



近代中国农业领域科研合作网络分析

□陈海珠 李树青 汪圣忠 包平*

摘要 收集整理了晚清民国时期《农业论文索引》中粮食作物专题研究论文作者合作情况,构建了有向合作网络。采用社会网络分析方法对这一科研合作网络进行实证分析,包括:整体网络分析、作者中心性分析、合作频次分析、凝聚子群分析、最大合作子群分析;并且结合文献计量方法对粮食作物研究专题所有高产作者发文量与其在合作网络中的度数、入度作了相关性分析。通过实证研究,对我国近代科学学者合作网络的整体状况和特征进行了总结并就该研究意义提出了一些思考。

关键词 合著分析 社会网络分析 科研合作网络 信息可视化 近代中国

分类号 G353

DOI 10.16603/j.issn1002-1027.2019.04.010

1 引言

社会网络分析(Social Network Analysis, SNA)是20世纪60年代以来社会学大师怀特(Harison White)及其后继者布尔曼(Boorman)、布里格(Brieger)和弗里曼(Linton Freeman)等人一手由数学的图形理论导演出来的一套数学分析方法^[1]。和致力于分析属性数据的统计学不同,社会网络分析是在研究分析关系数据中发展起来的新兴学科。社会网络要分析的关系,可以是人与人之间的情感关系、情报关系及咨询关系,也可以是组织与组织间的信息交换关系、财务关系以及联合研发关系等。由于它强调关系思维,关注对社会关系属性和结构的分析,提供了与传统的忽视结构的研究倾向不同的新的分析思路和范式,社会网络分析在近二三十年来迅速发展成为学术界的一个热点领域,得到了包括社会科学界以及数学、物理等自然科学界的广泛接纳与应用^[2]。

科研合作网络是科学家通过交流思想、共享知识、合作发表论文而形成的人际关系网络,它是一种典型的社会合作网络,同时也是具有高聚集性、无标度、小世界联结等性质的复杂网络,适合采用社会网络分析方法进行分析^[3-4]。社会网络分析方法在科学家之间的交流网络研究中运用的最早成果之一,是克兰(Crane, 1972)的《无形学院》(Invisible Col-

lege),她用这种观念来解释科学知识的增长,指出:科学中研究领域的合作者群体以及将这些合作者群体联系起来的交流网络形成了“无形学院”,“无形学院”对于统一研究领域和为领域提供凝聚力和方向是有帮助的^[5]。但是由于时代因素,克兰当时还未能使用图论以及社会计量学的系列测度概念来揭示其内在关系属性和结构。近些年来,国内外越来越多的学者尝试用社会网络分析的概念和测量技术来研究科研合作现象^[6]。例如,一些学者将社会网络分析法中的小团体分析、中心性分析、角色分析、网络密度等测度工具引入信息计量学中,对具体学科领域的科研合作现象进行分析。如塞得(Yasmin H. Said)等构建了许多著名学者的合作网络,基于作者合作网络的集聚特征,划分了作者合作的类型^[7];李亮、朱庆华从中心性分析、凝聚子群分析和核心-边缘结构分析3个角度来研究《情报学报》的合作现象,对国内情报学领域的合作现象进行实证研究^[8];雷银枝、宋歌将社会网络分析中的位置分析引入作者同被引分析,改进科学共同体测度方法,对我国经济学领域学术共同体科研合作与交流关系进行实证分析^[9]。除了对科研作者及群体的分析,运用社会网络分析法对机构科研合作现象开展实证研究的成果也屡见不鲜。如邱均平、瞿辉通过对生物多样性研究高产机构所构成的合作网络进行社会网络分析,揭示科研机构合作网络知识扩散规律,剖析我

* 通讯作者:包平,ORCID: 0000-0002-1131-5169,邮箱:baoping@njau.edu.cn。



国当前科研合作网络知识交流、扩散的特点^[10]。朱云霞,魏建香利用社会网络分析相关测度工具对 2005~2009 年国家社科基金项目文献的发表机构科研合作情况进行了量化分析、内容分析^[11]。

上述作者基本上是基于单一期刊或少数期刊(也有收集期刊论文数据自建题录数据库)来进行数据分析,并且选取的数据年限较短,十年以上、数十年以上的极少,这必然导致数据分析时的结果不够精确,无法完整反映某个学科领域的整体合作情况。另外,几乎所有的国内研究都集中在当代数据,对于历史研究数据,特别是采集近代学术研究数据开展社会网络分析的文章不多见。本研究通过建构中国近代粮食作物研究论文题录数据库,经过数据清洗,整理获取这一段时期所有作者的合作数据,而后综合运用社会网络分析法、数理统计分析等手段,对近代粮食作物研究领域的科研合作关系进行实证研究,试图将社会网络分析拓展到近代科学科研合作结构特点与规律研究,以及在科研合作意义探析方面作一个有益的探索。

2 研究数据及分析工具

数据质量是社会网络分析可靠性和有效性的关键因素之一。本研究采集数据的来源主要是三本索引图书,其中,由万国鼎先生指导、金陵大学农业图书研究部陈祖槃先生主编的《农业论文索引》初编和朱耀炳先生主编的《农业论文索引续编》都出版于 20 世纪 30 年代,收录了 1858~1934 年国内农业期刊(包含当时在国内出版的西文报刊)论文索引 43800 余条。第三本索引是南京农业大学王俊强老师编著的《民国时期农业论文索引》,主要收集的是 1935~1949 年这 14 年间国内所发表的农业论文。据研究,民国农业类相关期刊约为 320 种,专业期刊约为 200 多种^[12]。经笔者统计,《农业论文索引》初编收录期刊 320 种,《续编》收录期刊 553 种,《民国时期农业论文索引》收录期刊约 200 余种。前两种索引收录的期刊除了农业专业类期刊,还有不少社会、经济、时政以及综合类期刊,后一种索引收录的期刊都是农业专刊。可见选择这三本索引作为数据统计对象是合适的。为了构建数据库,笔者首先开展的一项工作是将这三本索引数字化,在其电子文档的基础上构建《农业论文索引》题录数据库,收录农业研究论文数据 48168 条(不含当时在国内出版的外文论文)。然后,根据研究主题的设定,从此数

据库中筛选出粮食作物专题研究数据 11690 条。该专题数据库涵盖了农学(育种、选种等)、土壤与肥料、水利、农机、谷物、植保等与粮食作物主题紧密相关的各核心范畴。另一方面,鉴于近代农业学校推行“教学-科研-推广”三位一体的办学体制,农业科研机构、农业试验、农业推广、灾荒等研究主题的文章也收录进该专题数据库里面。在建构粮食作物研究专题数据库的过程中,校对、勘误与查漏补缺是一个相当耗时的环节。特别是《民国时期农业论文索引》里收录的论文有不少发表时间不详,为此笔者通过与上海图书馆的《晚清民国期刊篇名数据库》以及北京大成公司推出的《大成老旧刊全文数据库》比对检索,补充了论文发表时间。剩下的仍然找不到时间的,进一步通过网络搜索,经过合理的推断,进行补充。最后,笔者从自建的《农业论文索引-粮食作物》(1897~1951)专题数据库 11690 条论文记录中提取所有涉及合作的 870 篇论文题录信息,在此基础上构建了 962 对作者两两合作关系数据,形成本研究的网络数据。

合作关系数据的确定是所有开展合著分析研究工作的前提。一般来说,若一篇文章有 $N(N \geq 2)$ 个作者,则认为存在 C_N^2 对关系,可视化图中就会出现 N 个节点 C_N^2 条边。这种设定即为合著无向网络,其特点是不区分作者的贡献度大小,默认所有作者对文章的贡献相同。这种做法侧重于对作者间整体关系和合作规律的考察,却不利于识别真正的核心作者。而相比之下,有向关系网络因为在“度”这个中心性指标之外还提供了“入度”和“出度”两个不同的测度指标,这样对作者之间的关系能做出区分,对作者的角色和声望的量化测度一般来说更加精确和恰当。鉴于国内学界绝大多数学科习惯性注重文章署名第一作者^[13],本研究突出第一作者的贡献度,以第一作者为中心节点,其他共同作者为从属节点,构建作者合著有向网络。即只构建第一作者和非第一的其他共同作者两两之间的有向合作关系,其他作者连线的箭头均指向第一作者,而其他共同作者(比如第 2/3 作者、第 2/4 作者、第 3/4 作者)之间不构建有向关系,无连线(边)。

本研究数据采用的社会网络分析可视化测度工具是 Gephi0.9.2 版。在使用软件之前,需要进行的一项数据预处理工作是将合作关系数据转化为 Gephi 能够直接读取的 CSV 数据。在用文本文件编辑器创建 CSV 格式合作关系数据时,笔者将一篇论



文的第一作者确立为目标节点(Target),非第一作者为源节点(Source),两两一对;如果论文有三个(包括3个)以上的作者,则这篇论文就产生两对及以上的合作关系。另外,由于这一段时期科技论文写作、产出的特点,存在许多“笔述、口译、录、记、讲、述”等形式的合作文章(这类文章总计88篇,约占总量的10%),为了消除语义识别的干扰,将这些字眼全部去除,统一视为单一性质的科研合著论文。这样录入的结果就是产生了一个“邻接名单”(linked list)数据表格,这个表格有两列,第一列是源节点名单,第二列是目标节点名单。这样的表格是多数社会网络分析软件读入的基本格式。在导入系统之后会自动转换为二值或多值形式的表。这个表就是软件内嵌的统计程序计算的基础,可以用它们来计算该网络的各种属性。

3 作者合著网络的实证分析

3.1 基于社会网络理论的描述性分析

3.1.1 网络整体概述

打开 Gephi 软件,导入作者合作数据,完成对节点、边、标签的图设置,运行 FR 算法(Fruchterman Reingold)布局,然后点击“统计”,显示主要参数:节点:1013,边:699,有向图,平均度=0.69,网络直径=5,图密度(网络密度)=0.001,连接组件=340,平均聚类系数=0.008,平均路径长度=1.31。全图概貌见图1(此处用弧代替有箭头的连线,顺时针表示箭头所指的方向)。据统计,《农业论文索引-粮食作物》专题数据库中署有署名作者4008位,而合著作者(节点)只有1013位,合著率只有25%,对比当代自然科学领域90%左右^[14]的合著率是很低的,而且合作关系(边)数量不多,合作范围不广(平均路径长度较短)。从图密度值、连接组件以及平均聚类系数来看,总体上作者之间的合作联系是非常稀疏的、规模分散细小的、缺乏凝聚力的。但从网络直径以及平均度可以看出,这个网络也具有典型的“小世界”、无标度、高聚集特性^[15],适合采用社会网络分析。下面运用社会网络分析理论和方法,分别从作者的中心性度量、作者两两之间合作关系强度、凝聚子群、最大连接子群四个不同层次的角度对此合著网络进行拓扑性质的分析。

3.1.2 重要作者的识别和发现

在社会网络分析中,识别重要的角色或位置是其关注的主要方面之一。节点的重要性或影响力一

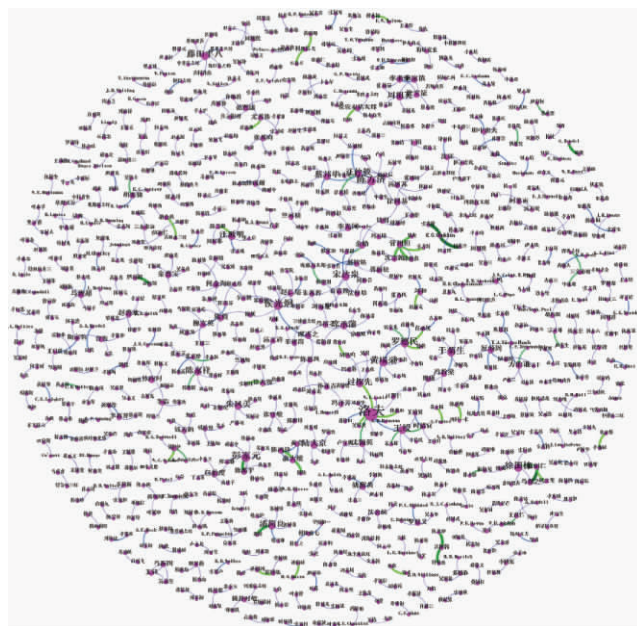


图1 1897年-1951年《农业论文索引-粮食作物》作者合作网络
般通过其“中心性”和“声望”属性指标来度量。无向网络可以通过节点的“度”考察其“中心性”,而“声望”指标只能在有向关系中通过节点的“入度”和“出度”的区分进行研究^[16]。联系社会网络分析理论,出度是扩张性的测度,而入度则是接收性或受欢迎程度的测度^[17],我们考虑论文合作关系的社会计量关系,假定多数情况下,一个有较大出度的作者是喜欢结交学术同道的人,一个有较大入度的作者是别人都喜欢与之交流合作的人,或者说,一个具有更高声望的作者^[18]。

(1) 入度值分布及高入度作者

本研究中网络节点的“入度”,就是他们作为第一作者产生的合作作者数量。运行 Gephi 软件中的“过滤”——“拓扑”——“在度的范围”,显示入度值为0-12。这说明数据库中第一作者合作最多的人数有12位。用鼠标单击并拖曳滑块设置选取的范围,进行过滤,可以得出不同合作者数量节点的总数及其占比。入度 ≥ 2 的节点有115个,在所有节点中所占的比例为11.35%;入度 ≥ 3 的节点数为44,所占比例为4.34%;入度 ≥ 4 的节点数为22,所占比例为2.17%;入度 ≥ 5 的节点数为13,所占比例为1.28%。这13位作者是作为第一作者产生论文合作关系数量最多的群体,也是整个科研交流与合作的核心力量(图2)。

鼠标切换到图形的“数据资料”,可以看到这13个节点的各项参数值,包括连入度、连出度、度、接近

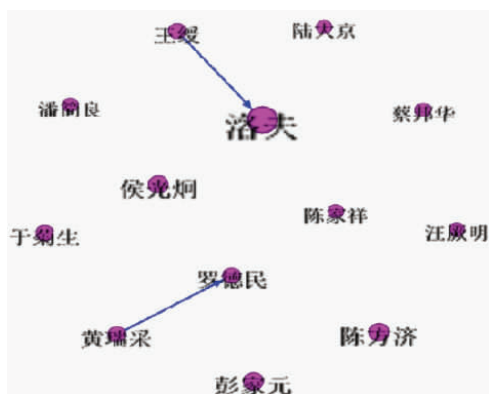


图2 入度值最高的13位作者

中心性 (Closeness Centrality)、中介中心性 (Betweenness Centrality) 以及所属的连接子群 (Component ID) 等。他们的入度值分别为: 洛夫 (H. H. Love) 第一, 12; 罗德民 (W. C. Lowdermilk)、侯光炯、彭象元并列第二, 均为 6; 其余作者均为 5。需要指出的是, 入度值只是反映第一作者产生的合作关系, 并不是合著的论文数量。这里面的作者“陆大京”实际上只合作撰写了一篇文章, 但由于这篇论文的合作者有 5 人, 是全部合作组中包含作者最多的一组, 因此合作次数比大多数仅跟 1 位或 2 位作者合作的数量反而多了。从整体上来看, 此数据库中合作发表论文的篇均作者数为 1.45, 大部分作者只跟一位作者合作发表, 跟三个以上不同作者合作发表的关系组只有 5 个, 占全部合作发表论文关系组的 0.7%, 不到百分之一。合作论文篇均作者数量的统计结果和前文由 Gephi 自动计算出来各入度值作者数量的分布是一致的, 表明入度分布的异质性较大, 反映出晚清民国这一时期科研合作的广度, 无论是科研人员的横向联系, 还是不同科研领域之间的交流合作范围, 都是比较有限的。仅有极少数有影响力的作者在促进学术的交流和发展中发挥着重要作用。

(2) 出度值分布

本研究合作网络中作者的“出度”, 是指非第一作者的作者参与的不同合作关系数量。运行 Gephi 软件中的“过滤”——“拓扑”——“超出度的范围”, 显示数据出度值范围为 0-6。这意味着, 整个网络中, 以非第一作者身份产生的合作关系最多为 6 个。各出度值节点数分布具体情况: 出度值 ≥ 2 , 有节点 75, 在全部节点中所占比例为 7.4%; 出度值 ≥ 3 , 有节点 21, 所占比例 2.07%; 出度值 ≥ 4 , 有节点 5, 所占比例为 0.49% (如图 3 所示)。最高出度值作者排名依序为: 藤田丰八, 6; 徐国栋, 5; 汪仲毅、徐树基

与宋达泉均为 4。对比前面的入度值, 可以看出, 高出度作者数量以及出度值变异程度要小得多。

通过研究我们发现, 出度和入度的 TOP 节点群没有交集, 这也验证了这两个测度的各自特点和某种程度上的差异性。如果不考虑第一作者和非第一作者的区别, 也就是说所有作者之间的关系是无向的, 没有出度和入度之分, 那就只能考察作者的中心性, 无法衡量作者的学术贡献度。



图3 出度值最高的5位作者

3.1.3 合著关系强度

在社会网络分析中, 边就是两个节点之间的连线。在本研究中, 两位作者之间的合作关系就构成一条边, 他们的合作次数就是“边的权重”。对他们合作关系的次数考察实际上就是对其合作关系的力度和强度的考察。在 Gephi 软件里, 运行 Gephi 软件中的“过滤”——“边”——“边的权重”, 显示值范围为: 1.0-7.0。分别设置不同的权重得到如下数据 (表 1):

表1 作者合作网络边的权重数量分布及其所占比例

边的权重	显示的边的数量	所占总边数的比例
2-7	143	20.46%
3-7	58	8.3%
4-7	37	5.29%
5-7	13	1.86%

切换到 Gephi 关于此图的统计资料页面, 可以发现, 在合作 4 次以上的 37 组边当中, 第一作者为外国学者的有 20 组, 占高频合作组的一半以上; 其中, 第一作者为欧美学者的有 15 组, 为日本学者的有 5 组。有数位外国学者和不同的作者合作了 2 次以上, 例如学者洛夫 (美国) 合作的高频组有 3 组, 罗德民 (美国) 有 2 组, 村越信夫 (日本) 有 2 组。

观察边的权重为 5-7 范围的可视图 (图 4) 和相应数据内容, 可以发现, 合作 5 次以上的 13 组边中, 第一作者为外国学者的只有 5 组, 分别是康克林 (E. G. Conklin) - 何定杰、康克林 - 张光耀、孟德尔 (G.



Mendel) - 顾复、博德 (R. H. Porter) - 俞大绂、魏庚 (R. G. Wiggins) - 沈宗翰。其中,康克林-何定杰、康克林-张光耀这两组边实际是三位作者合作了一篇译作,康克林是原著者,其他两位是译作者,文章题为《遗传与环境图》,在《学艺》杂志上连载7次。孟德尔-顾复,也是译作合作模式,他们合作的题名为《植物杂种之实验》,在《学艺》上连载6次。博德-俞大绂也是译作的合作模式,但是他们翻译了《小麦线虫病和他的防除法》以及《高粱粒黑穗病和防除的方法》等多篇文章。魏庚-沈宗翰这一合作组是第一作者为外国作者中唯一合作撰写学术论文的组合,但是他们仅合著了一篇论文《华北农业视察报告》,这篇论文在1930~1931年期间在《中华农学会报》《科学》《农林新报》等5种杂志上刊发、转载了,所以记合作频次为5。

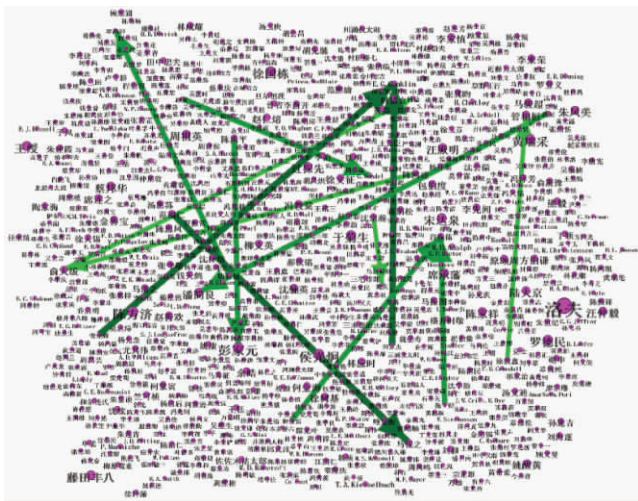


图4 合作频次最高的13组合作关系

而在中国本土作者构成的合作边中,排除一篇文章连载/转载的情况,真正合作高产的组是马保之-范福仁,他们二位合作撰写了《作物育种应有之认识》《谷类作物之植物性状对产量之相互关系》《广西引种美国杂交玉蜀黍结果简报》等7篇论文。其次是彭家元-陈禹平、潘简良-龚弼这两组,各自合作撰写了6篇题名不同的论文。再往下,管相桓-涂敦鑫撰写了5篇论文,俞大绂-陈鸿逵撰写了4篇论文。

综上所述,我国学者和国外学者广泛合作,但是从合作的频度、深度来讲,国际合作远逊于国内合作。由此可见,在中国近代农业粮食作物研究领域,从欧洲、美国、日本等西方科技先进国家引进、译介而形成的学术成果占了相当比重;但是在向科研纵深发展方面,我国本土的学者还是当仁不让地居于

主导地位。

3.1.4 合作网络中凝聚子群的探测分析

凝聚子群是社会网络分析关注的一个重要概念。所谓凝聚子群,是相互之间有着稳定、直接、强烈、频繁或正向联系的行动者子集^[19]。依据社会网络分析理论,K-核是侧重于节点度来研究凝聚子群的一种方法。K-核中的K表示在互相连接的节点群中,每个节点的度都不小于K值。本研究发现的凝聚子群均为2-核(在有向图中,如果两个节点之间有双向的边,也就是说每个节点都有一条出向边和一条入向边,那么这两个节点的K-核也是2-核)。具体运行 Gephi 软件里面的“过滤”——“拓扑”——“K-核”,将K值更改为2,系统过滤显示出有13个2-核的凝聚子群(图5)。

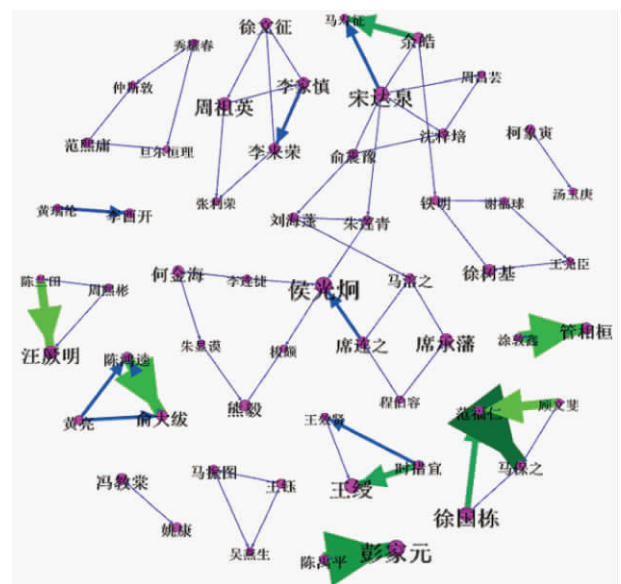


图5 作者合作网络中13个2-核凝聚子群

具体对这13个2-核凝聚子群展开分析,可以看到,有5个由双向边构成的2节点子群,4个3节点构成的2-核子群,2个4节点构成的2-核子群,1个5节点构成的2-核子群,以及1个22个节点构成的最大的2-核子群。这些凝聚子群出现的时间不同、规模不等,其凝聚程度和关联强度也是各不相同。比如马保之、范福仁、徐国栋、顾文斐这一个4节点组是联系直接、稳固,且合著次数最频繁的团体。而对比之下,秀耀春、仲斯敦、旦尔恒理、范熙庸这个同样的4节点组的合作关系就稀疏得多。联系到合作团体存续的时间,前者的合作年代为20世纪30、40年代,后者的合作年代为19世纪末20世纪初的晚清时期,两者之间凝聚程度的差别,在某种程度

上反映了我国近代科学合作研究在不同时期、不同发展阶段具有不同的特点。不过从合作关系的方向来看,这些凝聚子群基本上属于弱连接关系,即便有双向合作,合作的次数也没有完全相同的。

值得重点考察的是最大的 2-核子群。仔细研究发现,这个最大 2-核凝聚子群中许多节点的中介中心度较高,如宋达泉、侯光炯、朱莲青、余皓、何金海、熊毅、席承藩、席连之。社会网络分析大师弗里曼认为行动者的中介中心度表示的是一个点在多大程度上位于图中其他点的“中间”,起到“中介”作用。另一位大师伯特(Burt)用“结构洞”(structural holes)概念也对此进行了描述。当两个点以距离 2 而不是 1 相连的时候,就可以说二者之间存在一个结构洞^[20]。由此可见,中介中心度是测量位于两个连接路径之间的行动者角色作用的指标。具有较大中介中心度的作者,毫无疑问相对于其连接着的其他作者占据着重要的位置。在笔者统计的整体网络全部 1013 节点的中介中心性数据中,有 32 个节点中介中心性 ≥ 3.5 ,而这个凝聚子群拥有 8 个高中介中心度作者,占了全网络所有高中介性节点数的 25%。它实际上也发展为全合作网络中最大的连接子群,在笔者参考的 Gephi 使用教程里面被称为“巨人组件”^[21]。后面将对其展开专门分析(主要是关于作者的角色和地位分析)。不仅如此,这个最大的合作团体,也产生了 7 位高产作者(核心作者),他们是马寿征、余皓、宋达泉、侯光炯、熊毅、马溶之、席承藩,高产作者在所属团体中的占比为 32%。而所有 13 个 2-核合作团体共产生了 19 位高产作者,占团体总作者数 57 位的 33%。经统计,含 1013 位作者的整个合作网络共产生 124 位高产作者,高产作者占比是 12%。没有参加论文合作团体、开展独立研究的高产作者是 68 位,占有所有独立作者总数(2995)的 2%左右。可见凝聚子群中的核心作者比例远远高出一般合作团体(更勿论独立研究作者群)的核心作者比例。

从前面获得的网络整体参数可知:(1) 整个合作网络共存在 340 个连接子群,1013 位作者,平均每个连接子群的作者数约为 2.98 人;(2) 绝大多数合作团体规模很小,在 4 人以下,占有合作团体的 96.76%;(3) 合作团体规模在 5 人以上的共有 11 个,在 10 人以上的只有 6 个,其中最大规模的团体由 44 位作者组成,规模排名第二的团体由 33 位作者组成。而在这 340 个合作团体中,仅发现 13 个 2-

核凝聚子群,这表明中国近代粮食作物研究科研合作团体的规模总体上相当细小,而且呈高度分散性,内部缺乏凝聚力;全国范围内只零星分布着极少数具有较强向心力的合作团体,它们规模不等,成员之间的联系较为直接、稳定,合作频繁、密切。这些凝聚子群,尤其是那些合作关系密切、稳定,联系广泛、活跃的子群,在推动学术成果的高产、孕育核心作者方面产生了重要作用。

3.1.5 合作网络中最大连接子群的内在结构

由社会网络分析里图论的概念可知,最大连接子群是指在一个图中,占有节点数比例最高的连通子图。运行 Gephi 软件里面的“过滤”——“拓扑”——“巨人组件”,选择布局中的“Force Atlas2”对其进行设置,得到最大连接子群的图形(图6)。

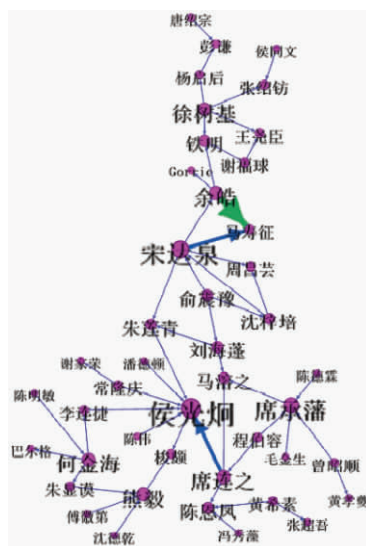


图 6 合著网络中最大的连接子群

这个子群包含节点 44 个,4.34%可视,边 51 条,7.3%可视。从图中可以看到,这个子群的节点度数以及边的权重比较均衡,彼此差异幅度不大。需要指出的是,这个节点群不是由一个时间节点连接而成的,是跨越了半个世纪的前后相继的联系、连接,有些节点在前期出现,但到了后期就湮没不见了。另外,正如前文所述,子群能发展到最大规模,依赖于众多高中介中心性节点的桥梁作用。但是在有向有值合著网络中,拥有高中介性的节点并不一定是当然的中心节点。从图论的节点中心性关于个体行动者角色和地位的等价性具备要素(如相同或相似的入度、出度,相同或相似的中介中心度或接近中心度等)分析,可以发现侯光炯、宋达泉、席承藩、熊毅等是度数较高的“明星”节点。但是我们通过构造节点之间的 1-模网

2019年第4期

大學圖書館學報



络无向二分数据矩阵,导入统计软件 SPSS22 对该最大连接子网的所有行动者进行分层聚类分析,得到的结果就和图论对个体角色测度的结论有所不同。在聚类显示的树状图(图 7)里,节点间连线越短表示他们之间的结构(地位)等价性越高。因此,侯光炯与谢家荣,宋达泉与刘海蓬、沈梓培,席承藩与黄孝夔,熊毅与何金海等,他们两两之间在合作网络中扮演的角色和占据的社会地位具有更大的相似性。

3.2 核心作者发文量与其合著论文数量呈现显著相关性

《农业论文索引-粮食作物》专题数据库中包含 4008 位作者以及他们贡献的 9428 篇论文(不含匿

名的论文)。平均每位作者发表论文 2.35 篇。撰写一篇论文的作者一共有 2560 位,约占作者总数的 63.9%,比较符合文献生产的洛特卡定律。本研究依据普赖斯定律衍生的核心作者公式确定该数据库中发表论文 8 篇以上的作者为核心作者,共计 192 位。这 192 位作者占总作者数的 4.8%,但是他们发表论文的数量有 2983 篇,占总发文量的 31.6%。笔者对这 4008 位作者做了一个发文量的排行榜;同时通过作者合著数据产生了一个合著作者排行榜。在此,本研究以发文量排行榜前 192 位作者为观测值,

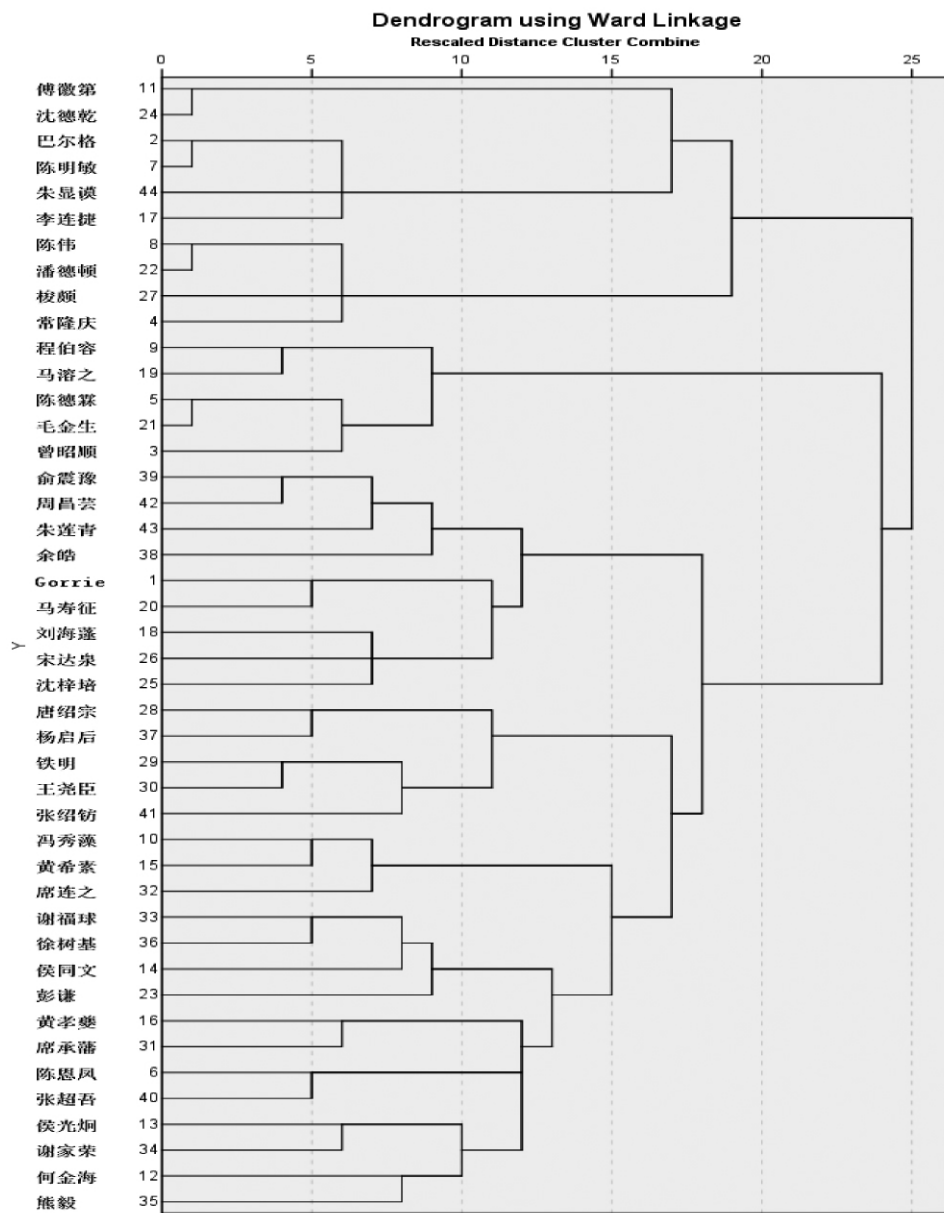


图 7 最大合作子群中作者结构等价性聚类分析谱系图



以发文量、入度、总度数为变量,运用 SPSS 软件做了发文量与合著网络总度数以及入度的两两相关性分析并进行假设检验。由于这些变量属于离散型数据,因此采用 Pearson 相关系数进行双变量相关性分析和双侧 t 检验。从相关性分析结果来看,发文数量与总度数的相关性略高于与入度的相关性,两组相关性都是在 $\alpha=0.01$ 的水平上显著相关(如表 2 所示)。总体上可见,作者的发文量与其合作发表论文的数量具有确定的正向相关性。在发文量排名前 50 的高产作者(表 3)当中,只有 10 位合作度为零的独立研究作者,其余都是合作研究作者;在 40 位合著高产作者中,有 22 位作者同时也是论文合作度数排名居于前 50 之列的高中心性作者,如著名学者藤田丰八(日本)、王绶、丁颖、彭家元、沈宗瀚、陈芳济、洛夫(美国)、过探先、朱凤美、胡先骕、俞大绂、原颂周、金善宝等。

表 2 核心作者发文量与度数、入度相关性分析结果

	文章数量	度
文章数量 皮尔森 (Pearson) 相关显著性 (双尾)	1	.363 **
N	192	192
度 皮尔森 (Pearson) 相关显著性 (双尾)	.363 **	1
N	192	192

** .相关性在 0.01 层上显著 (双尾)。

	文章数量	入度
文章数量 皮尔森 (Pearson) 相关显著性 (双尾)	1	.245 **
N	192	192
入度 皮尔森 (Pearson) 相关显著性 (双尾)	.245 **	1
N	192	192

** .相关性在 0.01 层上显著 (双尾)。

表 3 排名前 50 的作者发文量及其度中心性数值

作者	发文量	入度	出度	度	作者	发文量	入度	出度	度
藤田丰八	100	0	6	6	过探先	25	4	1	5
王绶	59	5	1	6	朱凤美	25	4	1	5
丁颖	54	3	0	3	万国鼎	25	1	0	1
李积新	54	1	1	2	李仪祉	24	0	0	0
彭家元	40	6	1	7	胡先骕	23	0	3	3
沈宗瀚	39	2	1	3	陈寿彭	23	0	0	0
周明懿	39	1	0	1	周长信	23	0	0	0
周建侯	38	2	0	2	俞大绂	22	2	2	4
凌道扬	35	1	0	1	刘和	22	3	0	3

作者	发文量	入度	出度	度	作者	发文量	入度	出度	度
陈方济	34	5	2	7	朱学曾	22	0	1	1
徐国栋	33	1	5	6	乔启明	21	1	1	2
薛德焯	32	0	0	0	顾复	21	0	1	1
洛夫	30	12	0	12	邵仲香	21	0	0	0
潘简良	30	5	0	5	徐世大	21	0	0	0
邓植仪	30	1	0	1	原颂周	20	4	0	4
莫定森	30	1	0	1	罗登义	20	1	2	3
汪仲毅	29	1	4	5	黄瑞采	19	5	1	6
邹钟琳	29	0	0	0	金善宝	19	4	0	4
马骏超	28	4	0	4	张巨伯	19	1	0	1
蔡邦华	27	5	0	5	吴福楨	18	1	1	2
蓝梦九	27	0	1	1	任明道	18	0	1	1
条衣	27	0	0	0	李映惠	18	0	0	0
任承统	26	1	1	2	范福仁	17	2	1	3
金孟肖	26	1	0	1	马保之	17	3	0	3
唐启宇	26	0	0	0	杨开渠	17	2	0	2

4 结论与讨论

本研究主要运用社会网络分析理论和方法对中国近代科学(以农业科学粮食作物这一具体领域为例)论文合著者群体规模、结构以及他们交流的模式、特征进行了实证分析。从中我们可以得出:中国近代科学科研活动已经产生由分散的合作者群体以及互相联系交叉的合作者群体组成的学术共同体,这对于科学的知识积累和增长是有显著推进作用的;享有高声誉/高中心性的核心学者在形成高水平合作研究群体中具有显著作用;这些核心领袖学者长时间保持了对合作小组的奉献精神是维持高水平合作研究的重要特征;高产作者往往比同一领域其他作者具有更多合作关系,而且,他们与其中某些合作者形成的长期紧密的合作关系,对于他们在领域中的深耕和拓展具有重要意义。在某些学科中,许多学者在一个以上的领域进行研究,与其他领域保持较高联系,这转面也促成了他们较高水平的学术产出。此外,这个学术共同体网络还是一个国际性的网络,参与这一网络最多的国家是美国和日本,但是这两个国家参与的研究领域以及合作的程度是具有明显区别的。

另外,本文从社会网络视角和统计学方法对合著网络的重要指标,如中心性、声望、凝聚子群、有价值关系、角色与地位等的分析与阐释,一方面可以修订以往同类主题定性研究中小样本的个案取向,在今后的典型个案研究中扩大关注范围,比如,以前被忽略的地区、机构、团体、作者等;另一方面也通过这一新的研究方法得到的结论的呈现,给其他研究路径



得出的结论提供了一个可以比较、相互补充的蓝本。

关于本研究的意义之几点说明:

(1) 首先,社会网络分析有一套既定的操作性定义与测量方法,研究者出于研究目标的考虑,将分散的个体聚集成一个特定的群体或者区域,据此确定的边界可能完全是人为的,该区域之外的关系被忽略了,就这一点来说,社会网络分析是对整体网的不完美的表达。

其次,社会网络分析关于某一具体研究对象往往提出了多种不同的模型和测度方法,因此研究者需要根据研究的目标和内容,进行取舍和界定,这在某种程度上也是一种人为的、先验的行为,需要研究者在理论和经验上给出可辩护的原因。

(2) 此外,科研合作关系研究的是科学生产过程中的共享关系,对合作关系的研究跟对引文关系研究一样,能够告诉我们大量科学专业的社会结构方面的信息。但是,正如社会网络分析权威学者指出(霍华德·怀特,White,2011)的那样,不能用它们来展示个体、院系的生产性、质量或影响,并且它们一点也不代替同行评审^[22]。

但是,本项研究也证实了“图或者社会关系图所提供数据的可视化显示经常能让研究者发现不如此就难以发现的模式^[23]。”因此,上述基于社会网络分析理论对中国近代科学科研合作现象的研究,是对中国近代科学科研群体学术合作与交流演进特点、模式与评价研究的有益补充,而且在将社会网络概念和方法整合普及到科学史、学术史等人文社会科学研究领域方面提供了一个探索性的尝试。

参考文献

- 1 罗家德.社会网分析讲义(第二版)[M].北京:社会科学文献出版社,2010:4-5.
- 2 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,2009:37.
- 3 王柏,吴巍,徐超群,吴斌.复杂网络可视化研究综述[J].计算机

- 科学,2007,34(4):17-23.
- 4 刘则渊,尹丽春,徐大伟.试论复杂网络分析方法在合作研究中的应用[J].科技管理研究,2005(12):267-269,273.
- 5 [美]黛安娜·克·无形学院[M].刘珺珺,顾昕,王德禄,译.北京:华夏出版社,1988:129.
- 6 庞弘桑,方曙,杨波,付鑫金.科研团队合作紧密度的分析研究——以大连理工大学 WISE 实验室为例[J].图书情报工作,2010,55(4):28-32,99.
- 7 Said Y H, Wegman E J, Sharabati W K, et al. RETRACTED: Social networks of author - coauthor relationships [J]. Computational Statistics & Data Analysis, 2008, 52(4): 2177-2184.
- 8 李亮,朱庆华.社会网络分析法在合著分析中的实证研究[J].情报科学,2008(4):549-555.
- 9 雷银枝,宋歌.科学共同体测度方法的改进与实证研究[J].科学学与科学技术管理,2011,32(4):13-19,97.
- 10 邱均平,翟辉.我国科研机构合作网络知识扩散研究——以“生物多样性”研究为例[J].图书情报知识,2011(6):5-11.
- 11 朱云霞,魏建香.我国高校社会科学领域科研合作网络分析[J].情报科学,32(3):144-149.
- 12 王雅戈,侯汉清,黄建年.民国文献数字化整理研究——以民国农业文献数字化整理为例[J].图书情报工作,2009(11):63-66,148.
- 13 雷雪,王立学,曾建勋.作者有向合著网络构建与分析[J].图书情报工作,2015,59(5):94-99.
- 14 刘彦庆.我国20所高校自然科学学报的统计分析[J].情报杂志,2000,19(2):69-71.
- 15 尹丽春.科学学知识图谱[M].大连:大连理工大学出版社,2008:5,13.
- 16 斯坦利·沃瑟曼,凯瑟琳·福斯特.社会网络分析:方法与应用[M].陈禹,孙彩虹,译.北京:中国人民大学出版社,2012:125.
- 17 同[16]:92.
- 18 同[16]:125.
- 19 同[16]:183.
- 20 约翰·斯科特.社会网络分析[M].刘军,译.重庆:重庆大学出版社,2016:95.
- 21 刘勇,杜一.网络数据可视化与分析利器:Gephi 中文教程[M].北京:电子工业出版社,2017:127.
- 22 同[20]:129.
- 23 同[16]:69.

作者单位:陈海珠,包平,南京农业大学中华农业文明研究院,南京,210095

陈海珠,南京财经大学图书馆,南京,210023

李树青,南京财经大学信息工程学院,南京,210023

汪圣忠,南京财经大学经济学院,南京,210023

收稿日期:2018年9月4日

An Empirical Study of Co-authorship Network in Modern Chinese Agricultural Science Using Social Network Analysis

Chen Haizhu Li Shuqing Wang Shengzhong Bao Ping

Abstract: The author cooperation in cereals field derived from Index of Agricultural Papers (1897-1951) has been collected and arranged and a network of co-authors has been established. Social Network Analysis combined with statistical analysis methods is used to analyze the scientific collaboration situation in Modern China. Current status and characteristics of Modern Chinese agricultural science co-authorship network have been summarized and the significance of such research has been also discussed in the end of this paper.

Keywords: Co-authorship Analysis; Social Network Analysis; Author Cooperation Networks; Information Visualization; Modern China