

基于多源数据融合的学科领域 专家研究轨迹和学科影响分析*

——以李政道教授为例

陈煦蔚 郭晶 袁继军

(上海交通大学图书馆, 上海 200240)

摘要: [目的/意义] 特定学科领域专家的学科影响力是情报学研究的重要课题, 其研究对象主要来自论文及相关数据库, 研究方法也主要基于图书情报学的方法。随着图书馆功能的扩展, 图档博等跨界融合, 数据的来源也变得多样化。本文拟探索特藏数据和情报学方法在图档博数字人文领域的跨界应用。[方法/过程] 本文以诺贝尔奖获得者李政道教授为案例, 基于李政道图书馆特藏资源, 融合 Web of Science 数据库资源和以诺贝尔奖等信息为代表的特征性信息资源, 综合分析李政道在中微子物理领域的研究轨迹和影响, 以揭示李政道与中微子物理学科发展的关系。[结果/结论] 利用不同的数据源, 互相关联、验证和补充, 可以使得挖掘角度更全面、揭示和展示层面更丰满。

关键词: 特藏资源 特征性资源 引文分析 多源数据 学科影响力 可视化

分类号: G203

DOI: 10.31193/SSAP.J.ISSN.2096-6695.2021.04.04

0 引言

图书馆学、档案学、博物馆学、科学史研究等对于学科发展和人物的影响力都有着各自的专业研究方法。图书情报学中, 往往通过论文及其引证关系等数据来进行学科领域发展或人物影响力分析, 如冉从敬等人将 CiteSpace、VOSviewer^[1] 相结合研究莎士比亚研究的发展轨迹; 自 MARX 等人提出 RPYS 方法后^[2], 李信等人探索了该方法在学科领域历史根源研究上的可行性^[3];

* 本文系上海交通大学“李政道与中国科教事业”2017年度专项课题“基于定量分析的宇称不守恒理论在粒子物理学领域贡献和影响力研究”(项目编号: TDLee2017001)的研究成果之一。

[作者简介] 陈煦蔚, 女, 特藏服务部副主任, 博士, 研究方向为特藏数据揭示利用、现代物理学史, Email: xwchen@sjtu.edu.cn; 郭晶 (ORCID: 0000-0001-8185-7003), 女, 副馆长, 博士, 研究方向为信息分析与资源组织、情报研究与利用, Email: jguo@sjtu.edu.cn; 袁继军, 女, 党委副书记, 硕士, 研究方向为档案学, Email: jiyuan@sjtu.edu.cn。

刘俊婉等人结合 Word2Vec 算法、TF-IDF 算法和 PageRank 算法对引证行为进行了分析研究, 改进了多学科、大量学者等场景下的学术影响力评价^[4]; 迟培娟等人比较研究了学术迹、F1000 评分、影响因子、被引次数、h 指数、ESI 高水平论文数量等评价指标, 用于评价个人学术影响力的效果和相关性^[5]; 池雪花等人以论文数据为基础, 在传统影响力回归方法上提出分段回归预测模型, 对学者学术影响力进行预测^[6]; 杨瑞仙等人基于 Web of Science (以下简称“WOS”) 数据库, 利用多元统计软件和社会网络分析软件分层次研究科学合作和论文影响力之间的相关关系^[7]; 匡登辉基于 WOS 数据库分析比较不同学科文献的使用差异, 从而利用用户学术行为数据测评学术论文影响力^[8]; 袁国华等人采用同行评议的方法对传统的文献计量评价方法进行了补充^[9]。博物馆学、档案学及科学史专业往往通过藏品、档案、口述历史等信息来考证历史事实及其影响脉络。随着高校图书馆越来越多地吸纳、保管各类以领域专家为主题的特藏资源, 其功能也逐步扩展。如何从跨领域的视角, 更好地对数据进行揭示、开放和利用, 成为跨界探讨的热点之一^[10]。

对于特定学者而言, 能够反映其研究轨迹与学术影响所涉及的数据来源和类型往往是多样的, 除了论文, 还包括信函、手稿以及相关档案资料所构成的特色馆藏体系。例如, 诺贝尔奖获得者李政道教授将其毕生积累和收藏的各类科学文献、研究手稿、通讯信件、科艺作品以及各类实物资源共 8 万余件捐赠给上海交通大学图书馆, 成为其独具特色的馆藏资源。图书馆对所有纸质特藏资源进行了数字化加工编目, 形成了 6 万余条元数据, 并通过网站进行发布。这些类型多样的特藏数据成为支撑研究者进一步分析和把握相关学者成就及学科领域进展的重要资料, 需要研究制定相关的分析方法和流程。

此外, 标志性的成果、荣誉或者具备权威性的学术团体、活动等信息往往对分析学者成就和影响力有着重要作用。由于这类信息往往具备非常明显的学科成就或者权威特征性, 且与传统的论文、著作等资源类型有着较大差异, 为便于区分, 本文将与其相关的资源定义为“特征性资源”, 并作为学者分析所采用的数据源之一。

本文从数字人文视角, 尝试以李政道图书馆特藏资源为基础, 综合利用 WOS 数据库资源和以诺贝尔奖网站资源为代表的特征性资源等^[13-18], 主要通过 EXCEL 的函数库构造出清理、筛选、统计、聚类、分析相关数据信息的计算方法, 通过其自身的图表功能将数据信息可视化, 辅助分析特定学者(本文以李政道为例)在学科领域的研究轨迹及影响力。最后, 在对多源数据进行统计和融合分析的基础上, 利用 Gephi 进行可视化呈现。

1 研究方法

因科学研究具有交叉复杂性, 本文作为探索性研究, 选取了李政道与中微子这一研究领域, 以“中微子”为检索词, 选取直接针对中微子的资源作为研究的样本。通过对这一典型案例的分析, 总结以特藏资源为基础分析特定学者研究轨迹与影响力的通用流程和方法。

数据选取和研究的主要步骤如下:

(1) 依靠自建的特藏数据库, 获取相关的特藏数据样本, 对李政道在中微子物理领域的活动轨迹做初步研究。

(2) 基于 WOS 数据库, 以论文数据为对象, 对中微子物理研究领域发展脉络做简单描绘, 以此作为整体背景。在该背景下, 有限选取并清理高被引论文作者的数据作为样本, 与特藏数据库中信件类资源的元数据中的人物数据进行关联分析, 初步揭示李政道与中微子研究领域内高被引论文作者之间的关联度。

(3) 选取李政道在中微子领域的一篇重要论文作为典型特藏资源样本, 通过引文分析揭示其在中微子物理大背景下的影响力程度。

(4) 结合诺贝尔物理学奖相关网络信息, 与特藏数据和论文数据进行融合, 提炼出可供展示发掘的特色与亮点。

(5) 综合以上数据, 形成整体结论。

2 特藏数据采集及初步分析

李政道图书馆的纸质资源分为 7 类, 包括: 论文、手稿、信函、讲义、图书、图像和其他。进行数字化加工后, 组织物理学专业的博士研究生对资源进行初步的元数据编写, 最后由专业人员编目并录入系统, 每一件形成一个元数据, 共计 6 万余条。因资料捐赠于 2011 年, 因此特藏资源所属时间截止于 2010 年。

特藏数据在制定元数据规范时, 科学主题词是中英文对照关联的, 因此可以用“中微子”作为中英文所有资源的检索词, 得到直接相关的特藏数据共 484 条。这些数据包含: 论文 209 条、手稿 162 条、信函 12 条、讲义 72 条、图书 2 条、图像 0 条、其他 27 条。

在这些数据中, 能追溯到其确切形成年份的有 187 条, 其中 1947 年的数据 1 条, 其余 186 条数据导出后, 用 EXCEL 函数进行统计, 其形成年份分布情况如图 1 所示。从图 1 可以看到, 共有三个高峰值点, 分别是 1957 年、1963 年和 2007 年。

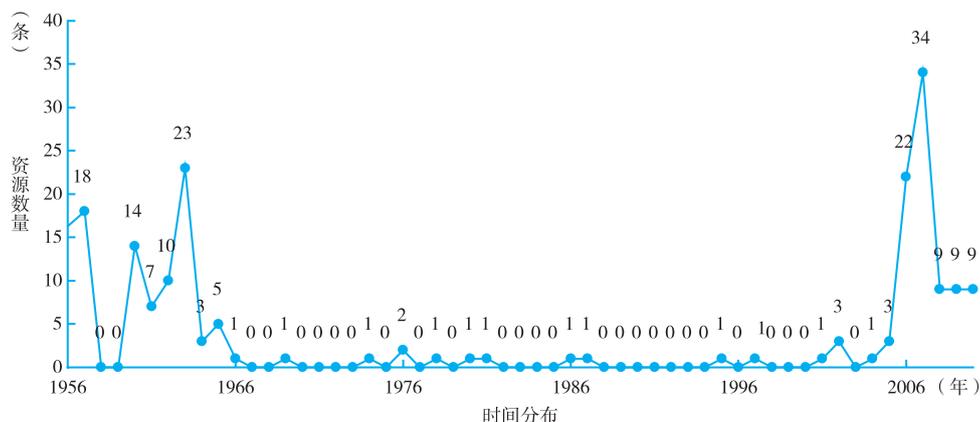


图 1 部分中微子特藏资源年份分布图

利用 EXCEL 函数对不同类型资源在不同年份的数量进行统计, 并进一步利用其散点图下的三维气泡图功能, 形成气泡分布图, 见图 2, 其中蓝色气泡为论文类资源, 棕色气泡为手稿类资

源,灰色气泡为讲义类资源,黄色气泡为其他类资源。通过气泡图,可以直观看到,早期的资源类型以论文类最多,到了21世纪后,资源类型则以手稿类最多,讲义类资源数量也明显增加。

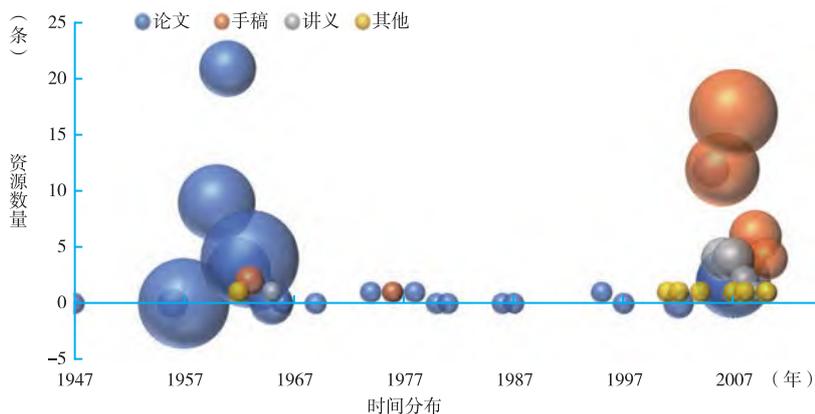


图2 不同类型资源在不同年份的数量分布气泡图

2.1 典型特藏资源样本分析

文章选取1947年的一条数据作为典型特藏资源样本进行内容分析,这是李政道在1947年所阅读并一直收藏的、ENRICO FERMI于30年代所著的一篇关于 β 衰变的论文。1947年,李政道正师从FERMI攻读博士研究生学位。追溯中微子物理发展历史,PAULI根据实验中无法解释的结论,预言了“一种粒子”的存在,此后,FERMI提出了 β 衰变的定量理论,从理论上再次预言并正式将其命名为“中微子”。可见,在中微子物理领域诞生初期,李政道即已经以博士研究生的身份步入其中。

2.2 第一个高峰值点分析

1957年,特藏数据中与“中微子”相关的数据量形成第一个高峰值点,其所属类型均为论文。1956年,李政道与杨振宁提出“弱相互作用下宇称不守恒”理论,并因此而获得1957年的诺贝尔物理学奖。这一突破性的发现表明,任何费米子只有其左手分量参与弱相互作用。基于此,李政道和杨振宁于1957年再次合作发表论文《宇称不守恒与中微子的两分量理论》^[11],首次提出关于中微子的两分量理论,这一理论成为此后逐步成型的粒子物理标准模型的一块奠基石。这些特藏数据正是反复修改和审校的论文。

2.3 第二个高峰值点分析

中微子虽被预言存在,却在一段时期内一直未被真正探测到。1960~1965年的特藏数据中,与中微子直接相关的数据有83条,包含论文数据75条、讲义数据5条、手稿数据2条和其他类数据1条。从1960年到1967年,李政道发表了10篇与中微子相关的论文,对高能中微子实验进行理论探讨,并对中微子的特性进行了研究,论文类的数据是这些发表论文的反复修订稿。其中,1960年发表的论文《可能的高能中微子实验的理论讨论》^[12],再次成为高能中微子实验的指导性的论文。讲义和手稿数据则显示,李政道在学术会议中公开讨论对中微子特性及其可能性

实验的理论研究。其他类数据则是美国布鲁克海文实验室 (Brookhaven National Laboratory, 简称 BNL) 在 AGS (alternating gradient synchrotron) 同步加速器上的实验记录, 可见此时 BNL 正在进行中微子相关的实验, 李政道也与其有着紧密关联。

2.4 第三个峰值点分析

2001 ~ 2010年, 与中微子相关的特藏数据共有 91 条, 其中手稿 43 条、论文 30 条、讲义 12 条、其他类数据 6 条。数据截止于 2010 年的一个可能的原因是, 李政道从 2011 年开始捐赠资源, 并集中整理了两批数据捐至图书馆, 其后陆续产生的数据尚未进入数字化特藏库中。

这个时期的李政道已经进入耄耋高龄, 他仍正式发表了 3 篇与中微子相关的论文。手稿数据主要与其研究相关, 显示李政道正在对中微子质量矩阵、中微子转换矩阵、中微子混合矩阵进行研究, 并与时间对称性破缺方面的研究联合在一起思考。讲义数据显示, 李政道在美国物理学会、欧洲核子研究中心、中国高等科学技术中心以及多个高校作了关于中微子的报告。值得关注的是, 据部分手稿和其他数据显示, 李政道高度关注大亚湾中微子实验, 在 2004 年、2010 年曾两次召开大亚湾相关的中微子实验研讨会, 并自 2001 年起多次以中微子为题作讲座报告。

可见, 特藏资源的年代分布、类型气泡分布图, 可以一定程度反映出李政道教授在中微子领域的研究轨迹及相关社会活动。

3 论文数据分析及与特藏数据的关联分析

3.1 中微子相关论文数据

为简化研究, 在此不考虑学科交叉的复杂性, 在 WOS 数据库中, 以 “neutrino” 为检索词, 在 Topic 和 Title 下分别检索, 得到论文数据共 66434 条 (该数据检索时间为 2021 年 11 月)。根据发表时间分析检索结果, 并将其导出至 EXCEL 表格, 以其年份分布情况生成折线图, 如图 3 所示。添加线性趋势线后, 可以看到, 论文数量整体呈上升趋势。早期, 在 1957 年和 1958 年有微小起伏; 此后, 在 1973 年前后有一高峰; 1998 年后, 在趋势线上方呈现振荡型快速发展的趋势。

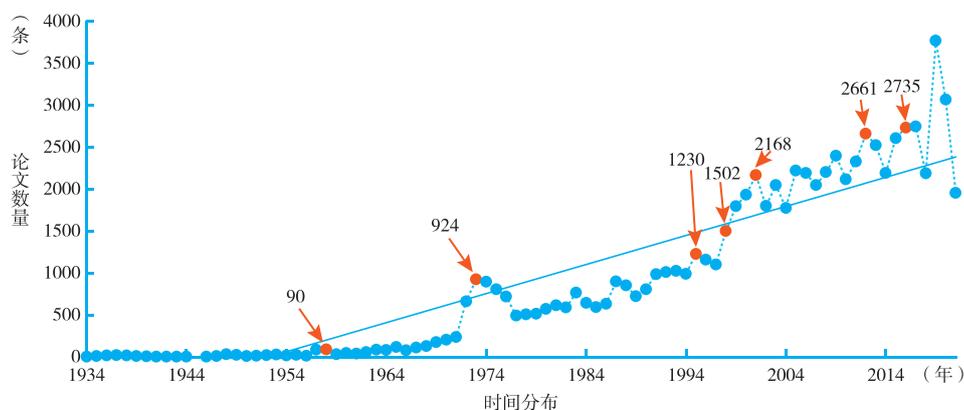


图 3 中微子相关论文发表年份分布图

个重复值, 保留了 628 个数据, 获取这 372 个重复值的全名信息, 对其进行人工验证, 去重误差率小于 2%, 主要是因为研究者来自不同国籍或种族, 取名的习惯多样。对于第二个误差来源, 由于未批量获取论文作者的全名信息, 无法全名校对, 有待进一步探讨。

3.3 重要论文的引文初步分析

以 1957 年发表的、首次提出关于中微子两分量理论的论文《宇称不守恒与中微子的两分量理论》为样本, 进行引文分析试验。

在 WOS 数据库核心合集中, 检索到该论文的引文 535 篇, 其发表年代分布如图 5 所示。1957 年和 1958 年是引用高峰期, 可见该论文在短时期内得到了极大的关注度, 并推动了研究, 也由此揭示了图 2 中在 1958 年附近的论文数量有小幅波峰的原因。

对引文的部分代表性作者关联研究, 将结合网络信息数据和特藏数据共同分析。

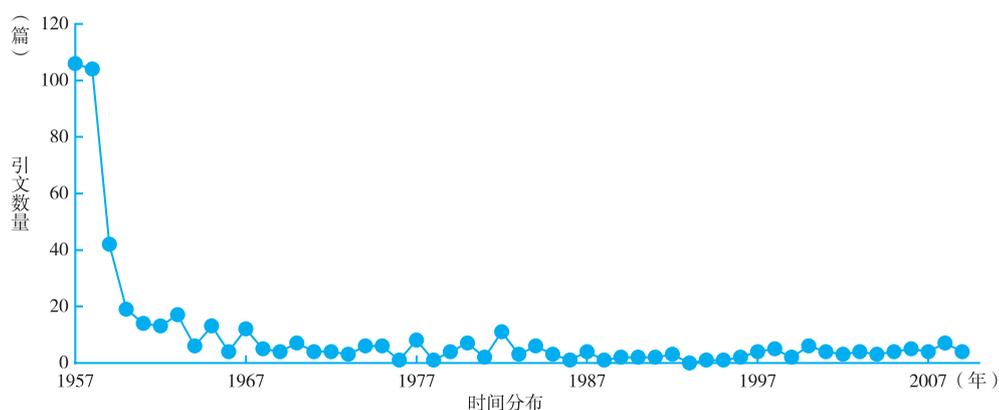


图 5 引文数量年份分布图

4 特征性资源融合分析

在信息时代, 从网络上往往能获取到学科领域专家的较为官方的信息, 比如其所在研究单位或团队在网上发布的信息、权威学术机构发布的信息等。这些可以作为对该学科领域专家进行研究的特征性资源。

诺贝尔物理学奖是物理学界最高荣誉之一, 诺贝尔奖官方网站详细记载了历年的诺贝尔奖得主及其得奖原因。选取诺贝尔奖官方网站的信息^[14]作为特征性资源的来源, 将诺贝尔物理学奖获奖者的信息与特藏数据、论文数据进行关联, 分析李政道与其他诺贝尔物理学奖得主及其成就的关系。

通过诺贝尔物理学奖的这一特征性资源与上述论文的被引文献作者数据进行关联比较, 发现有 7 位诺贝尔物理学奖获得者引用过该论文, 详见表 2。其中 1988 年和 1995 年授予的诺贝尔物理学奖是为了表彰其在中微子物理领域的成就。

表2 对样本论文施引的作者中诺贝尔奖得主列表

序号	获奖人	获奖年份	引用该论文次数
1	JULIAN SCHWINGER	1965	1
2	RICHARD PHILLIPS FEYNMAN	1965	1
3	LEON LEDERMAN	1988	2
4	MELVIN SCHWARTZ	1988	1
5	JACK STEINBERGER	1988	1
6	MARTIN L. PERL	1995	1
7	FREDERICK REINES	1995	1

在特藏数据库中，也藏有李政道与 STEINBERGER 等人合作的记录，以及与众多诺贝尔奖得主的关联信息。如，李政道曾从 STEINBERGER 的实验中意识到，弱相互作用下宇称很可能是不守恒的，由此引发他对宇称守恒的质疑，并最终获得 1957 年诺贝尔物理学奖；而此后，李政道引发了 STEINBERGER 的学生 SCHWARTZ 对两种中微子实验的构想，促使 SCHWARTZ 构想出产生中微子束的方法，最后实验成功，SCHWARTZ 和 STEINBERGER 也因此获得了 1988 年诺贝尔物理学奖。

特藏资源、论文和特征性资源，可通过人物这一数据信息关联起来，如图 6 所示。

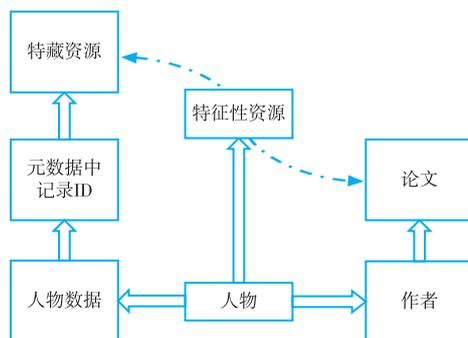


图6 特藏资源、论文、特征性资源关联关系示意图

通过上述分析，可以提炼出基于特色馆藏资源进行特定学者研究轨迹与影响力分析的一般方法和流程，具体如下：

(1) 首先从特藏资源和 WOS 数据库资源中，分别获取与选定学科领域直接相关的数据样本，并进行初步分析。根据特藏资源数据，得到该学者与这一学科领域相关的研究、社会活动等信息，初步勾勒出学者的活动足迹和数据分布图。根据 WOS 数据库所得论文数据，作出特定学科领域的研究热度趋势图以及重要论文的被引论文时间分布图，将特藏资源的数据分布图与两图的趋势和峰值进行比较，结合特藏资源的具体内容，分析学者的研究成果对学科发展的贡献。比如从文章的案例中就可以看到，李政道较早地进入了中微子物理学研究领域，并催生了第一次研究

小高峰, 21 世纪后, 又投入到中国大亚湾核电站相关的研讨及社会活动中。

(2) 再以要研究的学者信息作为基础数据, 比对特藏资源中某一由代表性资源类型所涉及的人物数据与高被引论文作者信息, 将具有相同人物信息的论文和特色资源关联, 分析该学者与该领域高被引论文作者的关联度, 可以看到学者与高被引论文作者这样一个学术群体的学术交流与合作状况。由于李政道特色资源中有 4 万余件信函, 在文章案例中选取了信函中人物数据与中微子物理领域高被引论文作者进行比对分析。

(3) 最后, 结合该学者的标志性的成果、荣誉或者该学科权威性学术团体、活动等信息, 获取与之相关的特征性资源, 进行影响力的关联分析, 显示学者在某一学科领域的贡献和成就。本文选取了诺贝尔奖获奖者的信息, 以论文被诺贝尔奖得主引用的情况来揭示论文的重要性, 并通过特藏资源来展现李政道与其他诺贝尔奖得主的关系。从诺贝尔物理学奖获得者这样一个高端学术群体的角度, 来揭示李政道在中微子物理领域的影响力。

至此, 基于多种数据来源, 对特定学者自身活动、论文数据、与特定学术群体的关联度进行了统计分析, 可以多角度揭示学者在某一学科领域的研究轨迹和关联影响, 充分挖掘特色馆藏资源在数据呈现和内容分析方面的多重价值。

5 可视化试验

Gephi 的可视化算法有两类, 一类为布局算法, 一类为统计算法。本研究在对数据进行统计分析的基础上, 利用 Gephi 的布局算法^[19], 对分析结果进行可视化布局。

结合本文第 4.2 节的内容可知, 被引频次数量排名前 1000 位论文的 2200 名作者中, 有 317 人与李政道有过信件往来。本试验的目标为: 展现特藏库中这 317 人与李政道的来往信函、这 317 人的信息、对应的高被引频次论文, 以及揭示以上三个来源数据之间的关联。

以上述三类数据本身作为 Gephi 成图所需的“节点”, 包括 976 件信函、317 人和 1000 篇论文。对于 976 件信函, 以其在特藏数据中的元数据的 ID 作为“节点”的 ID, 以其信函往来对象的人名为“节点”的属性值之一; 317 人以其姓名为“节点”的 ID 和属性值之一; 对 1000 篇论文进行编号, 以该编号作为其“节点”的 ID, 作者姓名设定为其属性值之一。

三类数据的“节点”间通过“有向边”连接: 特藏数据节点指向人物数据节点, 人物数据节点再指向关联的论文数据节点, 其指向关系如图 7 所示。

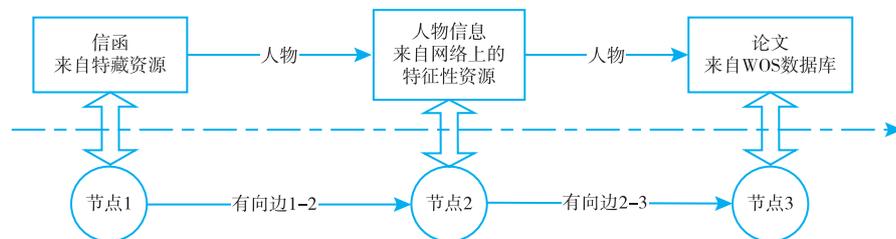


图 7 数据与节点关系指向示意图

为获取“边”数据，将三类数据信息导入 Matlab 软件中，编写程序对包含相同人物姓名字符串的数据节点进行关联，同时剔除与其他两类数据不具有关联性的论文数据节点，关联结果形成包含“有向边”信息的矩阵。按照 Gephi 对边数据文件的格式要求，将矩阵转化为列向量并导出为 CSV 文件，得到节点之间“有向边”的数据。

将节点和边信息导入 Gephi 中，默认生成的布局图如图 8 所示。其中，来自特藏数据的“信函”的节点设置为红色、大小 8；来自网络其他数据源的“人物”的节点设置为蓝色、大小 3；来自 WOS 数据库的“论文”的节点设置为黑色、大小 5。图 8 中，黑色的有向边由信函指向其涉及的人物；蓝色的有向边则由该人物进一步指向其所发表的论文。

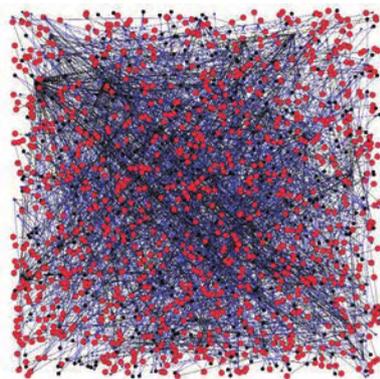


图 8 三类数据及其关系布局图

在 Gephi 中进一步对其运行重力布局算法 (force atlas)，计算每个节点在图中的合理位置，使得图更加紧凑，更加具有可读性。运行一段时间后的图形如图 9 所示。



图 9 重力布局算法下的三类数据及其关系布局图

运用 Gephi 对其关系进行初步布局，添加数据具体信息后，将具有可读性，可作为实际应用中进一步可视化的基础。

6 结论

本研究以李政道教授为例, 基于李政道图书馆特藏数据库, 结合 WOS 数据库和特征性资源, 对李政道在中微子物理领域的影响足迹进行综合分析, 利用不同数据源的特色, 互相关联、验证和补充, 从而使得挖掘角度更全面、可展示层面更丰满, 并通过 Gephi 进行了初步试验。在此基础上, 可以继续对其关联展示进行探究。

多源数据涉及到的数据量相对较大, 且往往具有不同的描述形式, 需要制定格式调整方案。本文在对格式调整的过程中, 使用了 EXCEL 自带的函数, 损失了一定的关联比对精确度, 这一损失可以通过更完善的计算机语言来避免。

特藏数据描述的精确度和完整度对研究结果有着很大影响, 而信函等科学手稿的专业性, 以及对数字化手稿进行文字识别的难度, 制约了对特藏数据的充分描述及其后续的揭示利用。对特藏数据进行更深度的揭示, 以及不同元数据标准之间更精准的映射和转化, 将有益于特藏资源与其他资源信息的互相关联与聚合, 从而得到更加客观、全面的分析结论, 并进行富含更多信息量的可视化展示。

致 谢

本文所用特藏数据来自上海交通大学李政道图书馆。感谢上海交通大学图书馆黄楠协助提供论文数据、杨昭在 EXCEL 软件上的支持, 以及施晓华的讨论建议。

【参考文献】

- [1] 冉从敬, 吕雅琦, 李新来, 黄海瑛. 莎士比亚研究领域引文可视化分析 [J]. 图书馆论坛, 2017(9):8-15.
- [2] MARX W, BORNMAN L, BARTH A, LEYDESDORFF L. Detecting the historical roots of research fields by reference publication year spectroscopy (RPYS) [J]. Journal of the Association for Information Science and Technology, 2014,65(4):751-764.
- [3] 李信, 陆伟, 李旭晖. 一种新兴的学科领域历史根源探究方法:RPYS [J]. 图书情报工作, 2016(20):70-76.
- [4] 刘俊婉, 杨波, 王菲菲. 基于引证行为与学术相似度的学者影响力领域排名方法研究 [J]. 数据分析与知识发现, 2018,2(4):59-70.
- [5] 迟培娟, 宋秀芳. 个人学术影响力评价方法比较分析——学术迹与 F1000 评分、影响因子、被引次数等指标的比较分析 [J]. 情报科学, 2018,36(12):9-12.
- [6] 池雪花, 章成志. 基于分段回归的学者学术影响力预测研究 [J]. 文献与数据学报, 2019,1(4):36-45.
- [7] 杨瑞仙, 李贤. 科学合作与论文影响力之间的相关性研究 [J]. 现代情报, 2019(4):125-133.
- [8] 匡登辉. 科学评价视角下文献使用比较研究——基于 Usage Count 的实证分析 [J]. 现代情报, 2019(4):115-124.
- [9] 袁国华, 寇晶晶, 张建勇, 韩正琪. 基于开放同行评议的学者影响力评价研究——以 F1000 为例 [J]. 图书情报工作, 2018,62(13):37-44.
- [10] 曾蕾, 王晓光, 范炜. 图档博领域的智慧数据及其在数字人文研究中的角色 [J]. 中国图书馆学报, 2018(44):17-34.
- [11] LEE T D, YANG C N. Parity nonconservation and a two-component theory of the neutrino [J]. Physical Review.

1957(105): 1671.

[12] LEE T D, YANG C N. Theoretical discussions on possible high energy neutrino experiments [J], Physical Review Letters. 1960(4):307.

[13] 李政道数字资源中心 [DB/OL]. [2021-02-10]. <http://tdlee.sjtu.edu.cn>.

[14] The nobel prize [DB/OL]. [2021-02-10]. <http://www.nobelprize.org>.

[15] 中国高等科学技术中心. 李政道科学论文选:上册 [M]. 上海:上海科学技术出版社, 2007.

[16] 中国高等科学技术中心. 李政道科学论文选:下册 [M]. 上海:上海科学技术出版社, 2007.

[17] 中国高等科学技术中心. 李政道教授八十华诞文集 [M]. 上海:上海科学技术出版社, 2009.

[18] 中国高等科学技术中心. 李政道文选:科学和人文 [M]. 上海:上海科学技术出版社, 2008.

[19] Gephi [DB/OL]. [2021-05-01]. <http://gephi.org>.

Study on the Academic Trace and Influence of the Subject Expert Based on Multi-source Data Fusion: A Case Study of Prof. Tsung-Dao Lee

CHEN Xuwei GUO Jing YUAN Jijun

(Shanghai Jiao Tong University Library, Shanghai 200240, China)

Abstract: [**Purpose/significance**] The academic influence of subject expert has always been an important research topic in information science. The research objects are mainly from papers and related databases, and the research methods are mainly based on the library and information studies. With the development of library's functions and the cross-border integration of libraries, archives and museums, the data sources have become diversified. This paper discusses a transboundary application of special collections and information scientific method based on multi-source data fusion. [**Method/process**] This paper, taking Nobel laureate Tsung-Dao Lee as an example, tries to analyze his academic trace and influence in the field of neutrino physics based on special collections, WOS database resources and characteristic information, in order to reveal the relationship between his academic influence and the progress of neutrino physics. [**Result/conclusion**] Using multi-source data can verify and supplement each other, and make the resources more fully explored and revealed.

Keywords: Special collections; Characteristic resources; Citation analysis; Multi-source data; Academic influence; Visualization

(本文责编: 孙龙慧)