·学习·发现·交流·

CiteSpace Ⅱ在新发传染病研究领域的应用

陈红光　刘民

【导读】 探讨可视化软件CiteSpace Ⅱ在新发传染病研究领域的应用。采用基于Java平台的知识图谱分析软件CiteSpace Ⅱ进行研究。用可视化分析软件通过对国别、机构、术语、被引文献、被引期刊等节点的分析，可明确开展新发传染病研究的参照对象，识别学科研究前沿热点，快速确定该领域的重要文献资料，掌握该领域的支持性期刊种类等。可视化技术分析工具及方法在新发传染病研究领域的应用具有一定适用性，也有其局限性，但对于医学科研人员宏观、准确、快速掌握本研究领域基础及研究进展具有一定帮助。

【关键词】 新发传染病；可视化

Application of visualization on emerging infectious disease based on CiteSpace Ⅱ

CHEN Hong-guang, LIU Min. Department of Epidemiology and Biostatistics, Peking University Health Science Center, Beijing 100191, China
Corresponding author: LIU Min, Email: liumin@bjmu.edu.cn

This work was supported by a grant from the National High Technology Research and Development Program of China (863 Program) (No. 2008AA02Z416).
【Key words】 Emerging infectious diseases; Visualization

CiteSpace Ⅱ可视化分析技术当前受到普遍关注[1,2]。该技术是由美国Drexel大学信息科学与技术学院研制，适合进行多元、分时、动态复杂网络分析。在我国该项技术曾被应用于科技期刊文献信息可视化分析[3,4]。本文则以近年来国际公共卫生领域比较关注的新发传染病为例，探讨CiteSpace Ⅱ可视化技术在医学领域的应用。

1. 资料与方法：

(1) 资料来源：*EMERGING INFECTIOUS DISEASES*(*EMERG INFECT DIS*)作为全球范围内新发传染病研究领域的重要期刊，在一定程度上能够引领全球该领域的研究重点及方向。因此本文以该期刊所构成的知识图谱为基础进行可视化分析。

(2)数据获取方法：在Web of Science(WoS)数据库中，以EMERGING INFECTIOUS DISEASES为检索词，检索范围为“出版物名称”。检索的时间跨度为数据库收录该期刊的所有年份(1995－2011年)，数据库选择SCI-EXPANDED、SSCI、A&HCI、CPCI-S、CPCI-SSH，整理出检索文献，以download*.txt为文件名，选择带参考文献的全著录格式下载并存储，数据获取时间为2011年9月26日。

(3)分析方法：选择CiteSpace Ⅱ为知识图谱分析工具[5]。该软件在对某一特定学科或技术领域进行分析时，通过关键词共现分析、文献共被引分析，绘制该学科或技术领域的科学知识图谱，探讨该学科领域的前沿热点及其演化过程。

研究领域的概念和可视化是基于信息科学中的两个概念，即研究前沿和知识基础间的时变对偶。“研究前沿”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是其在科学文献中(即由科学文献形成的演化网络)的引文和共引轨。在CiteSpace Ⅱ中，研究前沿是基于题目、摘要、系索词(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的突变专业术语而确定的，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。同时，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，后者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[1]。具体概念模型见图1。

本文所有生成图谱均按照网络路径简化算法(Pathfinder算法)进行处理。Pathfinder网络最初应用于作者共引分析[6]，尔后扩展到一般的共引分析。路径网络简化是依据一个三角不等式检验以决定是否保留某个连接，判断标准是一个单连接路径其长度不能超过多个连接路径的长度。采用Pathfinder精简后可减少连接数量使网络主要结构

DOI: 10.3760/cma.j.issn.0254-6450.2012.06.020
基金项目：国家高技术研究发展计划(863计划)(2008AA02Z416)
作者单位：100191 北京大学公共卫生学院流行病与卫生统计学教研室
通信作者：刘民，Email：liumin@bjmu.edu.cn

更清晰显现。另选择“time slicing”值为2，即选择每两年为一个时间间隔进行处理，以利于辨识新发传染病研究发展过程和学科前沿的动态模式，同时提高软件运行速度和准确。

2. 检索结果：按照以上检索策略共得到5719条文献记录，其中以“ARTICLE”和“LETTER”为主要文献类型，考虑到各种文献承载信息的重要程度及时间顺序，本研究选择原创性较强的“ARTICLE”及“PROCEEDINGS PAPER”作为可视化分析对象，最终得到文献3848篇（其中会议交流论文均被期刊收录，已在数据库中分类标识）进行分析。

（1）新发传染病研究领域的主要国家及机构：CiteSpace Ⅱ软件在可视化分析中，采用引文年代环的形式来表现引文历史，其中引文年轮的颜色代表相应的引文时间，年轮的厚度与某个时区内引文数量成正比，年代环之间的连线表明之间的共引或合作关系，节点的大小与发表文献的数量成正比。图2表明，分析新发传染病相关研究领域文献数量，居前列的国家依次为美国、法国、英国、加拿大、中国、荷兰、澳大利亚等，通过软件的引文突增功能（如图2中粉红色核心的圈，表明近期引文数量突增）可以发现美国佐治亚州、马里兰州、纽约等地区近期该研究比较活跃。同时还可以通过Google Earth对得到的数据进行模拟[7]，如图3中连线表示合作研究之间的关系或具有共同的研究方向，整体上看新发传染病研究领域呈现以美国和欧洲为中心的全球性交互网状分布，说明随着传染病发生、发展及控制的全球化趋势，其对应的研究之间也呈现全球化趋势，各国间合作紧密。

CiteSpace Ⅱ也可对研究机构进行分析。如图4所示，全球范围内该领域的排位比较靠前的研究机构包括美国疾病预防控制中心、法国巴斯德研究所、美国德克萨斯大学、荷兰国家公共卫生及环境研究院、美国哈佛大学、美国纽约州卫生科学中心及世界卫生组织等，以上研究机构均为各国负责传染病预防和控制研究的主体单位，各研究所之间合作也比较紧密。同样通过“引文突增”也可发现美国德克萨斯大学、中国疾病预防控制中心、美国纽约州卫生科学中心等机构近期研究比较活跃，这与各国近期传染病发生有关。

图5是通过对文献提取的术语进行聚类分析，以发现各自的研究历史和目前研究方向及不同研究

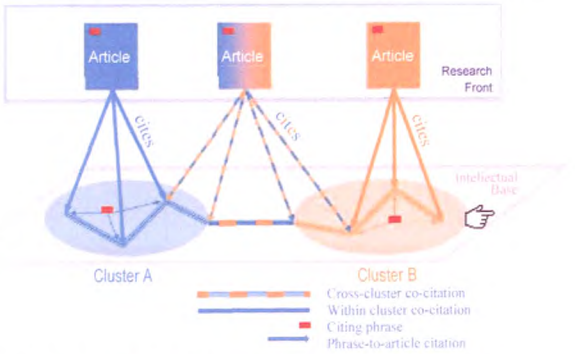

图1 CiteSpace Ⅱ的概念模型（以时间显示研究领域的演变）

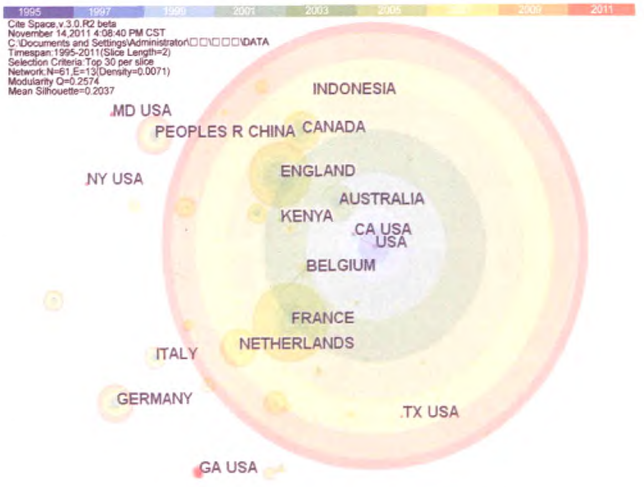

图2 新发传染病研究领域主要国家的分布

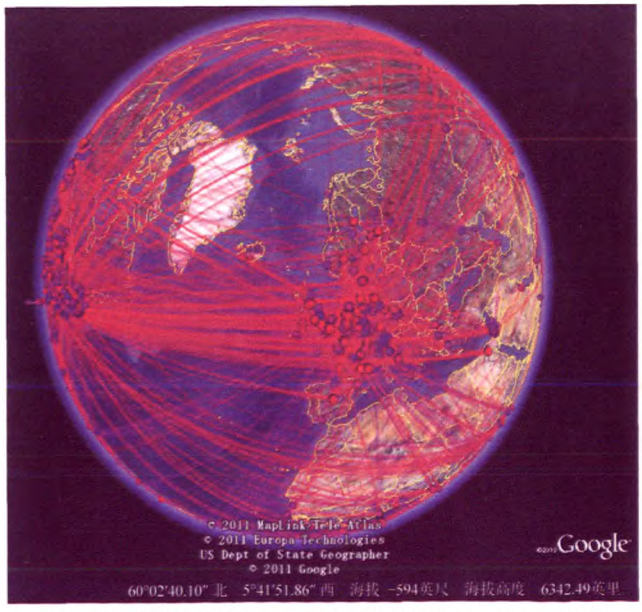

图3 全球新发传染病研究领域的网络分布（Google Earth模拟）

机构间重点方向的差异。如美国埃默里大学和华盛顿大学的研究重点为传染病治疗；美国纽约卫生科学中心、哈佛大学及英国健康保护局的重点研究为传染病蚊媒监测；美国疾病预防控制中心涉及的领域较宽，其中包括严重急性呼吸道综合征及发生在美国邮政工人的传染病等；美国德克萨斯大学主要为针对百日咳杆菌的研究等。还可通过时间轴来分析该机构在该领域的研究历史，如自2007年始针对环丙沙星治疗的敏感性及耐药性研究就成为关注点。此外该软件还可展示全球该领域比较活跃的研究团队。

（2）新发传染病研究领域的热点分析：在可视化网络中CiteSpace Ⅱ软件通过有意义的关键点辨认研究热点领域，是基于突显频率的关键词及术语分析揭示研究的热点。如图6显示针对默克尔细胞多瘤病毒研究、大环内酯类抗生素的耐药机制研究、可变数目串联重复位点研究、非结核分枝杆菌研究、16S rRNA基因的研究及耐药结核病研究等均是近期研究重点，代表新发传染病的研究前沿。通过图6中各种连线也可了解不同年代的研究之间明显的递呈关系，表明研究前沿与其基础之间关系非常紧密。此外还可通过对关键词的聚类分析获得新发传染病的研究热点。

（3）新发传染病研究领域的高被引文献分析：引用频次在一定程度上可反映该文献在新发传染病研究领域的价值。通过文献共被引网络图谱，可进行文献引用频次及排序分析。如图7显示，一篇关于分离菌株脉冲场凝胶电泳分型的文章（Tenover FC, 1995）其被引频次最高，从引文年轮可见至今仍被大量引述；其次是第一篇关于分子进化遗传分析软件MEGA4的文章（Koichiro Tamura, 2007），至今仍是非常重要的文献，说明该软件还是分子研究领域中非常重要的分析工具；图7中呈蓝色的聚类网络为1994年有关经水传播隐孢子虫感染的研究，但随着隐孢子虫感染的降低，此类研究在2000年后有所减少。图7还显示新发传染病各项研究间的分布相对分散，这与各国传染病的发生及发展有关。图7中粉红色圈表示通过“引文频率突增”确定的增速较快的被引文献，也反映了研究重点方向及趋势。除了以聚类簇的形式展示知识图谱外，还可通过时间线的形式展现某主题研究的历史轨迹及与其他主题的关系。

（4）新发传染病研究领域相关期刊：CiteSpace Ⅱ还可对被引文献的期刊进行分析，从中发现各研究领域与各学科之间的交叉合作关系。如图8显示，与*EMERG INFECT DIS*相关的期刊包括*JOURNAL OF CLINICAL MICROBIOLOGY*（*J CLIN MICROBIOL*）、*VIROLOGY*、*CLINICAL MICROBIOLOGY REVIEWS*（*CLIN MICROBIOL REV*）、*JOURNAL OF VIROLOGY*（*J VIROL*）等针对传染病病原学研究的杂志，以*JOURNAL OF INFECTIOUS DISEASES*（*J INFECT DIS*）、*CLINICAL INFECTIOUS DISEASES*（*CLIN INFECT DIS*）、

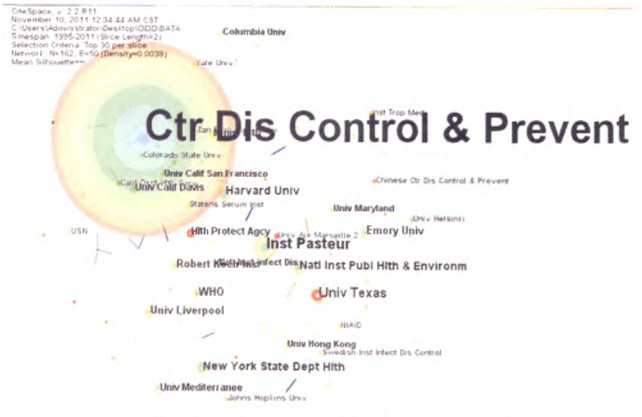

图4　新发传染病研究领域的主要科研机构分布

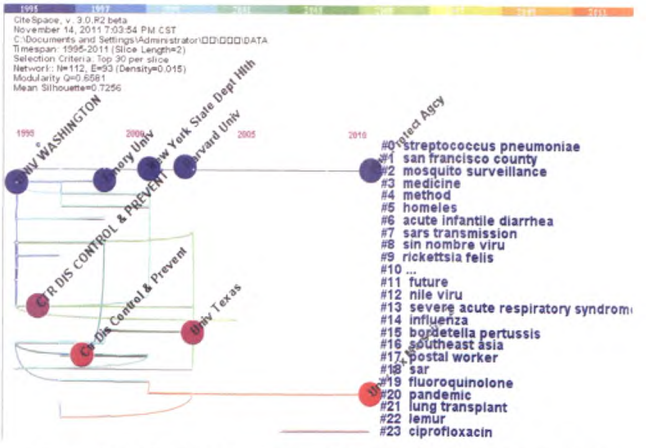

图5　新发传染病研究领域不同研究机构的研究历史及其方向

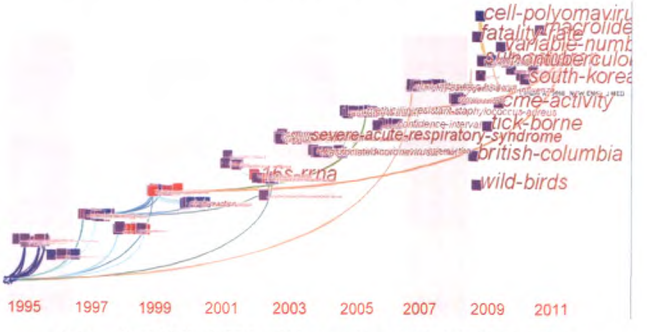

图6　新发传染病研究领域基于专业术语的研究热点分析

THE PEDIATRIC INFECTIOUS DISEASE JOURNAL（*PEDIA TR INFECT DIS J*）等为主的传染病专科杂志，以 *ANNALS OF INTERNAL MEDICINE*（*ANN INTERN MED*）、*ANTIMICROBIAL AGENTS AND CHEMOTHERAPY*（*ANTIMICROB AGENTS CH*）、*BRITISH MEDICAL JOURNAL*（*BRIT MED J*）、*THE JOURNAL OF THE AMERICAN MEDICAL ASSOCIATION*（*J AM MED ASSOC*）等为主的针对传染病治疗的杂志，以 *EPIDEMIOLOGY & INFECTION*（*EPIDEMIOL INFECT*）为代表的传染病流行病学杂志，以 *THE NEW ENGLAND JOURNAL OF MEDICINE*（*NEW ENGL J MED*）、*LANCET*、*SCIENCE* 等为代表的医学综合类高水平杂志，以及以美国疾病预防控制中心 *MORBIDITY AND MORTALITY WEEKLY REPORT*（*MMWR-MORBID MORTAL W*）为代表的信息类文献。从相关期刊分布可见其对新发传染病研究涉及领域较为宽泛，包括流行病学现场、病原学识别与发现、临床治疗等。这些期刊确实引领新发传染病研究的发展方向和趋势。

3. 分析：CiteSpace Ⅱ可视化技术是借助科学文献引文网络的可视化分析，监测科学期刊文献中出现的研究前沿、热点、趋势和动向的一种模式化探测和可视化研究。本文运用CiteSpace Ⅱ对新发传染病的研究领域进行可视化分析。

新发传染病是1995年国际上出现的一个新词汇，是指近30年在人群中新认识或新发现的并造成地域性或国际性公共卫生问题的传染病。本文通过对该领域全方位可视化研究，快速明确了开展新发传染病研究领域的主体国家、机构及团队，为掌握并跟踪该项研究的发展提供了参照对象；通过对关键词及术语的分析，准确把握该领域的研究重点及前沿方向，如与新发传染病有关病毒变异、耐药、动物疫源性传播及传染病疾病负担问题等；通过被引重点个案文献的分析及回顾，探讨该领域研究过程中的演化及发展（本文因选题为新发传染病，其学科演化历程并不明显，呈现一种分散性研究分布）；通过对该研究领域期刊的分析，为进一步研究提供参考依据。此外，该软件还可用于针对某领域学科的分析，通过分析该学科与其他学科间的交叉和融合，预测一些潜在的新兴学科或交叉学科的出现，对于新学科的培育和孵化具有一定的帮助。

本分析软件有一定局限性。如只能分析已发表的公开文献，对于不以文献为载体或为主的研究领域的分析则不适合，另外目前该软件只能开展基于英文文献的可视化分析。此外，本文知识图谱分析的结果也存在局限性。首先，受到数据库选择的限制，数据的不完整可能对结果有一定影响；其次，仅选用数据库中 *EMERG INFECT DIS* 进行分析，从文献来源上可能存在一定偏倚。但并不影响分析方法的使用及其结果的借鉴。

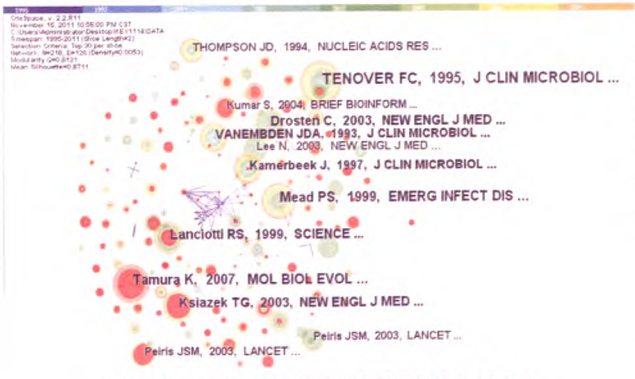

图7 新发传染病研究领域被引文献中心性分布

图8 新发传染病研究领域共被引文献频次分布

参 考 文 献

[1] Chen CM. CiteSpace Ⅱ: detecting and visualizing emerging trends and transient patterns in scientific literature. J Am Soc Inf Sci Technol, 2006, 57(3): 359-377.
[2] Synnestvedt MB, Chen C, Holmes JH, et al. CiteSpace Ⅱ: visualization and knowledge discovery in bibliographic databases. AMIA Annu Symp Proc, 2005: 724-728.
[3] Liu ZY, Wang XW. Information visualization of research fronts of ecological economics. J Southwest Forestry College, 2008, 28(4): 4-11. (in Chinese)
刘则渊，王贤文. 生态经济学研究前沿及其演进的可视化分析. 西南林学院学报，2008，28(4)：4-11.
[4] Hou JH, Chen Y. Research on visualization of the evolution of strategic management front. Studies in Science of Science, 2007, 25 Suppl 1: S15-21. (in Chinese)
侯剑华，陈悦. 战略管理学前沿演进可视化研究. 科学学研究，2007，25增刊1：15-21.
[5] http://cluster.cis.drexel.edu/~cchen/citespace.acess; 2010-9-10.
[6] White HD. Pathfinder networks and author co-citation analysis: a remapping of paradigmatic information scientists. J Am Soc Inf Sci Technol, 2003, 54(5): 423-434.
[7] Chen C, Zhu W. Tracing conceptual and geospatial diffusion of knowledge. LNCS, 2007, 4564: 265-274.

（收稿日期：2011-11-18）
（本文编辑：张林东）