

知识图谱在领域知识多维分析中的应用途径研究*

王思茗¹ 孙熊兰¹ 滕广青¹ 叶心² 栾宇¹

(1. 东北师范大学信息科学与技术学院, 长春 130117; 2. 长春市图书馆, 长春 130021)

摘要: 知识图谱为解决众多领域的现实问题提供了新的思想与方法, 有助于呈现和分析领域知识及其关联关系。研究工作采用图数据库技术构建领域知识图谱。从简单实体关系和多维复杂关系层面, 对知识图谱在领域知识多维分析中的应用途径进行探索。研究表明, 知识图谱能够从多维角度组织和分析领域知识与知识关联; 灵活地对领域知识及其关系实现可视化呈现; 能够方便存储和快速提取知识关系并进行推理。

关键词: 知识图谱; 实体关系; 领域知识; 多维分析

中图分类号: G255.76

DOI: 10.3772/j.issn.1673-2286.2019.03.003

21世纪以来, 随着网络科学的复兴, 以网络结构表征知识关联关系的思想已经被图书情报学界普遍接受和认可。人工智能领域中各类神经网络(CNN、RNN、DNN)^[1]的崛起, 无一不印证了网络思维在揭示多维复杂关系中的重要性。2012年5月, Google在其官方网站发布*Introducing the Knowledge Graph: Things, not Strings*, 首次提出了知识图谱(Knowledge Graph)的概念^[2]。知识图谱相比传统的知识网络具有更丰富的知识关联, 正在迅速应用于经济、医疗、金融、教育等诸多领域。与此同时, 图书情报学界也开始尝试知识图谱在本学科领域的应用。

本文基于知识图谱理论及其构建技术, 构建特定学科领域的知识图谱。在阐述知识图谱构建方法及过程的基础上, 基于知识图谱从简单实体间关系和多维复杂关系的视角进行研究分析, 探索和分析知识图谱在领域知识多维分析中可能的应用途径, 以期为基于知识图谱的多维知识分析提供参考和借鉴。

1 相关研究现状

传统的知识组织体系以树形结构为主, 当前风行世

界各大图书馆的杜威十进制^[3]就是典型的代表。树形结构强调非此即彼的隶属关系和等级森严的层级关系, 体系结构相对简单和僵化。进入21世纪以后, 领域本体^[4]凭借其多重继承关系和对语义网的支持, 一定程度上突破了传统树形结构的桎梏。Soergel^[5]认为, 本体作为概念与概念之间的语义地图, 有利于将多元学科、语种和文化的知识建立关联, 作为一种参考工具为知识组织体系的构建过程提供方向指引。Hautamäki^[6]通过对实体的“属性—值”描述, 丰富了领域本体中概念的内在逻辑和继承关系类型。滕广青等^[7]指出, 领域本体的相关研究沿着“本体构建—本体合并—本体集成—本体对应”这一总体路径向前发展。孙雨生等^[8]对本体研究和应用进行总结, 表明领域本体可以应用于知识表示、知识组织、知识存储、知识检索、知识导航、知识问答、知识推荐、用户兴趣建模、个性化定制等方面。尽管领域本体在知识组织与服务领域取得了较为显著的进展, 但是其仍然受制于严苛的等级关系。

随着网络科学理论和方法的突破, 网络思维和网络分析技术被引入知识组织与信息分析的研究领域。Martin等^[9]使用网络和聚类的思想构建知识网络, 将知识群落用冲击图进行层级展示, 进而分析领域知识的

*本研究得到国家自然科学基金面上项目“基于网络结构演化的Folksonomy模式中社群知识组织与知识涌现研究”(编号: 71473035)资助。

演化进程。Pham等^[10]将期刊论文和会议论文结合,借助论文中的引用关系分别构建期刊关系网络、主题关系网络、知识关系网络等多个不同维度的知识网络,并通过不同维度的网络交叉分析,挖掘知识间的引用行为,判识子学科的发展趋势。Ma等^[11]通过一致性分析、中心性分析、核心-边缘分析等网络分析方法,证实了标签知识网络同时具备小世界效应和无标度属性,并基于引用关系的发展演化过程建立了引文知识网络的时序演化模型^[12]。安宁等^[13]从Hub涌现的角度对知识网络在时间序列上的知识涌现进行分析,揭示了其中随机性与非随机性的交互。随着研究工作的深入,知识网络的相关研究在主题聚类^[14]、学科进展^[15]、科研合作^[16]、知识传递^[17]等方面取得了丰硕的成果,网络科学的理论与方法也得到学术界的普遍认可。随着数据科学的崛起和相关研究的深入,单一的网络分析在面临多维信息分析时的局限性逐渐显露出来。

在领域本体数据模式与知识网络结构关系的基础上,知识图谱^[2]应运而生。目前国内学术界对知识图谱的相关研究刚起步,尚处于引进和探索阶段^[18],虽然在模型构建方法^[19]等基础研究方面进行了探索和尝试,但同时也暴露出一些问题。对真正意义的知识图谱的认识尚不清晰,知识地图、知识网络、知识图谱等概念之间存在一定的混用现象^[20]。有鉴于此,本文将详细阐述知识图谱的构建过程,并对知识图谱相对于一般知识网络在领域知识多维分析中的可能途径进行探索和分析,以为知识图谱在领域知识多维分析中的应用提供参考和借鉴。

2 基础理论框架

随着数据科学的崛起以及领域本体与知识网络的应用,数据多样化与复杂的结构关系为知识间多维度关联关系的揭示提出了新的诉求。知识图谱可以形式化地对客观世界中不同类别的概念实体以及实体之间的关系进行表示,并通过属性对实体和实体间关系进行界定和描述。

在领域本体和知识网络中,实体(节点)的识别与判定往往需要对真实数据进行抽象化,尤其适用于相同或近似性质的事物或概念,实体间关系也被抽象化和简单化。知识图谱则可以容纳多种不同类型的实体,实体间关系更是种类繁多,加之附带的实体属性与关

系属性,知识图谱可以用于组织和架构现实社会中众多的事物及其关联关系。当前,知识图谱已经在语义检索、问答系统、金融分析等众多领域得到运用,并展露出巨大的应用前景。基于科技文献的相关数据构建领域知识图谱,能够为面向学科知识服务的多维知识分析提供新的途径。

本研究的目的在于构建学科领域知识图谱,并通过对知识图谱中不同维度的实体间关系分析,探索和展示知识图谱在多维信息分析中的应用途径和优势。因此,研究工作将基于图数据库技术构建领域知识图谱。知识图谱的具体构建步骤如下。①知识提取。知识提取包括实体提取、关系提取及属性提取,并将结构化、半结构化或非结构化数据与实体、关系及属性对齐^[21]。②知识图谱架构。将“实体—关系—实体”和“实体/关系—属性—属性值”以图结构的形式进行架构,构成包含全部数据的实体关系属性网络。③知识层面应用。对所构建的知识图谱进行知识层面的分析,从逻辑概念层次与图层次进行推理,挖掘潜在知识及其关联。

研究中,将基于上述步骤构建领域知识图谱,并基于知识图谱将领域知识分析从简单实体间关系分析推进到多维复杂关系分析,探索知识图谱在领域知识多维分析中应用的可能途径。

3 数据采集与知识图谱构建

3.1 研究数据

本研究根据2017年发布的*Journal Citation Reports*(《期刊引证报告》),利用Web of Science核心数据库检索“Information Science& Library Science”学科领域内Q1区的22种权威期刊,选取2014—2018年收录的期刊文献,期刊语种限定为英语,检索日期为2018年12月11日,所获数据及相关信息见表1。

数据显示共获得“Information Science& Library Science”学科5年内Q1区期刊论文7 010篇。被引量(篇均被引>11)和影响因子表明,研究中所采集的文献数据具有高被引和高影响因子的特征,一定程度上能够代表该学科领域的主要研究状况。此外,表1只列示了部分相关数据,限于篇幅和数据关系复杂性的原因,基金资助、地域、作者署名、刊发时间等数据并没有列示。更细致的多维信息将在领域知识图谱中体现。

表1 文献数据相关信息

序 号	期刊名称	发文数量/篇	总被引量/次	影响因子
1	<i>Mis Quarterly</i>	212	15 184	5.430
2	<i>Journal of Information Technology</i>	98	1 170	4.535
3	<i>International Journal of Information Management</i>	429	3 626	4.516
4	<i>Journal of Strategic Information Systems</i>	81	1 464	4.313
5	<i>Journal of the American Medical Informatics Association</i>	847	8 713	4.270
6	<i>Information Systems Journal</i>	127	1 580	4.267
7	<i>Government Information Quarterly</i>	274	2 837	4.009
8	<i>Journal of Computer-Mediated Communication</i>	158	4 181	4.000
9	<i>Information & Management</i>	381	6 406	3.890
10	<i>Telematics and Informatics</i>	533	1 574	3.789
11	<i>Journal of Informetrics</i>	373	2 334	3.484
12	<i>Information Processing & Management</i>	330	3 250	3.444
13	<i>International Journal of Computer-Supported Collaborative Learning</i>	80	721	3.273
14	<i>Social Science Computer Review</i>	225	1 631	3.253
15	<i>European Journal of Information Systems</i>	161	2 597	3.197
16	<i>Journal of the Association for Information Systems</i>	125	1 842	2.839
17	<i>Journal of the Association for Information Science and Technology</i>	780	1 904	2.835
18	<i>Journal of Management Information Systems</i>	191	4 560	2.744
19	<i>Journal of Knowledge Management</i>	560	3 318	2.551
20	<i>Journal of Enterprise Information Management</i>	168	961	2.482
21	<i>Research Evaluation</i>	156	1 049	2.449
22	<i>Qualitative Health Research</i>	721	7 660	2.413

3.2 实体与关系及属性提取

本研究假设实体集 $U=\{A, T, K, J, S, R\}$ 。其中，A代表作者，T代表文献标题，K代表关键词，J代表期刊，S代表地域，R代表文献所属的学科方向。关系集 $V=\{A_write_T, A_use_K, A_site_S, K_tag_T, T_published_J, T_belong_R, J_site_S\}$ 。其中，A_write_T表示作者A创作文献T，A_use_K代表作者A使用关键词K，A_site_S代表作者A归属于地域S（国家或地区），K_tag_T代表关键词K标注文献T，T_published_J代表文献T发表于期刊J，T_belong_R代表文献T归属学科方向R，J_site_S代表期刊J归属于地域S（国家或地区）。属性集 $O=\{Ax, Tf, ATy, TJy\}$ ，属性集中包含两类属性，分别为实体属性和关系属性。其中，Ax表示为作者A是否为通信作者的属性，Tf为文献T是否有基金资助，ATy表示作者A创作文献T的时间（首次投稿时间），TJy则表示为在文献T发表于期刊J（T_published_J）的这条关系中存在文献发表时间的

属性。相关实体、关系、属性的界定可以在具体的研究工作中根据任务需求有所调整。

（1）A_write_T。A_write_T表示作者A创作文献T，作者A与文献T为实体。考虑到同一篇文献可能由一名作者独立完成，也可能由多名作者共同创作完成，实体作者自身又具有是否为通信作者的属性，在知识图谱中通过布尔逻辑值是或否（True/False）表示作者是否为通信作者这一属性。在作者实体建立过程中，将是否为通信作者加入实体属性。同时，作者创作文献的这条关系中含有文献创作时间的属性。

（2）A_use_K。A_use_K表示作者A使用关键词K，作者A和关键词K作为连接关系两端的实体。关键词在一篇论文中至关重要，能够精准反映文章的主旨。为保证知识图谱中数据的完整性和准确性，本文所构建的知识图谱选取扩展关键词作为实体。

（3）A_site_S。A_site_S表示作者A归属于地域S。地域S为国家或地区，将地域作为实体加入知识图谱，有利于从地域的角度分析作者合作现象、学科地域

分布、区域研究现状和热点等,并能够在一定程度上化解作者重名问题。

(4) K_tag_T 。 K_tag_T 表示为关键词 K 标注文献 T 。通过对文献关键词进行分析,可以洞察所选学科领域权威期刊的研究热点,从而跟踪和把握领域知识发展方向。

(5) $T_published_J$ 。 $T_published_J$ 表示文献 T 发表于期刊 J ,文献 T 与期刊 J 作为实体。科学文献是否具有基金资助的属性,除此以外,连接文献与期刊的实体间关系附带发表时间的属性。

(6) T_belong_R 。 T_belong_R 表示文献 T 归属于学科方向 R 。许多研究工作在学科方向上存在交叉关系,将文献归属于各个学科方向,在一定程度上可以反映出研究领域中的学科交叉与融合,判识当前学科领域发展的趋势。在知识图谱视域下,知识并非严格隶属于某一固定学科方向,学科交融和知识交叉可以通过知识图谱得以呈现,有助于知识演化与科技创新的预测。

(7) J_site_S 。 J_site_S 表示为期刊 J 归属地域 S (国家或地区)。本文选取*Journal Citation Reports*(《期刊引证报告》)“Information Science& Library Science”学科领域的Q1区期刊,在某种程度上代表了图书情报学领域的权威地位。通过 J_site_S 可以推理出不同地域对学科知识研究与传播的贡献。

研究中基于不同的关系类型,分别对不同类型实体间关系包含的实体数量、实体间关系数量进行统计,所得结果见表2。

表2 实体与实体间关系数量

实体间关系	实体间关系含义	实体间关系数量/条	实体数量/个	实体数量/个
A_write_T	作者创作文献	24 203	A: 19 538	T: 7 010
A_use_K	作者使用关键词	193 199	A: 19 538	K: 9 674
A_site_S	作者归属地域	20 091	A: 19 538	S: 118
K_tag_T	关键词标注文献	49 246	K: 9 674	T: 7 010
T_belong_R	文献归属学科方向	17 142	T: 7 010	R: 10
$T_published_J$	文献发表于期刊	7 010	T: 7 010	J: 22
J_site_S	期刊归属地域	23	J: 22	S: 3

统计数据表明,研究工作共提取36 376个实体节点,310 918条实体间关系。其中 A_use_K (作者使用关

键词)在实体间关系中占比超过62%,是关系总量中最多的实体间关系。研究工作基于上述提取的实体、关系和相关属性,采用图数据库技术构建学科领域知识图谱。

3.3 知识图谱构建

在该学科领域内实体、关系和属性进行提取的基础上,综合领域本体的数据模式和知识网络的结构关系,借助图数据库技术,基于研究中所采集的数据及所提取的实体、关系、属性,采用开源的Neo4j图引擎构建图书情报学科领域的知识图谱,基本结构见图1。

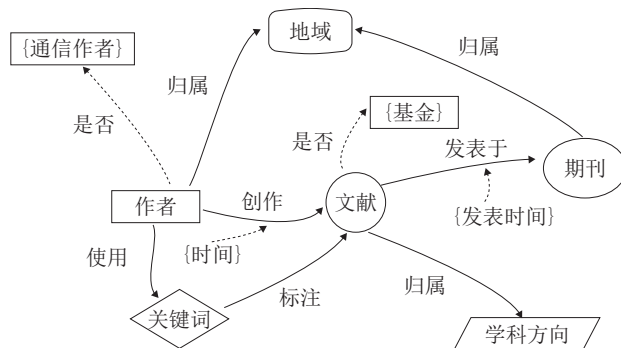


图1 领域知识图谱基本结构

在知识图谱基本结构中,由作者、文献、关键词、期刊、地域(国家或地区)、学科方向构成实体集合。实体间的关系由有向实线表示。实体属性用虚线与实体连接,其中作者实体带有是否为通信作者的属性,文献实体带有是否有基金资助的属性;关系属性用虚线与关系连接,分别为作者创作文献时间(首次投稿时间)和文献正式发表时间。从图1中可以获知的实体间关系信息有:作者实体创作了文献实体从而产生关系,作者实体具有是否为通信作者的属性,在作者创作文献的这条关系中具有作者创作时间的属性。作者使用关键词,关键词用来标注文献,从而作者与关键词、关键词与文献之间产生关系。文献在期刊上发表也产生一种关系,同时这条关系附带文献发表时间的属性。文献带有基金属性,文献与学科方向存在归属关系。此外,作者、期刊还分别与地域(国家或地区)存在归属关系。在此基础上,研究工作将对知识图谱在领域知识多维分析中的应用途径进行探索。

4 知识图谱在领域知识分析中的应用途径

4.1 简单实体间关系分析

本文所构建的领域知识图谱中,涉及两种不同实体的简单关系包含7种类型。从网络科学的视角看,简单实体间关系分析类似于2-模网络分析,并在2-模网络的基础上增加了实体属性和关系属性。研究工作选取文献归属学科方向(T_belong_R)、关键词标注文献(K_tag_T)、作者创作文献(A_write_T)3类简单实体间关系,对知识图谱在领域知识多维分析中的应用

进行分析。

4.1.1 文献归属学科方向的简单实体间关系

本文构建的知识图谱中包含7 010篇文献,共涉及10个学科方向。通过文献归属的学科方向(T_belong_R)产生17 142条关系。在知识交叉融合的大背景下,一篇文献可能同时归属于多个学科方向,不但能够揭示科研工作对不同学科知识的吸纳融合,还可以通过统计数据反映多学科之间的交叉程度。研究工作从所构建的知识图谱中提取了部分文献及其归属的学科方向(见图2)。

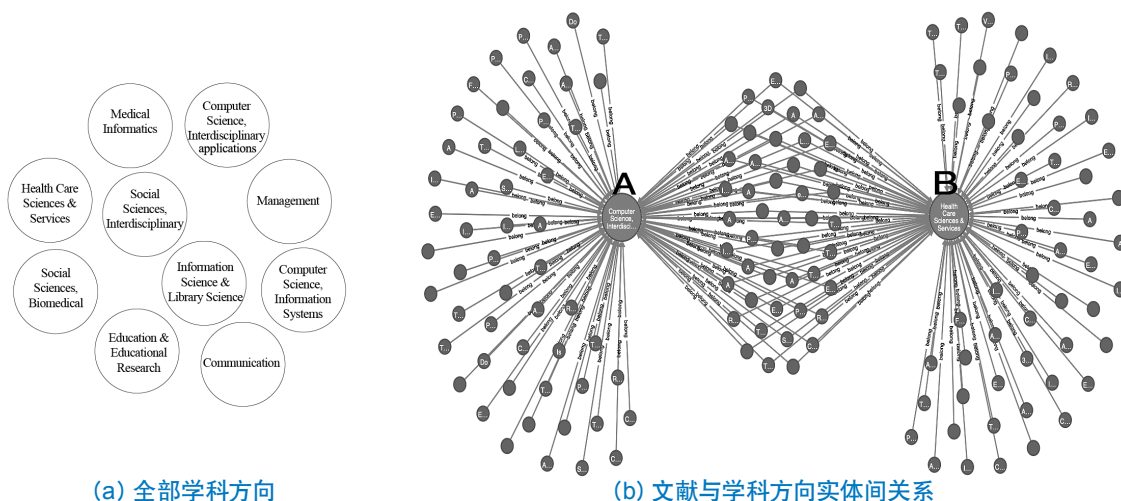


图2 文献归属学科方向的简单实体间关系

图2(a)为图书情报学科领域Q1区内22种核心期刊在2014—2018年所刊发文献归属的10个学科方向。图2(b)为以计算机跨学科应用(A: Computer Science, Interdisciplinary Applications)和卫生保健科学与服务(B: Health Care Sciences & Services)两个学科方向为例提取的部分文献。其中指向A、B的有向关系为文献归属的学科方向(T_belong_R)。图2(b)中,左右两侧文献分别归属于计算机跨学科应用(A)和卫生保健科学与服务(B)两个学科方向,位于中间位置的文献则同时归属于计算机跨学科应用(A)和卫生保健科学与服务(B)两个学科方向。限于篇幅原因,图中只提取部分文献,且未展示与其他学科的归属关系。在本研究构建的领域知识图谱中,文献的学科归属关系更加丰富。表3为知识图谱中全部文献涉及的学科方向。

在采集的图书情报学科领域的7 010篇文献中,有

表3 全部文献涉及的学科方向

研究方向	文献总量/篇
Information Science & Library Science	7 010
Computer Science, Information Systems	3 306
Computer Science, Interdisciplinary Applications	1 645
Management	1 482
Social Sciences, Interdisciplinary	996
Health Care Sciences & Services	847
Medical Informatics	847
Social Sciences, Biomedical	771
Communication	158
Education & Educational Research	80

3 306篇同时归属于计算机科学信息系统(Computer Science, Information Systems),1 645篇文献同时归属于计算机跨学科应用(Computer Science,

Interdisciplinary Applications), 1 482篇文献归属于管理学(Management)等。这表明, 2014—2018年, 图书情报学科方向与计算机科学信息系统学科方向的交叉融合现象非常显著(超过47%的文献同时涉及这两个学科方向的交叉)。

4.1.2 关键词标注文献的简单实体间关系

关键词为文献主题的语言凝练, 简洁准确地表达了文献的主旨。关键词不仅可以帮助读者总体上掌握文献的主旨, 读者也可以使用关键词作为检索词并根据自身的需求获取相关文献。本研究构建的领域知识图谱中共包含9 674个关键词, 7 010篇文献, 49 246条K_tag_T实体间关系。图3为部分关键词标注文献(K_tag_T)的局部知识图谱, 展示了关键词与文献之间以标注关系作为连接产生的关联。

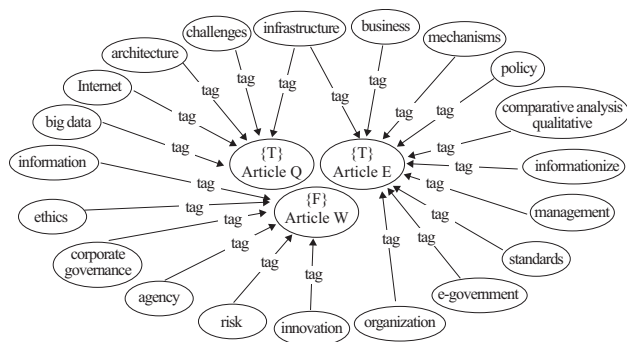


图3 关键词标注文献的简单实体间关系

其中, 外围椭圆形为关键词实体, 中心处为3篇文献实体。关键词标注文献(K_tag_T)关系由外围的关键词实体分别指向中心处的文献实体。Article Q、Article W、Article E分别被5个、6个和10个关键词标注。在一般的网络分析中, 可以借助关键词标注文献的简单实体间关系确定关键词或文献间的潜在关联或相似性。如通过关键词“internet”和“big data”共同标注Article Q的同现关系确定两个关键词的潜在关联并构建1-模共词网络; 通过关键词“infrastructure”分别标注了Article Q、Article E两篇文献, 发现两篇文献间的相似性等。这些分析思想与方法在以往的知识网络分析中经常被使用。

在知识图谱中, 即使简单的实体间关系也可以借助实体属性增加信息维度。从图3中可以发现, 文献实体具有基金资助属性, 其中Article Q、Article E均有基金

资助{T}, Article W则没有基金资助{F}。通过基金资助属性, 可以获得政府或社会团体对相关研究的政策支持与支持状况。传统的网络分析更多地关注网络节点之间的关系, 一定程度上忽视了网络节点(实体)自身所具备的性质, 而知识图谱则可以通过实体属性呈现更多维度的信息。

4.1.3 作者创作文献的简单实体间关系

本研究构建的领域知识图谱共包含7 010篇文献、19 538位作者, 以及24 203条作者创作文献(A_write_T)的实体间关系。研究工作抽取了知识图谱中部分作者创作文献(A_write_T)的关系(见图4)。

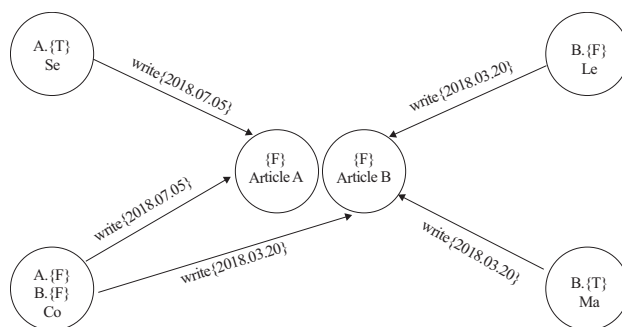


图4 作者创作文献的简单实体间关系

其中, Co、Se、Le、Ma为作者实体, Article A、Article B为文献实体。作者实体与文献实体通过创作关系产生连接。图3中借助实体属性能够扩展领域知识分析的信息维度, 事实上更多的属性能够将知识分析的维度进一步扩展。通过图4中的作者创作文献(A_write_T)关系可以发现, 作者Co与作者Le、Ma于{2018.03.20}共同撰写了Article B, 作者Ma为通信作者(通信作者属性为True: B.{T}); 作者Co与作者Se于{2018.07.05}共同撰写了Article A, 作者Se为通信作者。Article A与Article B均未获得基金资助(基金属性为False: {F})。显然, 基于知识图谱的分析, 借助实体属性和关系属性, 能够获得比普通网络分析更丰富的信息。如通信作者属性能够帮助研究者更准确地识别科研团队中的领军人物; 创作时间属性有利于在刊发周期漫长的环境下解决思想或方法首创的辨识问题等。

事实上, 研究中构建的领域知识图谱所包含的实体间关系的类型和规模远大于文中所抽取展示的部分。如果以经典网络分析中的2-模网络视角观察, 知识图谱中则包含多种相互关联的2-模网络, 复杂程度与

信息丰富程度远高于普通2-模网络。加之知识图谱的简单实体间关系可以方便地扩展到多维复杂实体间关系,因此可以基于知识图谱挖掘到更丰富的多维信息。

4.2 复杂实体间关系分析

复杂实体间关系可以理解为简单实体间关系的组合交叠。由于知识图谱的灵活性,不仅可以呈现出所查询的实体间关系,同时还会呈现所查询实体间的其他关联关系。这部分研究中,基于知识图谱提取并组合4种复杂实体间关系进行分析,以进一步探索知识图谱在领域知识多维分析中的可能途径。

4.2.1 作者、关键词、文献、期刊间的多维复杂关系

研究工作从领域知识图谱中抽取了作者、关键词、文献、期刊4种类型的部分实体及其关系。作者实体为Con、Cha、Ti、Ju、Sc、Bh、Pa、Re、Jo,关键词实体为“agency”和“scientometrics”, Article X为文献实体,期刊实体为“Information Systems Journal”。实体间关系则分别为作者使用关键词(A_use_K)、关键词标注文献(K_tag_T)、作者创作文献(A_write_T)及文献发表于期刊(T_published_J)。同时包括相应的实体属性和关系属性。抽取的局部知识图谱见图5。

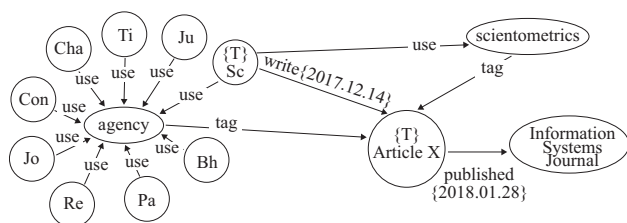


图5 作者、关键词、文献、期刊间的多维复杂关系

其中,9位作者分别在各自的文献中使用“agency”作为关键词,关键词“agency”又被作者Sc用于标注Article X。作者Sc于{2017.12.14}创作Article X,是该文献的通信作者(通信作者属性值为True: {T}),并同时使用了关键词“scientometrics”标注文献Article X。Article X受到基金资助(基金属性值为True: {T}),并于{2018.01.28}发表于期刊“Information Systems Journal”。图5中多个实体间关系形成多条有向关系路径,这种交叉关联的多维关系在传统的知识网络分析中难以体现,也正是知识图谱的优势所在。知识图谱完

全突破了一般1-模知识网络和2-模知识网络的结构关系,能够以更丰富的信息维度灵活地对知识及其关系进行呈现与表达。

4.2.2 作者、文献、学科、期刊间的多维复杂关系

研究工作基于所构建的领域知识图谱提取了4种实体(作者、文献、学科方向、期刊)与3种实体间关系(作者创作文献A_write_T、文献归属学科方向T_belong_R、文献发表于期刊T_published_J),组成局部知识图谱(见图6)。

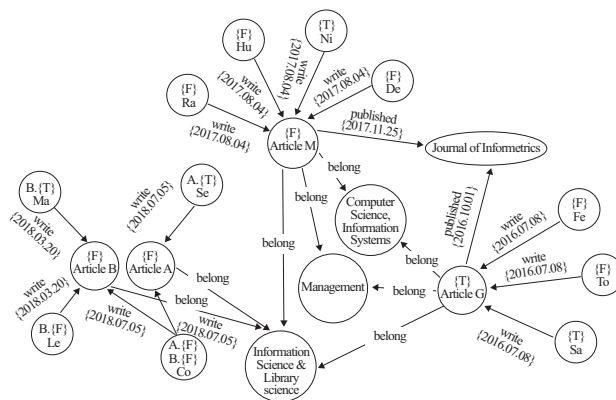


图6 作者、文献、学科、期刊间的多维复杂关系

其中,作者实体为Ma、Le、Co、Se、Ra、Hu、Ni、De、Fe、To、Sa,文献实体为Article A、Article B、Article M、Article G,学科方向实体为“Computer Science, Information Systems”“Management”和“Information Science & Library Science”,期刊实体为“Journal of Informetrics”。如图6所示,作者Co与作者Ma、Le于{2018.03.20}合作创作了Article B,作者Ma为通信作者;作者Co与作者Se于{2018.07.05}合作创作了Article A,作者Se为通信作者。Article A、Article B归属于“Information Science & Library Science”学科方向。作者Ra、Hu、Ni、De于{2017.08.04}共同创作了Article M,作者Sa、To、Fe于{2016.07.08}共同创作Article G并受到基金资助。Article M、Article G分别于{2017.11.25}和{2016.10.01}在“Journal of Informetrics”上发表,且共同归属于“Information Science & Library Science”、“Management”和“Computer Science, Information Systems”3个学科方向。图6同时可以视为图5局部知识图谱基础上的简单关系的扩展。仅就图6呈现的内容而言,不但可以发现Article A、Article B、Article M、Article G

在学科方向上具有一定的相关性, Article M、Article G所归属的学科方向和刊发期刊均相同,相关性更高;还可以发现传统的图书情报学期刊“Journal of Informetrics”近年来一定程度上倾向于“Information Science & Library Science”、“Management”和“Computer Science, Information Systems”等学科方向交叉的研究成果等信息。知识图谱中的多维复杂关系将领域知识分析扩展到更丰富的维度空间。

4.2.3 关键词、文献、作者、学科、期刊间的多维复杂关系

为了更突出地展现知识图谱在领域知识多维分析中的应用价值,研究工作进一步提取了更多类型的实体及其关系。基于领域知识图谱提取关键词、文献、作者、学科、期刊5类实体和实体间关系及相应的属性,组成的局部知识图谱如图7所示。

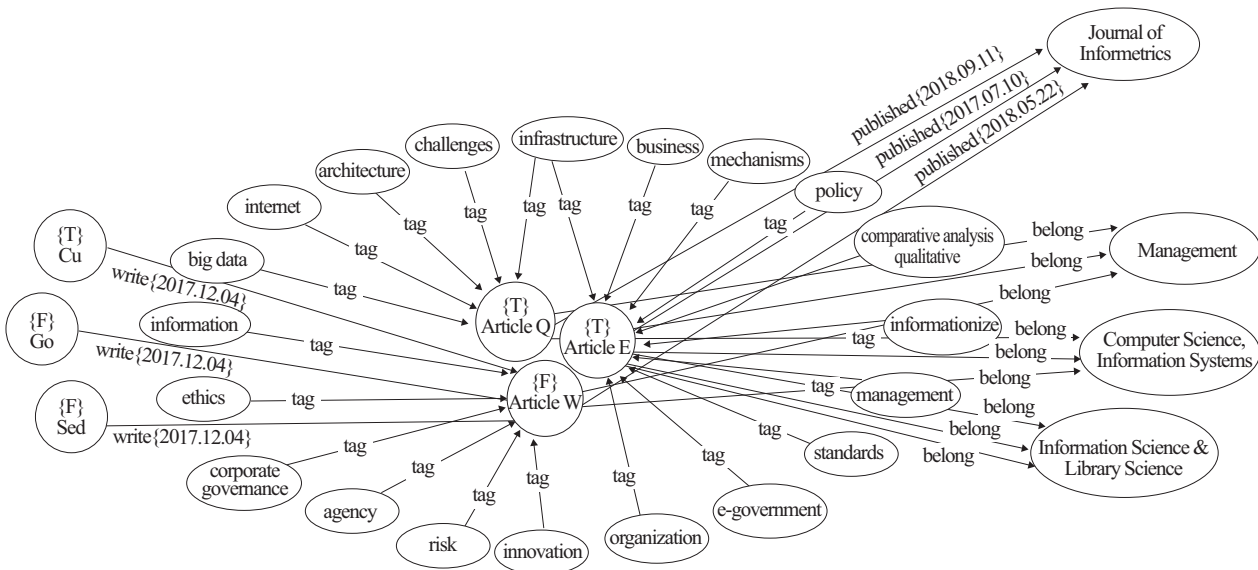


图7 关键词、文献、作者、学科、期刊间的多维复杂关系

图7可以视为图4局部知识图谱中简单实体关系的扩展。图7包含的5种实体间关系,分别是作者创作文献(A_write_T)、作者使用关键词(A_use_K)、关键词标注文献(K_tag_T)、文献归属学科方向(T_belongs_R)及文献发表于期刊(T_published_J)。此外还包括作者实体的通信作者属性、文献实体的基金属性,以及作者创作文献关系的时间属性和文献发表于期刊关系的时间属性。在关键词、文献、作者、学科和期刊的多维关系中,作者Cu以通信作者的身份与作者Go、Sed于{2017.12.04}合作创作了Article W,并于{2018.05.22}在“Journal of Informetrics”上发表。Article Q、Article E的研究工作获得了基金资助(文献实体的基金属性值为True: {T})。

表面上看(与图3相同的部分), Article Q、Article E、Article W的关键词存在较大差异,从关键词的维度看3篇文献的相关性不大。但是对图7中新增维度关系的进一步分析可知, Article W、Article Q、Article E虽然在关键词使用上存在差异,但3篇文献刊发于同一学术期刊、文献所归属的学科方向相同、同时涉及3个

学科方向的交叉等;甚至可以大致推断出3篇文献具有一定的学科相关性。显然,知识图谱能够从多维复杂关系的层面为领域知识多维分析提供支持与可能。

4.2.4 作者、关键词、文献、期刊、地域间的多维复杂关系

研究工作进一步从领域知识图谱中提取了5类实体和6种实体间关系。实体类型分别为作者、关键词、文献、期刊、地域;实体间关系类型为作者创作文献(A_write_T)、作者使用关键词(A_use_K)、关键词标注文献(K_tag_T)、文献发表于期刊(T_published_J)、文献归属于地域(A_site_S)、作者归属于地域(A_site_S)。获得在图5基础上扩展后的局部知识图谱如图8所示。

在多维复杂关系中共有8个关键词实体,其中6个关键词共同标注受到基金资助的Article Z;关键词“agency”标注2篇Article Y、Article X,并且与关键词

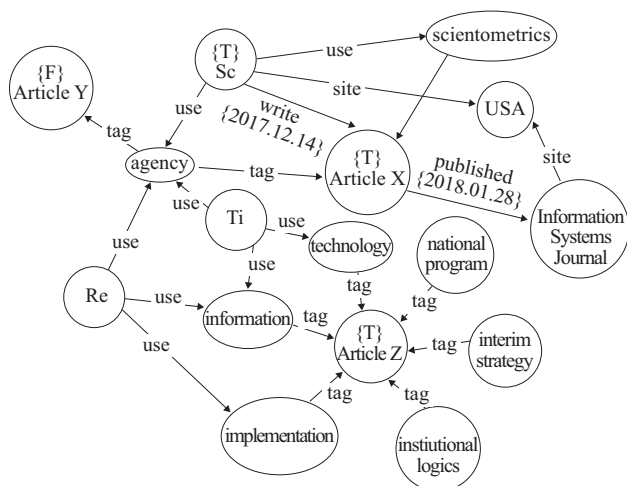


图8 作者、关键词、文献、期刊、地域间的多维复杂关系

“scientometrics”共同标注受基金资助的Article X。作者Sc、Ti、Re共同使用关键词“agency”，其中作者Sc作为通信作者使用关键词“agency”“scientometrics”标注了其于{2017.12.14}创作的Article X，该文献于{2018.01.28}刊发于期刊“Information Systems Journal”，该期刊与作者均归属于地域“USA”。

随着更多实体及实体间维度的加入，更多维度的信息得以呈现。图8中虽然仅展示了领域知识图谱的部分数据和部分实体及实体间关系，但其中包含的信息维度远远高于一般的知识网络。实际研究工作中，对领域全局的知识图谱进行分析，则可以从知识图谱当中挖掘获得更多的领域知识信息，借助其中作者、文献、关键词、期刊、地域、基金、时间等多维信息，可以对科研团队及领袖、学科研究状况、热点知识识别、区域优势分布、发展趋势预测等提供有力的支撑。

5 结论与讨论

本文采用图数据库技术，基于Web of Science核心数据库“Information Science & Library Science”学科领域内Q1区的22种权威期刊2014—2018年文献数据，构建领域知识图谱。从简单关系和多维复杂关系两个层面，对知识图谱在领域知识多维信息分析中的应用进行可视化分析与展示。通过上述分析可以得出以下结论。

(1) 知识图谱能够从多维角度组织和分析领域知识与知识关联。与一般的领域本体和知识网络的体系结构不同，知识图谱内含丰富的实体类型和实体间关系，并且实体和关系都可以携带自身的属性，信息携带

量极其丰富。知识图谱自身的体系结构不但适合处理结构化数据，更适合处理半结构化和非结构化数据。借助内部丰富的信息，知识图谱能够从多个维度对领域知识进行组织与分析，更适用于面向学科知识服务的多维信息分析任务。

(2) 知识图谱能够灵活地对领域知识及其关系实现可视化呈现。知识图谱不仅是一种信息组织与分析技术，同时也是一种信息可视化技术和手段。研究中不同维度的初步分析已经证实，知识图谱能够根据用户选择，将相关实体和实体间关系及相应属性实现可视化呈现，并且可以通过选择相应的实体或关系，灵活地组织可视化信息。知识图谱优秀的可视化功能，能够灵活方便地从海量数据中提取和呈现知识及其关联，有助于大数据背景下的领域知识分析与服务。

(3) 知识图谱能够方便存储和快速提取知识关系并进行推理。知识图谱能够方便地存储真实数据，灵活的查询指令也可以快速提取相关的信息。知识图谱同时还具有一定的知识推理功能。研究中的多维关系分析发现，通过对知识图谱内实体间关系的分析，可以推理和挖掘更多的潜在信息。因此，知识图谱不但在多维信息分析中具有优势，而且对于知识推理与知识挖掘同样具有优势。

本研究旨在通过构建领域知识图谱和不同维度层面的实体间关系分析，分析和探索知识图谱在领域知识多维分析中的可能途径。研究中也存在不足之处，限于本文目的和篇幅原因，研究中仅抽取部分数据和有限的实体及关系，没有基于全部数据开展更深入的具体分析与挖掘。未来的研究工作将基于知识图谱在领域知识多维分析中的优势展开更深入的研究。

参考文献

- [1] HINTON G E, SALAKHUTDINOV R R. Reducing the dimensionality of data with neural networks [J]. Science, 2006, 5786 (313): 504-507.
- [2] SINGHAL A. Introducing the Knowledge Graph: Things, not Strings [EB/OL]. [2019-01-08]. <http://googleblog.blogspot.ie/2012/05/introducing-knowledgegraph-things-not.html>.
- [3] FOUNTAIN J F. Dewey decimal classification, 21st Edition [J]. Library Collections Acquisitions & Technical Services, 2001, 25 (2): 236-236.
- [4] USCHOLD M, GRUNINGER M. Ontologies: Principles,

- methods and applications [J]. The Knowledge Engineering Review, 1996, 11 (2): 93-136.
- [5] SOERGEL D. The rise of ontologies or the reinvention of classification [J]. Journal of the American Society for Information Science, 1999, 50 (12): 1119-1120.
- [6] HAUTAMÄKI A. A conceptual space approach to semantic network [J]. Computer & Mathematics with Applications, 1992, 23 (23): 517-525.
- [7] 滕广青, 毕强. 国外本体协调研究前沿进展及热点分析 [J]. 中国图书馆学报, 2012, 38 (1): 113-120.
- [8] 孙雨生, 白璧娇. 国内基于本体的知识服务研究进展: 核心内容 [J]. 图书馆学研究, 2018 (16): 9-16.
- [9] MARTIN R, BERGSTROM C T, FABIO R. Mapping change in large networks [J]. PLoS ONE, 2010, 5 (1): e8694.
- [10] PHAM M C, KLAMMA R, JARKE M. Development of computer science disciplines: A social network analysis approach [J]. Social Network Analysis and Mining, 2011, 1 (4): 321-340.
- [11] MA F C, LI Y T. Utilising social network analysis to study the characteristics and functions of the co-occurrence network of online tags [J]. Online Information Review, 2014, 38 (2): 232-247.
- [12] 马费成, 刘向. 科学知识网络的演化模型 [J]. 系统工程理论与实践, 2013, 33 (2): 437-443.
- [13] 安宁, 滕广青, 白淑春, 等. 基于网络Hub的领域核心知识涌现研究 [J]. 图书情报工作, 2017, 61 (18): 98-106.
- [14] WENG L, MENCZER F. Topicality and impact in social media: Diverse messages, focused messengers [J]. PLoS ONE, 2015, 10 (2): e0118410.
- [15] MA Z, LEE Y, YU K H. Ten years of conflict management studies: Themes, concepts and relationships [J]. International Journal of Conflict, 2008, 19 (3): 234-248.
- [16] 李纲, 李春雅, 李翔. 基于社会网络分析的科研团队发现研究 [J]. 图书情报工作, 2014, 58 (7): 63-70, 82.
- [17] 安宁, 滕广青, 徐汉青, 等. 知识群落演化与知识传递模式研究 [J]. 图书馆学研究, 2018 (21): 76-85.
- [18] 孙雨生, 常凯月, 朱礼军. 大规模知识图谱及其应用研究 [J]. 情报理论与实践, 2018, 41 (11): 138-143.
- [19] 王颖, 钱力, 谢靖, 等. 科技大数据知识图谱构建模型与方法研究 [J]. 数据分析与知识发现, 2019, 3 (1): 15-26.
- [20] 徐雷, 潘珺. 知识网络等相关概念比较分析 [J]. 情报科学, 2017, 35 (12): 10-15.
- [21] COWIE J, LEHNERT W. Information extraction [J]. Communications of the ACM, 1996, 39 (1): 80-91.

作者简介

王思茗, 女, 1996年生, 硕士研究生, 研究方向: 知识组织与信息分析。

孙熊兰, 女, 1996年生, 硕士研究生, 研究方向: 知识组织与信息分析。

滕广青, 男, 1970年生, 教授, 博士生导师, 研究方向: 知识组织与信息分析, E-mail: tengguangqing@163.com。

叶心, 女, 1970年生, 副研究馆员, 研究方向: 知识组织与服务。

栾宇, 男, 1988年生, 硕士研究生, 研究方向: 信息分析与知识发现。

Application of Knowledge Graph in Multidimensional Analysis of Domain Knowledge

WANG SiMing¹ SUN XiongLan¹ TENG GuangQing¹ YE Xin² LUAN Yu¹

(1. School of Information Science and Technology, Northeast Normal University, Changchun 130117, China;

2. Changchun Library, Changchun 130021, China)

Abstract: Knowledge graph provide new ideas and methods for solving realistic problems in many fields, to help reveal and analyze that the domain knowledge and its associated relationships. The research work uses graphic database technology to construct the domain knowledge graph. From the perspective of simple entities relationship and multi-dimensional complex relationships, the application of knowledge graph in multi-dimensional analysis of domain knowledge is explored. The results show that knowledge graph can organize and analyze domain knowledge and knowledge relationship from the point of view of multi-dimension, it also can flexible visualize the domain knowledge and its relationships. Moreover, knowledge graph makes it easy to store and rapidly extract knowledge relationships and do inference researched.

Keywords: Knowledge Graph; Entities Relationship; Domain Knowledge; Multidimensional Analysis

(收稿日期: 2019-02-06)