

从知识图谱看南京医科大学科学研究发展 ——基于SCI(2002-2013)

陈静¹, 谢靖², 黄华兴³

(1. 南京医科大学图书馆, 江苏 南京 210029; 2. 南京中医药大学经贸管理学院信息管理系, 江苏 南京 210023; 3. 南京医科大学第一临床医学院, 江苏 南京 210029)

摘要:文章以南京医科大学作为检索对象, 获取2002~2013年SCI收录南京医科大学发文, 并通过知识图谱展现12年间南京医科大学科学研究的发展历程。文章主要通过作者发文、科研合作、学术热点分别构建知识图谱, 从学者、机构、关键词角度可视化地计量了南京医科大学的科研发展。

关键词: 科研发展; 南京医科大学; 知识图谱; SCI

中图分类号: G255

文献标识码: A

文章编号: 1671-0479(2014)06-464-005

doi: 10.7655/NYDXBSS20140610

知识图谱的起源与发展源于引文分析理论、复杂网络系统、社会网络分析的兴起以及信息可视化提供的技术支持。通过知识图谱, 可以直观地将学科内知识(术语、机构、作者等)重新组织。目前美国、荷兰等国外的学者对知识图谱的研究已形成了一些较为成熟的理论和方法, 陈超美^[1]利用其研制开发的可视化工具 CiteSpace 构建了弦理论、恐怖主义、大规模物种灭绝领域的知识图谱, 其研究成果对知识图谱的发展有着重要影响。刘则渊^[2-3]较早将知识图谱的理论和概念引入国内科学计量研究, 并与陈超美合作将开源可视化工具介绍给国内科学计量学研究界。赵蓉英等^[4-5]以 Web of Science 内图书馆主题的文献为研究对象, 通过 CiteSpace 对相关引文数据和主题词数据进行分析和处理。王宁等^[6]利用 CiteSpace 对生物安全相关文献期刊共被引、关键词共现、文献共被引以及国际合作等进行可视化分析。

一、资料与方法

文章以汤森路透的SCI作为数据源, 以“Nanjing Medical University”为检索词, 通过机构地址检索^[7] Web of Science 的核心数据集, 共获取2002~2013年

南京医科大学发表的相关科技文献数据7610篇。通过知识图谱工具 CiteSpace^[8]构建作者发文合作图谱、关键词共现图谱展现12年间南京医科大学科学研究发展, 并统计了12年间与南京医科大学有紧密合作的学术机构及其国家。

二、结果与分析

(一) 南京医科大学科研关键词共现图谱

本研究绘制2002~2013年南京医科大学SCI发文的关键词共现图谱, 进而揭示南京医科大学科学研究的知识结构及其演进, 相关算法为 Pathfinder。以2年为一个时间切片, 参数调整后可视化图谱见图1。

图1中按照频次的大小来控制节点标签的显示数量, 节点大小与其代表的关键词出现频次大小呈正比。关键词对应的圆环颜色代表年代信息, 随着年代的递进颜色由浅入深, 从里到外不同颜色圆环的厚度与关键词对应年份的出现频次呈正比。文章主要从频次、中介中心度、突变值反映关键词在共现网络中的重要度(表1)。中介中心度是关键词网络中起到衔接作用的术语, 而突变值则和年度信息有关, 能够反映

基金项目:南京医科大学科技发展基金“基于知识图谱的2002-2012年南京医科大学论文分析”(2012NJMU036), 江苏省教育科学“十二五”规划课题“我国高等医药院校学科知识图谱研究”(C-b/2013/01/028)阶段性成果

收稿日期:2014-10-26

作者简介:陈静(1979-), 女, 江苏南京人, 图书馆馆员, 研究方向为数字图书馆、科学计量。

该术语前后年间的热度变化,即中介中心度较高的关键词在研究中连接了其他高频术语,而突变值高的关键词表明该词在相关年份成为学术热点。

由于研究起始于 2002 年,所有 2002 年有突变值为“-”,其他无突变的关键词相应值为空。

根据图 1 和表 1,可以清晰地看出,2002~2013 年关键词频次高于 50 的有 33 个,其中文章已经将 breast

cancer 和 breast-cancer, lung cancer 和 lung-cancer, polymorphism 和 polymorphisms 进行了同类词合并。在这些关键词中频次较高的关键词,其中介中心度也相对较高,基本都大于 0.1。纵观图 1 和表 1,2002~2013 年南京医科大学的科技发文,主要关键词集中在癌症(cancer, breast-cancer, lung cancer, gastric cancer, carcinoma)、基因(gene-expression, microrna, polymorphism)、

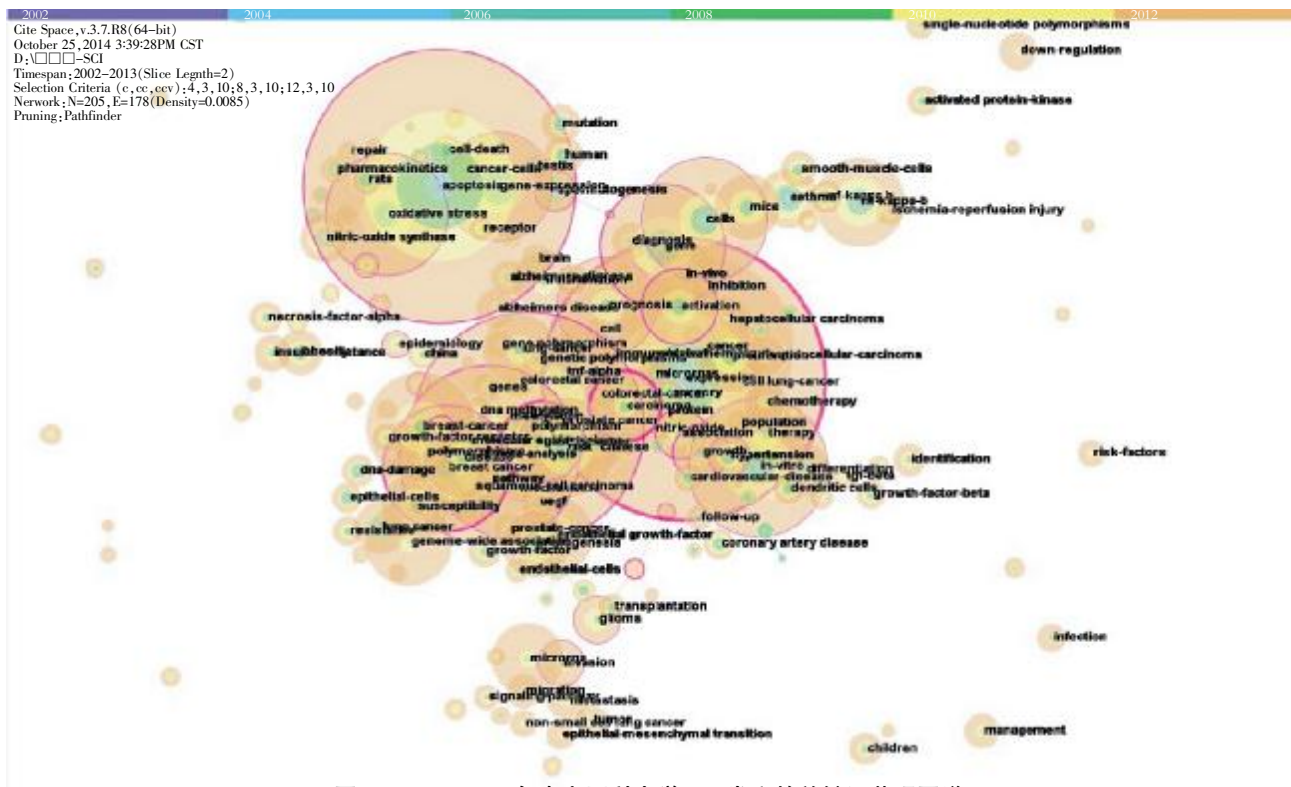


图 1 2002~2013 年南京医科大学 SCI 发文的关键词共现图谱

表 1 2002~2013 年南京医科大学 SCI 发文的热点关键词

关键词	频次	中介中心度	年	突变	关键词	频次	中介中心度	年	突变
expression	289	0.69	2002	-	endothelial growth-factor	83	0.03	2006	
apoptosis	283	0.21	2004		activation	80	0.36	2006	
polymorphism	305	0.22	2006	11.70	proliferation	79	0.05	2006	
breast cancer	224	0.48	2008	5.00	microrna	77	0.02	2010	
meta-analysis	185	0.23	2008		angiogenesis	75	0.05	2008	
in-vitro	146	0.11	2006		squamous-cell	74	0.00	2010	
lung cancer	136	0.03	2006		susceptibility	74	0.08	2008	
gene	131	0.35	2006	7.51	in-vivo	73	0.00	2008	
cells	131	0.13	2006		mice	73	0.10	2008	
oxidative stress	125	0.21	2006		carcinoma	72	0.76	2008	
gene-expression	113	0.18	2004		inflammation	68	0.00	2008	
risk	106	0.06	2006	3.37	asthma	66	0.08	2006	5.98
prognosis	105	0.15	2008		protein	63	0.02	2006	
Nf-Kappa-b	98	0.02	2008		differentiation	58	0.06	2006	5.59
genome-wide association	90	0.02	2010		growth	58	0.35	2010	
disease	88	0.03	2008		Chinese	54	0.03	2010	
gastric cancer	84	0.57	2008						

由于研究起始于 2002 年,所有 2002 年有突变值均为“-”,其他无突变的关键词相应值为空。

细胞 (apoptosis, endothelial growth-factor, proliferation, differentiation)等方面,表明近年来南京医科大学在肿瘤与遗传流行病学、分子细胞学、心血管疾病的早期诊断、遗传机制以及预后的情况分析等领域取得了突出的成绩,并在国际上获得了认可,成为学校科学研究的核心领域。而图 1 和表 1 中出现的 expression、association、mice、Chinese、therapy 等一般性描述词汇表示了专业研究的方法或手段,所以尽管频次、中介中心度较高,但是没有参考价值,不予讨论。同时可以发现突变值较高的 5 个关键词 (polymorphism, lung cancer, gene-expression, asthma, differentiation),突变的时间值多集中在 2006 年,说明这 5 个关键词的词频产生了量变,表明从 2006 年开始南京医科大学的科学研究在这五个方面取得了较大的进展。

(二)南京医科大学作者合作图谱

从作者合作角度计量 2002~2013 年南京医科大学作者合作情况,同时从机构和国家合作角度,计量 2002~2013 年南京医科大学与其他科研机构的合作情况,并在此基础上统计了与相关国家的合作情况(图 2)。需要说明的是,由于 SCI 发文中部分作者信息是以缩写形式出现,在数据处理时文章对这部分数据进行了重新清洗。

图 2 中节点大小与其所代表作者的被引频次呈

正比关系,可以清楚地看到 Hongbing Shen (沈洪兵)、Xinru Wang (王心如)、Zhibin Hu (胡志斌)、Zhengdong Zhang(张正东)等位于图谱的中心位置,而这些学者同时具有较高的中介中心度,基本在 0.1 以上,表明这些学者在南京医科大学的科学研究中占有核心位置,并在其各自的研究领域具有重要影响力。通过进一步的整理合并与优化调整,得到 2002~2013 年被引频次排名前 30 位的学者(表 2)。

根据作者的发文研究可以看出某机构科学研究的主要核心团队,并通过该团队了解机构的主要研究方向和领域。结合表 2 观察图 2,可以看出其中以 Hongbing Shen(沈洪兵)、Xinru Wang(王心如)为高产作者形成了一个较大的簇群,另外还有以 Zhengdong Zhang(张正东)、Jianwei Zhou(周建伟)、Kejiang Cao(曹克将)、Jiahao Sha(沙家豪)等为为核心的较小团体。进一步观察图 2 和表 2,可以得出:①沈洪兵、胡志斌为核心的团队,其成员还包括马红霞等,主要研究领域为慢性病分子流行病学、恶性肿瘤分子与遗传流行病学;②以王心如为核心的团队,其核心成员为夏彦恺,主要研究方向为环境与人类生殖健康;③以韩晓为核心的团队,主要研究分子生物学方面,研究为糖尿病发病机制;④以张正东为核心的团队,主要研究环境遗传与肿瘤发展机制;⑤以沙家豪为核心的团队,主要研究领域为生殖医学;⑥以

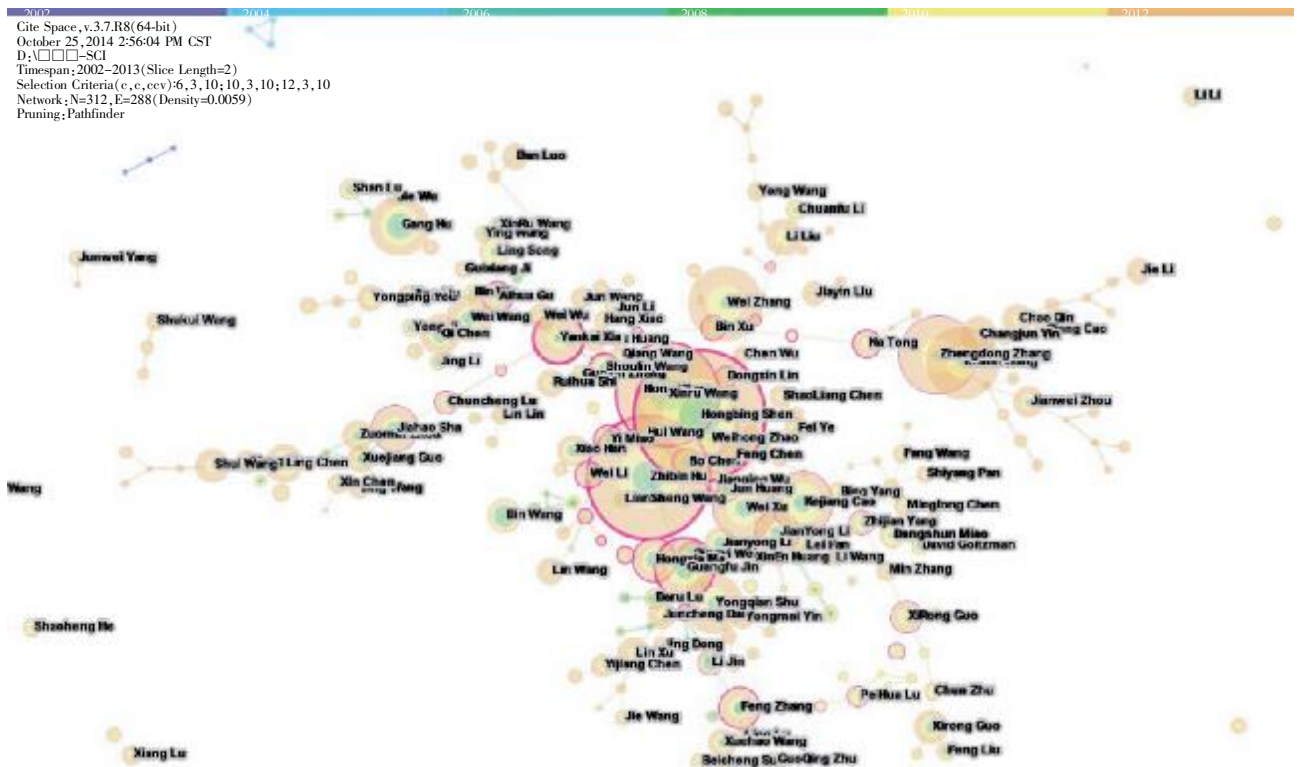


图 2 2002~2013 年南京医科大学 SCI 发文的作者合作图谱

表2 2002~2013年南京医科大学SCI发文的学者统计

作者	频次	中介中心度	作者	频次	中介中心度
Hongbing Shen	215	0.68	Jianwei Zhou	74	0.05
Xinru Wang	200	0.54	JianYong Li	71	0.10
Zhibin Hu	186	0.67	Ling Song	71	0.07
Gang Hu	130	0.04	Hongxia Ma	66	0.37
Jiahao Sha	123	0.15	Feng Zhang	65	0.21
Zhengdong Zhang	121	0.19	Qi Chen	65	0.03
Wei Zhang	119	0.00	Ling Chen	63	0.08
Meilin Wang	117	0.00	Changjun Yin	61	0.08
Kejiang Cao	117	0.14	Li Liu	61	0.01
Wei Xu	98	0.11	Bin Wang	58	0.00
Yongqian Shu	93	0.05	Shui Wang	57	0.04
Guangfu Jin	92	0.47	Feng Chen	56	0.06
Zuomin Zhou	92	0.05	Xirong Guo	55	0.02
Wei Wang	90	0.04	Lin Xu	54	0.05
Yankai Xia	76	0.63	Jianyong Li	54	0.00

曹克将为核心的团队,主要研究方向为心血管疾病。根据图2观察得出的高影响力作者和核心研究团队,结合其研究方向,可以得出南京医科大学的核心科研领域,这与前文根据关键词共现分析得出的主要研究领域的结论基本一致。同时研究图2可以看出合作图谱整体比较紧凑,表明南京医科大学为专业型研究院校,作者的研究领域相对交叉接近,合作关系比较密切;另外根据统计分析,可以发现其中公共卫生学院、基础医学院以及第一临床医学院是学校主要的科技发文机构。

研究发文的合作机构和国家,可以观察到某机构在其主要研究领域的交流频度和国际地位。通过表3和表4的研究,可以分析出2002~2013年国内合作最多的大学为南京大学,频次为451次。与南京医科大学合作的单位机构中,以江苏省内高校较多,这体现了我校在省内医学研究中的科研合作成效;国际间合作最多的为美国,频次达到了696次,表明我校的科研访学及学术交流的主要机构为美国高校和科研机构。通过以上数据,可以表明在2002~

表3 2002~2013年南京医科大学SCI发文的机构合作统计

机构	频次	机构	频次
Nanjing Univ	451	Univ Texas	60
Southeast Univ	263	Nanjing Normal Univ	59
Fudan Univ	141	Tongji Univ	56
China Pharmaceut Univ	136	Zhejiang Univ	49
Shanghai Jiao Tong Univ	132	Xuzhou Med Coll	48
Soochow Univ	125	Capital Med Univ	46
Chinese Acad Sci	94	Sun Yat Sen Univ	42
Jiangsu Univ	87	Peking Univ	39
Nantong Univ	70	Suzhou Univ	37
Chinese Acad Med Sci	61	Univ Hong Kong	36

表4 2002~2013年南京医科大学SCI发文的海外国家合作统计

国家	频次	国家	频次
USA	696	BELGIUM	25
JAPAN	94	AUSTRALIA	22
CANADA	78	SOUTH KOREA	18
ENGLAND	40	TAIWAN	15
GERMANY	34	ITALY	15
SINGAPORE	26	NEW ZEALAND	12

2013年我校与国内外的合作频繁加深,合作的深度和广度都有了相当大的拓展。

一个机构的研究概况和发展,可以通过其发文的关键词、作者、合作机构和地区等文献的统计分析出来。文章通过关键词共现图谱、作者合作图谱对SCI南京医科大学发文进行了可视化分析,可以看出2002~2013年南京医科大学科技论文的发文数量呈现几何级数的增长,直观揭示了我校近年来科研的重点研究方向、作者影响力以及国际合作状况等多重科研趋势脉络。目前只统计分析了关键词及作者,本研究后续将通过文章共引图谱等进一步揭示南京医科大学的科研发展情况,以期对近十年来的科技创新及整体学术实力有一个更加深入、完整的认识。

参考文献

- [1] 陈超美,陈悦,侯剑华,等. CiteSpace II:科学文献中新趋势与新动态的识别与可视化[J]. 情报学报,2009(3):401-421
- [2] 陈悦,刘则渊. 悄然兴起的科学知识图谱[J]. 科学学研究,2005,23(2):149-154
- [3] 刘则渊,王贤文,陈超美. 科学知识图谱方法及其在科技

- 情报中的应用[J]. 数字图书馆论坛, 2009 (10):14-34
- [4] 赵蓉英,王 静. 网络计量学研究热点与前沿的知识图谱分析[J]. 情报学报, 2011(4):424-434
- [5] 赵蓉英,王 菊. 图书馆学知识图谱分析[J]. 中国图书馆学报, 2011(3):40-50
- [6] 王 宁, 盛 立. 生物安全相关文献知识图谱分析[J]. 军事医学科学院院刊, 2010(4):370-372
- [7] 刘垣贞, 包祖军. 2004~2005 年美国 SCIE 数据库收录南京医科大学文献的定量分析 [J]. 南京医科大学学报:社会科学版, 2006, 6(2):167-170
- [8] CitespaceIII 3.7.R7.[EB/OL].[2014-10-24]. <http://cluster.ischool.drexel.edu/~cchen/citespace/download.html>

Visualizing research development of Nanjing Medical University based on SCI(2002-2013)

Chen Jing¹, Xie Jing², Huang HuaXing³

(1. Library of Nanjing Medical University, Nanjing 210029; 2. School of Economics and Management, Nanjing University of Chinese Medicine, Nanjing 210023;3. The First Clinical Medical College of Nanjing Medical University, Nanjing 210029, China)

Abstract: Based on the statistical data in SCI(2002-2013), the paper made visual analysis on the research evolution of Nanjing Medical University. The study showed the research evolution of Nanjing Medical University in the past 12 years from author analysis, research collaboration and research hotspots by mapping knowledge domain. From the point of view of scholars, institutions, keywords, we visually measured scientific research development of Nanjing Medical University.

Key words: research evolution; Nanjing Medical University; knowledge mapping;SCI