

基于科学知识图谱的加菲尔德学术影响分析

王金鹏 王慧杰

(华中师范大学信息管理系, 湖北武汉 430079)

摘要: 以科学知识图谱为研究视角, 在美国 ISI Web of Science 引文数据库中, 以“尤金·加菲尔德”为检索关键词, 进行被引作者检索, 并以检索出的所有文献构造原始数据集, 借助 Bibexcel、Ucinet、CiteSpace 等科学计量和信息可视化工具, 从一个较为新颖的角度更加直观和客观地分析尤金·加菲尔德博士的学术地位和影响, 得出引文索引和引文分析奠定了加菲尔德的学术地位和影响力等一系列结论。

关键词: 尤金·加菲尔德; 科学知识图谱; 科学计量学; Bibexcel; Ucinet; CiteSpace

中图分类号: G301

文献标识码: A

DOI: 10.3772/j.issn.1674-1544.2010.05.008

Analyzing Garfield's Academic Influence Based on the Perspective of Mapping Knowledge Domains

Wang Jinpeng, Wang Huijie

(Department of Information and Management, HuaZhong Normal University, Wuhan 430079)

Abstract: Based on the perspective of Mapping Knowledge Domains, this paper tries to explore the academic status and influence of Eugene Garfield more intuitively and more objectively. We use Bibexcel, Ucinet, SPSS and CiteSpace to analyze the data, which are downloaded from Web of Science.

Keywords: Eugene Garfield, mapping knowledge domains, scientometrics, Bibexcel, Ucinet, CiteSpace

1 引言

1955年, 美国著名情报学家尤金·加菲尔德(Eugene Garfield)博士在《科学》杂志上发表了《引文索引用于科学》的论文^[1], 系统地提出了用引文索引检索科技文献的新方法, 打破了分类法和主题法在检索方法中的垄断地位, 开创了从引文角度研究文献及科学发展动态的新领域。1963年, 他编制、出版了《科学引文索引》单卷本; 1964年, 美国 ISI 开始正式出版发行《科学

引文索引》(季刊); 1973年和1978年, 又相继出版了《社会科学引文索引》(SSCI)和《艺术和人文科学引文索引》(A&HCI)^[2]。由此, 引文索引这项富有创意的发明在图书情报界牢固地确定了自己的地位。引文索引不仅开创了一种新的检索途径, 增加了一种新的检索工具, 而且为文献计量学、信息计量学和科学计量学(“三计”学)等学科的发展开辟了道路。因其对世界科学研究的杰出贡献, 加菲尔德博士被科学界誉为“SCI之父”, 并于2006年荣获英国伦敦国际信息产业奖的终身成就奖^[3]。本文将从科学知识图谱的角

第一作者简介: 王金鹏(1985-), 男, 华中师范大学情报学专业硕士研究生, 主要研究方向: 科学计量、技术竞争情报。

收稿日期: 2010年5月31日。

度,借助引文和共被引分析理论以及 Bibexcel、Ucinet 和 CiteSpace 等软件,采用定性定量相结合的方法来探讨加菲尔德的学术地位和影响。

本文研究所用的数据来源于加菲尔德本人创建的 Web of Science 数据库。选择“被引参考文献检索”,以“Garfield E”为被引作者,检索了 1990–2009 年间引用加菲尔德论文的所有文献记录,共检出被引作者为“Garfield E”的文章共 342 篇,引用这些文章的文献共 462 篇。以这 462 篇文章构成原始数据集,进行计量分析并绘制知识图谱,将得出的一系列结论。

2 原始数据特征分析

2.1 关键词分析

利用 Bibexcel 软件^[4],统计原始数据集中的关键词频数,出现次数超过 10 次的关键词分布如表 1 所示。表中的关键词涵盖科学、引证、共被引、索引、引文分析、知识、绩效、生产力等。这些词语可以很好地描述原始数据的学科或领域属性。很明显,此数据集与“三计学”密切相关,领域的确定为进一步的分析奠定了基础。

选择出现频次高于等于 5 次的 36 个关键词为对象,利用 Netdraw 软件^[5],结合特征向量中介中心性,绘制了该数据集的高频关键词共现网络图谱(图 1)。

从图 1 中可以看出,Cocitation(共被引)处于网络的中心位置,其中介中心性最大,是网络中最重要节点,可见对共被引的研究是该领域的核心主题。其他重要的结点包括 Indicators(指

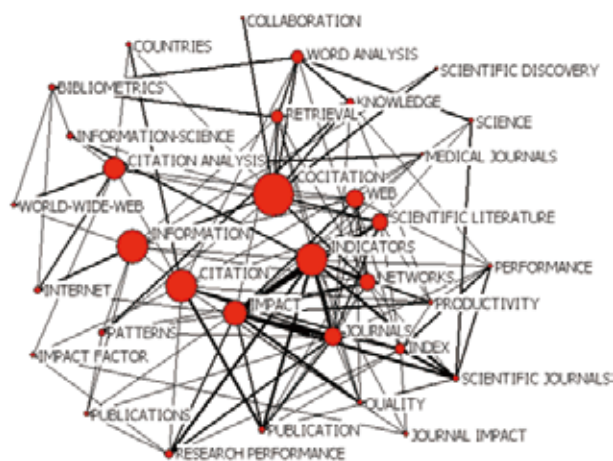


图 1 高频关键词共现网络图谱

标)、Citation(引证)、Information(信息)、Impact(影响)和 Citation Analysis(引证分析)等,它们是该领域的研究热点和重点。所有这些研究重点都建立在加菲尔德的贡献基础之上,即他发明的引文索引理论和 SCI 数据库。

2.2 被引情况分析

2.2.1 被引作者分析

利用 Bibexcel 的频次统计功能,获得被引次数超过 50 次的著者列表(表 2)。加菲尔德的被引次数(751)独占鳌头,远高于其他作者。此外,我们还可以发现一些熟悉的名字,斯莫尔(Small)、梅德(Moed)、格伦采尔(Glanzel)、怀特(White)、普赖斯(Price)、莫顿(Merton)等,他们都是“三计学”的大师级人物,其中不乏科学计量学“北斗七星”的成员^[6]。被引频次列表不仅说明他们在“三计”学方面的卓越造诣和开创性贡献,而且定量地刻画了他们的工作对后继研究

表 1 频次超过 10 的关键词列表

频次	英文	中文	频次	英文	中文
67	Science	科学	15	Information	信息
35	Impact	影响	14	Citation Analysis	引文分析
31	Indicators	指标	14	Networks	网络
24	Citation	引证	13	Publication	出版物
22	Journals	期刊	12	Knowledge	知识
17	Cocitation	共被引	10	Performance	绩效
16	Quality	质量	10	Productivity	生产力
15	Index	索引			

表2 被引次数超过50次的著者列表

频次	被引作者	频次	被引作者
751	Garfield E	68	Price DJD
127	Small H	68	Cronin B
91	Moed HF	67	Merton RK
75	Glanzel W	59	Leydesdorff L
71	White HD	53	Nederhof AJ
70	Milankovic M	52	Seglen PO

的影响以及其他研究者对其工作的认可和尊重。

2.2.2 作者共被引分析

利用 Bibexcel 和 Ucinet 软件处理^[7]，得到高被引作者共被引矩阵，通过 Netdraw 绘制高被引作者共被引网络图谱（图2）。图2中点的大小代表该点的度数（Degree），即该作者与其他作者之间联系的紧密程度；点的位置代表作者在领域中的地位 and 重要性，处于中心位置的点代表领域最重要的作者。由图2可知，代表加菲尔德的结点不但最大，而且处于中心位置，一方面说明他对其他作者的影响深远，另一方面说明他在该领域中处于相对核心的地位。

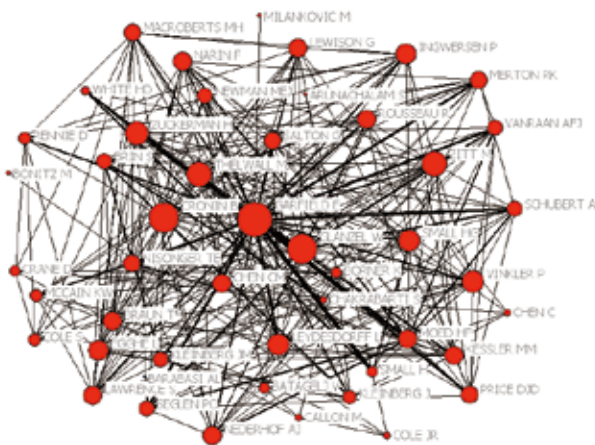


图2 高被引作者共被引网络图谱

2.2.3 被引文献分析

利用 Bibexcel，获得被引频次超过20次的文章列表（表3）。被引率排在前五位的文章除了第3篇作者是斯莫尔外，其他4篇均为加菲尔德发表的，当然这与我们的检索策略和数据构成有关。就加菲尔德个人而言，他于1972年、1955

年和1979年发表的3篇关于引文索引和引文分析的论文被引次数最多，对他的学术地位影响最大。

表3 被引次数超过20次的文献列表

频次	被引文献
52	Garfield E, 1972, V178, P471, SCIENCE
47	Garfield E, 1955, V122, P108, SCIENCE
38	Small H, 1973, V24, P265, J AM SOC INFORM SCI
35	Garfield E, 1979, CITATION INDEXING IT
34	Garfield E, 1979, CITATION INDEXING
25	Price DJD, 1965, V149, P510, SCIENCE
24	Kessler MM, 1963, V14, P10, AM DOC

选择被引频次超过10次的33篇文献，利用 Netdraw 软件，绘制成高被引文献共被引网络图谱（图3）。在图3中每个结点代表一篇引文，结点的大小代表中介中心性的大小，中介中心性大的那些节点，在不同的聚类网络中可以起到连接和过渡的桥梁作用，它们是网络中的中枢节点。可以发现，加菲尔德分别发表于1955年、1972年和1979年的3篇论文，是该网络中最重要的点，从网络结构角度再一次证明了加菲尔德对“三计”学的奠基性贡献。



图3 文献共被引网络图谱

2.2.4 多维尺度分析

多维尺度分析图是在信息科学中使用最为广泛的绘图技术，尤其是对文献可视化、作者共被引分析。利用 SPSS 的多维尺度、聚类分析等多元统计分析方法，绘制了该集合的作者共被引图

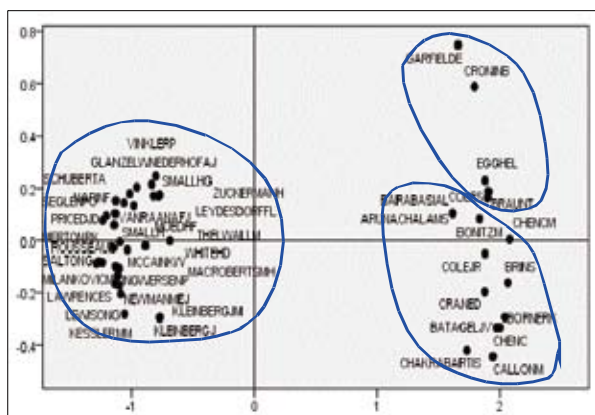


图4 被引作者群关系结构图

谱,如图4所示。根据作者群在多维尺度图上的距离远近以及作者聚集度和群体影响力可以将他们分为3类,即图中的3个线圈。

右上角的聚类以加菲尔德、克罗宁、埃格赫、布劳温为主,可以看出加菲尔德是源头,是其他学者研究和学术分支发展的基础;左边的较大的聚类包含了数据集中的大部分作者,他们的研究范围很广,既有理论研究也有应用研究;右下角的聚类包含了新近的研究热点——文献和知识可视化,其中华人学者陈超美(Chen CM)是“三计”学界冉冉升起的新星,在可视化方面有较高的造诣。

2.3 施引情况分析

2.3.1 发表年份分析

以原始数据集中的462篇施引文献绘制发表年份分布图,如图5中的实折线。以“Bibliometric OR Informetrics OR Scientometrics”为检索字段在WoS中按“主题”范围进行检索,共检索到发表在1990–2009年之间的文献467篇。对它们进行统计分析,得到主题词中至少含“三计”学之一的文献发表年份分布,如图5中的虚折线。从图中可以清楚地看到,两条折线的发展态势非常相似,可以推断,“三计”学的发展与施引文献之间有密切的联系。这间接说明加菲尔德对文献计量、信息计量以及科学计量学的巨大贡献和深远影响,正是他的研究工作促使了“三计”学的快速发展。

2.3.2 施引作者分析

统计原始数据集的施引作者频次,出现次数超过5次的如表4所示。他们是这个领域中的活

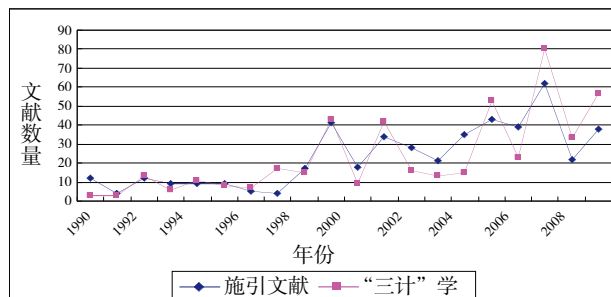


图5 施引文献与“三计”学文献年代分布图

跃分子,对领域发展的贡献较大。从表中作者各自的研究领域可以看出,引用加菲尔德论文的作者分布极其广泛,涉及“三计”学的所有分支。例如,网络计量学的重要代表人物赛尔沃(Thelwall)和鲁索(Rousseau),他们分别引用了加菲尔德论文6次和7次,可见网络计量学与“三计”学有着密切的联系,是新的前沿领域。再如,陈超美和麦肯(McCain)都是科学知识图谱和可视化领域的重要代表,他们分别引用了加菲尔德论文8次和6次。可见,正是有了加菲尔德的研究基础和卓越贡献,网络计量和科学知识图谱才有如今的发展。

表4 频次超过5次的施引作者分布表

频次	施引作者	频次	施引作者
12	Garfield E	6	Van Leeuwen TN
12	Moed HF	6	McCain KW
8	Chen CM	6	Bassecoulard E
8	Zitt M	6	Thelwall M
8	Giles CL	5	Scharnhorst A
7	Rousseau R	5	Small H

3 利用 CiteSpace 绘制知识图谱

如上所述,陈超美是知识可视化的重要代表人物,他发明的CiteSpace软件被广泛用于知识可视化和引文分析可视化^[8]。CiteSpace通过对科学文献题录数据的分析和处理,以图谱形式展现出科学发展的演化趋势以及前沿领域。在本小节中,作者利用此软件绘制了该数据集的知识图谱。

3.1 原始数据聚类分析

将数据集输入到CiteSpace中,选择时间分片

(Time Slicing) 为 2, 即将 1990–2009 共 20 年的数据分为 10 个分段, 每个分段中选择被引率最高的前 30 篇文献进行分析, 网络节点 (Node Types) 选择参考文献 (Cited Reference), 其他选项默认, 得到文章共被引聚类混合网络, 网络中共有节点 236 个, 连线 818 条, 如图 6 所示。

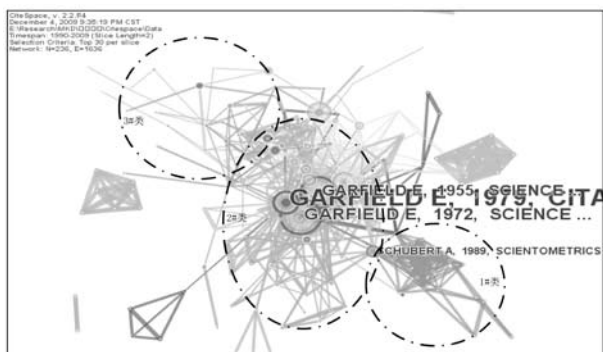


图 6 原始数据聚类图谱

从图 6 看出, 此聚类图的发展途径呈从右下角向左上角演化的趋势。按出现时间的先后顺序, 图谱中的聚类可分为 3 类: (1) 右下角的 1# 类聚类最先出现, 其研究内容聚焦于科学计量; (2) 中间的 2# 类聚类紧随其后, 引文索引是主要研究内容; (3) 左上角的 3# 类最后出现, 研究内容涵盖可视化、网络计量、社会网络等新领域。

图 6 中的深色字体代表重要性极高的关键文章, 是图谱中最重要的节点。节点的半径代表被引次数, 半径越大说明文献的地位越重要^[9]。由图 6 可知, 加菲尔德于 1955 年、1972 年和 1979 年分别发表在《Science》和《Citation Indexing》上的 3 篇文章是该数据集的关键节点, 是“三计学”的奠基性文章, 图谱直观地呈现了加菲尔德在该领域的核心地位。

3.2 时间线图谱分析

在 CiteSpace 中选择时间线 (Timeline) 聚类图, 可以看到各聚类之间随年代变化的演化情况, 如图 7 所示。

将图 7 的结果与时间分区对应分析, 可以粗略地将图 7 中的 13 个聚类分为三大类。

第一类为 #10–#13, 大约出现在 1994–1996 年之间, 是比较早的类, 且大都停止发展。

第二类为 #0–#4, 大约出现在 2000–2004



图 7 时间线图谱

年之间, 其中起点最早、持续时间最长的类是 #3scientific specialty 类。这个类的出现时间最早, 而且聚集着大量的高被引文献。从各聚类之间连线的关系可以看出, 其他聚类几乎都是从这个类衍生出来的, 它是该领域的基础类, 而加菲尔德是这个类中最重要的代表。

第三类为 #5–#9, 是 2004 年以后出现的新类, 也是当前的研究热点和前沿领域。其中, #9 类“community mining tool”比较重要, 从 20 世纪 60 年代至今始终保持着稳健的发展态势, 这个类所关注的领域在于信息挖掘、信息检索和可视化等。

4 结 论

笔者通过作者被引频次、文献被引频次以及施引情况分析, 揭示了加菲尔德及其研究对其所在领域的重要程度和卓著贡献; 通过社会网络分析, 特别是中介中心性的分析, 从网络结构方面分析了加菲尔德及其作品的重要地位和影响; 通过多维尺度分析和 CiteSpace 分析, 从历史演进的角度分析了加菲尔德及其研究在学术发展中的作用和角色。笔者认为, 上述方法可以为判定一位学者的学术地位和影响提供科学的解决思路和完善的操作范式。从这些分析可以得出如下结论。

(1) 加菲尔德所从事的研究主要集中在文献计量、信息计量和科学计量领域, 还涉及科学学和科学评价领域。

(2) 加菲尔德分别发表于 1955 年的、1972 年和 1979 年的 3 篇关于引文索引和引文分析的文章

奠定了他的学术地位和影响力,是他对该领域的最重要贡献,引文索引和引文分析是他最卓越的成就。

(3)加菲尔德的基础研究直接或间接地推动了领域其他分支的发展,网络计量学和科学知识图谱是很好的证明。而且,加菲尔德本人目前也在从事可视化方面的研究工作^[10],他于2003年开发的HistCite软件包是一款重要的引文历史可视化分析工具。

参考文献

- [1] Garfield E. Citation Indexes for Science: A New Dimension in Documentation through the Association of Ideas [J]. Science, 1955,122:108-111.
 - [2] Eugene Garfield. Citation Indexing [M]. Translated by Hou Hanqing, et al. Beijing:Beijing Library Press, 2004:23.(in Chinese)
〔尤金·加菲尔德.引文索引法的理论及应用[M].侯汉清,等,译.北京:北京图书馆出版社,2004:23.〕
 - [3] Web of Science [EB/OL]. [2009-12-12].http://www.sciencenet.cn/htmlnews/2009/9/223297.shtm.(in Chinese.)
 - 〔科学网[EB/OL]. [2009-12-12].http://www.sciencenet.cn/htmlnews/2009/9/223297.shtm.〕
 - [4] Bibexcel Website [EB/OL]. [2009-12-12].http://www8.umu.se/inforsk/Bibexcel.
 - [5] Netdraw Website[EB/OL]. [2009-12-12]. http://www.analytictech.com/downloadnd.htm.
 - [6] Liang Yongxia, Liu Zeyuan, et al. Front and Evolution Knowledge Mapping of Citation Analysis Domains[J]. Studies in Science of Science,2009(4):516-522. (in Chinese)
〔梁永霞,刘则渊,等.引文分析领域前沿与演化知识图谱[J].科学学研究,2009(4):516-522.〕
 - [7] Ucinet Website [EB/OL]. [2009-12-12]. http://www.analytictech.com/ucinet.
 - [8] CiteSpace Website [EB/OL]. [2009-12-12].http://cluster.cis.drexel.edu/~cchen/citespace.
 - [9] Chen C. CiteSpace II: Detecting and Visualizing Emerging Trends and Transient Patterns in Scientific Literature[J]. Journal of the American Society for Information Science and Technology,2006(3):359-377.
 - [10] Garfield Personal Website [EB/OL]. [2009-12-12]. http://www.garfield.library.upenn.edu.
-
- (上接第14页)
- nese)
- 〔胡媛,焦灵芝.免费开源 ROST 投稿审稿系统研究[EB/OL].[2010-09-18].http://www.sciencenet.cn/m/user_content.aspx?id=291380.〕
- [5] Hao Wei. Research and Implementation of Online Assignment System with Plagiarism Detection Function[D]. Shenyang: Northeastern University, 2006. (in Chinese)
〔郝伟.具有抄袭检测功能的在线作业系统的研究与实现[D].沈阳:东北大学,2006.〕
- [6] Hu Mingxiao. An Electronic Experiment Reports Management System with Plagiarism Detection[J]. Journal of Wenzhou University:Natural Sciences,2009,30(1):56-59. (in Chinese)
〔胡明晓.一种带有抄袭识别的电子版实验报告管理系统[J].温州大学学报:自然科学版,2009,30(1):56-59.〕
- [7] Wan Gang. Strengthen Integrity of Research and Creating a Good Academic Environment for Independent Innovation[EB/OL].[2010-09-18]. http://www.most.gov.cn/kjbgz/201004/t20100413_76734.htm. (in Chinese)
- 〔万钢.加强科研诚信建设 为自主创新营造良好的学术环境[EB/OL]. [2010-09-18]. http://www.most.gov.cn/kjbgz/201004/t20100413_76734.htm.〕
- [8] Guo Ling, Chen Yan. Immoral Phenomenon in Citing References and Some Measures to Fight[J]. ACTA Edictologica,2007,19(1): 37-41. (in Chinese)
〔郭玲,陈燕.参考文献著录中的学术道德缺失现象及其防范[J].编辑学报,2007,19(1): 37-41.〕
- [9] Metrics on the Effectiveness of Turnitin [EB/OL]. [2010-09-18].www.turnitin.com. (in Chinese)
〔Turnitin 系统功能[EB/OL].[2010-09-18].www.turnitin.com.〕
- [10] Nan Qiao. How Catch Plagiarism by Turnitin [EB/OL]. [2010-09-18].http://berlinfang.blog.163.com/blog/static/11667071620097463533453/. (in Chinese)
〔南桥.如何用 Turnitin 抓剽窃[EB/OL]. [2010-09-18]. http://berlinfang.blog.163.com/blog/static/11667071620097463533453/.〕