



开放科学
(资源服务)
标识码
(OSID)

投入产出视角下世界城市学科学术影响力的评价研究

——以物理学科为例

郭金忠 王申宏 刘晓玲 冯贺禾

新疆财经大学信息管理学院 乌鲁木齐 830012

摘要: [目的/意义] 从知识流动的角度, 探究世界城市学科的学术影响力, 为后续城市学科研究及投入产出模型的应用提供参考。[方法/过程] 以物理学科为例, 基于美国物理学会 (APS) 期刊论文数据集及其自有的物理学与天文学分类方案 (PACS), 构建城市物理学子领域分类体系及知识流网络, 考察城市科研产出分布情况, 并借助投入产出模型衡量城市物理学子领域的影响力, 最后将影响力排序列表与 PageRank、HITS 算法所得排序列表比较, 讨论不同算法特性对排序列表的影响。[局限] 受限于数据的可得性, 缺乏对宏观学科以及最新数据的分析。[结果/结论] 研究发现: ①城市在物理领域的科研产出存在较大异质性, 整体服从幂律分布, 且高科研产出城市主要位于欧洲、北美洲和东亚地区, 其中东亚地区主要为日本和中国; ②城市内部各物理学子领域间影响力差异显著, USA-Cambridge 的基本粒子物理和场 (10) 子领域学术影响力排名第一; ③投入产出模型综合考虑了引用的数量和质量, 且算法复杂度较低, 对引用结构的敏感性要低于 PageRank 和 HITS 算法, 为评价城市学科学术影响力提供了新的可靠分析范式。

关键词: 科学计量学; 影响力评价; 知识流网络; 投入产出模型; 地理分布; 可视化

中图分类号: G350

Research on the Evaluation of Academic Influence of the World Urban Discipline from the Perspective of Input and Output

—— Take Physics as an Example

GUO Jinzhong WANG Shenhong LIU Xiaoling FENG Hehe

School of Information Management, Xinjiang University of Finance and Economics, Urumqi 830012, China

基金项目 国家自然科学基金项目“科研人员校友网络对高校科研成果转化的影响及机理研究”(NSFC No.72264036); 中国科学院“西部青年学者”项目“国家科学领域的重要性及相互支撑和影响关系研究”(2020-XBQNXX-020); 新疆维吾尔自治区高校科研计划项目“科研评估中计量指标与方法的研究”(XJEDU2020Y033)。

作者简介 郭金忠(1985-), 博士, 副教授, 研究方向为科学计量评价、计算机仿真模拟, E-mail: guojz@xjufe.edu.cn; 王申宏(1995-), 硕士研究生, 研究方向为科学学与科学计量、数据管理与分析预测; 刘晓玲(1985-), 博士, 研究方向为科学情报与科技政策、高性能科学计算; 冯贺禾(1996-), 硕士研究生, 研究方向为科学学与科学计量。

引用格式 郭金忠, 王申宏, 刘晓玲, 等. 投入产出视角下世界城市学科学术影响力的评价研究: 以物理学科为例 [J]. 情报工程, 2022, 8(6): 88-102.

Abstract: [Objective/Significance] From the perspective of knowledge flow, this study explores the influence of the world urban discipline and provides a reference for the subsequent research in the urban discipline and the application of the input-output model. [Methods/Processes] Taking physics as an example, based on the American Physical Society (APS) journal paper dataset and its own Physics and Astronomy Classification Scheme (PACS), this article constructs a classification system and knowledge flow network for urban physics subfields, examines the distribution of urban research output, and measures the influence of urban physics subfields with the help of input-output models. Finally, the impact ranking list is compared with the ranking list obtained from PageRank and HITS algorithms, and the effects of different algorithm characteristics on the ranking list are discussed. [Limitations] Limited by the availability of data, there is a lack of analysis of macro disciplines and the latest data. [Results/Conclusions] The results show that: (1) There is a large heterogeneity in scientific research output of cities in the field of physics, and the overall obeys a power-law distribution, and the cities with high scientific research output are mainly located in Europe, North America and East Asia, of which East Asia is mainly Japan and China; (2) The influence of various physics subfields within cities varies significantly, and the academic influence of elementary particle physics and field (10) in Cambridge, USA ranks first; (3) The input-output model integrates the quantity and quality of citations, and the algorithm is less complex and less sensitive to the citation structure than the PageRank and HITS algorithms, providing a new and reliable analytical paradigm for measuring the influence of urban discipline.

Keywords: Scientometrics; impact evaluation; knowledge flow network; input-output model; geographic distribution; visualization

引言

随着知识经济时代的到来和全球化水平的不断提高,城市作为创新活动的聚集地和实现载体,其网络地位比规模地位更加重要^[1],也因此涌现出一批空间计量学、地理学、科学学等领域的学者在城市层面构建引文网络、合作网络和创新网络,以此探寻城市的知识地位、合作地位和创新地位。然而,目前针对城市学科的微观研究却略显逊色。一方面,城市地址数据的庞大和不规范性使得数据收集和处理工作复杂而繁琐,且语言的隔阂以及“城市”这一地理单元规模的不同也给相关研究带来阻碍;另一方面,尽管学科分类可以划分不同的学科领域,但不同学科分类体系的差异同样为研究增加了不确定性。但是,结合知识流的特征和内涵评价城市学科的学术影响力具有重要意义,不仅有助于城市管理者 and 政策制定者有效识别

城市科学核心优势,从而更好的制定管理、投资战略,而且能够用城市学科网络揭示国家知识流动格局,为城市协调均衡发展提供新思路,为探索国家协调创新路径助力。基于此,本文借鉴经济学中的投入产出模型探索世界城市学科的学术影响力,旨在丰富对城市学科的有限探索。

1 相关文献综述

1.1 城市网络相关研究

科技创新是抢占科学前沿高地的重要动力,党的十九大报告强调要加快建设创新型国家,完善科技强国布局。因此黄卫对加强我国面向世界科技强国的基础研究布局发表若干见解^[2]。Guo等^[3]定量的衡量各个国家学科的学术影响力,揭示了国家学科间的知识流动格局。城市

是国家参与全球创新竞争的重要地理单元,在国家创新体系中的地位不断提升。城市网络已经成为解释城市空间组织结构的一种新范式,而知识流动作为典型的城际互动关系之一^[4],吸引着不同学科领域的关注,故越来越多的研究经由科研论文和专利的引用、合作、转让关系探索城市间知识溢出和互动。Zhang等^[5]基于APS数据库建立城市引用关系网络,定义知识扩散代理算法和科学生产排名算法,并确定物理知识生产和消费的关键城市。Csomós^[6]构建世界城市国际合作网络,对城市国际合作的效率做出回答。马海涛^[7]提出四种城市间创新关系建构和网络模拟方法,反映城市创新关系的不同方面。曹湛等^[8]以WoS合著论文数据为基础,从空间结构、拓扑特征及驱动因素三方面分析全球526个主要城市知识合作网络的演化。可以看出,当前研究多停留在以城市间知识流动为研究对象,并未过多细化研究的粒度。因国家内部的学术影响力和科研资源分配等存在地区性、学科性差异,故将城市学科作为研究的立足点较国家学科更具代表性。因此,本文依托于现有国家学科层面相关研究以及城市网络研究日趋成熟的基础上,深入学科层面对城市知识网络开展微观研究,为国家学科发展提供更细致的全息画像。

1.2 学科领域相关研究

学科分类是对具有一定关联学科的归类,最终分类结果构成学科分类体系。出于不同的应用目的,产生了不同的学科分类体系。一个科学的学科分类体系不仅结构完整、逻辑缜密,并且在理论上有助于人们对科学的从属关系有

整体的认知和把握,在实践中能够为科技政策和科技发展规划以及科研成果统计和管理服务。目前,国内外学者围绕学科领域进行了广泛的研究。陈雨^[9]开发了分类体系差异性测度软件,并对WoS和Scopus学科分类对学科学术竞争力评价结果的影响展开分析。王文平^[10]探讨中国国际科技合作对学科发展的影响并发现新兴的、应用性较强的领域跨学科研究程度较高。Sun等^[11]依据APS中的PACS代码划分了现代物理学子领域,探究其内外部知识的演化并发现知识流趋于同质化,表明跨学科研究的重要性。叶春雷^[12]基于WoS学科分类提出一种综合的跨学科态势分析方法,识别主题研究领域涉及的核心学科类别及其研究热点内容。闫慧等^[13]对国内外图情档领域研究热点进行比较,并探索了学术热点与学科发展的关系。不难发现,目前的研究大多局限于不同学科分类体系差异性的影响、跨学科研究及某一领域的学术热点等,而关于不同学科在不同计量实体间(科研人员、机构、国家和地区等)的横向比较相对较少,因此不利于发现差距和不足^[14]。鉴于此,本文在城市网络和学科领域两个中观研究的基础上,深入城市学科间的知识互动,对城市学科的学术影响力进行评价,比较不同城市学科间的差距并探索城市重点学科领域,为城市科学发展寻求方向。

1.3 影响力评价及计数方法相关研究

文献中评价作者、期刊或机构学术影响力最直接的方法是基于其所载论文的被引频次,然而这一方法没有对发文数量进行标准化^[15],且未区分施引文章的重要性。随着PageRank算

法的出现和广泛应用,利用 PageRank 算法评价学术影响力的方法应运而生,并衍生出各种变形。Chen 等^[16]使用 PageRank 算法评估论文的相对重要性。Ding^[17]将加权 PageRank 算法应用于作者引文网络,发现作者的知名度和声望均与加权 PageRank 高度相关。Zeng 等^[18]在量化科研论文的影响中详尽的总结和介绍了 PageRank 及其变形。尽管从基于计数的方法到基于网络的方法是科学计量学一次里程碑式的发展^[15],但 PageRank 算法只能回答学术实体的重要性问题而无法解答学术实体间的相互影响关系。

为能在考虑引用数量和质量的同时,一贯且定量的对学科领域的相对重要性和相互影响关系做出回答,Shen 等^[19]创造性地将经济学中的投入产出模型应用于物理学子领域。但文中使用的是后向投入产出关系矩阵 B,即衡量的是施引领域的相对影响力,而从学科领域的角度看,因被引用端是知识的生产者和贡献者,故考虑其影响力更具合理性与指导意义。Guo 等^[3]将初始投入产出矩阵所有元素加 1 并修正上述文章中的 B 矩阵,转而使用衡量被引端重要性的前向投入产出关系矩阵 F,并将其应用于国家学科层面。本文将城市学科的知识流网络转化为初始投入产出矩阵后不作任何处理,直接作为投入产出模型的输入,以此探究投入产出模型在更精细范围内应用的可行性及中国城市学科的发展现状,即将某个城市的某个学科视为模型中的一个部门,其引用其他城市学科的量视为投入,被其他城市学科引用的量视为产出,通过虚拟消去法探究城市学科的学术影响力,并将结果与 PageRank、HITS 算法比较,

以此考察投入产出方法的评价效果。

此外,诸多学者针对在科学计量研究中应使用何种计数方法这一问题开展研究。但琮洁^[20]依据不同计数法下国际合作论文中机构的贡献度,发现不同计数方法的综合运用,可以多角度反映机构国际合作论文的研究现状。张丽华等^[21]基于评价科研人员时不同计数方法的影响,认为计数方法的选择要与评价目的要结合。陈莉玥等^[22]基于信誉值分配将计数方法分为全计数法与分数计数法两大类,并比较计数方法在论文指标、引文指标、网络指标三个方面的差异。而由于全计数法的便利性,大多数研究尤其是空间计量学的文章中,多使用全计数法^[23]。因目前学界没有一致的结论,考虑到本文重点验证方法可行性的初始目的以及在城市层面细化论文作者地址信息的较大工作量,因此选择全计数法,即文中涉及的发文数量、被引频次及投入产出矩阵的元素均采用全计数法。

2 数据与研究方法

2.1 数据收集与处理

基于数据的可获得性,本文依托 APS 数据集开展相关研究。APS 于 1975 年开始采用 PACS 代码用于论文分类,并于 2016 年启用了新的分类方案 PhySH^[24],即物理主题词语,但因其正处于不断更新且逐渐完善的过程中,因此本文收集了 1980-2013 年发表在 APS 期刊的 379310 篇论文的信息(包括论文的 DOI、发表时间、PACS 代码、作者地址信息和参考文献及其上述信息)。在分析中,“城市”选取位于国家(州/地区等)和作者所属机构名称之间

的空间单元，而非空间计量学在城市相关研究中常用的大都市圈。在此基础上统计每个国家和城市的发文数量，其中发文数量超过 1000 篇的国家有 43 个，城市有 255 个。将在美国、日本、英国、德国、意大利和法国 6 个国家中发

文数量不足 1000 篇的城市归类为“国家二级区域名称 -Others”的形式以离散数据，在其他 37 个国家中的归类为“国家名称 -Others”，发文数量不足 1000 篇的部分归类为“Other-Others”，共得到 448 个城市地址。

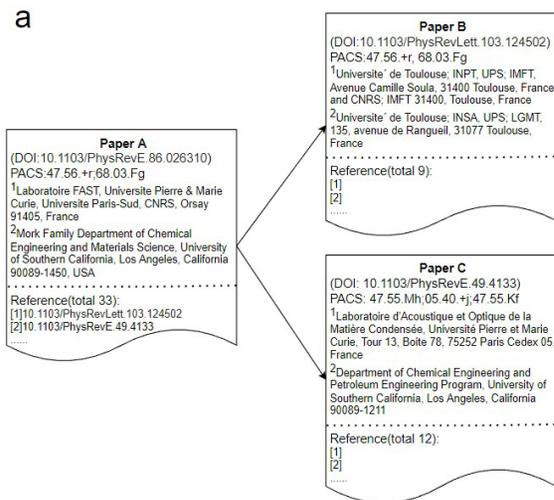


图 1a 论文地址、PACS 代码信息以及论文间引用关系，其中论文 A 引用了论文 B 和 C

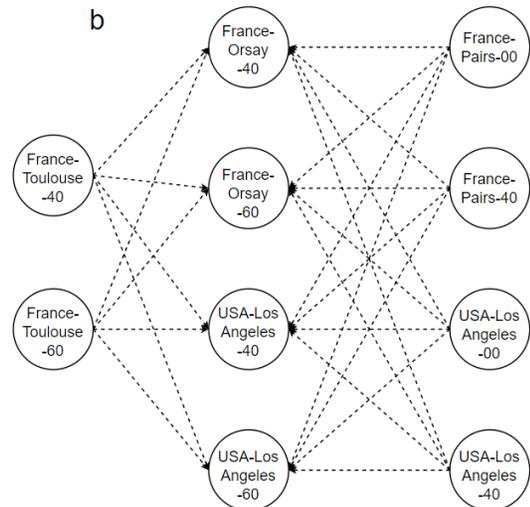


图 1b 由图 1a 中论文间引用关系得到的城市学科间知识流网络

APS 数据库编制了具有多层次结构的 PACS 代码用于论文分类，如图 1a 所示，最高级别的 PACS 代码将现代物理学划分为 00, 10, 20, 30, 40, 50, 60, 70, 80, 90 共 10 个主要的子领域^[25]，本文借此将论文划分到不同的物理学子领域，并以此代表论文的所属学科。每篇论文所属的城市和学科均确定以后，经配对组合得到论文所属的“城市 - 学科”，并根据论文间 6039964 条引用关系建立城市学科间知识流网络，如图 1b 所示，由此最终得到节点数为 4346，链接数为 4483299 的加权有向网络。

平衡关系时所使用的的方法，并已应用于中国物流业碳排放测算^[26]、新冠肺炎冲击路径分析^[27]与科普能力评价^[28]。在城市学科层面借鉴投入产出模型的思想，将每一个城市学科视为一个部门开展相应的分析^[29]。由上文建立的知识流网络，得到对应的投入产出矩阵 X ，如表 1 所示，矩阵的行为被引的城市学科 i ，列为施引的城市学科 j ，矩阵中的元素 x_{ij} 表示 i 传递给 j 的知识量，即网络中对应链接的权重。

2.2 研究方法

投入产出模型是研究国民经济各部门间

需要指出，城市学科的投入产出模型和传统的投入产出模型有两点不同：①文中投入产出模型是封闭系统，即不存在传统意义上的“增加值”行、“最终需求”列等，只存在各城市

学科间知识的“中间投入 - 中间使用”矩阵；
②在传统的投入产出模型中，每个部门的总投

入等于其总产出，而本文中城市学科 i 的总施引量等于其总被引量的概率极小。

表 1 城市学科间的投入产出矩阵

From To	France-Toulouse-40	France-Toulouse-60	France-Orsay-40	France-Orsay-60	USA-Los Angeles-00	USA-Los Angeles-40	USA-Los Angeles-60	France-Pairs-00	France-Pairs-40
France-Toulouse-40			1	1		1	1		
France-Toulouse-60			1	1		1	1		
France-Orsay-40									
France-Orsay-60									
USA-Los Angeles-00			1	1		1	1		
USA-Los Angeles-40			1	1		1	1		
USA-Los Angeles-60									
France-Pairs-00			1	1		1	1		
France-Pairs-40			1	1		1	1		

注：未填写的空格处为 0。

对投入产出矩阵 X 的元素 x_{ij} 做式 (1) 所示处理以得到前向投入产出关系矩阵 F ，矩阵的元素 f_{ij} 表示城市学科领域 i 每施引一次，带来的城市学科领域 j 的引用量：

$$f_{ij} = \frac{x_{ij}}{\sum_{l \in U} x_{li}} \quad (1)$$

其中， U 是城市学科领域 i 引用的所有城市学科集合。

F 矩阵构建完毕后采用虚拟消去法，即假设系统中没有某个城市学科 k ，检查整个系统的结构所发生的变化。变化越大，则表示城市学科 k 的影响力越大。若系统中没有某个城市学科 k ，即从 F 矩阵中去掉城市学科 k 所在的第 k 行 k 列，再计算新矩阵 $F^{(-k)}$ 的最大特征值。由于原始投入产出关系矩阵 F 的最大特征值为 1，且去掉城市学科 k 后系统的最大特征值 $\lambda^{(-k)}$ 是

小于等于 1 的，因此可由去掉城市学科 k 前后特征值的损失变化，衡量城市学科 k 的影响力。故定义投入产出学术影响力指标 IOF 如下：

$$IOF_F^k = 1 - \lambda^{(-k)} \quad (2)$$

其中， $\lambda^{(-k)}$ 是矩阵 $F^{(-k)}$ 的最大特征值， IOF_F^k 越大，则表示城市学科 k 越重要。

3 研究结果

3.1 APS 中各城市科研产出情况

科研产出一般通过发文数量或专利数量衡量^[30]。为考察城市在物理学领域的科研产出水平，本文统计各城市在 APS 中发表论文的数量，并分别从整体视角和高产出的局部视角探索城市科研产出的分布情况。

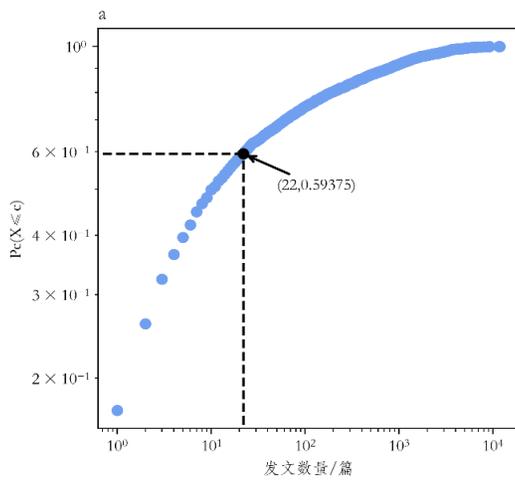


图 2a 城市发文数量的累计概率分布

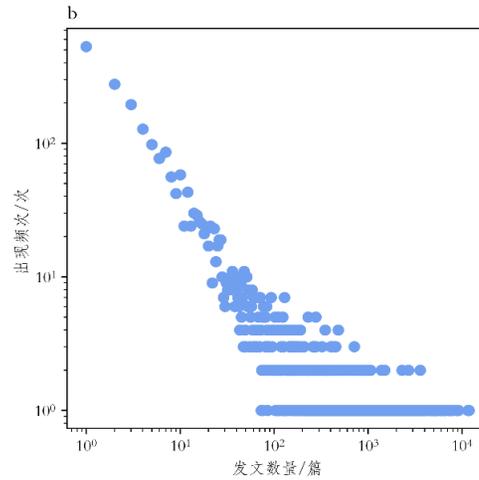


图 2b 城市发文数量的频次分布

图 2 为绘制的城市发文数量累计概率分布图及频次分布图，从整体看分析涉及 3104 个城市。其中在 1980-2013 年发文数量为 1 的城市有 531 个，占城市数量的 17.11%，发文数量不大于 10 篇的城市有 1548 个，占城市数量的 49.87%。同时，发文数量超过 10000 篇的城市也屈指可数，仅有日本的东京市、中国的北京市及美国的剑桥市。如表 2a 所示，发文数量排

名 20 的城市已锐减至 5493 篇。图 2a 中累计概率在发文数量较小时增长速度较快，当发文数量大于 22 篇时增长趋于缓慢。同样，图 2b 中出现明显的拖尾现象，故发文数量在城市间存在较大的异质性，即大部分城市发文数量较少，而少部分城市发文数量较大，整体服从幂律分布且符合二八法则，因此聚焦于高科研产出城市进一步分析。

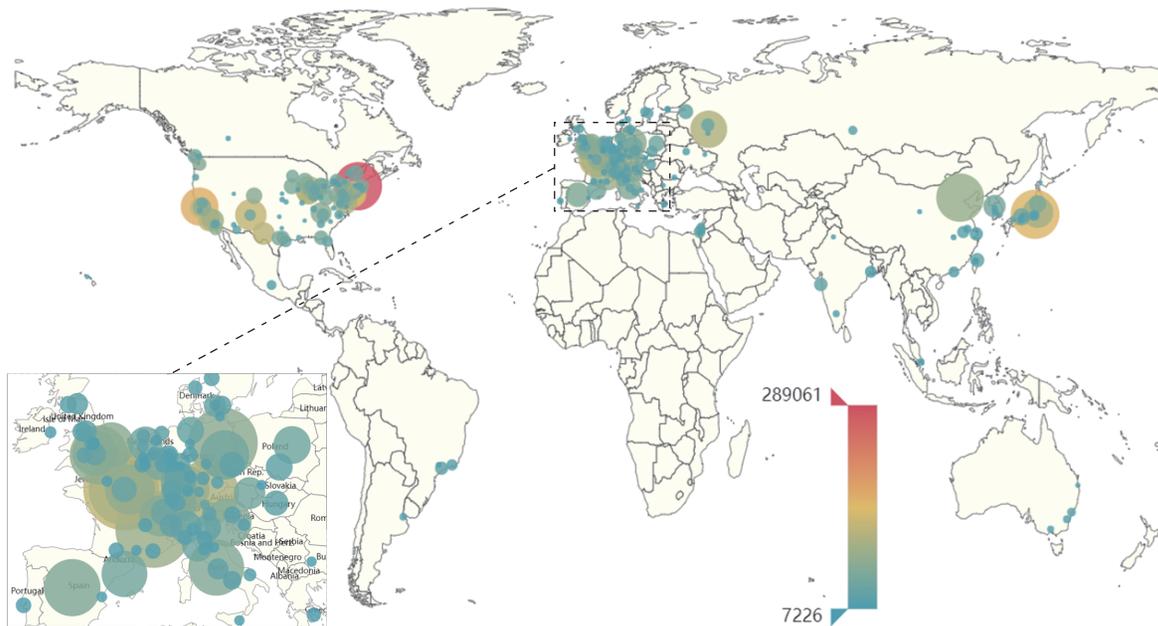


图 3 APS 中发文数量超过 1000 篇的城市的空间地理分布格局

表 2a APS 中发文数量排名前 20 的城市

排名	城市	发文数量
1	Japan-Tokyo	11833
2	China-Beijing	11627
3	USA-Cambridge	11598
4	USA-Berkeley	9168
5	Russia-Moscow	8975
6	France-Paris	8689
7	France-Grenoble	7969
8	USA-Los Alamos	7625
9	Japan-Tsukuba	7454
10	France-Orsay	7339
11	Germany-Berlin	7283
12	Germany-Garching bei München	6780
13	USA-Lemont	6561
14	USA-New York	6474
15	USA-Princeton	6209
16	UK-London	6111
17	Spain-Madrid	5927
18	USA-College Park	5923
19	Italy-Rome	5821
20	USA-Urbana	5493

表 2b 高科研产出城市中占比排名前 20 的国家

排名	国家	城市数量	占比
1	USA	85	0.3333
2	Germany	27	0.1059
3	Italy	16	0.0627
4	Japan	15	0.0588
5	France	11	0.0431
6	UK	11	0.0431
7	China (含港澳台)	10	0.0392
8	Canada	8	0.0314
9	Netherlands	7	0.0275
10	Israel	6	0.0235
11	Russia	5	0.0196
12	Australia	4	0.0157
13	India	4	0.0157
14	Korea	4	0.0157
15	Sweden	4	0.0157
16	Switzerland	4	0.0157
17	Belgium	3	0.0118
18	Brazil	3	0.0118
19	Denmark	3	0.0118
20	Spain	3	0.0118

本文将高科研产出城市划定为在 APS 中发文数量超过 1000 篇的城市，共 255 个城市参与分析。在地图上对各个城市进行地理定位，用点的大小表示城市的发文数量，颜色表示被引频次，在空间上呈现这些城市的地理分布情况，由此通过 APS 数据确定世界上哪些国家拥有物理学产出较高的城市。如图 3 所示，绝大多数具有高科研产出的城市位于三个主要的地理区域：44.31% 的城市位于欧洲；36.86% 位于北美洲，尤其是美国的东海岸；15.68% 位于亚洲，这与 Csomós 对 Scopus 数据库观察到的三个主要的地理区域吻合。同时，位于这些区域的城市论文被引频次也十分突出，且与发文

数量的 Pearson 相关系数为 0.8685。结合表 2b 可知，美国有 85 个高科研产出城市，占比高达 33.33%，因此在地图上呈现出广泛的分布。此外，在澳大利亚东海岸、印度、南美洲东部观察到零星分布，代表着这些地方物理科研产出的主要来源，而位于中国的高科研产出城市有北京、上海、兰州、武汉、南京、合肥、香港、台北、桃园、新竹。

3.2 由投入产出模型得到的城市学科学术影响力

由式 (2) 计算出 4346 个城市学科学术影响力的得分 IOF，将得分从高至低排序得

到城市学科的影响力排名情况，并绘制由投入产出模型得到的 IOF 排名与发文数量排名以及被引频次排名之间的相关图，如图 4 所示，图中点的颜色代表各城市学科所归属的不同国家。图 4b 中，大部分的点分布在 1:1 的对角线上，而 IOF 排名和发文数量排名相关图中的点分布则比较分散，但两图均呈现出一定的正相关性。采用 Spearman 等级相关系数考察两组排名的相关性分别为 0.9010 和 0.9984，与图中观察到的结果一致，后者的相关性要明显强于前者。

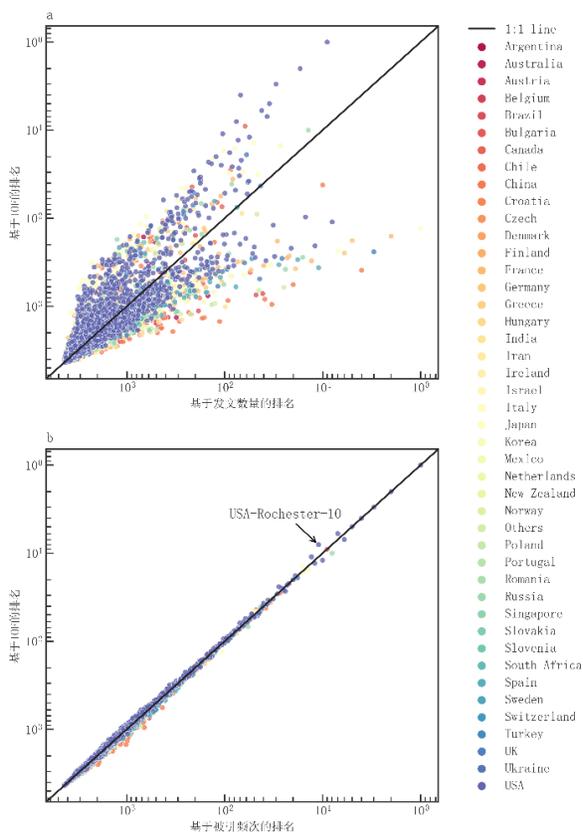


图 4 投入产出学术影响力指标 IOF 排名与发文数量排名的相关图以及与被引频次排名的相关图

此外，由图 4b 中的点分布在对角线周围，尽管能说明投入产出方法所得排名和被引频次

排名较为接近，但并不能因此认为被引频次比投入产出方法统计分析简单而选择被引频次衡量学术影响力，仍要选择投入产出方法的原因有：首先，两种方法具有本质区别。被引频次衡量的是城市学科领域的直接影响力，而投入产出方法考虑的是总影响效应，即直接影响和间接影响。其次，图 4b 在呈现两者相关性时，为更好的展现“头部”城市学科的排名差异，纵横坐标轴均采用对数形式，因而会放大两者间的线性关系，即实际分布要比所呈现的分散。因此，投入产出方法要优于被引频次，且通过图 4b 还可以判断哪些城市学科分别被高估或低估，如 USA-Rochester-10 位于对角线的左上方，即它的 IOF 排名高于被引频次的排名，若只单纯的依据被引频次对其评价，则会低估它的实际学术影响力。

从学科层面看，基本粒子物理与场（10）子领域处于绝对领先的地位。其中学术影响力排名前 97 的城市学科均属于 10 领域，且 10 领域处于城市各学科第一位的城市占城市数量的 61.38%。一个可能的解释是，城市层面 10 领域被引用的频次较高，因此在全计数法中占据优势，在结果中突出。为此，对所有学科领域（00-90）在城市层面的被引频次进行统计，发现 10 领域总被引频次高居第一位，且是其对应的施引文章的城市数量、学科数量及其所属被引文章的被引频次综合作用的结果。

表 3 为影响力排名前 20 的城市学科概况。从城市层面看，除去第 4 名是在美国二级行政区域得克萨斯州上合并的数据，其他均为行政意义上的城市，而其中有 12 个位于美国，虽然这与 APS 数据库归属美国有一定的原因，但也

能够见得美国在学术影响力上的主导地位；从学科层面看，前 20 名均属于 10 领域。同时值得注意的是，城市各学科间学术影响力差异较

大，美国剑桥市的 10 领域位列首位，而其 50（气体、等离子体和放电物理）领域则排名 1909 位，展现了城市内部学科发展的不均衡。

表 3 投入产出学术影响力指标 IOF 排名前 20 的城市学科概况

IOF排名	城市学科	IOF值	被引频次	被引频次排名	发文数量	发文数量排名
1	USA-Cambridge-10	0.007189	821579	1	3385	9
2	USA-Berkeley-10	0.007021	803793	2	2829	17
3	USA-Batavia-10	0.006092	683549	3	2463	30
4	Texas-Others-10	0.005888	678695	4	1943	69
5	USA-Chicago-10	0.005584	626833	5	2301	35
6	USA-New York-10	0.005454	612888	7	2222	42
7	USA-Madison-10	0.005453	622571	6	2272	37
8	USA-Rochester-10	0.005246	582831	11	1850	76
9	Canada-Montreal-10	0.005231	598322	9	2026	62
10	Russia-Moscow-10	0.005221	606047	8	2892	15
11	USA-Philadelphia-10	0.005132	578756	13	1623	105
12	USA-Columbus-10	0.005115	586037	10	2069	56
13	USA-Los Angeles-10	0.005108	582758	12	1892	73
14	Japan-Tsukuba-10	0.005085	570246	16	2106	51
15	Korea-Seoul-10	0.005054	577195	14	2494	27
16	Italy-Frascati-10	0.005000	570974	15	1794	79
17	Italy-Pisa-10	0.004827	549081	17	1647	103
18	USA-Urbana-10	0.004790	530654	19	1721	89
19	UK-London-10	0.004620	537391	18	2038	60
20	USA-New Haven-10	0.004588	517840	20	1534	123

图 5 绘制了中国 10 个高科研产出城市及国外 6 个重点城市在不同物理子领域下的排名情况，以此了解中国城市学科的发展现状。当城市在某个子领域中的影响力越高时，图中对应位置所堆叠的面积则越大。在城市学科综合排名中，台北市与北京市的 10 领域位列中国城市学科的前二，但北京市在不同子

领域的影响力较台北市更为均衡，且 20（核物理）领域在所有城市中排 15 位，高于其在 10 领域中的排名 28 位。中国其他高科研产出城市在各子领域中的排名则相对靠后。而与国外的 6 个重点城市相比，中国城市学科的影响力还存在一定的差距，具有较大的发展空间。

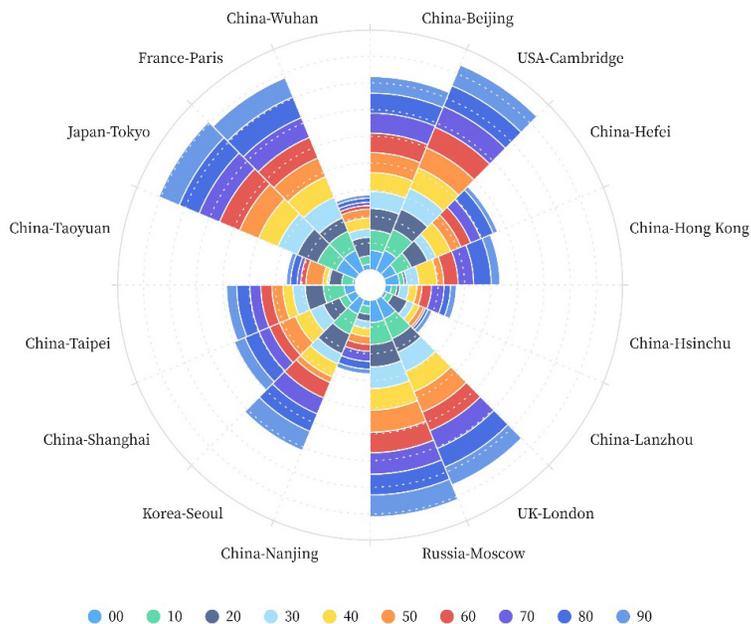


图5 中国10个高科研产出城市及国外6个重点城市在不同物理子领域下的影响力排名情况

3.3 与不同算法排序列表的比较

当基于复杂网络分析方法的科学学研究出现以后,以PageRank、HITS算法为代表的结构化指标相继涌现并产生了各种变形,因此本文选取上述两种传统的算法与投入产出方法进行比较,以验证投入产出方法的有效性。PageRank算法作为互联网网页重要度的计算方法,最初是由Google用作网页排序,并逐步在度量标准文献重要性^[31]、核心专利发现^[32]、文献检索排序^[33]等方面有了广泛的应用。HITS算法是通过两个权值,即内容权威度(Authority)和链接权威度(Hub)计算网页的得分。因PageRank算法只提供一个值,此值可以认为是网页的Authority值^[34]。基于本文2.1中所建立的知识流网络及投入产出矩阵 X ,由式(3)和式(4)分别进行迭代计算,当前后两次迭代的数值差全部小于 $1 \times e^{-15}$ 时停止,此时得到各个

节点的PageRank和Authority值,即为各城市学科的学术影响力得分。据此绘制城市学科基于IOF排名分别与基于PageRank算法的排名、基于HITS算法的Authority值排名的相关图,如图6所示,并计算相应的Spearman等级相关系数分别为0.9728和0.8283。

$$PR(i) = \frac{1-d}{N} + d \sum_{j \in U} \frac{PR(j) \cdot x_{ij}}{\sum_{l \in R} x_{lj}} \quad (3)$$

其中, d 为阻尼因子,一般取0.85; N 为网络中的节点数; $PR(i)$ 、 $PR(j)$ 分别为节点 i 、 j 的PageRank值; U 为节点 i 指向的所有节点的集合, R 是指向节点 j 的所有节点的集合。

$$Authority^{(k)} = X \cdot hub^{(k-1)}$$

$$hub^{(k)} = X^T \cdot Authority^{(k-1)} \quad (4)$$

$Authority^{(k)}$ 和 $hub^{(k)}$ 分别表示在 k 次迭代计算得出的Authority值与Hub值,式中 X^T 表示 X 的转置矩阵。

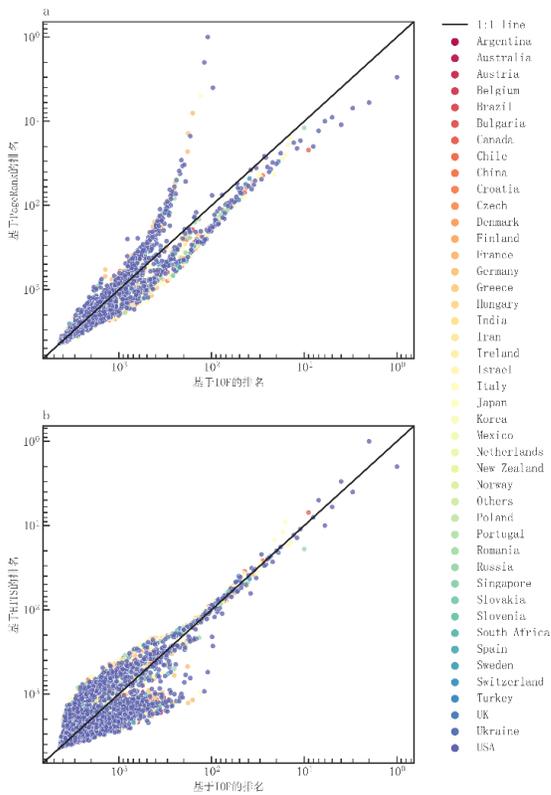


图 6 投入产出学术影响力指标 IOF 排名与 PageRank 算法排名的相关图以及 HITS 算法排名的相关图

为观察各方法中排名靠前的城市学科的一致性情况，本文重点选取在投入产出方法中排名前 50 的城市学科在基于被引频次、PageRank 和 HITS 算法下的排名分布情况。如图 7 所示，IOF 排序列表和 HITS 排序列表的前 50 与被引频次排序列表基本一致，为此数据集中的高被引城市学科，和 IOF 排序差异最大的是 PageR-

ank 算法，投入产出方法的前 50 名分布在 PageRank 算法的前 80 名中，其中 Japan-Osaka-10 排名差异最大，在两种方法中分别排第 45 名和第 80 名。为此，研究每种方法下排名前 100 的城市学科的被引频次分布和平均被引频次情况，如表 4 所示。PageRank 算法排名前 100 的城市学科中只有 39 个被引频次介于 30-50 万次之间，有 36 个的被引频次少于 30 万次，这可以从侧面说明重要的引用会剧烈影响排序列表的顺序。同样从平均被引频次角度可以看出，投入产出排序列表与被引频次排序列表最相似，而 PageRank 排序列表与被引频次排序列表差异最大。

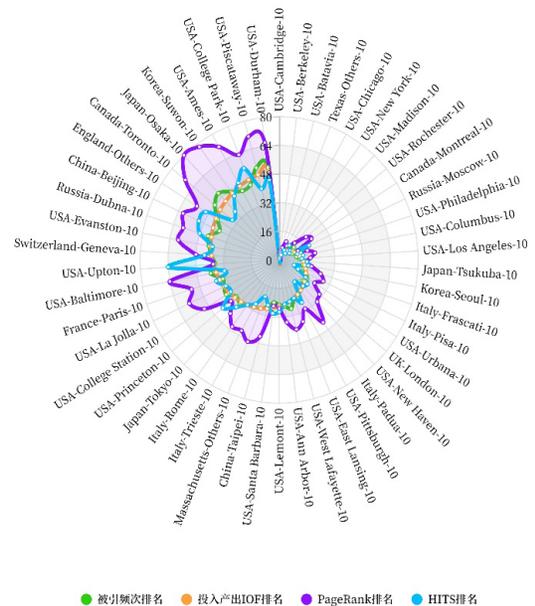


图 7 投入产出学术影响力指标 IOF 排名前 50 的城市学科在被引频次、PageRank、HITS 算法下的排名分布

表 4 各方法排序前 100 的城市学科的被引频次分布和平均被引频次

方法	被引频次分布			平均被引频次
	≥500000	300000-500000	≤300000	
被引频次	25	52	23	413665.20
投入产出	25	52	23	413460.81
PageRank	25	39	36	357565.58
HITS-A	25	52	23	411917.25

表 5 各算法排序前 100 中共有城市学科的数量

方法	被引频次	投入产出	PageRank	HITS-A
被引频次	—			
投入产出	98	—		
PageRank	67	67	—	
HITS-A	94	94	64	—

如表 5 所示, PageRank 算法与其他三种方法排名列表前 100 的共有城市学科数都较少, 表明其与其他方法的差异较大, 其中与 HITS-A 算法差异最大, 共有城市学科数仅有 64 个。因此, PageRank 算法得到的排序列表前列中会包括来自不同领域的城市学科, 而不仅仅是 10 领域, 故更适用于全局排序。HITS 算法提供 Authority 和 Hub 两个分数, 考虑了内容的相关性, 更适用于同一领域内的排序, 但计算效率要低于 PageRank 算法。而投入产出方法与被引频次的共有城市学科数为 98, 表明其与被引频次排序方法高度相似, 说明其对引用结构的敏感度较低, 即它对引用重要性的区分效果没有 PageRank 算法和 HITS 算法明显^[34]。因 PageRank 算法中的阻尼因子 d 为可变概率, 故其结果依赖于 d 值的设定, 而投入产出方法所得的学术影响力为确定型结果, 同时投入产出模型可以定量评估城市学科间的相互影响关系^[19], 因此, 投入产出方法要比 PageRank 算法的结果更稳定、准确, 且超越了 PageRank、HITS 算法解决问题的范围, 具有一定的优势。

综上, 与只考虑被引频次的做法相比, 投入产出、PageRank 和 HITS 算法综合考虑引用的数量和质量的做法更为合理, 能够更好的反映城市学科的学术影响力, 因而投入产出方法为评价城市学科的影响力提供了一种新的分析

范式, 但选择算法时要考虑算法的适用场景以及研究的目的做出决定。

4 结论与展望

4.1 结论与讨论

本文将经济学中的投入产出模型结合知识网络的背景与内涵创造性地应用于城市学科, 并通过 APS 数据集对方法的可实施性与评价效果进行验证和考察。首先对城市科研产出分布情况进行探究, 发现城市间发文数量呈现非均衡特征, 整体服从幂律分布。发文数量超过 1000 篇的城市共 255 个, 通过地理定位发现其多位于北美洲、欧洲和亚洲, 尤其是美国和德国呈现出广泛的分布, 且这些城市的论文被引频次也相当突出; 其次, 借助投入产出模型衡量各城市学科的影响力, 研究发现城市内部各学科间影响力差异较大, 其中 10 领域在城市各学科中排名第一的城市占城市数量的 61.38%, 且影响力排名靠前的城市学科多为美国城市的基本粒子物理与场领域。北京市的学科发展相对均衡, 中国的城市学科具有较大的发展空间; 最后, 将投入产出方法得到的排序列表与 PageRank 及 HITS 算法所得排序列表比较, 并讨论了每种方法下城市学科的共有数和平均被引频次等特点。由于三种方法均综合考虑了引用的数量和质量, 因此要优于仅仅依据被引频次的方法, 而投入产出模型对于引用结构的敏感度要低于 PageRank、HITS 算法, 即不会赋予重要引用以过分的权重, 且不受外界参数的影响, 能得到更稳定、准确的结果, 同时能定量回答城市学科领域间的相互影响关系问题, 为评价

城市学科的影响力提供一种新的、可靠的分析框架和范式。

基于研究结果可以看出,中国城市1980-2013年间在APS中的发文数量尚可,尤其是北京,仅次于日本东京市处于所有城市的第二名,且北京和台北市10领域的影响力处在前1%左右,学术影响力不容小觑。而近些年中国的科研产出更是呈井喷式增长,北京已连续三年蝉联全球科研城市首位^[35]。如今,中国已经具备了孕育世界科学中心所需要的经济条件和物质基础,2020年12月至2021年6月,京津冀、粤港澳大湾区、长三角三个综合类国家技术创新中心分别在北京、广州和上海揭牌成立,同时《北京市“十四五”时期国际科技创新中心建设规划》中也明确指出,到2025年北京将建设成为世界科学中心和创新高地^[35],因此在这难得的“机会窗口”,中国应把握历史机遇,发挥科学优势,做好城市重点布局工作,并积极破解挑战性难题,抢占前沿研究与颠覆性技术创新先机,掌握科学话语权。

4.2 研究局限与展望

本研究存在一定的局限性并有待于进一步研究,主要表现在:第一,由于数据的可获得性,本文基于APS数据库1980-2013年内的数据开展分析,缺乏对宏观学科及最新数据的分析;第二,使用的计数方法单一,未探讨不同计数方法间的差异;第三,文章仅从引文分析视角以数量化的思维对城市学科的学术影响力做出评价,后续应考虑基础学科学术成果的创新性及突破性等因素综合评价城市学科的学术影响力。但投入产出方法适用于任何可以提取

城市信息和具有学科分类的数据集,如WoS、Scopus、OpenAlex等,因此笔者未来将会在医学、数学、生物学等宏观学科门类及不同计数方法下探究城市学科的知识流动,并将引入时间因素,结合各城市的相关科技政策,研究城市学科的影响力随时间的变化,从中发掘重点城市的核心优势,为城市学科的发展做出更多贡献;同时将定量的探究城市学科间的支撑或竞争关系,为中国的学科布局、科学研究以及科创中心建设提供参考。

参 考 文 献

- [1] 马海涛. 知识流动空间的城市关系建构与创新网络模拟[J]. 地理学报, 2020, 75(4): 708-721.
- [2] 黄卫. 加强我国面向世界科技强国的基础研究基本布局和若干思考[J]. 中国软科学, 2017(8): 1-8.
- [3] Guo J Z, Liu X L, Yang L Y, et al. Are Contributions from Chinese Physicists Undercited? [J]. Journal of Data and Information Science, 2019, 4(4): 86-97.
- [4] 桂钦昌, 杜德斌, 刘承良, 等. 全球城市知识流动网络的结构特征与影响因素[J]. 地理研究, 2021, 40(5): 1320-1337.
- [5] Zhang Q, Perra N, Gonçalves B, et al. Characterizing scientific production and consumption in physics [J]. Scientific reports, 2013, 3(4): 1-9.
- [6] Csomós, Gyrgy, Lengyel B. Mapping the efficiency of international scientific collaboration between cities worldwide [J]. 2019: 1-4.
- [7] 马海涛. 基于知识流动的中国城市网络研究进展与展望[J]. 经济地理, 2016, 36(11): 207-213.
- [8] 曹湛, 戴靓, 吴康, 等. 全球城市知识合作网络演化的结构特征与驱动因素[J]. 地理研究, 2022, 41(4): 1072-1091.
- [9] 陈雨. WoS与Scopus学科分类对学科学术竞争力评价结果的影响研究[D]. 中国农业大学, 2018.
- [10] 王文平. 基于科学计量的中国国际科技合作模式及影响研究[D]. 北京理工大学, 2014.
- [11] Sun Y, Latora V. The evolution of knowledge within and across fields in modern physics [J]. Scientific

- reports, 2020, 10(1): 1-9.
- [12] 叶春蕾. 基于 Web of Science 学科分类的主题研究领域跨学科态势分析方法研究 [J]. 图书情报工作, 2018, 62(2): 127-134.
- [13] 闫慧, 陈慧彤. 国内外图情档领域学术热点比较研究——兼论学术热点与学科发展的协同关系 [J]. 情报资料工作, 2022, 43(2): 5-13.
- [14] 赵乃瑄, 卞方方, 蒋旭牧, 朱晓峰. 国际比较视域下的中国图情领域学科话语权演变及地位研究 [J/OL]. 情报理论与实践: 1-17 [2022-05-25]. <http://kns.cnki.net/kcms/detail/11.1762.G3.20220406.1431.002.html>.
- [15] 步一, 许家伟, 黄文彬. 基于引文的科学文献定量评价: 引文影响力指标评述 [J]. 图书情报知识, 2021, 38(6): 47-59.
- [16] Chen P, Xie H, Maslov S, et al. Finding scientific gems with Google's PageRank algorithm [J]. 2007, 1(1): 8-15.
- [17] Ding Y. Applying weighted PageRank to author citation networks [J]. Journal of the Association for Information Science & Technology, 2011, 62(2): 236-245.
- [18] Zeng A, Shen Z S, Zhou J L, et al. The science of science: From the perspective of complex systems [J]. Physics Reports, 2017, 714(714-715).
- [19] Shen Z S, Yang L Y, Pei J S, et al. Interrelations among scientific fields and their relative influences revealed by an input-output analysis [J]. Journal of Informetrics, 2016, 10(1): 82-97.
- [20] 但琼洁. 不同计数方法在国际合作论文中分析机构贡献度的应用 [J]. 情报工程, 2021, 7(3): 32-40.
- [21] 张丽华, 田丹, 曲建升. 计数方法在科研人员评价过程中的影响研究 [J]. 情报杂志, 2019, 38(9): 171-179.
- [22] 陈莉玥, 杨立英, 丁洁兰. 科学计量研究中全计数法与分数计数法研究综述 [J]. 图书情报工作, 2018, 62(23): 132-141.
- [23] Csomós, Gyrgy. On the challenges ahead of spatial scientometrics focusing on the city level [J]. Aslib Journal of Information Management, 2020, 72(1): 67-87.
- [24] Smith A. From PACS to PhySH [J]. Nature Reviews Physics, 2019, 1(1): 8-11.
- [25] 赵蓉, 赵基明. 物理学与天文学分类系统简介 [J]. 物理, 2002(9): 581-583.
- [26] 王丽萍, 刘明浩. 基于投入产出法的中国物流业碳排放测算及影响因素研究 [J]. 资源科学, 2018, 40(1): 195-206.
- [27] 刘世锦, 韩阳, 王大伟. 基于投入产出架构的新冠肺炎疫情冲击路径分析与应对政策 [J]. 管理世界, 2020, 36(5): 1-12.
- [28] 单孟丽, 张思光. 基于投入产出视角的科普能力评价研究 [J]. 中国科技资源导刊, 2022, 54(2): 55-64.
- [29] MILLER. Input-Output Analysis: Foundations and Extensions [M]. Input-output analysis: foundations and extensions, 2009: 1-54.
- [30] 李兰芳, 陈云伟, 张雪, 邓勇. 论空间科学计量学的研究与应用 [J/OL]. 农业图书情报学报: 1-12 [2022-05-25]. DOI: 10.13998/j.cnki.issn1002-1248.22-0078.
- [31] 李涛, 汪光阳. PageRank 在度量标准文献重要性中的研究 [J]. 苏州科技大学学报 (自然科学版), 2017, 34(2): 59-62.
- [32] 张欣, 马瑞敏. 基于改进 PageRank 算法的核心专利发现研究 [J]. 图书情报工作, 2018, 62(10): 106-115.
- [33] 汪志伟, 邹艳妮, 吴舒霞. PageRank 算法应用在文献检索排序中的研究及改进 [J]. 情报理论与实践, 2016, 39(11): 126-130.
- [34] 苏成, Hee-Sop KIM. 基于 PageRank, HITS 和 SALSA 算法的学术论文评价 [J]. 情报杂志, 2015, 34(6): 48-54.
- [35] 张航, 李祺瑶. 北京将建成世界科学中心和创新高地 [N]. 北京日报, 2021-11-25.