

## 基于 NEViewer 的医学科研主题演化可视化分析

孙 静<sup>1</sup> 程齐凯<sup>2</sup> 张 雯<sup>3</sup>

[摘要]获取 22 种学报类医学期刊 2005-2013 年文献的题录信息,建立高频词的共词网络,利用网络演化可视化分析软件 NEViewer 对医学科研主题进行演化分析,发现我国医学科研集中在基础研究与应用研究层面,多发病病种(如糖尿病、肿瘤等)的防治研究及技术方法的研究是两大主流方向,医学科研主题将在较长时期内保持稳定发展状态,最后对热点医学主题进行了阐释。

[关键词]共词网络;关键词;NEViewer;医学主题;可视化分析;学报

[中图分类号]G254.2;G255.2;R-058

[文献标志码]A

[文章编号]1671-3982(2014)10-0056-05

### NEViewer-based visual analysis of medical scientific research topics evolution

SUN Jing<sup>1</sup>, CHENG Qi-kai<sup>2</sup>, ZHANG Wen<sup>3</sup>

(1. Taishan Medical College Information Engineering School, Tai'an 271016, Shandong Province, China; 2. Wuhan University Information Management School, Wuhan 430072, Hubei Province, China; 3. Taishan Medical College Nursing School, Tai'an 271016, Shandong Province, China)

[Abstract] A co-words network for high frequency words was established using the title information of papers published in 22 medical journals from 2005 to 2013. The evolution of medical scientific research topics was analyzed with the visual analysis software NEViewer, which showed that medical research in China was focused on basic and application studies with the prevention and treatment of frequently encountered diseases and the associated technical methods as the two main directions, and that medical scientific research topics would remain their stable status in a rather long period. The hot medical topics were explained.

[Key words] Co-words network; Key words; NEViewer; Medical headings; Visual analysis; Academic journal

科学研究的成果通常以学术文献的形式呈现,而文献中的关键词则是文献主题内容的概括和体现。当两个能够表达某一学科领域研究主题或研究方向的关键词同时出现在同一文献中,说明这两个词之间有一定的内在关系。以关键词为节点,共

现关系为边,构建出由关键词及其共现关系形成的网络——“共词网络”,揭示了特定学科领域的科研主题及其相互关系。不同时间段的共词网络存在差异,体现了共词网络的演化过程,同时也揭示了研究主题的演化<sup>[1]</sup>。

可视化是将数据、信息和知识转化为看得见的表现形式并对数据形成更深层次认识的过程。作为一种可以放大人类感知的数据、信息、知识的表示方法,可视化日益受到重视并得到越来越广泛的应用<sup>[2]</sup>。NEViewer(Network Evolution Viewer)是一款由武汉大学信息管理学院开发的用于复杂网络演化可视化分析的软件。

笔者统计了 2005-2013 年在 22 种学报类医学

[基金项目]国家社科基金重大项目“基于语义的馆藏资源深度聚合与可视化展示研究”(11&ZD152)之“语义化馆藏资源深度聚合知识服务研究”

[作者单位]1.泰山医学院信息工程学院,山东泰安 271016; 2.武汉大学信息管理学院,湖北武汉 430072; 3.泰山医学院护理学院,山东泰安 271016

[作者简介]孙 静(1980-)女,山东莱芜人,硕士,讲师,参编教材 1 部,发文 2 篇。

核心期刊发表论文的关键词,建立不同时间段的共词网络,利用网络演化与趋势探测系统 NEViewer 识别我国医学领域不同时期科研主题,并对其进行演化分析。

### 1 数据获取

学报是指专门进行学术研究成果报道的学术类期刊,信息蕴含量高、情报价值大,基本能够反映我国医学领域的科研水平。《中文核心期刊目录总

览》列出了综合性医药卫生类的 37 种核心期刊,其中 22 种期刊为大学学报。

选择 CNKI 中国知网-期刊数据库,设置检索条件为:来源期刊依次为北京大学学报(医学版)等 22 种大学学报期刊(表 1),期刊年限为 2005-2013 年。对采集的题录信息进行筛选,去除会议通知、征稿启事等非学术文献信息,获取有效题录 62 075 条,并将其导入文献管理软件 NoteExpress。

表 1 22 种学报类期刊及其文献数量

序号	期刊名称	文献数量/篇	序号	期刊名称	文献数量/篇
1	北京大学学报(医学版)	1624	12	南京医科大学学报(自然科学版)	3940
2	浙江大学学报(医学版)	1001	13	中国医科大学学报	2991
3	第三军医大学学报	6966	14	武汉大学学报(医学版)	1482
4	西安交通大学学报(医学版)	1771	15	郑州大学学报(医学版)	3820
5	四川大学学报(医学版)	2432	16	哈尔滨医科大学学报	1838
6	南方医科大学学报	5534	17	中国医学科学院学报	1513
7	吉林大学学报(医学版)	3673	18	中山大学学报(医学科学版)	2277
8	上海交通大学学报(医学版)	3520	19	安徽医科大学学报	2663
9	第二军医大学学报	3382	20	中南大学学报(医学版)	1911
10	重庆医科大学学报	3536	21	华中科技大学学报(医学版)	1978
11	山东大学学报(医学版)	2563	22	复旦学报(医学版)	1660

### 2 构建共词网络

为了建立时序共词网络,以三年为一个时间段,将获取题录信息划分为 2005-2007 年、2008-2010 年、2011-2013 年三个时间段。由于获取的关键词数据量巨大,过多的数据会降低共词网络分析的效果,因此需要对数据进行一定的预处理。

提取每个时间段的关键词信息并进行词频统计。医学涵盖了基础医学、临床医学、内科学、外科学等多个学科,而各领域的研究内容也是繁杂而精细的,因此词频统计出现了典型的长词尾现象。在所有 72 124 个关键词中,词频为 1 的关键词比例高达 66.9%。为了保证共词网络分析的效果,本文对低词频的关键词进行过滤。此外,“大鼠”、“患者”、“治疗”等词汇作为通用词出现于不同的领域中,虽然具有较高的词频但会严重影响社区划分的效果,因此也对这种通用词汇进行了过滤处理。最终获得各时间段的高频关键词(词频大于或等于 5):2005-2007 年 2 888 个,2008-2010 年 3 046 个,2011-2013 年 2 836 个。

按照上述过滤策略对关键词进行处理后,将各个时间段的高频关键词按照其共现关系构建共词网络,构建的 3 个共词网络节点数分别为 2 888、3 046 和 2 836,边数分别为 35 509、37 206 和 32 431,符合 NEViewer 的网络分析规模。

### 3 演化分析

网络演化与趋势探测系统 NEViewer 的数据基础是时序共词网络,因此将不同时间段的共词网络存储为 nwb 格式文件,并放置于同一文件夹内,在 NEViewer 内加载文件后即可进行探测演化。对上述不同时间段的共词网络进行主题识别与演化分析的结果见图 1。

NEViewer 以冲积图的形式展示了医学研究领域的热点主题及演化过程。图 1 中,矩形颜色块表示社区,两个时间段之间的曲线形色块表示演化的过程,而矩形块的高度则表示社区的节点规模,前一个时间段的主题(社区)与后一时间段的主题(社区)之间存在的演化关系可通过色块的融合、分化加以体现<sup>[3]</sup>。

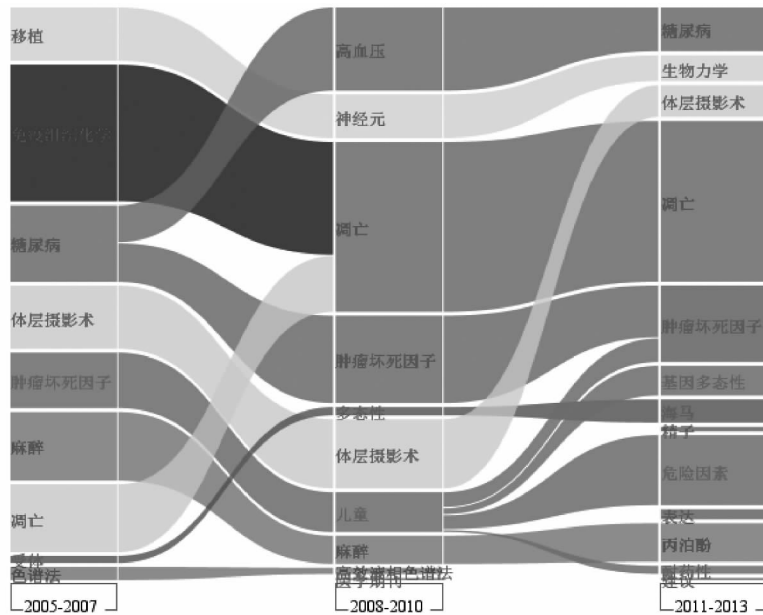


图 1 医学科研主题演化冲积图

### 3.1 主题识别

在复杂社会网络中,社区是一个极为重要的概念,它由一群高度聚集、联系紧密的节点聚集而成。共词网络中,联系紧密的关键词节点也构成了社区,不同的词汇社区代表了特定的学科主题及研究方向。社区划分依赖于社区发现算法,学术界提出了多种有效的社区发现算法,如 Blondel 分区算法、Newman MM 分区算法、Ball overlapping 分区算法等。在 NEViewer 系统中,选择 Blondel 算法对各个时间段的共词网络社区进行发现,并选择社区中的核心节点关键词为社区命名(主题标识)。在选择核心节点关键词时,系统采用 Z-Value 值这个指标,可以衡量网络节点与其他节点联系的紧密性,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<sup>[1]</sup>。图 1 比较清晰地呈现了不同时间段的热点科研主题,如 2005-2007 年的移植、免疫组织化学、糖尿病、体层摄影术、肿瘤坏死因子、麻醉、凋亡等。

在社区划分即主题识别之后,NEViewer 提供了主题排序功能,排序越靠前的社区越靠近图形的顶端。系统提供两种排序方法,一种是社区节点数量,即按照学科主题的词汇规模排序,节点数量越多,排序越靠前;第二种是根据社区中心度,将社区看作节点,计算社区在网络中的度值,度值越大,排序越靠前。图 1 为按照中心度进行主题排序的结果。

### 3.2 演化分析

在共词网络中,网络社区并不是一成不变的。网络社区的变化既包括社区自身内部节点、关系和结构的变化,也包括社区间位置和关系的变化。如图 1 中“糖尿病”在位置上显示出明显的上升趋势,而“凋亡”不仅在位置上不断靠前,社区规模更是大幅度增加。

不同时间段网络社区间关系的变化则体现了学科主题的演化。为不同学科主题建立合理的联系就是为所有网络社区寻找合适的前驱和后继,而寻找社区前驱和后继实质上是一个度量社区间相似程度的问题。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演化关系。判断度量社区相似性有三种基本方法:节点重合度,关系相似性和前两个指标的结合,NEViewer 系统主要采用了节点重合度这一指标。图 1 为使用节点比重后向寻找相似度阈值设定为 0.3,得到的演化分析结果。

图 1 显示出学科主题的演化主要表现为三类:同一主题在不同时间段的延续,如“体层摄影术”;主题分化,如 2005-2007 年时间段的“糖尿病”逐渐分化为 2008-2010 年时间段的“高血压”和“肿瘤坏死因子”;主题融合,如 2005-2007 年时间段的“免疫组织化学”和“凋亡”融合为“凋亡”。

### 3.3 结果讨论

对医学领域的关键词建立共词网络、主题识别与演化分析之后,发现我国医学研究主要集中在基础研究层面,常见病、多发病病种(如糖尿病、肿瘤等)和研究方法(如免疫组化化学、细胞凋亡等)的研究是两大主流方向。随着时间的推移,研究热点和研究方法也有细微的变化,如随着肿瘤的相关研究逐渐深入,对器官移植的研究呈下降的趋势,研究方法由最初的免疫组织化学病理阶段逐渐过渡到分子生物学水平,同时细胞凋亡的研究以及基因的研究开始凸显。下面对发现的热点医学主题进行简单阐释。

#### 3.3.1 移植

在医学上,应用自体或异体的正常细胞、组织或器官置换病变的或功能缺损的细胞、组织或器官,以维持和重建机体的生理功能,这种治疗方法称为细胞移植、组织移植或器官移植<sup>[4]</sup>。由此定义可以看出,移植是一个涉及面较广的概念,更多的关键词可以与其建立直接的连接关系。作为核心节点关键词,“移植”高居图形顶端。观察移植代表社区内的关键词,我们发现词频较高的关键词为细胞培养、组织工程、移植、骨髓间充质肝细胞、内皮细胞等,因此将本社区的研究主题归结为移植免疫。在移植免疫研究中,成体干细胞的可塑性引起了人们的极大兴趣,其中骨髓间充质干细胞的功能作用和临床应用更成为医学界普遍关注的研究热点<sup>[5]</sup>。

在演化图形中,移植逐渐演化到神经元和生物力学社区,尽管主题命名采用了 Z-value 值较大的神经元和生物力学,但高频关键词依然为骨髓间充质肝细胞、细胞培养、移植等。因此,可以判定移植免疫自 2005-2007 年段至 2008-2010 年段、2011-2013 年段,一直是医学研究领域的热点。

#### 3.3.2 免疫组织化学和细胞凋亡

免疫组织化学和细胞凋亡是 2005-2007 年的另外两个研究热点。免疫组织化学是应用抗原和抗体结合的原理,检测细胞内多肽、蛋白质等大分子物质的分布。这种方法特异性强、敏感度高、发展迅速、应用广泛,成为生物学和医学众多学科的重要研究手段,尤其在肿瘤研究领域发挥强大作用<sup>[6]</sup>。

细胞凋亡成为研究热点,取决于细胞凋亡与临床病毒的密切关系,特别是与肿瘤之间的密切关系。

免疫组织化学和细胞凋亡在 2008-2010 年和 2011-2013 年时间段都融合到更大规模的细胞凋亡社区。细胞凋亡社区的高频关键词为凋亡、肿瘤、癌,说明肿瘤在 2005-2013 年一直是医学研究的热点问题,并在研究过程中采用了免疫组织化学技术及细胞凋亡等诸多研究方法。

#### 3.3.3 糖尿病和高血压

我国糖尿病与高血压的发病率逐年提高,两种疾病的防治亦是医学界的重要研究课题。二者作为同源性疾病,临床上许多高血压病人。经常伴有糖尿病,而糖尿病也较多地伴有高血压,因此二者在图中显示出交叉演化趋势。

糖尿病是一种以血糖代谢紊乱为特点的常见慢性疾病,其中 90% 以上为 2 型,而胰岛素抵抗又是 2 型糖尿病的重要发病机制。研究显示,血浆中某些炎性因子(如 C 反应蛋白、白细胞介素-6、肿瘤坏死因子- $\alpha$ )浓度的升高与胰岛素抵抗的发生密切相关,而肿瘤坏死因子- $\alpha$  正日益成为研究热点<sup>[7]</sup>。

#### 3.3.4 肿瘤坏死因子

肿瘤坏死因子一直是医学界研究比较深入、应用比较广泛的内容,在儿童疾病中也发挥着重要作用,因此与儿童疾病社区建立了演化关系。随着对儿童疾病研究的细化和深入,儿童疾病社区逐渐分化到 2011-2013 年的肿瘤坏死因子、基因多态性、危险因素等社区,显示对儿童疾病的研究开始由治疗向预防过渡,医学工作者更加强调对儿童疾病的产生、并发症、后遗症等危险因素的识别。

#### 3.3.5 体层摄影术

该社区显示的是以 CT 技术为代表的现代医学影像技术。医学影像由于含有极其丰富的人体信息,能以非常直观的形式展示人体内部的组织结构、形态或脏器的功能等,因此医学影像已成为医学研究及临床诊断中最活跃的领域之一。

#### 3.3.6 麻醉

麻醉学也是医学研究的重要领域,包括新药物、新技术和麻醉学基础与应用研究等。丙泊酚由于具有麻醉诱导起效快、苏醒迅速且功能恢复完善、不良反应少等优点而被广泛应用于临床麻醉。它对器官的保护作用与副作用(认知功能损害、脂代谢紊乱等)成为麻醉学近年来的研究热点<sup>[8]</sup>。

#### 4 总结

以医学期刊论文为基础,通过文献计量学的方法发现医学领域的科研结构及发展状态的文献并不少,但大多都集中在对论文发表时间、数量、作者机构等文献外围信息的统计分析<sup>[9]</sup>。少有的关于文献内容的研究也着眼于医学某一具体领域,分析方法多采用聚类分析、因子分析等方法<sup>[10]</sup>。本文深入到医学论文的研究内容本身,从复杂网络的角度,利用 NEViewer 实现对综合医学研究主题的发现识别、演化分析,并对其进行了可视化展示。

在研究过程中我们也发现一些不足之处。由于医学涵盖多个领域,研究内容繁杂,关键词数量众多,分布广泛,使得划分后的社区所代表的主题并不集中,而 NEViewer 系统仅仅采用 Z-value 值对社区进行命名也未能准确的表达社区的全部内容。

#### 【参考文献】

[1] 程齐凯,王晓光.一种基于共词网络社区的科研主题演化分析框架[J].图书情报工作,2013,57(8):91-96.  
 [2] 洪文学,王金甲.可视化和可视化分析学[J].燕山大学学报,

2010,34(2):95-99,105.

[3] 王晓光,程齐凯.基于 NEViewer 的学科主题演化可视化分析[J].情报学报,2013,32(9):900-911.  
 [4] 同种异型抗原的递呈与识别机制[EB/OL].[2014-04-17].http://www.docin.com/p-13773663.html.  
 [5] 武京国,曹谊林,肖 苒.骨髓间充质干细胞的免疫学特性及其同种异体的应用研究进展[J].组织工程与重建外科杂志,2012,8(1):49-52.  
 [6] 360 百科.免疫组织化学[EB/OL].[2014-04-17].http://baike.so.com/doc/6589807.html.  
 [7] 张天弼,冯慧斌,林 斌,等.2 型糖尿病合并高血压患者胰岛素抵抗与 C 反应蛋白、肿瘤坏死因子的关系研究[J].中国实用医药,2011,6(27):6-7.  
 [8] 俞卫锋.麻醉学科研究的现状与热点[J].现代实用医学,2013,25(2):121-123.  
 [9] 吕艳华,于 琦,贺培风.我国医学信息学研究分析[J].中华医学图书情报杂志,2013,22(6):25-35.  
 [10] 夏 怡,王德斌,柴 静,等.2002-2011 年 PubMed 收录的"肿瘤和网页"文献计量分析[J].预防医学情报杂志,2013,29(12):1097-1102.

[收稿日期:2014-06-20]

[本文编辑:刘 娜]

(上接第 40 页)

#### 5.3 开发资源,建立地方特色数据库

重新审视我国远程医学教育现状,研究分析教学资源开发上存在的问题,尤其是优质教学课件的开发、利用方面,需理清思路,探索其教学资源开发的新机制<sup>[5]</sup>。高校出台配套政策,鼓励优秀教师带头表率,一方面在充分调研和把握农村医务人员资源需求的基础充实课件,注重内容的针对性、实用性;另一方面,充分发挥网络优势,开发、整合、优化远程医学教育资源,建立适合农村医务人员需求的地方特色数据库。

#### 5.4 学习成果的认可

学习成果的认可是制约农村远程医学教育发展的重要因素。目前我国的远程教育未能像国外一样可获得相应的学历。以山西医科大学远程医学教育为例,起始阶段不提供学习成果的认可,只在达到一定程度后才给予类似成教学历的认定。这在一定程度上打击了参加远程教育医务人员的学习积极性。

#### 6 结语

本次调查表明,高等医学院校建立面向农村开

展远程教育是可行的,并且具有一定的紧迫性和重要性。面向农村开展远程医学教育是我国医学学科服务的趋势和方向,有利于提升乡镇社区基层医务人员的基本素质和医疗水平,改善农村医疗条件,提高农民的健康水平。

#### 【参考文献】

[1] 卫生部关于加强在职卫生技术人员现代远程医学教育工作的通知[EB/OL].(2004-09-17)[2014-05-20].http://www.moh.gov.cn/mohbgt/pw10402/200804/26666.shtml.  
 [2] 孟 群,姒健敏,董德刚,等.远程医学教育和远程医疗在美国大学中的发展状况[J].中华医学教育杂志,2007,27(4):1-4.  
 [3] 郭常平,茅晓延.我国远程医学教育的实践与思考[J].中国高等医学教育,2006(1):27-28,73.  
 [4] 王沁萍,曲成毅.远程医学教育的现状与进展[J].中国医学教育技术,2012,26(2):119-123.  
 [5] 安建民,蔡力民.美国远程医学教育的 Q-CATS 评估模式[J].国外医学:医学教育分册,2002,23(4):29-30.

[收稿日期:2014-06-09]

[本文编辑:黄思敏]