

作者文献耦合分析与作者关键词耦合分析比较研究: *Scientometrics* 实证分析*

宋艳辉 武夷山

摘要 本文以 *Scientometrics* 1991—2010 年间刊载的 2,045 篇论文作为数据样本,分两个时间段研究这 20 年间科学计量学的知识结构与演进状况,并以此来研究作者文献耦合分析法(ABCA)与作者关键词耦合分析法(AKCA)在揭示学科领域知识结构方面的异同。研究发现,2001—2010 年间的科学计量学的知识结构要比 1991—2000 年间的更加清晰明朗,其研究主题之间更为亲密,相互作用力明显要强。作者排名相关分析、研究主题探测、余弦相似度计算、研究主题变迁等均显示,AKCA 与 ABCA 存在高度相关性;通过因子分析的模型拟合、研究主题的探测与变迁分析,又显示二者略有不同,ABCA 可以探寻到比 AKCA 更多的研究主题,AKCA 比 ABCA 能显示更多的信号来反映学科的技术突破以及研究前沿的进展。因此 ABCA 与 AKCA 不可互相替代,二者结合起来是探寻学科知识结构及其发展的理想研究方法。图 4。表 6。参考文献 25。

关键词 作者文献耦合分析 作者关键词耦合分析 科学计量学

分类号 G350

A Comparative Study on Author Bibliographic-coupling Analysis and Author Keyword-coupling Analysis Based on *Scientometrics*

Song Yanhui & Wu Yishan

ABSTRACT Taking 2,045 papers published in *Scientometrics* from 1991 to 2010 as a sample and by dividing them into 2 groups according to different time periods, we study the knowledge structure and evolution of *Scientometrics* during these 20 years, and then the two methods for disciplinary structure analysis: author bibliographic-coupling analysis(ABCA) and author keyword-coupling analysis(AKCA). We find that the knowledge structure of *Scientometrics* between 2001 and 2010 is more clear than that between 1991 and 2000, the relationship among research topics of the former is closer and the interaction force of the former is stronger. Related analysis of author ranking, exploration of research topics, calculation of cosine similarity and changes of research topics all show that there are both high correlation and slight difference between ABCA and AKCA with the former exploring more research topics and the latter displaying more signals to reflect technological breakthroughs and the development of research front in the disciplines. A combination of the two methods is proposed in exploring disciplinary structure and development. 4 figs. 6 tabs. 25 refs.

KEY WORDS Author bibliographic-coupling analysis. Author keyword-coupling analysis. *Scientometrics*.

* 本文系浙江省高校人文社科“决策科学与创新管理”重点研究基地资助项目“引文分析视角下的信息管理学科可视化研究”(编号:RWSKZD04—201207)和国家社科基金项目“中外图书情报学科知识图谱比较研究”(编号:11CTQ015)的研究成果之一。

通讯作者:宋艳辉,Email:syh687@163.com

1 引言

1973年,美国情报学家 Small^[1]首次提出了文献共被引(Co-citation)的概念,作为测度文献间关系程度的一种研究方法。与 Small 在同一时间提出该理念的还有苏联情报学家 Marshakova^[2]。1981年,White 与 Griffith^[3]将文献共被引拓展至作者层面,形成了作者共被引分析(Author Co-citation Analysis,ACA)的研究方法。随后,众多学者对 ACA 进行系统而全面的研究。这些研究中不仅包含对该方法本身的改进,如第一作者共被引、全部作者共被引^[4-6],矩阵对角线如何设定、矩阵是否需要转化、如何转化^[7-8],共被引强度计算方法^[9]等;还包含将 ACA 应用于某一科学进行领域探测的研究,如 Pamela^[10]将 ACA 应用于社会生态学领域、White^[11]应用于情报科学领域、Chen^[12]应用于数字图书馆等。这一系列的研究成果丰富了共被引的理论与方法,促进该学科的发展,并使得 ACA 作为一种分析方法日渐成熟。

与文献共被引的发展态势相比,文献耦合(Bibliographic coupling,BC)作为与文献共被引形成对称关系的研究方法,其发展却相形见绌。而实际上,文献耦合的提出比文献共被引早了整整10年的时间。1963年,美国麻省理工学院教授 Kessler 在对《物理评论》期刊进行引文分析研究时发现,越是学科、专业内容相近的论文,其参考文献中相同文献的数量就越多。他把引证同一篇论文的两篇或多篇论文称为耦合论文(coupled papers),并把它们之间的这种关系称为文献耦合^[13]。BC 同样是测度文献间关系程度的一种研究方法,却并未像共被引一样获得长足的发展。直到2008年,Zhao^[14]等人将 BC 拓展至作者文献耦合分析(Author Bibliographic Coupling Analysis,ABCA),以情报学领域为例来进行实证研究,并将 ABCA 与 ACA 进行充分对比,以探析二者之间的异同点。但之后的几年,Zhao 等人并没有继续进行 ABCA 的研究,又转而进行 ACA 的研究。

根据耦合的原理,既然两篇或多篇文献可以

引用同样一篇参考文献而发生相互关联,从而体现文献之间的某种隐含关系;那么,同样作为分析单元的文献关键词若被不同文章共用,也可以将不同作者的文章关联起来。通过关键词耦合起来的文献之间,以及这些文献的作者之间会隐含着一种什么样的关系?这两种关系之间又会有什么异同呢?本文试图予以探讨。本文研究主要有两个目的:①作者关键词耦合分析(Author Keyword-coupling Analysis,AKCA)是否是一种有效的耦合分析方法?它与另一种耦合分析方法 ABCA 在揭示学科领域的知识结构方面有哪些异同点?②在1991—2000年与2001—2010年这两个时间段,科学计量学呈现一种什么样的知识结构,以及科学计量学在这20年间是如何演进的?

2 方法论

2.1 数据采集与数据清洗

本研究有两个假设:①ABCA 是一种可信的、有效的研究学科领域知识结构的方法;②科学计量学领域高活力作者的文献基本可以显示科学计量学的研究状况。为说明此研究,我们引入 AKCA,将它运用于科学计量学领域,并将其结果与 ABCA 的结果进行对比分析。

本文的数据源为 Web of science,从该网络数据库分别检索并下载 *Scientometrics* 所刊载的1991—2000年间以及2001—2010年间的数据库,精炼结果仅保留 Article 与 Proceedings,分别得到科学计量学在1991—2000年间的790篇论文数据以及在2001—2000年间的1,255篇论文数据,数据包含题名(TI)、作者(AU)、摘要(AB)、关键词(ID)、参考文献(CR)等信息。将经过采集、清洗完毕的数据分作两个数据集存储,即1991—2000的数据集与2001—2010的数据集,以便于运用 AKCA 与 ABCA 研究科学计量学的知识结构及其在这20年间的演进状况。我们开发了耦合系统软件来处理该数据集,并引导该软件运行生成便于进一步进行因子分析的数据结构形式。

2.2 方法与相关研究

2.2.1 耦合分析

在经典的作者共被引分析(ACA)中,选择多少位作者并没有严格的限制^[15]。同ACA一样,作者耦合分析也没有约定俗成的做法,而且,由于这方面的研究较少,其分歧甚至比ACA更大。本文认为,在进行作者耦合分析时,对结果影响最大的就是作者耦合频次的计算以及选择所要研究科学领域的代表性作者。

2.2.1.1 计算思路

关于作者文献耦合的计算方法有两个不同的视角,颇具代表性的是Leydesdorff和Zhao的算法。Leydesdorff将其耦合分析软件挂接在个人学术网站上^[16],马瑞敏推算其算法思想为:两个作者(分别为A和B)的耦合首先是两个文献的耦合,即文献的耦合是基础,先求出两篇文献的耦合次数,再求出这两篇文献的作者之间的耦合次数,然后累加。Zhao^[14]提供的两个作者文献耦合的计算方法是:将某个作者(只考虑第一作者)所有论文的参考文献作为一个集合,然后和另一个作者所有论文的参考文献进行比较,找出共同的参考文献次数即为这两个作者之间的文献耦合次数。由于一篇文献在某个作者的参考文献中出现的次数不止一次,比如在A作者的参考文献中出现N次,同时在B作者的参考文献中出现M次,则这篇文献要赋以权值,具体为 $\min(N, M)$ 。Leydesdorff的方法是每增加一篇文献便要和目标作者的所有文章的参考文献进行匹配,因此效率比较低,而且Leydesdorff仅仅提供了该软件,并未用它进行任何的实证研究。本文更加倾向于Zhao的处理方法,并且将这种方法推广至AKCA,每一个关键词相当于ABCA中的一篇引文。

2.2.1.2 代表作者选择

具体选择思路是:首先统计数据集中的高发文作者,提取发文大于两篇的作者;接着计算这些高产作者之间的耦合频次,计算每位作者的平均耦合频次(计算方法:作者与其他作者的耦合频次总和除以高产作者数目,再减去1);根据平均耦合频次对作者进行排名,抽取排名前100的作者,作为

代表作者进行耦合矩阵的构建。至于选择多少位作者作为代表作者,也没有约定俗成的做法^[15]。White与McCain^[11]在进行作者共被引分析时选择了120位作者,当时由于数据库功能的限制,他们直接从DIALOG中检索获得作者的共被引数据。Zhao等人效仿White,同时也是为了与White的研究结果进行比较,也选择了120位代表作者进行作者共被引分析和作者文献耦合分析,他们的数据量是4,422篇论文。在Zhao的一系列相关研究中,都是沿用120位作者的做法,其数据样本量一般也都比本文的大,如Scopus收录的情报学领域的3,824篇论文^[17],Web of science收录的“XML”领域的2,475篇论文^[18]。而本文的数据样本量仅有2,045篇论文,因此本文选取100位作者作为代表作者。两个时间段、两种不同方法筛选出四组不同的作者集(都由100位作者组成)所发表的论文数量之和占我们数据样本量的比例如下:ABCA1991—2000为46%,AKCA1991—2000为44%,ABCA2001—2010为35%,AKCA2001—2010为36%。这几个数值都远远高于Zhao^[14]等人统计的25%的比值。因此,我们选取的作者比Zhao等人的更有代表性,这100位作者完全可以反映出科学计量学在这两个时间段的研究状况。

2.2.2 因子分析

分别构造出100位作者的作者文献耦合矩阵和作者关键词耦合矩阵,得到四个不同的矩阵。矩阵的对角线值为作者的平均耦合频次减去1(消除自耦合)。矩阵导入SPSS进行因子分析,因子萃取选择主成分分析(principal component analysis, PCA)。因子数量是通过检查方差的总解释度、公共因子和相关残差确定^[19]。因子分析模型采取直接斜交转轴法。从理论上讲,一个学科尤其是一个二级学科的各个研究主题(因子)之间不可能是孤立的,而是存在一定相关关系的。因此,直接斜交转轴法可以更好地反映现实状况。

2.2.3 可视化

很早就有学者选择因子分析结果中载荷值超过某一自行设定阈值的载荷作者进行分析^[3,20]。早期的学者在因子结构可视化方面一般选择表格

形式或者多维尺度图谱(MDS maps)^[11,15],这种可视化展示形式虽然展示的信息量够大,但是占用的空间也很大,且不够形象直观。Zhao^[21]等人在借鉴前人的基础上,引入了一种新的可视化方法,以更加浓缩的社会网络图谱形式呈现各因子及其高载荷作者。该网络图谱的绘制是通过 Pajek 执行的 Kamada – Kawai 图谱分布算法(Kamada – Kawai graph layout algorithm)实现的,这是一种以发明者名字命名的可视化算法。本文的研究借鉴 Zhao 等人的可视化展示形式,但区别于 Zhao 的是,本文所绘制的图谱是通过执行 UCINET 中集成的 NETDRAW 软件实现的。该图谱依然可以以较小的空间呈现尽可能多的信息,且展示效果直观明了。

在本文的图谱中,作者以方形节点表示,因子以圆形节点表示。方形节点与圆形节点之间连接,表示作者在该因子上承载一定的荷值,载荷值越大,连线就会越粗;连线的色度也表明载荷值的大小,浅灰色的连线表示载荷值较小,黑色的连线表示载荷值较大。需要说明的是,只有作者在因子上的载荷值大于 0.3 才会被选入图谱并与因子建立连接。

图谱中节点的大小也表示不同的含义,是通过计算载荷数值而得出的结果。作者节点(方形)的大小与作者所承载在因子上的载荷值成正比;因子节点(圆形)的大小与连接在该因子上的载荷数值大于 0.3 的作者的载荷值之和成正比。图谱中节点的色度表示它们之间互相连接的次数,即网络

中节点的点度(degree)。网络节点之间的距离远近可以近似地表示它们之间的亲密程度。如果两个因子之间距离较近,则这两个研究主题之间很可能有较强的相关性。

3 结果和讨论

3.1 作者排名

无论是作者文献耦合还是作者关键词耦合,作者能够与其他作者建立较高的耦合关系就表示该作者具有较高的研究活力。根据作者跟其他作者的平均耦合频次对作者进行排名,然后对排名结果进行相关分析,可以建立四对相关分析,如表 1 所示。对 1991—2000 年的 136 位作者分别进行 ABCA 和 AKCA 分析,并根据排名各自选取 100 位代表作者,在这 100 位代表作者中有 79 位相同作者,对这 79 位作者在 AKCA 与 ABCA 中的排名进行相关分析;同理,可以从 2001—2010 年的 203 位作者中提取 68 位相同作者进行 AKCA 与 ABCA 排名的相关分析。如此,分别得到相关系数 0.396 和 0.398。另外,还可以分别对两个时间段的所有作者的 AKCA 与 ABCA 排名进行相关分析,得到相关系数 0.407 和 0.520。结果显示,作者文献耦合与作者关键词耦合之间并不是没有关系,而是呈现一种弱正相关性,并且随着样本数据量的增大,这种相关性也在变强。

表 1 ABCA 与 AKCA 作者平均耦合排名相关分析

	1991—2000 ABCA 共有作者	2001—2010 ABCA 共有作者	1991—2000 ABCA 所有作者	2001—2010 ABCA 所有作者
1991—2000 AKCA 共有作者	0.396			
2001—2010 AKCA 共有作者		0.398		
1991—2000 AKCA 所有作者			0.407	
2001—2010 AKCA 所有作者				0.520

在作者文献耦合与作者关键词耦合的 100 位代表作者中有很高的重合率(1991—2000 年 79% 的重合,2001—2010 年 68% 重合),即在作者文献耦合中的高耦合频次作者依然是作者关键词耦合的高耦合频次作者,这进一步论证了 ABCA 与 AKCA 存在着相关性。至于 1991—2000 年与 2001—2010 年的科学计量学耦合情况的比较,虽然是 79% > 68%,却不足以说明这种相关性在减弱,因为针对 1991—2000 年的统计,是从 136 位作者中选取 100 位代表作者;而针对 2001—2010 年的统计,是从 203 位作者中选取 100 位代表作者。事实上,ABCA 与 AKCA 的这种相关性在 2001—2010 年略有增强,因为,在分析过程中我们若如 1991—2000 年的一样,同样仅选定前 136 位作者作为分析基础,结

果显示有 80 位共有作者,即 80% 的重合率,略大于 79%。

3.2 因子模型拟合分析

将我们构建的 ABCA 与 AKCA 的作者耦合矩阵进行相似矩阵的转化,导入 SPSS 进行因子分析。因子模型的拟合结果如表 2 所示。从整体结果上看,耦合分析模型拟合优度非常理想,例如,1991—2000 年的 ABCA 中,16 个因子可以解释 93.76% 的总方差;只有 47 个数值的观察值和预测值的相关差异大于 0.05,几乎 100% 的残差都小于 0.05;公因子变动的数值最小为 0.76,最大为 0.99,小于 0.7 的公因子为 0(或 0%),小于 0.8 的公因子为 1(或 1%),小于 0.9 的公因子也只有 18(或 18%)。

表 2 因子模型拟合结果

耦合分析	因子数	总方差 解释度(%)	非冗余残差 >0.05(%)	公因子			
				变动范围	<0.7(%)	<0.8(%)	<0.9(%)
1991—2000ABCA	16	93.76	47(0)	0.76—0.99	0(0)	1(1)	18(18)
1991—2000AKCA	8	97.17	15(0)	0.86—0.998	0(0)	0(0)	2(2)
2001—2010ABCA	11	93.86	71(1)	0.76—0.995	0(0)	2(2)	15(15)
2001—2010AKCA	10	94.55	37(0)	0.81—0.99	0(0)	0(0)	8(8)

两个时间段都显示,作者关键词耦合比作者文献耦合拟合结果更为理想,这表示作者关键词耦合的结果比作者文献耦合的结果更易于解释。1991—2000 年的 AKCA 分析结果显示,作者耦合分析仅用 8 个因子就可以解释高达 97.17% 的总方差,比 ABCA 用 16 个因子解释 93.76% 好很多;最小公因子数值为 0.86,远远大于 0.76;98% 的公因子数值都大于 0.9,而 ABCA 大于 0.9 的公因子数仅占 88%。2001—2010 年的数据仍然显示 AKCA 好于 ABCA,AKCA 的 10 个因子可以解释 94.55% 的总方差,几乎 100% 的残差绝对值都是小于 0.05;ABCA 则是 11 个因子揭示 93.86% 的总方差,99% 的残差绝对值小于 0.05;AKCA 的公因子最小值 0.81 大于 ABCA 的 0.76;ABCA 小于 0.8 的公因子数占 2%,小于 0.9 的公因子数占 15%,而 AKCA 没有小于 0.8 的公因子,小于 0.9 的公因子也仅占 8%。

3.3 科学计量学(1991—2000)

针对科学计量学在 1991—2000 年间的数据库构建的作者文献耦合矩阵与作者关键词耦合矩阵,分别进行因子分析,以主成份方法萃取因子,并进行直接斜交旋转,然后可视化其结构矩阵。

3.3.1 科学计量学(1991—2000)知识结构

文献耦合在揭示学科领域知识结构方面已经受到越来越多学者的认同,并得到实证分析。因此,我们以作者文献耦合分析方法来呈现科学计量学的知识结构。经过因子分析,我们探测到科学计量学(1991—2000)共 16 个因子(见表 3)。载荷数表示研究主题(因子)在学科内的活跃度,载荷数是因子上载荷大于 0.3 的代表作者数,其计算方法不仅计算首要载荷作者数,同时计算次要载荷作者数,我们认为次要作者也可以表示研究主题的活跃性,忽略次要载荷作者数的做法不太合理^[21-22]。各因子上最高载荷表示它所代表的各因子(研究主

题)在学科内的显著度。因子标签的标注是一项比较困难的工作,为此我们充分考虑耦合的原理,即耦合的发生是作者共同引用了一篇文献或者一个关键词。我们检查各因子中高载荷作者以及高耦合强度作者的共同研究主题,其中 ABCA 主要检查源数据集中高耦合强度作者的耦合文献以及它们的题名;AKCA 主要检查源数据集中高耦合强度作者的耦合关键词。我们发现,各因子中高载荷作者跟高耦合强度作者往往保持一致,而且各因子中与最高载荷作者发生最高耦合强度的作者往往也是因子中第二、三的高载荷作者,如 ABCA 引文分析 Luukkonen,T(1)与 Wouters,P(3)、学科领域计量 Jain,A(1)与 Karki,M M S(2)、专利分析 Tijssen,R

J W(1)与 Meyer,M(3); AKCA 科学计量指标 SANCHO,R(1)与 WHITNEY,G(2)、期刊计量指标 PICHAPPAN,P(1)与 Egghe,L(2)、医学计量 Breimer,L H(1)与 Zhang,H Q(2)等。这使得因子的标注工作变得相对容易很多,只需要检查这些高载荷与高耦合强度作者的共同研究主题,即可确定因子标签。因子的确定主要是根据模式矩阵,因为模式矩阵中的作者体现了作者对因子的独特贡献。如果模式矩阵的载荷作者过少,无法进行标注时再借助结构矩阵。通过以上方法就可以确定 16 个因子的标签,其中有 3 个因子无论是在模式矩阵还是结构矩阵中都无法找到高载荷作者,因此无法准确地确定其因子标签,我们以“未查明”来表示。

表 3 ABCA 因子及其载荷(1991—2000)

因子	载荷数	最高载荷	因子	载荷数	最高载荷
国际合作	23	0.97	科学知识图谱	8	0.83
科学合作	16	0.75	专利分析	7	0.94
期刊影响因子	14	0.84	文献计量应用研究	6	0.94
科学创新	13	0.85	文献计量基础 规律研究	4	1.04
引文分析	12	0.70	期刊评价	1	0.31
共词分析	12	0.88	未查明	3	0.42
学科领域计量	10	0.92	未查明	1	0.31
期刊引证	9	0.68	未查明	0	0.00

结构矩阵中的作者载荷体现了作者之间以及因子之间的相关性,也就是说这些载荷不仅承载了作者与因子之间的相关性,也包含了各因子之间的相互作用^[19]。因此,我们认为结构矩阵可以更充分地体现学科领域的结构特性,因子分析结果的 16 个因子的结构矩阵可视化结果如图 1 所示。如果在图谱中央划一根虚线,则科学计量学(1991—2000)研究被分为两个研究区域,左侧的区域明显表现为科学计量学的文献研究,右侧的区域则表现为科学计量学的政策研究。文献研究是对科学计量学基本的、传统的问题、方法以及规律的研究,主要包含:期刊影响因子、引文分析、期刊引证、期刊评价、共词分析、文献计量基础规律研究等。政策研究是科学计量学与科学学的实践问题、方法以及规律的应用等,主要包含:国际合作、科学合作、科

学创新、文献计量应用研究、学科领域计量、专利分析等。科学知识图谱,由引文分析衍生而来,似乎正在成为连接文献研究与政策研究的一个领域。在科学计量学(1991—2000)中,科学知识图谱还是一个规模较小的研究主题。但下文的研究显示,在科学计量学(2001—2010)中,该研究已经发展为一个主流研究主题,而且与引文分析研究变得更加紧密相关。

3.3.2 ABCA 与 AKCA 对比分析

3.3.2.1 主题探测

对科学计量学(1991—2000)的数据进行文献关键词耦合分析所探测到的因子及其载荷如表 4 所示, AKCA 的因子分析结构矩阵的可视化结果如图 2 所示。分别比较表 3 与表 4、图 1 与图 2,我们可以发现:

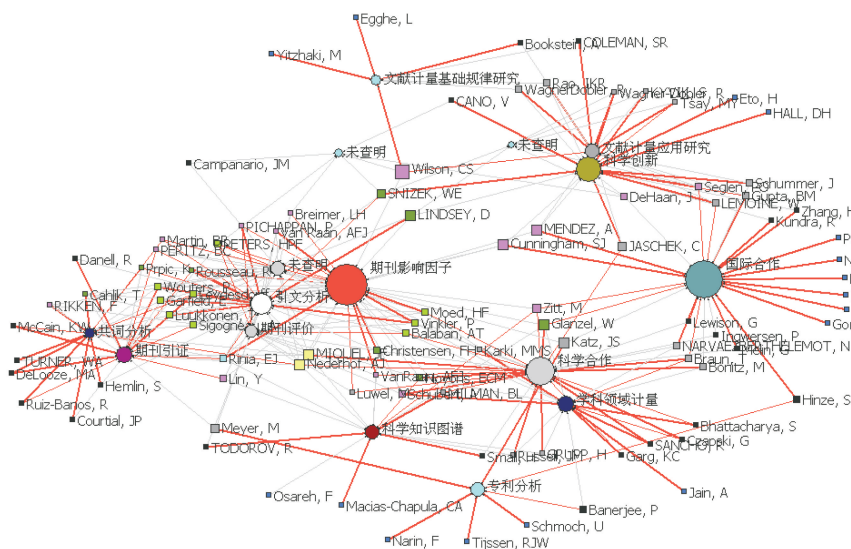


表4 AKCA 因子及其载荷(1991—2000)

因子	载荷数	最高载荷	因子	载荷数	最高载荷
科学计量指标	33	0.92	区域计量	14	0.94
科学合作与分布	29	0.85	科学与技术	11	0.92
科研合作	20	0.93	医学计量	6	1.02
期刊计量指标	19	0.99	未查明	4	0.38

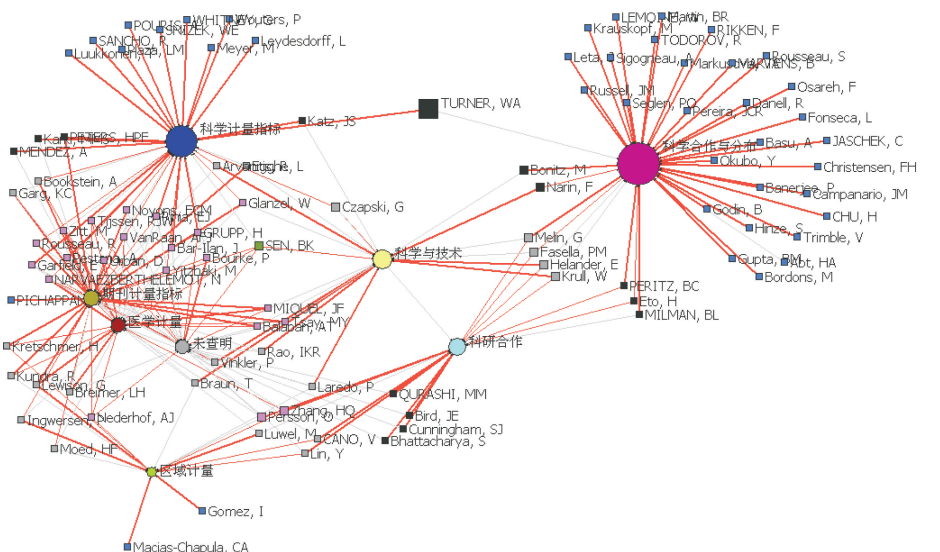


图2 AKCA 结果(科学计量学 1991—2000)

(1) ABCA 可以探寻到比 AKCA 更多的研究主题。针对科学计量学 1991—2000 年的数据, ABCA 探寻出了 16 个研究主题, 而 AKCA 仅探寻到 8 个研究主题。其原因有: ①耦合的基础数据项(引文、关键词等)数量不同。ABCA 在计算耦合频次时比 AKCA 依赖更多的数据项。统计 136 位作者的引文数量, 它们的引文数量总和为 5,576, 平均每位作者的引文数量为 41; 而统计 136 位作者的关键词数量, 其数值远远小于引文数量, 关键词数量总和为 935, 平均每位作者的关键词为 6.88。统计科学计量学 2001—2010 年间的数, 依然是这种现象, 203 位作者的引文数量总和与关键词数量总和分别为 16,737 和 2,982; 平均每位作者的引文数量与关键词分别为 82.45 和 14.69。不可否认的是, 数据项越多, 发生耦合的可能性就越大, 但也更易产生大的变数。②耦合发生的学科差异。ABCA 所依赖的引文数据不仅包含本学科的文献还依赖其它众多学科的文献, 而 AKCA 依赖的关键词一般都是反映文章实质内容的本学科领域范围内的分类主题词。③年代追溯性的差异。文献耦合关系一旦确立, 这两篇文献的耦合频次便不会随时间发生变化, 这一点不仅适用于 ABCA, 也同样适用于 AKCA。但是作者文献耦合发生的追溯性更强, 可以跨越数个年代。也就是说, 虽然文献耦合发生是在文献出版的时间, 但致使耦合发生是可以追溯到数个年代之前的文献。AKCA 则在这些方面表现得很弱。因此, 我们认为, 以上三点原因使得 ABCA 增加了耦合发生的多变性和不稳定性, 当文献耦合上升到作者耦合时就会增加作者耦合的多变, 致使作者呈现多样性, 因而降低了因子对总方差的解释度, 使得 ABCA 的模型拟合不及 AKCA 理想, 要充分解释其总方差就比 AKCA 需要更多的因子。

(2) 在研究热点主题发现上, ABCA 与 AKCA 的探测结果基本是一致的。以科学计量学(1990—2000)为例, ABCA 探测到的第一与第二的研究主题是“国际合作”与“科学合作”, 这与 AKCA 中探测到的排名第二与第三的研究主题“科学合作与分布”与“科研合作”基本是对应的, 虽然在表述或者合作方式上有所差异, 实质上都是关于合作的研究

主题。ABCA 中排名第三的“期刊影响因子”与 AKCA 中排名第一的“科学计量指标”、第四的“期刊计量指标”有对应关系。期刊影响因子应属于期刊计量指标的研究范畴, 而二者又是科学计量指标的一个方面, 因此, 三者之间具有一定的隶属关系。由此看来, 针对科学计量学 1991—2000 年间的数, ABCA 探测到的前三位的研究主题与 AKCA 探测到的前四位的研究主题基本一致。这也进一步说明, ABCA 与 AKCA 存在着一定的相关性, 在 ABCA 中建立高耦合关系的作者, 在 AKCA 中也极易建立高的耦合关系。

(3) AKCA 同样具有一定的预测性。AKCA 的这种特性我们会在下文结合科学计量学(2001—2010)的数据加以说明。

3.3.2.2 耦合矩阵余弦相似度比较

前文已经比较 ABCA 与 AKCA 中作者的排名存在弱相关性, 并论证 ABCA 与 AKCA 具有一定的相关性。这种相关性还可以通过比较 ABCA 与 AKCA 中两个耦合矩阵的余弦相似度得到进一步论证。将本文构建的科学计量学(1991—2000)的 100×100 耦合原始矩阵导入 SPSS, 计算得出 ABCA 与 AKCA 的余弦相似度为 0.398。筛选出二者共同拥有的 79 位作者重新构建 79×79 矩阵, 计算得出二者的余弦相似度为 0.504。以同样方法计算科学计量学(2001—2010)的耦合矩阵, 得出 ABCA 与 AKCA 的余弦相似度为 0.526, 二者共同拥有的 68 位作者构建的 68×68 耦合矩阵的余弦相似度为 0.704。

从这四个数值上看, ABCA 与 AKCA 具有一定的相似度, 实际上更具有统计意义的是 ABCA 与 AKCA 过程中共有作者耦合矩阵的相似度。由于作者耦合方式(文献、关键词)的不同必然导致不同的作者被选入代表作者, 从而进入我们的耦合矩阵。因此, 只有两种耦合矩阵中共有作者的耦合频次在多大程度上相似, 才能更好地说明两种耦合矩阵以及 ABCA 与 AKCA 这两种耦合方式的相似度有多大。结果表明, 两个时间段共有代表作者矩阵的余弦相似度分别为 0.504 和 0.704, 这说明 ABCA 与 AKCA 这两种不同耦合方式下产生的不同结果

具有很大的相似性,也同时表明这种相似性在科学计量学(2001—2010)中比在科学计量学(1991—2000)中体现得更为明显。而且,这个结论也与前文中的研究结论保持高度一致,前文中根据作者跟其他作者的平均耦合频次对作者进行排名,然后对排名结果进行相关分析,无论是 AKCA 与 ABCA 的共有作者还是所有作者,都表明科学计量学第二个时间段中的 ABCA 与 AKCA 的相关性比第一个时间段更强。

3.4 科学计量学(2001—2010)

采用同样的方法对科学计量学(2001—2010)作者文献耦合矩阵与作者关键词耦合矩阵分别进行因子分析,然后可视化其结构矩阵。

3.4.1 科学计量学(2001—2010)知识结构

经过因子分析,我们探测到科学计量学(2001—2010)共 11 个因子(见表 5)。检查高载荷与高耦合强度作者的共同研究主题,确定因子标签。该阶段二者依然保持高度的一致,

如 H 指数的 Ye,F Y(1)与 Liu,Y X(2),科学知识图谱的 Porter,A L(1)与 Rafols,I(2),区域合作的 Boshoff,N(1)与 Sooryamoorthy,R(2),科学与技术的 Van Looy,B(1)与 Verbeek,A(2),期刊影响因子的 Campanario,J M(1)与 Sombatsompop,N(2),科研生产力的 Lariviere,V(1)与 Archambault,E(3)等。该阶段科学计量学最活跃的研究主题是“H 指数”。最小的因子“网络计量”因其载荷数为 0,是检查结构矩阵作者载荷确定的因子标签。

表 5 ABCA 因子及其载荷(2001—2010)

因子	载荷数	最高载荷	因子	载荷数	最高载荷
H 指数	24	0.99	科研生产力	10	0.95
科学与技术	22	0.93	学科交互融合	4	0.45
学科合作	21	0.79	科学知识图谱	2	0.35
期刊影响因子	16	0.81	社会网络分析	1	0.33
区域合作	12	0.86	网络计量	0	0.00
引文分析与可视化	12	0.73			

因子分析结果的 11 个因子的结构矩阵可视化结果如图 3 所示。同样在图谱的中央划一根虚线,科学计量学(2001—2010)的研究较为明显地分为四个研究领域:①科学与技术指标,包含“期刊影响因子”和“H 指数”两个研究因子。②科学合作研究,包含“区域合作”、“学科合作”和“科研生产力”三个研究因子。③科学与技术交融研究,包含“科学与技术”和“学科交互融合”两个研究因子。④引文分析与可视化研究,包含“科学知识图谱”、“社会网络分析”和“引文分析与可视化”三个研究因子。我们可以发现,研究领域①与研究领域②交织在一起,没有明显的界限。实际上,研究领域①与研究领域②分别属于科学计量学 1991—2000 年

间所论述的文献研究与政策研究范畴,这说明该时期科学计量学的发展不再像前 10 年所表现出的文献研究与政策研究、理论问题与实践问题相分离,而是更加注重二者之间的融合,呈现一种良性的发展态势。在科学计量学 1991—2000 年间,科学知识图谱源自引文分析,且并未形成一定规模。在 2001—2010 年间,科学知识图谱与可视化研究俨然已成为科学计量学的两个独立的研究热点领域,更重要的是,该领域各个研究主题之间的联系更加紧密,相互作用与影响在扩大。“网络计量”就是由该领域衍生出来的一个较新的研究主题,检查该因子下的作者及其共同研究的主题发现,该主题主要集中在“网络链接分析”的研究,我们认为网络链

接分析属于网络计量的研究内容。

经过比较可以看出,经过一段时间的发展,科学计量学(2001—2010)的结构比科学计量学(1991—2000)的结构更加清晰明朗,各个研究主题之间不再孤立,而是更加融合贯通。

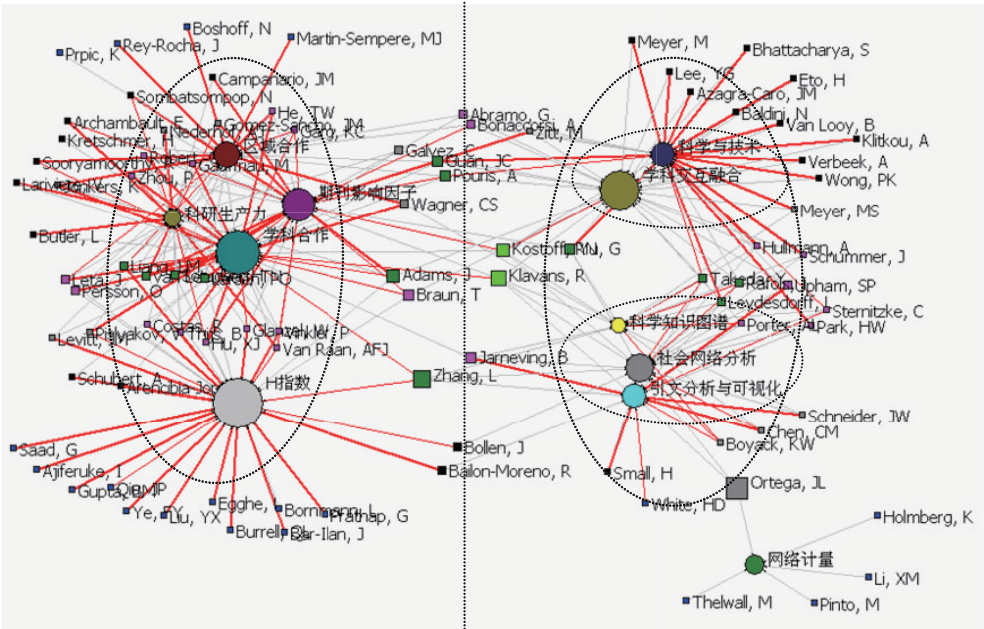


图3 ABCA 结果(科学计量学 2001—2010)

3.4.2 ABCA 与 AKCA 综合对比

科学计量学(2001—2010)进行文献关键词耦合分析所探测到的因子及其载荷如表6所示。AKCA的因子分析结构矩阵的可视化结果如图4所示。

前文中我们阐明了ABCA可以比AKCA探测到更多的研究主题,并对其原因进行了详细分析。在该阶段,该论断可以得到进一步的论证,ABCA探测

到11个因子,AKCA则探测到10个因子。在研究主题发现上,该阶段的研究结果与前文中论述的研究结论也保持一致。比较表6、图4与表5、图3,可以看到,在研究热点主题发现上AKCA的研究热点主题依次是:“影响因子”、“科学与技术”、“期刊指标”,ABCA的研究热点主题依次是:“H指数”、“科学与技术”、“学科合作”、“期刊影响因子”。除了“学科合作”外,其它研究主题基本一致。

表6 AKCA 因子及其载荷(2001—2010)

因子	载荷数	最高载荷	因子	载荷数	最高载荷
影响因子	31	0.79	科学指标与模型	11	1.01
科学与技术	29	0.86	科学交流模式	9	0.66
期刊指标	14	0.77	引文分析	9	0.97
网络计量与影响因子	13	0.98	未查明	1	0.37
科学指标与生产力	13	1.03	未查明	1	0.40

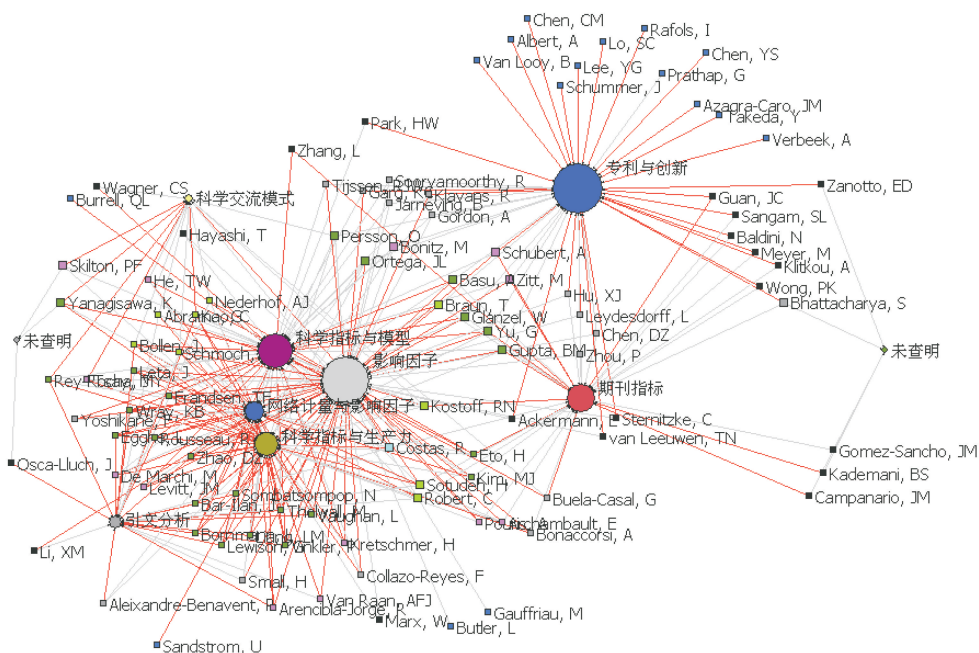


图4 AKCA 结果(科学计量学 2001—2010)

Kuusi 与 Meyer^[23] 研究发现,文献耦合分析很适合用于预测,例如运用于预测技术突破。另外一些学者也论证了耦合分析可以显示微弱的信号来进行研究前沿的探测,这是由于耦合分析并不像共被引分析选取过去的高被引文献,而是选取最近发表不久的有影响的文献作为分析基础^[24-25]。Strotmann 与 Zhao^[14] 等人坚持将作者共被引分析(ACA)与作者文献耦合(ABCA)分析结合起来预测学科发展,他们认为 ACA 未探测到而 ABCA 探测到的研究主题很有可能成为学科未来发展的趋势。比较表 3、图 1、表 4、图 2、表 5、图 3,本文认为,AKCA 也同样对研究领域的发展有一定的预测性,而且这种预测性并不会比 ABCA 弱。本研究的过程是,分别以 ABCA 方法揭示科学计量学(1991—2000)与科学计量学(2001—2010)的知识结构及其演进过程,并以该领域的的数据为样本展示 ABCA 与 AKCA 这两种分析方法的异同。在研究过程中发现,与科学计量学(2001—2010)的知识结构(表 5、图 3)更为接近的是 AKCA 揭示的科学计量学(1991—2000)的知识结构(表 4、图 2)。在 AKCA

结果(科学计量学 1991—2000)中探测到的研究主题在 ABCA 结果(科学计量学 2001—2010)中基本都可以找到相同或者相关的研究主题。例如,对比表 4 和表 3,AKCA 中的“科学计量指标”、“期刊计量指标”与 ABCA 中“H 指数”、“期刊影响因子”相对应;“科学与技术”与 ABCA 中“科学与技术”一致;“科研合作”、“科学合作与分布”对应于 ABCA 中的“学科合作”、“区域合作”;“区域计量”对应于 ABCA 中“区域合作”;只有“医学计量”未找到与之相对应的主题。因此可以认为,科学计量学(2001—2010)的知识结构在科学计量学(1991—2000)的 AKCA 中早已有所体现,换言之,科学计量学(1991—2000)经过作者关键词耦合分析可以在某种程度上预示科学计量学(2001—2010)的某些结构特性。因此,AKCA 是有一定预测性的,它可以预测学科在下一阶段的发展趋势。

另外,我们认为,ABCA 与 AKCA 结合起来会是一种探寻学科知识结构及其发展的研究方法。在研究过程中,我们发现在 ABCA 结果(科学计量学 1991—2000)与 AKCA 结果(科学计量学 1991—

2000)中都出现的研究领域在科学计量学(2001—2010)中也有出现,而仅仅在 ABCA 结果(科学计量学 1991—2000)或者 AKCA 结果(科学计量学 1991—2000)中出现的研究领域在科学计量学(2001—2010)中出现的可能性较小。科学合作研究与科学技术指标这两大研究领域在第一个时间段的 ABCA 结果(表 3、图 1)与 AKCA 结果(表 4、图 2)中都有出现,并在第二个时间段成为科学计量学的两个主流研究领域;而仅仅在 ABCA 结果中出现的“科学创新”、“共词分析”、“学科领域计量”、“专利分析”、“文献计量应用研究”、“文献基础规律研究”以及在 AKCA 中出现的“医学计量”都未在科学计量学(2001—2010)的主题探测中被发现。而仅仅在 ABCA 中出现的“引文分析”与“科学知识图谱”相互作用形成了科学计量学在 2001—2010 时间段的主流研究领域,仅仅在 AKCA 中出现的“科学与技术”也逐渐演变为科学计量学在 2001—2010 时间段的另一主流研究领域,这并不影响我们的结论。

毫无疑问,在 2001—2010 年间,“科学与技术”已经成为科学计量学的一个重要的研究领域,因为它不仅在 ABCA 结果中出现,也出现在 AKCA 的分析结果中。该领域的重要性在 AKCA 结果(科学计量学 1991—2000)中就已有所体现,在 AKCA 结果中它位于图谱的中央,成为了联系科学合作与分布研究、科学计量指标、期刊计量指标的一个关键节点因子(图 2)。我们分析,该因子的重要性主要跟 International Conference on Science and Technology Indicators(科学与技术指标国际会议)的密集召开有很大关系。在 1991—2000 年间,《科学计量学》收录了两届该会议的论文(第 4、5 届),而在 2001—2010 年间《科学计量学》连续收录了五届该会议的论文(第 6、7、8、9、10 届)。这一方面说明国际会议对于推动学科发展甚至仅仅是学科某一方向发展的重要作用;也从另一方面论证了我们的因子探析结果与实际情况基本吻合,我们的研究结论是可信的。

4 结语

本文引入作者关键词耦合分析,结合作者文

献耦合分析来揭示科学计量学 1991—2010 年间的研究状况。以 10 年为限,将这 20 年分为两个时间段,分别揭示这两个时间段的知识结构及其演进,并以这 20 年间的的数据为样本分析比较 ABCA 与 AKCA 这两种分析方法的异同。

研究发现,在 1991—2000 年间,科学计量学的研究可以划分为文献研究与政策研究,而这两个研究领域以及研究主题之间的相互作用较弱。科学计量学在 2001—2010 年间的发展中,其知识结构变得更加清晰明朗,其研究主题之间变得更加亲密,相互作用力也明显加强。该阶段有四个较为明显的研究区域:科学与技术指标、科学合作研究、科学与技术交融研究、引文分析与可视化研究。

本文对 ABCA 与 AKCA 这两种分析方法,从作者排名相关分析、研究主题探测、余弦相似度计算、研究主题变迁等角度,以实证的方式,论证 ABCA 与 AKCA 间存在高度相关性。对作者的平均耦合频次排名进行相关分析发现 ABCA 与 AKCA 是相关的,这种相关性会随着样本的增大而有变强的趋势;在研究热点主题探测上,ABCA 与 AKCA 的结果基本是一致的;计算 ABCA 与 AKCA 构建的耦合矩阵的余弦相似度,尤其是提取二者共有作者构造新的耦合矩阵并进行余弦相似度的计算,表明这两种矩阵具有很强的相关性;两个时间段、两种分析方法的研究主题迁移状况表明,AKCA 与 ABCA 一样具有一定的学科发展的预测性。本文在论证 ABCA 与 AKCA 存在着众多相同或者相似之处的同时,也发现二者之间有些许不同。在因子分析时,对二者进行相同的参数设置,并完全采取相同的分析过程,结果却显示 AKCA 比 ABCA 具有更为理想的因子分析模型拟合结果。在研究主题探测方面,ABCA 可以探寻到比 AKCA 更多的研究主题,我们将其原因归纳为三点:耦合的数据基础的数量不同;耦合发生的学科差异;年代追溯性的差异。

本文认为,ABCA 与 AKCA 二者不可以完全互为取代。虽然 ABCA 的实证文献较少,但这种方法很早便被提出,其基本原理和思想也深受广大学者的认同;而本文的研究显示,AKCA 似乎能

比 ABCA 显示更多的信号来反映学科的技术突破以及研究前沿的发展。因此,ABCA 与 AKCA 结合起来会是探寻学科知识结构及其发展的一种理想研究方法。

参考文献

- [1] Small H. Cocitation in the scientific literature: A new measure of the relationship between two documents[J]. *Journal of the American Society for Information Science*, 1973, 24(4): 265–269.
- [2] Marshakova S I. System of document connections based on references[J]. *Nauch – Techn. Inform*, 1973, 2 (6): 3–8.
- [3] 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.
- [4] Persson O. All author citations versus first author citations[J]. *Scientometrics*, 2001, 50(2): 339–344.
- [5] Schneider J W, Larssen B, Ingwersen P. Comparative study between first and all-author co-citation analysis based on citation indexes generated from XML data[C]// *Proceedings of the Eleventh International Conference of the International Society for Scientometrics and Informetrics*, June 25–27, Spain, 2007: 696–707.
- [6] Eom S. All author cocitation analysis and first author cocitation analysis: A comparative empirical investigation[J]. *Journal of Informetrics*, 2008, 2(1): 53–64.
- [7] Ahlgren P, Jarneving B, Rousseau R. Requirements for a cocitation 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.
- [8] White H D. Author cocitation analysis and Pearson's r [J]. *Journal of the American Society for Information Science and Technology*, 2003, 54(13): 1250–1259.
- [9] Rousseau R, Zuccala A. A classification of author co-citations: Definitions and search strategies[J]. *Journal of the American Society for Information Science and Technology*, 2004, 55(6): 513–629.
- [10] Pamela E S. Scholarly communication as a socioecological system[J]. *Scientometrics*, 2001, 51(3): 573–605.
- [11] White H D, McCain K W. Visualizing a discipline: An author cocitation analysis of information science, 1972–1995 [J]. *Journal of the American Society for Information Science*, 1998, 49(4): 327–355.
- [12] Chen C M. Visualizing semantic spaces and author co-citation networks in digital libraries[J]. *Information Processing & Management*, 1999, 35(3): 401–420.
- [13] Kessler M M. Bibliographic coupling between scientific papers[J]. *American Documentation*, 1963, 14(1): 10–25.
- [14] Zhao D Z. Evolution of research activities and intellectual influences in information science 1996–2005: Introducing author bibliographic Coupling Analysis [J]. *Journal of the American Society for Information Science and Technology*, 2008, 59(13): 2070–2086.
- [15] McCain K W. Mapping authors in intellectual space: A technical overview[J]. *Journal of the American Society for Information Science*, 1990, 41(6): 433–443.
- [16] 马瑞敏. 基于作者学术关系的科学交流研究[D]. 武汉: 武汉大学, 2009. (Ma RuiMin. A research on scientific communication on the basis of author academic relationships[D]. Wuhan: Wuhan University, 2009.)
- [17] Zhao D Z, Strotmann A. Comparing all-author and first-author co-citation analyses of Information Science[J]. *Journal of Informetrics*, 2008, 2(3): 229–239.
- [18] Zhao D Z, Strotmann A. Information science during the first decade of the Web: An enriched author co-citation analysis [J]. *Journal of the American Society for Information Science and Technology*, 2008, 59(6): 916–937.
- [19] Hair J F, Anderson R E, Tatham R L, et al. *Multivariate data analysis* (5th ed.) [M]. Upper Saddle River, NJ: Prentice Hall, 1998.

小词典

Altmetrics

Altmetrics 一词近来受到业内学者关注。该词来源于 alternative + metrics, 由美国图情专业博士生 Jason Priem 2010 年最早使用。有人将其翻译成“替代计量学”, 有人翻译成“另类计量”或者“可选计量”, 目前尚未形成确定的翻译。“替代”的解释不完全贴切, 而且也不一定成为“学”, 它应该是相对于影响因子而言的另一种计量方法。

传统的文章引用率和期刊影响因子这样的评价指标很有局限性, 这些指标通常需要有几个月或者是几年的积累, 而 altmetrics 可以快速收集目标数据, 让研究人员几乎实时看到文章、数据集或博客通过各级学术生态系统的传播情况。Altmetrics 常用于文章的计量, 但是也可以用于人、期刊、图书、数据集、网站等的评价, 所计量的指标包括浏览(网页浏览和全文下载)量、话题讨论量、收藏(书签)量、引用量等。

Altmetrics 的出现, 说明社交媒体在学术交流中的作用越来越受到人们的重视, 图书馆员和图书馆学者应该重视这样的变化, 调整自己的服务和研究方向。

资料来源

1. 维基百科, Altmetrics, <http://en.wikipedia.org/wiki/Altmetrics>.
2. Altmetrics 一词的出现再次引发科研影响力评价的问题, <http://blog.sciencenet.cn/blog-1035376-725891.html>.

(三牛)

-
- [20] Eom S B, Farris R S. The contributions of organizational science to the development of decision support systems research subspecialties[J]. Journal of the American Society for Information Science, 1996, 47(12): 941-952.
 - [21] Zhao D Z, Strotmann A. Can citation analysis of web publications better detect research fronts[J]. Journal of the American Society for Information Science and Technology, 2007, 58(9): 1285-1302.
 - [22] Zhao D Z. Towards all-author co-citation analysis[J]. Information Processing & Management, 2006, 42(6): 1578-1591.
 - [23] Kuusi O, Meyer M. Anticipating technological breakthroughs: Using bibliographic coupling to explore the nanotubes paradigm[J]. Scientometrics, 2007, 70(3): 759-777.
 - [24] Glänzel W, Czerwon H J. A new methodological approach to bibliographic coupling and its application to the national, regional, and institutional level[J]. Scientometrics, 1996, 37(2): 195-221.
 - [25] Bassecouard E, Lelu A, Zitt M. Mapping nanosciences by citation flows: A preliminary analysis[J]. Scientometrics, 2007, 70(3): 859-880.

宋艳辉 杭州电子科技大学管理学院讲师, 中国科学技术信息研究所博士后。

通讯地址: 杭州市下沙高教园区杭州电子科技大学管理学院。邮编: 310018。

武夷山 中国科学技术信息研究所副所长, 研究员, 博士生导师。

通讯地址: 北京市复兴路 15 号中信所情报方法研究中心。邮编: 100038。

(收稿日期: 2013-03-28; 修回日期: 2013-06-20)