文章编号：2095-2163(2023)05-0032-08 中图分类号：TP399 文献标志码：A

知识图谱在医学领域的研究现状分析

郑增亮[1]，蔡晓琼[1]，苏前敏[1]，黄继汉[2]
（1 上海工程技术大学 电子电气工程学院，上海 201620；2 上海中医药大学 药物临床研究中心，上海 201203）

摘 要：本文针对国内外知识图谱在医学领域的研究进行可视化分析，对比国内和国外研究的热点和异同，以期推动中国知识图谱在医学领域的研究。以 CNKI 和 Web Of Science 上刊载的“知识图谱在医学领域研究”主题相关核心文献作为数据来源，运用 CiteSpace 可视化软件进行文献计量分析。从时间序列上看，知识图谱在医学领域的研究已引起国内外学者的广泛关注，该领域的发文量随着时间推移，呈现不断增长的趋势。新的方法、技术如大数据、人工智能，深度学习不断应用到医学领域的知识图谱中，但国内外知识图谱在医学领域方面的研究侧重点不同，国内侧重于理论研究，国外侧重于实际应用。

关键词：知识图谱；医学领域；可视化

Analysis of the current research status of knowledge graph in the medical field

ZHENG Zengliang[1], CAI Xiaoqiong[1], SU Qianmin[1], HUANG Jihan[2]
(1 College of Electrical and Electronic Engineering, Shanghai University of Engineering Science, Shanghai 201620, China;
2 Center for Drug Clinical Research, Shanghai University of Chinese Medicine, Shanghai 201203, China)

【Abstract】 This paper presents a visual analysis of domestic and foreign research on knowledge graphs in medicine, comparing the hotspots and similarities between domestic and foreign research, with a view to promoting the research on knowledge graphs in medicine in China. The core literature related to the topic of " knowledge graphs in medicine" published in CNKI and Web Of Science was used as the data source, and the bibliometric analysis was conducted using CiteSpace visualization software. In terms of time series, the research on knowledge graphs in medicine has attracted widespread attention from scholars at home and abroad, and the number of articles published in this field has shown a growing trend over time. New methods and technologies such as big data, artificial intelligence, and deep learning have been continuously applied to knowledge mapping in the medical field, but the focus of research on knowledge mapping in the medical field is different at home and abroad, with China focusing on theoretical research and foreign countries on practical applications.

【Key words】 knowledge graph; medicine; visualization

0 引 言

随着移动互联网、物联网、云计算等技术的不断发展，数据的类型和规模以前所未有的速度增长，社会各个领域都步入大数据时代[1]。在医学领域，伴随着医学信息化系统的发展，积累了规模可观的医学大数据，但这些数据并没有发挥应有的价值，如何从巨量复杂的数据中快速提取最有价值的信息，是制约当前医学大数据分析的关键问题[2]。近年来，知识图谱在工业界和学术界都得到了广泛的应用，成为最有效的知识集成方法之一[3]。知识图谱作为一种新型的知识表示形式，可以对错综复杂的文本数据进行有效的加工、处理、整合，转化为简单、清晰的三元组，最后聚合大量的知识，从而实现知识的快速响应和推理。

一个完整的知识图谱的构建需要经历知识建模、知识存储、知识抽取、知识融合、知识计算和知识应用等阶段[4]。近年来，Freebase 和 DBpedia 这样的大型知识图谱在众多下游应用中发挥了重要作用，引发了学术界和工业界的广泛关注。

为了更全面分析知识图谱在医学领域的研究现状和趋势、对比研究热点，本文通过检索 CNKI 和

基金项目：“十三五”国家科技重大专项(2018ZX09711001-009-011)；科技创新 2030 重大项目(2020AAA0109300)。
作者简介：郑增亮(1996-)，男，硕士研究生，主要研究方向：知识图谱、大数据；苏前敏(1974-)，男，博士，副教授，硕士生导师，主要研究方向：生物医学信息处理、智能信息处理。
通讯作者：苏前敏 Email：suqm@sues.edu.cn
收稿日期：2022-05-28

Web Of Science 中 2012~2021 年与知识图谱在医学领域研究主题相关的核心期刊为数据来源，导入 CiteSpace 软件进行文献计量可视化分析，旨在为中国的知识图谱在医学领域的研究提供参考建议。

1　数据与方法

1.1　数据来源

中国知网（CNKI）是目前世界上最大的连续动态更新的学术期刊全文数据库，因此对 CNKI 数据库的学术期刊进行检索。2012 年 5 月 17 日，Google 正式提出了知识图谱（Knowledge Graph）的概念，其初衷是为了优化搜索引擎返回的结果，增强用户搜索质量及体验，2013 年以后开始在学术界和业界普及[5]。故本文高级检索条件设置为：主题=知识图谱，检索时间设置为：2012~2021 年，来源类别设置为：北大核心期刊、CSSCI 期刊及 CSCD 期刊，根据检索结果，继续在检索结果中检索，设置主题="医学"or 主题="医疗"or 主题="疾病"，总计 220 条数据。

以科学引文数据库 Web of Science（WoS）核心合集为数据源，基本检索条件 1 设置为："主题 = Knowledge Graph；文献类型 = Article，Review；语种 = English；自定义年份：2012-01-01 到 2021-12-31"；基本检索条件 2 设置为："主题 = Knowledge Map *"，其余检索条件同条件 1；基本检索条件 3 设置为："主题 = Medical *"，其它条件同条件 1。条件 1 检索到数据 6 019 条，条件 2 检索到数据 25 439 条，条件 3 检索到数据 493 292 条。根据条件 1、2、3 检索的结果进行高级检索，高级检索条件 4：(#1) OR (#2)；高级检索条件 5：(#3) AND #4。高级检索条件 4 检索到数据 30 784 条，高级检索条件 5 检索到数据 1 251 条。由于选择了精确匹配且在检索条件中限定了文献类型，而 WoS 数据库入库时也对文献类型进行了筛选分类，故检索获得的 1 251 篇文献全部纳入本研究。

1.2　研究方法

本文以中国知网（CNKI）和 Web of Science 数据库核心合集收录的相关文献为研究对象，对国内外"知识图谱在医学领域研究"相关文献进行分析探究；利用文献分析工具 CiteSpace 对国内外该领域的研究现状和研究热点进行可视化分析；最后，综合对比国内外该领域研究现状和研究热点，提出相关建议。

1.3　检索结果

截止 2021 年 11 月 6 日，从 CNKI 核心期刊库检索出相关的文献 220 条，国内医学领域应用知识图谱的研究较少，从 Web of Science 核心期刊数据库检索出相关文献 1 251 篇，相对于国内的研究，国外在该领域的研究投入较多。

2　国内知识图谱在医学领域研究现状和热点分析

2.1　发文量

CNKI 检索出该领域研究的学术论文 220 篇，从时间序列上来看，2012~2021 年，国内知识图谱在医学领域研究整体发文量呈增长趋势如图 1 所示。2012~2014 年该领域发文量增长缓慢，原因为国内知识图谱在医学领域研究处于起步阶段；2014 年以后，该领域发文量增长速度较快；2020 年达 53 篇，增长率高达 70.9% 学科领域的发文量在一定程度上可以反映该学科的发展程度和研究水平，该数据表明国内知识图谱在医学领域正处于较快发展阶段，知识图谱研究已引起了相关研究者的关注。

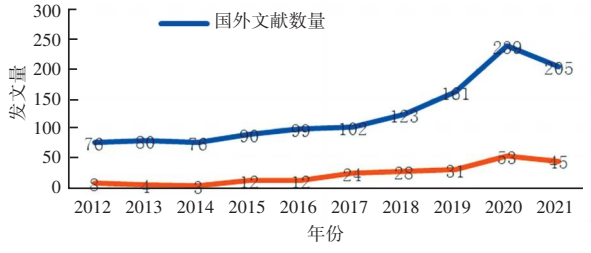

图 1　发文量随时间变化趋势

Fig. 1　Trends in the number of articles published over time

2.2　作者和研究机构分析

对作者和研究机构进行分析，有助于整体把握中国知识图谱在医学领域开展研究的作者和机构分布态势。利用 Citespace 软件进行可视化分析，获得该领域研究者的合作关系如图 2 所示，节点半径越大表示相应发文量越多。

对论文发表的作者进行统计分析见表 1。表中列出了知识图谱在医学领域研究发表论文数量前 10 位的作者。普莱斯定律能够有效评价学者研究成果的影响力，定律指出相同主题中论文数量的一半是由具有较高生产力的作者群体所写，并且作者集合的数量约等于所有作者总数的平方根，计算公式(1)：

$$Mp = 0.749\sqrt{Npmax} \quad (1)$$

式中 $Npmax$ 表示发文量。

按取整原则，发文量在 2 篇或 2 篇以上的论文作者为核心作者。

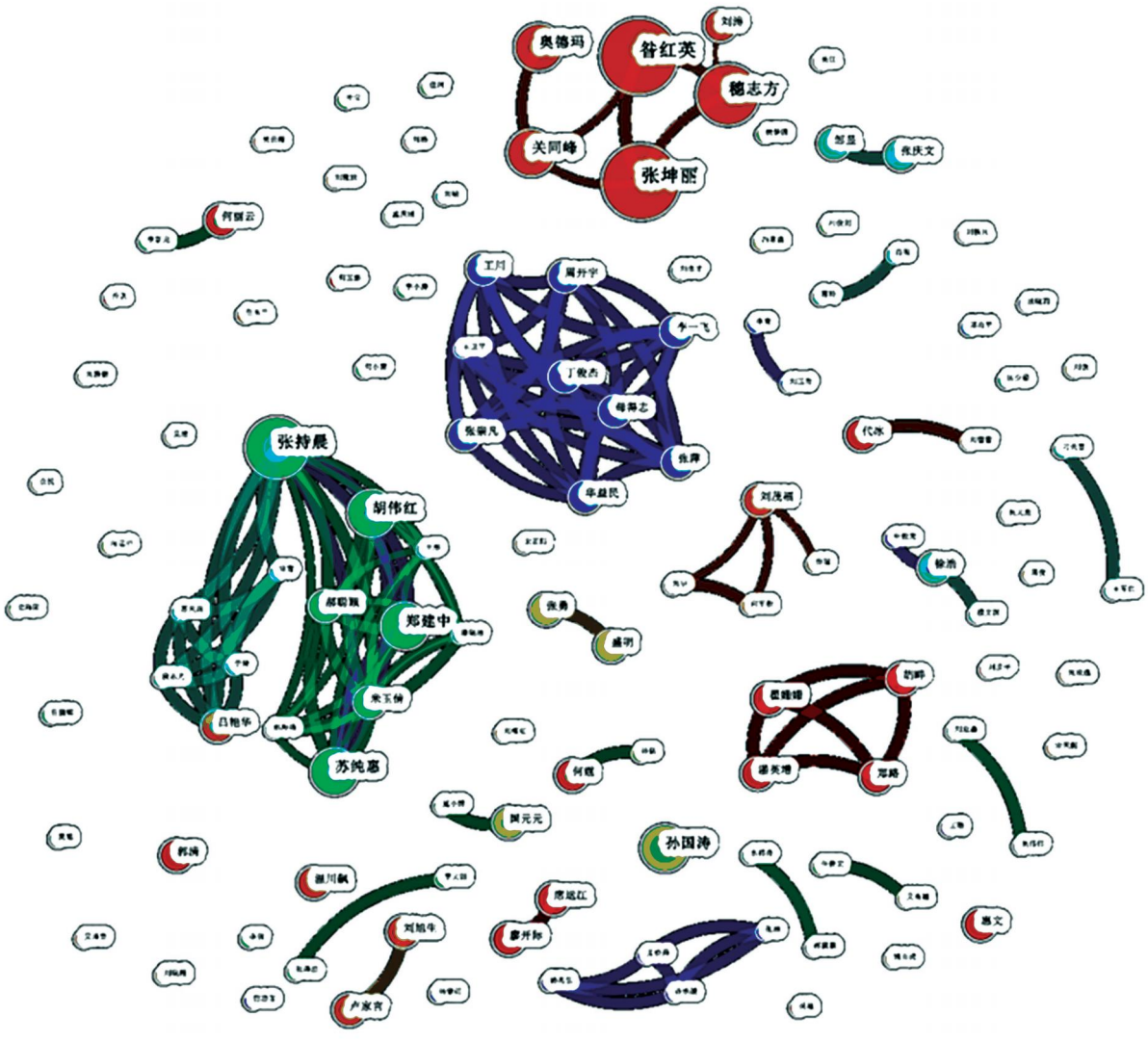

图 2 国内作者合作关系图

Fig. 2 Domestic author partnership chart

表 1 国内作者发文量统计

Tab. 1 Statistics on the number of articles published by domestic authors

发文量(篇)	首次发文时间	作者
5	2020	昝红英
5	2020	张坤丽
4	2020	穗志方
4	2014	张持晨
3	2020	关同峰
3	2017	孙国涛
3	2014	苏纯惠
3	2014	郑建中
3	2019	奥德玛
3	2014	胡伟红

利用 Citespace 进行可视化分析,获得机构合作关系图如图 3 所示,图中节点半径越大表示该机构与其他机构合作次数越多、发文量越多。由图 3 可知,中国知识图谱在医学领域的研究主要集中在高校和研究所,且主要集中于信息情报工程学院和医学院,其中郑州大学信息工程学院和鹏城实验室发文量最多,说明这两所研究机构对知识图谱在医学领域的研究比较重视,而且合作密切,在该领域科研力量强大;其次是中国中医科学院中医临床基础医学研究所、华中科技大学同济医学院医药卫生管理学院、华南理工大学工商管理学院等。

2.3 国内研究热点和研究前沿分析

研究热点和研究前沿常来源于新的科学发现或学科进展,是科学研究中最先进、最有发展潜力的研

究主题或研究领域[6]。关键词词频共现可揭示文献所属领域研究主题的热点分布并揭示其内在联系和演进规律[7]。利用 Citespace 绘制关键词词频共现时序图如图 4 所示，进而展现知识图谱在医学领域研究热点和趋势。时序图节点的大小代表出现频次，频次较多的关键词或名词短语在一定程度上代表该领域的研究热点[8]。关键词时序图中关键词表示该关键词首次出现的时间，字体或节点大小客观反映知识图谱在医学领域研究持续的热度，节点越大说明该方向研究持续的热度越久。

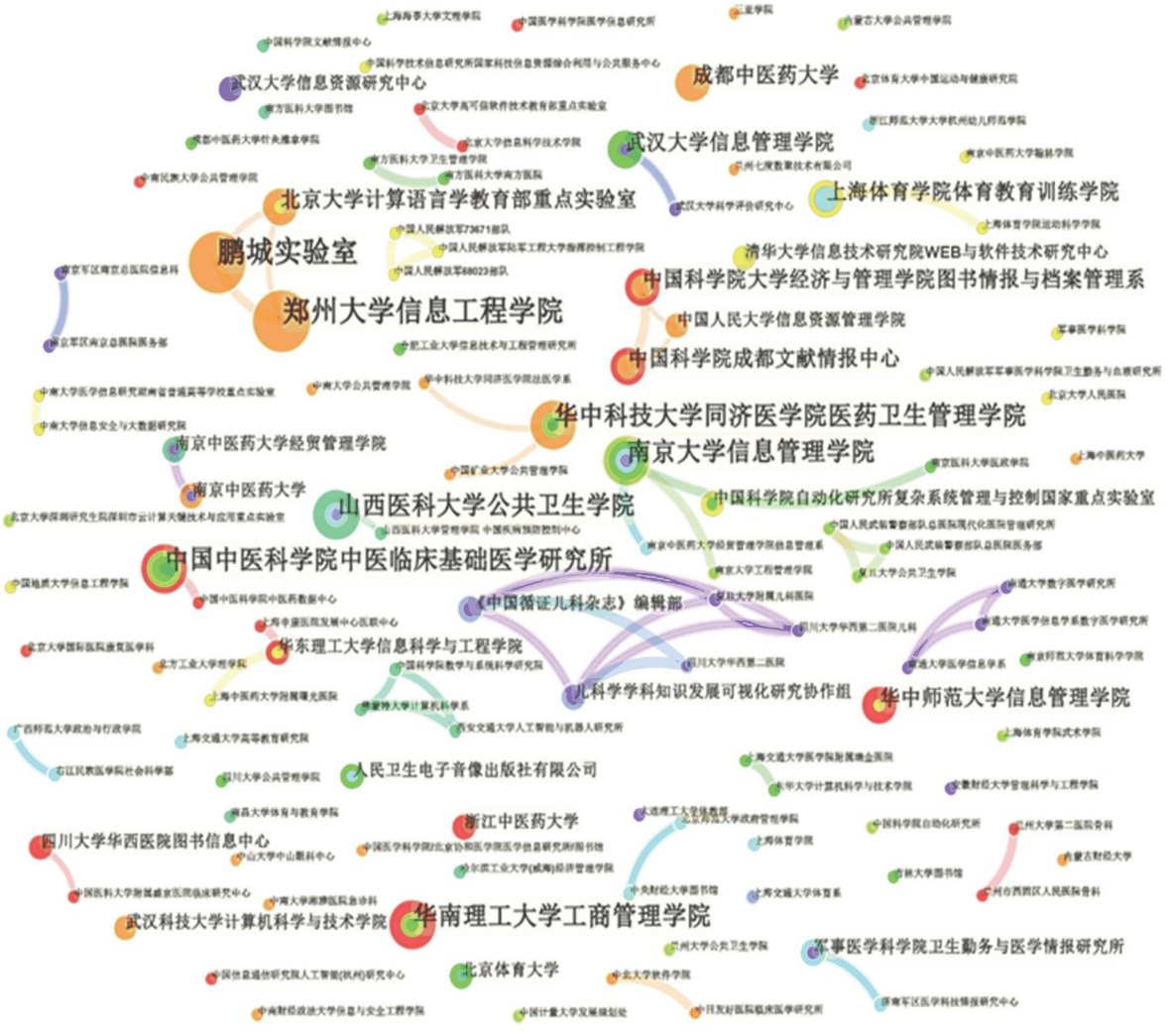

图 3　机构合作关系图谱

Fig. 3　Institution cooperation map

图 4 从左向右时间从 2012 年依次递增，最大的节点是“知识图谱”，表明“知识图谱”热度在 2012 年一直持续；其次是“研究热点”，“可视化”，“共词分析”方面的热度比较持久；在“大数据”、“人工智能”词条出现后，“实体抽取”、“实体关系”和“实体识别”等关键词集中涌现，深度学习也应用于医学领域的知识图谱研究，说明随着前沿技术的应用，医学领域知识图谱的研究有了更深层次的发展；近年来知识图谱开始应用于“医养结合”、“临终关怀”、“养老院”等相关的养老服务，说明养老方向是近年国内医学知识图谱研究的一个趋势。

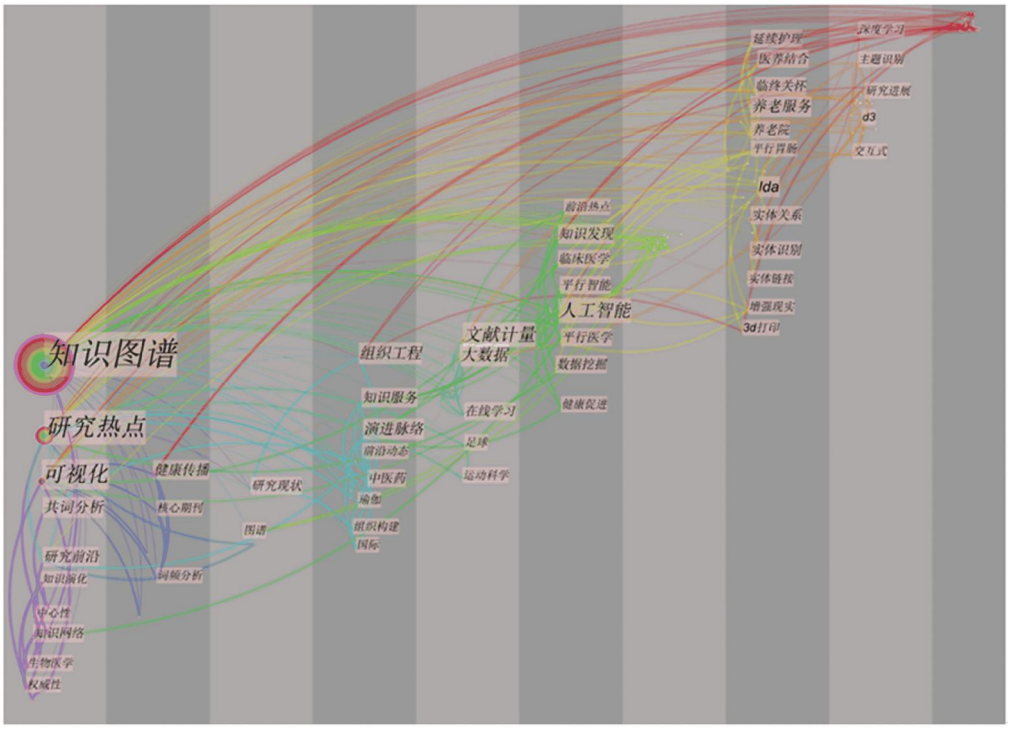

图 4 国内研究关键词时序图

Fig. 4 Timeline of domestic research keywords

3 国际研究热点和研究前沿分析

3.1 发文量

Web of Science 数据库中检索出知识图谱在医学领域研究方面的文献 1 251 篇。从时间序列上来看,2012-2021 年,国外知识图谱在医学领域的研究的发文量整体呈增长趋势,每年的发文量总体大于国内的发文量,在 2020 年增长最快,增长率为 48.45%。总体表明,国外知识图谱在医学领域的研究正处于不断发展的阶段。

3.2 作者和研究机构分析

对国外高产作者进行统计,见表 2。依据普莱斯定律,发文量在 2 篇或 2 篇以上的论文作者为核心作者,共计 77 位,共发表论文 162 篇,占所有论文总数的 12.95% ,表明领域内合作度较小,作者发文都集中在自己的小圈子。可见国外在该领域研究的高产作者带头作用还未形成,且排名前十的作者中中国学者占据了 6 位,表明国内知识图谱在医学领域的研究处于国际前沿。

利用 Citespace 进行可视化分析,获得国外该领域研究者的合作关系图以及国外机构合作关系图,如图 5、图 6 所示。由图 5 可知,国外作者间的合作度比较低,倾向于在自己的圈子中开展研究;由图 6 可知,国外知识图谱在医学领域的研究机构主要集中在高校,加拿大多伦多大学(University of Toronto)发文量最多,其次依次是加拿大的麦克马斯特大学(McMaster University)、美国的约翰斯·霍普金斯大学(Johns Hopkins University)、加拿大的麦吉尔大学(McGill University)等。在发文量前十的国外机构中,加拿大的高校占据四席,且排名前二的都是隶属于加拿大的机构,表明加拿大高校在该领域的研究投入较多,在国际处于领先地位。

表 2 国外作者发文量统计

Tab. 2 Statistics on the number of articles published by foreign authors

发文量(篇)	首次发文时间	作者
4	2014	CLOVIS FOGUEM
4	2014	BERNARD KAMSUFOGUEM
3	2012	ELPINIKI I PAPAGEORGIOU
3	2020	BUZHOU TANG
3	2020	YANG LI
3	2020	JUN YAN
2	2021	TAO LIU
2	2020	ZHEYU WANG
2	2015	ADAM LEE GORDON
2	2018	AILIAN ZHANG

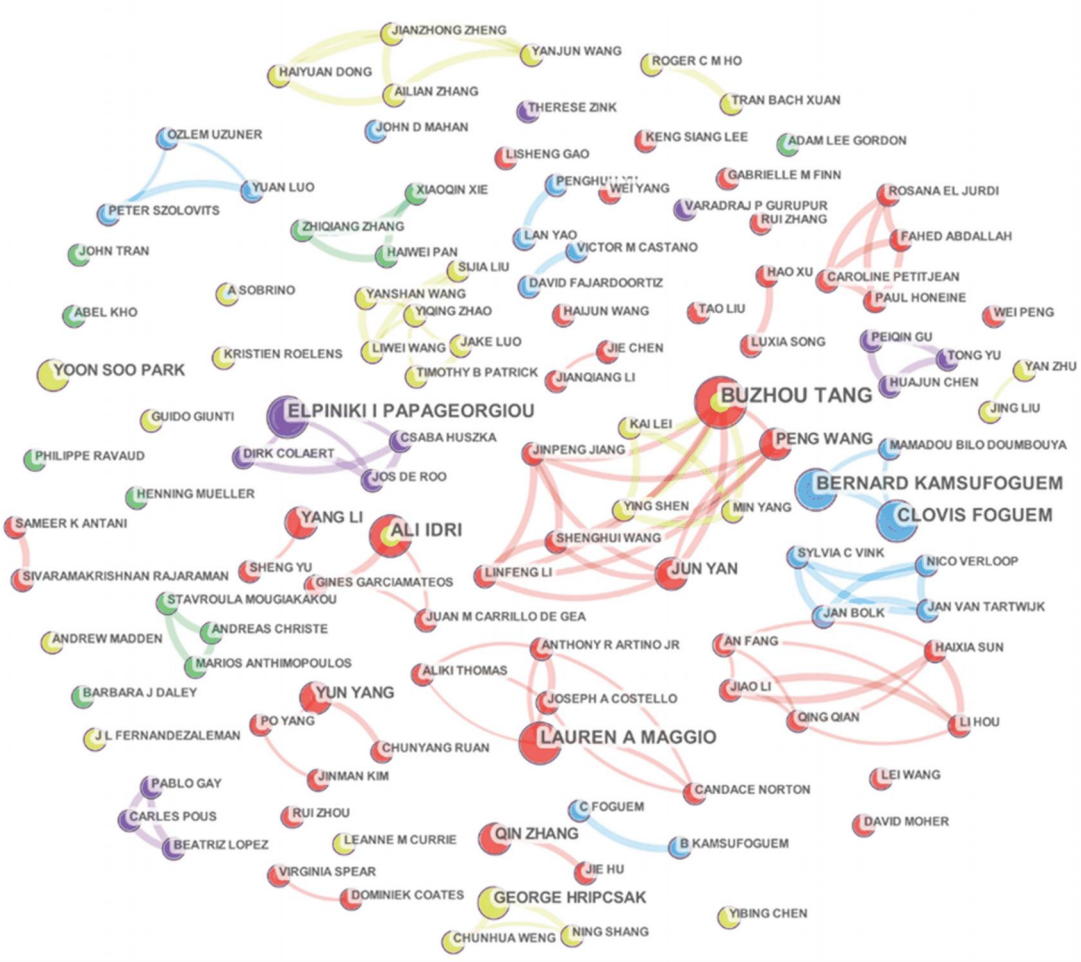

图 5　国外作者合作关系图

Fig. 5　Foreign authorcollaboration chart

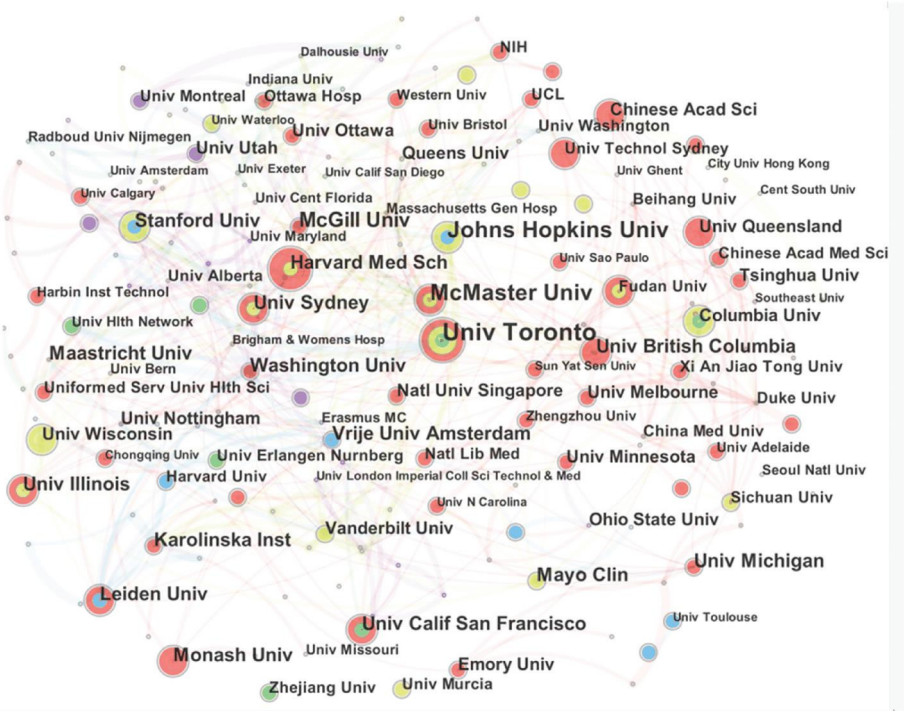

图 6　国外机构合作关系图

Fig. 6　Foreign institution partnership chart

3.3 研究热点和前沿分析

利用 CiteSpace 构建关键词共现时序图，构建的关键词共现时序图包括 339 个节点，1 786 条连线如图 7 所示。可以看到 knowledge（知识）、system（系统）、Care（护理）、model（模型）、management（管理）、education（教育）、medical education（医学教育）、disease（疾病）、classification（分类）、impact（影响）10 个热点词汇，显示当前国外在该领域的研究主题比较广。与国内的发展趋势相近，在 2019 年以后，出现了大数据、人工智能、预测等词汇，表明当前国外的知识图谱在医学领域的研究延伸到了技术应用的深层次领域。

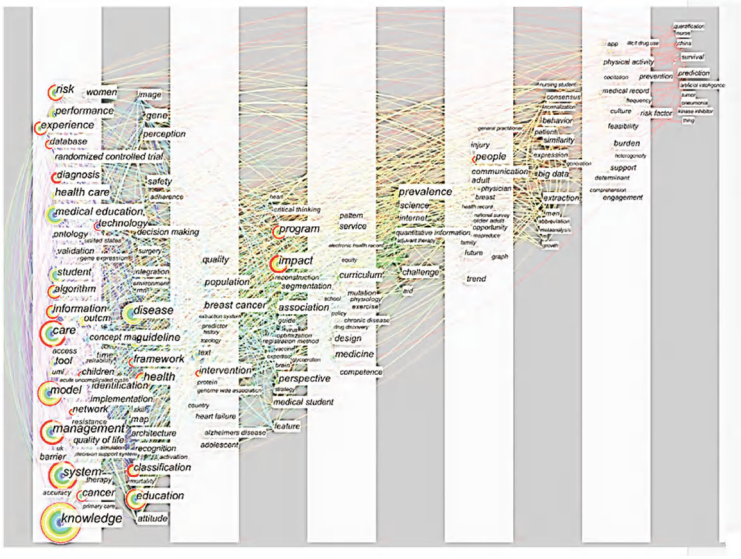

图 7 国外研究关键词共现时序图

Fig. 7 Foreign research keyword co-occurrence time series chart

4 国内外知识图谱在医学领域研究对比分析

一篇文献的研究主题、研究方法等集中体现在关键词上，因此对一学科研究热点的探析可通过统计关键词的方法来进行研究，在 CiteSpace 可视化图谱中，突变词是指在较短时间内出现较多或使用频次增长率明显提高的词，可以反映出该领域的前沿动态[9]。关键词突现度可以反映一段时间内影响力较大的研究领域[10]。利用 CiteSpace 绘制关键词突显图来综合分析该领域的研究热点如图 8 所示。

Top 23 Keywords with the Strongest Citation Bursts

Keywords	Year	Strength	Begin	End	2012 - 2021
resistance	2012	2.5	2012	2015	
architecture	2012	3.37	2013	2014	
information	2012	2.62	2013	2013	
disease	2012	3.15	2014	2015	
population	2012	2.94	2014	2016	
management	2012	2.79	2014	2014	
concept map	2012	5.79	2015	2016	
student	2012	4.28	2015	2017	
medical education	2012	3.75	2015	2017	
framework	2012	2.64	2015	2016	
curriculum	2012	3.66	2016	2018	
design	2012	3.14	2016	2019	
medicine	2012	2.67	2016	2018	
service	2012	2.6	2016	2017	
prevalence	2012	3.11	2017	2021	
outcm	2012	3.89	2018	2019	
adult	2012	2.76	2018	2019	
trend	2012	2.76	2018	2019	
science	2012	3.21	2019	2019	
patient	2012	3.08	2019	2021	
intervention	2012	2.72	2019	2021	
extraction	2012	2.7	2019	2021	
classification	2012	3.47	2020	2021	

Top 20 Keywords with the Strongest Citation Bursts

Keywords	Year	Strength	Begin	End	2012 - 2021
可视化	2012	2.25	2012	2014	
多元统计	2012	1.19	2012	2014	
研究前沿	2012	1.02	2012	2015	
研究热点	2012	1.51	2013	2015	
研究现状	2012	1.13	2013	2015	
在线学习	2012	1.27	2016	2016	
大数据	2012	1.14	2016	2016	
前沿	2012	1.36	2017	2018	
数据挖掘	2012	1.25	2017	2019	
热点	2012	1.16	2017	2017	
知识融合	2012	1.14	2018	2018	
实体关系	2012	1.87	2019	2021	
实体链接	2012	1.11	2019	2021	
全科医生	2012	1.11	2019	2021	
养老服务	2012	1.11	2019	2019	
深度学习	2012	1.97	2020	2021	
文献计量	2012	1.67	2020	2021	
研究进展	2012	1.02	2020	2021	
图数据库	2012	1.02	2020	2021	
命名实体	2012	1.02	2020	2021	

图 8 关键词突显

Fig. 8 Keyword highlight

由图 8 可知，在研究内容方面，在知识图谱概念提出的前期，该领域“可视化”、“体系机构”、“学习”、“统计”等词出现较多，表明知识图谱研究初期，知识图谱在医学领域的研究大部分工作是利用知识图谱进行医学数据的统计。近几年，国内该领域逐渐出现“命名实体”、“抽取”、“分类”、“图数据

库”等名词，可见随着科学技术的发展，医学领域的知识图谱正逐步构建起来。纵观国外知识图谱在医学领域的研究，从图概念、医学教育到药物的研发、病人的干预模式，而国内的研究则主要集中在统计和数据挖掘分析，值得注意的是中国知识图谱在养老服务中的研究比较深入。

在研究深度方面，该领域的研究初期，国内外的研究热点主要集中在“统计”、“图概念”、“学习教育”领域的研究。随着时间推移，国外学者研究的主要方向在于知识图谱在“疾病”、“药物”、“干预方式”等领域的研究，国内主要注重于“大数据”、“数据挖掘”、“养老服务”领域的研究。最近研究的趋势都倾向于“人工智能|”，“实体抽取”，“深度学习”等领域，表明知识图谱在医学领域的研究步入更深层次的阶段。

在研究方向方面，国内知识图谱在医学领域研究关键词出现频次最高的为研究热点可视化（26 次）、文献计量（13 次）、深度学习（10 次）、研究前沿（6 次）、人工智能（6 次）、实体关系（5 次）、大数据（4 次）；国外关键词出现频次最高的为“system（系统）”（70 次）、“care（护理）”（59 次）、“model（模型）”（55 次）、“management（管理）”（54 次）、“education（教育）”（42 次）、“medical education（医学教育）”（39 次）、“disease（疾病）”（39 次）、“classification（分类）”（38 次），说明国内的研究侧重于利用知识图谱相关技术进行医学领域知识的分析，并将前沿的技术应用到知识图谱中，而国外的研究侧重于把知识图谱应用到具体相关的应用，使其发挥实际作用，即国内知识图谱在医学领域的研究侧重于学术理论研究，国外研究侧重于实际应用。

5　结束语

本研究借助文献计量学方法和 Citespace 软件，对 2012~2021 年 CNKI 和 Web of Science 核心数据库中收录的、以“知识图谱在医学领域研究”为主题的研究文献，从发表时间、作者机构及前沿热点视角进行统计分析，探讨国内外学者对于知识图谱在医学领域研究异同点，得出以下结论。

从时间序列上看，知识图谱在医学领域的研究已引起国内外学者的广泛关注，该领域的发文量正随着时间推移，呈现不断增长的趋势，并且国内外在该方面的研究逐渐步入更深层次的技术领域，新的方法技术正不断应用到医学领域的知识图谱中，包括“人工智能”、“大数据技术”、“深度学习”，最近几年“实体抽取”、“实体融合”、“图数据库”等关键词不断涌出，表明医学领域的知识图谱正在逐步被构建。随着人工智能、大数据技术、机器学习和知识图谱逐步融合，构建完善的医学领域知识图谱，必定在医学辅助决策、辅助诊断、智慧医疗等方面发挥积极作用。

从该领域作者发文量和作者所属机构的合作情况来看，该领域还未形成具有带头作用的机构或团体，在该领域的研究合作度较低，知识图谱在医学领域还有广阔的发展空间，各机构间加强合作，扩展自己的合作圈是在该领域快速取得成果的有效途径。领域发文最多的前 10 作者中，国内的作者占据一多半，足以展现出中国知识图谱在医学领域的研究处于国际领先水平，中国许多优秀的学者倾向于把研究成果优先发表于国外的核心期刊中。在研究机构中，加拿大高校在该领域的研究投入较多，在该领域的科研实力较强。

国内外知识图谱在医学领域方面的研究侧重点不同，国内学者在该领域的研究处于世界领先地位，未来利用大数据、人工智能、深度学习技术推进医学领域知识图谱的构建当前知识图谱在医学领域的研究趋势。国内学者加强合作，积极探索理论和应用相结合的方式方法，进一步深化研究，必然推动中国医学领域的全面发展。

参考文献

[1] LI G, LIU Y, CAI H. Research on application of big data in medical industry[C]//2018 3rd International Conference on Smart City and Systems Engineering (ICSCSE). IEEE, 2018: 763-765.

[2] 袁凯琦，邓扬，陈道源，等. 医学知识图谱构建技术与研究进展[J]. 计算机应用研究，2018，35(7):8.

[3] YAN J, WANG C, CHENG W, et al. A retrospective of knowledge graphs[J]. Frontiers of Computer Science, 2018, 12(1): 55-74.

[4] 朱超宇，刘雷. 基于知识图谱的医学决策支持应用综述[J]. 数据分析与知识发现，2020，4(12):26-32.

[5] 段宏. 知识图谱构建技术综述[J]. 计算机研究与发展，2016，53(3):19.

[6] 陈仕吉. 科学研究前沿探测方法综述[J]. 现代图书情报技术，2009(9): 28-33.

[7] 孙雨生，陈卫. 我国网格服务研究进展——基于 CNKI（2003-2012）的文献计量与知识图谱分析[J]. 现代情报，2013，33(7): 102-111.

[8] 安传艳，李同昇，翟洲燕，等. 1992-2016 年中国乡村旅游研究特征与趋势——基于 CiteSpace 知识图谱分析[J]. 地理科学进展，2018，37(9):30-44.

[9] 寇继虹，楼雯. 概念图研究演进的知识图谱分析[J]. 图书情报知识，2012(2):117-123.

[10] 李静，朱继民，武松. 我国医学统计学课程研究热点及趋势的知识图谱分析[J]. 中国卫生统计，2020，37(2):284-286.