知识图谱研究中的软件引用和扩散分析*

于晓彤 潘雪莲 华薇娜 （南京大学信息管理学院 江苏 210023）

摘 要 文章先采用内容分析法对国内知识图谱研究论文中的软件引用情况进行深入细致分析，再利用扩散广度和扩散速度等知识扩散指标对论文中发现的高频使用的10个知识图谱软件在学术交流系统中的扩散特征进行探索性研究。结果表明：国内知识图谱研究对软件的依赖程度不断提升，2017年提及软件的论文占比超过90%，然而这并未带来软件引用缺失状况的改善，软件引用缺失依然严重；知识图谱研究人员接受和使用新软件的速度在不断加快；知识图谱软件起初的扩散比较缓慢，随着时间的推移而不断加快，目前仍未见扩散放缓；知识图谱软件较早地出现在图书馆、情报与文献学、管理学、新闻学与传播学和计算机科学领域，并在图书馆、情报与文献学、教育学和管理学研究中获得最广泛的应用。

关键词 软件引用 引用分析 软件扩散 知识扩散 学术评价

Analysis of Software Citation and Diffusion in Knowledge Mapping Research

Yu Xiaotong Pan Xuelian Hua Weina （School of Information Management, Nanjing University, Jiangsu, 210023）

Abstract The paper first examines the citation of software in Chinese core journal papers on knowledge mapping using a content analysis approach, and then explores the diffusion characteristics of the top ten most frequently used knowledge mapping software tools with several diffusion indicators such as diffusion breadth and diffusion speed. The findings are as follows: (1) Knowledge mapping research increasingly relies on software and the proportion of papers about knowledge mapping using software is more than 90% up to 2017. However, the state of software citation practice does not become better and the uncitedness of software in domestic knowledge mapping research is still noticeable. (2) The speed of new software acceptance and usage of Chinese researchers in the field of knowledge mapping is increasing. (3) The knowledge mapping software tools have been adopted slowly in early years, but afterward the diffusion speed grows rapidly up to now. (4) The knowledge mapping software tools have occurred earlier in the fields of library, information and document science, management, journalism and communication, and computer science. And they have been more frequently used in the research of library, information and document science, education and management.

Keywords software citation, citation analysis, software diffusion, knowledge diffusion, academic evaluation

1 引言

随着数据密集型科学研究范式的兴起，软件在科学研究中发挥着越来越重要的作用，因为数据本身不能做任何事，它依赖于软件工具的处理和分析[1-2]。已有调查显示，超过90%的研究者使用软件进行科学研究[3]。事实上，很多研究者甚至需要自主开发软件解决科学问题[4-5]。然而，长久以来，软件往往被认为是科学研究的副产品，其学术价值一直被低估甚至被忽略[6-7]。直到最近，随着一些科研资助机构和科研评价

* 本文系国家自然科学基金青年项目“基于全文本数据的软件实体抽取与学术影响力研究”（编号：71704077）的研究成果之一。

组织将软件认定为有效的科研成果[8-9]，图书情报学领域学者才逐渐开始关注软件的学术价值及其测度。Howison和Bullard[10]对生物学核心期刊论文中的软件可见性进行分析，发现65%的论文提及了软件，但仅有44%的被提及软件获得了正式引用。Li等[1]对PLOS ONE期刊论文中R软件的引用情况进行调查，发现仅有37.6%的论文按照R软件的官方引用说明进行引用。针对目前软件引用缺失严重的现状，一些学者尝试通过制定软件引用标准来推进引用行为的规范化[11]，而另外一些学者尝试用被引次数之外的评价指标，如使用频次、下载次数来测度软件的学术影响力[12-13]。

知识扩散是科学发展的一个基石，已经成为图书情报学领域的一个重要研究主题[14-15]。图情领域的很多学者从引用角度探索知识扩散的特征、模式和规律[16-20]，他们通常将引用视作从被引对象到引用对象的知识流，被引对象和引用对象则被视为知识扩散的来源和目标。知识扩散研究对象已经涉及论文[19]、专利[20]、手册[21]和数据库[22]等科研成果。然而，目前还少有研究将软件这一科研成果作为研究对象，从提及角度来分析知识扩散的特征和规律。在本研究中，我们将被提及的软件和提及软件的论文分别视为知识扩散的来源和目标，尝试用知识扩散测度指标来分析软件的学术影响力和扩散特征，以期揭示软件的学术价值、拓展目前以文献等科研成果为基本单元的知识扩散研究。

本文首先采用内容分析法对我国知识图谱研究中的软件提及和引用情况进行统计分析，再依据发现的知识图谱软件检索出更多的提及这些软件的核心期刊论文，最后基于收集到的数据，利用知识扩散测度指标探索软件在学术交流系统中的扩散特征。选择知识图谱研究，一方面是因为知识图谱研究是信息计量学领域的一个重要研究主题，另一方面是因为知识图谱研究较为依赖软件，一些专门的知识图谱工具已经在学术界获得了广泛使用[23-24]。探索分析知识图谱研究中的软件提及、引用和扩散情况，对深入认识软件的学术价值、推进软件引用规范化、促进软件共享和再利用以及拓展知识扩散研究范围等都有着重要意义。

2 相关研究

2.1 国内知识图谱研究

知识图谱是显示科学知识的结构关系和动态发展进程的一种图形[25]。自此概念传入中国以来，国内学者在知识图谱的理论、方法、工具和应用等诸多方面展开了研究。早期，学者们围绕知识图谱的概念辨析、理论基础和发展历程等方面进行研究。陈悦等[26]对科学知识图谱的概念、发展历程、原理与分类进行归纳总结。秦长江和侯汉清[27]对知识图谱的概念、国外知识图谱研究的发展历程以及构建知识图谱的理论、关键技术进行了概述。随着知识图谱研究的深入，学者们开始关注知识图谱的绘制方法与工具，并借助这些工具开展知识图谱的应用研究。杨思洛和韩瑞珍[28]介绍了知识图谱的绘制流程以及九种国外知识图谱绘制工具。刘启元和叶鹰[29]设计开发出文献题录信息统计分析工具软件SATI，该软件可以将Web of Science、知网、万方和维普数据库的题录数据处理成可导入Ucinet、SPSS等分析工具的格式，进而可利用这些工具生成相应的知识图谱。肖明等[30]介绍了知识图谱分析的一般流程并从软件支持的数据格式、构建关系句子所支持的方法等方面对Bibexcel、CiteSpace、HistCite、Network Workbench Tool、Gephi、Pajek、Sci2 Tool、SciMAT等12个知识图谱软件进行对比分析。大连理工大学刘则渊的研究团队应用知识图谱软件CiteSpace分析纳米技术研究、国际物流研究以及中国科学技术方法论研究的研究前沿和演进脉络[31-33]。宗乾进等[34-36]利用Ucinet、Netdraw、CiteSpace等知识图谱软件对我国图书馆学、情报学、档案学等多个学科的研究趋势进行探究。

2.2 软件引用及影响力评价研究

软件的开发者、用户以及科研资助机构都对软件的使用情况和影响力感兴趣。对于软件开发者来说，一方面可以通过了解软件的使用情况来确定应该对自己的软件进行何种修改和扩展，另一方面可以通过用户数量、类型和软件对他人科学研究的贡献来了解自己的学术影响力[37]。Trainer等[38]的研究发现软件的广泛使用数据有助于科学家获得科研资助。对用户来说，一方面可以根据本领域内他人使用的软件来改变自己对软件的选择，另一方面可以依据软件的使用频次和影响力来判断其质量[39]。已有研究发现，科研工作者倾向于选择使用被本领域内其他研究者广泛使用的软件，他们认为获得广泛使用是软件质量高的一个表现[40]。科研资助机构可以依据其所资助的软件所产生的影响来判断资源配置的合理程度[41]。因此，一些学者呼吁用具体的评价指标来测度软件的影响力[42]。由于被引次数被广泛用来测度出版物的学术影响力，有学

者提出用被引次数评价数据、软件等实体的影响力[43]。然而，对软件引用现状的一些调查发现，目前软件引用缺失、引用随意的现象普遍存在[1-2,10,44]。在这种情况下，用被引次数这单一评价指标来测度软件的学术影响力是不恰当的。Smith等[11]制定软件引用标准，以推进软件引用规范化。赵蓉英等[45]则提出用下载量、被引指标和复用指标来评价软件的学术影响力。

2.3 知识扩散研究

1962年，Rogers[46]提出了创新扩散理论（Diffusion of innovations），这是一个旨在解释新思想和新技术怎样、如何以及以何种速度传播的理论。他将创新扩散定义为“一种创新经由特定渠道随着时间推移在一个社会系统的成员中传播的过程”[47]。他认为，累积创新采纳人数随时间变化呈现出S型曲线特征，即创新的扩散刚开始比较缓慢，当采纳人数达到一定数量后，扩散过程迅速加快，这个过程一直延续到系统中有可能采纳创新的人大多都已采纳创新，之后扩散放缓，直至达到饱和状态，采纳创新人数不再增加。创新扩散理论自提出以来，已引起来自不同领域的很多研究者的注意。图书情报学领域的一些学者认为，科学交流系统中的基于学术引用的科学思想扩散过程类似社会系统中基于用户使用的创新扩散过程[48]。基于此，很多研究者开展了基于引用的知识扩散研究。Liu和Rousseau [48]对Charles Kuen Kao获得诺贝尔奖的论文Dielectric-fibre Surface Waveguides for Optical Frequencies在1966～2009年间的被引数据进行研究，发现该论文在科学交流系统中通过学术引用进行的扩散满足S型曲线分布。宋歌[15]对结构洞理论的扩散过程进行研究，发现结构洞理论的知识扩散过程呈现S型曲线特征。Van Leeuwen和Tijssen[49]对学科间的知识扩散进行研究，发现一个学科的论文倾向于引用本学科或相近学科的论文。与此同时，Rinia等[50]研究发现一篇论文在其所属领域内获得引用往往早于在其他领域获得引用。Liu等[19,51]已提出期刊扩散广度、学科扩散深度等量化指标来描述出版物的扩散特征、测度出版物的影响力。此外，Yu等[22]将数据库视作一种知识实体并研究其使用和扩散情况。尽管有如此多的知识扩散研究，但关于学术交流系统中的软件扩散研究还很少。

3 数据与方法

3.1 数据采集

本文选取中文社会科学引文索引（CSSCI）和中国科学引文数据库（CSCD）作为来源数据库，检索字段限定在题名、关键词和摘要，时间限定为2005～2017年（这是因为被视为国内知识图谱研究开端的论文《悄然兴起的科学知识图谱》发表在2005年[25]）。首先，我们依据陈超美[52]和曹树金等[53]在知识图谱研究综述性论文中使用的检索词选取了“知识图谱”“信息可视化”“知识可视化”“共被引”“共现”等进行第一轮检索（根据词间关系进行了组合运算）。虽然两篇论文都将常用的知识图谱软件如CiteSpace、HistCite等作为检索词，但是本研究的第一轮检索未将知识图谱软件作为检索词。在对第一轮检索结果去重之后，再依据杨思洛和韩瑞珍[28]对狭义知识图谱的限定删除涉及地理知识图谱和仿真可视化等非狭义知识图谱研究的学术性论文，最终获得1804篇论文。为了便于表述，将本轮检索获得的1804篇论文称为软件引用数据集，用于探索知识图谱研究对软件的依赖程度以及软件引用行为特征。然后，我们对获得的1804篇论文进行人工标注，并依据标注结果统计出高频软件。需要指出的是，我们以篇为单位统计软件出现频次，也就是说，即使某个软件在一篇论文中出现多次，其提及频次仍为1。最终从1804篇论文中发现194种软件，其中，仅有12种软件被30篇以上的论文提及。排除Excel和SPSS这样的通用软件后，剩下的10种高频知识图谱软件（即CiteSpace、Ucinet、Netdraw、Pajek、Bibexcel、Bicomb、VOSviewer、SATI、HistCite、TDA）将作为本研究第二轮检索的检索词。最后，我们用上述高频知识图谱软件名称进行第二轮检索（即在CSSCI和CSCD数据库中检索主题位置出现上述高频使用软件名称的论文），并将检索结果和软件引用数据集中提及这些高频知识图谱软件的论文进行合并去重，共获得2592篇论文，称之为软件扩散数据集，用于探索知识图谱软件的扩散过程。此外，为研究知识图谱软件学科级的扩散广度、扩散速度和加速度，将上述2592篇论文依其期刊所属学科分类：先依CSSCI期刊所属学科对其论文进行分类，再将剩余CSCD期刊论文依其学科体系分类（当某刊隶属多个学

科时，以该刊多数论文所属学科为准）。数据采集工作于2018年3月完成。

3.2 软件提及和引用特征编码

本文采用内容分析法对1804篇知识图谱研究论文中的软件提及和引用特征进行统计分析。内容分析法是一种对显性内容进行客观、系统和定量描述的研究方法[54]，是图书情报学领域的一个重要研究方法[55]。首先，依据已有研究[10]构建软件提及和引用特征编码表（见表1）。然后，由四位编码员依据编码表对1804篇知识图谱研究论文进行标注。最后，对编码结果进行统计分析。为保证编码质量，在经过培训的编码员独立进行标注之前，我们随机抽取20篇论文让四位编码员对其中四个主要指标分别进行独立编码，采用统计工具 ReCal3(http://dfreelon.org/utils/recalfront/recal3/#doc)[56]计算Krippendorff's Alpha值来检验编码员间的信度。四个指标的Alpha值分别为0.792、0.817、0.899和0.909均大于可接受信度0.7[54]，说明四位编码员的标注结果一致性较好。需要指出的是，只要期刊论文中出现软件名称，即认为该论文提及了软件。此外，本研究以期刊论文中明确提及的软件作为研究对象，那些实际被使用但未被明确提及的软件不在研究之列。

表1 软件提及和引用特征编码框架

类别	编码	定义说明	示例/解释
提及	论文号	人工分配的论文编号	XU1、XU2、XU3……
	软件名	软件的名称	“VOSviewer是一款专门设计用于构建可视化知识图谱的软件，适合于构建大型复杂网络图谱”中的VOSviewer。
	软件位置	软件在论文中出现的位置	软件名称出现在论文的标题、摘要、关键词或正文时，对其位置进行记录。
引用	软件引用	提及的软件有参考文献标注	“本文使用知识图谱软件VOSviewer[22]分析我国图书情报学研究的热点和趋势”中的VOSviewer软件获得引用。
	引用出版物	引用论文、图书等正式出版物	“本文使用知识图谱工具VOSviewer[22]分析我国图书情报学研究的热点和趋势。”若标注[22]对应的是“Van Eck, N. J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. Scientometrics, 84(2), 523-538.”，则表示引用的是出版物。
	引用网址	引用软件的存储地址	同上，若标注[22]对应的是“Van Eck, N. J., & Waltman, L. VOSviewer. [2016-01-15]. http://www.vosviewer.com”，则表示引用的是网址。
	引用手册/指南	引用软件的使用指南、手册	同上，若标注[22]对应的是“Van Eck, N. J., & Waltman, L. VOSviewer manual. [2016-01-15]. http://www.vosviewer.com/download/f-z2w2.pdf”，则表示引用的是手册。

3.3 软件扩散测度指标

根据Liu和Rousseau[19]、宋歌[15]提出的知识创新扩散测度指标，本文提出如下指标来测度软件的学术影响力、描述软件扩散的基本特征：

（1）扩散广度（diffusion breadth, db）：分为论文级、期刊级和学科级扩散广度三个指标。其中，论文扩散广度（paper diffusion breadth, pdb）指提及软件的论文数，期刊扩散广度（journal diffusion breadth, jdb）指提及软件的论文发表的期刊数，学科扩散广度（domain diffusion breadth, ddb）指提及软件的论文发表的期刊所属的学科数。

（2）扩散时间（diffusion time, dt）：统计测度时间与软件发布时间差。例如，软件Sci2发布于2009年，若在2010年进行统计测度，其扩散时间值为2。

（3）扩散速度（diffusion speed, ds）：分为论文级、期刊级和学科级平均扩散速度三个指标。其中，论文扩散速度（average diffusion speed over papers）即为pdb/dt。类似的，期刊扩散速度（average diffusion speed over journals）和学科扩散速度（average diffusion speed over domains）分别为jdb/dt、ddb/dt。

（4）扩散加速度（diffusion acceleration, da）：扩散速度变化量与发生这一变化所用时间的比值。例如，软件Sci2在2010和2011年的论文级扩散速度分别为ds2010和ds2011，那么Sci2在2011年的扩散加速度da2011=(ds2011- ds2011)/(2011-2010)。

（5）扩散延时（diffusion delay, dd）：某软件出现在某学科领域的时间与该软件发布时间之间的时间差。

4 结果与讨论

4.1 知识图谱研究中的软件提及和引用分析

对软件引用数据集中的1804篇论文的标注结果进行统计，发现共有1548篇论文提及了软件，占比为85.8%，远高于我们之前研究发现的图书情报学领域13.9%的使用软件论文占比，也高于生物学领域65.6%的使用软件论文占比[10]。图1显示了我国知识图谱研究论文数量和提及软件论文占比的逐年变化情况。从中可以看出，我国知识图谱研究论文数量逐年递增，近两年增长迅猛，并且在2006～2017年间提及软件的论文占比总体呈上升趋势，论文占比已经从2006年的23.5%上升到2017年的92.5%，增幅显著。在2017年的399篇知识图谱研究论文中，仅有30篇未提及软件。在这30篇论文中，还有4篇实际上使用软件进行了可视化分析却未在论文中提及使用的软件。可见，知识图谱研究对软件的依赖程度非常高，从事知识图

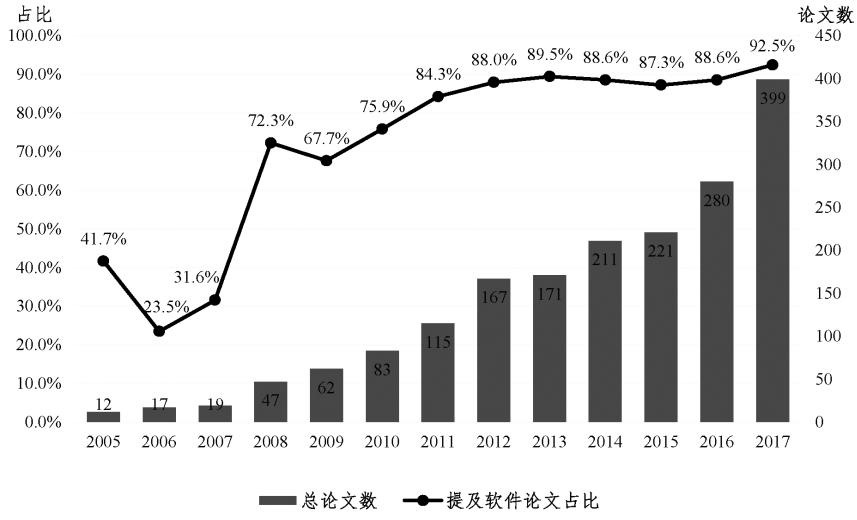

图1 提及软件论文占比的逐年变化情况

谱研究的新人有必要加强相关软件使用和开发的学习。与此同时,我们还对1548篇知识图谱研究论文提及软件的位置进行分析,发现36.4%的论文在文章主题(题名、关键词和摘要)位置提及软件。出现在论文主题位置的词往往比正文中的词对论文更为重要[57],超过35%的论文在主题位置提及软件在一定程度上说明软件工具对知识图谱研究的重要性。

软件引用不仅可以帮助读者定位和获取软件,有利于促进软件共享和再利用,还可以增加软件研发者的被认可度,便于对软件研发者的贡献进行科学评价[58]。对软件引用数据集中的软件引用情况进行统计分析,发现1804篇论文中提及软件3127次,只有18%标注了参考文献,软件平均引用缺失率高达0.82。在这种情况下,用被引次数作为测度软件影响力的唯一指标并不恰当。图2展示了2005~2017年13年间的软件引用缺失率变化情况。从图2可以看出,软件引用缺失率在0.6~1.0之间浮动,无明显下降趋势。可见,虽然国内知识图谱研究论文量自2005年以来增长显著,但软件引用缺失状况却无明显改善。

我们还对软件引用的参考文献类型进行分析,发现我国知识图谱领域研究者更倾向于引用软件相关出版物,引用比例高达85.4%,远高于软件网址(14%)和指南/手册(0.6%),这或许与学术界有引用出版物的传统有关。与此同时,我们还对软件引用缺失率与软件出现的位置之间的关系进行研究。我们先将提及软件的1548篇论文分为"主题位置"提及和"未在主题位置"提及软件两组,再利用SPSS 20.0[59]对上述两组论文的软件引用缺失率进行分析。其中"主题位置"提及软件组的软件引用缺失率为0.58,"未在主题位置"提及软件组的软件引用缺失率为0.80。运用卡方检验进行组间差异比较,卡方值=11.314,P值=0.001<0.05,两组的软件引用率有显著差异,说明主题中出现的软件更易获得正式引用。

4.2 知识图谱软件扩散分析

本文还基于软件扩散数据集测度10种高频知识图谱软件的扩散延时、扩散广度、扩散速度和扩散加速度,以探索分析软件在学术交流系统中的扩散特征和模式。

4.2.1 软件扩散延时

下页表2列出了高频使用的10个知识图谱软件的发布时间、最早出现在CSSCI/CSCD期刊论文中的时间和扩散延时。从中可以看出,2000年之前研发的国外

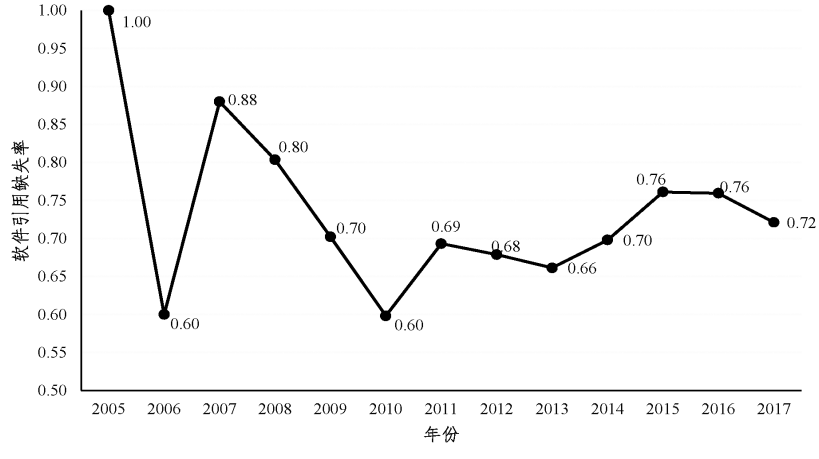

图2 软件引用缺失率变化趋势

表2　知识图谱软件的扩散延时

序号	软件名称	发布时间	出现时间	扩散延时
1	Bibexcel	1992	2006	14
2	Pajek	1996	2005	9
3	Netdraw	1997	2007	10
4	Ucinet	2000	2006	6
5	HistCite	2001	2005	4
6	CiteSpace	2004	2007	3
7	TDA	2005	2008	3
8	Bicomb	2010	2010	0
9	VOSviewer	2010	2011	1
10	SATI	2012	2012	0

知识图谱软件Bibexcel、Netdraw和Pajek均经过了10年左右的时间才被国内学者应用到自己的研究中；2000～2009年间发布的国外软件Ucinet、HistCite、CiteSpace和TDA的扩散延时缩短到3～6年；2010年及其之后发布的软件VOSviewer、Bicomb和SATI的扩散延时进一步缩短到1年甚至1年以内。可见，国内知识图谱研究者应用新软件的速度在逐渐加快。我们还发现，在扩散延时最短的三个软件中，Bicomb和SATI这两个国产软件的扩散延时短于国外软件VOSviewer。

4.2.2　软件扩散广度

图3展示了10个知识图谱软件在2005～2017年间的论文扩散广度。从图中可以看出，10个知识图谱软件的论文扩散广度均在不断增长，增幅十分显著。其中，CiteSpace的论文扩散广度已从2007年的2篇增长到2017年的1274篇，增长了636倍，增幅远高于其他9个软件。CiteSpace的累积使用频次最高，这或许是因为该软件自身的功能较为完善，既可以处理Web of Science、Scopus、Pubmed、NSF、Derwent等外文数据库的数据，也可以处理中文数据库CSSCI和CNKI的数据。除此之外，其开发者陈超美在中国的宣传培训以

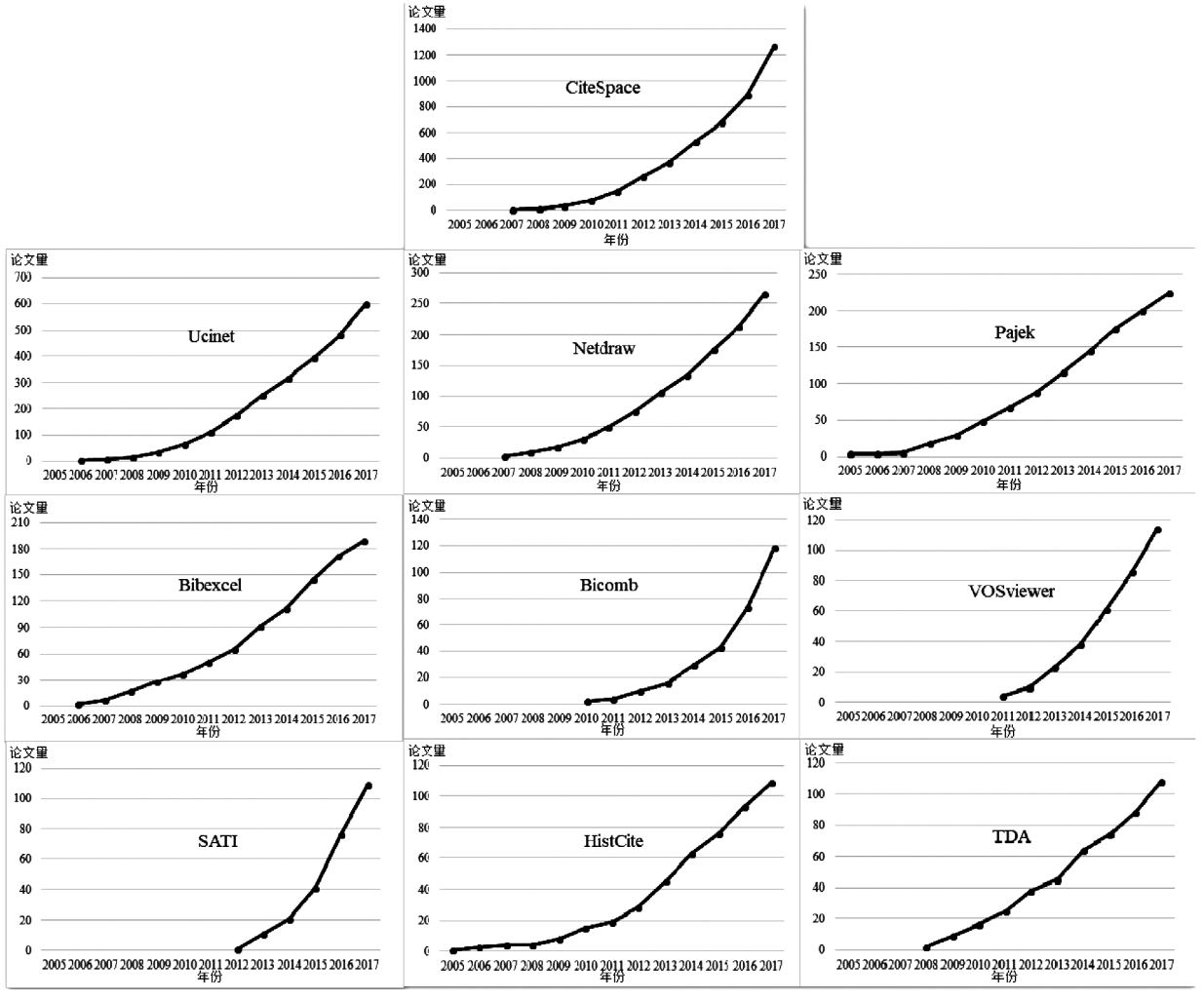

图3　知识图谱软件的论文扩散广度

表3 知识图谱软件的期刊扩散广度

软件名称＼年份	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
CiteSpace			2	8	16	29	51	76	103	135	161	211	307
Ucinet		2	5	13	20	35	57	75	94	113	128	155	199
Netdraw			3	8	13	21	31	42	55	64	77	89	106
Pajek	3	3	3	9	12	20	26	36	41	46	56	68	80
Bibexcel		2	5	13	16	20	23	29	39	47	59	71	80
Bicomb						2	3	8	12	23	31	46	76
VOSviewer							4	9	14	20	27	36	49
SATI								1	7	15	24	33	53
HistCite	1	3	4	4	8	13	14	17	25	31	35	40	50
TDA				2	8	15	22	27	30	40	43	46	57

及其拥有较多中文版本的使用手册/指南可能也使得它比其他国外知识图谱软件在国内获得更为广泛的使用[30]。HistCite虽然较早出现在国内知识图谱研究中，但在13年间仅有103篇论文提及该软件，少于SATI和TDA之外的其他软件。这或许与HistCite功能较为简单且只能处理Web of Science数据库数据有关[30]。此外，从图中还可以看出，到2017年，CiteSpace、Ucinet、Netdraw、Pajek和Bibexcel均已被150篇以上的论文使用，而Bicomb、VOSviewer、SATI、HistCite和TDA的累积使用频次均不足120。

同时，我们对10个知识图谱软件在2005～2017年间的期刊扩散广度也进行了计算。由表3可知，10个知识图谱软件的期刊扩散广度呈明显增长趋势。其中，CiteSpace的期刊扩散广度增幅最大，VOSviewer的期刊扩散广度增幅最小。从表3还能发现，到2017年，VOSviewer、SATI、HistCite和TDA的期刊扩散广度小于其他6个软件。此外，计算结果显示，这些知识图谱软件的学科扩散广度也呈现明显增长趋势。其中，CiteSpace的增长尤为显著，其学科扩散广度已从2007年的2个增长到2017年的40个，增幅超过其他9个软件。而Bicomb、VOSviewer、SATI和HistCite的学科扩散广度分别为17、18、14和17个，小于其他6个软件(均大于18)。

我们还对这10个知识图谱软件在学科领域内的分布情况进行分析。结果发现，这些软件频繁出现在图书馆、情报与文献学、教育学和管理学研究领域，而在其他诸如宗教学、艺术学、历史学、社会学、心理学、医学、工程学等领域出现频次较低。其中，Bicomb在教育学领域获得最多使用，其他9个软件在图书馆、情报与文献学领域出现最为频繁。同时，我们还对较早应用这些知识图谱软件的学科领域进行分析。从表4可以看出，除了SATI和HistCite最早出现在管理学领域，其他8个软件均于其出现在CSSCI/CSCD期刊论文中的最早年份就被应用于图书馆、情报与文献学研究。此外，对HistCite、CiteSpace和VOSviewer这三个软件的主要技术论文在Web of Science中的被引频次进行统计发现，它们获得最多引用的技术论文均发表在图书情报学期刊上。同时，鉴于这三个软件的开发者都是图书情报学专家，可以认为这三个软件起源于图书情报学领域。在此基础上可以看出，这些软件更有可能被频繁用于其来源领域，且它们的扩散过程呈现从来源领域扩散到相近领域再到更多不同领域的特征。这些研究发现与先前的出版物扩散研究结果[48-49]类似。可

表4 较早出现知识图谱软件的学科领域

软件名称	较早出现知识图谱软件的学科领域				
	A	B	C	D	E
CiteSpace	图书馆、情报与文献学/管理学	哲学/新闻学与传播学	教育学	高校综合性学报/体育学	普通内科医学/经济学
Ucinet	图书馆、情报与文献学/新闻学与传播学	高校综合性学报/计算机科学	教育学/管理学	政治学	经济学/遥感学/环境科学
Netdraw	图书馆、情报与文献学/计算机科学/管理学	教育学/经济学	公共、环境与职业健康	运输学/地质学/体育学	环境科学/工程学
Pajek	图书馆、情报与文献学/新闻学与传播学	计算机科学	管理学	高校综合性学报/教育学	环境科学
Bibexcel	图书馆、情报与文献学	教育学/管理学	高校综合性学报/新闻学与传播学	哲学	农业科学/眼科学/人文经济地理
Bicomb	图书馆、情报与文献学	高校综合性学报/普通内科医学	教育学/人文经济地理	管理学	儿科学/神经科学
VOSviewer	图书馆、情报与文献学/普通内科医学	人文经济地理	高校综合性学报/地质学/公共、环境与职业健康/新闻学与传播学	教育学	民族学与文化学/农业科学/经济学
SATI	管理学	图书馆、情报与文献学/教育学	新闻学与传播学	农业科学/经济学	科学技术/矿物学/体育学
HistCite	管理学	图书馆、情报与文献学	新闻学与传播学	普通内科医学/高校综合性学报	农业科学/植物科学/教育学/人文经济地理
TDA	图书馆、情报与文献学	物理学/工程学	普通内科医学/环境科学/遥感学/新闻学与传播学	地质学/矿物学	高校综合性学报

注：A栏代表最早出现相关软件的学科领域，B栏晚于A栏，C栏晚于B栏，D、E栏同理；在单元格中以/来划分同一年出现的多个学科。

见，在学术交流系统中，基于使用的软件扩散过程类似于基于引用的知识扩散过程。

4.2.3 软件扩散速度和加速度

表5列出了10个知识图谱软件的论文扩散速度和扩散加速度，括号内的数值表示扩散加速度。由表5可知，10个知识图谱软件的扩散速度一开始都比较缓慢，然后随着时间的推移不断增加，到2017年达到历史最高值。由于10个软件近五年的扩散加速度均大于0且无明显下降趋势，可以预见这些软件将被更多的研究者接受和使用，它们的扩散速度在未来一段时间内还将继续增长。同时还发现，Pajek、HistCite、TDA、Bibexcel、VOSviewer和SATI的论文扩散加速度分别在2010年、2013年、2014年、2015年和2016年达到最大值，此后各软件扩散加速度均有所放缓。在这10个软件中，CiteSpace、Ucinet、SATI、Bicomb 和 VOSviewer 在 2017 年的论文扩散速度和扩散加速度均高于其他5个软件。其中，SATI、Bicomb和VOSviewer虽然晚于其他软件出现，但是它们的扩散速度在短短几年内已经赶上并超过多个早期同类软件。可见，国内知识图谱研究人员对新软件的接受和使用速度不断加快。这一结论可以为知识图谱研究人员，特别是新进入的研究人员选择相关软件提供参考和帮助。

此外，我们还对10个知识图谱软件期刊级和学科级的扩散速度、扩散加速度进行了计算。结果显示，10个软件的期刊扩散速度总体均呈上升趋势，2017年达到最高值。其中，CiteSpace、Ucinet、Bicomb、SATI 和 VOSviewer 在 2017 年的期刊扩散速度分别为 21.9、11.1、9.5、8.8 和 6.1，高于其他5个软件。然而，这10个软件在期刊扩散加速度上并未呈现上升趋势。在学科级扩散速度方面，CiteSpace、Ucinet、Pajek、Bibexcel 和 HistCite 均呈总体上升趋势，在2017年分别达到2.9、1.8、0.9、0.7和1.0；Bicomb、VOSviewer和SATI未呈总体上升趋势，但在2017年达到最大速度2.1、2.3和2.3；TDA和Netdraw分别在2011年和2016年达到最大速度1.7和1.3。同时还发现，CiteSpace、Bicomb、VOSviewer和SATI近两年的学科扩散加速度均大于0.2，可以预见上述4个软件将被应用到更多的学科领域中。

表5　知识图谱软件的论文扩散速度和加速度

软件名称	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
CiteSpace			0.5 (0)	2.2 (1.7)	5.7 (3.5)	10.6 (4.9)	18.3 (7.7)	28.8 (10.5)	36.9 (8.1)	48.0 (11.1)	56.4 (8.4)	68.8 (12.4)	91.0 (22.2)
Ucinet		0.3 (0)	0.6 (0.3)	1.8 (1.2)	3.5 (1.7)	5.9 (2.4)	9.4 (3.5)	13.2 (3.8)	17.9 (4.7)	21.2 (3.3)	24.6 (3.4)	28.5 (3.9)	33.3 (4.8)
Netdraw			0.3 (0)	0.8 (0.5)	1.4 (0.6)	2.2 (0.8)	3.2 (1)	4.7 (1.5)	6.2 (1.5)	7.4 (1.2)	9.2 (1.8)	10.7 (1.5)	12.6 (1.9)
Pajek	0.3 (0)	0.3 (0)	0.4 (0.1)	1.3 (0.9)	2.0 (0.7)	3.2 (1.2)	4.3 (1.1)	5.2 (0.9)	6.4 (1.2)	7.6 (1.2)	8.8 (1.2)	9.6 (0.8)	10.2 (0.6)
Bibexcel		0.1 (0)	0.4 (0.3)	1.0 (0.6)	1.6 (0.6)	1.9 (0.3)	2.5 (0.6)	3.1 (0.6)	4.1(1)	4.9 (0.8)	6.0 (1.1)	6.8 (0.8)	7.3 (0.5)
Bicomb						2.0 (0)	2.0 (0)	3.3 (1.3)	4.0 (0.7)	5.8 (1.8)	7.2 (1.4)	10.4 (3.2)	14.8 (4.4)
VOSviewer							2.0 (0)	3.3 (1.3)	5.8 (2.5)	7.6 (1.8)	10.2 (2.6)	12.3 (2.1)	14.3 (2)
SATI								1.0 (0)	5.5 (4.5)	7.0 (1.5)	10.3 (3.3)	15.2 (4.9)	18.2 (3)
HistCite	0.2 (0)	0.5 (0.3)	0.6 (0.1)	0.5 (−0.1)	0.9 (0.4)	1.5 (0.6)	1.7 (0.2)	2.3 (0.6)	3.5 (1.2)	4.5 (1)	5.1 (0.6)	5.8 (0.7)	6.4 (0.8)
TDA				0.5 (0)	1.8 (1.3)	2.7 (0.9)	3.6 (0.9)	4.8 (1.2)	5.0 (0.2)	6.4 (1.4)	6.7 (0.3)	7.3 (0.6)	8.3 (1)

注释：括号内数值代表软件扩散加速度。

5　结论与展望

本文先采用内容分析法对国内知识图谱研究论文中的软件使用和引用情况进行深入细致分析，再利用知识扩散测度指标对论文中发现的高频使用软件在学术交流系统中的扩散特征和模式进行探索性研究。结果显示，国内知识图谱研究对软件的依赖程度不断提升，2017年提及软件的论文比例高达92.5%，然而这并未带来软件引用缺失状况的改善，软件引用缺失依然严重。这可能是因为国内学术界和出版界尚未认识到软件引用的重要性、目前国内尚缺乏统一的软件引用规范。研究还发现，知识图谱研究人员接受和使用新软件的速度在不断加快，在知识图谱研究之初，国内学者对国外知识图谱软件接受程度较低，知识图谱软件的扩散比较缓慢，随着时间的推移扩散速度不断加快，目前仍未见扩散放缓。依据创新扩散理论可以预见，出现较晚的知识图谱软件（如CiteSpace、VOSviewer、SATI等）将被更多的研究者快速接受和使用，直到它们被知识图谱研究领域中的大部分研究者熟知和使用，之后扩散将放缓，逐渐趋于饱和状态。此外，我们还发现，一些知识图谱软件在学术交流系统中的扩散过程呈现出由其来源领域（图书馆、情报与文献学）扩散到相近领域再到其他差异较大的学科领域的特征。同时，这些知识图谱软

件往往在其来源领域获得最为广泛的应用。上述结果表明，学术交流系统中基于使用的软件扩散过程与该系统中基于引用的知识扩散过程类似。

需要说明的是，一些CSCD期刊的学科类别具有多学科交叉的特征，本文将这些期刊分到该刊大部分论文所属的类别（而CSSCI期刊都有唯一的学科分类）。上述分类操作虽然可能会造成软件扩散的学科级测度指标值偏离实际值，但因为样本中这类论文占比较少，研究结果与实际偏差不大。

本研究为测度软件影响力、探索软件扩散特征提供了新的思路和方法，其研究结论一方面进一步证实了软件在科学研究中发挥着重要作用，但研究者在论文中提及和引用软件时却表现出很大的随意性；另一方面有助于我们理解软件在学术交流生态系统中的地位和扩散特征，有助于我们更为完整地认识知识生产和扩散过程。针对国内软件引用缺失严重、尚无统一的软件引用规范的现状，我们有必要通过建立有效的软件引用机制、根据国外已有的软件引用规范制定我国的软件引用标准、培养国内学者软件引用意识来推进软件引用规范化、促进软件传播和共享。在软件识别与扩散方面，还有很多有趣的问题有待进一步探索和研究。比如，如何利用机器学习方法高效地从全文本数据中自动识别出软件实体？软件扩散过程是否呈S型曲线特征？软件在其来源领域和其他领域内的扩散速度是否存在差异？不同类型软件（如商业软件和开源软件、通用软件和专业软件）的扩散特征是否有所不同？为什么一些软件的扩散广度和扩散速度能在同类软件中保持长时间的优势？这些问题的解决，将有助于我们深入理解知识生产和扩散过程，有助于提高学术交流和科研活动的效率。

致谢：感谢崔明、徐潇洁和丁笑舒三位同学为本研究标注部分数据。

参考文献

[1] Li K, Yan E, Feng Y. How is R cited in research outputs? Structure, impacts, and citation standard[J]. Journal of Informetrics, 2017, 11(4):989-1002.

[2] Pan X, Yan E, Hua W. Disciplinary differences of software use and impact in scientific literature[J]. Scientometrics, 2016, 109(3):1-18.

[3] Hettrick S, Antonioletti M, Carr L, et al. UK research software survey 2014[EB/OL].[2017- 10- 16].https://zenodo.org/record/14809.

[4] Hanney S, Frame I, Grant J, et al. Using categorisations of citations when assessing the outcomes from health research[J]. Scientometrics, 2005, 65(3): 357-379.

[5] Prabhu P, Kim H, Oh T, et al. A survey of the practice of computational science[C].High Performance Computing, Networking, Storage and Analysis (SC), 2011 International Conference for. IEEE, 2011: 1-12.

[6] Lou H, Kirkpatrick A E. Assessing open source software as a scholarly contribution[J]. Communications of the Acm, 2009, 52(12):126-129.

[7] Piwowar H. Altmetrics: value all research products[J]. Nature, 2013, 493(7431):159.

[8] NSF. GPG Summary of Changes[EB/OL].[2017-10-16].https://www.nsf.gov/pubs/policydocs/ pappguide/nsf13001/gpg_sigchanges.jsp.

[9] Research Excellence Framework. Output Information Requirements[EB/OL]. [2017- 10- 16]. http://www.ref.ac.uk/about/guidance/submittingresearchoutputs/

[10] Howison J, Bullard J. Software in the scientific literature: problems with seeing, finding, and using software mentioned in the biology literature[J]. Journal of the Association for Information Science & Technology, 2015, 67(9):2137-2155.

[11] Smith A M, Katz D S, Niemeyer K E. Software citation principles[J]. PeerJ Computer Science, 2016, 2: e86.

[12] Pan X, Yan E, Wang Q, et al. Assessing the impact of software on science: a bootstrapped learning of software entities in full-text papers[J]. Journal of Informetrics, 2015, 9(4): 860-871.

[13] Thelwall M, Kousha K. Academic software downloads from Google code: useful usage indicators?[J]. Information Research: An International Electronic Journal, 2016, 21(1): n1.

[14] Liu Y, Rousseau R. Towards a representation of diffusion and interaction of scientific ideas: the case of fiber optics communication[J]. Information Processing & Management, 2012, 48(4): 791-801.

[15] 宋 歌. 学术创新的扩散过程研究[J]. 中国图书馆学报, 2015, 41(1):62-75.

[16] Rousseau R. Robert Fairthorne and the empirical power laws [J]. Journal of Documentation, 2005, 61(2): 194-202.

[17] Zhao R, Wu S. The network pattern of journal knowledge transfer in library and information science in China[J]. Knowledge Organization, 2014, 41(4):276-287.

[18] Yan E. Disciplinary knowledge production and diffusion in science[J]. Journal of the Association for Information Science and Technology, 2016, 67(9): 2223-2245.

[19] Liu Y, Rousseau R. Knowledge diffusion through publications

and citations: a case study using ESI-fields as unit of diffusion [J]. Journal of the Association for Information Science and Technology, 2010, 61(2): 340-351.

[20] Nomaler Ö, Verspagen B. Knowledge flows, patent citations and the impact of science on technology[J]. Economic Systems Research, 2008, 20(4): 339-366.

[21] Milojević S, Sugimoto C R, Larivière V, et al. The role of handbooks in knowledge creation and diffusion: a case of science and technology studies[J]. Journal of Informetrics, 2014, 8(3): 693-709.

[22] Yu Q, Ding Y, Song M, et al. Tracing database usage: detecting main paths in database link networks[J]. Journal of Informetrics, 2015, 9(1): 1-15.

[23] B?rner K, Chen C, Boyack K W. Visualizing knowledge domains[J]. Annual Review of Information Science and Technology, 2003, 37(1): 179-255.

[24] Cobo M J, López-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 & Technology, 2011, 62(7):1382-1402.

[25] 陈 悦, 刘则渊. 悄然兴起的科学知识图谱[J]. 科学学研究, 2005, 23(2):149-154.

[26] 陈 悦, 刘则渊, 陈 劲,等. 科学知识图谱的发展历程[J]. 科学学研究, 2008, 26(3):449-460.

[27] 秦长江, 侯汉清. 知识图谱——信息管理与知识管理的新领域[J]. 大学图书馆学报, 2009, 27(1):30-37.

[28] 杨思洛, 韩瑞珍. 国外知识图谱绘制的方法与工具分析[J]. 图书情报知识, 2012(6):101-109.

[29] 刘启元, 叶 鹰. 文献题录信息挖掘技术方法及其软件SATI的实现——以中外图书情报学为例[J]. 信息资源管理学报, 2012(1):50-58.

[30] 肖 明, 邱小花, 黄 界,等. 知识图谱工具比较研究[J]. 图书馆杂志, 2013, 32(3):61-69.

[31] 侯剑华, 刘则渊. 纳米技术研究前沿及其演化的可视化分析[J]. 科学学与科学技术管理, 2009, 30(5):23-30.

[32] 刘则渊, 陈立新. 国际物流研究领域的知识可视化分析[J]. 图书情报工作, 2010, 54(8):140-143.

[33] 刘则渊, 张春博. 中国科学技术方法论研究三十年回顾——基于期刊文献的科学计量分析[J]. 科学技术哲学研究, 2014 (4):86-93.

[34] 宗乾进, 袁勤俭, 沈洪洲. 2001—2010年我国图书馆学研究知识图谱——基于知识图谱的当代学科发展动向研究[J]. 国家图书馆学刊, 2012, 21(2):84-91.

[35] 宗乾进, 袁勤俭, 沈洪洲,等. 2001—2010年国内情报学研究回顾与展望——基于知识图谱的当代学科发展动向研究[J]. 情报资料工作, 2012, 33(1):10-15.

[36] 宗乾进, 袁勤俭. 回顾与展望:近十年我国档案学研究全景透视[J]. 档案学通讯, 2012(2):12-16.

[37] Howison J, Deelman E, McLennan M J, et al. Understanding the scientific software ecosystem and its impact: current and future measures[J]. Research Evaluation, 2015, 24(4):454-470.

[38] Trainer E H, Chaihirunkarn C, Kalyanasundaram A, et al. From personal tool to community resource: what´s the extra work and who will do it?[C]// ACM Conference on Computer Supported Cooperative Work & Social Computing. ACM, 2015.

[39] 潘雪莲. 软件实体的自动抽取和学术影响力研究[D]. 南京:南京大学, 2016.

[40] Huang X, Ding X, Lee C P, et al. Meanings and boundaries of scientific software sharing[C] .Conference on Computer Supported Cooperative Work. ACM, 2013:423-434.

[41] Joppa L N, Mcinerny G, Harper R, et al. Troubling trends in scientific software use[J]. Science, 2013, 340(6134):814-815.

[42] Niemeyer K E, Smith A M, Katz D S. The challenge and promise of software citation for credit, identification, discovery, and reuse[J]. Journal of Data and Information Quality (JDIQ), 2016, 7(4): 16-20.

[43] Ding Y, Song M, Han J, et al. Entitymetrics: measuring the impact of entities[J]. Plos One, 2013, 8(8):e71416.

[44] 王 雪, 马胜利, 佘曾溧, 等. 科学数据的引用行为及其影响力研究[J]. 情报学报, 2016, 35(11): 1132-1139.

[45] 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.

[46] Rogers E M. Diffusion of Innovations[M]. New York: Free Press, 1962.

[47] Rogers E M.Diffusion of Innovations[M].5th New York. Free Press, 2003.

[48] Liu Y, Rousseau R. Towards a representation of diffusion and interaction of scientific ideas: the case of fiber optics communication[J]. Information Processing & Management, 2012, 48(4): 791-801.

[49] Van Leeuwen T, Tijssen R. Interdisciplinary dynamics of modern science: analysis of cross-disciplinary citation flows[J]. Research Evaluation, 2000, 9(3): 183-187.

[50] Rinia E J, Van Leeuwen T N, Bruins E E W, et al. Citation delay in interdisciplinary knowledge exchange[J]. Scientometrics, 2001, 51(1): 293-309.

[51] Faber Frandsen T, Rousseau R, Rowlands I. Diffusion factors

[J]. Journal of Documentation, 2006, 62(1): 58-72.

[52] Chen C. Science mapping: a systematic review of the literature [J]. Journal of Data and Information Science, 2017, 2(2): 1-40.

[53] 曹树金, 吴育冰, 韦景竹, 等. 知识图谱研究的脉络, 流派与趋势——基于 SSCI 与 CSSCI 期刊论文的计量与可视化[J]. 中国图书馆学报, 2015, 41(5): 16-34.

[54] Ford J M. Content analysis: an introduction to its methodology [J]. Personnel Psychology, 2004, 57(4): 1110-1113.

[55] Chu H. Research methods in library and information science: a content analysis[J]. Library & Information Science Research, 2015, 37(1): 36-41.

[56] Freelon D G. ReCal: intercoder reliability calculation as a web service[J]. International Journal of Internet Science, 2010, 5 (1): 20-33.

[57] Zhou B, Pei J, Tang Z. A spamicity approach to web spam detection[C] .Proceedings of the 2008 SIAM International Conference on Data Mining. Society for Industrial and Applied Mathematics, 2008: 277-288.

[58] Soito L, Hwang L J. Citations for software: providing identification, access and recognition for research software[J]. International Journal of Digital Curation, 2016, 11(2): 48-63.

[59] SPSS 20.0. IBM SPSS Software[EB/OL]. [2017-10-16]. https://www.ibm.com/analytics/us/en/technology/spss/

[作者简介]于晓彤，女，1994年生，南京大学信息管理学院硕士研究生。

潘雪莲，女，1983年生，南京大学信息管理学院助理研究员。

华薇娜，女，1955年生，南京大学信息管理学院教授，博士生导师。

收稿日期：2018-09-06