

新世纪以来知识图谱软件扩散规律研究

——以2000–2019年Web of Science核心合集为例

朱猛男¹,张耀军²

(1. 上海出版印刷高等专科学校 信息化办公室; 2. 上海出版印刷高等专科学校 规划发展处, 上海 200093)

摘要:为探索知识图谱软件的扩散路径及扩散影响力,预测其发展趋势,基于创新扩散理论,运用文献计量法和主路径分析法等数据+图谱的方法,根据扩散曲线、扩散路径和扩散网络探索其扩散规律。实证分析得出,知识图谱软件的扩散过程基本符合S型曲线规律。随着软件的更新与迭代,其扩散路径分化出两条关键主路径,主路径上有4个核心人物(Garfield E、Merigo J M、Van Eck N J和Chen C M)和3个重要事件(HistCite、CiteSpace和VOSviewer)。

关键词:知识图谱软件;创新扩散;扩散曲线;主路径分析

DOI: 10.11907/rjdk.211295

中图分类号: TP301

文献标识码: A

开放科学(资源服务)标识码(OSID):

文章编号: 1672-7800(2022)001-0176-06



Research on Diffusion Regularity of Knowledge Domain Visualization Software Since the New Century: A Case Study of Web of Science Core Collection from 2000 to 2019

ZHU Meng-nan¹, ZHANG Yao-jun²

(1. Informationization Office; 2. Planning and Development Department, Shanghai Publishing and Printing College, Shanghai 200093, China)

Abstract: To explore the process and effect of the diffusion of knowledge domain visualization software, and predict its future development trend; Based on the “diffusion of innovation” theory, take “data + graph” methods such as bibliometric method and main path analysis method to explore the diffusion law of knowledge domain visualization software in terms of diffusion curve, diffusion paths, and diffusion network. The results show that the knowledge domain visualization software’s diffusion process conforms to the S-curve rule. With the software’s update and iteration, the diffusion path of knowledge domain visualization software is divided into two key main paths, there are four core characters (Garfield E, Merigo J M, Van Eck, N J and Chen C M) and three important events (HistCite, CiteSpace and VOSviewer).

Key Words: knowledge domain visualization software; diffusion of innovation; diffusion curve; main path analysis

0 引言

软件工具对于科学研究具有重要辅助作用,可用于问题识别、数据存储、数据分析、数据建模、可视化结果呈现等。然而,长久以来,软件的学术价值一直被低估甚至被忽略^[1]。在科研大数据时代,基于科学文献挖掘研究内容的知识关联、结构关系与发展趋势已成为一种新方向,科学知识图谱的理论、技术与方法应运而生,为科学研究提

供了一种新范式。知识图谱软件运用数据分析技术和信息可视化技术展示学科领域的发展进程和结构概况,辅助文献计量分析和可视化研究,逐渐受到学术领域的重视,一些学者指出软件应被视为学术研究产出。据不完全统计,目前国内外开发的免费使用的知识图谱软件有几十种,其中 CiteSpace、VOSviewer、SCI 2、BibExcel、HistCite、CitNetExplorer、Gephi、Pajek、NetDraw、Ucinet、CRExplorer较为流行。

国内外对知识图谱软件的研究持续升温,从理论、绘

收稿日期: 2021-03-05

作者简介: 朱猛男(1993-),女,硕士,上海出版印刷高等专科学校信息化办公室助理工程师,研究方向为信息计量与知识传播;张耀军(1987-),男,硕士,上海出版印刷高等专科学校规划发展处助理研究员,研究方向为职业教育。

制方法、实证分析、应用等诸多方面展开研究,研究领域已扩展到计量学、工程学、生物学、管理学、文学等学科。如 Liao 等^[2]使用 GraphPad Prism 5、VOSviewer 和 CiteSpace 软件对医学大数据 (Medical Big Data, MBD) 进行文献计量分析,从多层面可视化展现 MBD 的发展现状和趋势;Yeung 等^[3]将 299 602 篇有关抗氧化剂的文献题录及引用文献导入 VOSviewer 软件进行文献计量分析,通过标签视图的形式可视化展示研究关键词,并绘制出关键期刊的引文网络标签图谱;徐建国等^[4]基于 CiteSpace 和 VOSviewer 绘制出国内深度学习领域相关文献的知识图谱,通过分析关键词时区、突变词检测、关键词共现权重等图谱,发现该领域的研究热点主要集中在卷积神经网络模型及其应用、深度学习模型创新改进等方面。

目前,较多学者开始关注软件工具的学术价值及影响力,将学术文献作为载体研究软件的扩散规律和扩散模式^[5-6],还有一些学者尝试采用被引次数、使用频次、下载次数等指标定量分析软件的影响力^[7]。如 Pan 等^[8]以扩散广度、扩散时间、扩散速度为指标衡量 CiteSpace、HistCite 和 VOSviewer 的使用、引用和传播情况;Cobo 等^[9]对 9 个代表性知识图谱软件的核心功能进行比较研究,运用科学计量和绘制知识图谱的方法突出不同软件工具之间的主要差异和协同合作关系。然而学术界对知识图谱软件扩散的定量研究较少,缺少对软件扩散特征、学科扩散规律、扩散网络以及扩散路径等多维度的研究。

本文基于创新扩散理论,对 2000–2019 年 Web of Science 核心合集中提及知识图谱软件的论文进行实证分析。首先对文献进行计量,绘制出知识图谱软件的扩散曲线;然后运用主路径分析法从引文网络中析出关键主路径,识别出扩散过程中的重要人物和重要事件,揭示重要软件的传播影响力;最后构建主路径扩散网络,梳理扩散脉络。探索知识图谱软件的扩散规律可以丰富和发展软件技术扩散的理论和研究方法,追溯软件扩散轨迹,了解软件发展过程,加深学术界对知识图谱软件学术价值的认知,促使其更新迭代与持续创新扩散。

1 研究设计

1.1 数据来源

以 Web of Science 核心合集为数据来源,对 CiteSpace、HistCite、VOSviewer、BibExcel、Ucinet、Pajek、Gephi、NetDraw 共 8 种主流知识图谱软件的名称或扩展词进行主题检索,检索表达式为:TS=(CiteSpace) OR TS=(“Cite Space”) OR TS=(HistCite) OR TS=(“Hist Cite”) OR TS=(VOSviewer) OR TS=(“VOS viewer”) OR TS=(BibExcel) OR TS=(“Bib Excel”) OR TS=(Ucinet) OR TS=(Pajek) OR TS=(Gephi) OR TS=(NetDraw) OR TS=(“Net Draw”),时间跨度为 2000–2019 年。共检索得到 1 494 篇文献,以此作为实证分析的数据集。

1.2 研究方法

创新扩散理论指出创新的扩散是一种社会过程,需要

经过一定的时间,通过某种渠道在特定的社会系统中进行传播。大多数创新的扩散过程符合 S 型曲线规律,即某项创新的累积采纳数随时间变化呈现为 S 型趋势。扩散过程可分为 3 个阶段,即起步阶段、起飞阶段和稳定阶段^[10]。知识图谱软件是一种技术创新,其扩散可定义为:在科学研究系统中,知识图谱软件被应用到研究论文中,在一定的时间内通过科学文献或专利等载体进行知识交流,逐渐被研究学者等潜在使用者所接受。

知识图谱软件被研究学者采纳的过程类似于科学文献的被引过程,本研究从提及知识图谱软件的文献数量、文献内容和文献间的引用关系方面研究知识图谱软件的扩散规律。将 1 494 篇文献导入 VOSviewer 软件建立原始引文网络,其中 789 篇文献之间存在引用关系。使用 Pajek 软件进行主路径提取,深入研究关键文献和重要知识图谱软件,借助 Gephi 工具绘制主路径成分引文网络,采用社会网络分析法分析扩散网络的结构特征、扩散中心性。

2 知识图谱软件的扩散规律分析

2.1 知识图谱软件的扩散曲线

将论文累积数量表示为累计采纳数,每年新增的文献数量定义为采纳速率,绘制出 2000–2019 年知识图谱软件总体扩散曲线,如图 1 所示,基本显示为一条 S 型扩散曲线。近 20 年提及知识图谱软件的文献数量稳步增长,增幅较为稳定。2000–2011 年采纳速率较低,只有少数学者的研究论文提及知识图谱软件,累积采纳曲线坡度平缓,扩散速度慢,这一阶段为扩散起步阶段;2012 年起扩散曲线上升趋势明显,且没有回落趋势,因此 2012 年开始进入扩散起飞阶段;2012 年以后扩散采纳速率持续增长,没有出现扩散曲线的拐点,尚且不能描述 S 型曲线的全貌。

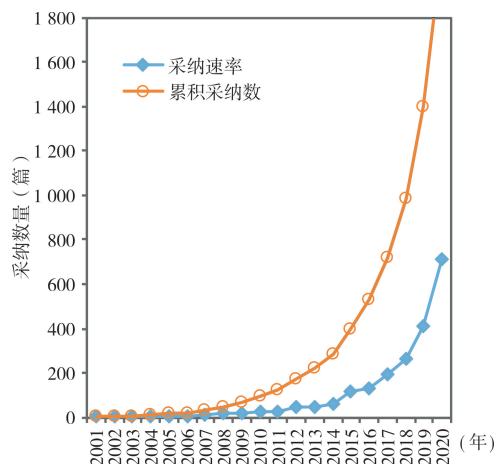


Fig. 1 Overall diffusion curve of knowledge domain visualization software

图 1 知识图谱软件总体扩散曲线

Bass 模型主要用于时间序列预测,知识图谱软件根据时间顺序向外扩散,呈现出上升或下降的变化趋势,因此可采用 Bass 模型对知识图谱软件的扩散趋势进行预测。Bass 模型的一般表达式为:

$$N(t) = m \left[\frac{1 - e^{-(p+q)t}}{1 + \frac{q}{p} e^{-(p+q)t}} \right] \quad (1)$$

以发文年作为自变量t,如2000年为第一年,则t=1,将累积采纳数作为因变量N(t),采用非线性最小二乘法,利用SPSS软件进行回归分析,得到估计参数值m=88 964,p=0.000 044 19,q=0.286 5。然后利用MATLAB软件对扩散曲线进行拟合,拟合图见图2,参数估计值如表1所示,可决系数R²的值接近于1,表示曲线拟合效果良好。根据Bass模型对知识图谱软件未来30年的累计采纳数进行预测,得出知识图谱软件的扩散过程符合S型曲线规律,预计2036年进入扩散稳定阶段。

Table 1 Estimation of cumulative adoption number diffusion model parameters based on Bass model

表1 基于Bass模型的累积采纳数扩散模型参数估计

参数	估计值
m	195 194
p	0.000 007 3
q	0.285 148 84
R ²	0.999 5

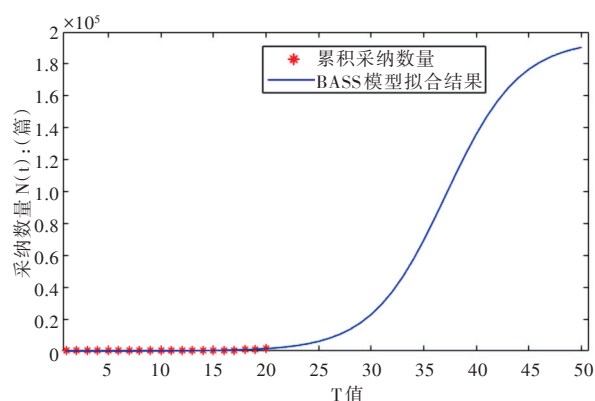


Fig. 2 Bass model fitting curve

图2 Bass模型拟合曲线图

2.2 知识图谱软件的扩散主路径分析

扩散路径是扩散过程的重要体现,从原始引文网络中析出的扩散主路径有两条(见图3),分别为路径1:3→6→9→30→41→68→161→176→189→201→225→298→334→366→372→420→447→627→704→709→766;路径2:3→6→10→41→68→161→176→189→201→225→298→334→366→372→420→447→627→704→709→766。图中节点标签采用“文献编号+第一作者+发文年”表示,箭头方向表

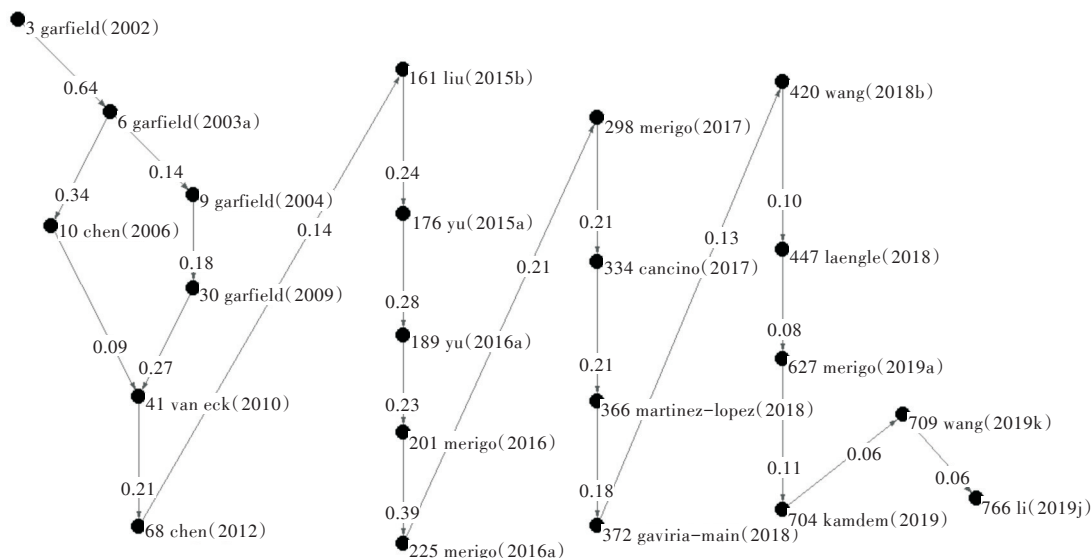


Fig. 3 Diffusion main path diagram of knowledge domain visualization software

图3 知识图谱软件扩散主路径图

示知识图谱软件的扩散方向,连线上的值表示权重,主路径上共有22篇关键文献,详见表2。可以看出,扩散主路径结构简单、样式单一,没有复杂的分支结构。

主路径文献的第一作者中, Garfield E出现了4次, Merigo J M出现4次, Chen C M和 Yu D J各出现2次;主路径文献的其他作者中, Pudovkin A I、Istomin V S、Cancino C A、Wang X X、Xu Z S各出现2次, Merigo J M出现10次。表2中,编号41的文献被引频次为1 063,编号10的文献被引842次,对应的第一作者分别为 Van Eck N J和 Chen C M。综上分析可知,在知识图谱软件的扩散过程中核心人物有4位,分别为 Garfield E、Merigo J M、Van Eck N J和 Chen C

M。Garfield E是美国著名的情报学家,编制出版了《科学引文索引》,被誉为“SCI之父”,其致力于研究可视化图谱并开发了HistCite软件;Van Eck N J和 Waltman L合作开发了VOSviewer软件;Chen C M(陈朝美)开发了CiteSpace软件;另有核心作者Merigo J M。

主路径关键文献最早于2002年出现在计算机科学、信息与图书馆研究领域,2012年扩散到生物技术与应用微生物学、医学研究和实验等领域,2016年逐渐扩散至其他研究领域,以工程学为主,还涉及商业与经济学、社会科学、运营研究与管理、化学、食品科学与工程、营养学与膳食学以及自动化及控制系统等学科领域。

Table 2 List of key literatures on the main path of knowledge domain visualization software diffusion
表 2 知识图谱软件扩散主路径关键文献

编号	标题	遍历权值	第一作者	发文年	被引频次
3	Algorithmic citation-linked historiography—mapping the literature of science	0.789 668	Garfield E	2002	19
6	Why do we need algorithmic historiography?	0.635 554	Garfield E	2003	92
9	Historiographic mapping of knowledge domains literature	0.286 182	Garfield E	2004	120
10	CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature	0.341 703	Chen C M	2006	842
30	From the science of science to scientometrics visualizing the history of science with HistCite software	0.266 794	Garfield E	2009	101
41	Software survey:vosviewer, a computer program for bibliometric mapping	0.355 414	Van Eck N J	2010	1 063
68	Emerging trends in regenerative medicine: a scientometric analysis in citespace	0.475 837	Chen C M	2012	146
161	Visualizing the intellectual structure and evolution of innovation systems research: a bibliometric analysis	0.240 774	Liu	2015	60
176	A scientometrics review on aggregation operator research	0.590 723	Yu D J	2015	64
189	Mapping development of linguistic decision making studies	0.512 393	Yu D J	2016	54
201	A bibliometric analysis of the first thirty years of the international journal of intelligent systems	0.386 041	Merigo J M	2016	0
225	Academic research in innovation: a country analysis	0.416 518	Merigo J M	2016	51
298	Thirty years of the international journal of intelligent systems: a bibliometric review	0.403 871	Merigo J M	2017	41
334	Forty years of computers & industrial engineering: a bibliometric analysis	0.583 356	Cancino C A	2017	55
366	Fifty years of european journal of marketing: a bibliometric analysis	0.349 452	Martinez-Lopez F J	2018	37
372	Twenty years of journal of knowledge management: a bibliometric analysis	0.350 768	Gaviria-Marin M	2018	20
420	A bibliometric analysis of the first twenty-five years of the international journal of uncertainty, fuzziness and knowledge-based systems	0.391 109	Wang W R	2018	26
447	Twenty-five years of group decision and negotiation: a bibliometric overview	0.198 227	Laengle S	2018	9
627	Forty years of safety science: a bibliometric overview	0.263 855	Merigo J M	2019	4
704	Research trends in food chemistry: a bibliometric review of its 40 years anniversary(1976–2016)	0.120 011	Kamdem J P	2019	4
709	Bibliometric analysis on research trends of international journal of computers communications & control	0.060 469	Wang X X	2019	1
766	Studies in informatics and control: a bibliometric analysis from 2008 to 2019	0.060 917	Li Y	2019	0

主路径上关键文献使用频次较多的知识图谱软件为 HistCite、CiteSpace 和 VOSviewer, 是扩散主路径上的 3 个重要事件。结合扩散曲线的分析结果, 将扩散主路径分为 2002–2010 年和 2012–2019 年两个阶段。

2.2.1 第一阶段: 2002–2010 年

Garfield E 于 2001 年开发出 HistCite 软件, 随后发表一些实证分析论文推广 HistCite, 其中文献[11](对应表 2 编号 3)发表于 2002 年, 遍历权值最高, 介绍了 HistCite 软件的主要功能, 展示了引文数目列表、引文关系以及“gene flow”的引文编年史图。HistCite 经过优化功能逐渐完善, 书目列表和可视化图谱比上一版本更加美化。文献[12](对应表 2 编号 6)采用研究主题、期刊、作者等数据实证展示文献列表和图谱。文献[13](对应表 2 编号 9)使用 HistCite 对特定研究领域的知识进行实证分析, 将文献计量分析与可视化图谱相结合。文献[14](对应表 2 编号 30)考察了从科学到科学计量学的历史发展轨迹, 论证强调了科学计量学的重要性。综上可知, 正是得益于 Garfield 对文献计量学和可视化研究的贡献和影响, 知识图谱研究才有如今的发展。

2003 年, 陈超美教授基于 Java 平台开发了 CiteSpace 这

款用于分析科学文献的可视化软件, 并于 2004 年发表了论文《Searching for intellectual turning points: progressive knowledge domain visualization》, 该篇文献^[15]在 Web of Science 核心合集中被引 355 次。其对 CiteSpace 软件识别关键节点的方法进行了详细阐述, 介绍了软件的核心功能与实现方式, 并实证展示了共被引网络图谱。2006 年, 陈教授发表了文献[16](对应表 2 编号 10), 被引 842 次, 为高被引论文, 该篇文献通过恐怖主义和大规模灭绝两个案例实证了 CiteSpaceII 的聚群视图和时区视图功能。CiteSpace 自对外发布以来不断更新迭代, 软件使用简单、功能强大, 被研究者广泛使用, 产出的研究论文也随之快速扩散, 由最初的计算机科学领域传播到多个学科领域。

VOSviewer 软件于 2009 年被开发出来, 文献[17](对应表 2 编号 41)于 2010 年发表, 其作者即为 VOSviewer 软件的开发者 Van Eck N J 和 Waltman L。该篇文献被主路径上 14 篇文献引用, 是扩散过程中的一个重要文献。其介绍了 VOSviewer 的 4 种视图, 阐述了可视化图谱的技术实现方式, 同时实证展现了 5 000 个主要科技期刊的共引标签视图, 以及计算机科学和经济学领域之间的文献分布视图。VOSviewer 既能分析标题和摘要中的共现词汇, 又能以网

期刊、作者、机构、国家等不同维度的可视化图谱,版本从 CiteSpaceI 到 CiteSpaceV 实现多次再创新,促使 CiteSpace 软件循序渐进地扩散;VOSviewer 软件是后发突起,在扩散中后期被大范围采纳,占据扩散主路径成分的半壁江山(占比 61.2%)。2014–2019 年,提及 VOSviewer 软件的研究论文数量、被引频次均占领先地位,可以认为 VOSviewer 是知识图谱软件扩散路径上的获胜者。

3 结论

本文以 Web of Science 核心合集在 2000–2019 年间提及知识图谱软件工具的研究论文为研究对象,主要运用主路径分析法和社会网络分析法进行研究,得出知识图谱软件的扩散规律如下:①知识图谱软件在 2000–2010 年处于扩散起步阶段,2012 年开始进入起飞阶段,扩散速度仍持续增长,通过建立 Bass 模型进行验证可知,知识图谱软件的扩散过程基本符合 S 型曲线规律,并预计在 2036 年进入扩散稳定阶段;②随着时间的推移,知识图谱软件扩散路径分化出两条关键主路径,主路径上有 4 个核心人物(Garfield E、Merigo J M、Van Eck N J 和 Chen C M)和 3 个重要事件(HistCite、CiteSpace 和 VOSviewer)。结合 S 型曲线规律可将扩散过程划分为两个阶段:2002–2010 年为第一阶段,主要介绍知识图谱软件的功能和展示实证分析图谱;2012–2019 年为第二阶段,该阶段较少出现新开发出来的软件,大多是优化已有软件功能,升级用户使用体验,但是知识图谱软件的实证分析研究成果逐渐增多;③8 个知识图谱软件中 HistCite、CiteSpace 和 VOSviewer 是扩散的主力军,其中 VOSviewer 在扩散网络中占据重要地位。

4 结语

本文以知识图谱软件为研究对象,丰富了知识扩散的研究内容,创新性地对主路径成分文献使用的知识图谱软件工具进行聚类,识别出关键软件和核心人物。但研究仍存在一定的局限性:由于软件评测尚未有统一标准,数据来源单一,选取的 8 个主流知识图谱软件不能充分说明整个知识图谱软件的扩散情况。此外,根据知识图谱软件名称检索获得的文献不能完全保证使用知识图谱软件开展实证分析,也有可能是对知识图谱软件进行理论分析和文献综述。本文采用主路径分析法析出的关键文献是粗线条的,未能直接反映出知识图谱软件的扩散归集,也没有添加时间轴分析其随时间的演化过程。后续研究将扩大数据范围,优化检索表达式,将主路径分析法与文献内容分析法等多种方法相结合,从学科、作者、期刊、机构等多个角度展开全面细致的研究。

参考文献:

[1] PIWOWAR H. Altmetrics: value all research products[J]. *Nature*, 2013, 493: 159.

- [2] LIAO H C, TANG M, LUO L, et al. A bibliometric analysis and visualization of medical big data research[J]. *Sustainability*, 2018, 10(1): 1–18.
- [3] YEUNG A W K, TZVETKOV N T, EI-TAWIL O S, et al. Antioxidants: scientific literature landscape analysis[J]. *Oxidative Medicine and Cellular Longevity*, 2019, 2019: 8278454.
- [4] XU J G, LIU Y H, LIU M F. Research progress and hotspots analysis in the field of deep learning in China: comprehensive application based on CiteSpace and VOSviewer[J]. *Software Guide*, 2021, 20(1): 234–237.
徐建国,刘泳慧,刘梦凡.国内深度学习领域研究进展与热点分析——基于 CiteSpace 与 VOSviewer 的综合应用[J]. *软件导刊*, 2021, 20(1): 234–237.
- [5] YU X T, PAN X L, HUA W N. Analysis of software citation and diffusion in knowledge mapping research[J]. *Information and Documentation Services*, 2019, 40(2): 19–29.
于晓彤,潘雪莲,华微娜.知识图谱研究中的软件引用和扩散分析[J]. *情报资料工作*, 2019, 40(2): 19–29.
- [6] ZHAO R Y, WEI M K, WANG S Z. Evaluation of academic impact of open source software based on Altmetrics[J]. *Journal of Library Science in China*, 2017, 43(2): 80–95.
赵蓉英,魏明坤,汪少震.基于 Altmetrics 的开源软件学术影响力评价研究[J]. *中国图书馆学报*, 2017, 43(2): 80–95.
- [7] THELWALL M, KOUSHA K. Academic software downloads from Google code: useful usage indicators?[J]. *Information Research: An International Electronic Journal*, 2016, 21(1): 1–14.
- [8] PAN X, YAN E, WANG Q, et al. Assessing the impact of software on science: a bootstrapped learning of software entities in fulltext papers[J]. *Journal of Informetrics*, 2015, 9(4): 860–871.
- [9] COBO M J, LOPEZ-HERRERA A G, HERRERA-VIEDMA E, et al. Science mapping software tools: review, analysis, and cooperative study among tools[J]. *Journal of the American Society for Information Science and Technology*, 2011, 62(7): 1382–1402.
- [10] ROGERS E M. Diffusion of innovations[M]. Translated by XIN X. Beijing: Central Compilation & Translation Press, 2002: 10.
埃弗雷特·M. 罗杰斯. 创新的扩散[M]. 辛欣, 译. 北京: 中央编译出版社, 2002: 10.
- [11] GARFIELD E, PUDOVKIN A I, ISTOMIN V S. Algorithmic citation-linked historiography——mapping the literature of science[J]. *Proceedings of the American Society for Information Science and Technology*, 2002, 39(1): 14–24.
- [12] GARFIELD E, PUDOVKIN A I, ISTOMIN V S. Why do we need algorithmic historiography?[J]. *Journal of the American Society for Information Science and Technology*, 2003, 54(5): 400–412.
- [13] GARFIELD E. Historiographic mapping of knowledge domains literature[J]. *Journal of Information Science*, 2004, 30(2): 119–145.
- [14] GARFIELD E. From the science of science to scientometrics visualizing the history of science with HistCite software[J]. *Journal of Informetrics*, 2009, 3(3): 173–179.
- [15] CHEN C M. Searching for intellectual turning points: progressive knowledge domain visualization[J]. *Proceedings of the National Academy of Sciences*, 2004, 101(suppl): 5303–5310.
- [16] CHEN C M. 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, 57(3): 359–377.
- [17] VANECK N J, WALTMAN L. Software survey: VOSviewer, a computer program for bibliometric mapping[J]. *Scientometrics*, 2010, 84(2): 523–538.