我们将这样的  $\{p^{(i)}\}$  称作关键点、转折点或支点。

在 CiteSpace I 中<sup>[36]</sup>,用户只能通过视觉观察找到网络中连接不同聚类的点,进而确定关键点。这个方法的优点是不需要其他的计算,但缺点是无法确保找到所有的关键点,除非费劲地观察网络中的所有节点。

CiteSpace II 使用户更容易地辨认关键点。明显的视图效果使用户很容易地看到那些具有较高中间中心性的节点<sup>[2]</sup>。由于关键点是通过计算辨认并呈现出来的,因而在可视化网络中用紫色的圈来突出显示。图论辨认关键点可以使研究者对整个网络的操作减少到仅对那些少数有解释意义的关键点进行研究。

### 3.4 操作步骤

CiteSpace II 的操作步骤如下,本文通过具体实例来说明其新功能。

(1) 运用尽可能广泛的专业术语来确定一个知识领域。这是为了确保接下来的分析能涵盖一个知识领域的全部内容。(例如:大规模生物集群灭绝)

(2) 收集数据。目前,数据主要来源于 Web of Science。CiteSpace II 也能够处理直接从 PubMed (一个化学文献的基本数据库) 下载的文献记录。用步骤 1 确定的主题词 (Topical terms) 检索出 Web of Science 中的所有相关文献记录,包括题目、摘要和

被引文献。每个文献记录代表一篇引文( Citing article ) ,而在每条记录中的参考文献被称作被引文献( Cited article )。

(3) 提取研究前沿术语。首先从数据库的引文题目、摘要、系索词( Descriptors ,指标引文献主题的单元词或词组 )和标识符中检索  $N$  元文法(  $N$ -grams )或专业术语。目前最多使用 4 个单词或是四个词组。例如“ terrorism ”和“ PTSD ”都是有效的专业术语。出现频次增长率快速增加的专业术语将被确定为研究前沿术语。

(4) 时区分割( Time slicing )。在 CiteSpace II 中 ,用户要明确整个时间跨度和单个时间分区的长度。

(5) 阈值选择。CiteSpace II 允许用户在引文数量、共引频次和共引系数三个层次上 ,前、中、后三个时间分区中分别设定阈值 ,其余的由线性内插值来决定。本项研究构建了研究前沿术语和知识基础文章的二分网络 ,这种网络包括三种类型的连接 :①研究前沿术语的共现 ②知识基础文章的共被引 和③研究前沿术语引用知识基础文章。

(6) 精简和合并。在 CiteSpace II 中 ,Pathfinder network scaling 方法是对网络精简的默认选项<sup>[33,36]</sup> ,用户可以选择是否精简图形。Pathfinder network scaling 是一个相当费时的算法。CiteSpace II 采用并行算法来同时处理多重网络以缩短整个等待时间。不同时间分区的网络  $G_i=(V_i, E_i)$  和  $G_j=(V_j, E_j)$  可能会有相互重叠的点和边。例如,  $V_i \cap V_j \neq \emptyset$  and  $E_i \cap E_j \neq \emptyset$ 。CiteSpace II 在网络的重叠区 ,将所有点组成一个整体单元 ,并选择那些没有违背三角不等式的连接 ,从而合并单个网络。用户可以选择是否将这个合并的整体网进行精简。

(7) 显示。CiteSpace II 生成显示标准视图和时区视图。

(8) 可视检测。CiteSpace II 能使用户以多条途径同这个可视化系统进行交互。用户可以根据运算法则设置参数 ,以控制可视化属性和标识的显示。图 2 显示了一篇文章节点的可视属性。

(9) 验证关键点。突出的关键点的作用可以通过咨询该领域的专家( 例如 ,关键点文章的作者 ) ,或是通过查阅文献( 比如包含关键点文章引文的段落 ) 来得以验证。能够自动概括关键点意义的工具开发是一个特别有趣的研究方向。数字图书馆 ,自动文本总结 ,机器学习等都是很有前途的研究领域。

### 3.5 实证研究

我们用两个研究领域的实例来说明 CiteSpace II 新特征 :大规模生物集群灭绝研究( 1981 ~ 2003 年 ) 和恐怖主义研究( 1990 ~ 2003 年 )。这两个领域都由于引发事件而导致了重大的变化。

#### ● 大规模生物集群灭绝( 1981 ~ 2003 年 )

地球上曾经发生过五次重要的大规模生物集群灭绝事件。最著名的是发生在 6500 万年前的“ K-T 灭绝事件”( 是指介于白垩纪( cretaceousperiod ,简称为 K)与第三纪( tertiaryperiod ,简称为 T)之间 ) ,它直接导致了恐龙的灭绝。大规模灭绝的原因是灾变论学者和渐进论学者一直争论的话题。灾变论强调灾难事件引发生物种群灭绝 ,而渐变论认为通常种群灭绝是一个漫长的演变过程 ,单一事件并不能决定其过程。早在 80 年代初提出的用以解释 K-T 灭绝的撞击说( impact theory )<sup>[40]</sup> 是最为著名的灾变理论 ,它认为巨大的陨石撞击地球导致了物种的灭绝。1990 年已证实了创击坑的存在。本实例研究

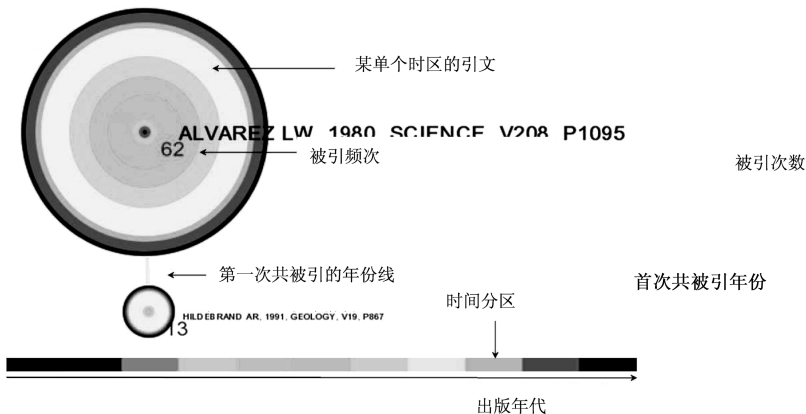


图 2 引文年环代表这篇文章的引文历史。引文年轮的颜色代表相应的引文时间。一个年轮的厚度与某个时间分区内引文数量成比例。节点中心旁的数字代表整个时间跨度内的被引次数



的目的是看看运用撞击说相关的专业文献能够在多大程度上探测到新趋势和突变。

有许多关键的问题需要通过实例研究来回答。什么是生物种群灭绝研究的最新前沿?在过去20年中撞击说的研究前沿是如何演进的?最热门的研究前沿术语是什么?哪些文章同这些专业术语有关?

在 Web of Science 引文索引数据库中以“mass extinction”为主题词检索到的从1981年到2003年间发表的有关生物种群灭绝的文章,并将其输入到 CiteSpace II 系统中。主题检索的范围包括每条文献记录的四个主题领域:题目、摘要、系索词和标识符,并仅限于英语文章。由此得到由771条记录构成的数据库,从这些记录的四个主题领域中检测到333个前沿术语。将1981~2003年这23年跨度分为12个时间分区(每2年一个分区),阈值分别设置为(2, 1, 10)、(3, 1, 0)和(3, 2, 10)。

#### ● 恐怖主义(1990~2003年)

1995年的俄克拉荷马城爆炸和2001年的9.11恐怖主义袭击是最具杀伤力的恐怖主义事件。与生物种群灭绝事件不同的是,其每个事件都可能改变研究的进程。科学共同体如何应对这些事件和衍生的结果?这个研究领域的新兴研究前沿是什么?他们同较早的研究前沿是怎样联系的?

恐怖主义研究(1990~2003年)的数据库是由在 Web of Science 中,以“terrorism”为主题词检索到的关于恐怖主义的1776条文章记录组成,并在其主题领域检测到1108条研究前沿术语。1990~2003年整个时间段被分成7个两年段。最后合并这七个网络成一个全景网络,用以描绘整个过去14年间的主导研究问题的变化。

## 4 结果

这个部分由两部分组成:大规模灭绝(1981~2004年)和恐怖主义(1990~2003年)。需要说明的是: $D$ :一组文章; $T$ :一组专业术语; $G(T_1, T_2, \dots, T_N)$ :节点类型 $\{T^{(k)}\}$ 的多重网络,如 $G(D)$ 和 $G(T, D)$ ;  $\Pi\{G_i\}$ :将“抓拍网络” $\{G_i\}$ 合并而成的网络, $\Pi$ 是合并操作; $PF(G)$ - $G$ 的 Pathfinder 网络。

### 4.1 大规模生物种群灭绝(1981~2004)

我们运用 CiteSpace II 生成了生物种群灭绝数据库的四个视图。①共被引文章的合并网络  $\Pi\{G_i$ ,

$(D)\}$ ,②经 Pathfinder 裁剪的个体共引网的合并网络  $\Pi\{PF(G_i(D))\}$ ,③经 Pathfinder 剪枝的混合网络  $PF(G(T, D))$ ,④经 Pathfinder 裁剪的混合网络  $PF(G(T, D))$ 的时区视图。

表3显示了515篇文章的共引网络  $\Pi\{G_i(D)\}$  的最初组配(Configuration)。值得注意的是在前两个时间分区里没有选择出来文章。CiteSpace II 用了不到10秒的时间完成这个过程。

表3 515篇文章共引网络结构组配

每2年分区	c	cc	ccv	文章数量	节点数量	连线数量
1981~1982	2	1	0.15	64	0	0
1983~1984	2	1	0.19	27	0	0
1985~1986	2	1	0.23	163	7	14
1987~1988	2	2	0.28	205	8	16
1989~1990	2	2	0.32	331	7	10
1991~1992	2	2	0.36	1 687	123	353
1993~1994	3	3	0.40	2 004	31	59
1995~1996	3	3	0.38	2 908	47	30
1997~1998	3	3	0.36	4 503	142	522
1999~2000	3	3	0.34	5 994	156	436
2001~2002	3	3	0.32	7 431	300	1 181
2003~2004	4	4	0.30	7 029	101	243
合计				32 409	92(515)	2 864

视图中最突出的文章是 Alvarez-1980,其内容正是前面提到的碰撞说(图3)。中心区域被五个相互联系的稠密聚类围绕着。在 CiteSpace II 中,用户可以选择某个项目聚类,并在 PubMed 中找到相匹配的记录。医学主题词(Medical Subject Headings, MeSH)中的词都是医学文献中最常用的,因而可以用来标注相应的聚类。图3中每个长方形角上的数字是指被选择区域中文献的数目。例如,专业术语“地质沉积”(geologic Sediment)在所选择区域中的100篇文章中的8篇中出现。由于 PubMed 根据标题域,主要涵盖的是医学文献,所以可视化网络中的文章不一定都能在 PubMed 中找到。

$\Pi\{G_i(D)\}$ 网络是将多个时间分区网络合并而成的。共引网络不能突出显示出研究前沿,但它们显示了研究前沿的发展轨迹。因此,这样的网络不能对诸如“为什么出现了某个特别的共被引文章聚

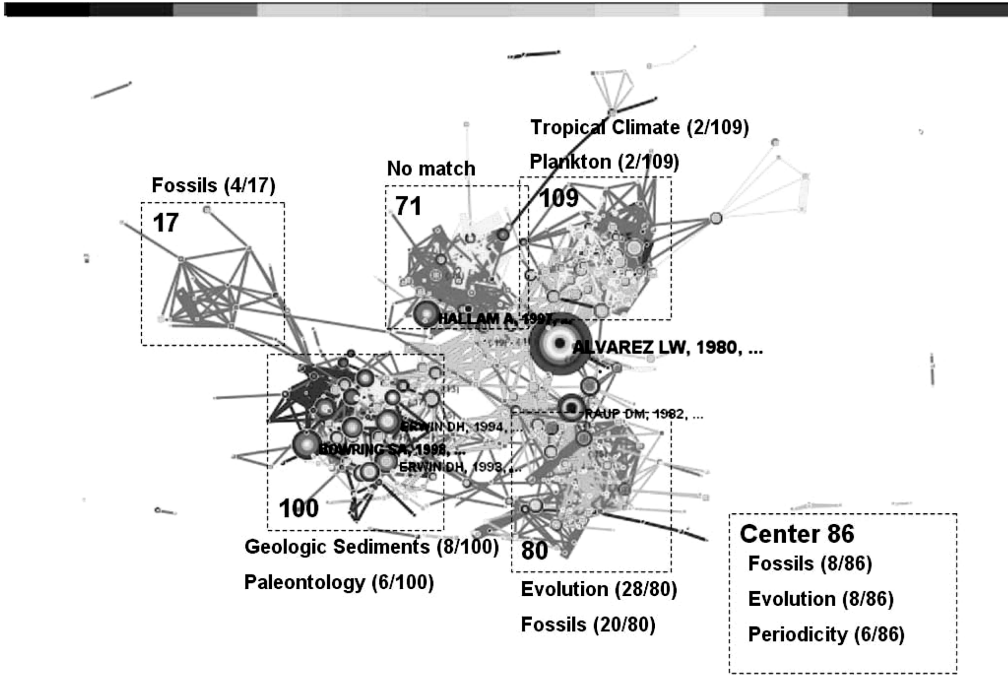


图 3 关于生物种群灭绝的 515 篇文章共引网络时区视图(12 个 2 年分区)。5 个聚类，中心区域标注有 MeSH 指定的专业术语，MeSH 专业术语是根据需要从 PubMed 中检索到的。每个长方形角上的数字代表这个聚类中文章的数量。最新的聚类位于左下角，由 100 篇文章组成的

类？涉及到了哪些研究前沿？”这类问题做出直接的回答。

第二组搭配生成了由 541 篇文章组成的网络  $\Pi \{PF(G, D)\}$ ，它合并了个体共被引文章 Pathfinder 网络。前时区阈值设置较低，这个过程用了 24.6 秒完成，结果见表 4。

表 4 第二组搭配，包括路径搜索剪枝和低阈值

每 2 年分区	c	cc	ccv	文章数量	单节点数量	连线数量
1981 ~ 1982	1	1	0.35	64	62	1 555
1983 ~ 1984	1	1	0.35	27	26	325
1985 ~ 1986	1	1	0.35	163	156	2 964
1987 ~ 1988	2	2	0.35	205	8	16
1989 ~ 1990	2	2	0.35	331	7	10
1991 ~ 1992	2	2	0.35	1 687	123	353
1993 ~ 1994	3	3	0.35	2 004	31	60
1995 ~ 1996	3	3	0.35	2 907	47	30
1997 ~ 1998	3	3	0.35	4 497	142	523
1999 ~ 2000	4	4	0.35	5 990	59	68
2001 ~ 2002	5	5	0.35	7 427	145	364
2003 ~ 2004	5	5	0.35	7 091	153	93
合计				32 393	850(541)	6 361

低阈值设置使得 1981 ~ 1982 年和 1983 ~ 1984 年时区的被选择文章增多了，尽管这些新被选的文章在数据库中仅被引用了一次，但使得共被引更为广泛。在 1981 ~ 1982 年间，62 篇文章之间有 1 555 个共引连接。同样的，在 1985 ~ 1986 年间，在 156 篇文章间有 2 964 个共引连接。从图 4 中可以看出，这些文章除了与 Alvarez-1980 共引外，它们也有 1 或 2 次共引连接。位于视图右边的聚类明显是最近形成的，因为这个聚类中的年轮由红色主导，红色是表示被引年代是 2003 ~ 2004 年间。这个聚类中高被引的文章包括 Bowring-1998, Renne-1995, Hallam-1997, 和 Wignall-1992。Raup-1982 是在早些形成的以 Alvarez-1980 为中心聚类中第二高被引文章。

第三组搭配是显示研究前沿术语和知识基础文章的混合网络。图 5 上方的聚类是最近形成的。这个聚类包括诸如 Late Frasnian(晚泥盆世弗拉斯期)，Atmospheric CO<sub>2</sub>(大气 CO<sub>2</sub>)和 carbon cycle(碳循环)研究前沿术语。中心领域是关于 Cretaceous-Tertiary Boundary(白垩纪之末与第三纪之初的交界)和 Late Cretaceous(后白垩纪)为术语的研究前沿，它们是撞击说的核心概念。

除了以上 3 个视图，图 6 显示了一个强调时间联系的时区视图。时区视图揭示了随时间形成的 3

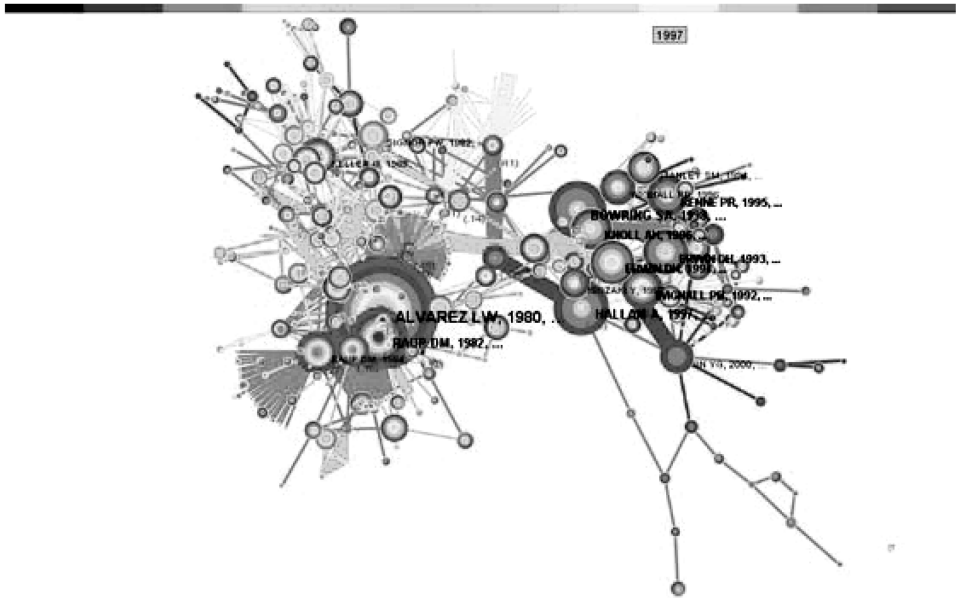


图4 基于12个 Pathfinder 精简‘抓拍’网络的541篇文章共被引网络。前三个分区显示在视图中。最突出的聚类也是最新形成的(图右侧的聚类)

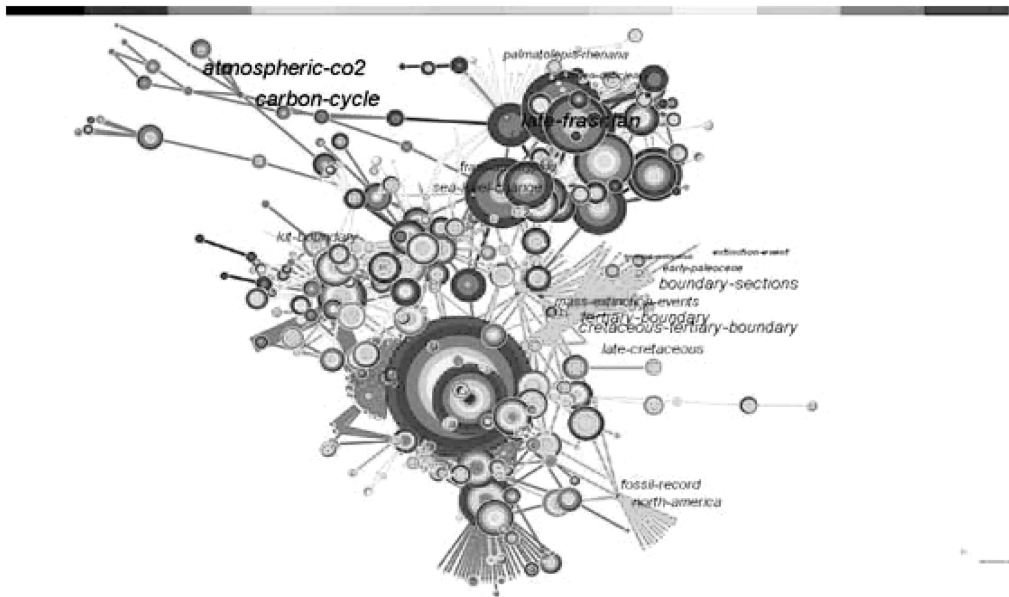


图5 被引文献和引文专业术语的569点混合网络(用时28秒)

个显著的聚类。每个聚类都对应于生物种群灭绝研究的一个学科方向。在视图下方充满绿色线条的聚类就是 K-T 灭绝争论分支,这个分支方向开始于 Alvarez-1980,在随后的10年中一直保持着强势,直到1991~1992年开始削减,直至1995~1996年后消失。从引文最外环的颜色可以看出,这个聚类中的许多文章在2003~2004年间仍被引用。最近的研究前沿术语有 North America(北美), fossil record(化石记录)和 Late Cretaceous(后白垩纪)。

视图中间是黄色连接的聚类,从20世纪80年代开始,结束于1997年。这个方向中有文章 Johnson-1985 和 Copper-1986,其规模比 K-T 灭绝方向小。由于这个方向中有晚泥盆世弗拉期(Late Frasnian)前沿术语,我们就将其称作晚泥盆世弗拉期方向。在地质学中,是后泥盆纪(3.86亿~3.75亿年前)的一个时期,发生在晚泥盆世弗拉期和法门期(Famennian)之交的后泥盆纪生物灭绝事件是古生代以来五大生物大规模灭绝事件之一。对这次大

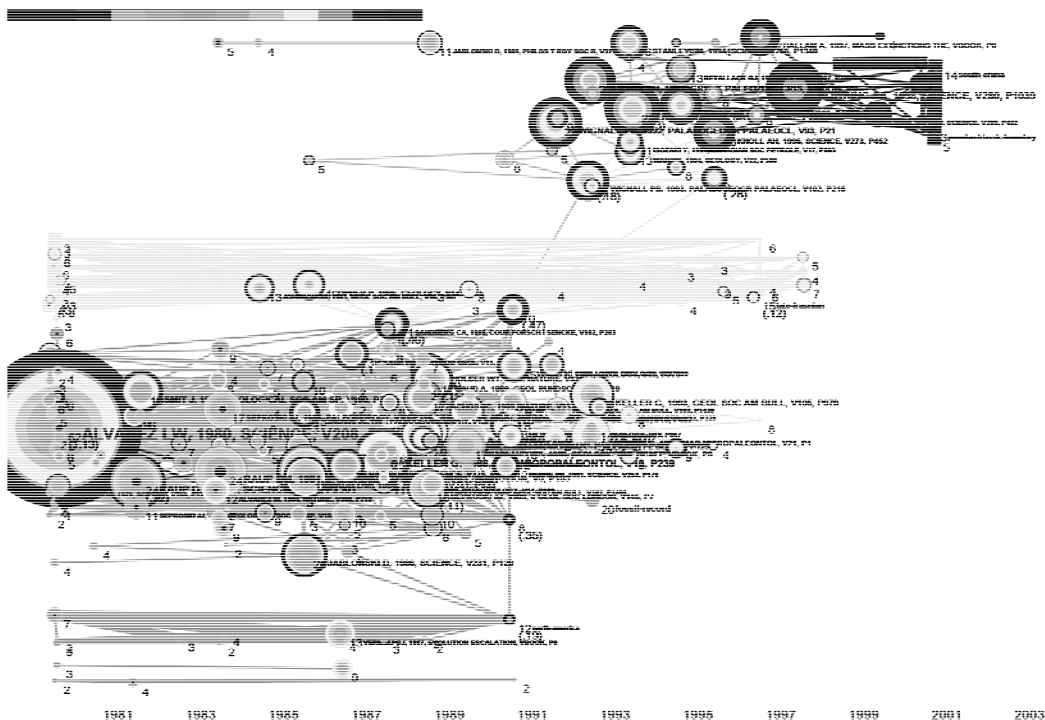
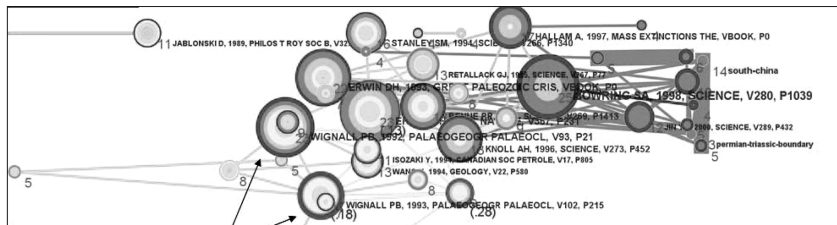


图 6 大规模灭绝研究的时区视图 表现出 3 条主要研究方向 :K-T 灭绝 , 晚泥盆世弗拉斯期(中间聚类)和二叠纪灭绝(上聚类)



问题 1: 这个显示了当前关于大规模灭绝的一个研究方向。你的两篇文章(1992, 1993)在这条线索(相关专业术语“Permian-Triassic Boundary/SouthChina”)的起始处。这个线索与下面的两个线索有什么联系?

问题 2: “Late Frasnian”的重要文章是 Copper 1986 和 Johnson 1985, 这是一条什么样的线索? 为什么看上去像要中断?

问题 3: Alvarez 1980: 这条线索是不是在 1990 年代中期结束的(1995 年后没有发现被引论文)

图 7 送给 Dr. Wignall 的附有三个问题的时区视图,他是视图中两个关键点论文的作者

规模灭绝的解释是全球变冷、海平面降低和隕星撞击。

视图上部是橙色连接聚类,也代表一个演进方向,它开始于 Wignall<sup>[41]</sup>,由 Bowring-1998 推动,最近 Becker-2001<sup>[42]</sup>又进一步推进。Permian-Triassic-

boundary(二叠纪-三叠纪-分界)和 South China(中国南部)是在 2001 年检测到的两个研究前沿术语。Becker-2001<sup>[42]</sup>是一篇以“二叠纪-三叠纪-分界线的影 响事件”为标题的文章,在网上对“Becker, Permian, and Triassic”作简单搜索,表明 Becker-2001

为二叠纪大规模灭绝的主要影响提供了证据。Becker 等最近在《科学》杂志上发表了“Bedout(是一个撞击坑的名字)澳大利亚西北部海边的一个可能的撞击坑”一文,这篇文章对二叠纪大规模灭绝研究的重要价值可以与 Chicxulub 撞击坑对 K-T 撞击理论价值相当,Chicxulub 撞击坑的发现极大地增加了 K-T 撞击理论的可靠性<sup>[43]</sup>。许多科学家受到这些成功解谜的鼓舞,开始应用这个撞击说来解决其他大规模灭绝。寻找撞击坑便是下一步工作,有关辨认二叠纪-三叠纪边界的撞击坑的研究已经吸引了许多研究者。当前的研究前沿就是在这样的背景下演变而来的。

我们将附有 3 个问题的时区视图(图 7)送给英国 Leeds 大学的 Dr. Paul Wignall,他是在二叠纪灭绝方向<sup>[41, 44]</sup>中关键点文章的作者,我们让他来检验一下上述的结果。他证实了大量的 K-T 争论由于 Chicxulub 撞击坑的发现而从 90 年代中期开始消退:“许多人仍旧在研究这个灭绝事件,但是更多的人将研究兴趣已经转移到二叠纪末的大规模灭绝,这就是为什么大家一直对我的研究感兴趣。”Late Frasnian 是于 1997 年在这个研究方向上首次发现的专业术语。至于“为什么中间的这条线索看来快要结束”这个问题,Wignall 的解释是,这条线索是同 Copper 的有关泥盆纪后的大规模灭绝研究相关,研究线索不太活跃主要是由于缺乏足够的数据库。

总之,时区视图把该领域专家的认识表现在一幅长期发展趋势图上。K-T 灭绝研究方向的终点时间恰是二叠纪灭绝研究方向的起始点。是不是因为 K-T 灭绝争论的结果鼓舞了科学家? Becker 等最近的一系列研究表现出了同 80 年代的 K-T 争论极大的相似性,尤其在撞击理论和寻找撞击坑方面,这有助于揭示这个方向今后的演变<sup>[43]</sup>。

同早期的可视化图像相比,CiteSpace II 提高了图像的清晰度和可解读性。2002 年大规模生物种群灭绝的实例研究<sup>[45]</sup>,是通过共引网络演化的可视化手段,重构科学领域发展历史,使其不仅能按年代显示,还能回顾性显示。被引频次最高的文章表示成最高的共引柱,主流研究方向是通过基于主成分分析(PCA)的群组颜色表示的。三维模型(一个二维的共引网络作为基本图和一维引文词条)的解释主要参考 Walter Alvarez 的著作<sup>[46]</sup>。尽管 3 维模型确实生动的再现了引文的增长,但很难在这种模型中识别出新趋势。CiteSpace II 视图不仅能突出显示诸如 Alvarez 等的奠基性文章,更重要的是明确地标

出了从一个研究方向到另一个研究方向转变过程中的关键点,即转折点。基于网络拓扑性质识别关键点具有实际价值,例如,我们可以用这种办法很快找出某领域的几个最重要的专家,并就专业问题向他们请教。在下面的实例研究中,许多领域的专家就是通过这种方法识别出来的。

#### 4.2 恐怖主义(1990 ~ 2003 年)

合并 7 个“抓拍”得到由 335 个节点组成的专业术语和文章混合网络。阈值设置见表 5。1.6GHz 的奔腾处理器和 1GB 内存的 IBM ThinkPad 环境下,CiteSpace II 用 8 秒钟生成了一个混合网络。

CiteSpace II 应用并行 Pathfinder 程序以缩短整个运行时间。在这个实例中,CiteSpace II 需要 20 秒来完成从每个时间分区里提取一个 Pathfinder 网络。如果其能代表足够的信息,这被认为是合理的。

表 5 恐怖主义数据库 1990 ~ 2003 年可视化的 CiteSpace 组配

每 2 年分区	c	cc	ccv	文章数量	单节点数量	连线数量
1990 ~ 1991	2	1	0.10	1 455	15	27
1992 ~ 1993	2	1	0.11	1 372	32	137
1994 ~ 1995	2	1	0.12	1 527	12	19
1996 ~ 1997	2	1	0.14	2 735	109	1 504
1998 ~ 1999	3	2	0.15	3 037	31	75
2000 ~ 2001	3	2	0.18	5 518	47	211
2002 ~ 2003	3	2	0.22	18 366	182	649
合计				34 010	428(335)	2 622

恐怖主义研究视图有三个明显的聚类(图 8):  
 ①聚类 A( $C_A$ ),恐怖主义爆炸中的身体外伤(左上,绿色),前沿术语有 body injuries(身体外伤)和 terrorist bombing(恐怖爆炸);  
 ②聚类 B( $C_B$ ),与生化武器威胁相关的卫生保健(右上,黄-橘色),前沿术语有 health care(卫生保健)和 chemical weapons(化学武器);  
 ③聚类 C( $C_C$ ),9.11 恐怖袭击事件对心理和精神的影响(中下,橘色),前沿术语有 September 11(9.11),United States(美国)和 post-traumatic stress disorder(创伤后应激障碍)。

$C_A$  表示恐怖爆炸中的躯体受损学科专业,尤指外科。Cooper-1983<sup>[47]</sup>, Frykberg-1988<sup>[48]</sup>, 和 Katz-1989<sup>[49]</sup>都是构成知识基础的文章。 $C_B$  表示卫生保健学科专业,由左半部亮黄色和右半边桔色两部分

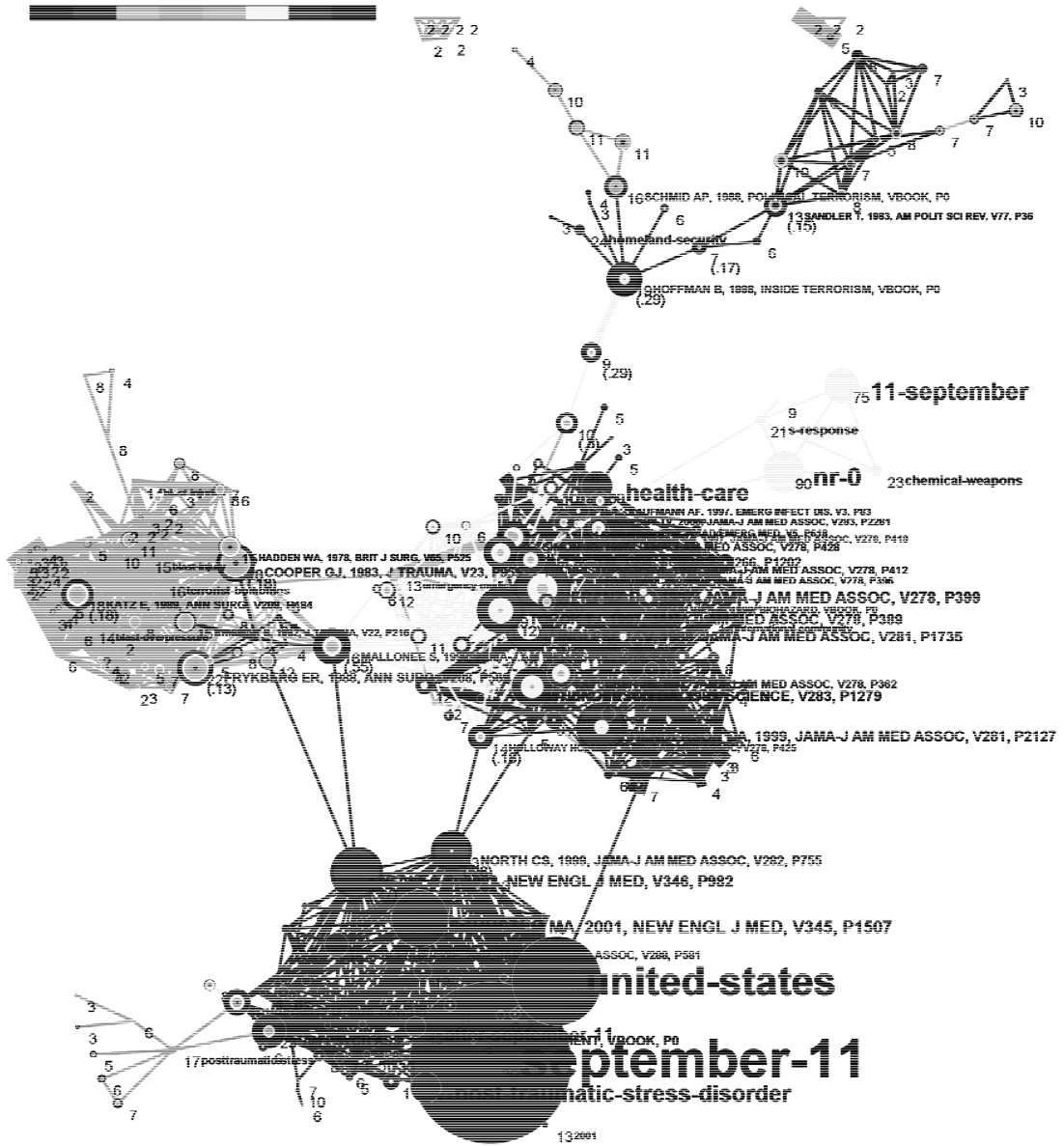


图 8 在由 335 个节点组成的被引文章(知识基础)和引文专业术语(知识前沿)混合网络中出现 3 个明显的聚类：  
 C<sub>A</sub> 恐怖主义爆炸中的身体外伤(左上 绿色)，C<sub>B</sub> 与生化武器威胁相关的卫生保健(右上 黄-桔色)，  
 C<sub>C</sub> 911 恐怖袭击事件对心理和精神的影响(中下 桔色)

构成左半部的主导前沿术语是 emergency medical (急救), 右半部的是 Health care(卫生保健), 两个部分都与作为应对生化武器引起的潜在大规模灾难的卫生保健和紧急系统的脆弱性相关。

C<sub>A</sub> 和 C<sub>B</sub> 通过一个关键点 Mallonee-1996<sup>[50]</sup> (中心性 = 55%) 相联, 这篇文章对由 1995 年俄克拉荷马城爆炸引起的灾祸作了全面的研究。因为这是一篇关于恐怖爆炸的人体外伤治疗和灾难处理的文章, 故与 C<sub>A</sub> 相联。另一方面, 这篇文章中的伤病和灾难是起因于俄克拉荷马城爆炸, 故也同 C<sub>B</sub> 相联。

C<sub>C</sub> 表示 9.11 恐怖袭击导致的心理和精神紊乱学科专业。PTSD 研究线索包括了 2001 年 PTSD 调查 Schuster-2001<sup>[51]</sup>, 2002 年 PTSD 调查 Galea-2002<sup>[52]</sup> 和 1999 年俄克拉荷马城爆炸 PTSD 调查 North-1999<sup>[53]</sup> 的研究, 其中 Galea-2002 和 North-1999 被确认为关键点。C<sub>C</sub> 仅通过 Galea-2002 与 C<sub>A</sub> 相连, 也仅通过 North-1999 与 C<sub>B</sub> 相连。C<sub>B</sub> 和 C<sub>C</sub> 是通过 Holloway-1997<sup>[54]</sup> 和 North-1999 之间的关键而联系起来的。C<sub>B</sub> 中的 Holloway-1997 是一篇关于应对可能的大规模破坏性恐怖袭击的卫生保健准备的文章,

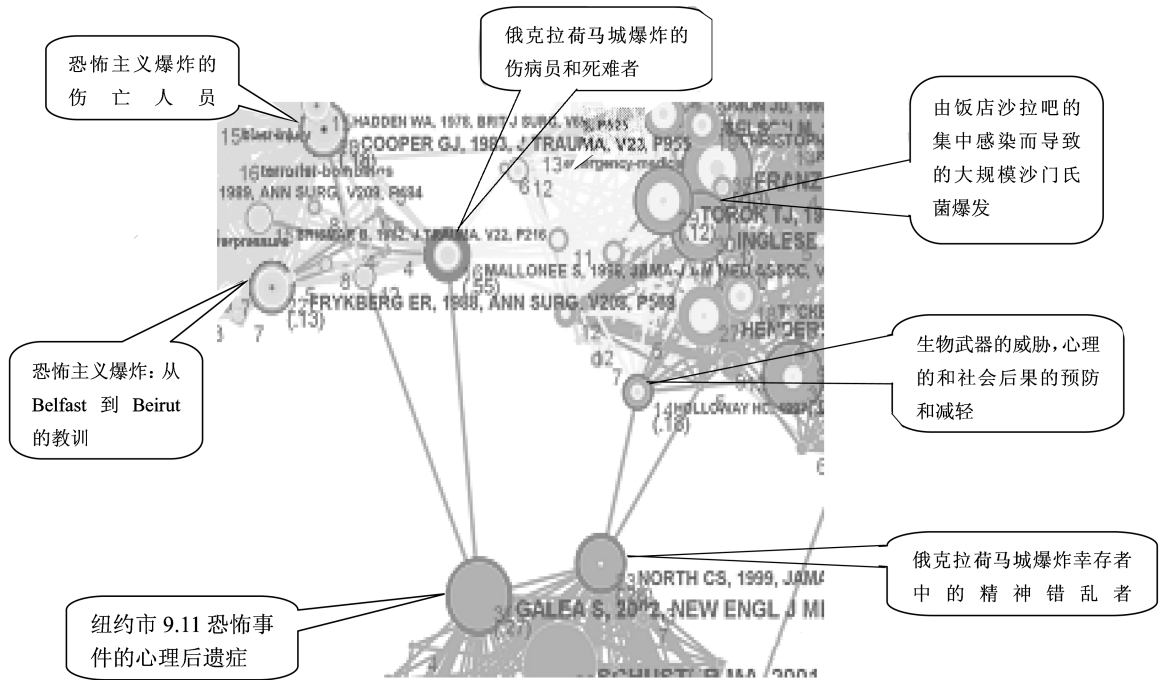


图 9 恐怖主义研究领域的三个主要前沿

从心理学角度看, Holloway-1997 对应于恐怖主义研究的 PTSD 一面。C<sub>B</sub> 和 C<sub>C</sub> 间的另外一个连接是通过一个词: United States(美国)。

如果我们注意引文的内容, North-1999 在整个图中的重要性就变得更清晰了。人们一眼就能直观的找出值得细看的文章。CiteSpace II 允许用户多层次地观察细节。最高层次自动显示最突出的项目, CiteSpace II 用户可以通过选择不同阈值来调整信息显示程度。在第二个层次上, 用户可以通过单击一个网络中的节点提取出单线文献的细节, 还可以通过双击节点获得更详细的信息。

图 9 是图 8 的一个局部图。这个视图显示了几个位于三个聚类(C<sub>A</sub>, C<sub>B</sub>, C<sub>C</sub>)间和关键点和关键点文章题目。其中的两个点同 1995 年俄克拉荷马城爆炸有关系: 关于身体外伤的 Mallonee-1996 和关于幸存者中 PTSD 的 North-1999。9.11 后的研究极大地参考了俄克拉荷马城爆炸的经历和发现。另一方面, 如同 North-1999 一样, 研究焦点从俄克拉荷马城爆炸的冲击波幸存者转向更广泛的人群。9.11 袭击的 PTSD 研究<sup>[52, 55]</sup>, 大众媒体更关注的是纽约没有直接受到 9.11 袭击的人群。CiteSpace II 使我们能够从学科研究的角度注意到 North-1999 和 Galea-2002 之间的重要关联性。这种关联的准确含义在

万方数据

随后的专家(即关键点文章的作者)调查中得以确定。调查结果如下。

CiteSpace II 生成了一个恐怖主义研究的时区视图。限于篇幅, 我们只概述最主要的研究结果。在视图中, 出现了三条主要的研究线索。俄克拉荷马城爆炸线索于 1997 年结束。最近, 在生物恐怖主义研究前沿中发现“health care”(卫生保健)和“after 9.11”(9.11 后)两个专业术语, 它们最早于 2003 年被检测到。PTSD 线索被拉长, 贯穿于整个时间跨区, 这表明尽管 PTSD 研究只是最近了在恐怖主义研究中突显出来, 但其已有很长的研究历史了。PTSD 研究范围宽泛的紧张性刺激(Stressor)如压力的起因, 比如自然灾害和恐怖主义, 9.11 袭击事件拓宽了传统 PTSD 的范围。

表 6 列出了出现在 1990 年到 2003 年间的研究前沿术语。同 9.11 恐怖袭击相关的术语有 2001 年涌现的 Chemical-weapon(化学武器)Anthraxvaccine(炭疽热), 2002 年涌现的 New-York(纽约)和 posttraumatic stress disorder(创伤后应激障碍), 2003 年涌现的 health care(卫生保健)和 after-September(9 月后)。同俄克拉荷马城爆炸相关的术语有, 1996 年涌现的炸伤(blast-injuries)和联邦大楼(federal-building)。注意 PTSD 在 1996 年被拼写为“posttraumatic”, 而 2002 年, 即 9.11 后, 被拼写成

表 6 1990 ~ 2003 年间涌现的研究前沿术语

术语	频次	术语	频次	术语	频次
2003		1999		1995	
Health-care	47	Biological-weapon( s )	16	Peace-process	10
After-September	32	Emergency-medical	11	Couple-violence	7
Care-workers	9	Emergency-management	6	Emergency-room	7
Attacks	8	1998		Common-couple	6
security	7	Organophosphate-poisoning	9	Sarin-vapor	6
2002		1997		Major-trama	5
New-York	111	Terrorist-bombings	15	1994	
Post-traumatic-stesee-disorder	70	Blast-overpressure	14	National-security	8
North-Korea	30	Warfare-agents	10	1993	
Homeland-secutity	24	Tokyo-subway	9	United-states	21
International-community	14	Mass-disaster	8	International-news	6
Bin-laden	13	Biological-terrorism	7	Criminal-justice	5
Health-protection	13	Gastrointestinal-tract	6	Nuclear-weapons	4
2001		Lung-injury	6	1992	
11-september	75	Trarmatic-events	6	Northern-Ireland	8
Chemical-weapons	23	Buenos-Aires	3	Gulf-war	6
America 's-response	20	1996		Political-violence	6
Anthrax-vaccine	11	Blast-injuries( injury )	29	United-Kingdom	6
2000		Mental-health	19	Blast-loading	5
Law-enforcement	14	Posttraumatic-stress	17	Blast-lung	2
Task-force	14	Federal-building	6	Iowa-state	2
New-terrorism	7				

“post-traumatic”。“Mental-health”(精神健康)是同 PTSD 密切相关的概念,这两个专业术语在 1996 年俄克拉荷马城爆炸后出现频繁。

## 5 检 验

关键点的本质可以通过两种方式评估,①将关键点文章的作者视为该领域的专家,向其发放调查问卷,②对关键点文章中被引和共被引的段落进行内容分析。限于篇幅,我们仅说明前一种方式。

我们通过电子邮件把 4 个问题发送给关键点文章作者。其目的在于证实这些关键点确实是知识转化、焦点转移、或知识领域图景改变过程中的关键点。通过问卷,我们向每位专家询问他们的关键点

文章的意义,以及和邻接关键点文章联系的本质。附有高分辨率视图链接的调查问卷发放给了 8 位专家,其中 3 位专家由于电子邮箱地址变更而没有收到邮件,其余 5 位专家给与了回复并回答了所有问题。限于篇幅,我们仅详细讨论其中一位专家的回答,对其他专家的回答给与概括。从专家的回答中,我们获得更详细的信息。

Galea-2002 和 North-1999 是在 PTSD 簇中的两篇关键点文章。Dr. Sandro Galea 是 Galea-2002 的第一作者,其对问卷的回答如下。

1. 你的这篇文章中所研究问题的意义是什么?

我想我们的文章是第一篇运用有效的方法对 911 恐怖袭击后导致的心理紊乱进行研究的文献。911 袭击后,人们对此有了浓厚的兴趣,我们的文章



为预测潜在的袭击后果提供了一些实证的理论基础。

## 2. 你的文章对后来的研究有什么重要影响？

911 恐怖主义袭击后,发表了大量关于袭击的心理学影响和恐怖主义内涵的文章。我认为我们文章首先使人们认识到恐怖主义事件对人类神经健康的重要影响,其次清晰地列举出在大规模灾难后人们普遍(不仅仅是直接受害者)存在心理紊乱的事实。

3. 你的文章经常与另外两篇关键点文章同时被引用(North, C. S. et al. (1999) Psychiatric disorders among survivors of the Oklahoma City bombing. *JAMA*, 282(8), 755 ~ 762. 和 Mallonee, S. et al. (1996) Physical injuries and fatalities resulting from the Oklahoma City bombing. *JAMA*, 276(5), 382 ~ 387.),你能解释一下这种联系的实质吗?

我想“真正有联系”的是这两篇文章中的第一篇。9.11 事件被认为是俄城爆炸之后的“第二次”(发生在美国本土的)大规模恐怖袭击,这样开始的时候许多有关 911 的文章自然大量引用俄城爆炸的文献。North 的论文是在俄城爆炸后的“经典”文献,这样人们就不可避免地要比较这两个(非常不同)论文。

4. 在你看来,你的文章是不是代表了研究焦点的转移或方向的改变?

我想我们的文章明确提出在恐怖主义后“城中的每个人”都面临 PTSD 的风险,这是一个不可争辩的事实,我们扩展了人们原本理解的“直接经受”灾难的范围。在这点上,这篇文章确实代表了研究方向的改变。尽管还有些争议,但随后的文章还是接受了这个新观点。

其他专家也提供了他们的评论,包括  $C_A$  中关键点文章 Frykberg-1988 的作者 Dr. Eric Frykberg 和关键点文章 Sandler-1983<sup>[56]</sup> 的作者 Dr. Todd Sandler。Dr. Carol North 接受了 30 分钟的电话访谈,提供了宝贵见解和解释。

Frykberg 说明了  $C_A$  中 Frykberg-1988 的意义:“作为一篇对上百篇有关爆炸事件的文献综述, Frykberg-1988 首次通过分析恐怖主义爆炸带来的伤病灾难的严重和模式,退引出管理的一般原理,即一旦确认了某种灾害模式,就有可能进行相应的准备和计划来改善结果”。Frykberg-1988 提出了医药管理的一种新的结果测量手段,即临界死亡率(critical mortality)文章的结论是,因为准确快速确认少数需

万方数据

要紧急治疗的伤病员,对救助紧急危险中的生命是非常重要的,因此伤病员的鉴别分类是灾难过后处理的关键因素。

North-1999 是在 PTSD 聚类中的两个突破点之一。North 在电话访谈中解释到,尽管 PTSD 聚类在 9.11 事件后快速显现出来,但实际上它已近有了很长时间的研 究历史。早期的 PTSD 研究同越南战争、海湾战争和自然灾害(如地震和飓风)相关。恐怖主义在 PTSD 研究中是一个相对较新的紧张刺激类型。恐怖主义的 PTSD 研究的突然高涨主要是由于 9.11 后出现了大量的 PTSD 研究成果,并同早期的俄城爆炸的 PTSD 研究一起,创建了恐怖主义研究中的一个重要方向,即 PTSD 研究。

North 也注意到  $C_B$  中的生物恐怖主义研究前沿是非常活跃的领域。既然生物恐怖主义被认为是恐怖主义的一种类型,那么从一般的恐怖主义研究获得的经验也可应用于生物恐怖主义。如时区视图中所示,  $C_B$  中的前沿术语 health-care(卫生保健)属于当前活跃的研究前沿。North 根据他们研究的人口聚类也提出了一种区分 PTSD 研究的方法。例如, North-1999 研究了俄城爆炸现场中的直接伤病受害者。Galea-2002 基于曼哈顿的全部人口来进行抽样研究,这要比世贸中心现场受伤人数要多。Schuster-2001 在关于恐怖主义引发 PTSD 爆发的研究中,将人口建立在更宽泛的美利坚合众国的基础上。Schlenger-2002 研究的人数增长到远远超过纽约人口。

## 6 讨论和今后的工作

CiteSpace II 系统适用的用户群广泛,特别是科学家、科技政策研究者和搞科研的学生都可以用它来进行科学学科的发展趋势和发展过程中的重要变化的探测和可视化研究。它可以通过从研究前沿到知识基础间的时间映射来探讨学科发展的动力学机制。计算技术和可视化技术的创造性的融合极大地促进了知识领域可视化研究。以上两个实例根据专家的验证,其研究结果是令人鼓舞的。

将研究前沿和知识基础完全整合在同一个可视网络中有三个实际优点。首先,使用涌现主题术语(surged topical terms)要比使用出现频次最高的主题词(title words)更适于探测学科发展的新兴趋势和突然变化,在可视化网络中,研究前沿术同知识基础文章之间的连接是明显的,这种设计显示出研究前沿

和知识基础间的紧凑结合。第二,研究前沿术语很自然地用来标注学科专业。第三,它克服了基于词集标注方法的一般缺点,集合的词集也许不能汇聚成集中的主题。挑选出现频次突然增加的专业术语是特别适合表征当前的研究前沿的。

分治法(divide-and-conquer)有效地发掘出学科演化的动力机制。通过计算辨认关键点大大减轻了用户的理解负担。在恐怖主义研究中,我们通过专家证实了这些关键点。更多评价研究是用来调查可转化中心性矩阵的效果。

Pathfinder算法从网络中提取出最显著的模式(Pattern),但却不能伸缩自如。CiteSpace II应用了Pathfinder算法的并行版本,尽管这仍需要6 000秒来处理14个网络并将它们合并成1 704个节点的合并网络,但已经充分利用了网络缩放组件。我们正在运用聚类计算技术开发平行处理方案,目前在单个PC机上优化构型限制在十几个200~300个节点的单网合并的可视网络,这个过程在十几秒内完成。

开发一个能使用户在各种水平上研究科学文献的无缝环境是一个长期的研究工作。突破点可以作为帮助用户追踪一个学科的发展轨迹的路标。另外一项工作是将引文检索和共引检索同全文文本处理技术整合起来,这样的整合使用户不仅能基于引文索引辨识重要的模式和联系,而且能准确获取原文中的第一手信息。这个研究思路在一定意义上实现了<sup>[57]</sup>在1945年的设想。

CiteSpace II将学科演化建立在研究前沿和知识基础间的时间变量双重性基础上,作为一个为方便学科演化研究而设计的系统,作为一个便于研究学科演化研究而设计的系统,还有许多可能拓展的方向。例如,科学文献(文章)和技术发明(专利)间的连接已经成为政策研究、科学技术指标和知识扩散等研究的中心议题<sup>[58,59]</sup>,然而,仍旧还有许多工作需要去做,用户需要那些能够把海量的数据转化成清晰和有指导意义的信息工具。

政府投资机构与科学家产出间的连接日益重要。辨识科研投资和各种奖励基金的发展趋势对科学家非常重要。搞清科研投资趋势和和现存智力基础间的联系可以极大地提高大学和研究实验室的竞争力。除了水平扩张的机会,一些应用领域特别需要趋势探测和题目追踪的工具,包括生物信息领域和国家安全相关领域,如智能分析。

一个长期研究工作就是要开发深层的知识发现和数据挖掘技术并同知识领域可视化相结合。

Swanson研究发现逻辑上相关的不交互文献<sup>[60,61]</sup>和假说模型<sup>[62]</sup>是有特别关系的。

由于不是所有文章都可以获得全文,我们无法获得足够多的引文文章。这就提出了一个实际问题,即分析者要获取什么程度的文献。理想程度是,从一个综合的数字图书馆获取所有的引文和被引文章的全文,实际上最封闭的数据源包括ACM数字图书馆和IEEE浏览器。更可行的是,用户可以从某个领域的一些核心期刊中获取全文。开放数据库arXive和CiteSeer都是很好的数据库。今后我们计划将拓展和完善这个工具以进行趋势分析。

## 7 结 论

CiteSpace II对知识领域新兴发展趋势和突然变化的探测和可视化分析的新功能已经产生了有前途的并令人鼓舞的结果。主要的发现如下:

(1)研究兴趣的骤然兴起是一个研究前沿兴起的信号。

(2)术语和文章的异质网络综合地表现了一个学科专业发展的动态机制。

(3)研究前沿术语是聚类的信息标签。

(4)引用环可视图生动形象并可解释。

(5)中间中心性矩阵辨识语义上鲜明的突破点。

本文也明确了今后研究中的挑战和机会。主要的挑战概括如下:

(1)通过实证研究对使用跳变专业术语表征聚类的内涵相比于用词集表征的优势和劣势做出评价。

(2)通过中间中心性识别突破点和通过转换中心性矩阵辨识的实证比较。

(3)除了开发统计方法,还要通过修改和整合自动文本总结和自然语言处理技术开发无缝环境使得可视浏览和科学的日常行动能更好地整合在一起。

(4)将研究前沿和知识基础的可视化与全文数字图书馆整合起来。

(5)应用和比较趋势探测和主题追踪技术。

(6)将研究扩展到其他数据类型,包括专利、基金白皮书和奖励基金等。

CiteSpace II作为一个工具,能够使分析者较为容易地对科学领域进行定量和定性的研究。从知识贡献上来说,这项研究表明了这个能增进我们对一个研究领域认识的实际工具的潜在优势。这项研究也提出了今后需要工作的方向与解决的问题。这个

研究融合从各个学科汲取的知识,将孤立的各种技术转化成整合环境。

我们论证了可视化整合方法的潜能,分析了研究前沿的进化。我们特别研究了有关探测暂态研究前沿中新兴趋势和突然变化的多种实例。结果表明,这是一个有前途的研究思路,各学科的使用者都会受益。

## 致谢

十分感谢以下各位专家的慷慨帮助,他们是英格兰 Leeds University 进行大规模秘诀研究的 Dr. Paul Wignall 和华盛顿大学医学系精神病学教授 Dr. Carol North;纽约医学院城市流行病学研究中心的 Dr. Sandro Galea;Florida 大学的外科的教授 Dr. Eric Frykberg;南加利福尼亚大学国际关系学院,进行恐怖主义研究的 Dr. Todd Sandler。同时感谢美国信息科学技术学会 2002 年引文研究基金资助和美国科学信息所。CiteSpace 可以在以下网址获取: <http://cluster.cis.drexel.edu/~cchen/citespace>。

## 参 考 文 献

- [ 1 ] Kleinberg J. Bursty and hierarchical structure in streams. Proceedings of Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining ( pp. 91-101 ), Edmonton, Alberta, Canada :ACM Press 2002.
- [ 2 ] Freeman L C. Centrality in social networks : Conceptual clarification. Social Networks ,1979 ,1 215-239.
- [ 3 ] Price D D. Networks of scientific papers. Science ,1965 ,149 510-515.
- [ 4 ] van Raan A. On growth , ageing , and fractal differentiation of science. Scientometrics ,2000 ,47( 2 ) 347-362.
- [ 5 ] Abt H A. Why some papers have long citation lifetimes. Nature ,1998 ,395 756-757.
- [ 6 ] Bak P , Chen K. Self-organized criticality. Scientific American ,1991 264( 1 ) 46-53.
- [ 7 ] Chen C , Morris S. Visualizing evolving networks : Minimum spanning trees versus pathfinder networks. Proceedings of IEEE Symposium on Information Visualization ( pp. 67-74 ) , Seattle, Washington :IEEE Computer Society Press ,2003.
- [ 8 ] Small H G , Griffith B C. The structure of scientific literatures i : Identifying and graphing specialties. Science Studies ,1974 4 :17-40.
- [ 9 ] Garfield E. Scientography : Mapping the tracks of science. Current Contents : Social & Behavioural Sciences ,1994 7 ( 45 ) 5-10.
- [ 10 ] Garfield E H , S I , Torpie R J. The use of citation data in writing the history of science. Philadelphia : Institute for Scientific Information ,1964.
- [ 11 ] Morris S A , Yen G , Wu Z , et al. Timeline visualization of research fronts. Journal of the American Society for Information Science and Technology ,2003 ,55( 5 ) :413-422.
- [ 12 ] Persson O. The intellectual base and research fronts of jasis 1986-1990. Journal of the American Society for Information Science ,1994 45( 1 ) 31-38.
- [ 13 ] Burton R E , Kebler R W. The ' half-life ' of some scientific and technical literatures. American Documentation ,1960 ,11 :18-22.
- [ 14 ] Braam R R , Moed H F , Raan A F J v. Mapping of science by combined co-citation and word analysis ii : Dynamical aspects. Journal of the American Society for Information Science ,1991 42( 4 ) 252-266.
- [ 15 ] Kessler M M. Bibliographic coupling between scientific papers. American Documentation ,1963 ,14 :10-25.
- [ 16 ] Small H G. A co-citation model of a scientific specialty : A longitudinal study of collagen research. Social Studies of Science ,1977 7 :139-166.
- [ 17 ] Small H. Visualizing science by citation mapping. Journal of the American Society for Information Science ,1999b ,50 ( 9 ) :799-813.
- [ 18 ] Small H. Paradigms , citations , and maps of science : A personal history. Journal of the American Society for Information Science and Technology ,2003 ,54( 5 ) :394-399.
- [ 19 ] Popescul A , Flake G W , Lawrence S , et al. Clustering and identifying temporal trends in document databases. Proceedings of IEEE Advanced in Digital Libraries ( ADL 2000 ) ( pp. 173-182 ) , Washington , DC 2000.
- [ 20 ] Roy S , Gevry D , Pottenger W M. Methodologies for trend detection in textual data mining. Proceedings of Proceedings of the Textmine ' 02 Workshop at the Second SIAM International Conference on Data Mining ( pp. Washington , DC 2002.
- [ 21 ] Erten C , Harding P J , Kobourov S G , et al. Exploring the computing literature using temporal graph visualization ( Technical Report TR0304 ) : University of Arizona 2003.
- [ 22 ] Kontostathis A , Galitsky L , Pottenger W M , et al. A survey of emerging trend detection in textual data mining // Berry M. A comprehensive survey of text mining. Heidelberg , Germany : Springer-Verlag 2003.
- [ 23 ] Allan J , Papka R , Lavrenko V. Online new event detection and tracking. Proceedings of ACM SIGIR ,1998 37-45.
- [ 24 ] Swan R , Allan J. Extracting significant time varying features

- from text. Proceedings of Eighth International Conference on Information Knowledge Management ( CIKM '99 ) pp. 38-45 , Kansas City , Missouri :ACM ,1999.
- [ 25 ] Havre S , Hetzler E , Whitney P , et al. Themriver : Visualizing thematic changes in large document collections. IEEE Transactions on Visualization and Computer Graphics , 2002 , 8 ( 1 ) :9-20.
- [ 26 ] Griffith B C , Small H , Stonehill J A , et al. The structure of scientific literatures ii : Towards a macro- and microstructure for science. Science Studies , 1974 , 4 :339-365.
- [ 27 ] Small H. A passage through science : Crossing disciplinary boundaries. Library Trends , 1999a , 48 ( 1 ) :72-108.
- [ 28 ] Granovetter M. Strength of weak ties. American Journal of Sociology , 1973 , 8 :1360-1380.
- [ 29 ] Burt R S. Structural holes : The social structure of competition. Cambridge , Massachusetts :Harvard University Press ,1992.
- [ 30 ] Girvan M , Newman M E J. Community structure in social and biological networks. Proc. Natl Acad. Sci. USA , 2002 , 99 :8271-8276.
- [ 31 ] Garfield E , Pudovkin A I , Istomin V S. Why do we need algorithmic historiography ? Journal of the American Society for Information Science and Technology , 2003 , 54 ( 5 ) :400-412.
- [ 32 ] Small H. The synthesis of specialty narratives from co-citation clusters. Journal of the American Society for Information Science , 1986 , 37 ( 3 ) :97-110.
- [ 33 ] Schvaneveldt R W. Pathfinder associative networks : Studies in knowledge organization. Norwood , New Jersey : Ablex Publishing Corporations , 1990.
- [ 34 ] Chen C. Visualising semantic spaces and author co-citation networks in digital libraries. Information Processing and Management , 1999 , 35 ( 2 ) :401-420.
- [ 35 ] Chen C , Paul R J. Visualizing a knowledge domain 's intellectual structure. Computer , 2001 , 34 ( 3 ) :65-71.
- [ 36 ] Chen C. Searching for intellectual turning points : Progressive knowledge domain visualization. Proc. Natl. Acad. Sci. USA , 2004 , 101 :5303-5310.
- [ 37 ] Small H , Greenlee E. Citation context analysis and the structure of paradigms. Journal of Documentation , 1980 , 36 ( Sept ) :183-196.
- [ 38 ] Boyack K W , Wylie B N , Davidson G S. Domain visualization using vxsight for science and technology management. Journal of the American Society for Information Science and Technology , 2002 , 53 ( 9 ) :764-774.
- [ 39 ] Chen C , Kuljis J. The rising landscape : A visual exploration of superstring revolutions in physics. Journal of the American Society for Information Science and Technology , 2003 , 54 ( 5 ) :435-446.
- [ 40 ] Alvarez L W , Alvarez W , Asaro F , et al. Extraterrestrial cause for the cretaceous-tertiary extinction. Science , 1980 , 208 ( 4448 ) :1095-1098.
- [ 41 ] Wignall P B H A. Anoxia as a cause of the permian/triassic extinction : Facies evidence from northern italy and the western united states. Palaeogeography Palaeoclimatology Palaeoecology , 1992 , 93 :21-46.
- [ 42 ] Becker L , Poreda R J , Hunt A G , et al. Impact event at the permian-triassic boundary : Evidence from extraterrestrial noble gases in fullerenes. Science , 2001 , 291 ( 5508 ) :1530-1533.
- [ 43 ] Becker L , Poreda R J , Basu A R , et al. Bedout : A possible endpermian impact crater offshore of northwestern australia. Science , 2004 , 304 ( 5676 ) :1469-1476.
- [ 44 ] Wignall P B , Hallam A. Griesbachian ( earliest triassic ) palaeoenvironmental changes in the salt range , pakistan and south-east china and their bearing on the permo-triassic mass extinction. Palaeogeography , Palaeoclimatology , Palaeoecology , 1993 , 102 :215-237.
- [ 45 ] Chen C , Cribbin T , Macredie R , et al. Visualizing and tracking the growth of competing paradigms : Two case studies. Journal of the American Society for Information Science and Technology , 2002 , 53 ( 8 ) :678-689.
- [ 46 ] Alvarez W. T. Rex and the crater of doom. New York : Vintage Books , 1997.
- [ 47 ] Cooper G J , Maynard R L , Cross N L , et al. Casualties from terrorist bombings. J Trauma , 1983 , 23 ( 11 ) :955-967.
- [ 48 ] Frykberg E R , Tepas J J r , Alexander R H. The 1983 beirut airport terrorist bombing : Injury patterns and implications for disaster management. Am Surg. , 1989 , 55 ( 3 ) :134-141.
- [ 49 ] Katz E , Ofek B , Adler J , et al. Primary blast injury after a bomb explosion in a civilian bus. Ann Surg. , 1989 , 209 ( 4 ) :484-488.
- [ 50 ] Mallonee S S S , Stennies G , Waxweiler R , et al. Physical injuries and fatalities resulting from the Oklahoma city bombing. JAMA , 1996 , 276 ( 5 ) :382-387.
- [ 51 ] Schuster M A , Stein B D , Jaycox L H , et al. A national survey of stress reactions after the september 11 , 2001 , terrorist attacks. N Engl J Med , 2001 , 345 ( 20 ) :1507-1512.
- [ 52 ] Galea S , Ahern J , Resnick H , et al. Psychological sequelae of the september 11 terrorist attacks in new york city. N Engl J Med , 2002 , 346 ( 13 ) :982-987.
- [ 53 ] North C S , Nixon S J , Shariat S , et al. Psychiatric disorders among survivors of the oklahoma city bombing. Journal of

- the American Medical Association , 1999 , 282 ( 8 ) : 755-762.
- [ 54 ] Holloway H C , Norwood A E , Fullerton C S , et al. The threat of biological weapons : Prophylaxis and mitigation of psychological and social consequences. JAMA , 1997 , 278 ( 5 ) : 425-427.
- [ 55 ] Schlenger W E , Caddell J M , Ebert L , et al. Psychological reactions to terrorist attacks. JAMA , 2002 , 288 ( 5 ) : 581-588.
- [ 56 ] Sandler T , Tschirhart J T , Cauley J. A theoretical analysis of transnational terrorism. American Political Science Review , 1983 , 77 ( 4 ) : 36-54.
- [ 57 ] Bush V. As we may think. The Atlantic Monthly , 1945 , 176 ( 1 ) : 101-108.
- [ 58 ] Carpenter M P , Cooper M , Narin F. Linkage between basic research literature and patents. Research Management , 1980 , 23 ( 2 ) : 30-35.
- [ 59 ] Tijssen R J W , Vanraan A F J. Mapping changes in science and technology - bibliometric cooccurrence analysis of the r-and-d literature. Evaluation Review , 1994 , 18 ( 1 ) : 98-115.
- [ 60 ] Swanson D R. Undiscovered public knowledge. Library Quarterly , 1986 , 56 ( 2 ) : 103-118.
- [ 61 ] Swanson D R. Online search for logically-related noninteractive medical literatures - a systematic trial-and-error strategy. Journal of the American Society for Information Science , 1989 , 40 ( 5 ) : 356-358.
- [ 62 ] Smalheiser N R , Swanson D R. Using arrowsmith : A computer-assisted approach to formulating and assessing scientific hypotheses. Computer Methods and Programs in Biomedicine , 1998 , 57 ( 3 ) : 149-153.

( 责任编辑 王建平 )