

我国高校科研合作的变化趋势： 复杂网络分析的视角

康乐，陈晓宇

[摘要] 本文在 Web of Science 合作发表数据的基础上(包括 A&HCI、SSCI 以及 SCI EXPANDED 论文)，利用复杂网络分析，计算了我国 110 所“211 院校”的度中心性、近中心性、介中心性指标，并讨论了不同层次、专业类型、区域高校之间这些指标的分布差异，以及它们在最近几十年的变化趋势。文章的主要发现是：从分布差异上看，我国“985 院校”在合作网络中处于核心地位，而且，综合院校以及东部院校也是如此。从变化趋势上看，我国高校的科研合作网络直到本世纪初才逐渐形成，并且理科的合作网络已经先于文科网络进入了成熟阶段。

[关键词] 科研合作；复杂网络分析；“211 院校”

一、引言

现代科学的一个重要特征是合作变得日益广泛与频繁——不同机构、不同城市，甚至不同国家的科学家，利用各种技术手段“聚集”在一起，初期集体讨论制定研究计划，中期分工协作负责不同环节，最后共同发表相关科研成果。

我国的科研事业发展也是如此。1956 年我国首次颁发自然科学一等奖时，华罗庚(中科院数学所)、吴文俊(中科院数学所)、钱学森(中科院力学所)分别以独立完成的“典型域上的多元复变数函数论”“示性类及示嵌类的研

[收稿日期] 2018-12-07

[基金项目] 教育部人文社会科学重点研究基地重大项目“经济新常态下的教育财政研究”(16JJD880002)；广东省科技计划项目“科研规律下的政府资助科研项目经费管理制度改革研究”(2015A080802013)。

[作者简介] 康乐，北京大学中国教育财政科学研究所，电子邮箱地址：lekang@pku.edu.cn；陈晓宇(通讯作者)，北京大学教育学院，电子邮箱地址：xychen@gse.pku.edu.cn。

究”“工程控制论”获得这一殊荣。但是到了1982年，“人工全合成牛胰岛素研究”“大庆油田发现过程中的地球科学工作”等自然科学一等奖便涉及了多个单位的多位科学家。2017年的两项自然科学一等奖——“水稻高产优质性状形成的分子机理及品种设计”和“聚集诱导发光”，均由来自3家科研单位的5位科学家集体完成。

更为细致的数据也证明了这一点。Web of Science 网站提供了 A&HCI、SSCI、SCI-EXPANDED 论文的的作者单位统计，在此基础上可以计算“篇均合作单位”这一指标：假定某高校在一年的时间内发表了 M 篇文章，同时其它单位又在这 M 篇文章的署名中出现了 N 次，则 N/M 就是“篇均合作单位”。当所有论文均为独立完成时， $N/M=0$ ；当很多论文中由两个或者更多单位共同完成时， N/M 会接近甚至超过 1。图 1 显示了 1978—2016 年，我国 4 所著名院校这一指标的变化趋势。从中可以发现，无论是人文与社会科学研究还是自然科学研究，这 4 所院校的“篇均合作单位”都在迅速增长之中。

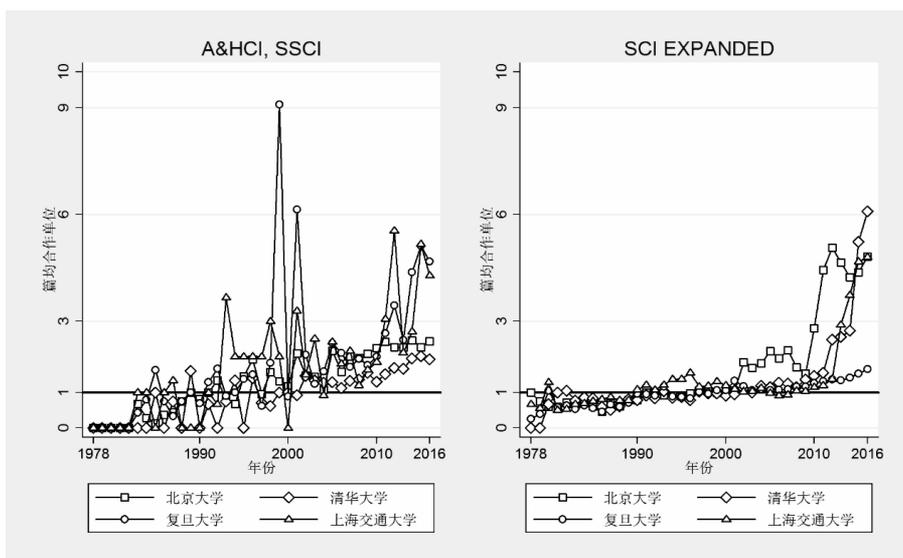


图 1 北京大学等 4 所高校英文发表中“篇均合作单位”指标的变化趋势

数据来源：Web of Science

为什么合作在现代科学中如此盛行？赵君和廖建桥(2013)总结了三个关键的驱动因素：首先是为了提高科研效率，因为很多时候从潜在合作者那里获取相关数据、资源或者某项技术，比自己掌握更为容易，例如考古工作者利用自然科学的碳 14 技术进行测年；其次是为了降低科研成本，这一点往往出现在使用昂贵设备的某些自然科学领域，例如使用加速器的粒子物理学、

使用电镜的结构生物学等等；再次是在资历较深与资历较浅的合作者之间，存在着一种“指导性”的合作，大量导师与研究生之间的合作就属于此类。

另外在文章中，两位作者还讨论了为什么“合作发表”数据往往被用来研究“科研合作”。首先，合作者之间的关系是固定的，客观的合作发表数据可以被其他学者运用并进行验证；其次，通过这一数据来做研究的成本较低，可行性强，客观数据的统计结论比案例研究或问卷调查得到的结论更有意义；再次，文献计量学研究具有时间滞后性的特点，不会反过来影响科研合作状况。

正因为如此，可以在合作发表数据的基础上，利用基于图论的复杂网络分析对科研合作进行定量研究。例如经典的思路是将论文作者定义为网络中的节点；如果两位作者共同署名发表过论文，则认为这两个节点间存在一条边；甚至还可以考虑作者的排序，从而定义边的不同方向。

事实上，国内外的学者们已经开展了很多这方面的工作。Newman (2000)首先利用复杂网络讨论了生物医学领域 MEDLINE 数据库、物理学领域 Los Alamos e-Print Archive 数据库、计算机科学领域 NCSTRL 数据库中的科研合作现象，指出将共同发表论文视为两位科学家之间的边，那么科研合作事实上是一个小世界(small world)。随机地选出两位学者，他们之间会以通过其他学者以非常短的路径连接起来。之后 Barabási 等(2001)也利用复杂网络讨论了 1991—1998 年间数学以及神经科学电子数据库中的共同发表，尤其是合作网络演化的情况。Ramasco 等(2004)则讨论了物理学领域 Los Alamos e-Print Archive 数据库中的合作网络及其自组织(self-organization)现象，并将其与 IMDB 电影数据库中的演员网络、Fortune 1000 数据库中的公司董事会网络做了对比。

最近几年来，国内学者也陆续使用这一方法研究我国的科研合作发表。在院校的层面上，邱均平和温芳芳(2011)选取了我国 39 所“985 工程”高校作为研究样本，借鉴社会网络分析工具和方法，从整体网络、网络密度、核心与边缘结构、网络节点中心性等几个方面对高校之间的科研合作关系进行了全方位、多角度的计量分析。郭崇慧和王佳嘉(2013)则收集了 1996—2011 年我国 39 所“985 工程”高校在 WOS 和 CNKI 数据库中的年度校际合作发文数量，运用复杂网络与社会网络分析方法，从整体、个体和社区三个方面分别对国内和国际期刊校际科研合作网络的网络图、发文量、合著率、节点度、点介数、核心与边缘结构、社区结构等进行了讨论。除此之外，党亚茹等(2013)对环境科学、焦璨等(2014)对心理学等细分领域的合作发表也做了专门研究。胡晓辉和杜德斌(2012)、刘承良等(2017)则将焦点聚集在城市以及

国家的层次上,从更为宏观的角度讨论了城市与城市、国家与国家之间科研合作的网络结构。

本文延续邱均平和温芳芳(2011)以及郭崇慧和王佳嘉(2013)的思路,利用复杂网络分析,主要是度中心性、近中心性、介中心性的计算来讨论我国高校之间的科研合作。有所不同的是,一方面在数据选取的范围上,我们将涉及的院校从39所“985院校”扩大到了110所“211院校”,这样可以从更为全面的角度,刻画出我国高校科研合作网络的全貌。因为在39所“985院校”与剩下71所“211院校”之间,以及71所“211院校”内部,依然存在着一数量量的科研合作^①。另一方面在讨论深入的程度,我们将重点讨论各个院校网络中心性指标的分布差异,包括不同层次、不同专业类型、不同区域的院校,其各自的中心性指标高低如何?以及如果考虑时间趋势,那么这些中心性指标在最近几十年中又有什么变化?

本文分为五个部分。第一部分是引言,第二部分是复杂网络分析的相关概念及计算方法,第三部分是文章涉及的数据,第四部分是各中心性指标的计算结果及相关讨论,第五部分是文章的总结。

二、复杂网络分析

(一)相关概念

网络由节点以及节点之间的边构成,前者代表参与网络的各个成员,后者则描述了它们之间的关系。利用Web of Science网站提供的文科英文论文数据(包括A&HCI以及SSCI),我们绘制了1978—2016年间我国“211院校”之间的文科英文合作网络,如图2所示。需要说明的是,为了绘图以及表达的方便,在本章中我们“暂时”设定了一个50篇的阈值。例如浙江大学并未出现在图2中,但事实上浙江大学与北京大学、上海交通大学也都合作发表过49篇文科英文论文。在文章稍后更为详细的计算部分,我们将去掉这个限制。另外,图2中合作发表最多的是北京大学与北京师范大学、复旦大学与上海交通大学,它们的合作发表数量都是148篇^②。

从图2中可以直观地发现:

首先,有一类院校的连接边数更多一些。例如北京大学与清华大学、上

^① 但是在110所“211院校”之外,进行合作研究并发表A&HCI、SSCI以及SCI EXPANDED论文的普通院校就相对较少了。

^② 事实上,图2中节点大小以及连接边宽度也是由(合作)发表数量所决定的。

程度要弱一些。

上面所说的连接边数、与其他节点的距离以及是否垄断其它节点之间的连接,从不同侧面定性地刻画了节点在网络中的影响力,接下来我们继续从定量的角度,先后引入度中心性(degree)、临近中心性(closeness)以及介中心性(betweenness),更为准确地对节点影响力进行度量。

(二)计算方法

首先,度中心性计算的是给定结点的连接边数。我们考虑 n 个节点构成的图, i 为给定的需要计算的结点, j 为不同于 i 的其它结点,点 i 和点 j 之间存在边时 L_{ij} 取 1, 否则取 0, 此时度中心性 DC 由以下公式决定:

$$DC_i = \sum_{j=1, j \neq i}^n L_{ij} \quad (1)$$

在考虑节点数目 n 之后, DC 相对值为:

$$nDC_i = \frac{\sum_{j=1, j \neq i}^n L_{ij}}{n-1} \times 100 \quad (2)$$

其次,临近中心性计算的是给定节点到所有其它节点距离和的倒数。记 D_{ij} 为点 i 和点 j 之间的最短距离, 此时临近中心性 CC 由以下公式决定:

$$CC_i = \frac{1}{\sum_{j=1, j \neq i}^n D_{ij}} \quad (3)$$

在考虑节点数目 n 之后, CC 相对值为:

$$nCC_i = \frac{n-1}{\sum_{j=1, j \neq i}^n D_{ij}} \times 100 \quad (4)$$

最后,介中心性计算的是给定结点位于其它两点最短路径上的比例之和。记 s 和 t 为不同于 i 的一对节点, $\sigma(s, t)$ 为它们之间最短路径的数量, 其中 $\sigma(s, t | i)$ 经过了点 i , 则 i 点位于 s 和 t 最短路径上的比例为 $\sigma(s, t | i) / \sigma(s, t)$, 此时介中心性 BC 由以下公式决定:

$$BC_i = \sum_{s=1, s \neq i, t \neq s}^n \frac{\sigma(s, t | i)}{\sigma(s, t)} \quad (5)$$

在考虑节点数目 n 之后, BC 相对值为:

$$nBC_i = \frac{2}{(n-1)(n-2)} \sum_{s=1, s \neq i, t \neq s}^n \frac{\sigma(s, t | i)}{\sigma(s, t)} \times 100 \quad (6)$$

钟柯等(2012)指出, 以上三种中心性指标描述了节点位置影响力的不同侧面: 一个节点的度中心性越高说明该节点与其他节点的联系越广泛; 邻近中心性是一种对节点在整个网络中不受其他节点控制的衡量方法, 评价该节

点在空间位置上是否占有优势；介中心性则衡量了一个点的连通潜力，反映了该节点对网络信息流动的影响能力。

基于图2，表1提供了各高校文科英文论文的网络中心性计算结果。我们发现北京大学在三个指标上都具有非常显著的优势，这与北京大学在文科研究中的领先地位是一致的。

表1 示意性结果：1978—2016年间我国“211院校”之间
文科英文论文的合作网络中心性(阈值为50篇)

院校	度中心性	相对值	临近中心性	相对值	介中心性	相对值
北京大学	9.00	60.00	0.0476	71.43	72.00	68.57
中国人民大学	3.00	20.00	0.0313	46.88	0.00	0.00
清华大学	3.00	20.00	0.0313	46.88	0.00	0.00
北京师范大学	4.00	26.67	0.0323	48.39	14.00	13.33
中央财经大学	1.00	6.67	0.0286	42.86	0.00	0.00
复旦大学	4.00	26.67	0.0345	51.72	7.33	6.98
同济大学	2.00	13.33	0.0256	38.46	0.00	0.00
上海交通大学	5.00	33.33	0.0357	53.57	20.67	19.68
武汉大学	2.00	13.33	0.0303	45.46	14.00	13.33
华中科技大学	1.00	6.67	0.0213	31.92	0.00	0.00
中南大学	1.00	6.67	0.0238	35.71	0.00	0.00
中山大学	4.00	26.67	0.0345	51.72	14.67	13.97
暨南大学	1.00	6.67	0.0233	34.88	0.00	0.00
西南大学	1.00	6.67	0.0222	33.33	0.00	0.00
西安交通大学	1.00	6.67	0.0238	35.71	0.00	0.00
四川大学	4.00	26.67	0.0357	53.57	15.33	14.60

数据来源：根据 Web of Science 网站数据计算

最后需要说明的是，常见的网络结构可以分为无向图和有向图。在无向图中不区分连接的方向，因此只要简单统计边数即可，而在有向图中需要区分方向。来自 Web of Science 的数据没有提供合作论文的署名顺序，我们无法对合作发表区分连接方向，在这里只能考虑无向图。另外在示意性的图2中，我们考虑的是1978—2016年的合作发表加总，因此设定了50篇这样一个很高的阈值，即拥有50篇或者更多的合作发表记为存在1条边，稍后我们将逐年计算各个院校的中心性指标，因此具体的阈值将有所不同。

三、数据来源

本章使用的数据来自 Web of Science 网站。

Web of Science 网站提供了 A&HCI、SSCI 和 SCI Expanded 三项不同的论文统计——与之前的做法一致，我们将 A&HCI 与 SSCI 进行加总，称之为文科英文发表，SCI Expanded 则称之为理科英文发表。如表 2 所示，1983 年北京大学一共有 15 篇论文被 A&HCI 与 SSCI 收录，这些论文中还有 8 个非北京大学的署名单位出现过。表 3 则显示，1979 年北京大学一共有 28 篇论文被 SCI Expanded 收入，这些论文中还有 13 个非北京大学的署名单位出现过。

表 2 1983 年北京大学 A&HCI 与 SSCI 论文的合作院校列表

非北京大学的署名单位	出现次数
University of California Berkeley	2
University of California System	2
Beijing Municipal Research Institute of Environment	1
Chinese Academy of Social Sciences	1
Chinese Research Academy of Environmental Sciences	1
Michigan State University	1
N China Inst Comp Technol	1
Railway Minist	1

数据来源：Web of Science 网站

表 3 1979 年北京大学 SCI EXPANDED 论文的合作院校列表

非北京大学的署名单位	出现次数
Chinese Academy of Sciences	6
Nanjing Institute of Astronomical Optics Technology Naoc CAS	2
Nanjing University	2
Purple Mountain Observatory CAS	2
Beijing Inst Canc Res	1
Gansu Res Inst New Med Pharmacol	1
Inst Atom Peking	1
Municipal Hosp 4	1

续表

非北京大学的署名单位	出现次数
Shanghai Normal University	1
Southwestern Institute of Physics China	1
University of Maryland College Park	1
University System of Maryland	1
Yunnan Astronomical Observatory Naoc CAS	1

数据来源：Web of Science 网站

根据以上数据就可以构建本文关心的合作发表网络。

整个数据采集工作主要分两步进行：首先，我们确定一个所关心的院校列表，就本文而言也就是全部的“211 院校”，并利用其英文名称检索不同年份的合作伙伴列表；第二，对这些伙伴列表进行筛选，仅仅保留合作伙伴同样是“211 院校”的记录，由此形成有效的、网络中的边。例如表 2 事实上就不构成任何有效的边，因为涉及的 8 个非北京大学署名单位都不是“211 院校”；但表 3 中的南京大学为“211 院校”，因此北京大学与南京大学之间就构成了边。完成这些工作之后，我们就得到了不同年份以 110 所“211 院校”为节点，各自合作发表为边的复杂网络。在此基础之上，我们就可以计算各个院校的中心度指标。

另外还有几点技术细节需要说明：第一，我们之所以仅仅考虑“211 院校”，主要是因为非“211 院校”的英文发表不多，特别是人文社会科学领域，中间由合作发表构成的、有效的边还会更少。第二，1978 年以来我国高校经历了多次更名与合并，Web of Science 网站已经对此进行了处理，例如 1983 年原始署名为 BEIJING MED COLL(北京医学院)的文章，经处理后已经统一归到了北京大学名下。第三，与之前示意性的计算不同，我们此时修改了 50 篇门槛值的设定。考虑到人文社会科学与自然科学整体发表规模的差异，我们在前者中设置的年度门槛值为 1 篇、5 篇，后者为 10 篇、50 篇。第四，当无论如何高校 i 也无法与高校 j 建立连接时(此时网络为非连通的)，我们假设其距离为 110 而非无穷大。

最终我们采集了 1979—2016 年间，110 所“211 院校”39 年的合作发表数据。

四、计算结果及讨论

(一) 计算结果

表4报告了相关院校网络中心性指标,包括相对度中心性、近中心性、介中心性计算结果的描述统计。

表4 科研合作网络中心性计算结果的描述统计

变量	观测值	均值	标准差	最小值	最大值	均值	标准差	最小值	最大值
文科英文		年度门槛值: 1 篇				年度门槛值 5 篇			
度中心性	4, 290	2.086	5.894	0	52.68	0.214	1.259	0	19.64
近中心性	4, 290	9.736	17.29	0.885	69.03	2.699	9.074	0.885	100
介中心性	4, 290	0.255	1.088	0	19.34	0.0257	0.252	0	9.191
理科英文		年度门槛值: 10 篇				年度门槛值: 50 篇			
度中心性	4, 290	2.733	7.222	0	63.39	4.808	11.11	0	75.89
近中心性	4, 290	11.69	19.15	0.885	100	15.72	22.94	0.885	83.33
介中心性	4, 290	0.264	1.180	0	18.31	0.268	1.123	0	19.75
分类	观测值					百分比			
“985”院校	1480					34.5%			
“211”院校	2810					56.5%			
东部院校	2420					56.4%			
中部院校	935					21.8%			
西部院校	935					21.8%			

数据来源: 根据 Web of Science 网站数据计算

就文科英文论文而言,在年度门槛值为1篇的假设下,各院校相对度中心性、近中心性、介中心性的均值分别为2.068、9.736、0.255;在年度门槛值为5篇的假设下,这3个中心性指标的均值分别为0.214、2.699、0.0257。就理科英文论文而言,在年度门槛值为10篇的假设下,各院校相对度中心性、近中心性、介中心性的均值分别为2.733、11.69、0.264;在年度门槛值为50篇的假设下,这3个中心性指标的均值分别为4.808、15.72、0.268。另外在整个样本中,“985院校”和非“985院校”的观测值分别为1480与2810,分别占34.5%和65.5%;东部、中部、西部院校则分别为2420、

935 与 935，分别占 56.4%、21.8% 和 21.8%。

(二) 分布差异

接下来讨论这些院校网络中心性指标的分布差异。首先，不同层次的院校，包括“985 院校”和“211 院校”，各自的中心性指标高低如何？其次，不同专业类型的院校，包括综合院校、理工院校和文科院校，它们之间的中心性指标是否存在差异？第三，不同区域的院校，包括东部、中部和西部院校，它们之间的中心性指标是否也不同？最后，如果考虑时间趋势，那么这些中心性指标在最近若干年又有什么变化？

具体地，考虑以下的回归方程：

$$\text{Centrality}_{it} = \beta_0 + \beta_1 \text{university_class}_i + \beta_2 \text{university_type}_i + \beta_3 \text{region}_i + \beta_4 \text{year}_t + u_{it} \quad (7)$$

其中 Centrality 代表中心性，包括度中心性、近中心性和介中心性，基于文科英文论文以及理科英文论文计算；university_class 代表院校层次，university_type 代表院校类型，region 代表所在地区，year 则代表对应的年份。这里需要指出的是，按照教育部的分类标准，我国院校可以分为 12 个专业类型，为了简明起见，文章对这 12 个专业类型进行了归并，其中综合院校包括原综合院校、师范院校，理工院校包括原理工院校、农林院校、林业院校和医药院校，文科院校则包括原文学、财经院校、政法院校、艺术院校、体育院校和民族院校。

表 5 提供了文科英文论文的回归结果。就院校层次而言：假设年度门槛值为 1 篇，“985 院校”在度中心性、近中心性、介中心性三个回归中的估计系数分别为 2.53、3.18、0.51；假设年度门槛值为 5 篇，估计系数则分别为 0.41、2.92、0.06。这些估计系数均在 1% 的显著性水平上显著，说明在 A&HCI 和 SSCI 的合作发表中，“985 院校”的地位显著高于作为基准组的非“985 院校”。

就专业类型而言：假设年度门槛值为 1 篇，理工院校、文科院校在三个回归中的估计系数分别为 -1.07、-1.06、-0.23 以及 -0.80、-0.76、-0.21；假设年度门槛值为 5 篇，估计系数则分别为 -0.27、-1.84、-0.04 以及 -0.17、-0.85、-0.03。这些估计系数同样均在 1% 的显著性水平上显著，说明在 A&HCI 和 SSCI 的合作发表中，作为基准组的综合院校地位显著高于文科院校，文科院校又高于理工院校。

就地区分布而言：假设年度门槛值为 1 篇，中部院校、西部院校在三个回归中的估计系数分别为 -0.61、-0.97、-0.14 以及 -1.10、-1.78、-0.24；假设年度门槛值为 5 篇，估计系数则分别为 -0.13、-1.29、-0.02

以及-0.16、-1.29、-0.03。这些估计系数同样均在1%的显著性水平上显著,说明在A&HCI和SSCI的合作发表中,作为基准组的东部院校地位显著高于中部院校,中部院校又高于西部院校。

表5 文科英文论文分布差异的回归结果

	(1)	(2)	(3)	(4)	(5)	(6)
变量	度中心性	近中心性	介中心性	度中心性	近中心性	介中心性
“985 院校”	2.53*** (0.15)	3.18*** (0.22)	0.51*** (0.04)	0.41*** (0.04)	2.92*** (0.29)	0.06*** (0.01)
理科院校	-1.07*** (0.14)	-1.06*** (0.23)	-0.23*** (0.03)	-0.27*** (0.04)	-1.84*** (0.27)	-0.04*** (0.01)
文科院校	-0.80*** (0.20)	-0.76* (0.40)	-0.21*** (0.04)	-0.17*** (0.04)	-0.95*** (0.36)	-0.03*** (0.01)
中部院校	-0.61*** (0.15)	-0.97*** (0.27)	-0.14*** (0.03)	-0.13*** (0.04)	-1.29*** (0.26)	-0.02*** (0.01)
西部院校	-1.10*** (0.15)	-1.78*** (0.28)	-0.24*** (0.03)	-0.16*** (0.04)	-1.29*** (0.31)	-0.03*** (0.01)
常数项	0.05 (0.18)	0.93*** (0.26)	0.03 (0.04)	0.05 (0.04)	1.34*** (0.27)	0.01** (0.01)
年度门槛值	1 篇	1 篇	1 篇	5 篇	5 篇	5 篇
观测值	4, 290	4, 290	4, 290	4, 290	4, 290	4, 290
R ²	0.53	0.84	0.21	0.21	0.23	0.12

注:括号内为稳健的标准误;*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$ 。

表6提供了理科英文论文的回归结果。就院校层次而言:假设年度门槛值为10篇,“985院校”在度中心性、近中心性、介中心性三个回归中的估计系数分别为3.95、4.84、0.66;假设年度门槛值为5篇,估计系数则分别为5.81、4.44、0.64。这些估计系数均在1%的显著性水平上显著,说明在SCI Expanded的合作发表中,“985院校”的地位显著高于作为基准组的非“985院校”。

就专业类型而言:假设年度门槛值为10篇,理科院校、文科院校在三个回归中的估计系数分别为-1.29、-0.29、-0.09,以及-2.78、-10.60、-0.18;假设年度门槛值为50篇,估计系数则分别为-1.19、-0.19、-0.06,以及-4.48、-11.90、-0.18。这些估计系数大多都在1%的显著性水平上显著,说明在SCI Expanded的合作发表中,作为基准组的综合院校地

位显著高于理科院校，理科院校又高于文科院校。

就地区分布而言：假设年度门槛值为10篇，中部院校、西部院校在三个回归中的估计系数分别为-0.20、-0.35、-0.08，以及-1.56、-3.27、-0.16；假设年度门槛值为5篇，估计系数则分别为-0.46、-0.36、-0.07，以及-2.09、-2.00、-0.17。其中西部院校的估计系数都在1%的显著性水平上显著，说明在SCI Expanded的合作发表中，作为基准组的东部院校地位略高于中部院校，但不是特别显著，但与此同时又显著地高于西部院校。

表6 理科英文论文分布差异的回归结果

	(1)	(2)	(3)	(4)	(5)	(6)
变量	度中心性	近中心性	介中心性	度中心性	近中心性	介中心性
“985 院校”	3.95*** (0.20)	4.84*** (0.29)	0.66*** (0.04)	5.81*** (0.27)	4.44*** (0.24)	0.64*** (0.04)
理科院校	-1.29*** (0.18)	-0.29 (0.27)	-0.09** (0.04)	-1.19*** (0.25)	-0.19 (0.23)	-0.06 (0.04)
文科院校	-2.78*** (0.26)	-10.60*** (0.77)	-0.18*** (0.03)	-4.48*** (0.41)	-11.90*** (0.84)	-0.18*** (0.03)
中部院校	-0.20 (0.20)	-0.35 (0.35)	-0.08** (0.03)	-0.46 (0.28)	-0.36 (0.30)	-0.07** (0.03)
西部院校	-1.56*** (0.19)	-3.27*** (0.36)	-0.16*** (0.04)	-2.09*** (0.26)	-2.00*** (0.31)	-0.17*** (0.03)
常数项	-0.12 (0.27)	1.29*** (0.49)	-0.12*** (0.04)	-0.44 (0.39)	1.25** (0.48)	-0.12*** (0.04)
年度门槛值	10 篇	10 篇	10 篇	50 篇	50 篇	50 篇
观测值	4, 290	4, 290	4, 290	4, 290	4, 290	4, 290
R ²	0.47	0.76	0.17	0.58	0.85	0.19

注：括号内为稳健的标准误；*** $p < 0.01$ ，** $p < 0.05$ ，* $p < 0.10$ 。

(三)变化趋势

图3提供了三个中心性指标在时间尺度上的变化，均基于文科英文，也就是A&HCI和SSCI论文计算。其中，左侧为度中心性，中间为近中心性，右侧为介中心性。横坐标是年份，纵坐标是之前省略了的、以1978年为基准组其他各年份虚拟变量的回归系数。我们发现从1979年到20世纪末，三个中心性指标几乎都没有任何增长，真正的转折点是在21世纪的前几年，尤其

在大约 2005 年以后，度中心性、近中心性、介中心性三个指标集体得到了迅速提高。

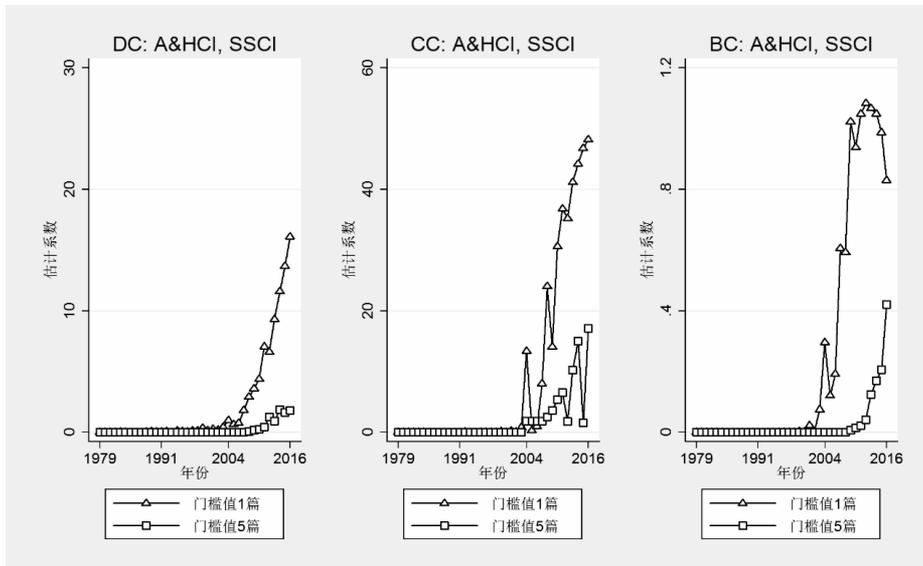


图3 基于 A&HCI、SSCI 英文论文的度中心性、近中心性、介中心性的变化趋势
数据来源：根据 Web of Science 网站数据计算

类似地，图 4 同样提供了三个中心性指标在时间尺度上的变化，但均基于理科英文，也就是 SCI Expanded 论文计算。其中，左侧为度中心性，中间为近中心性，右侧为介中心性。横坐标是年份，纵坐标是之前省略了的、以 1978 年为基准组其他各年份虚拟变量的回归系数。与之前的 A&HCI 和 SSCI 论文非常类似，我们发现从 1979 年到 20 世纪末，三个中心性指标几乎也都没有任何增长，真正的转折点同样是在 21 世纪的前几年，度中心性、近中心性、介中心性三个指标集体得到了迅速提高。

(四) 介中心性的停滞

图 4 中值得关注的是，理科论文的平均介中心性在迅速增长之后出现了停滞，甚至近些年来还有下降的趋势——这跟介中心性的定义有关，也恰恰反映了我国自然科学方面的合作研究已经进入了较为成熟的阶段。

在早期，我国高校的合作研究非常不发达，也没有形成真正有意义的网络，因此绝大部分高校的介中心性都非常低。中期这一局面逐渐得到改善，一部分优势地位的高校开始兴起，它们之间也形成了局部的合作网络，并且垄断了其他劣势地位高校之间的连接，这就会造成优势高校以及所有高校平均下来的介中心性都得到了提高。后期处于劣势地位的高校也逐渐赶迎头赶

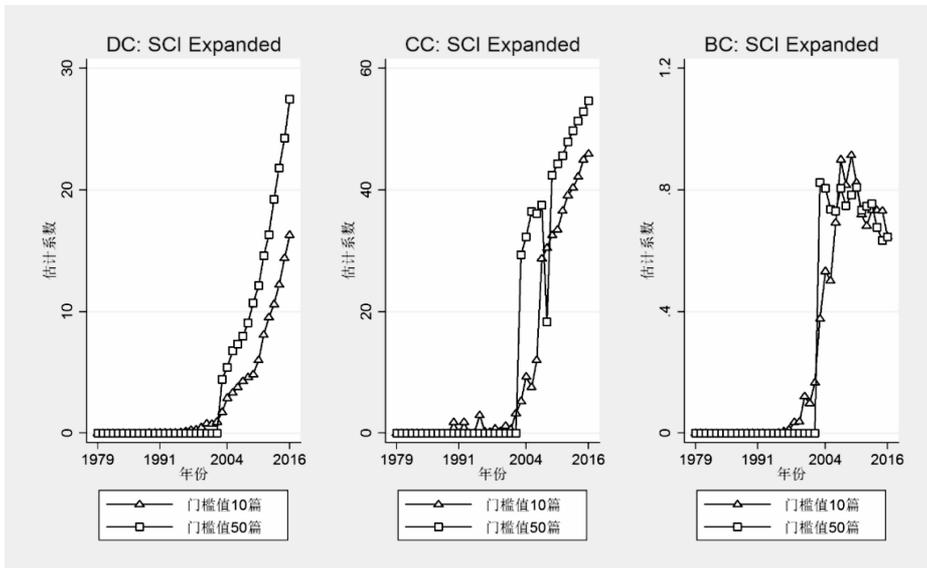


图4 基于 SCI Expanded 论文的度中心性、近中心性、介中心性的变化趋势

数据来源：根据 Web of Science 网站数据计算

上，它们之间也开始寻求合作并建立了直接连接，此时优势高校自然也就丧失了之前垄断其他高校连接的地位，它们以及所有高校平均下来的介中心性都会逐渐下降。

特别地，考虑一个由 N 所高校组成的合作网络，假设早期两两之间没有任何直接连接，那么平均介中心性也就是 0；中期平均介中心性会增长到 0 以上；后期假定两两之间全部建立起了直接连接，那么也就没有任何节点可以垄断其他节点之间的最短连接，此时平均介中心性又会回到 0。因此介中心性的趋势类似于一个倒 U 型曲线。但与此同时，度中心性和近中心性则会一直增长，直到等于理论上的最大值为止^①。

从这个角度而言，我国人文社会科学的发展程度还很不够。正如前面提到的，大量非“985 院校”甚至还未跨过进入门槛，其英文发表还非常稀少，更遑论劣势高校之间的英文发表合作。所以如图 3 所示，如果将合作发表的门槛值定为容易达到的 1 篇，那么平均介中心性的停滞事实上已经发生，而将门槛值定为较难达到的 5 篇，则平均介中心性依然在增长。

^① 度中心性的理论最大值为 $n-1$ ，近中心性的理论最大值为 $1/(n-1)$ ，两者的相对最大值均为 100。

五、总结

现代科学的一个重要特征是合作变得日益广泛与频繁。利用 Web of Science 提供的数据计算“篇均合作单位”这一指标,我们发现中国的科研事业发展也是如此。如何定量地研究这些科研合作网络?现有文献提供了两个非常重要的思路,一个是收集合作发表的数据,一个是采用复杂网络分析的方法。

本文在 Web of Science 合作发表数据的基础上,利用复杂网络分析,计算了我国 110 所“211 院校”的度中心性、近中心性、介中心性指标,并讨论了不同层次、专业类型、区域高校之间这些指标的分布差异,以及它们在最近几十年中的变化趋势。文章的主要发现是:

首先,从分布差异上看,我国“985 院校”在合作网络中处于核心地位,而且,综合院校以及东部院校也是如此。具体而言,我们发现无论是以 A&HCI、SSCI 衡量的文科发表,还是以 SCI EXPANDED 衡量的理科发表,在度中心性、近中心性、以及介中心性这三个指标中,“985 院校”均显著高于作为基准组的非“985 院校”。这意味着我国高校的科研合作网络是围绕着“985 院校”构成的;或者说,“985 院校”是科研合作网络的核心,而非 985 院校”则位于网络的边缘。除此之外,综合院校、东部院校的中心性指标也相对较高,因此在以专业类型、分布区域划分的科研合作网络中,综合院校与东部院校也都处于核心位置。从某种意义上说,要想进一步提高我国科研事业的发展水平,在激励机制上必须鼓励这些高校更好地发挥辐射、带动作用。另外,在资源分配上,也无需过于担心它们获得了更多的资源。这些高校位于科研合作网络的核心,因此它们科研事业的飞速发展也会带动边缘高校的进步。

其次,从变化趋势上看,我国高校的科研合作网络直到 21 世纪初才开始形成,并且理科的合作网络已经先于文科网络进入成熟阶段。具体而言,我们发现无论是以 A&HCI、SSCI 衡量的文科发表,还是以 SCI EXPANDED 衡量的理科发表,在 1979 年一直到 20 世纪末这段时间内,度中心性、近中心性、以及介中心性这三个指标几乎都没有任何增长。真正的转折点是在 21 世纪的前几年,尤其在大约 2005 年以后,这三个指标集体才得到了突飞猛进式的提高。另一方面,如果仔细考察介中心性,我们发现理科论文的平均介中心性在迅速增长之后出现了停滞,甚至近些年来还有下降的趋势。按照介中心性的定义,这说明某一所高校垄断另外两所高校之间连接的情况正在减

少，或者说那两所高校之间也逐渐产生了直接连接。这恰恰说明了我国高校间的理科合作网络已经逐渐进入成熟阶段。相比而言，我国人文社会科学的发展程度还远远不够，科研合作网络仍处于发育之中，从这个角度来讲，我们要注重鼓励和引导各个高校在人文社会科学领域的合作，进一步提高其科研事业的发展水平^①。

当然，本文不可避免地存在一些不足和未来研究可以进一步挖掘的地方。首先，我们仅仅考虑了 Web of Science 统计的英文发表。事实上，郭崇慧和王佳嘉(2013)在研究中还考虑了来自 CNKI 的中文论文。如果研究 110 所“211 院校”中文发表的合作网络，讨论起中心性指标的分布差异与变化趋势，并将其与英文发表作对比，相信会得到更有意义的结论。其次，我们仅仅比较了中心性指标的在各组院校之间的差异，然后得到哪些高校位于网络中心，哪些高校又位于网络边缘的结论。事实上，整个科研合作网络还拥有比这些结论更为丰富的内容——例如实际的高校空间分布如何影响抽象的合作网络结构？外生的政策冲击，如 20 世纪末“985 工程”的出台是否对网络结构造成了影响？这些都是未来研究的发展方向。

[参考文献]

- 党亚茹、李雪娇、宋素珍，2013：《环境科学研究合作网络分析》，《情报杂志》第 11 期。
- 郭崇慧、王佳嘉，2013：《“985 工程”高校校际科研合作网络研究》，《科研管理》第 51 期。
- 胡晓辉、杜德斌，2012：《长三角城市间科学合作的社会网络动态分析》，《科技与经济》第 1 期。
- 焦璨、张楠楠、张敏强等，2014：《基于社会网络分析的心理学科科研人员合作网络研究》，《吉林大学社会科学学报》第 4 期。
- 刘承良、桂钦昌、段德忠等，2017：《全球科研论文合作网络的结构异质性及其邻近性机理》，《地理学报》第 4 期。
- 邱均平、温芳芳，2011：《我国“985 工程”高校科研合作网络研究》，《情报学报》第 7 期。
- 赵君、廖建桥，2013：《科研合作研究综述》，《科学管理研究》第 2 期。
- 钟柯、肖昱、许珺等，2012：《基于列车运行网络的中国城市中心性分析》，《地球信息科学学报》第 1 期。

^① 需要注意的是，人文社会科学本身的特点也是合作网络不够发达的原因之一。例如 2018 年 12 月出版的 *American Sociological Review* 第 83 卷第 6 期、*American Economic Review* 第 108 卷第 12 期，其发表论文的平均作者数量为 1.6、2.5；而 2018 年 12 月 26 日出版的 *Journal of the American Chemical Society*、12 月 27 日出版的 *The New England Journal of Medicine*，其平均作者数量则达到了 8.4、21.4。当然，如果考虑的是作者单位数量而不是作者数量，这个差距会小一些。

- Barabási, A. L., H. Jeong and Z. Néda, 2001, Evolution of the Social Network of Scientific Collaborations. *Physical Statistical Mechanics & Its Applications*, 311(3): 590–614.
- Newman, M. E. J., 2000, The Structure of Scientific Collaboration Networks. *Proceedings of the National Academy of Sciences of the United States of America*, 98(2): 404–9.
- Ramasco, J. J., S. N. Dorogovtsev and R. Pastorsatorras, 2004, Self-organization of Collaboration Networks. *Physical Review E Statistical Nonlinear & Soft Matter Physics*, 70(2): 036106.

Complex Network of Scientific Research Collaboration in China's Higher Education Institutions

KANG Le¹, CHEN Xiao-yu²

(1. China Institute for Educational Finance Research, Peking University

2. Graduate School of Education/Institute of Economics of Education, Peking University)

Abstract: Based on the data provided by Web of Science (including A&HCI, SSCI and SCI EXPANDED publications), this article uses the complex network analysis to calculate the indicators of degree centrality, closeness centrality and betweenness centrality of 110 “211 project universities” in China, and then discusses the differences of these indicators between universities differentiating from each other in terms of their levels, specialized types, and regions, as well as their developing trends in recent decades. The main findings are: First of all, from the perspective of distribution differences, China’s “985 project universities” are at the core of the cooperative network, so are the comprehensive institutions and Eastern institutions. Secondly, from the perspective of developing trends, it’s not until the beginning of 21st century that the research cooperation network of Chinese universities take shape, within which the cooperation network of sciences research has passed into the mature stage ahead of that of arts, humanities, and social sciences research.

Key words: scientific research collaboration; complex network analysis; “211 project universities”

(责任编辑: 刘泽云 责任校对: 刘泽云 胡咏梅)