

中国传统音乐知识图谱建构研究

滕 腾

(集美大学 音乐学院, 福建 厦门 361021)

[摘要] 大数据时代, 数据数量与种类越来越多, 数据之间的关系越来越复杂, 数据的应用场景也从传统的统计分析, 到现在基于大数据的机器学习、知识图谱等。数据的深度挖掘使用、数据对业务的促进、数据在行业运营等方面均发挥着越来越重要的作用。因此, 本研究通过对传统音乐的元数据标注与叙词管理, 叙词关系案例阐释, 并借助 CiteSpace、Graphviz 软件对传统音乐进行引文可视化和图结构分析。希望通过挖掘海量传统音乐文献数据背后隐含的知识和技术信息、文献之间关联信息, 即“知识图谱”, 研究分析传统音乐学科发展趋势, 洞见传统音乐学术社交网络与全球视野下其深层规律与动态, 以此促进传统音乐的学术生态发展。

[关键词] 知识图谱; 中国传统音乐; 叙词管理; CiteSpace; Graphviz

[中图分类号] J60

[文献标识码] A

[文章编号] 1008-889X (2024) 01-0075-09

回溯二千多年前, 古希腊哲学三贤开创了伟大的逻辑思维时代, 其思想直接或间接地影响了从符号主义到知识工程再到知识图谱的历史延承与发展^[1]。知识图谱 (Knowledge Graph) 的早期理念源于“万维网之父”蒂姆·伯纳斯·李 (Tim Berners-Lee) 关于语义网 (The Semantic Web) 的设想, 旨在采用图结构 (Graph Structure) 来建模和记录世界万物之间的关联关系和知识, 使互联网上的信息更加有结构、可理解和可操作, 从而使机器能够理解和响应用户查询, 提供更加精准和相关的结果。自 2012 年 Google 发布知识图谱以来, 知识图谱技术飞速发展, 其理论体系日趋完善, 相关技术已经在搜索引擎、智能问答、语言理解、推荐计算、大数据决策分析等众多领域得到广泛的实际应用。知识图谱旨在从数据中识别、发现和推断事物之间的复杂关系, 是事物关系的可计算模型。知识图谱代表万维网发展的一种理想, 也是人工智能与互联网深度融合的产物^[2]。

人类已进入人工智能时代。人工智能在众多方面都发挥了重要作用, 也加速了其在全球范围

内的发展。所谓“智”指的是智慧, 表现为知识; “能”指的是解决问题的能力, 主要表现为基于知识的推理能力和运用知识解决问题的能力。因此, 知识是智能的基础和核心, 知识的建模、表示、获取、融合、推理、赋能等是人工智能及其相关应用的核心问题, 知识图谱是其有效的技术实现手段。其技术实施的过程中, 需要根据知识的内在结构与关联数据对事物进行元数据、叙词表的标注与管理。

一、叙词表管理中的传统音乐

叙词表^[3]¹⁶³是一种知识结构, 能够用来揭示概念之间的关系, 帮助用户理解相应知识领域的主题关系; 是一种词汇控制工具, 使得标引人员和用户能够使用统一的词语进行标引和检索; 是一种术语参照系统, 通过参考它, 用户在检索知识的过程中能够随时扩大或缩小检索范围; 是一种动态的词汇集合, 任何新生成的概念或术语, 都能够以用户可以理解的方式纳入词汇系统中相应的位置^[4]⁹⁻¹⁰。

[收稿日期] 2023-09-18

[基金项目] 福建省社会科学基金项目 (FJ2021B197); 福建省教育厅项目 (JAS21111); 全国社会科学基金教育学一般课题 (BLA230248)

[作者简介] 滕腾 (1981—), 女, 内蒙古赤峰人, 副教授, 博士, 主要从事传统音乐研究。

(一) 关于叙词、叙词表的定义与国内外发展概况

1. 叙词与叙词表的定义。叙词(Thesaurus)是一种重要的信息检索工具,通常用于图书馆、档案馆、数据库等信息存储和检索系统。它们包含了一组经过仔细选择的词和短语,这些词和短语用于表示信息资源中的主题或概念。叙词的主要特性包括概念性、描述性、组配性、动态性、直观性以及语义关联性。从信息检索的角度来说,叙词是作为一种检索语言使用,所以,在一些场合也称为叙词法^{[4]1}。所谓叙词表(Thesaurus),根据《汉语叙词表编制规则》(GB13190-91)中的定义,它是“一种将文献、标引人员或用户的自然语言转换成规范化语言的术语控制工具,是概括各部门或某一学科领域并由词义相关、族性相关的术语组成的可以不断补充的规范化的词典”^{[5]1}。有些人也将叙词表称为主题词表,为了方便对其进行定义,专家学者通常选取某一角度作为切入点,如功能、结构等。从功能角度看,叙词表的主要任务是将标引人员或用户使用的自然语言转换成规范化的系统语言。这种转换过程可以提高信息检索的准确性和效率,因为它确保了同一概念的不同说法都被统一到一个标准的表达方式上。从结构角度看,叙词表是一个反映特定知识领域的,由词义相关、语义相关的术语组成的规范化词典。它不仅包含了这个领域的主要术语,还定义了这些术语之间的关系。由于知识领域是随着新的研究和发现不断发展的,因此叙词表也需要不断地进行更新和补充^{[4]9}。

2. 关于国内外叙词表的发展概况。国外叙词法的研究开展得较早,可以追溯到19世纪下半叶。在一些欧美国家,叙词表被作为一种主要的检索语言。而我国叙词表的发展历史可以追溯到古代“用韵以统字,用字以系事”。我国明代永乐年间(1403—1408)编制的《永乐大典》,就是带有主题法性质的检索工具。1934年诞生了我国第一部主题法词表是沈祖荣的《标题总录》^{[4]35},我国叙词表的发展主要体现在航空、电子、医学、环境、军事等诸多领域。但是,有关音乐方面的叙词表目前只有中国艺术研究院编著的《艺术科学叙词表》^[6]和赵锦、戴维民主编的《音像资料叙词表》这两部^[7]。

(二) 关于叙词表中叙词之间的关系与示例

叙词表中叙词之间主要有3类关系、6种含义。3类关系分别是:等同关系、属分关系、相关关系。6种含义分别包括:用、代、分、属、族、参,用符号表示分别为:Y、D、F、S、Z、C。

1. 叙词表中叙词之间的关系

(1) 等同关系,也称同一关系、用代关系,使用Y、D两种符号。Y用于非叙词之下,用来指向对应的叙词;D用于正式的叙词之下,用来指向被替代的非叙词。比如,在音乐中俗称的“钢伴”,这一称呼是仅限在音乐领域内的约定俗成,但是对于正式叙词而言,显而易见“钢伴”这个词属于非正式叙词。所以,此处将用这种形式表达它们直接的形式:

钢琴	伴奏钢伴
D 钢伴	Y 钢琴伴奏
民间艺术家	民间艺人
D 民间艺人	Y 民间艺术家

等同关系除用来提示同义词、准同义词之间的关系外,还可以适用于组代关系,以提高标引的一致性,降低标引的难度。等同关系的优点在于:无论使用同义词中的哪一个词语进行搜索,搜索的结果均相同。也就是说,不会因为用户的用词习惯以及对相关领域概念、术语的陌生而导致检索不到需要的信息,从而使得数据库使用体验不佳。

(2) 等级关系,也称属分关系,用F、S、Z符号表示。F为分项指示符,表示该项叙词为款目词(Entry terms),是指在款目单元的各种成分中,一款目中起标目作用的词称为款目词,款目词是叙词款目和非叙词款目的核心和排检的依据;S为属项指示符,表示该项叙词为款目词的上位词;Z为首词符号,表示该项叙词为族首词,亦即一组等级关系的叙词中外延最宽的词。等级关系是描述传统音乐概念之间的层级关系的必要手段之一。等级关系主要用于下列3种叙词之间:(1)具有属种关系的叙词之间。(2)具有整部关系的叙词之间,包括地理位置中的整部关系、人体系统与器官以及某些事物中的整部关系、组织机构及其分支中的整部关系。(3)用于部分集合概念与个体概念之间。本研究以传统

音乐作为实例，对叙词管理做进一步解读。

(3) 相关关系，亦称类缘关系，用 C 符号加以连接，与属分关系的处理相同。除了等同关系与等级关系以外的都可以做相关关系的处理。比如，近义关系的叙词之间，反对或矛盾关系的叙词之间，事物与其部分、成分的叙词之间，因果关系的叙词之间，各种事物与材料、性质的叙词之间，事物、操作与其工具的叙词之间，职业与职业工作者的叙词之间，学科或研究领域与其对象的叙词之间。值得注意的是，相关关系的类型繁多，很难将其一一列举。对相关关系的处理一般要有一定的控制，通常只有一词“作为标引词与另一词具有较强的提示作用”时，才作为相关关系处理^{[4]43}。

总之，传统音乐叙词表试图提供一种方式，使得某一主题领域的术语能够以规范的方式加以应用。制定传统音乐叙词表，它能够有效梳理传统音乐的相关概念和术语，使得大量的传统音乐文献能够得到系统的归纳与整理，从而为传统音乐本体构建奠定基础。传统音乐以其元数据标注与叙词管理的方式纳入到数字人文系统，透过 CiteSpace 引文可视化分析工具厘清传统音乐脉络与学术生态。

二、中国传统音乐“可视化”与案例分析

(一) 传统音乐 CiteSpace 引文可视化分析

引文可视化分析软件 CiteSpace 通过可视化的手段来呈现引文网络，揭示文章、作者、期刊、机构、国家和地区之间的关系以及研究主题的演变过程^{[8]27-31}。因此，也将通过此类方法分析得到的可视化图形称为“科学知识图谱”。CiteSpace 软件主要包括共引分析、共现分析、热点和趋势分析、网络可视化与突破性研究识别等功能。笔者以传统音乐为关键词收集数据进行共现分析与合作网络分析（见图 1、图 2）。

本研究以 CSSCI 作为文献数据检索来源，通过高级检索功能，使用主题词“传统音乐”，时间过滤器选择 1999 年至 2020 年 8 月 10 日，检索条件设置为“精确”，在中国知网（CNKI）得到检索结果 661 条。文献信息包括题目、作者、摘要、关键词、机构等字段。笔者将这些文献数据，输入 CiteSpace 可视化分析软件，对 1999 年至 2020 年 8 月 10 日期间我国传统音乐的传承与发展进行分析，将传统音乐的历史演变与发展动态等方面的内容完整地、科学地予以展现（见表 1、表 2）。

表 1 传统音乐文献作者与机构

次数	发文作者机构	首现年份	次数	发文作者机构	首现年份
48	中央音乐学院	1999	4	首都师范大学	2007
44	中国音乐学院	1999	4	西北师范大学	2002
38	中国艺术研究院	1999	4	西藏大学	2007
31	上海音乐学院	1998	3	安徽师范大学	2007
27	福建师范大学	2004	3	东北师范大学	2005
23	武汉音乐学院	1999	3	河南大学	2004
21	南京艺术学院	2005	3	鲁东大学	2011
9	中国传媒大学	2012	3	人民音乐杂志社	2007
7	南京师范大学	2002	3	山东艺术学院	2012
7	山东师范大学	2006	3	山西大学	2000
7	天津音乐学院	1998	3	皖西学院	2016
7	西安音乐学院	2012	3	西南大学	2011
7	中央民族大学	2007	3	浙江艺术职业学院	2008

表 2 传统音乐文献作者首现年份与次数

次数	作者	首现年份	次数	作者	首现年份	次数	作者	首现年份	次数	作者	首现年份
22	王耀华	1999	6	许 璐	2008	3	蓝雪霏	2005	3	向 文	2011
17	蔡际洲	2003	5	杨 红	1999	3	陈应时	2002	3	徐天祥	2007
14	项 阳	2001	5	张君仁	2004	3	陈 鹏	2017	3	刘富琳	1999
14	乔建中	1998	5	施 咏	2005	3	毛继增	2006	3	马 达	2011
11	杜亚雄	1998	5	刘承华	2002	3	张振涛	1998	3	博特乐图	2011
10	萧 梅	2003	5	王 州	2007	3	田 青	1999	3	杨曦帆	2010
10	杨民康	2005	4	徐元勇	2001	3	郭树芸	2011	3	程晖晖	2005
8	张伯瑜	2000	4	孙星群	1998	3	杨善武	1999	2	李彦荣	2002
7	张天彤	2006	4	张玉榛	2010	3	袁 玥	2003	2	孟凡玉	2008
7	管建华	1999	4	洛 秦	2009	3	赵书峰	2011	2	田耀农	2000
7	冯光钰	1998	4	伍国栋	2006	3	张 宁	2008	2	薛艺兵	2008
7	樊祖荫	2000	4	姚艺君	2003	3	赵 敏	2016	2	黄 婉	2008
6	蒲亨强	2003	4	刘再生	1999	3	王安潮	2010	2	黄 虎	2004
6	崔学荣	2009	4	李延红	2006	3	李 昕	1999	2	吴 凡	2004

1. 传统音乐研究的作者 CiteSpace 引文网络分析。中国知网的文献检索中，每一篇文献检索后，都会附有一个引文网络。所谓引文网络就是由文献间引用和被引用的关系构成的集合。这种集合以“节点文献”作为开端，分别以“二级参考文献”“参考文献”“引证文献”“二级引证文献”“共引文献”“同被引文献”构成集合。

透过 CNKI 的引文网络，笔者试图对传统音乐研究的作者与研究机构引文网络进行分析。由于研究机构与该研究的作者单位是一一对应的，又因为大部分人文社科类文献的作者没有合作的现象，其作者单位之间的合作网络线条少之又少。本研究以作者的文献的引文网络做具体论述（见图 1）。

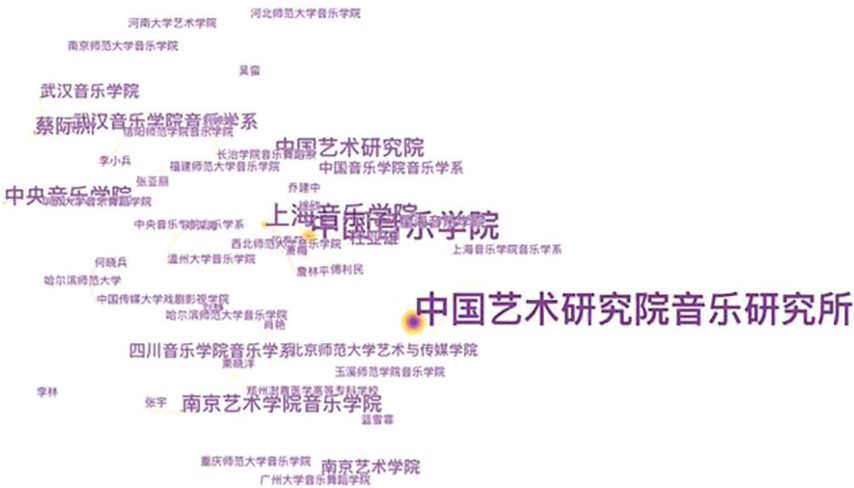


图 1 传统音乐文献的作者研究机构 CiteSpace 引文网络分析

在 661 篇论文中，作者之间的引用与被引关系，可以揭示作者及其论文的研究影响力与重要性以及作者和团队之间的学科交叉与合作。此处以 661 篇文献为数据，节点选择作者（Author），按每一年归类分析（时间切片设置为 1），选取

头部 50% 接点（top N 参数 50%），使用 CiteSpace 软件分析，得到作者共现知识图谱（见图 2），一共有 118 位作者（ $N=118$ ），作者间关联数为 1 006（ $E=1\ 006$ ），即有 118 个作者结点和 1 006 条作者之间的连线。

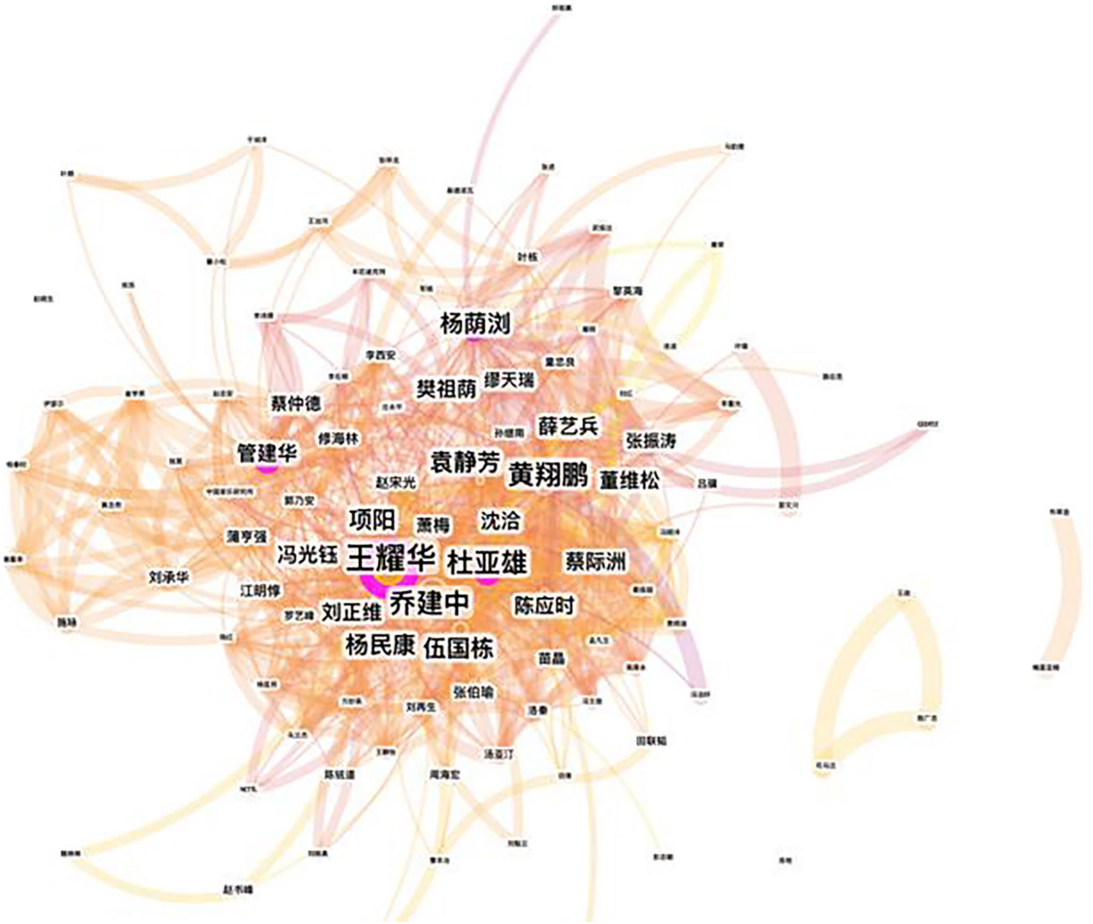


图 2 传统音乐文献作者之间的引用与被引关系 CiteSpace 分析

2. 传统音乐研究的热点分析。为更直观地展示高频关键词，本研究以关键词频次大于 4 次作为筛选条件，一共得到 25 个主题关键词，分别为传统音乐、中国传统音乐、民族音乐学、音乐文化、音乐教育、传统音乐文化、民间音乐、非物质文化遗产、民族音乐、音乐艺术、传承、学科建设、中国传统音乐理论、少数民族、中国传统音乐文化、音乐研究、董维松、民族传统音乐、音乐人类学、中国传统音乐研究、王耀华、发展、中国传统音乐结构、中国传统音乐学、传统音乐理论。中介中心度大于 0.12 的 7 个关键

词，分别为传统音乐、中国传统音乐、民族音乐学、音乐文化、音乐教育、民间音乐、传承（具体见表 3）。综合分析图 3 和表 3，发现传统音乐研究的动向集中在传统音乐、中国传统音乐、民族音乐学、音乐文化等几大主题。
论文的关键词是一组用于概括和描述论文主题、内容和关注点的术语或短语。关键词应该具有代表性，涵盖论文的核心概念，并且能够准确地反映研究领域和内容。此处以 661 个样本文献为源数据，节点类型选择关键词（keyword），按每一年归类分析（时间切片设置为 1），选取头

数 206 ($E = 206$), 即有 128 个关键词结点和 206 条关键词之间的连线。

表3 传统音乐文献热点分析-关键词突发指数与出现频次

频次	突发指数	中介中心度	关键词	频次	突发指数	中介中心度	关键词
202		0.38	传统音乐	6		0.01	中国传统音乐理论
124		0.51	中国传统音乐	6		0	少数民族
38	6.58	0.21	民族音乐学	6		0	中国传统音乐文化
38	4.82	0.17	音乐文化	6		0.04	音乐研究
23		0.25	音乐教育	6		0	董维松
21	5.5	0.04	传统音乐文化	5		0.03	民族传统音乐
19		0.12	民间音乐	5		0.01	音乐人类学
14		0	非物质文化遗产	4		0	中国传统音乐研究
12	3.86	0.04	民族音乐	4		0	王耀华
11	4.45	0.03	音乐艺术	4		0	发展
10	5.32	0.2	传承	4		0	中国传统音乐结构
8		0.01	学科建设	4		0	中国传统音乐学
7	4.46	0.03	少数民族传统音乐	4		0.02	传统音乐理论



图3 传统音乐文献-主题词分析

款开源的图形可视化软件，它能够绘制关系图、流程图、数据结构图、网络图、树形图等各种图

形，在机器学习、网络、软件工程、数据库、网页设计等技术领域有着重要的应用。因此，本研究针对南音传承人与川剧艺术家的生态发展，通过 Graphviz 软件绘制其内在的关系^{[9]14}（见图 4、图 5）。

1. 通过南音资源数据抽取，以南音传承人的多重脉络，构建南音资源中关于人、物、组织、事件等实体及其之间的关联关系，按照不同维度组织数据资源，为南音资源的深度挖掘和揭示内在规律提供技术支持。图 4 以南音传承人为

例，通过本体与元数据标注，运用语义分析、智能检索、叙词等技术，构建南音传承人为中心的社交网络，包括师徒、同事、同学、校友、夫妻、亲缘等关系实体之间部分关联的知识图谱可视化结果。通过图形可视化分析也可清晰地展示出某位特定南音传承人的各类相关信息等^{[10]53-60}。构建南音关联数据知识图谱，以可视化的形式展现数字化传统音乐关系脉络，拓展开发和创新利用传统音乐资源，实现传统音乐资源承传与使用的延续（见图 4）。

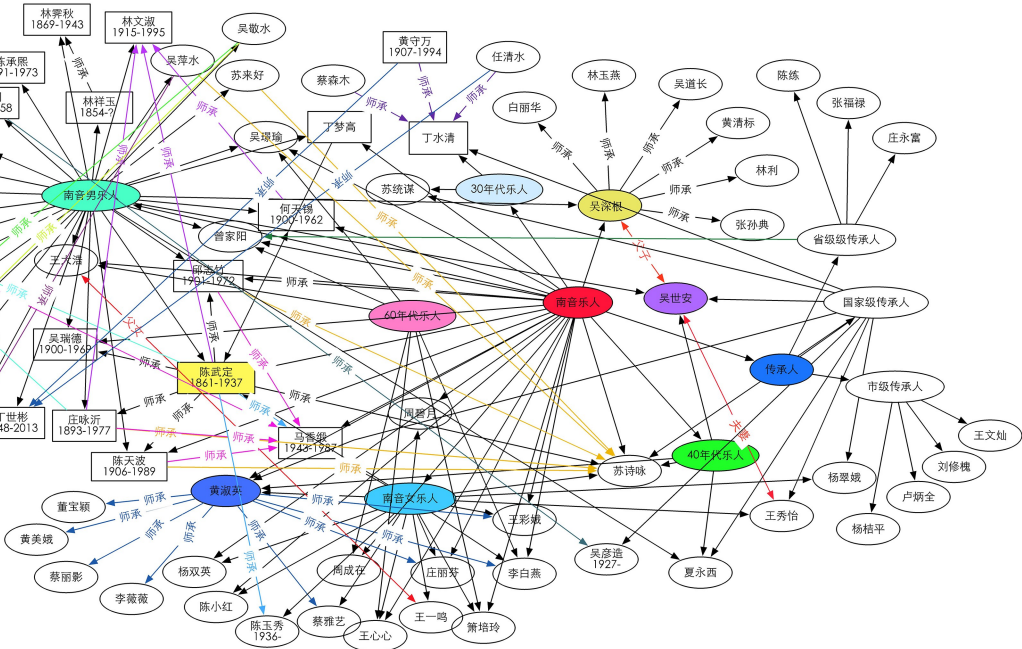


图 4 南音传承人知识图谱 Graphviz 可视化结果

2. 笔者在中国知网以“口述史”作为关键词进行查阅，其根据智能自助标注的方式，发现单独的分组“川剧艺术家”，包含论文 19 篇。分别是：万平、冯乃光的《川剧艺术家口述史（梅花奖得主卷）之沈铁梅篇》，程建、忠万平的《川剧老艺术家口述史（四川卷续）之魏益新篇》，万平、刘咏涛的《川剧老艺术家口述史（四川卷）之余开元篇》，胡艳、万平的《川剧艺术家口述史（四川卷）之周正荣篇》，万平、严铭、冯乃光的《川剧老艺术家口述史（四川卷）之汪洋篇》，万平、冯乃光、李启后的《川剧艺术家口述史（梅花奖得主卷）之马文锦篇》，万平、冯乃光、李启后的《川剧艺术家口

述史（梅花奖得主卷）之孙勇波篇》，严铭、万平、张萍的《川剧艺术家口述史（梅花奖得主卷）之崔光丽篇》，严铭、万平、冯乃光的《川剧艺术家口述史（梅花奖得主卷）之刘谊篇》，万平、叶俊莉、李启后的《川剧艺术家口述史（梅花奖得主卷）之吴熙篇》，万平、叶俊莉的《川剧艺术家口述史（梅花奖得主卷）之蒋淑梅篇》，万平、叶俊莉、郭奥的《川剧艺术家口述史（梅花奖得主卷）之陈智林篇》，万平、叶俊莉、李启后的《川剧艺术家口述史（梅花奖得主卷）之黄荣华篇》，胡艳、万平、林琳的《追寻戏曲艺术传统与现代审美理念的契合——川剧老艺术家口述史（四川卷）之任庭芳篇》，刘咏

涛、万平、林琳的《川剧老艺术家口述史 (四川卷续) 之许明耻篇》, 万平、马丽的《川剧老艺术家口述史 (成都卷) 之竞艳篇》, 殷晓燕、吴念群的《川剧老艺术家口述史 (成都卷) 之唐思敏篇》, 李轼华、蔡少波的《川剧老艺术家口述史 (成都卷) 之阳荣华篇》, 胡艳、林琳的《川剧老艺术家口述史 (四川卷) 之左清飞篇》。通过阅读这 19 篇期刊论文, 笔者获得诸多川剧艺术家人物与曲目信息。

关于川剧艺术家人物与剧目 (图 5 是以川剧艺术家为起点), 均是从各类文献中提取到零散、孤立的信息, 难以加以利用, 并最终形成系统的知识。如何让这些信息形成有机的整体, 从而有利于深入地研究工作, 可以借助本体理论,

将非结构化的文本信息中蕴含的结构化知识提取、整理出来, 放入数据库中。音乐本体 (music ontology) 是一个以信息科学本体论原理为基础, 旨在定义音乐领域并支持音乐知识推理的词汇表。它提供结构化的共享词表和术语集, 用于描述音乐信息, 例如艺术家、专辑、音轨、演出、编曲等, 并连接广泛的音乐相关数据, 最终建立一个方便分类和使用音乐信息的语义网。通过使用音乐本体, 可以实现对音乐领域知识的语义建模和推理。它可以帮助计算机系统理解音乐数据, 并进行智能的音乐分析、搜索、推荐等任务。此外, 基于音乐本体的语义网还可以促进音乐领域的协作和数据共享, 为音乐研究、音乐教育等领域提供更多可能性。

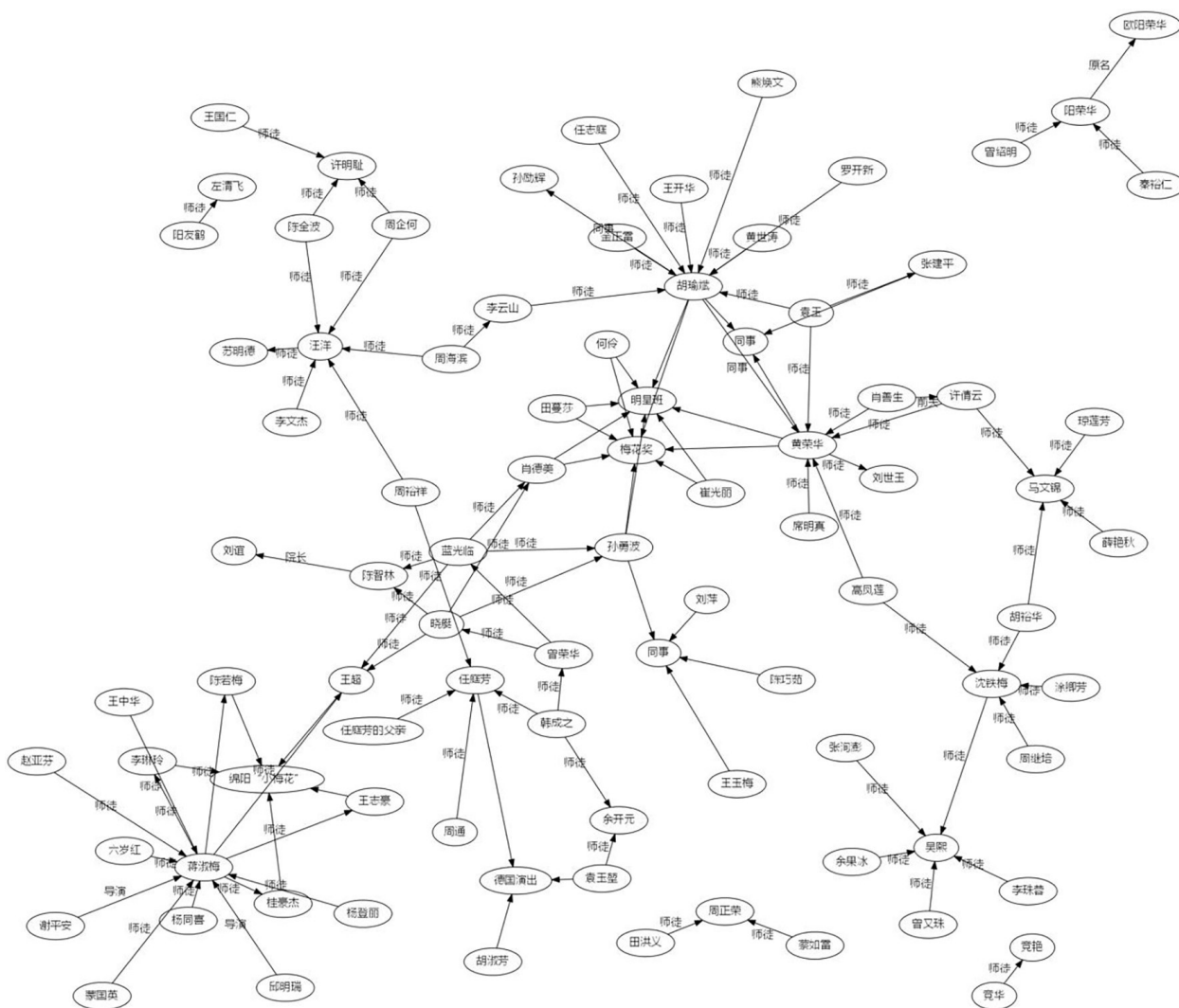


图 5 川剧艺术家 Graphviz 图结构可视化结果

[参考文献]

三、结 语

本研究试图采用知识图谱的方法,有效帮助研究者实现对传统音乐数据的识别、厘清、挖掘、推断。最终对中国传统音乐的形态学、美学、史学、乐种学、结构学、乐器学、乐谱学、社会学、心理学和演奏演唱理论等,以及传统音乐的教学模式与方法、教育机构的组成、课程与教材设置、创作与表演形式、传播与传承体系、代表人物等内在的复杂、系统、结构关系进行话语体系整体建构。知识图谱属于应用型学科的范畴,如何解决传统音乐中杂糅的脉络与谱系的多重现实问题,显然是一项大规模知识工程,本研究旨在为传统音乐历史延承与发展的知识梳理开辟一条新路径。

- [1] 肖仰华等. 知识图谱: 概念与技术 [M]. 北京: 电子工业出版社, 2019.
- [2] 王昊奋, 漆桂林, 陈华钧. 知识图谱: 方法、实践与应用 [M]. 北京: 电子工业出版社, 2019.
- [3] 滕腾. 蒙古族音乐数据库的设计与制作 [M]. 北京: 文化艺术出版社, 2021: 163.
- [4] 邓仲华, 赵又霖, 黎春兰, 等. 多语种叙词本体 [M]. 武汉: 武汉大学出版社, 2011.
- [5] 国家技术监督局. 汉语叙词表编制规则 [S]. 1991: 1.
- [6] 中国艺术研究院. 艺术科学叙词表 [M]. 北京: 文化艺术出版社, 1992.
- [7] 赵锦, 戴维民. 音像资料叙词表 [M]. 北京: 中国计量出版社, 1995.
- [8] 滕腾. 数字时代下音乐教育学术社交功能研究 [J]. 中国音乐教育, 2020 (3): 27-31.
- [9] 焦天源. 基于算子的代码自动生成技术研究 with 实现 [D]. 郑州: 郑州大学, 2021.
- [10] 舒忠梅. 数字人文背景下的档案知识图谱构建研究 [J]. 山西档案, 2020 (2): 53-60.

Research on the Construction of Traditional Chinese Music Knowledge Graph

TENG Teng

(Music College, Jimei University, Xiamen 361021, China)

Abstract: In the era of big data, the amount and the type of data are increasing, and the relationships between data are becoming more and more complex. The application scenarios of data have also changed from traditional statistical analysis to machine learning and knowledge graphs based on big data. The in-depth mining and use of data are playing an increasingly important role in all aspects of industry operations. By using metadata annotation and thesaurus management of traditional music, the paper explains the thesaurus relationship cases, and employs CiteSpace and Graphviz software to conduct citation visualization and graph structure analysis of traditional music. It is hoped that by mining the implicit knowledge and technical information behind the massive traditional music literature data, as well as the associated information between documents, a “knowledge graph” can be constructed to research and analyze the development trends and predictions of traditional music disciplines, and to gain insights into the deep rules and dynamics of traditional music academic social networks and global perspectives, thereby to promote the development of the academic ecology of traditional music.

Key words: knowledge graph; traditional Chinese music; thesaurus management; CiteSpace; Graphviz

(责任编辑 张永汀)