

# 基于期刊论文的作者合作特征及其 对科研产出的影响 ——以国际医学信息学领域高产作者为例

张 雪<sup>1,2</sup>, 张志强<sup>1</sup>, 陈秀娟<sup>1,2</sup>

(1. 中国科学院成都文献情报中心, 成都 610041;

2. 中国科学院大学经济与管理学院图书情报与档案管理系, 北京 100190)

**摘 要** 为了解科研活动中科研合作与学术产出的特点并提出改进科研合作模式的建议, 使科研工作者在注重合作意识的基础上, 更好地选择科研合作伙伴, 建立高效研究合作网络, 提高学术产出, 促进科学发展, 本文以国际医学信息学领域200位高产作者为例, 通过文献调研, 综合运用社会网络分析、文献计量学、统计学等相关分析方法, 首先对科研产出计量指标如发文量、篇均被引数量、h指数、重要论文数, 以及科研合作计量指标如点度中心度、中介中心性、结构洞、合作能力指数、合作度、合作率、合作系数等分别进行相关分析, 然后采用主成分分析法提取公因子, 最后以新构造的科研产出计量指标为因变量、科研合作计量指标为自变量进行多元统计分析。结果表明, 医学信息学领域科研合作现象越来越普遍, 不同合作模式下, 学者科研产出分布存在差异; 作者固定合作人数增多并不能反映与其合作人群的异质程度高, 故研究者若想提高其科研产出, 可注重在与合作伙伴建立稳定、高频的合作模式的基础上, 努力与不同学科背景的研究者合作, 增加合作伙伴的多样性。

**关键词** 科研合作; 科研产出; 医学信息学; 相关性分析; 因子分析; 回归分析

## Collaborative Features of Authors Based on Academic Journal Papers and Their Influence on Scientific Research Output — Taking Highly Published Authors of International Medical Informatics as an Example

Zhang Xue<sup>1,2</sup>, Zhang Zhiqiang<sup>1</sup> and Chen Xiujuan<sup>1,2</sup>

(1. Chengdu Library and Information Center, Chinese Academy of Sciences, Chengdu 610041; 2. Department of Library, Information and Archives Management, School of economics and management, University of Chinese Academy of Sciences, Beijing 100190)

**Abstract:** In order to realize the characteristics of collaboration and output in scientific research, propose improvements to research collaboration models, increase collaboration awareness, help scientific researchers choose better research partners, establish an efficient research collaboration network, improve academic research output, and promote scientific

收稿日期: 2018-10-17; 修回日期: 2018-11-12

基金项目: 国家社会科学基金重点项目“面向领域知识发现的学科信息学理论与应用研究”(17ATQ008); 中国科学院信息化专项“面向干细胞领域知识发现的科研信息化应用”(XXH13506-203)。

作者简介: 张雪, 女, 1994年生, 博士研究生, 主要研究方向为学科信息学与领域知识发现; 张志强, 男, 1964年生, 博士, 主任, 研究员, 博士生导师, 主要研究方向为科技战略与规划、科技政策与管理、科学学、科学计量学与科技评价等, E-mail: zhangzq@clas.ac.cn; 陈秀娟, 女, 1989年生, 博士研究生, 主要研究方向为科学计量学。

development. This study uses 200 high-yielding authors in the field of international medical informatics as an example to conduct a literature review, and social network, bibliometric, statistical analyses. First, a correlation analysis is carried out to investigate the relationships among measurement indicators of scientific research output, such as the number of articles, number of articles cited, h-index, and number of papers whose citation exceeds a certain eigenvalue  $n$ , and scientific research collaboration indicators such as point degree, intermediary center, structural hole, collaboration ability index, collaboration degree, collaboration rate, and collaboration coefficient. Then, the principal component analysis method is adopted to extract the common factors. Finally, using the newly constructed scientific research output measurement index as the dependent variable and the scientific research collaboration measurement index as the independent variable, a multivariate statistical analysis is undertaken. The results show that the phenomenon of scientific research collaboration in the field of medical informatics is increasingly common. Under different collaboration modes, there are differences in the distribution of scientific research output of scholars. The increase in the number of fixed collaborations between authors does not reflect the high degree of heterogeneity among the cooperating groups. If researchers want to improve scientific research output, they should focus on establishing a stable and high-frequency collaboration model with partners and strive to cooperate with researchers from different disciplines to improve partner diversity.

**Key words:** scientific research collaboration; scientific research output; medical informatics; correlation analysis; factor analysis; regression analysis

## 1 引言

控制论创始人维纳曾说过,爱迪生个人发明创造的时代已经过去了,现在已进入科学合作的时代<sup>[1]</sup>。随着学科交叉广度和深度的不断增强,科学研究对象也变得错综复杂,许多科研任务或技术成果仅靠一人之力是难以完成的。因此,无论是在科学技术领域还是人文社会科学领域,科研活动的合作化和集体化或团队化趋势不断增强,合作研究的比例呈稳步上升趋势<sup>[2]</sup>。2017年7月,国家评估中心和科睿唯安联合发布的《中国国际科技合作现状报告》指出,中国的科研合作中心度从“十一五”时期的全球第10位,提升到“十二五”时期的全球第7位,2个5年间中国科研合作中心度的绝对增幅为1.4,位列包括美国在内的20个国家中的第1位,合作网络关系呈现更为紧密的趋势<sup>[3]</sup>。Beaver等<sup>[4]</sup>在 *Scientometrics* 上撰文指出,对17世纪以后合著关系的研究表明,科学家们持合作态度一般能导致科研成果增多,且能提高科学家的活动范围和名望,也就是说,科研合作的程度与科研产出及科研水平存在“量”与“质”的等比关系<sup>[5]</sup>。国内外许多专家对科学合作及科研产出的关系进行了多方面的研究,从研究层次上来看,科学合作的研究可以分为宏观(国家之间)、中观(科研院所或研发企业之间)以及微观(科学家个人及研究团队之间)等3个层次,其中宏观方面的研究有 Sooryamoorthy<sup>[6]</sup>通过对近年来南非科学家科研产出的被引情况进行分析,得出南非科学家和外国作者撰写的论文比南非科学家自己的合作论文引用次数更高; Abramo

等<sup>[7]</sup>对意大利大学的学者进行统计分析,发现其科研产出质量及数量均与国际合作程度呈正相关关系。中观方面的研究有 Abramo 等<sup>[8]</sup>发现与单一机构合作相比,意大利大学2001—2003年跨机构合作产出平均发表在影响因子较高的期刊上;卢东<sup>[9]</sup>以我国内地图书情报学领域15家高产学术机构为主要分析对象,以香港地区作为参照组分析中国高产学术机构不同科研合著现状,分析发现学术机构国际合著程度越高,其自身的科研质量越高,但机构内的合著程度越高,将不利于提高科研质量。微观方面的研究有邱均平等<sup>[10]</sup>对“图书情报档案学”专业50位高产作者的合作程度与科研产出数量及学术影响力进行 Spearman 相关分析,得出作者合作程度与科研产出数量之间不存在有意义的正相关关系,但作者的合作程度与其科研产出的学术影响力之间存在显著正相关关系;李纲等<sup>[11]</sup>以39个国家自然科学基金创新研究群体为样本,探讨学术带头人的合作特征及其对科研产出的影响,发现学术带头人的合作规模、合作强度与科研产出呈正相关关系,合作程度与科研产出呈负相关关系,合作稳定性和署名模式则对科研产出无显著影响。针对某一研究领域的科研合作与产出文献也不尽少数,例如,田江等<sup>[12]</sup>以供应链为例,探究科研成果与合作的关联性,得出供应链子网中网络密度和科研产出数量呈现显著负相关,以及合作率与科研产出质量也呈现出负相关;李文聪等<sup>[13]</sup>分析了2000—2014年干细胞研究领域国家间科研合作网络的结构与演化特征,发现干细胞领域国际科研合作规模不断扩大,一张高度连通世界各国的科研合作网络已基本形成;

He等<sup>[14]</sup>通过对新西兰大学65名生物医学科学家进行调研,得出在论文层面,大学内部合作和国际合作均与文章质量呈正相关关系,而在科学家层面,只有国际合作与科学家未来的研究成果呈正相关关系;Glänzel等<sup>[15]</sup>分析了生命科学领域30个国家的科研产出情况,发现跨国发文通常比同一国家的科研产出论文具有更多的被引量。

纵观已有相关研究文献,对于作者合作行为的研究,国内外都主要是宏观层次、数量上的分析,对于微观层次、个体层次的合作研究很少<sup>[16]</sup>,大多只反映了科研合作与科研产出的相关关系,合作特征对科研产出产生的影响并没有进一步分析,且合作与产出指标选择单一。本文主要探讨微观层面的科研合作与学术产出的关系,以2008—2017年国际医学信息学10种高影响力英文期刊为例,采用普赖斯定律确定高产作者人数共200位,分别统计各项科研产出及合作指标,分析各项指标相关关系,采用因子分析分别提取科研产出及合作的主成分,以新构造的科研产出指标为因变量、科研合作指标为自变量,采用多元线性回归方法进一步分析医学信息学领域高产作者的各项合作特征对其科研产出的影响。

## 2 数据来源、测度指标与方法

### 2.1 数据来源

#### 2.1.1 数据获取

由美国科学情报研究所(Institute for Scientific Information, ISI)编辑出版的JCR(Journal Citation Reports)是在多年积累的SCI数据库的基础上,专门对期刊的引证与被引证关系进行系统归类、整理、分析,并通过计算机自动处理编辑而成,是国

际上公认的权威期刊评价工具<sup>[17]</sup>。在JCR 2016年版中限定学科为Medical Informatics,共检索到24种医学信息学相关期刊,其影响因子的平均值为2.29,通过阅读期刊简介和相关文献,剔除相关度较低期刊及影响因子低于平均值期刊14种,保留10种核心期刊(表1)作为本研究文献数据集来源。在Web of Science核心合集中,以10种期刊刊名为检索词,检索字段之间关系为“OR”,时间限定为2008—2017,文献类型为Article、Proceedings Papers、Review共3种,共检索得到11823篇文献。

#### 2.1.2 高产作者筛选

经统计,本论文涉及的11823篇医学信息学领域论文共出现38939位不同作者,如对每个作者进行统计分析,不仅数据量大不易获取,且分析意义不大,无法凸显核心作者。普赖斯<sup>[18]</sup>指出,某一领域全部科学家总人数开平方,所得到的人数撰写了全部科学论文的50%,根据普赖斯定律,前197位作者应该为高产作者,平均发文为11篇左右,本文选定发文在11篇以上的作者共200位作为分析对象,下文每位作者的合作及论文产出计量指标均以近十年的发文为统计基础。

## 2.2 科研合作与科研产出计量指标

### 2.2.1 科研合作计量指标

朱丽娟等<sup>[19]</sup>建议计量科研合作可将合作者的数量、合作者的质量和作者间合作关系的质量均加以考虑,本文在调研相关文献的基础上,对这3方面因素进行分析,共选取7个指标作为学者合作的计量指标,这7个指标的内涵及计算方法如下:

(1)点度中心度:指网络中节点与其他有联系节点的总数目,可以准确衡量科研团队中与某个作者存在

表1 医学信息学领域10种高影响力英文期刊

序号	刊名	ISSN	出版国	IF值	总被引频次
1	Journal of Medical Internet Research	1438-8871	Canada	5.175	8927
2	JMIR mHealth and uHealth	2291-5222	Canada	4.636	825
3	Journal of The American Medical Informatics Association	1067-5027	England	3.698	7655
4	IEEE Journal of Biomedical and Health Informatics	2168-2194	USA	3.451	1933
5	International Journal of Medical Informatics	1386-5056	Ireland	3.21	4069
6	Health Informatics Journal	1460-4582	England	3.021	536
7	Journal of Biomedical Informatics	1532-0464	USA	2.753	4886
8	Computer Methods and Programs in Biomedicine	0169-2607	Netherlands	2.503	4565
9	Journal of Medical Systems	0148-5598	USA	2.456	3239
10	Medical Decision Making	0272-989X	USA	2.362	4282

合作关系的其他作者的数量,计算公式为

$$C_D(n_i) = d(n_i) = \sum_j X_{ij}$$

式中,  $X_{ij}$  表示节点  $n_i$  与  $n_j$  之间的连接数目<sup>[20]</sup>, 点度中心度越大, 说明该作者与他人的联系越紧密, 使用 UCINET 中 Network-Centrality and Power-Degree 可获得。

(2) 中介中心性: 指网络中所有节点对的最短路径之中经过该节点的数量, 即该节点在所有的接地线中有多少条是处于中间位置, 反映了某作者在其他两作者合作中所起的桥梁作用, 计算公式为

$$C_B = \sum_{j < k} g_{jk}(n_i) / g_{jk}$$

式中,  $g_{jk}$  表示点  $j$  和  $k$  之间存在的 shortest path 数目,  $g_{jk}(n)$  表示包含行动者  $n_i$  的 2 个行动者之间的 shortest path 数目<sup>[20]</sup>, 中介中心性越大, 说明该作者对信息的来源控制力越强, 中介中心性较大的点去掉之后, 网络会分裂成孤立的 2 个, 使用 UCINET 中 Network-Centrality and Power-Freeman-Betweenness-Node Betweenness 可获得。

(3) 结构洞: Burt 指出, “非冗余的联系人被结构洞所连接, 一个结构洞是两个行动者之间的非冗余的联系”, 计算公式为

$$P_{ij} = (a_{ij} + a_{ji}) / \sum_k (a_{ik} + a_{ki})$$

式中,  $a_{ij}$  是指  $i$  和  $j$  两点间的边的属性值 (权重),  $p_{ij}$  表示  $i$  与  $j$  联系的强度。结构洞能为其占据者获取 “信息优势” 和 “控制优势” 提供机会, 从而比网络中其他位置上的成员更具有竞争优势<sup>[21]</sup>, 使用 UCINET 中 Network-Ego Networks -Structural Holes 可获得。

(4) 合作能力指数: 指作者所拥有的稳定合作关系的合作作者数, 将某一作者的合作者按照其与该作者的合作频次降序排列, 合作者序号首次大于其与该作者的合作频次时, 该序号减 1 为该作者的合作能力指数<sup>[22-23]</sup>。

(5) 合作度: 指某一作者所发表的所有论文中的作者总数与某一作者所发表的论文总数之比<sup>[10]</sup>, 通过 Web of Science 分析检索结果功能统计可得。

(6) 合作率: 指某一作者所发表的所有论文中的合作论文数与某一作者所发表的论文总数之比, 反映了某一种期刊或者某一学科的论文作者的合作智能发挥程度, 数值越高, 合作智能发挥越充分<sup>[10]</sup>。

(7) 合作系数: 由于合作度和合作率指标不够全面, 1988 年西安大略大学图书馆与情报学学院 Ajiferuke 等<sup>[24]</sup> 对合作度和合作率指标进行了比较研

究, 并在此基础上提出了合作系数, 指作者所发表论文中合作者的产出贡献比例, 计算公式为

$$CC = 1 - \left( \sum_{j=1}^q f_j / j \right) / N$$

式中,  $f_j$  为作者数为  $j$  的论文数量,  $N$  为作者发表论文总量<sup>[24]</sup>。

### 2.2.2 科研产出计量指标

本文以发文量作为科技成果 “量” 的量化, 篇均被引数量、h 指数、重要论文数作为科技成果 “质” 的量化, 这 4 个指标的内涵及计算方法如下:

(1) 发文量: 指每个作者所发表的论文总数, 在 Web of Science 中以 “作者姓名” AND (“目前工作单位 OR 以前工作单位”) 为检索式进行检索可得每个作者发表论文数。

(2) 篇均被引数量: 是每个作者被引数量与发文量之比, 被引数量在 Web of Science 中根据引文分析报告可得, 一般认为被引量越高, 作者所发表的论文质量越高。

(3) h 指数: h 指数等于作者发表了 h 篇文章且每篇至少被引 h 次的论文数量。同样首先获得每个作者发文情况, 然后根据引文分析报告可得每个作者 h 指数。

(4) 重要论文数: 指引文数超过某一特征值  $n$  的论文的数量,  $n$  的选取是任意的, 主观性较强<sup>[25]</sup>, 本文将  $n$  值定义为篇均被引量, 故每个作者重要论文数即其所有文章中被引量大于等于其篇均被引数量的个数。

## 2.3 研究方法

本文主要研究作者科研合作及产出指标的相关关系及合作特征对产出的影响, 故首先利用文献计量、社会网络分析方法对作者合作及产出特征、作者合作网络特征的各项指标进行统计; 其次, 使用 Spearman 相关分析对合作及产出各项指标进行分析, 因科研合作指标、科研产出指标之间可能存在相关性, 而因子分析法的基本目标是将比较密切相关的几个变量归类, 每 1 类变量构成 1 个因子, 以较少的几个因子来反映初始数据的大部分信息<sup>[26]</sup>, 故使用主成分分析法提取潜在的公因子进行回归分析; 最后, 以新提取的科研产出因子为因变量, 科研合作因子为自变量, 运用多元回归线性方法分析科研合作特征对产出的影响, 所使用的分析软件为 SPSS 23.0。

### 3 多元线性回归分析指标确立

#### 3.1 科研合作计量指标体系构建

为能够恰当地解释多元线性回归模型的系数, 各个指标之间应相互独立, 故指标之间是否存在多重共线性是需要首先考虑的问题, 论文一共统计了 7 个科研合作计量指标, 下文首先采用 KMO 检验和 Bartlett 检验、Spearman 相关系数矩阵检查指标之间是否存在共线性, 然后采用主成分分析法提取公因子, 在此基础上建立科研合作计量指标体系。

(1) KMO 检验和 Bartlett 检验、Spearman 相关分析。KMO 值越接近 1, Bartlett 球度检验的 Sig 值越小于显著水平 0.05, 意味着变量间的相关性越强, 原有变量越适合作因子分析, 经统计, KMO 值为 0.657, 显著性为 0.000, 说明这 7 个变量之间相关性较强, 符合因子分析的条件; 相关分析是分析客观事物之间关系的数量分析方法, 采用 SPSS 计算 7 个变量之间的相关系数, 结果如表 2 所示。

从表 2 可看出: ①除合作度与合作系数外, 合作率与其他指标之间不存在相关关系 ( $P>0.05$ ), 这是因为在统计数据时, 发现大多数作者的合作率均为 1, 即作者大多以合作形式发文, 虽合作率与合作度、合作系数存在正相关关系, 但关联性也较弱, 为避免过多 1 造成误差, 后续分析摘除合作率; ②合作能力指数与中介中心性、结构洞之间均不存在相关关系 ( $P>0.05$ ), 合作能力指数反映某作者所拥有的稳定合作伙伴数目, 中介中心性与结

构洞均反映合作的多样性, 即作者合作网络异质性程度, 说明随着作者固定合作人数增多, 固定合作人群异质性可能会增大 (与更多不同机构作者合作), 也可能会减小 (加强与同一机构作者合作)。除此之外其他指标两两之间均具有相关性, 其中, 合作度与合作系数的相关性最强, 为 0.857, 即随着作者合作人数的增多, 合作者在论文产出中所占贡献比例也越高。

(2) 因子分析。根据以上分析可知, 7 个指标之间存在信息冗余, 为满足下文回归分析各变量之间独立性的要求需要采用因子分析降低变量维度、去除信息冗余, 构建作者科研合作计量指标体系。根据碎石图及“特征值>1 的因子保留, 特征值<1 的因子舍弃”的原则, 且“提取的因子累计百分比应达到 60% 以上”<sup>[27]</sup>, 有 3 个主成分被提取, 假设点度中心度~合作系数分别为 X1~X6, 新提取的因子分别为 F1~F3, 由表 3 可知,  $F1=0.713 \times X1+0.887 \times X2-0.036 \times X3+0.220 \times X4+0.084 \times X5-0.801 \times X6$ , F2、F3 类似。其中成分 1 主要解释点度中心度、中介中心性、结构洞, 方差贡献百分比为 33.22%, 反映了作者合作人数的多少以及合作网络的异质性程度; 成分 2 主要解释合作度、合作系数, 方差贡献百分比为 31.33%, 反映了作者合作的广度与深度; 成分 3 主要解释合作能力指数, 方差贡献百分比为 21.56%, 反映了作者所建立的稳定合作关系。根据每个成分的特点, 分别将其命名为合作多样性、合作程度与强度、合作稳定性。

表 2 科研合作计量指标的相关关系分析

		点度中心度	中介中心性	结构洞	合作能力指数	合作度	合作率	合作系数
点度中心度	相关系数	1.000	0.615**	-0.842**	0.401**	0.670**	0.050	0.524**
	显著性(双尾)		0.000	0.000	0.000	0.000	0.358	0.000
中介中心性	相关系数	0.615**	1.000	-0.719**	-0.045	0.304**	-0.074	0.146*
	显著性(双尾)	0.000		0.000	0.529	0.000	0.296	0.039
结构洞	相关系数	-0.824**	-0.719**	1.000	-0.021	-0.452**	0.040	-0.248**
	显著性(双尾)	0.000	0.000		0.770	0.000	0.575	0.000
合作能力指数	相关系数	0.401**	-0.045	-0.021	1.000	0.469**	0.131	0.532**
	显著性(双尾)	0.000	0.529	0.770		0.000	0.064	0.000
合作度	相关系数	0.670*	0.304**	-0.452**	0.469**	1.000	0.170*	0.857**
	显著性(双尾)	0.000	0.000	0.000	0.000		0.016	0.000
合作率	相关系数	0.050	-0.074	0.040	0.131	0.170*	1.000	0.383**
	显著性(双尾)	0.358	0.296	0.575	0.064	0.016		0.000
合作系数	相关系数	0.524**	0.146*	-0.248**	0.532**	0.857**	0.383**	1.000
	显著性(双尾)	0.000	0.039	0.000	0.000	0.000	0.000	

注: \*\*表示在显著性水平为 0.01 时, 变量之间线性关系显著; \*表示在显著性水平为 0.05 时, 变量之间线性关系显著。

表3 医学信息学领域文献合作计量指标旋转因子成分矩阵

指标	成分		
	1	2	3
点度中心度	<u>0.713</u>	0.227	0.576
中介中心性	<u>0.887</u>	-0.033	0.017
合作能力指数	-0.036	0.328	<u>0.916</u>
合作度	0.220	<u>0.881</u>	0.241
合作系数	0.084	<u>0.901</u>	0.245
结构洞	<u>-0.801</u>	-0.363	0.070

注:提取方法为主成分分析法,旋转方法为凯撒正态化最大方差法。旋转在5次迭代后已收敛。

### 3.2 科研产出计量指标确立

与科研合作计量指标确定类似,首先采用KMO检验和Bartlett检验、Spearman相关分析检验科研产出计量指标是否存在信息冗余,其中KMO值为0.608, Bartlett检验显著性为0.000,意味着变量间相关性较强, Spearman相关分析如表4所示,其中篇均被引数量与重要论文数量不存在相关关系( $P>0.05$ ),重要论文是被引量高于篇均被引数量的论文个数,变化趋势并不因篇均被引数量的高低而呈现规律变化,说明医学信息学领域作者发文质量参

表4 科研产出计量指标的相关关系分析

		发文量	篇均被引数量	h指数	重要论文数量
发文量	相关系数	1.000	0.174*	0.673**	0.645**
	显著性(双尾)		0.014	0.000	0.000
篇均被引数量	相关系数	0.174*	1.000	0.634**	-0.082
	显著性(双尾)	0.014		0.000	0.248
h指数	相关系数	0.673**	0.634**	1.000	0.526**
	显著性(双尾)	0.000	0.000		0.000
重要论文数量	相关系数	0.645**	-0.082	0.526**	1.000
	显著性(双尾)	0.000	0.248	0.000	

注:\*\*表示在显著性水平为0.01时,变量之间线性关系显著;\*表示在显著性水平为0.05时,变量之间线性关系显著。

表5 科研合作计量指标与产出的相关分析

		点度中心度	中介中心性	结构洞	合作能力指数	合作度	合作系数
发文量	相关系数	0.585**	0.262**	-0.339**	0.516**	0.117	0.146*
	显著性(双尾)	0.000	0.000	0.000	0.000	0.098	0.039
篇均被引数量	相关系数	0.229**	0.064	-0.204**	0.178*	0.117	0.096
	显著性(双尾)	0.001	0.372	0.004	0.012	0.098	0.178
h指数	相关系数	0.521**	0.218**	-0.358**	0.494**	0.204**	0.226**
	显著性(双尾)	0.000	0.002	0.000	0.000	0.004	0.001
重要论文数量	相关系数	0.434**	0.200**	-0.254**	0.364**	0.145*	0.168*
	显著性(双尾)	0.000	0.005	0.000	0.000	0.041	0.017

注:\*\*表示在显著性水平为0.01时,变量之间线性关系显著;\*表示在显著性水平为0.05时,变量之间线性关系显著。

差不齐,有的可能只有几篇被引量较高,有的质量均衡,故重要论文数趋势也不尽相同,其余各个指标之间均存在相关关系( $P<0.05$ ),其中发文量与h指数的相关性最强,为0.673。科研产出计量指标之间的相关关系分析说明各指标之间存在重叠,需要采用主成分分析进一步提取公因子,以便消除冗余,经分析,一共提取一个因子,方差贡献百分比为66.30%,新提取的因子 $Y=0.947\times\text{发文量}+0.259\times\text{篇均被引数量}+0.931\times\text{h指数}++0.906\times\text{重要论文数量}$ 。

## 4 科研合作与科研产出的关系分析

### 4.1 相关分析

使用SPSS的Spearman相关分析对以上7个科研合作计量指标,4个科研产出计量指标进行分析,结果如表5所示,可得:①发文量、篇均被引数量、重要论文数量与合作度、合作系数之间均无相关关系( $P>0.01$ ),合作度和合作系数是合作程度指标,考量了合作者的多少及其在合著论文中贡献比例大小,除h指数外,其余科研产出指标与其无相关关系,说明某作者合作人数增多及合作者在科研产出中贡献比例高并不一定会提高科研产出整体的“质”与“量”,还需综合考量合作者的学术水平与地位;②篇均被引数量与中介中心性之间无相关关系( $P>0.05$ ),中介中心性是作者合作多样性指标,反映了作者学术圈、人脉圈的宽广,作者的中介中心性不影响篇均被引数量,这可能是由署名模式及其他合作者的质量导致;③科研产出计量指标与点度中心度、结构洞之间均存在相关关系,说明科研合作多样性程度高,不仅有利于科研产出量的增多,也会有助于科研产出质的提升;④h指数是一个综合评价指标,对科研产出的考察较为全

面, 既可评估科研人员的科研产出数量, 又可评价其科研产出质量, 它与所有科研合作计量指标均存在相关关系, 说明作者合作多样性程度高, 合作稳定性强, 合作程度大可以提高高影响力文章的数量。

### 4.2 回归分析

采用多元线性回归的方法, 以提取的科研产出因子 Y 为因变量, 以提取的科研合作因子 F1、F2、F3 为自变量, 分析这些自变量对因变量的影响是否显著。在 SPSS23 输出的结果中 F 值为 201.679, 通过了显著性检验,  $R^2$  为 0.755, 说明模型的拟合效果较好; 回归系数检验方面, 合作多样性、合作程度与强度、合作稳定性的 Sig 均小于 0.05, 说明这 3 个自变量对科研产出均具有影响; 此外, 删除的学生化残差、杠杆值、库克距离、DW 统计量以及 VIF 值均通过检验, 表明此模型不存在异常值及强影响点, 变量之间也不存在自相关及多重共线性。综上所述, 模型的拟合效果较好, 结合表 6, 回归方程为:  $Y=0.401 \times F1-0.081 \times F2+0.767 \times F3-8.312 \times 10^{-16}$ , 根据上文中 F1、F2、F3 主要解释的变量可知, 科研产出与合作多样性、合作稳定性存在正相关关系, 与合作程度与强度存在负相关关系, 其中与合作稳定性之间相关性最强。为了进一步分析合作多样性、合作稳定性、合作程度与强度与科研产出的关系, 下文以科研产出因子 Y 为因变量, 分别对 F1、F2、F3 因子中各项代表性指标与科研产出的关系进行评估。

表 6 科研合作与产出回归分析结果

模型	未标准化系数		标准化系数	t	显著性
	B	标准误差	Beta		
(常量)	-8.312×10 <sup>-16</sup>	0.035		0.000	1.000
1 X1	0.401	0.035	0.401	11.339	0.000
X2	-0.081	0.035	-0.081	-2.286	0.023
X3	0.767	0.035	0.767	21.708	0.000

#### 4.2.1 合作多样性对科研产出的影响

表 7 是三个合作多样性指标分别与科研产出的回归结果, 从表可看出, 三个指标与科研产出的模型均通过显著性检验, 但  $R^2$  值 (下同) 整体较小, 这是由于我们选择单一指标对因变量做回归, 解释变量解释能力不足。点度中心度、中介中心性与科

研产出有显著正影响, 结构洞与科研产出有显著负影响 (本文结构洞使用限制度 constraint 来衡量, 越小表示该节点在网络中越重要), 且结构洞的回归系数最大, 为 4.184, 表示某作者在网络中限制度每减少一个单位, 该作者的论文产出增加 4.184 次, 点度中心度的回归系数最小, 说明某作者仅仅合作规模大对科研产出并不会产生很大的影响, 若合作规模扩大且合作者之间互不认识则将有效提高该作者的科研产出。

表 7 合作多样性与科研产出回归分析结果

	点度中心度	中介中心性	结构洞
回归系数	0.021	0.104	-4.184
$R^2$	0.629	0.091	0.100
Sig	0.000	0.000	0.000
常量	-1.240	-0.302	0.506

#### 4.2.2 合作稳定性对科研产出的影响

表 8 测度了合作稳定性与科研产出的影响程度, 因前文中合作稳定性的主因子为合作能力指数, 故此主要分析合作能力指数与科研产出之间的回归关系。经过 F 检验, 模型显著, 合作能力指数系数为 0.497, 表明某作者每增加一个固定合作人数, 该作者的科研产出增加 0.497 次。作者合作稳定性越高, 网络的抗毁性就越强, 但是合作网络也会存在一定风险, 比如, 当某位团队成员退出科研团队时, 其他团队成员间的合作距离就会增大, 甚至导致合作关系的中断<sup>[11]</sup>。

表 8 合作稳定性与科研产出回归分析结果

	合作能力指数
回归系数	0.497
$R^2$	0.352
Sig	0.000
常量	-1.716

#### 4.2.3 合作程度与强度对科研产出的影响

表 9 给出了合作度、合作系数对科研产出的影响情况, 合作度与科研产出之间的关系未通过显著性检验, 这也与上文  $\alpha=0.05$  水平下, 发文量、篇均被引数量与合作度之间无相关关系吻合, 说明作者合作程度大小对科研产出无显著影响, 合作系数对科研产出有显著正影响, 表明合作系数每增加 1 个, 科研产出增加 2.761 次, 但与上文整体回归模型中合作程度与强度与科研产出存在负相关关系违背, 这可能是由于合作度单一指标对科研产出无显

著影响，而整体回归主要是将其与合作系数进行加权一起对科研产出做回归的缘故。

表9 合作程度与强度与科研产出回归分析结果

	合作度	合作系数
回归系数	0.069	2.761
$R^2$	0.018	0.037
Sig	0.057	0.007
常量	-0.442	-2.184

## 5 结论与展望

本文以国际医学信息学领域的200位高产作者为例，首先采用KMO检验和Bartlett检验、Spearman相关系数矩阵分析了科研论文产出、科研合作各指标之间的相关性，确定科研合作与产出指标之间存在信息冗余，因此需要因子分析提取公因子，构建科研合作与产出指标体系以进一步分析它们之间的关系；其次采用主成分分析提取4个科研产出计量指标、6个科研合作计量指标中的公因子；最后在分析科研产出与科研合作的相关关系基础上，采用多元统计分析，论证了合作指标对科研产出的影响。本文得出的结论有：①所收集的样本中92%学者的合作率均为1，说明科研合作现象越来越普遍，学者要想有好的学术成果，必须培养和加强自己的科研合作意识。②合作能力指数与中介中心性、结构洞之间均不存在相关关系，即作者固定合作人数增多并不代表这些合作者的异质性也会提高，多数情况下相同学术背景内部成员同质性高于异质性。③篇均被引数量与重要论文数量不存在相关关系，说明医学信息学领域作者发文质量参差不齐，有的作者可能只有几篇被引量较高，有的作者则整体质量均衡。④合作程度与深度指标合作度、合作系数与发文量、篇均被引数量、重要论文数量之间均无相关关系，说明合著文章中篇均作者数多并不会显著提高科研产出，还需综合考虑合作者的学术地位与水平。⑤点度中心度、结构洞与所有科研产出计量指标均存在相关关系，故科研人员在扩大其合作者规模的基础上积极与不同专业背景的研究者进行合作，可以提高合作网络的异质性。h指数与所有科研合作计量指标均存在相关关系，它是测度科研产出的一个较为全面的指标，所以科研人员应在增加其产出数量的同时提高论文质量。⑥若某作者想提高其科研产出数量，可注重在与合作伙伴建立稳定、高频的合作模式的基础上，提高合作

伙伴的多样性，主动与不同学术背景的学者合作，可以促进其发文“质”与“量”的提高。

本文是基于数据统计分析基础上得到的结论，但在数据样本预处理、指标的选择等方面难免存在不足，因此也是需要进一步完善研究的。首先，是以医学信息学领域国际10种高影响力英文期刊发文为数据源，根据普赖斯定律统计出该领域的前200位高产作者。这样有可能不仅样本量较少，而且限定了学科领域，统计结果不一定可以外推到其他学科领域。其次，指标选择主要是通过文献调研确定的，可能会遗漏部分统计指标。最后，本文在统计样本数据时忽略了科研产出相对于科研合作的时间滞后性，因此在接下来的研究中需要针对上述问题做进一步的探究。

## 参 考 文 献

- [1] 张碧辉,王平.科学社会学[M].第1版.北京:人民出版社,1990:246-258.
- [2] 郑海燕.SSCI和A&HCI收录中国人文社会科学合著论文统计分析(1995~2004年)[J].社会科学管理与评论,2007(4):47-55.
- [3] National Center for Science and Technology Evaluation, Clarivate Analytics. China's International scientific research collaboration—A bibliometric analysis[R]. China: National Center for Science and Technology Evaluation, Clarivate Analytics, 2018:1-69.
- [4] Beaver D D, Rosen R. Studies in scientific collaboration[J]. Scientometrics, 1978, 1(1): 65-84.
- [5] 李霖,贺凤兰.从期刊论文的合著现象看高水平研究成果的产出[J].图书馆工作与研究,2007(2):26-28.
- [6] Sooryamoorthy R. Do types of collaboration change citation? Collaboration and citation patterns of South African science publications[J]. Scientometrics, 2009, 81(1): 177.
- [7] Abramo G, D'Angelo C A, Solazzi M. The relationship between scientists' research performance and the degree of internationalization of their research[J]. Scientometrics, 2011, 86(3): 629-643.
- [8] Abramo G, D'Angelo C A, Costa F D, et al. University - industry collaboration in Italy: A bibliometric examination[J]. Technovation, 2009, 29(6): 498-507.
- [9] 卢东.图书情报学领域作者合著方式对科研质量的影响分析——基于SciVal对中国高产机构的样本分析[J].现代情报,2015,35(9):111-115.
- [10] 邱均平,温芳芳.作者合作程度与科研产出的相关性分析——基于“图书情报档案学”高产作者的计量分析[J].科技进步与对策,2011,28(5):1-5.
- [11] 李纲,刘先红.科研团队中学术带头人的合作特征及其对科研产出的影响[J].情报理论与实践,2016,39(6):70-75.
- [12] 田江,段妮.科研成果与合作的关联性研究-以供应链为例[C]//

- 全国科技评价学术研讨会. 2015.
- [13] 李文聪, 沙思颖, 高雅, 等. 从合著网络及其变化看干细胞研究领域的国际科研合作[J]. 数学的实践与认识, 2016, 46(10): 68-76.
- [14] He Z L, Geng X S, Campbell-Hunt C. Research collaboration and research output: A longitudinal study of 65 biomedical scientists in a New Zealand university[J]. Research Policy, 2009, 38(2): 306-317.
- [15] Glänzel W, Lange C D. A distributional approach to multinationality measures of international scientific collaboration[J]. Scientometrics, 2002, 54(1): 75-89.
- [16] 邱均平, 马凤. 微观层次的引用与合作之间关系的研究——以三位中国图书情报学核心作者为例[J]. 情报理论与实践, 2011, 34(10): 43-47.
- [17] 邱均平, 温芳芳. 近五年来图书情报学研究热点与前沿的可视化分析——基于13种高影响力外文源刊的计量研究[J]. 中国图书馆学报, 2011, 37(2): 51-60.
- [18] 王伟. 信息计量学及其医学应用[M]. 北京: 人民卫生出版社, 2014.
- [19] 朱丽娟, 李丽娜. 科研合作计量指标研究述评[J]. 情报杂志, 2013, 32(6): 76-79.
- [20] 林聚任. 社会网络分析:理论、方法与应用[M]. 北京: 北京师范大学出版社, 2009.
- [21] 姜鑫. 基于“结构洞”视角的组织社会网络内隐性知识共享研究[J]. 情报资料工作, 2012, 33(1): 32-36.
- [22] Schubert A. A Hirsch-type index of co-author partnership ability [J]. Scientometrics, 2012, 91(1): 303-308.
- [23] Schubert A. Jazz discometrics—A network approach[J]. Journal of Informetrics, 2012, 6(4): 480-484.
- [24] Ajiferuke I, Burell Q, Tague J. Collaborative coefficient: A single measure of the degree of collaboration in research[J]. Scientometrics, 1988, 14(5-6): 421-433.
- [25] 高志, 张志强. 个人学术影响力定量评价方法研究综述[J]. 情报理论与实践, 2016, 39(1): 133-138.
- [26] 陈秀娟, 郭进京. 科研人员选择巨型OA期刊发文的影响因素分析——以PLoS ONE为例[J]. 中国科技期刊研究, 2018, 29(6): 586-591.
- [27] 余世英, 陈芳芳. 基于共词分析的数字出版研究热点分析[J]. 科技与出版, 2013(8): 81-85.

(责任编辑 魏瑞斌)