

复合合作强度指数构建及应用研究*

■ 陈云伟 邓勇 陈方 丁陈君 郑颖 刘春江 方曙

中国科学院成都文献情报中心 成都 610041

摘要: [目的/意义]构建一个基于科学家合作行为来评价科学家科研表现的复合合作强度(CCS)指数,该指数的特点在于不仅反映作者合作网络的特征,而且揭示科学家的合作机构及分布特征。[方法/过程]基于科学家的合作与科研表现的正相关性,提出六大假设构建 CCS 指数理论模型。选取来自中国科学院 2007 - 2011 年间发表 SCI 论文在 30 篇以上的 40 位工业生物技术领域科学家为实证分析对象,并对实证结果与合作网络度数、H 指数和篇均被引频次进行比较。[结果/结论]CCS 指数与合作网络度数具有较高的相关性,而与 H 指数和篇均被引频次的相关性很弱,CCS 指数的价值在于揭示科学家的合作者及合作机构的分布广度、深度和密度,反映科学家的国际合作参与度。这些指标之间存在互补关系,可进一步丰富科学家评价的指标,有助于开展科学家的科研表现评价和专家发现工作。

关键词: 复合合作强度 合作 作者合作 科研表现

分类号: G250

DOI: 10.13266/j.issn.0252-3116.2015.13.014

1 引言

当前国际范围内的科研合作越来越广泛,科研人员对合作的重视程度也日益提高。构建一个合作团队的目的是集合一群拥有不同思想、技能或资源的个体,进而激发创造力、促进旧问题的解决并鼓舞新思想的诞生^[1]。科研合作行为可以通过多种形式来揭示,如合作申请项目,合作开展科学研究等。在定量角度,科研合作的强弱可以通过合作撰写的论文或申请的专利数量来表征,这也是当前图书情报领域开展合作研究的重要手段之一。J. Whitfield 在 2008 年指出,在科学研究领域已经呈现出科研合作日益增加的态势,而这种特征将决定一个团队的成功^[2]。他指出,最近几十年里,单作者论文的比例呈现出下降的趋势,几乎所有原创性研究论文都有多个作者^[2]。那么,合作是否能反映作者的科研表现呢?近年来已有一些研究工作对合作与科学家的科研表现之间的关系进行了研究,主要的研究工作和发现简介如下:

S. Wuchty 等在 2007 年通过研究 1975 - 2005 年间发表的 420 万篇论文发现,学院间的合作与产出更高水平的科研成果有关^[3]。P. Panzarasa 和 T. Opsahl 通过分析商业与管理领域 8 360 位作者发表的 9 325 篇论文发现,来自机构间合作的论文比机构内合作的论文平均拥有更高的篇均被引频次^[4]。G. Abramo 等在 2011 年研究了来自意大利大学的学者在 2001 - 2005 年之间的国际合作关系,发现了学者的科研表现与国际化水平的正相关关系^[5]。M. Kato 和 A. Ando 在 2013 年针对化学领域学者的研究也得到了相近的结果,他们指出近年来国际合作论文数量显著增长,与本国内部工作相比,跨国合作论文获得了更多的被引频次^[6]。

在动态水平,G. Palla 等在 2007 年通过分析物理学领域的作者合作网络发现,大的团队如果可以动态地不断更新团队成员,该团队将维持得更加长久^[7]。R. Guimerà 等在 2005 年对艺术和科学两个领域的研究均发现,团队的组织机制同时决定了合作网络的结

* 本文系中国科学院知识创新工程重要方向项目“工业生物技术知识服务研究与应用”(项目编号:KSCX2-KW-G-9)和国家高技术研究发展计划(863 计划“微生物数字资源知识管理系统构建及关键技术研究”(项目编号:2014AA021503)研究成果之一。

作者简介:陈云伟(ORCID:0000-0002-6597-7416),副研究员,博士,E-mail:chenyw@clas.ac.cn;邓勇(ORCID:0000-0001-9179-0500),研究员,硕士;陈方(ORCID:0000-0001-9060-784X),副研究员,博士;丁陈君(ORCID:0000-0003-1403-2372),副研究员,博士;郑颖(ORCID:0000-0001-6503-2212),副研究员,博士;刘春江(ORCID:0000-0001-8934-339X),助理研究员,硕士;方曙(ORCID:0000-0003-4584-7574),研究员,博士。

收稿日期:2015-06-01 修回日期:2015-06-18 本文起止页码:96-103 本文责任编辑:刘远颖

构以及团队的科研表现^[1]。

以上研究表明, 科学家的合作结构与其科研表现间存在正相关关系, 科学家合作水平的高低在一定程度上可以反映其科研表现的强弱。本文将在此基础上构建一个基于科学家合作水平的指数并用于评价科学家的科研表现, 该指数对挖掘以及评价特定领域内的领军科学家具有非常重要的参考价值^[8]。

2 相关工作

目前, 在科学家的科研表现评价领域, 除了传统基于论文数、引文数等定量方法外, 许多新方法已被大量研究并用于评价实践, 根据评价所关注的对象可将这些评价方法分为3类: 基于内容的方法、基于链接结构的方法以及二者的整合方法^[8]。基于内容的方法主要专注于评估某特定主题的对象和用户查询之间的相关性, 却忽略了对象之间的社会关系; 基于链接结构的方法, 如 PageRank 和 HITS 等, 存在的一个普遍不足是无法处理主题漂移问题; 为了克服这两种方法的不足, 又有许多学者对上述两种方法进行了整合。下面重点对与本文研究相关的基于合作的、特别是作者合作网络的科学家的科研评价工作进行介绍。

首先, 基于作者合作论文数量的文献计量分析是合作分析的一种传统方法, G. Melin 和 O. Persson 在 1996 年就曾指出, 作者合作论文的文献计量分析是用于研究和揭示合作网络的结构和变化等合作问题的有效工具, 他们对合作与作者合作的关系进行了讨论^[9]。

其次, 作者合作网络分析在情报研究领域已经得到了广泛研究, E. Otte 和 R. Rousseau 在 2002 年阐述了社会网络分析的基础及其对情报学领域的影响, 他们指出核心科学家奠定了作者合作网络的基础^[10]。K. Borner 等在 2005 年基于论文数及被引频次数据研究了加权作者合作网络, 利用 4 种不同的中心度测度指标鉴别了作者的影响力并对加权作者合作网络进行了可视化分析^[11]。A. Abbasi 等在 2011 年指出, 研究人员的社会网络可被用来预测他们未来的科研表现, 他们发现学者的科研表现与标准化的度数中心度、效率和平均合作强度成正比^[12]。C. McCarty 等在 2013 年对作者的合作网络与 h 指数的比较发现, 作者获得最高的 h 指数需要与许多合作者合作, 其中一些合作者本身也拥有较高的 h 指数^[13]。Chen Yunwei 等在 2013 年基于 *Scientometrics* 期刊 1978-2010 年发表的所有论文的作者合作演化网络挖掘出了科学计量学领域的重要学者^[14]。

第三, 许多研究致力于利用诸如 PageRank 等新指标对学者进行排名, 以克服传统指数的局限性。例如 Liu Xiaoming 等在 2005 年开发了一个名为 Author-Rank 的 PageRank 改良算法, 以作者合作网络中链接的权重来测量作者的影响力^[15]。E. Yan 和 Ding Ying 设计了一个考虑引用和共引网络拓扑结构的加权 PageRank 指数用来评估作者的影响力^[16]。

通过上述研究工作可见, 当前围绕合作行为的科学家评价工作普遍还是基于传统的计量方法和科学家合作网络的方法, 虽然也有一些工作考虑了基于论文数量或被引频次等参数对合作关系进行加权分析, 也有利用诸如 PageRank 或其改良形式等基于链接的方法进行科学家排名研究。但这些工作大多都是在作者层面或整合论文数和引文数层面的研究, 很少有考虑到合作者的地区或机构分布情况, 也很少考虑科学家的合作者和合作机构在其论文中的分布特征。事实上, 即使两个科学家在计量或作者合作网络方面拥有相同的特征, 他们的科研表现也可能因合作者在论文中的分布结构以及合作机构的结构和分布特征的不同而不同。目前从基于合作者及合作机构的组成结构以及在论文中的分布特征角度开展科学家科研表现评价的工作还鲜有报道, 有必要构建整合上述多种因素的复合评价指标, 更加深入地揭示科学家的科研合作水平, 克服仅基于合作论文数量和作者合作网络度数开展科研表现评价的不足, 作为现有科学家科研表现评价手段的有益补充, 进而用于相对更加准确和全面地评价科学家的科研表现。

为此, 本文将基于现有科研评价工作, 构建了一个用于评价科学家科研表现的复合作强度指数 (Compound Collaboration Strength, CCS), 其特点在于除了考虑作者合作网络因素外, 还考虑合作者和合作机构在论文中的分布特征。

3 复合作强度指数 (CCS) 理论模型

3.1 科学家合作结构解析

从论文的著录信息中不难发现, 某一作者的合作者信息中还包括合作者所在机构以及这些合作机构在论文中的分布情况, 这些信息间至少蕴含 3 个偶联的无权二分网络, 分别是作者-机构网络、论文-机构网络以及作者-论文网络。二分网络已在一些研究工作中被用于链接预测工作且展现出了更好的效果^[17-18], 因此, 本文将借鉴二分网络的思想构建一个基于三维偶联二分网络的复合作强度指数来评估作者的科研

合作水平,进而用于评估科学家的科研表现。例如,图 1 中演示的是两个作者 A 和 B 的论文及合作信息,作者 A 和 B 都有 10 篇论文和 5 位合作者,因此基于传统的定量以及作者合作网络中心度(见图 2)分析,他们拥有相同的得分。然而,图 2 却清晰显示出 A 的合作网络比 B 更稠密,看似 A 的科研合作水平强于 B,但该如何测度呢?真实情况确实如此吗?

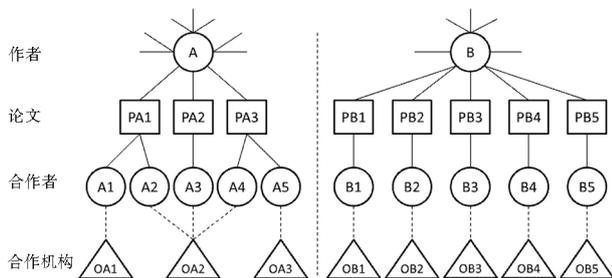


图 1 作者 A 和 B 的耦合合作网络

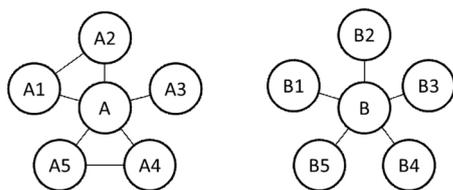


图 2 作者 A 和 B 的 Ego 合作网络

事实上,尽管 A 和 B 都拥有 5 位合作者,但在评价二者的科研合作水平时,他们的合作者对其科研合作水平的贡献度却不同。A 的 5 位合作者来自 3 家不同的机构且总计合作发表了 3 篇论文,而 B 的 5 位合作者来自 5 家不同机构且总计发表了 5 篇合作论文。从这个角度来看,B 的合作者更具多元性。除此之外,A 仅有 30% 的论文有合作者,而 B 有 50% 的论文都有合作者。

3.2 理论假设与模型

通过上述科学家合作结构分析发现,两个拥有相同数量合作者的作者在科研合作水平上也可能是不同的,还需从合作者来源机构、合作论文所占比例以及合作机构在论文中的分布等多个角度进行全面分析。那么,科学家的科研合作水平的评估可以表述为如公式(1)所示的数学问题,表 1 对公式中各参数的含义进行了说明。

$$CCS(i) \sim \begin{matrix} NCA(c) \\ NCO(o) \\ CDCA(pc/n) \\ CDCO(po/n) \\ COH(e) \\ CD(d) \end{matrix} \quad (1)$$

表 1 CCS 指数各参数说明

参数	含义
NCA (c)	指某科学家拥有的合作作者数量, $c = C(i) $
NCO (o)	指某科学家拥有的合作机构数量
CDCA (pc/n)	合作作者深度 指某科学家论文的篇均合作作者数
CDCO (po/n)	合作机构深度 指某科学家论文的篇均合作机构数
COH (e)	合作机构均度,用方差衡量某科学家所有论文的合作机构数量是否均匀分布, $C_e =$ 某科学家每篇论文所拥有的合作机构数量的 STDEV. S 标准差
CD (d)	合作机构密度,指拥有合作机构的文章数量占该科学家所有文章总数的比例

表 2 列举了 A 和 B 的各个参数得分,详细呈现了 A 和 B 合作的差异,然而,为了回答是 A 还是 B 拥有更高的科研合作水平,公式(1)还无法给出量化的数值。

因此,需要设计一个可计算的公式来对公式(1)中的各个参数进行计算并整合。为此本文提出以下假设,作为量化科学家 CCS 得分的理论基础。

表 2 作者 A 和 B 各参数得分

i	n	NCA (c)	NCO (o)	CDCA (pc/n)	CDCO (po/n)	COH (e)	CD (d)	CCS ₁	CCS ₂	CCS ₃
A	10	5	3	0.5	0.5	0.85	0.3	0.34	0.16	0.39
B	10	5	5	0.5	0.5	0.53	0.5	1.18	0.61	0.78

假设:

- 科学家拥有的合作作者数量越多,即 c 值越高,其合作强度越强;
- 科学家拥有的合作机构数量越多,即 o 值越高,其合作强度越强;

- 科学家的论文篇均合作作者数越多,即 pc/n 值越高,其合作强度越强;
- 科学家的论文篇均合作机构数越多,即 po/n 值越高,其合作强度越强;
- 科学家的合作机构数量在所有论文中的分布

越均匀,即 e 值越低,其合作强度越强;

- 科学家拥有合作机构的文章数量占其所有文章总数的比例越高,即 d 值越高,其合作强度越强。

在假定上述假设成立的前提下,本文构建了3种形式的 CCS 计算公式来用于计算科学家的 CCS 得分,分别是公式(2)、(3)和(4)。

$$CCS_1 = \sqrt{c} \sqrt{o} \frac{pc}{n} \frac{po}{n} \frac{1}{e} d \quad (2)$$

$$CCS_2 = \ln c \ln o \frac{pc}{n} \frac{po}{n} \frac{1}{e} d \quad (3)$$

$$CCS_3 = \sqrt{\ln c \ln o \frac{pc}{n} \frac{po}{n} \frac{1}{e} d} \quad (4)$$

实际上,上述3个 CCS 计算公式也是历经多次实证分析试验的,其中基于 CCS₃ 的科学家表现评价结果得到专家普遍认可,揭示了差距,又没有过大地放大差距。作者 A 和 B 通过3个公式计算获得的 CCS 得分已列在表2中,可见,基于 CCS₁ 和 CCS₂ 两个公式计算获得的科学家间的差异较大,而基于 CCS₃ 计算的科学家的得分差异相对较小。在实际应用中,可以根据具体分析需要,审慎选取上述公式或进一步对公式进行优化。

3.3 CCS 指数的特点、应用潜力与不足

CCS 指数的主要特点是在评价科学家的科研合作水平时,不仅考虑了传统方法所采用的合作者的数量及比例等参数,还考虑了合作机构这一影响因素,把合作机构的数量和比例、拥有合作机构的论文所占比例以及合作机构在论文中的分布特征等都作为参数纳入 CCS 指数中。与仅基于合作者进行科研合作水平评价相比,纳入合作机构有关参数后的 CCS 指数所能揭示的信息更多,可以更加突出科学家与机构外部合作者的合作情况。

在应用方面,CCS 指数中的6个参数均可以基于 ISI-WOS 等数据库下载的原始论文数据通过常用的文

献计量和统计软件计算获得,如 TDA、EXCEL 等,再根据 CCS 公式整合起来即可计算获得 CCS 得分。为了便于读者开展基于 CCS 的科研合作水平分析,本文开发了 CCS + 软件来实现对 CCS 指数得分的自动计算。此外,除了对基于论文的分析外,理论上 CCS 指数还可以拓展到基于专利的分析,用于评估发明人或专利权人的科研合作表现,甚至进一步用于评估专利转移的水平和预测专利转移潜力。

尽管如此,当前 CCS 指数所考虑的因素还仅仅是能反映科学家科研合作水平的众多因素中的有限几个,本文的目标是希望能通过有限的参数来尽可能地揭示科研合作表现。本文构建的 CCS 指数依旧有较为明显的改进空间,例如合作机构还可以区分本国机构和他国机构的不同,且可以考虑机构科研水平的差异。

4 实证研究

4.1 数据来源

本文选择中国科学院从事工业生物技术研究且在 2007-2011 年发表论文在 30 篇以上的科学家为实证分析对象,论文数据来自汤森路透的 Web of Science 数据库,涵盖 articles、proceedings papers、letters 和 reviews 4 种文献类型,数据下载日期为 2013 年 5 月。所有数据下载后利用 Thomson Data Analyzer (TDA) 工具进行清洗,以确保每位科学家论文数据的准确性,最终获得 40 位科学家的数据用于实证分析。

4.2 结果与分析

40 位科学家来自中国科学院的 13 个研究所,他们总体上拥有相近的科学环境,因此,不同的合作特征或许可以反映他们科研合作水平的不同,也进而可以用于评估他们科研表现的差异。表 3 统计了 40 位科学家的 CCS 得分以及指标中各个参数的得分情况。

表 3 40 位科学家的 CCS 得分

科学家	N	c	pc/n	o	po/n	e	d	CCS ₁	CCS ₂	CCS ₃
曹坤芳	38	68	3.4	30	1.7	1.8	0.7	97.4	30.9	5.6
车永胜	37	65	5.2	14	0.9	1.2	0.5	64.3	23.5	4.8
陈洪章	48	43	1.5	8	0.2	0.5	0.1	1.6	0.7	0.8
戴玉成	42	93	3.5	54	1.9	6.4	0.7	49.1	12.5	3.5
杜昱光	50	131	5.2	21	0.7	0.7	0.6	169.0	47.8	6.9
郭晨	39	52	4.7	15	0.5	0.8	0.3	28.3	10.8	3.3
郭良栋	40	113	4.9	42	1.6	2.7	0.6	115.4	29.6	5.4
黄宏文	32	43	2.8	11	0.4	0.8	0.3	11.1	4.6	2.1
姜卫红	32	100	6.5	16	1.1	0.8	0.8	256.9	82.0	9.1

(续表 3)

科学家	N	c	pc/n	o	po/n	e	d	CCS ₁	CCS ₂	CCS ₃
匡廷云	36	135	5.6	24	1.4	1.1	0.8	336.9	92.3	9.6
李会泉	32	44	3.2	15	0.6	1.2	0.4	15.7	6.3	2.5
李亦学	99	374	7.1	78	2.5	1.4	1.0	2 110.1	318.9	17.9
李寅	32	140	7.2	30	1.2	1.4	0.7	282.1	73.2	8.6
刘春朝	43	65	3.1	17	1.0	1.2	0.5	41.2	14.7	3.8
刘双江	57	130	4.8	26	0.9	1.0	0.5	128.4	35.0	5.9
刘杏忠	54	132	4.3	33	1.2	0.9	0.8	274.5	71.0	8.4
卢从明	31	102	7.0	14	0.6	1.1	0.4	59.8	19.3	4.4
马光辉	85	153	4.9	42	0.8	0.9	0.6	197.7	46.4	6.8
马隆龙	32	59	4.7	7	0.3	0.5	0.3	12.9	5.0	2.2
马小军	34	84	5.9	11	0.6	0.6	0.5	94.5	33.0	5.7
马延和	68	209	6.8	44	1.8	1.7	0.8	548.8	115.7	10.8
沈世华	32	98	5.0	19	0.9	1.0	0.6	120.9	37.8	6.2
苏志国	111	192	4.6	45	0.8	0.9	0.6	226.1	48.7	7.0
谭华荣	30	64	5.1	5	0.3	0.6	0.2	7.6	2.8	1.7
魏孝义	46	103	4.8	19	0.9	0.9	0.6	114.8	35.4	6.0
吴创之	31	53	4.4	10	0.5	0.6	0.5	43.5	17.3	4.2
邢建民	35	39	4.9	11	0.5	0.6	0.4	30.5	12.9	3.6
熊成东	31	46	4.4	8	0.7	0.6	0.6	64.4	26.7	5.2
徐杰	58	71	4.7	13	0.4	0.7	0.3	31.0	11.2	3.3
许国旺	85	263	7.2	66	1.7	1.3	0.8	1 009.4	178.9	13.4
薛红卫	30	79	3.7	11	0.5	1.0	0.4	19.8	7.0	2.7
袁振宏	30	56	4.5	11	0.4	0.6	0.3	20.3	7.9	2.8
张丽华	58	73	6.0	16	0.4	0.7	0.3	38.3	13.3	3.6
张隽	62	138	5.3	37	1.2	1.0	0.7	332.2	82.7	9.1
张先恩	72	184	6.5	40	1.0	0.9	0.6	370.1	83.0	9.1
张治平	54	137	7.1	28	1.0	1.0	0.6	300.5	79.5	8.9
赵国屏	42	221	8.0	31	2.6	1.7	0.9	958.6	214.7	14.7
赵宗保	40	56	3.4	11	0.7	0.8	0.6	42.7	16.6	4.1
种康	44	186	6.8	37	1.4	1.3	0.7	442.6	100.7	10.0
邹汉法	107	187	5.6	35	0.7	0.9	0.5	198.1	45.5	6.7

图 3 呈现了通过 3 种 CCS 算法计算获得的 40 位科学家 CCS 得分的分布情况, 基于 CCS₁ 算法的 40 位

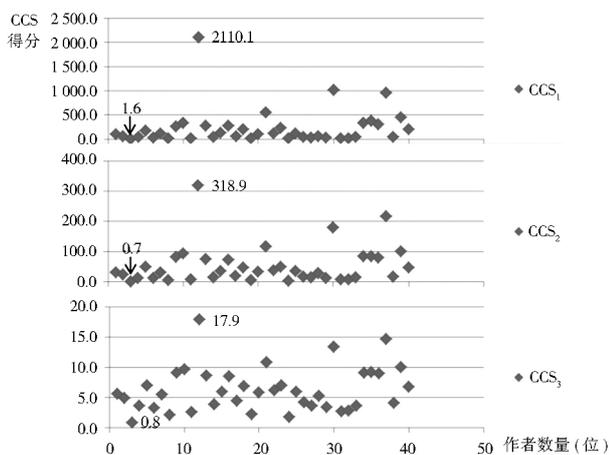


图 3 3 种 CCS 算法的得分分布

科学家的 CCS 值分布在从 1.6 到 2 110.1 的区间, 最大值是最小值的 1 318.8 倍, CCS₂ 和 CCS₃ 的最大值与最小值比分别为 455.6 倍和 22.4 倍。此外, 本文采用标准差的方法来衡量 3 种算法得分的变化和离散程度, 统计结果见表 4。CCS₁ 和 CCS₂ 得分的标准差均高于均值, 只有 CCS₃ 的标准差较小且低于相应均值, 说明 CCS₃ 算法获得的 CCS 值更趋向于总体期望值。这些分析表明, CCS₃ 是基于合作行为来评价科学家科研合作水平的较优选择。

表 4 3 种 CCS 得分算法的标准差

统计指标	CCS ₁	CCS ₂	CCS ₃
均值	231.7	52.4	6.3
标准差	381.3	63.7	3.7

为了揭示 CCS 指数的特点, 本文将 CCS 指数与当前应用较为广泛的几个用于科学家评价的指标进行比

较,包括合作网络度数、篇均被引频次和 h 指数。表 5 比较了 40 位科学家的 CCS_3 、网络度数、篇均被引频次以及 h 指数,表 6 计算了这 4 个指数的相关性。比较可见, CCS_3 指数与篇均被引频次和 h 指数均无显著相关性,显著性水平均高于 0.05, CCS 指数与度数具有较强的正相关性。图 4 对 CCS_3 与度数、篇均被引频次和 h 指数的相关关系进行了可视化展示,随着 CCS_3 得分的增长,网络度数也总体呈现出线性增长,而篇均被引

频次和 h 指数的得分未见明显的增长,没有发现较强的相关性。该结论与 CCS 指数的理论基础相符,即 CCS 指数揭示的是科研合作水平,因此与同样基于合作网络的度数呈现较高的正相关性。以上分析表明,在评价科学家的科研表现方面, CCS 指数与篇均被引频次以及 h 指数拥有显著不同的功能,而与网络度数的功能则相近。

表 5 CCS_3 与篇均被引频次和 h 指数的比较

科学家	$CCS_3 \downarrow$	篇均被引频次	h 指数	度数
李亦学	17.9	17.2	21	374
赵国屏	14.7	12.4	11	221
许国旺	13.4	15.0	23	263
马延和	10.8	6.1	12	209
种康	10.0	20.6	18	186
匡廷云	9.6	4.6	7	135
张先恩	9.1	7.6	14	184
张德	9.1	7.5	11	138
姜卫红	9.1	9.5	10	100
张治平	8.9	8.4	13	137
李寅	8.6	12.5	11	140
刘杏忠	8.4	5.5	10	132
苏志国	7.0	7.9	15	192
杜昱光	6.9	9.5	12	131
马光辉	6.8	8.5	14	153
邹汉法	6.7	20.8	25	187
沈世华	6.2	9.1	11	98
魏孝义	6.0	5.0	8	103
刘双江	5.9	7.4	11	130
马小军	5.7	6.6	9	84

科学家	$CCS_3 \downarrow$	篇均被引频次	h 指数	度数
曹坤芳	5.6	8.0	10	68
郭良栋	5.4	15.3	16	113
熊成东	5.2	4.6	6	46
车永胜	4.8	16.2	15	65
卢从明	4.4	17.3	13	102
吴创之	4.2	9.2	11	53
赵宗保	4.1	22.0	14	56
刘春朝	3.8	13.4	15	65
张丽华	3.6	9.1	12	73
邢建民	3.6	7.9	9	39
戴玉成	3.5	21.0	10	93
徐杰	3.3	8.6	13	71
郭晨	3.3	18.3	16	52
袁振宏	2.8	11.2	10	56
薛红卫	2.7	16.4	13	79
李会泉	2.5	8.5	9	44
马隆龙	2.2	7.6	8	59
黄宏文	2.1	5.7	7	43
谭华荣	1.7	6.0	9	64
陈洪章	0.8	11.2	14	43

注: 度数是指科学家合作网络中每位科学家拥有的合作者数

表 6 CCS_3 与篇均被引频次、h 指数和度数的 spearman 相关性

比较组	相关性	显著性(双侧)
CCS_3 - 篇均被引	-0.044	0.788
CCS_3 - h 指数	0.264	0.100
CCS_3 - 度数	0.875	0.000

5 讨论与展望

本文通过系统的文献调研发现,科学家的科研表现与其合作行为之间存在显著的正相关性,相关研究成果已经在引言部分进行介绍和说明。基于此理论基础,本文构建了一个基于合作行为来评价科学家科研合作水平的复合作强度 CCS 指数,进而用于评估科学家的科研表现。本文构建的 CCS 指数整

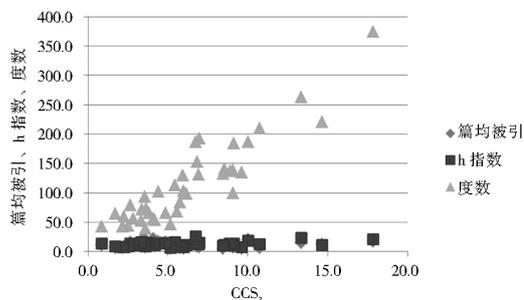


图 4 CCS_3 与篇均被引频次和 h 指数的相关性比较

合了与科学家合作行为相关的 6 个参数,包括合作作者数量、合作机构数量、篇均合作作者数量、篇均合作机构数量、以样本 STDEV. S 标准差表征的每篇论文合作机构分布的异质性以及拥有合作机构论文所占比例。 CCS 指数从作者合作网络、合作机构及分布

的角度来表征科学家的合作行为。基于 CCS 指数的算法,在本文的研究框架下,科学家的 CCS 得分越高,其科研表现越强,但尚不能基于 CCS 得分的差距来量化科研表现的差距,仅用于揭示科学家的科研表现存在差异。

在 CCS 指数的理论框架部分,本文研究的一个重大挑战在于如何有效地整合这 6 个参数以更加客观地反映科学家科研合作水平的差异。为此,实证研究部分对来自中国科学院工业生物技术领域的 40 位科学家进行了比较分析,通过比较发现,CCS₃ 与另外两个公式相比,利用其计算的科学家得分相对更为集中,差异较小,总体在一个量级水平,可能是较优选择。不过,如何选择不同的 CCS 计算公式还需根据实际工作进行选择。事实上,即使是目前看似为较优选择的 CCS₃,仍还可以根据需要进行修正和调整,以提升其揭示科学家科研合作水平的效果,增强其在评估科学家的科研表现方面的功效。

此外,CCS 指数还可以去除合作作者数量以及篇均合作作者数量这两个参数而升级为用于评价科研机构的 CCS 指数,来表征科研机构的科研合作水平,进而用于评价科研机构的科研表现。基于此,在本文分析的基础上,笔者开发了一款名为 CCS+ 的软件,专门用于计算科研机构和科学家的 CCS 指数,读者可以免费下载使用(网址 <http://159.226.140.42/ccs/ccs.html>),包括安装程序、说明文件和本文理论模型研究部分采用的作者 A 和 B 试用文件。

从 CCS 与作者合作网络度数、篇均被引频次以及 h 指数的相关性分析结果来看,CCS 指数可以作为当前科学家科研评价指数的有益补充,丰富当前科研评价的手段,对提高专家发现及评价的效率和效果、促进科研合作、提升科研水平及影响力、知识共享等具有重要价值。需要指出的是,本文比较的篇均被引频次和 h 指数均未区分自引和他引,在后续工作中,可以开展 CCS 与基于他引的被引频次和 h 指数的比较分析工作,以进一步观察 CCS 指数与它们的关系及差异。

最后需要指出的是,在构建 CCS 指数时,除了本文所考虑的合作机构外,合作者的学科、国家或地区等因素也是不同科研合作水平的体现,合作机构的科研水平差异也会影响科学家的科研合作水平。因此,

CCS 指数还有改进的空间,在未来的工作中,还可以将学科、区域的多样性等因素也考虑进来,并将其转化成参数整合到 CCS 指数中。

参考文献:

- [1] Guimerà R, Uzzi B, Spira J, et al. Team assembly mechanisms determine collaboration network structure and team performance [J]. *Science* 2005, 308(5722): 697-702.
- [2] Whitfield J. Collaboration: Group theory [J]. *Nature* 2008, 455(7214): 720-723.
- [3] Wuchty S, Jones B F, Uzzi B. The increasing dominance of teams in production of knowledge [J]. *Science* 2007, 316(5827): 1036-1039.
- [4] Panzarasa P, Opsahl T. Patterns of scientific collaboration in business and management: The effects of network structure and interdisciplinarity on research performance [EB/OL]. [2014-03-08]. http://www.ifr.ac.uk/netsci08/Download/CT_Abstracts/CT314.pdf.
- [5] 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.
- [6] Kato M, Ando A. The relationship between research performance and international collaboration in chemistry [J]. *Scientometrics*, 2013, 97(3): 535-553.
- [7] Palla G, Barabási A L, Vicsek T. Quantifying social group evolution [J]. *Nature* 2007, 446(7136): 664-667.
- [8] Lin Lili, Xu Zhuoming, Ding Ying, et al. Finding topic-level experts in scholarly networks [J]. *Scientometrics*, 2013, 97(3): 797-819.
- [9] Melin G, Persson O. Studying research collaboration using co-authorships [J]. *Scientometrics*, 1996, 36(3): 363-377.
- [10] Otte E, Rousseau R. Social network analysis: A powerful strategy, also for the information sciences [J]. *Journal of Information Science* 2002, 28(6): 441-453.
- [11] Börner K, Dall'Asta L, Ke Weimao, et al. Studying the emerging global brain: Analyzing and visualizing the impact of co-authorship teams [J]. *Complexity: Special Issue on Understanding Complex Systems* 2005, 10(4): 57-67.
- [12] Abbasi A, Altmann J, Hossain L. Identifying the effects of co-authorship networks on the performance of scholars: A correlation and regression analysis of performance measures and social network analysis measures [J]. *Journal of Informetrics* 2011, 5(4): 594-607.
- [13] McCarty C, Jawitz J, Hopkins A, et al. Predicting author h-index using characteristics of the co-author network [J]. *Scientometrics*, 2013, 96(2): 467-483.

[14] Chen Yunwei , Borner K , Fang Shu. Evolving collaboration networks in Scientometrics in 1978 - 2010: A micro-macro analysis [J]. Scientometrics 2013 ,95(3) :1051 - 1070.

[15] Liu Xiaoming , Bollen J , Nelson M L , et al. Co-authorship networks in the digital libraryresearch community [J]. Information Processing and Management ,2005 ,41(6) : 1462 - 1480.

[16] Yan E , Ding Ying. Discovering author impact: A PageRank perspective [J]. Information Processing and Management , 2011 , 47 (1) : 125 - 134.

[17] Guns R. Bipartite networks for link prediction: Can they improve prediction performance [C]//Proceedings of the ISSI 2011. Durban: ISSI 2011 Conference Organising Committee ,2011: 249 - 260.

[18] Zhou Yanbo , Lu Linyuan , Li Menghui. Quantifying the influence

of scientists and their publications: Distinguishing between prestige and popularity [J]. New Journal of Physics ,2012 , 14 (3) : 033033.

作者贡献说明:

陈云伟: 指数设计与文章撰写;
 邓勇: 指数设计;
 陈方: 实证分析专家选取;
 丁陈君: 实证分析专家数据分析;
 郑颖: 实证分析专家数据分析;
 刘春江: CCS + 软件开发;
 方曙: 理论指导。

Research on Construction and Application of CCS Index

Chen Yunwei Deng Yong Chen Fang Ding Chenjun Zheng Ying Liu Chunjiang Fang Shu
 Chengdu Library of the Chinese Academy of Sciences , Chengdu 610041

Abstract: [Purpose/significance] This paper constructs the CCS index for evaluating scientists' performance based on their cooperation behaviors , which is a comprehensive index reflecting not only the features of co-authors network , but also the ones of the co-authors affiliations and distributions. [Method/process] Based on the positive correlation between scientists' collaboration and performances , a CCS index theoretical model is constructed following six hypotheses. 40 authors from the Chinese Academy of Sciences in the field of industrial biotechnology who has at least 30 publications during the period of 2007 - 2011 are used for case study. It makes a comparison of empirical results and co-author network degrees , H index , times cited per paper. [Result/conclusion] This paper finds that while CCS index has high positive correlation to co-author network degrees , it has weak correlation to H index and times cited per paper. Therefore , CCS index can uncover the distribution breadth , depth and density of scientists' cooperators and cooperation agencies , and respond the international cooperation participation of scientists. These indexes could be mutually complementary when used for experts finding and scientist evaluation , and can further enrich the scientist evaluation index.

Keywords: CCS collaboration co-author performance

下 期 要 目

□ 英美数据管理计划与高校图书馆服务

(陈秀娟 胡卉 吴鸣)

□ 图书情报机构在国家智库建设中的使命担当与服务创新

(王红)

□ 国内外图书馆开源软件研究现状与展望

(陈丽君 毕强)

□ 基于公共与非公共检索计算机的 OPAC 检索行为比较研究

(黄崑 张路路 郭建峰)

□ 新媒体时代网络评论观点信息发现机制研究

(曾润喜 王君泽 杜洪涛)

□ 基于 LDA 的科技创新主题语义识别研究

(祝娜 杨京 王效岳等)