

作者共现网络的科学研究结构揭示能力比较研究*

邱均平 董克

摘 要 作者共现网络在科学研究领域结构分析中具有非常重要的作用,本文构建了作者合作、作者共被引、作者文献耦合、作者关键词耦合及作者期刊耦合五种网络,利用层次聚类和网络关联分析比较不同共现网络揭示科学研究结构的能力。通过层次聚类发现:作者文献耦合网络的准确性和精确性最高,作者共被引网络和作者合作网络其次,作者关键词耦合网络的结果容易受到干扰,作者期刊耦合网络的分析效果较差。通过网络关联分析发现:作者合作网络与其他网络的相关性都相对较低,与其他网络发现的角度不同;作者文献耦合网络与其他几种网络之间的相关性最高,对其他网络分析结果的替代性较强,多种方法结合运用更能反映科学研究的实际结构。本文还讨论了共现网络子结构研究中凝聚子群分析及中心性指标的适用性问题。图5。表4。参考文献16。

关键词 作者共现网络 科学结构 子结构分析 关联分析 数据清洗

分类号 G350

A Comparative Study on the Ability of Author Co-occurrence Network in Revealing Scientific Structure

Qiu Junping & Dong Ke

ABSTRACT Co-occurrence network at author-level plays a pivotal role in the structural analysis of intellectual and scientific domains. Five kinds of co-occurrence networks at author-level were constructed in this paper: co-authorship, author co-citation, author bibliographic coupling, keyword-based author coupling, journal-based author coupling. Comparative analysis of hierarchical clustering results showed that author bibliographic coupling, followed by author co-citation, co-authorship, keyword-based author coupling and journal-based author coupling, was proved to be the most accurate and effective way for discovering sub-structure of scientific research. QAP correlation analysis showed that the correlation of co-authorship with other networks was the lowest, and had a different occurrence perspective from other networks, while the correlation of author bibliographic coupling with other networks was the highest and had stronger substitutability for results of other networks. The multiple use of networks is suggested as a better choice for finding real structure of scientific research domains. Finally the paper discusses some specific problems about the fitness of cohesive subgroups analysis and centrality indicators for sub-network analysis in author co-occurrence networks. 5 figs. 4 tabs. 16 refs.

KEY WORDS Author co-occurrence network. Scientific structure. Sub-network analysis. Correlation analysis. Data cleanup.

* 本文系国家社科基金重大项目“基于语义的馆藏资源深度聚合与可视化展示研究”(编号:11&ZD152)的研究成果之一。

通讯作者:邱均平,Email: jpqiu@whu.edu.cn

1 引言

文献计量中的共现关系表现为多个层面和多种类型,有学者将基于网络分析的文献计量分解为集合、共现网络、方法三个层面,集合层面包括论文、作者、期刊和机构,共现网络包括引用、共被引、耦合、合作等,方法层分为宏观、中观与微观^[1]。文献计量学是研究科学交流的主要方法^[2],而作者则是学术交流的主体^[3],无论是论文、期刊还是机构之间都无法主动发生关联,其基本依托的是作者之间不同类型的学术交流,作者层面的共现网络是各种类型共现关系分析的基础,处于核心位置;而从文献计量的层面来看,科学研究的领域结构也主要体现在科学论文作者之间共现网络的结构。

2 基本问题阐述

2.1 作者共现网络

作者共现网络主要包括合作、作者共被引和作者文献耦合。Newman^[4]认为作者合作网络不存在规模受限、数据过于实验室化、数据获取存在主观性等缺陷,因此是研究社会网络的理想载体;作者共被引网络起源于 White 和 Griffith^[5]的研究,目前的热点包括了相似性测度^[6]、结果的可视化^[7]等内容;作者文献耦合是由文献耦合拓展而来的一个新概念,目前的相关研究还比较少^[8]。耦合形式的共现关系拓展性很强,从形式上来看,作者之间通过使用相同的关键词能够产生作者关键词耦合,同在某些固定期刊上发表文献则会产生作者期刊耦合,这些耦合形式如何反应科学研究的知识结构同样具有重要的探索意义。因此,本研究构建了作者合作、作者共被引、作者文献耦合、作者关键词耦合、作者期刊耦合五种作者共现网络作为研究对象。在目前的研究中,作者共现网络在科学知识结构发现中的应用已十分广泛,但对各种方法之间的差异、方法适用性的遴选标准的研究则相对缺乏;通过层次聚类获得网络结构分析结果,比较不同网络在科学研究结构揭示能力上的特征,进一步利用

相关性分析挖掘不同类型作者共现关系之间的潜在关联,有利于探索如何综合运用作者共现网络更客观地反映科学研究的实际知识结构。

2.2 网络构建及关系强度

由矩阵所表现的网络,从结构上可以分成1模网络和2模网络^[9]。1模网络指仅对一个行动者集合进行测量所构成的网络,2模网络测量的是两个行动者集合或者一个行动者集合与一个事件集的关系,本研究中的五种网络都是1模网络。五种网络本身的差异比较明显,事实上其内在的共性也极强,其网络形成的过程可用通过2模网络向1模网络转化的过程进行统一描述。假设矩阵A为行动者集合*事件集合所构成的2模网络矩阵,行为行动者集合列为事件集合,定义 A^T 为A的转置矩阵,P为行动者之间的1模矩阵,则有:

$$P = A \times A^T \quad (\text{公式1})$$

假设行动者集合为作者,如果这里的事件是作者发表的文献,得到的矩阵P为作者合作网络;如果是引用这些作者所发表成果的文献,则矩阵P为作者共被引网络;如果是作者发表论文中的参考文献,则矩阵P为作者文献耦合网络;如果是关键词,则P为基于词的作者耦合网络(也有文献称之为作者关键词耦合);如果是期刊,则P为基于期刊的作者耦合网络。公式1实际上是对称型共现关系网络的统一构建公式,一般文献计量分析中对称形式的共现网络如共词等都可以用公式1进行构建。

共现网络构建的另一个关键问题是计算共现关系强度,主要体现在矩阵中每个位置的取值方法。马瑞敏^[8]将目前常用的取值方法归纳为三种:简单算法,即不考虑重复出现的情况,如同一作者在文献中多次引用相同的文献,在计算该作者的参考文献集时进行去重,在此基础上计算两两作者间的累计共现次数;最小值算法,考虑重复出现的情况,每出现一次均计一次,取两两作者在同一施引文献、参考文献或关键词上的较小值,然后进行累计;组合算法,它和最小值算法的不同点在于取两两作者在同一施引文献、参考文献或关键词上的积,然后进行累计。相比较而言,最小值算法更能

反映作者之间的实际相似程度,因此本文主要采用了最小值算法。由于最小值算法可以直接在 2 模网络向 1 模网络转化的过程中实施^[10],可以避免直接获取作者共现关系矩阵时对每个位置进行逐一判断,因此也大大提高了数据处理和分析的效率。

3 研究设计

3.1 样本选取与数据清洗

本研究选取了 2011 版 JCR 中收录的 31 种影响因子大于 1 的图书情报领域期刊为基本的数据来源^[11],在 Web of Science(WOS)平台上下载了这 31 种期刊所刊载的论文、会议论文、综述三种类型的文献为基本分析对象,检索时间为 2012 年 12 月 5 日。期刊改名的情况在数据采集过程中也被考虑到,如 JASIS 改名为 JASIST。通过对数据的基本清洗,共获取文献 25,652 篇,参考文献记录 499,986

条,其中第一作者 14,875 名,全部作者 32,268 名,被引作者 214,887 名。

作者层面的共现关系数据获取的最大困难是数据噪声的干扰。WOS 已经意识到相应的问题,因此近年来采取了相关的手段进行改进,如在作者姓名的简称后补充全称,增加作者的联系信息(主要是 E-mail 和机构信息)以提高作者身份确定的准确率等。在 WOS 系统中,所有被收录的文献如果被引用,则出现在参考文献中时均会有超链接,但完全通过系统进行检索来获取共现矩阵中每个位置的值并不现实,100 名作者的共现矩阵就要进行 10,000 次检索。文本处理仍是获取共现关系矩阵的主要途径,因此数据清洗工作显得尤为重要。根据研究需要,首先对作者字段和被引作者字段存在的问题进行分析,发现作者字段同人异名的情况较少,而在被引作者标注中问题则比较严重,表 1 以作者 James J. Cimino 为例展示了三种主要不规范问题,这些问题主要集中在被引作者姓名的标注中。

表 1 作者姓名著录的不规范问题

作者字段常用形式	被引作者字段常用形式	存在问题	不规范形式
James J. Cimino	Cimino J J	空格的误用	Cimino JJ
		全称、简称的混用	Cimino James J
		省略符号“.”的使用	Cimino J. J

除此以外,还存在姓名标注颠倒的情况,如 White HD 和 White DH 等,主要在中韩学者中体现较为明显,本研究中涉及较少。针对上述提到的几种主要问题,利用正则表达式对样本数据进行清洗,得到第一作者 14,151 名,全部作者 30,908 名、被引作者 172,295 名。由于 WOS 数据库中参考文献仅著录第一作者,因此选择作者样本时主要考虑第一作者发文数及被引频次;根据以往进行计量分析及可视化的经验,许多相关论文均认为节点规模在 100 左右网络可视化效果最优,因此本研究确定发文数量 10 篇以上、被引 160 次以上的 98 位作者之间的五种共现关系为分析对象。在作者关键词耦合关系构建时,主要通过通过对 keywords、keywords plus 及 title 字段中出现次数在五次以上的词进行

清洗,最终获取 1,484 个关键词用于网络的构建。

3.2 方法与工具

共现关系的子结构实际上反映了学科研究的子领域。在子结构分析方面,采用目前在共现分析中被广泛使用的层次聚类法对 98 个作者依据其相似度进行聚类,最终得到不同类型共现关系所代表的领域子结构,并通过对比,分析不同类型的共现关系在学科领域结构揭示能力上的差异;在聚类过程中则主要通过计算不同作者之间的余弦系数作为聚类的依据。作者共现关系之间的相关性实际上反映了学科内部不同交流模式之间的潜在关联,通过对不同类型的共现关系进行相关性分析,能够测量不同共现关系之间的嵌入程度,分析方法主要

采用 QAP (二次指派程序)^[12]。QAP 算法根据输入的两个矩阵计算出一个初始的 Person 相关系数,然后将其与随机置换多次后的 Person 相关系数进行不断的比较,p 值为多次(本研究设定为 5,000 次)随机置换后所得相关系数等于或大于最初相关系数的数量占全部相关系数的比例,根据一般惯例,本研究设定 p 值在 0.05 水平上显著。

研究过程中采用的工具主要包括汤姆森数据分析器 TDA、社会科学统计软件包 SPSS、社会网络分析软件 Ucinet 以及可视化软件 VOSViewer^[13]。前三者在以往的研究中被广泛使用,而 VOSViewer 则是目前在计量分析中应用较为广泛的可视化软

件,该软件由荷兰莱顿大学科学技术研究中心(CTWS)资助开发,主要运用了基于相似性的可视化(Visualization of Similarities)思想,其可视化结果比较利于本研究的解释。

4 网络结构分析

4.1 网络概况

密度是指网络中实际存在的连线占所有可能出现的连线的百分比,表 2 显示了五种类型网络中边的数量及网络整体的密度。

表 2 网络的规模和密度

网络类型	边数量	密度
作者合作	81	0.0170
作者共被引	3,092	0.6505
作者文献耦合	3,457	0.7273
作者关键词耦合	2,518	0.5298
作者期刊耦合	3,344	0.6964

由于五种网络都是由 98 个作者节点所组成的,因此不存在由于网络规模本身的变化对密度产生的影响。五种网络中除了作者合作网络外其他几种密度均非常大,网络中的连线数量很多,许多网络分析及可视化专家均认为 100 数量级及以上规模的网络如果密度较大,很难通过可视化的方式对网络的细节进行描述,而 VOSViewer 能够较好地淡化由于边过多而对可视化结果的影响。

4.2 网络子结构

图 1 是合作网络结构分析的可视化结果。网络中节点的大小表示作者的度(即发文量),编号代表了不同的聚类结果。整个网络中一共有五个明显的聚类,其中最大的一个团体类 1 包含了 38 个作者,分为三个比较明显的部分:第一部分包括了 Egghe L, Leydesdroff L, Braun T, Rousseau T, Thelwall M 等 LIS 领域中的高产作者,成员数量很多,其研究的内容主要包括文献计量、信息计量、科学计量和网络计量;

第二部分则以 Ingwersen P, Croft W B 为核心,成员包括了 Borgman C L, Belkin N J, Rebertson S E 等人,这部分作者研究的主题主要是围绕信息检索;第三部分以 Spink A 为核心,包括 Wilson TD, Ford N, Ellis D, Cole C 等人,主要研究与信息行为相关的内容。三个部分的衔接人物研究内容的交叉性也非常明显,如 Ingwersen P 曾于 2005 年因其在科学计量学和网络计量学方面的贡献而被授予普赖斯奖,同时他也是一位信息检索领域的教授;又如 Saracevic T 的研究同时涉及信息检索与用户信息行为分析。此外,规模较大的还有以医学信息学研究为主的团队聚类 2,该类以 Friedman C 和 Cimino J J 为核心;聚类 3 集中了四名作者,包括 Venkatesh V, Dennis A R, Agarwal R, Rai A, 他们大多都是商学院的教授,研究的主要领域为管理信息系统;聚类 4 包含了 Grover V 为核心的四位作者,这几位作者的研究则明显偏重于信息技术方向;聚类 5 则包含了大量彼此之间不存在合作关系的作者。

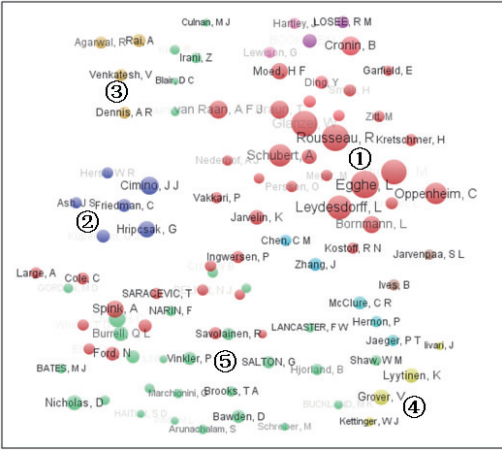


图1 合作网络结构

作者共被引结构分析的可视化结果如图2所示。图中度数较大的节点主要包括了Garfield E, Leydesdorff L, Thelwall M, Egghe L, Salton G, Braun T和Kostoff R N等,不同聚类间的边界也较为清晰。包含作者数量最多的聚类1结果十分零散,如包括了Cimino J J等生物信息学领域的专家和以Venkatesh V为中心的管理信息系统领域的专家,该聚类中作者群整体特点并不突出,经过分析发现这种情况的产生主要是由于这些作者相对于其他聚类中的作者而言被引次数较低。聚类2则主要是由信息计量学和科学计量学领域的专家构成,彼此之间的联系较强,相关作者的研究主题包括了引文分析、文献计量指标、网络计量、专利计量等。聚类3以Salton G为核心,集中了信息检索领域的主要研究人员,而聚类4中则主要集中了以Kostoff R N为中心的文本挖掘和知识发现等研究领域的作者。除聚类1以外,其他三个聚类中作者研究内容的确定均比较容易。

作者文献耦合网络的聚类结果如图3所示。与共被引相比,比较明显的区别是Egghe L, Rousseau R和Bermann L等六个作者构成了聚类2,与以Leydesdorff L为核心的聚类1分离开。聚类1中的核心作者群体则更多地关注应用计量学方法进行学科结构分析的研究,而聚类2中的六位作者因大量引用关于h-index研究、文献计量中的数

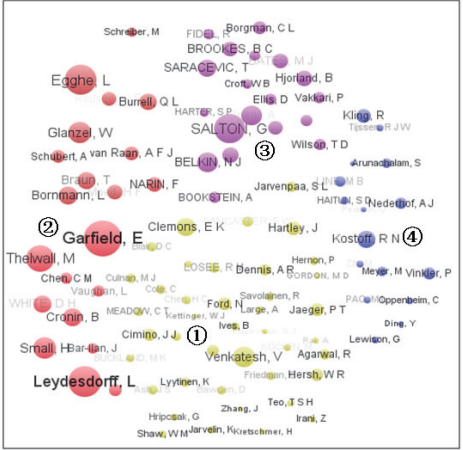


图2 共被引网络结构

学定律方面的文章而耦合在一起。Small H, White H D等作者构成聚类3,这几个作者大多数来自美国的大学或研究机构,如爵硕大学,ISI公司,印第安纳大学伯明顿分校等,彼此之间通过大量关于共现分析主题的文章耦合在了一起。这三个聚类大致上都在进行与计量学相关的研究,但是彼此之间的研究主题在更细微的层面上有所区别。Thelwall M等五个作者构成的聚类4规模较小,但是由于影响力较大,在聚类分析的可视化结果中也相对明显,这群作者主要通过研究网络计量、链接分析、网络环境下信息计量学应用方面的文章耦合在一起。位于顶端Salton G为中心的聚类5主要是信息检索技术领域的学者,以Cimino J J为中心的医学信息学研究群体聚类6和以Spink A等以用户研究和信息行为研究为主的群体聚类7,在作者文献耦合网络中也很好地被区分了出来。聚类8中作者主要属于管理学领域,由于耦合强度依赖于作者所发表文章参考文献交集的大小,这些作者均拥有很多相同数量的参考文献,因此这些节点度均较大。

作者关键词耦合是使用作者在表达自己研究内容所选用专业术语的交集来描述作者研究内容的相似性。本研究中关键词主要从样本数据的作者关键词、附加关键词以及题名词中进行抽取获得的。图4是作者关键词耦合网络结构的可视化结果,作者所归属的主题领域区分也十分明显。与前

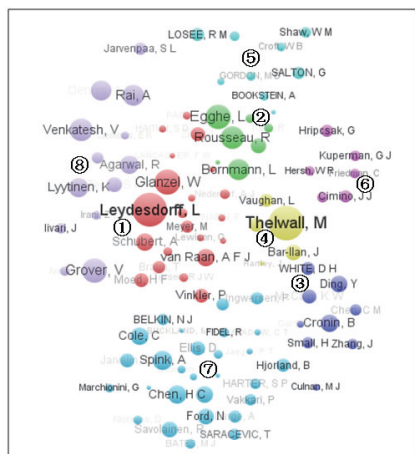


图3 作者文献耦合网络结构

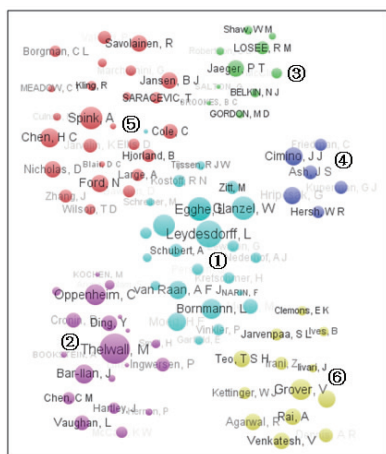


图4 作者关键词耦合网络结构

面的几个网络分析结果相比,聚类6所代表的管理信息系统领域作者群体变得更大,这是因为大量研究信息系统的作者由于使用相同的词汇也被集中到该类中。在作者文献耦合网络中区分出来的Egghe L, Rousseau R等几个作者重新合并到了计量学研究的聚类1中。Thelwall M本人研究的主要领域虽然为网络计量学,但他作为合作者的一些论文也涉及了引文分析、影响因子等研究内容,所以这个聚类2的可区分度就不像作者文献耦合的分析结果那么明显。以Cimino J J为核心的生物信息学群体聚类4、以Spink A为核心的用户研究和信息行为研究群体聚类5与作者文献耦合网络的分析结果十分相似。聚类3中的作者主要是进行信息检索方面的研究,但其他一些研究人员由于使用如information指代较为宽泛的关键词,导致聚类3中的作者群体研究特征被淡化。

图5显示了作者期刊耦合的结构分析结果,管理信息系统研究群体聚类4,医学信息学研究群体聚类5被比较明显地区分出来;聚类6中三名作者主要是Journal of Information的作者,研究内容之间没有明显的相似性;聚类7中的三名作者则主要是进行政务信息研究的学者,绝大多数论文基本上都发表在期刊Government Information Quarterly上。其他聚类在研究领域上的区分程度则十分模糊,无法做出清晰的界定。

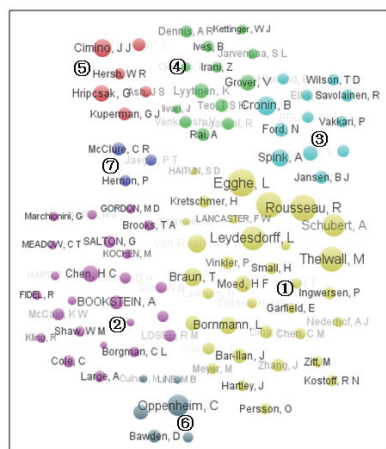


图5 作者期刊耦合结构分析

4.3 结果对比

从结构揭示结果来看,五种网络的子结构分析结果在某些地方具有一致性,例如医学信息学和管理信息系统领域的研究群体,由于这两个研究圈子在98名作者中相对封闭,且绝对影响力较大,因此各种网络都能够最先识别出这两类作者群体。除此以外,各网络的科学研究结构揭示能力则参差不一。作者之间合作关系一般都是在相识的关系上产生的,因此作者合作网络最接近于真实的社会网络,关系产生的要求也最强,往往也最稀疏;同时,由于两个作者之间即使属于同一个研究领域,但也不能保证彼此之间一定会产生合作关系,因此作者

合作网络在科学研究结构的揭示能力上略有不足。在引用基础上产生的共被引网络和真实的社会网络之间的差异是显见的,共被引网络代表了作者共同被他人的认可程度,而事实上两个作者之间很可能根本没有进行过学术交流,因此利用作者共被引网络进行学术交流的分析并不十分合适,但这并不影响作者共被引网络揭示学科结构。

作者文献耦合网络对科学研究结构的揭示能力最强,对于相似度很高的团体也能够根据其在耦合特征上的不同做出更为细致的区分,例如在本研究所选用的样本中,以计量学为主要研究内容的作者群体被分解为两个主题方向略有差别的更小的类,这种区别在其他几种网络的结构分析中很难被揭示出来。作者关键词耦合的分析结果和作者文献耦合的结果很相似,但如情报领域的信息系统研究和管理学领域的管理信息系统研究的作者们由于均使用类似于“系统”等词汇而被聚到一起,事实上这两个群体之间的区别可能大于彼此之间的共性。因此作者关键词耦合的局限也显而易见,其科学结构揭示能力很容易受到所属学科专业词汇使用特征的影响,作者关键词耦合应用于自然科学和工程科学领域的分析准确度更高。

根据加菲尔德^[14]的研究,某一学科数量极少的核心期刊集中了该学科的大部分高质量文献,而这些集中的论文又能够反映期刊的学科归属,虽然本文数据来源的31种期刊在ISI JCR-SSCI edition中都被归入了Information Science and Library Science,但有12种是跨学科期刊,涉及六个学科;从期刊所涉及的主题来看,即便是以科学计量学命名的期刊*Scientometrics*也并未局限于科学计量学单个主题。因此,作者期刊耦合的结果难以将作者

按照其研究领域进行十分细致的区分,图5的可视化结果仅仅能够区分出整个网络中存在明显差别的个别群体。

4.4 相关性分析

相关性分析能够更深层地挖掘各网络结构之间的关联及差异。QAP分析的结果如表3所示,从整体上看,所有网络间相关性的p值均在0.01水平上,各网络之间的相关性显著。五种作者共现网络产生的机理不同,作者之间的共被引网络基于他人对作者成果共同的认可;其他几种网络则是基于作者自身的选择,合作网络是作者选择共同研究的对象而产生的,作者文献耦合网络是通过作者在进行创作时选择之前的研究成果来支撑自己研究的情况所产生的,作者关键词耦合则是通过作者选择特定的专业术语来表示自己的研究成果所产生的,而基于期刊的作者耦合则是作者通过选择自己成果所发表的途径所产生的,但这些差异中同时隐含了一个基本的相同点,即均是基于作者据其研究需求进行的科学选择,因此作者之间的共现关系存在明显的潜在关联。从具体结果来看,作者合作网络和作者文献耦合网络关联程度最高,和作者关键词耦合网络之间相关性最低。

为了更进一步说明情况,QAP相关性的排序结果如表4所示,举例来说,表中的第一行可以解读为:对于作者合作网络而言,它和作者文献耦合网络之间最相关,其次是共被引网络,然后是作者期刊耦合网络,最后是作者关键词耦合网络;第二列可以解读为,相对于其他四种作者共现网络,作者合作网络与它们均最不相关。

表3 网络间QAP相关性检验结果

网络	作者合作	作者共被引	作者文献耦合	作者关键词耦合	作者期刊耦合
作者合作	1	0.353	0.473	0.220	0.306
作者共被引	0.353	1	0.641	0.364	0.513
作者文献耦合	0.473	0.641	1	0.70	0.586
作者关键词耦合	0.220	0.364	0.70	1	0.584
作者期刊耦合	0.306	0.513	0.586	0.584	1

表 4 网络间 QAP 相关性检验排序

网络	作者合作	作者共被引	作者文献耦合	作者关键词耦合	作者期刊耦合
作者合作	—	2	1	4	3
作者共被引	4	—	1	3	2
作者文献耦合	4	2	—	1	3
作者关键词耦合	4	3	1	—	2
作者期刊耦合	4	3	1	2	—

QAP 分析结果中有两个特征十分明显：

其一，作者合作网络与其他网络之间的潜在关联程度均最低。这个结果与 Yan 和 Ding^[15] 利用余弦距离所测机构层面网络相关性分析结果很相似，说明作者合作网络确实与其他网络之间存在明显的差别，利用合作网络进行分析的结果和其他网络进行分析结果之间存在的替代性较低。

其二，作者文献耦合网络是与其他四种网络最相关的网络。作者文献耦合与基于关键词的作者耦合的相关系数最大，达到 0.7，这是因为作者在进行科学研究时，所选用的参考文献与自身的研究极为相关，属于已有的知识基础，而利用词或者专业术语对自己的研究进行描述，这种描述本身可以认为是对以往研究基础进一步拓展后的总结，因此这种结果的出现也具有一定的必然性。作者文献耦合与作者共被引网络之间的相关性也很高，被引用体现了作者在整个研究主题发展过程中起到的基础作用，因此作者共被引揭示的学科结构是回溯的；作者文献耦合更倾向于揭示研究前沿的结构，相对于同一个作者群体而言，其研究内容从历史的角度来看应该存在明显的继承性，故两种网络之间的相关系数较高。

5 讨论

关于作者共现网络研究在分析方法上还有一些值得探讨的问题。网络之间的关联分析是对网络直接对比结果进一步解释的重要途径，QAP 方法适用于不同类型的作者共现关系^[11]。在子结构分析时，目前社会网络分析法的“凝聚子群分析”

应用也十分广泛，方法包括组元、k-核、集圈等，同时辅之以中心度指标进行结构解释。本文在分析过程中也尝试使用凝聚子群方法，但通过对比发现聚类分析的结果相对容易解释，而凝聚子群方法是否适用则不确定，因此应用于作者共现关系分析需要注意以下两个问题：

(1) 传统社会网络大多数是无权的 0-1 网络，k-核等指标主要是探索节点的结构性特征，因此分析这样的网络优势明显，因为这种网络中的节点结构特征往往就反映了节点的本身属性特征；作者之间的共现网络大多数情况下为权值网络，且由于经常选择高影响力或者高产样本进行实证研究，网络中关系的高权值的情况愈加明显，结构性特征是否能够真实反映节点特征，不同权值间的区分度有多大，这些均难以确定，因此类似于 k-核、集圈等指标的适用性很难保证。

(2) 用于辅助解释结构的指标如接近中心性、中介中心性在作者共现关系中并不一定适用。接近中心性主要反映某个节点与其他节点之间产生关联的能力，计算的基础是某个作者与其他作者之间的“距离”；中介中心性主要反映作者在多大程度上处于其他节点发生关联的中间路径上，其计算基础是某个作者在网路中起到的中介作用程度^[16]。在作者共现关系中，由网络分析所得到的“距离”和中介程度并不能够真实反映作者之间的关联，因为学术环境中，作者之间发生联系的基础发生了改变，例如一个作者选择其他人进行合作时，并不一定需要中间人进行过渡；在引用某个作者的成果时，优先产生影响的是信息获取渠道，而这种渠道并不基于作者之间的“距离”，因此这些

结构辅助解释指标的使用同样需要谨慎的甄别。

6 总结

为了尽可能反映客观实际,本文进行了详细的研究设计,总结了目前 WOS 数据库中作者、被引作者字段及关键词著录的主要问题,针对这些问题进行了数据清洗。在此基础上,构建了五种类型的作者共现网络,通过层次聚类与相关性分析方法对不同类型的作者共现关系揭示科学研究结构的能力进行了比较。通过层次聚类发现,作者合作关系在研究科学交流上的优势比较明显,但利用它进行科学研究结构的揭示则需要仔细甄别其适用性;作者共被引关系的科学研究结构揭示能力适中,在识别学科领域的研究结构上有其独特之处;作者文献耦合关系的科学研究结构揭示能力具有十分明显的优势,其结构分析的准确性和精确性最高;作者关键词耦合关系的科学研究结构揭示能力非常容易受到学科领域用词特点的影响;作者期刊耦合的

科学研究结构揭示能力虽然较差,但某些时候则能区分出一些容易被忽视的群体差异。

从网络之间的关联角度来看,作者合作网络分析的结果与其他网络之间的关联程度均最低,这说明利用合作网络进行分析的结果和其他网络进行分析结果之间存在的替代性较低,发现的角度不同;而作者文献耦合网络与其他网络之间的关联程度最高,这说明作者文献耦合网络的分析结果很容易涵盖其他网络的分析结果。在进行科学研究结构分析时,需要仔细甄别所选样本群体本身的特征,对于某些特殊的情况,例如对非正式交流渠道比较畅通的群体进行分析时,作者共被引网络就不一定合适,因为这个群体很可能合作程度会比较高,而非正式交流可能会对作者之间的互相引用产生替代作用,此时合作网络反映的结构差异可能更为合理。因此,根据分析对象的特点,分析多种作者共现网络的特点并结合使用,有助于较好地反映科学研究的实际情况。

参考文献

[1] Yan E J, Ding Y. A framework of studying scholarly networks [EB/OL]. [2012 - 12 - 05]. http://ella.slis.indiana.edu/~eyan/papers/QualPaper_ScholarlyNetworks_ErjiaYan.pdf.

[2] Borgman C L. Bibliometrics and scholarly communication[J]. Communication Research, 1989, 16 (5): 583 - 599.

[3] Borgman C L, Furne J. Scholarly communication and bibliometrics[J]. Annual Review of Information Science and Technology, 2002, 36: 3 - 72.

[4] Newman M E J. The structure of scientific collaboration networks[J]. PNAS, 2001, 98(2): 404 - 409.

[5] White H D, Griffith B C. Author cocitation: A literature measure of intellectual structure[J]. Journal of the American Society for Information Science, 1981, 32(3): 163 - 171.

[6] Ahlgren P, Jarneving B, Rousseau R. Requirements for a co-citation similarity measure, with special reference to Pearson's correlation coefficient[J]. Journal of the American Society for Information Science and Technology, 2003, 54 (6): 550 - 560.

[7] Chen C M, Paul R J, Keefe B O. Fitting the jigsaw of citation: Information visualization in domain analysis[J]. Journal of the American Society for Information Science and Technology, 2001, 52(4): 315 - 330.

[8] 马瑞敏,倪超群. 作者耦合分析:一种新学科知识结构发现方法的探索性研究[J]. 中国图书馆学报, 2012, 38(3): 4 - 11. (Ma Ruimin, Ni Chaoqun. Author coupling analysis: An exploratory study on a new approach to discover intellectual structure of a discipline[J]. Journal of Library Science In China, 2012, 38(3): 4 - 11.)

[9] Nooy D W, Mrvar A, Batagelj V. Exploratory social network analysis with Pajek[M]. New York: Cambridge University Press, 2005: 101 - 119.

小资料

梁启超推荐沈从文当图书馆管理员

1923年8月,沈从文(1902~1988)从湘西来到北京。他经常到北大听课,到京师图书馆阅览书籍。1924年底,沈从文以“休芸芸”笔名在《晨报》副刊发表了系列作品,北大林宰平教授读之大为欣赏,曾向梁启超谈及沈从文的困难处境。梁对这位自学成才的湘西青年也非常赏识。1925年8月,经梁启超向熊希龄推荐,23岁的沈从文来到熊希龄创办的香山慈幼院任图书管理员,(符家钦:《梁启超、朱光潜称赞沈从文》//萧乾主编:《史迹文踪》,47-48页,北京,中华书局,2005)月薪20元。(陈明远:《文化人的经济生活》,108页,上海,文汇出版社,2005)11月,沈从文受熊希龄派遣到北京大学图书馆在职进修,师从袁同礼学习图书编目等方面的业务知识,住在北大附近的公寓中,直到翌年二三月间才返回香山慈幼院。做图书馆管理员期间,沈因发表一些文学作品,讽刺了香山慈幼院图书馆的上司和香山上层人物,1926年8月底不得不辞去慈幼院图书馆的工作。(吴世勇:《沈从文年谱》,27-33页,天津,天津人民出版社,2006)今香山的绿云舫遗址上有一小白楼,即为民国时期所建香山慈幼院图书馆的旧址。

(王子舟 整理)

-
- [10] Robert A H, Riddle M. Introduction to social network methods[M/OL]. CA: University of California, Riverside Press, 2005: Chapter 17. [2012-12-20]. http://faculty.ucr.edu/~hanneman/nettext/C17_Two_mode.html.
- [11] Thomson Reuters. JCR-SSCI edition[EB/OL]. [2012-12-01]. <http://webofknowledge.com/JCR>.
- [12] White H D, Wellman B, Nazer N. Does citation reflect social structure? Longitudinal evidence from the "GloboNet" interdisciplinary research group[J]. Journal of the American Society for Information Science and Technology, 2004, 55(2), 111-126.
- [13] Eck N J, Waltman L. Software survey: VOSviewer, a computer program for bibliometric mapping[J]. Scientometrics, 2010, 84(2), 523-538.
- [14] Garfield E. Citation analysis as a tool in journal evaluation[J]. Science, 1972, 178 (4060), 471-479.
- [15] Yan E J, Ding Y. Scholarly network similarities: How bibliographic coupling networks, citation networks, co-citation networks, topical networks, co-authorship networks, and co-word networks relate to each other[J]. Journal of the American Society for Information Science and Technology, 2012, 63(7): 1313-1326.
- [16] Wasserman S, Faust K. Social network analysis: Methods and applications[M]. New York: Cambridge University Press, 1994: 169-215.

邱均平 武汉大学信息管理学院教授,博士生导师。通讯地址:武汉市珞珈山。邮编:430072。

董克 武汉大学信息管理学院博士研究生。通讯地址同上。

(收稿日期:2013-03-27;修回日期:2013-04-18)