

基于 *SCI-E* 的国际医学信息学领域中国科研队伍探析

曹霞

赵颖颖

(中国医科大学附属盛京医院图书馆 沈阳 110004) (中国医科大学图书馆 沈阳 110013)

〔摘要〕 基于 *SCI-E* 运用文献计量学和社会网络分析法对近 5 年国际医学信息学领域的中国科研队伍情况进行分析, 挖掘高产作者、高影响力作者、高频合作团体和合著网络重要节点作者, 对研究结果进行总结并提出相应建议。

〔关键词〕 医学信息学; 中国科研队伍; 文献计量学; 社会网络分析; UCINET

〔中图分类号〕 R-056 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673-6036.2018.04.003

Discussion and Analysis of Chinese Scientific Research Teams in the Field of International Medical Informatics Based on *SCI-E*

CAO Xia, Library of Shengjing Hospital of China Medical University, Shenyang 110004, China; ZHAO Ying-ying, Library of China Medical University, Shenyang 110013, China

〔Abstract〕 By making use of bibliometrics, social network analysis based on *SCI-E*, the paper carries out analysis of Chinese scientific research teams in the field of international medical informatics over the past five years, seeks high yield authors, high impact authors, high frequency cooperative groups and important nodes of co-authorship network, summarizes study results and offers corresponding proposals.

〔Keywords〕 Medical informatics; Chinese scientific research teams; Bibliometrics; Social network analysis; UCINET

1 引言

医学信息学 (Medical Informatics) 最早起源于 20 世纪 50 年代, 于 20 世纪 70 年代开始正式成为一门科学, 是以医学、信息管理学、系统论和计算机科学技术为理论基础, 以医药信息化管理为目标, 涵盖医学信息采集、存储与传输、生物医学信号处理、医学图形和图像处理、医学知识系统和智能系统、计算机仿真和医药信息工程等众多概念和

范畴的一门新兴交叉学科^[1]。随着科学发展的全球化, 越来越普遍, 合著论文成为科学界产出的最大组成部分, 对于汇集新思路、共享新成果、培养复合人才等方面有着十分重要的作用^[2]。在国际医学信息研究领域, 合作研究已成为科研新趋势, 中国科研人员要想在医学信息学领域中占据一席之地, 必须紧跟国际科研大方向, 加强国内外、机构内外和作者之间的合作交流。本文基于 *SCI-E* 运用文献计量学和社会网络分析相关理论及可视化方法了解国际医学信息学领域中国作者发文和合作情况, 揭示作者合著网络图谱结构特性, 挖掘高频合作团体和找出网络关键节点, 旨在鼓励我国医学信息学研究人员在增加科研产出和提高学术影响力的同

〔修回日期〕 2018-04-19

〔作者简介〕 曹霞, 馆员, 发表论文 10 篇。

时,为科研工作者寻求正确的合作途径、合作方式、高水平合作提供一定的参考,扩大知识交流共享,从而提高医学信息科研效率。

2 数据来源及研究方法

2.1 数据来源

本文以《科学引文索引(扩展版)》(*Science Citation Index Expanded, SCI-E*)核心合集为数据来源,检索中国科研人员 2013-2017 年发表的 Web of Science 中类别为"Medical Informatics"论文,具体检索策略为:AD=(Peoples R China) AND PY=(2013 OR 2014 OR 2015 OR 2016 OR 2017) AND WC=(Medical Informatics),共检索获得 1 012 条数据,其中第 1 作者或通讯作者地址为中国的共有 833 条数据,具体检索日期为 2018 年 3 月 15 日。

2.2 研究方法

2.2.1 文献计量学(Bibliometrics) 用数学和统计学等方法研究文献动态特性的一门学科,运用定量或半定量的方法分析研究论文的数量、质量、被引情况以及作者等计量学指标^[3]。文献计量学方法的出现可以回溯到 20 世纪初叶,1917 年动物学教授科尔和博物馆馆长伊尔斯两人合作分析 1543-1860 年《科学进展》杂志中欧洲各国的解剖学文章,这是文献计量学的首次研究。现在文献计量学已成为情报学和文献学领域的一门重要学科分支,同时也展现出突出的方法理论价值^[4]。本文运用文献计量学方法,在普赖斯定律的理论基础上遴选高产、有影响力作者^[5]。

2.2.2 社会网络分析(Social Network Analysis, SNA) 主要是对社会网络中行为者之间的关系进行量化研究,揭示研究对象间的网络结构关系。早期的社会网络分析起源于数学领域的图论和矩阵理论,通过网络描述客观世界起源于 1736 年

德国数学家欧拉使用图论解决哥尼斯堡七桥问题^[6],随着一系列可视化分析软件的研发和升级,社会网络分析方法对结构关系的揭示能力超过以往的任何一种理论和方法^[7],本文运用社会分析软件——UCINET^[8]及其 Netdraw 绘图工具构建作者合著网络,揭示作者合著关系,发现高频合作群体、网络重要作者节点并分析其合作研究内容。

3 结果与分析

3.1 高产作者

运用 Excel 对作者进行去重,统计得出 2013-2017 年以第 1 作者或通讯作者身份发文的中国作者共有 741 人,发表 1 篇论文的作者有 670 位,占作者总数的 90.42%;发表两篇论文的作者有 59 位,占 7.96%;发表 3 篇论文的作者有 9 位,占 1.21%;发表 4 篇论文的作者有两位,占 0.27%;发表 10 篇论文的作者有 1 位,占 0.13%。其中发文最高的前 3 位作者分别是浙江大学的黄正行(10 篇)、北京大学的雷健波(4 篇)和仲恺农业工程学院的蔡肯(4 篇)。根据普赖斯定律,在某一研究领域一半的论文由该领域中所有作者的平方根的这部分作者撰写完成。具体计算公式为: $m=0.749\sqrt{N}$, m 为核心作者中最低产作者发文篇数, N 为最高产作者发文篇数,即核心作者中发文量最低的作者发文篇数,等于发文量最多作者论文篇数的平方根的 0.749 倍,本研究最高产作者论文数为 10 篇,计算得出核心作者群中最低产作者论文数应为 2.37 篇,取最大整数为 3 篇。统计数据表明发文 3 篇及以上的作者只有 12 位,共计发文 45 篇,占作者总数的 1.62%,占论文总数的 5.4%,远低于总论文数的 50%,说明中国医学信息学领域暂未形成稳定的核心作者队伍,对该领域的研究还不成熟完善。

3.2 高被引作者

论文被引频次指某一篇文章被其他文章引用的次数，是衡量作者在某一学科领域学术影响力的重要指标，被引频次越高代表论文质量越高并且受到的关注程度也越大。本文统计的作者被引频次是指第 1 作者发表的所有论文的累积被引总频次。作者整体被引情况，见图 1。其中累积被引频次超过 40 次的高影响力作者情况，见表 1。

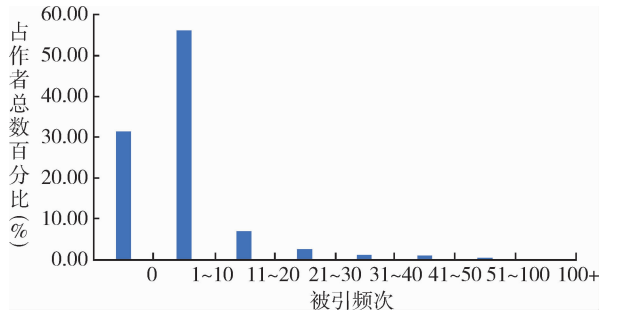


图 1 作者整体被引情况

表 1 累积被引频次 40 次以上作者

序号	作者	单位	地区	被引频次	发文量（篇）	篇均被引
1	黄正行	浙江大学	浙江	117	10	11. 7
2	姜奇	西安电子科技大学	陕西	83	3	27. 67
3	邓朝华	华中科技大学	湖北	61	2	30. 5
4	李智军	华南理工大学	广东	53	2	26. 5
5	叶军	绍兴文理学院	浙江	48	2	24
6	朱斐	苏州大学	江苏	48	1	48
7	李鹏	山东大学	山东	47	3	15. 67
8	温振庶	中科院上海生命科学研究院	上海	42	1	42
9	成娟	中国科学技术大学	安徽	42	1	42
10	CHEN Hao	香港中文大学	香港	41	1	41
11	马昕	山东大学	山东	41	1	41

被引 40 次以上论文，见表 2。邓朝华的文章主要基于价值态度行为模型、计划行为理论和老龄化特征结构的研究模型，分析移动健康服务在中老年人群中的接受度，文章中的移动健康服务也是健康

信息学的一门重要分支。姜奇主要研究远程医疗信息安全隐私问题，提出一种高效保护用户隐私并增强的认证方案，实现用户信息的匿名性和不可追踪性。

表 2 被引 40 次以上论文

序号	作者	机构	标题	被引频次
1	邓朝华	华中科技大学	Comparison of the middle – aged and older users’adoption of mobile health services in China	58
2	姜奇	西安电子科技大学	A Privacy Enhanced Authentication Scheme for Telecare Medical Information Systems	49
3	朱斐	苏州大学	Biomedical text mining and its applications in cancer research	48
4	温振庶	中科院上海生命科学 研究院	An integrated approach to identify causal network modules of complex diseases with applica- tion to colorectal cancer	42
5	成娟	中国科学技术大学	A Framework for Daily Activity Monitoring and Fall Detection Based on Surface Electromyog- raphy and Accelerometer Signals	42
6	CHEN Hao	香港中文大学	Standard Plane Localization in Fetal Ultrasound via Domain Transferred Deep Neural Net- works	41
7	马昕	山东大学	Depth – Based Human Fall Detection via Shape Features and Improved Extreme Learning Ma- chine	41

3.3 作者合作分析

3.3.1 概述 本研究共涉及 3 118 位作者（包括第 2、第 3 作者等所有合著者），篇均 3.74 位作者；合著论文共有 818 篇，整体合著率为 98.2%。由表

3 可知近 5 年论文数呈明显上升趋势，篇均作者数均在 4 人以上，其中 2~5 人的合作形式最为普遍，文献合著率均超过 95%，这说明我国高水平医学信息学论文作者越来越重视科研合作交流，越来越多的研究论文以合著的形式出现。

表 3 医学信息学作者合著情况

年份	1 人		2~5 人		6~10 人		11~15 人		15 人		论文	作者	篇均	合著	合著
	篇	%	篇	%	篇	%	篇	%	篇	%	合计(篇)	合计(人)	作者	论文(篇)	率(%)
2013	2	1.9	74	70.48	28	26.67	1	0.95	0	0	105	432	4.11	103	98.1
2014	4	3.45	69	59.48	41	35.34	2	1.72	0	0	116	524	4.52	112	96.55
2015	5	3.14	106	66.67	39	24.53	7	4.4	2	1.26	159	676	4.25	154	96.86
2016	3	1.51	134	67.34	60	30.15	2	1.01	0	0	199	859	4.32	196	98.49
2017	1	0.39	171	67.32	76	29.92	6	2.36	0	0	254	1168	4.6	253	99.61

3.3.2 作者合著网络分析 利用 R 语言^[9]进行编程，构建 3 118*3 118 规模的作者合著矩阵，将矩阵导入 UCINET 软件中，运用其 Netdraw 程序绘制作者合著网络图谱，见图 2。图左侧有 9 个孤立节点，说明该 9 位作者独著发文，没有与任何作者发生合作。通过 UCINET 相应程序计算每 1 年以及 5 年整体合著网络的密度、聚类系数、平均距离和平均点度中心度 4 个网络指标，对合著网络拓扑性质进行分析，对比近年来医学信息学作者合著网络结构演化，2013–2017 年作者合著网络指标对比，见

表 4。由表 4 可知 5 年来作者合作次数基本上逐年增加，子网络数和最大子网络的节点数规模越来越大。1979 年 Freeman 提出点度中心度概念^[10]，即某节点拥有的连线数量反映该作者节点与网络中其他作者合作的次数，作者点度中心度值越大，说明该作者与其他作者合作次数越多，表 4 中的平均点度中心度是指网络中所有节点的平均度数，反映网络节点的整体合作度^[11]。表 4 中 5 年整体平均点度中心度是 6.458，说明医学信息学领域存在一部分发文较多的高产作者，近 5 年连续有论文发表。

表 4 2013–2017 年作者合著网络指标对比

指标	2013–2017 年	2013 年	2014 年	2015 年	2016 年	2017 年
作者合作总次数	10 094	1 078	1 352	2 271	2 229	3 164
子网络数	347	87	99	119	147	186
最大连通子网络节点数	968	16	16	30	21	77
孤立节点（独著）	9	1	2	4	3	1
平均点度中心度	6.458	4.991	5.141	6.719	5.190	5.409
网络密度	0.002 1	0.011 6	0.009 8	0.010 0	0.006 0	0.004 6
平均路径长度	9.704	1.166	1.123	1.396	1.346	2.249
聚类系数	0.933	0.982	0.984	0.973	0.968	0.969

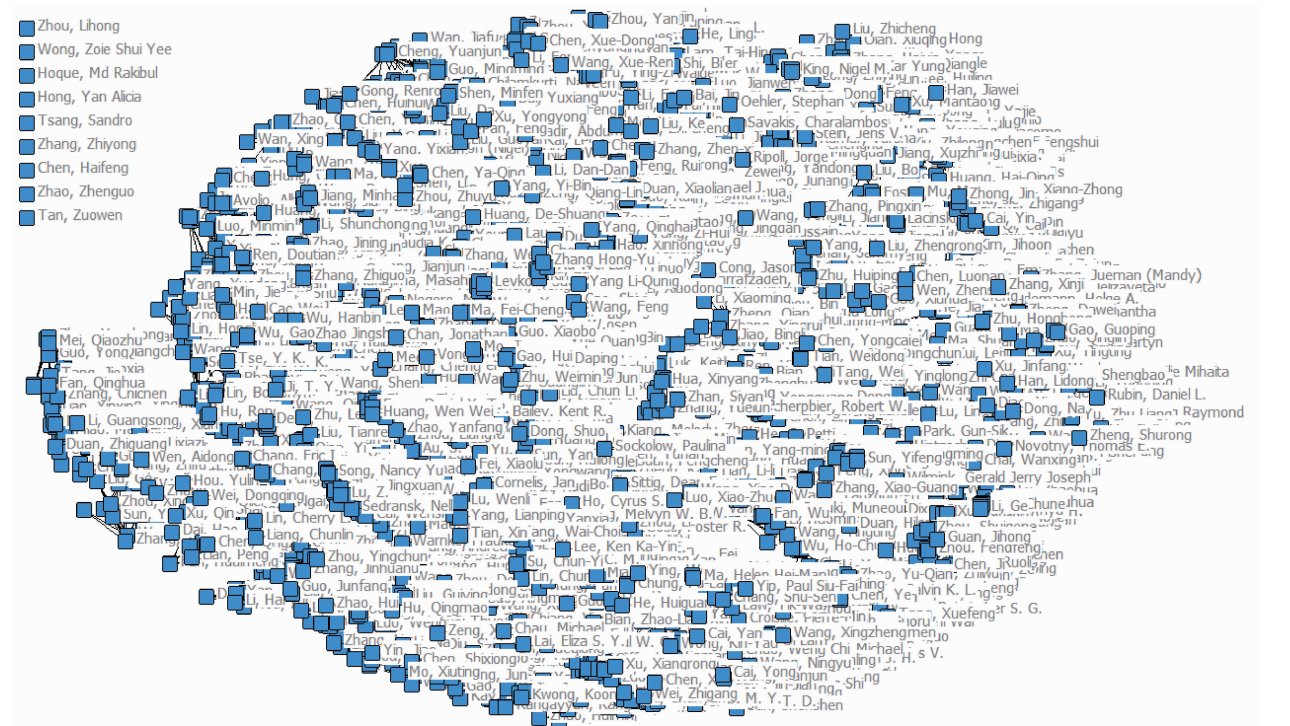


图2 3 118 位作者合著网络图谱

(1) 网络密度是指网络中实际存在的边的数量与理论上可能存在的边的最大值的比例，用于评价网络中节点之间关系的稀疏程度。单年以及5年的网络密度都不到0.1，说明医学信息学领域作者合作网络结构松散不连通，作者间交流不频繁。

(2) 网络的平均距离是指网络中所有节点最短距离的均值，是网络的可达性评价指标。Stanley Milgram提出6度分隔理论，即社会中任意两个人只要通过6个人就能相互认识^[12]。由于计算网络节点间的最短距离针对连通网络，因此在计算作者合著网络平均距离时提取最大连通子网络的平均距离^[13]。2013-2017年平均路径长度明显小于6，任意两位作者只需通过2~3位作者就能取得联系，5年整体网络规模大，平均距离为9.704，说明最大连通子网络比较畅通，随着作者合作频繁，作者之间越来越熟悉，也更容易取得联系。

(3) 聚集系数是指网络局部集团化的程度，是用来确定一个节点与邻居节点组成一个完整图的数量，值越大说明网络的局部凝聚性越强。聚类系数的取值范围在0~1之间，值为0则说明网络全是孤立节点，值为1则说明网络高度连通，任意两节点之间都存在联系^[14]。单年及

5年的聚类系数都在0~1之间，且特别接近1。Duncan Watts 和 Stevenstrogatz 提出小世界网络理论，认为具有聚类系数高和平均距离短特征的复杂网络可以称作小世界网络^[15]。可以看出2013-2017年每年中国科研工作者在医学信息学领域的合作网络具有明显的小世界特征，在连通网络内节点之间结构紧密，作者之间交流合作较多，信息传递速率高。

3.3.3 高频作者合作团体分析 为探寻中国科研队伍在国际医学信息学领域的高频作者合作网络，将3 118位作者的共现矩阵导入UCINET，在Netdraw绘图程序中删除孤立节点，发现其中合作次数5次及以上的团体有5组，见图3。节点大小代表作者合作总次数的多少，连线粗细代表两两作者合作次数的多少，线条之上的数字表示两作者间的具体合作次数。通过研究发现合作次数最多团体是第1组，其中浙江大学的黄正行、段会龙与中国人民解放军总医院的董蔚分别合作10次，是本领域内合作次数最高的前两对作者，该团队主要研究应用概率主题模型来发现临床路径模式和潜在治疗模式。其次是第2组，来自北京大学医学信息中心的张兴

厅、文栋和雷健波团队，分别 7 次，该团队论文主要致力于对穿戴式医疗健康设备的评估、通过社交媒体获取健康信息的调查等。第 4 组是浙江大学的田雨和李劲松，两位作者共计合作 7 次，主要研究基于临床大数据和移动机制的医疗决策支持系统。高频合作作者团队一般来自同一机构，甚至是同一科室，成员间多为同事或师生关系，但机构之间的高频合作比较少见。

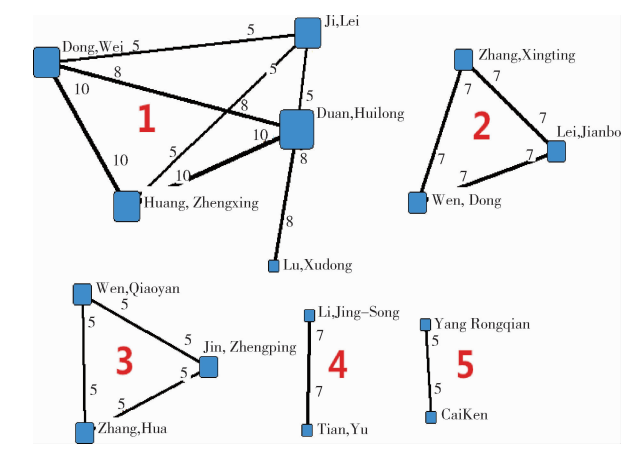


图 3 合作次数 5 次及以上作者团队

3.3.4 重要作者节点分析 网络中心度指标可以评价作者节点在网络中的位置和重要性，其中点度中心度、中间中心度和接近中心度是评价合著网络节点重要性的 3 个最经典指标，点度中心度是某节点拥有的连线数量，度值越大说明合作者越多，该作者越重要。中间中心度是指经过某节点的最短路径的数量，可以衡量该节点控制网络资源的能力，某节点中间中心度越大，说明该节点可控制的信息传递能力越强，该节点也就越处于网络的核心^[16]。接近中心度是指某节点到达网络中其余节点的最短路径的总和，因此接近中心度越小，该节点到达其余节点的距离越短，越处于网络的中心^[17]。通过 Netdraw 程序，综合 3 个中心度指标结果，可找出我国医学信息学核心作者，见图 4，方框内标记的北京大学医学信息中心的雷健波、浙江大学的段会龙和首都儿科研究所的王玮等 8 位作者位于网络的核心位置，通过这些核心作者可以缩短其他作者之间联系的距离，为连接其他作者起到重要的桥梁作用，为提高网络的连通度做出巨大贡献，很好地促进网络信息的交流传播。

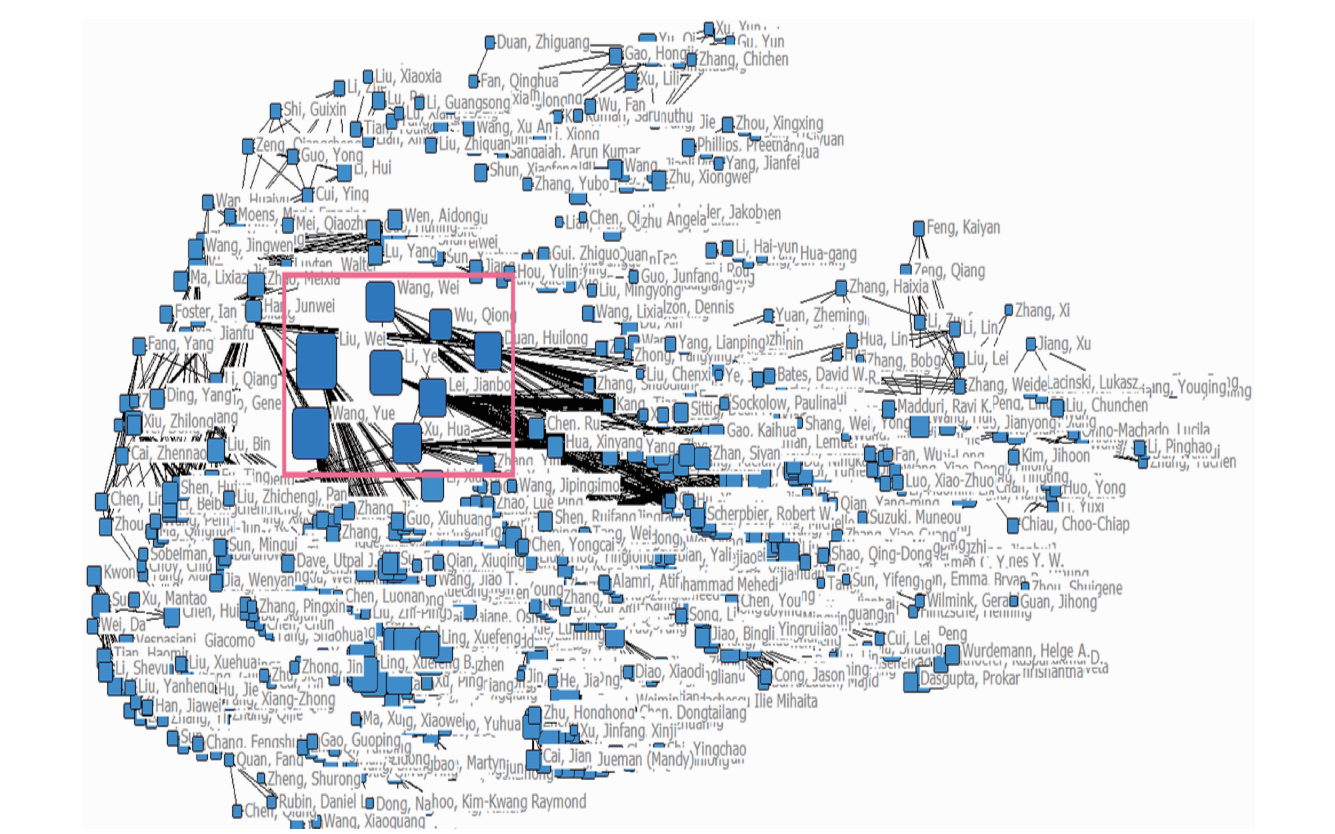


图 4 合著网络核心作者图谱

4 结论与建议

4.1 结论

本文综合运用文献计量学和社会网络分析方法对医学信息学领域的中国科研队伍进行分析,挖掘出高产作者、高影响力作者、高频合作团体和合著网络重要节点作者,研究结果发现在国际医学信息学领域中国科研队伍发文量增长趋势明显,高产和高影响力作者在华东、华北、华中、华南地区分布均匀,但西南地区科研产出稀少,区域合作较少,发文机构主要集中在高等综合院校。整体上中国医学信息学领域高产、高影响力作者和同时具备高中心度、中间中心度和低接近中心度的作者较少,缺乏稳定的高产核心作者群,高频合作团体具备一定的规模,但合作对象、广度和深度有待提高。作者合著率、文献合著率、篇均作者数和作者合作规模逐渐上升,合著子网络内部作者合作频繁,但是子网络之间合作较少,联系不紧密,信息传输渠道不通畅。

4.2 增加医学信息学科研产出,提高国际影响力

近 5 年间中国科研队伍在国际医学信息学领域的发文量增长缓慢,暂未形成稳定成熟的高产核心作者队伍,90.42%的作者仅发表 1 篇论文,一方面说明高产作者偏少,另一方面也说明有一大批新兴作者开始加入到医学信息科学的研究中,因此应积极培养这些新研究者,鼓励其继续深入研究,加大科研产出,壮大中国医学信息研究队伍。高产和高影响力作者地域分布不均衡,应积极推动西南地区的学科发展,形成各地区协同均衡发展良好科研态势,继续提高我国作者在国际医学信息学领域的学术影响力。

4.3 增强内部交流,扩大外部合作

目前国内作者在国际医学信息学领域的合著网络聚类系数高,但网络密度较低,平均距离偏长,即使在最大的连通子网络内,作者之间的交流渠道并不顺畅,说明本领域整体网络结构不紧密,作者

之间合作不频繁。且存在众多子网络,缺乏连接各子网络扩大外部联系的媒介作者,因此医学信息学领域内的中国科研人员在加强与熟悉团队合作的同时,应积极扩大与外界研究者的联系,即国内、国际的有关机构。

4.4 积极推动跨地域、跨学科合作

本领域高频合作者大多数以同事或师生关系进行机构内部合作,缺乏外部合作交流,这不利于学科的长期发展。高产、高影响和网络核心作者多来自高等综合院校。医学信息学科具有医学和信息科学双重特点,为跨学科合作提供良好条件,因此在研究过程中有效加强与医学、信息学和计算机科学等的跨专业、跨学科合作,在巩固科学专业性的基础上注入更多的科研新方法,对促进医学信息学科的发展和提升中国科研队伍国际影响力具有十分重要的作用。

参考文献

- 1 叶明全,吴少云. 信息管理与信息系统(医学)专业课程体系设置[J]. 医学信息学杂志, 2007, 28(5): 524-526.
- 2 林莉. 科研论文合著网络结构与合作关系研究[D]. 长春: 吉林大学, 2010: 1-2.
- 3 若谷. 情报科学术语荟萃[J]. 情报理论与实践, 1987, 10(3): 38-39.
- 4 冯茜,王晓鸿. 文献计量学的过去、现在和未来[J]. 兰州大学学报(自然科学版), 2000, 36(6): 155-161.
- 5 马费成,宋恩梅. 信息管理学基础. 第2版[M]. 武汉: 武汉大学出版社, 2011: 97-99.
- 6 曹霞,崔雷. 合著网络评价指标与文献计量学评价指标相关性研究[J]. 中华医学图书情报杂志, 2016, 25(2): 20-26, 74.
- 7 滕广青,牟冬梅,任晶. 国外社会网络分析在文献计量领域的应用研究[J]. 情报资料工作, 2014, (1): 47-51.
- 8 刘军. 整体网分析讲义—UCINET 软件应用. 第2版[M]. 上海: 格致出版社, 2014: 166.
- 9 高涛,肖楠,陈钢. R 语言实战[M]. 北京: 人民邮电出版社, 2013: 22.
- 10 Freeman U L C. Centrality in Social Networks Conceptual Clarification[J]. Social Networks, 1979, 1(3): 215-239.

- 11 开滨. 近十年来《情报科学》作者合作可视化分析 [J]. 情报科学, 2016, 34 (8): 110-115.
- 12 Guare J. Modern Plays Six Degrees of Separation [M]. New York: Methuen Drama, 2010: 12.
- 13 吕鹏辉, 刘盛博. 学科知识网络实证研究 (IV) 合作网络的结构与特征分析 [J]. 情报学报, 2014, 33 (4): 367-373.
- 14 温磊. 基于复杂网络的供应链建模与仿真研究 [M]. 保定: 河北大学出版社, 2012: 30-33.
- 15 Milgram S. The Small World Problem [J]. Psychology Today, 1967, 1 (1): 61-67.
- 16 Freeman L C. Centrality in Social Networks Conceptual Clarification [J]. Social Network, 1979, 1 (3): 215-239.
- 17 Sabidussi G. The Centrality Index of A Graph [J]. Psychometrika, 1966, 31 (4): 581-603.

2018 年《医学信息学杂志》编辑 出版重点选题计划

2018 年本刊将继续以“学术性、前瞻性、实践性”为特色, 及时追踪并深入报道国内外医学信息学领域前沿热点, 反映学科研究动态, 展示学科应用成果, 引领学科发展方向。现对 2018 年度编辑出版重点选题策划如下:

一、医药卫生体制改革与医药卫生信息化

1 “互联网+”环境下医药卫生发展的新方向、新举措; 2 医药卫生信息化发展规划与战略; 3 信息化助力医疗服务体系、医疗保障体系、公共卫生服务体系建设的技术方案与典型案例; 4 医疗卫生信息相关标准研究与应用; 5 医疗卫生信息化相关法律法规。

二、医学信息技术

1 人工智能在医疗卫生领域的研究与应用; 2 健康医疗大数据的管理、挖掘及应用创新; 3 移动互联网在医疗卫生领域的具体应用及技术实现; 4 精准医学与个性化医疗技术研究与应用; 5 物联网、智慧医疗、远程医疗服务与健康管

三、医学信息研究

1 医学信息学基础理论及方法研究; 2 医学科技创新体系和发展战略; 3 公民健康素养培养及健康促进; 4 医学智库研究与智库服务; 5 医药卫生知识发现技术与实现。

四、医学信息组织与利用

1 “互联网+”环境下医学图书馆的创新举措; 2 人工智能技术及其在医学图书馆中的应用; 3 需求与技术双驱动下的数字资源建设与知识服务; 4 医学数字文献、数据管理与长期保存研究; 5 医学图书馆区域合作及资源共享模式研究。

五、医学信息教育

1 “互联网+”环境下医学信息专科、本科、研究生教育及继续教育面临的挑战、改革与实践创新; 2 医学信息素养教育; 3 国外医学信息学教育的先进理念综述。

(《医学信息学杂志》编辑部)