文章编号: 1003- 2053(2005) 02- 0149- 06

悄然兴起的科学知识图谱

陈 悦, 刘则渊
(大连理工大学人文社会科学学院,辽宁大连 116024)

摘 要: 科学知识图谱是显示科学知识的发展进程与结构关系的一种图形。它的悄然兴起, 一方面是揭示科学知识及其活动规律的科学计量学从数学表达转向图形表达的产物, 另一方面又是显示科学知识地理分布的知识地图转向以图象展现知识结构关系与演进规律的结果。这里, 在介绍有关科学知识图谱基本概念的基础上, 从数据库、数据格式及存取, 数据分析算法, 可视化和互动设计, 科学计量学等方面阐述了有关科学知识地图绘制的最新进展, 并展望了其应用前景。其进展表明, 无论是对于科学技术研究, 还是对于企业技术创新, 科学知识图谱都是一种有效的知识管理工具。

关键词: 科学知识图谱; 科学计量学; 知识地图; 信息可视化
中图分类号: G301 文献标识码: A

现代科学技术的突飞猛进, 并伴随国际互联网络发展而在世界上的迅速传播, 导致全球知识呈爆炸式的增长, 由此也带来了知识与信息选择的困难。一个旨在将知识和信息中令人注目的最前沿领域和学科制高点, 以可视化的图像直观地展现出来的新兴交叉学科悄然兴起。它把现代科学技术知识的复杂领域通过数据挖掘、信息处理、知识计量和图形绘制而显示出来, 使研究人员得以在世界知识版图中了解自己研究领域的所在位置, 对于如何选择感兴趣的新领域也不再困难。这个以科学学为基础, 涉及应用数学、信息科学及计算机科学诸学科交叉的领域, 是科学计量学和信息计量学的新发展。这个极其重要、并有着广阔应用前景的交叉领域被称作“Mapping Knowledge Domains”。我们根据其性质和特征, 倾向于译为“科学知识图谱绘制”。

1 科学知识图谱的由来和概念

科学知识图谱, 是显示科学知识的发展进程与结构关系的一种图形。由于它是以科学知识为计量研究对象的, 所以属于科学计量学 (scientometrics) 的范畴。当它在以数学方程式表达科学发展规律的基础上, 进而以曲线形式将科学发展规律绘制成二维图形时, 便成为最初的知识图谱。从这个意义上说, 用定量统计方法发现科学知识指数增长规律的科学计量学奠基人普赖斯 (Derek J de Solla Price)[1], 也是科学知识图谱的早期开拓者。著名德国科学计量学家赫尔德若·克里奇默 (Hildrun Kretschmer) 关于科学合作的三维空间模型研究[2][3], 大大地推动了科学知识图谱的发展。因此, 知识图谱绘制是科学计量学的发展与创新。

科学知识图谱又同绘图学和地图学有一定关系, 但知识图谱的概念与知识地图的概念并不完全相同, “地图”是一个以二维或三维空间形式显示地形和人类活动及相关特征的地理学概念, 知识地图最初是表现科学技术活动与知识的地理分布状况的地图; 而“图谱”是图像以一定空间形式在一定时间范围中展现与变化的系统概念, 虽然可以把知识地图作为知识图谱的一种形式, 但知识图谱比知识地图更能揭示知识之间的联系及知识的进化规律。

知识地图的概念, 在狭义上就是表达科学技术知识或一般知识资源地理分布状况的地图。美国国家科学基金会早在 1970年代就出版了关于科学基金的地理分布的报告, 并论述了科技分布对地区经济的影响。由此科学研究地理学、高技术地理学作为经济地理学分支在 1980年代得到发展[4]。美国

收稿日期: 2004- 09- 14 修回日期: 2004- 10- 22
基金项目: 国家自然科学基金资助项目(7031027)
作者简介: 陈 悦 (1975—), 女, 辽宁大连人, 讲师, 博士生, 研究方向为科学学理论与技术创新管理。
刘则渊 (1940—), 男, 湖北恩施人, 教授、博士生导师, 研究方向为科学学理论与技术创新管理。

捷运公司最早的知识地图是一张展示知识资源地理分布的美国地图[5], 这就是知识地图的雏形。之后, 带有索引号或用其他方式表示层次关系的表格和文件, 以及用来表示信息资源与各部门或人员之间关系的信息资源管理表和信息资源地理分布图, 都是知识地图的早期形式。随着信息技术的迅速发展, 知识地图进入了电子时代, 在 Internet和 Intranet上普遍使用的超文本链接和应用链接就是知识地图的简单形式。这时, 很多绘制知识地图的工具应运而生, 如 LotusNotes, IBM 的 KnowledgeX 和微软的 Visio等, 它们都是基于数据库来绘制知识地图, 有利于知识地图的动态更新和扩展, 这就突破了局限于描述知识地理分布的知识地图界限, 并逐渐演化为涵义与内容更加广泛的知识图谱了。

有关知识地图的定义有多种。例如, 维尔(E. F. Vail)将知识地图定义为"可视化地显示获得的信息及其相互关系, 它促使不同背景的使用者在各个具体层面上进行知识的有效交流和学习。在这样的地图中包括的知识项目有文本、图表、模型和数字"。而为企业提供知识地图解决方案的萨拉蒙德(Salamander)组织将知识地图定义为"对企业的积极的可视化的描述"。他们的定义都强调了知识地图的功能。借鉴知识地图的定义, 本文将知识图谱定义为可视化的描述人类随时间拥有的知识资源及其载体, 绘制、挖掘、分析和显示科学技术知识以及它们之间的相互联系, 在组织内创造知识共享的环境以促进科学技术研究的合作和深入。

知识图谱所描绘的对象主要包括: (1)从事科学技术活动和作为知识载体的人, 包括科学家、技术专家、项目组、实践团体或某一知识领域共同体; (2)显性或编码化的知识, 如论文、专利、所学课程、数据库或类似的应用等; (3) 过程或方法, 包括研究问题和解决问题的过程或方法、组织的业务流程, 以及相关的知识投入等。

2 知识图谱绘制的最新进展

科学技术的发展已经达到这样一种状态, 即必须将已有的零碎知识整合起来, 科学技术才能继续发展。由此, 美国科学院组织了大批专家从事这个新的交叉科学领域的研究, 并于 2003年 5月 9—10日在加利福尼亚 Irvine大学的国家科学院贝克曼(Beckman)中心举办了主题为"知识图谱测绘"的大型学术研讨会, 出席会议的研究者来自于计算机领域、信息与认知科学领域、数学领域、地理学领域及其他领域, 各个学者、专家从不同方面介绍了他们有关知识图谱的最新研究成果, 共发表了 20多篇学术论文。这些论文应用各种计量方法与制图工具分析和处理不断增长和进化的知识数据, 绘制出各种类型的人类科学知识图谱。本文归纳出如下几个方面。

2.1 数据库、数据格式和存取

以美国科学情报研究所(ISI)名誉所长加菲尔德(Eugene Garfield)博士为首的科学团体创建了一系列关于知识域(动态系统 smallworld/milgram, 信息可视化 tufte, 共引 small, 引文耦合 kessler, 和科学计量学 *scientometrics*)资料数据库。Garfield博士认为"引文数据的使用在书写科学的历史", 由此利用他们开发的 HistCite软件包, 通过 ISI光盘引文索引(SCI, SSCI, and/or AHCI)形成某一学科发展的历时的图谱。HistCite系统是一个很好的引文历史分析工具, 当在 WoS上显示出一个有标记的列表时, 对每一个源文件都生成包括所有被引文献的专家文件, 这些引文收集被存储成由 HistCite处理生成的 ASCII文件, 用以产生历时代和其他类型表格, 以及显示出在本收集之内和之外被引用最多的文献的编年图表。

Google公司的莫尼卡·汉金格尔(Manika Henzingger)和 NEC的斯蒂文·劳伦斯(Steve Lawrence)认为大量的信息网络是瘫痪的, 即不能提供现成的信息, 必须要对它们进行数据格式和存取方式上的加工, 以获取有用的信息, 由此介绍了《伊利诺斯研究》(Illinois Research, 美国期刊网)的超级链接分析, 如何从网上抽取样本, 进而打造出环球网图表模型和历时间的聚焦式缓慢动态模型, 最终发现科学研究共同体。

康奈尔大学保罗·金斯帕格(PaulGinsparg)介绍了 http://aXiv.org网站的基本状况, 即按照物理学、数学、非线性科学、计算机科学四个大类进行分类, 并且在其内部细分为很多学科, 以便于访问该网站的学者寻找自己所需的学科的文章, 其中介绍了有关文章分类与支持向量机制的数据库构建方法, 以此来说明机制学习方法怎样用来分析, 构建, 维护和发展一个大的再现学术文献文集, 这个方法可以训练一个支持向量机制文本分离器以从一个大的资源中摘录出新出现的研究领域。

2 2　数据分析算法

圣达菲 (Santa Fe) 研究所马克· 纽曼 (Mark Newman) 在《合作网络和科学共同体的结构》一文中分析了生物医学、物理学和数学三个科学合作网络的结构, 揭示出不同研究领域的相似和差异。指出数据库网络的节点是科学家, 如果两位科学家共同完成一篇论文, 那么这两位科学家互相关联。应用这些网络可以回答许多关于合作模式的问题, 如: 作者论文的数量、与其他作者合作的数量、网络中科学家之间的象征性距离, 以及合作模式如何随时间和学科的变化而变化等。文章中也总结了其它有关共同合作模式的研究成果。

科罗拉多大学托马斯· 兰道尔 (Thomas K. Landauer) 在《从段落到图表》一文中认为, 整个文章语义内容的相似性能够比题目, 摘要或摘要提供更吸引人的东西, 潜在的语义分析提供了一种有效地用于反映相似词和词语组合意思的维度压缩方法, 文章中例举了许多例子来说明在可视化中潜在语义分析的应用, 并提出未来需求的远景。

斯坦福大学托马斯· 格瑞菲斯 (Thomas L. Griffiths) 和加利福尼亚大学马克· 斯蒂韦尔斯 (Mark Steyvers) 在《寻找科学主题》一文中, 阐述了知识图谱中有关主题动态化的问题。文章介绍了布雷 (N. Blei) 和乔丹 (Jordan) 的文本生成模型, 接着引入马尔可夫链蒙特卡罗运算法则进行推论, 并运用这一法则分析 PNAS的摘要, 应用贝叶斯定理模型选择法来确定主题的数目。这样萃取出来的主题抓住了资料中的有意义的结构, 与作者提供的种类名称是一致的, 并略述了这一分析的进一步应用, 包括通过检验当时的动态以及摘要来阐明语境, 进而鉴别"热门主题"。

华盛顿大学伊丽娜· 埃洛舍娃 (Elena Erosheva), 卡耐基梅隆大学斯蒂芬· 费恩伯格 (Stephen Fienberg)、约翰· 拉菲尔德 (John Lafferty), 汤姆· 闵卡 (Tom Minka) 合写的《科学刊物的混合成员模型》一文介绍了混合成员模型, 并将注重文献的混合成员模型与 PNAS的数据材料共同应用于亚瑟 (Arthur) 的科学图谱领域, 把研究焦点放在生物科学领域, 对摘要和注释部分进行分析。模型的建立应用了聚类分析的方法, 选取 PNAS数据库中的生命科学类科学出版物为样本, 抽取生命科学类出版物中的 19个主题作为研究样本, 将其进行聚类, 在进行包括潜在变量标准及样品方案等在内的一系列假设的基础上, 建立了以参考书目和文字为元素的科学出版物的混合元素模型。用此模型对 PNAS数据库中 19个生命科学类科学出版物进行分析, 并应用贝叶斯定理和包括变化的近似算法及遗传算法在内的直接的近似值方法对模型分析的结果进行评估, 从变化近似算法和遗传算法的八个方面获得一些比较结果。

康奈尔大学乔· 克林伯格 (Jon Kleinberg) 介绍了网址内容随时间呈现信息爆发流。为了组织好这种即时变化的信息, 他提出了网址的数据分级用法, 创立了网址内容的即时编辑器, 以使网络内容得以即时升级。

德雷克塞尔大学凯瑟琳· 迈克卡因 (Katherine W. McCain) 介绍了绘制知识地图的两种方法——作者共引分析 (author cocitation analysis, ACA) 和知识诱出—卡片分类法 (knowledge elicitation-card sorting KE), 从而产生 ACA簇群 & 地图, PFNet作者网络, 卡片分类簇群 & 地图, 并对两种方法产生的结果进行了比较, 得出 ACA (地图, PFNets) 和 KE (卡片分类), 提供了软件工程的完整视角, 提供了交叉确认, 所以将文献计量和知识诱出技术合并是一种非常有效的绘制知识图谱的方法。

科罗拉多大学的西蒙· 丹尼斯 (Simon Dennis) 介绍了句段范例模型简称 SP模型 (syntagmatic paradigmatic model), SP模型是基于记忆的句子加工过程的记载, 用来自动从未注解文章中提取命题信息。SP模型假设人们存储了大量句例。当要理解一个新句子时, 类似的句子会从记忆中取回并和新句子根据字符串编辑理论排在一起。可以认为这个排列是对句子的外延理解。采用这种提取命题信息的方法, 模型不仅能通过清晰地陈述文章中的相关事实而回答出问题, 而且能够利用巧合推断, 这里含蓄推断是作为机制中的新兴特性发生的。

康奈尔大学的乔纳森· 埃赞 (Jonathan Aizen) 等人认为, 网站能够提供给人们许多方便, 诸如时事、电影、音乐等信息并支持下载功能, 而判断一个网站的建立是否成功则应主要统计其点击率。他们在"有关网络点击率"的文章中引入了"平均点击率"概念, 以求对某网站的评价结果更为客观、有效。文章通过构建数学模型、举例等方法介绍了以下几种评价网站的新方法: 网站的平均点击率; 随时间跟踪兴趣; 匹配外部事件的兴趣变化等。同时还介绍、说明了评价网站的一些相关工作, 在文章最后

还给出了一些有利于网站评价的指导性建议。

2 3　可视化和互动设计

新罕布什尔大学数字可视化研究办公室主任克林· 韦尔 (Colin W are)在《视觉思维和视觉注意目标》一文中, 介绍了视觉思维的体系结构, 搜寻模式、二进制编码理论, 运转着的存储器的组成结构, 注意力的探照灯模型, 视觉思维的嵌入过程, 知识的成本等。

德雷克塞尔大学晁辰美 (Chao Chenmei) 对知识可视化进行了大量有益的探索, 尤其是针对科学中的范式、悖论和疑难等。

宾夕法尼亚州立大学的阿兰· 迈克伊奇仁 (A l an M. M acEachren)等人在《地理科学概念可视化的建构与分享》一文中, 认为科学知识的表达必须反映知识结构的动态特性及科学概念间的不断进化着的关系网络, 文章阐述了用来支持科学知识的建构、联系、修正及应用的动态的、可视化的理论与工具, 尤其是以人与环境相互作用的领域为例推导出获得及研究可以支撑各种科学合作的基础概念的理论工具。这种可视化的方法结合了地理学的可视化技术与概念图谱的工具, 以使知识结构化这些工具帮助单个的研究者描述其知识的建构过程以使其相关的合作的团队可以将其综合成普遍的概念, 进而帮助在相关领域研究的科学家可以共用一个知识网络的途径来提高自己的研究。

印第安纳大学凯蒂· 鲍娜尔 (Katy Börner), 杰格尔· 迈鲁 (Jeegar T. M aru) 和罗伯特· 高登斯顿 (Robert L Goldstone)在《作者与论文网络的同时演化》的一文中, 回顾了对科学演化的结构和动力学原理, 在此基础上, 引入了一种通用的过程模型, 揭示出合著和引文网络同时增长的规律, 并以 PNAS数据库中时长 20年的数据进行验证。但是, 这个模型存在着引文的幂律分布导致的系统性误差。为解决这个误差, 作者建立了 TARL模型 (主题、退化、循环联系)进行解释。这个模型包含着作者和论文在主题方面的分隔、作者在引用近期文献时所产生的误差、以及作者对所阅读文章中的参考文献的依赖心理。

印第安纳大学科坦· 梅恩 (Ketan K. M ane)和凯蒂· 鲍娜尔 (Katy Börner)在《PNAS的主题和主题爆炸图谱研究》一文中, 演示了克林伯格 (K lein-berg)的爆炸探查算法的应用和共生词分析法和图示技术, 以绘制主要研究主题和发展趋势的发现地图。用这一方法来分析和绘制 PNAS在 1982~ 2001年间中所发表的所有论文的二维共生主题词网络图, 如下图所示。六位该领域专家就所得到的最后地图进行了研讨, 试图重建 PNAS所蕴藏着的科学主要研究领域的演化图。文章通过对 1982~ 2002年间发表于 PNAS的 47072篇文章构成的大规模数据库的文本分析, 描述了高频词和爆炸词的识别, 分析这些词之间最重要的相关性, 并绘制平面图来显示和解释主要研究主题和它们的动态过程。

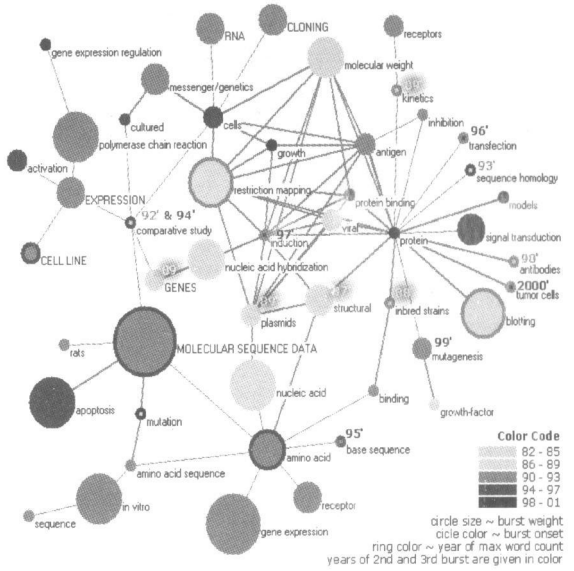

图 1　1982— 2001年 PNAS发表论文的二维共生主题词网络图

注: 这个网络中每个关节点都代表 50个高频词和爆炸词之一, 关节点圆的大小尺寸代表了这个词所达到的最大爆炸水平。颜色则代表了词被经常使用和达到最大爆炸水平的年份。五个时间段和代表颜色分别是: 1982~ 1985年, 绿色; 1986~ 1989年, 黄色; 1990~ 1993年, 橙色; 1994~ 1997年, 红色; 1998~ 2001年, 黑色。某一词出现频率最大年份和最开始爆炸年份通过圆边缘和内部的颜色分别表示。例如, 分子序列数据由一个相当大的圆来表示, 圆的内部是橙色的而边缘是红色的, 这显示该词在 1990~ 1993年间突然最大程度爆炸, 而其后 1994~ 1997年间是使用最频繁年份。

印第安纳大学情报计量学院菲利普· 麦克泽尔 (Filippo M enczer)认为, 当试图解释 WWW 的无标度布局并运用链接分析来分辨重要信息来源时, 如何集中控制文献网络已显得至关重要。现有的信息网络增长模型研究集中于链接的结构但却忽略了节点的内容。现存模型不能阐明信息网络的某些重要特征, 即所链接的文献之间的文本近似性状况。由此, 他设计了一个更现实的模型, 它可以通过网络的

知名度和内容来完成链接。网页和文本文献的网络分布状况都可以通过这个模型来准确预测。

2.4　科学计量学

科学知识图谱的绘制已成为科学计量学的前沿,在这方面作为创建国际科学计量学与情报计量学学会(International Society for Scientometrics and Informetrics,简称 ISSI)的首倡者及第一任会长克里奇默(Dr. H. Kretschmer)应是代表人物。她将科学计量学、情报计量学与心理学等科学研究方法相结合取得了一系列的成果,特别是对科学技术合作的计量学研究方面取得了重大突破,建立了科学技术合作的三维空间模型,并使之形成可视化的图像,走在了科学学和科学计量学的最前沿,在国际科学计量学界产生了广泛的影响。

克里奇默博士在"国际合作网络的构架"研究中,将新的数学方法引入科学计量学领域,根据 Metzger心理学中的构型(configuration)理论,借助非线性函数形象地描述了科学家合著网络构型的三维图形,揭示出高层次人才比低层次人员更容易合作而取得更多成果,呈现"物以类聚"而非"文人相轻",令人耳目一新。在"合著网络构型的三维图形"研究中,从非线性函数得到的三维图象,并用来描述科学合作的社会网络构型,并建立函数: $Z = Const \cdot (A+1)^{\alpha} \cdot (Amax - A + 1)^{\beta} \cdot (B+1)^{\gamma} \cdot (Bmax - B + 1)^{\delta}$ ($A = |X - Y|$,互补项 $B = |X + Y|$)。在"WEB上合作的可见度(visibility of collaboration on the web)"研究中,以 64个被检索 COLLNET成员的参考文献和 WEB数据为样本,通过 Google(完全线性)和 Alltheweb(非连续线性)搜索,获得 WEB可见率多著者出版物的分布。

3　应用及前景展望

知识图谱的测绘由于计算机处理能力的提高和大量电子化形式出现的出版物、专利、授权和其它数据而日益便利。知识图谱研究仍处于起始阶段,在可预见的未来还不能作为人类判断、搜索和决策的替代品,但目前一些工具已用来帮助人类完成某些重要方面的判断和决策。尤其是他们能够模拟人类的数据分析模式,而这是其他方法无法办到的。

CHI研究公司的弗朗西斯·纳林(Francis Narin)利用图谱的方法分析了科学论文、技术专利和财政绩效之间的转化关系,并介绍 CHI研究是一个高度专业的引文研究顾问,它进行三种类型的文献计量:科学论文引文分析,技术专利分析和引用论文专利与股票价值的联系。

Pacific Northwest国家实验室的本斯·海特兹乐(Beth Hetzler)阐述了可视化分析的价值,分析者的环境,分析的环境,分析者和分析环境的尴尬处境及信息超载的内涵,介绍了 IN-SPIRE基本工具的使用和功能,如图 2所示。

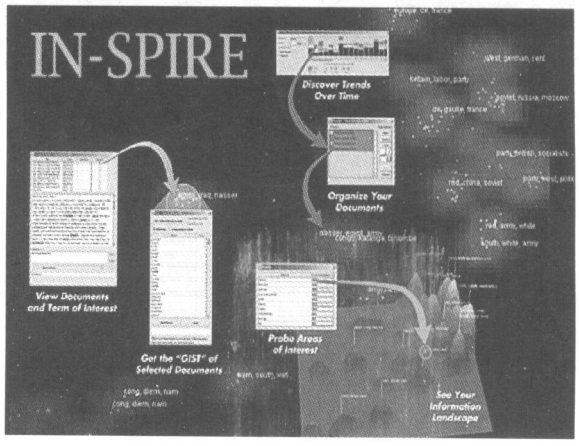

图 2　IN-SPIRE基本工具的使用及功能

注: IN-SPIRE发现工具整合了交互式信息可视化与询问功能。Galaxy可视化技术(右上角),其中的点代表文件和围绕在中心主题而形成的簇团,如同天上的星系。Theme View可视化技术(右下角)提供了一种能更快获取数据收集视觉效果的方法。用户可以看到一个简洁的地图,最高突起点代表在收集中最热门的主题。

资料来源: http://in-spire.pnl.gov/about.html

微软的苏珊·杜梅恩(Susan Dumain)介绍了一种能够提高搜索效率的算法和界面,这种方法能够超越普通的列表,帮助主体搜索,并能进行信息分析和信息发现。并以 SWISH系统, GridViz系统, SIS系统为实例加以说明。

Sandia国家实验室的凯文·包雅克(Kevin Boyack)介绍了基于指数的 PANS描述中的指数由三个来源: NSF和 OECD的工作指数、有关基金及其影响研究的指数和大规模输入—输出指数,利用 ISI/SCIE, Medline, NIA和 PNAS数据库进行数据融合,分析了基金类型,资助的种类及其影响,进而利用百分比统计及聚类方法测绘出受资金影响最大的图表,应用 VxInsight(™)软件(VxInsight(™)是一个在大型数据库中挖掘关系的有用工具。大多数据恢复和数据挖掘软件只是在数据库中发现信息,他们只

能表示数据元素。而 V_x Insight$^{(TM)}$能够帮助分析者揭示具有战略意义联系和模式,从而使之成为一个重要的知识管理工具)测绘出下图,此地图能反映出时间要素,即反映出科学是如何随着时间向前发展的。

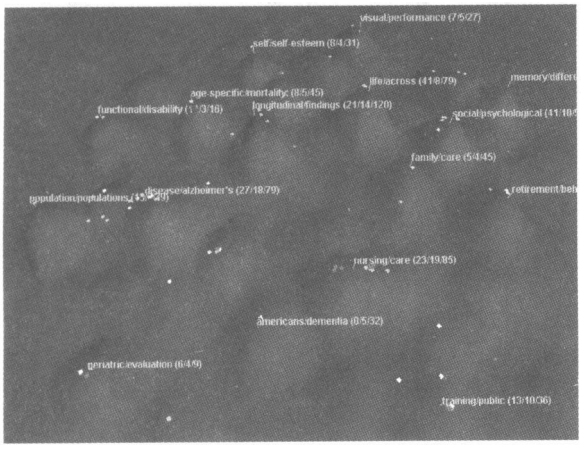

图 3　期刊文章数量与资助基金的关系

注:这张图显示了各类期刊文章的数量与资助基金之间的动态关系,较大的突起代表较多的文章,颜色较浅的代表资助较小,颜色较深的代表资助较大。

资料来源: Kevin W Boyack, Katy Börner. Indicator-Assisted Evaluation and Funding of Research: Visualizing the Influence of Grants on the Number and Citation Counts of Research Papers [J]. ASIST, 54(5): 447- 461.

另外,知识图谱测绘技术还能够用于以下几个方面:(1)明确主要研究领域、专家、机构、授权、出版物、期刊和特定研究领域的其他关键主题词,以及这些主题词之间的内部联系;(2)明确各研究领域之间的知识输入与知识输出;(3)科学研究领域的动态化(如增长速度,多样化等);(4)信息生产和传播中的经济因素;(5)科学社会网络;(6)明确战略的作用和政府项目的应用研究。

企业目前为了保持竞争优势,正在利用第一代数据分析和可视化技术控制知识和技术的流动(这是知识管理的最本质所在)。通过科学家和研究者的努力以及包括政府机构,企业和社会中每一个对科学进步感兴趣的人员的合作,我们相信当前的科学图谱技术能够得到显著提高和更成功的应用。

参考文献:

[1] D Price. Science Since Babylon[M]. Yale University Press, 1961.

[2] H Kretschmer. Coauthorship networks of invisible colleges and institutionalized communities[J]. SCIENTOMETRICS, 1994, 30(1): 363- 369.

[3] H Kretschmer. Types of two-dimensional and three-dimensional collaboration patterns[A]. C A Macias-Chapula. Proceedings of the Seventh Conference of the International Society for Scientometrics and Informetrics[C]. Mexico: Colima, 1999. 244- 257.

[4] A J Scott, M Storper. High technology industry and regional development: a theoretical critique and reconstruction[J]. International Social Science Journal, 1987, (1): 215- 230.

[5] 洛埃特·雷迭斯多夫. 科学计量学的挑战:可交流的发展、测度和自组织[M]. 北京: 科学技术文献出版社, 2003.

The rise of mapping knowledge domain

CHEN Yue, LIU Ze-yuan

(School of Humanities and Social Sciences, Dalian University of Technology, Dalian 116024, China)

Abstract: Mapping knowledge domain is a kind of graph showing the relationship between evolution and structure for science knowledge. Its rise shows that, one hand, the research direction for the scientometrics revealing the law of sciences knowledge and science activity is turning from mathematics expression to graph expression, on the other hand, the display for knowledge map is turning from geography distribution to knowledge structure relationship and evolution rule. Basing on the basic concepts of knowledge map, this paper introduce the lasted improvement in mapping knowledge domain on three aspects, which are Data bases& data format& access, Data analysis algorithms, Visualization& interactive design and scientometrics. Its Promising applications showing that mapping knowledge domain is an effective knowledge management tool to science& technology research and enterprise technology innovation.

Key words: mapping knowledge domain; scientometrics; knowledge map; information visualization