**备注提示：内文体例与格式**

**章 第一章 1** 每一章之间有两行空距

……

代码块通常须缩进展示，并在上下方与正文的其他部分有一行的空距

**基于中华传统音乐馆藏资源的“知识库”建设**

**和“知识地图”研究（CTM 2.0）**

（2.0 目录及重点批注）

**1 中华传统音乐文化知识库一期及相关研究回溯**

（1）现存的问题：语义网站点未真正打通、后端知识推理尚未激活、可视化使用平台效果不足、未开通图书馆员维护平台，等等

（2）重点回顾二期迭代方案

（3）对一期技术选型优势的介绍：即灵活性和可拓展性；知识图谱所代表的符号主义AI+大语言模型所代表的连接主义AI；产品性优势的介绍，比如，曲目的演出的细颗粒度、元数据+数据——双层架构的产品定位

（4）LinkedMusic研究方案的启发（这个方案主要就是在线音乐数据库资源共建技术方案）

（5）根本问题意识：有了大模型，我们为什么仍要迂回地做知识图谱？

**2 中华传统音乐文化知识库二期开发成果述介与分析**

已解决前述“现存的问题”

**2.1 增补数据与检索分析**

2.1.1 基于大模型的（乐器）知识图谱自动化智能知识抽取研究

2.1.2 增补数据与检索示例

首先，提出“语义网”作为数据库，在检索方面的重要优势，在图书馆业务环境下的关联数据发布等。然后：

2.1.2.1 新版本体(2.0)的简介（若干版本；看看能否植入SHACL及对VOID的数据分析）

2.1.2.2 东方乐器数据（+图片）的植入、局内人视角的检索案例

（空节点）

2.1.2.3 关联Wikidata（维基百科对应的知识图谱）的示例（乐种、乐器）与跨数据库检索示例

2.1.2.4 关联中国音乐学院实体乐器博物馆乐器编目

（5）乐器分类法（霍萨分类法+“概念->单件”垂直分类体系）

（6）激活推理机的数据增补（含owl:sameAs所链接的语义网站点）

推理机：……探索其他高级推理的效用，如2.3.1.1用到对称关系

（7）新增中华传统音乐馆藏资源链接等

**2.2 可视化与后端平台功能介绍与案例分析**

2.2.1 前端可视化网络分析功能（2.0版）

如路径查询等

2.2.2 后端（如图书馆员维护平台）

增、删、改、查、术语维护、关联数据发布平台

**2.3 “知识地图”案例拓展**

增补地理信息数据，可用以检索我国任意一级行政单位（村、寨除外）的乐种、乐器等信息

2.3.1 检索案例分析

2.3.1.1 地域毗邻关系检索：检索相邻地域的乐种分布

2.3.1.2 地理坐标距离检索：精确检索乐种间的地理距离、特定地域范围的乐种、资源等

2.3.2 知识地图研究意义小结

基于规则的数据补齐[[1]](#footnote-0)：目前只做了一部分……实现了，可以自动化地查询中国任意地域的乐种、乐器等分布信息，并比照中华传统音乐资源库的收录情况，回答这样的问题：某地有什么样的乐种、乐器，在我馆是否有收藏……

**3 基于大语言模型的知识库智能检索系统（后端）开发**

**3.1 研究的问题、意义与背景综述**

知识库建设最终的目的是服务于信息检索。然而，一方面，网站中所涉的检索技术始终无法脱离页面设计的局限性；另一方面，前论述中绝大多数的信息检索都依赖于专业的检索语言SPARQL——这会让一般的音乐用户望而却步。而本章要解决的问题有二：

1. 将用户提出自然语言问题通过大语言模型自动转化为SPARQL，获取检索结果并做自动阐释：这样，用户就不一定得会用SPARQL才能做语义检索
2. 此外，启发（音乐数字人文研究者）掌握基于SPARQL的结构化检索语言，乃至熟悉音乐知识组织技术

**3.2 本体驱动的NLQ2SPARQL技术流分析**

3.2.1 本体编辑与激活基于本体的知识推理

3.2.2 本体切片

3.2.3 基于本体的实体（类和属性）提取

3.2.4 子图拼装

3.2.5 基于子图的NLQ2SPARQL

（NLQ即natural language question。即将自然语言问题转化成查询语句）

3.2.6 检索增强生成与基于本体的音乐推荐

例如，将本体本身看成一个知识地图，然后进行音乐推荐

**3.3 具体检索案例分析：中华传统音乐馆藏资源+知识库**

**4 总结与展望**

研究的不足与改进方向、未来规划：

1. 已有数据存在的问题：乐器数据有待优化，比如，存在中文IRI的情况；关于对原文本中，“或”关系或“且”关系的抽取存在问题

（1）优化NLQ2SPARQL工作流，强化检索增强生成的功能

（2）智能体开发：强化交互、启发用户与推荐系统

（3）细粒度检索：乐谱+音频

（4）多模态知识图谱\*

（5）本体映射与多数据库开放、关联、共享、共建（基于VOID生成Ontology的思想的联邦查询，LinkedMusic案例）

结语：大模型+知识库+提示词工程+RAG

**基于中华传统音乐馆藏资源的“知识库”建设**

**和“知识地图”研究**

（2.0 正文）

**1 中华传统音乐文化知识库一期及相关研究回溯**

中华传统音乐文化知识库是在中国音乐学院图书馆已建成的中华传统音乐文化资源库基础上，欲采用关联数据和知识图谱技术，进一步打造的检索更便利、内容更丰富，亦同时彰显相应技术思想乃至潜在的音乐信息管理和数字人文学科或专业特征的一个研发项目。之所以称其为知识库，在于强调：（1）不同于一般的资源库或数据库可能将关注点集中在可视听的资源本身，本项目充分关注及应用元数据技术和规范音乐元数据资源；（2）鉴于“知识的要义或在于知识结构”，本知识库将突出关联数据、知识图谱技术搭建的知识网络，将音乐文化的跨域关联通过网络关系和结构形式化的表达出来，从而为深度的知识检索和知识网络的广度扩展提供基础；（3）从技术要素来看，典型的知识库应当是在一般数据库的基础上，“植入”一个“元数据本体”（metadata ontology），从而为更科学、有效的知识组织、知识共享、共建、交换、复用，知识推理、语义检索、知识网络可视化、基于本体的智能问答系统开发，等等，提供坚实的基础。

就此，这样的知识库建设，在本研究的语境中，基本上等同于知识图谱。而知识图谱的技术选型之一是关联数据（linked data），这是尤其适用于图书馆业务环境的技术范式。本知识库建设目前存在两个阶段：一是由我校中国乐派研究院2021年进站博士后曹军军（合作导师：付晓东教授）在我馆馆长、技术部、数字人文工作坊等部门、团队的支持下，完成的一期（2022年1月-2024年2月）知识库项目，亦为博士后科研课题成果。如下简称1.0版，随附有研究报告《“数字人文”视角下中华传统音乐文化知识库的构建研究》（曹军军，2024；对该报告简称1.0版报告）。二是在我校图书馆校内科研专项支持下的二期建设，即对应2.0版，该版在1.0基础上做了进一步优化，本研究报告则是2.0版成果的集中体现。就此，对相应的研究背景、概念界定、研究设计、元数据本体工程、检索设计等，除了特定段落的强调之外，不做赘述介绍。

目前，就1.0版知识库和报告研发中存在的问题做基本重申介绍，尤其将重提在2.0版中解决的问题。同时，针对1.0版报告的5.2.1知识库二期迭代的建设方案纲要部分，本版直接延续了原改进计划（有改动）。具体而言：（1）探索与试验更灵活的知识图谱可视化方案。（2）检索功能的增强与优化，并涉及更灵活、开放的数据发布和数据互操作。（3）开发后端（元）数据维护平台，同时，进一步建成东方乐器知识库。（4）进一步体现知识推理功能，尤其是基于规则的知识推理。（5）其他探索，例如，开发智能问答系统，探索基于关联数据的多种数据库共享、共建。

本研究的技术基石是关联数据，它也是所谓语义网（semantic web）的实现基础。作为一门技术，它是否能和通常所谓的音乐科技形成交融或纳入音乐交叉学科领域，成为其有分量的一个部类？这是本项目团队于此予以探索的。然而，无独有偶，麦吉尔大学舒立克音乐学院音乐科技系的分布式数字音乐档案与图书馆实验室（DDMAL）于2022年发起了一个名为“关联音乐：互联全球在线音乐数据库与检索增强”项目。该项目同样采用了关联数据技术，对本项目研究有直接的借鉴或启发。就发展至目前阶段而言，其核心要素主要体现在两方面：第一，将已有的全球范围内重点数据库（例如音乐学科领域的RISM、音乐产业领域的MusicBrainz）与Wikidata形成数据映射并转换成linked data数据格式，存储在Open Link Virtuoso的“数据湖”（data lake）中。其对Wikidata的数据映射即类似于与Wikidata 的元数据本体词表形成映射，以达到各个数据库之间的相互贯通。第二，在相互贯通或数据湖形成的基础上，可执行灵活的数据检索，这种灵活性将超越任何数据库网站前端的界面功能限制；进而，还可实现跨多数据库的“联邦查询”。这些查询语言是linked data专属的SPARQL语言，在1.0版报告中已有充分的案例介绍。重点则是，如何利用大语言模型将用户的自然语言问题（Natural Language Question）转化成SPARQL，即深化探索NLQ2SPARQL的可行性路径，从而能极大程度降低音乐类用户在检索基于关联数据的知识库时的应用门槛（比如，不需要掌握本体技术、SPARQL检索技术），并增强灵活性。

在本项目团队与DDMAL实验室的对话沟通中，既形成共识，又各自持关于“音乐类图书馆资源共建”的底层技术保留意见。就共识而言，本项目高度认同NLQ2SPARQL研发的重大意义，并认为，这是符号主义人工智能和连接主义人工智能融合的关键突破口之一。所以，本项目将单辟一章，探索技术实现的基本可操作方案。就保留意见而言，LinkedMusic负责人Ichiro Fujinaga倾向于避开元数据本体技术的加持，譬如，将所有数据库向Wikidata映射就流露了这样的倾向，因为Wikidata没有一个“显式”的元数据本体。而本项目团队认为，尤其是从数字人文的视角来看，本体的重要性将体现在本研究报告的诸多方面，如知识组织、知识抽取、知识推理，也尤其体现在NLQ2SPARQL的一个本体驱动的提示词工程的技术分支上。未来或将用于本体映射，而根本地助力于“互联全球在线音乐数据库与检索增强”的目的。

本章的最后，须捎带回顾1.0版报告文（5.2.3的余论）末提出的一个命题，即知识图谱如何和大语言模型（深度学习）同时各自充当成为人工智能的左膀右臂。在本研究报告的下文，我们能随时看到这种左膀右臂的互助操作，欲以探求更具思想和实效的知识库研发和应用。

**2 中华传统音乐文化知识库二期开发成果述介与分析**

**2.1 增补数据与检索分析**

1.0版知识库的数据规模为，含近30万三元组数据，而2.0版截至2025年4月，含三元组数据42.4万条。

**2.1.1 基于大模型的（乐器）知识图谱自动化智能知识抽取研究**

与1.0版知识库的技术策略相比，该版进一步地借助大语言模型技术（chatGpt4o）抽取知识图谱数据，从而大程度地提高了数据收集的效率。以半结构化的“中国乐器大辞典”的前身文件作为语料库[[2]](#footnote-1)，通过python程序流实现如下步骤：

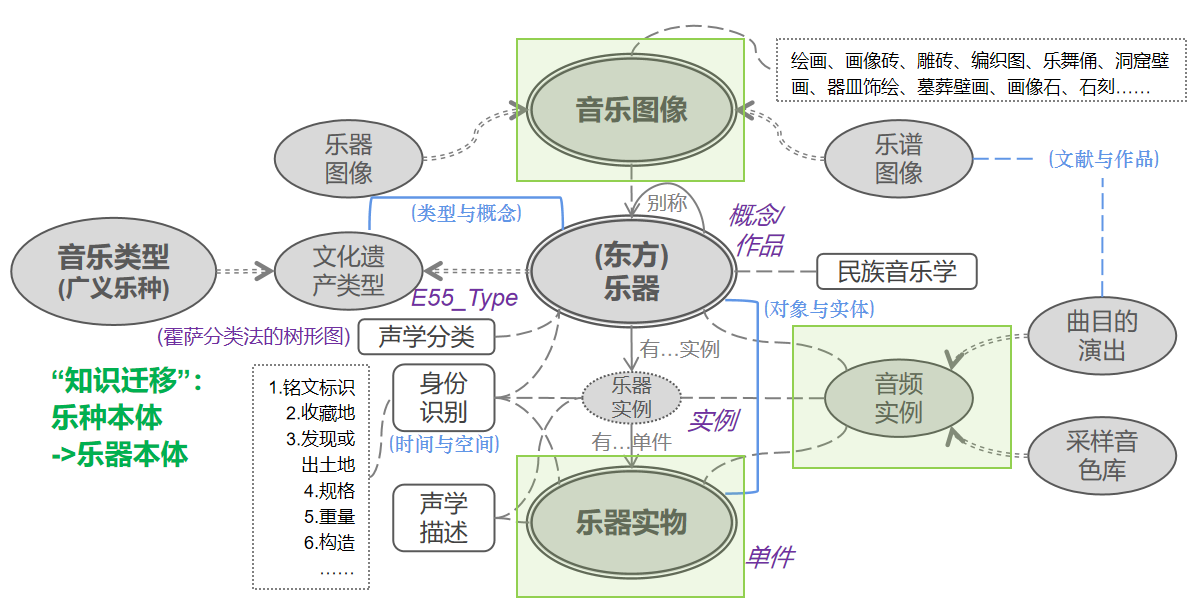
1. 准备一个本体片段（Instrument\_context\_A.ttl、Instrument\_context\_B.ttl，考虑到大语言模型的“消化”能力，在实际操作中，将本体片段分成A、B 两块，如下相应的任务也分为A、B两组），其包含要映射的元素，即类和属性，以进行知识提取。在这种情况下，本体是当作数据框架（schema）而用的。在准备阶段，选择相应的类和属性，以及它们的rdfs:label和rdfs:comment。 对于属性，重要的是要明确rdfs:domain和rdfs:range。此过程旨在告知LLM要用作从非结构化语料库中提取知识三元组的参考的知识组织结构。
2. 向LLM提供示例语料库（example\_A.csv、example\_B.csv）和相应的提取结果（aRightOutcome\_A.ttl、aRightOutcome\_B.ttl）。 来自步骤（1）的内容，结合详细的示例语料库及其正确提取的三元组，构成了少量样本（few-shot）训练上下文（context），称为“提示词”（prompt）。 示例语料库包括各种清晰、一致和全面的案例。
3. 提示LLM提取RDF三元组，并进行人工验证和迭代。 然后，通过调用ChatGPT API并执行python脚本（prompt\_A\_forLoop.py、prompt\_B\_forLoop.py）工作流程，向LLM提供更广泛的语料库，以提取和输出三元组。然而，最终的数据都要经过数字人文工作坊学员的校验把关，以达到专业上的正确。

**2.1.2 增补数据与检索示例**

本知识库采用关联数据（或参考1.0报告中的语义网）技术，其根本优势在于开放、关联、可共享、可扩展与灵活。如下将根据2.0增补数据，结合这些特点进行示例解析。

2.1.2.1 新版元数据本体介绍

鉴于2.0版的工作相当程度地关照了东方乐器方面的数据资料，新版元数据本体共有3个版本[[3]](#footnote-2)，其中，（1）最详实的是名为oriental\_instrument&music(EasternMusicInstruments)的版本，其也在官网的列表浏览和可视化浏览页面可见（可在后者页面直接下载）；（2）用于知识推理的本体（withAdditionalAnnotation）；（3）简化版本体（withAdditionalAnnotation...simplified...），用于NLQ2SPARQL试验。对于（2）（3），在后文会详细介绍。对于（1）详实版本体，可参考如下概念导图：



图

该图的核心特点在于参考了BibFrame本体模型。该模型认为，对图书的分类不仅仅是扁平化地展开，而是可以垂直地从三个层面展开，即作品层（概念层）—实例层—单件层。本项目同样地参考了这种思路，认为乐器的分类也可以垂直地从这三个层面展开，或者严谨地说，至少可以分为概念层（对应上图的东方乐器、OWL文档中的ctm:OrientalMusicalInstrument，其属于cidoc-crm:E55\_Type所规定的受控词汇表的范畴）、单件层（对应上图的乐器实物、OWL文档中的ctm:ItemOfInstrument）两个层面。至于中间的层面，乐器实例层，可作为过渡的层面考虑。这三个层面自上而下地反映了一种分类视角的从抽象到具体的变迁，譬如，概念层侧重于民族音乐学的视角，也受声学分类的思想影响，而具体的声学描述，主要适用于实例层、单件层。三层之间也是上下贯通的，如一个概念可以有一些实例（bf:hasInstance），而一些实例也可以有一些单件（bf:hasItem）。具体可参见该OWL文档中的Individuals示例：如，扬琴作为一个概念，它对应的实例可以是401扬琴，而这个实例还可能对应中国音乐学院图书馆的乐器博物馆中收藏的一件具体的扬琴。

目前，这种分类法仅通过此版本体得到初步规划与操作，而在学理上可初步推测其应用价值：不同的层面可以分化出不同的元数据方案，比如，实例层可以反映一个乐器类目的标准化信息，如基本规格（ctm:specification，下辖wdt:P2043—长度、wdt:P2048—高度、wdt:P2049—宽度）、管乐器的吹孔数目（ctm:nO\_OfFingerHoles）、弦乐器的琴弦数目（ctm:nO\_OfStrings）。对于尤其是传统的乐器，对一个概念类目用如上标准化的信息规约起来，可能是不合适的，所以，区别出概念层和实例层是有意义的。另一方面，一个概念类目，也不一定能找到对应的标准化实例，尤其是民间的很多手工艺乐器制品；即使是一个标准化的实例类目，经年累月及人为使用后，也会产生个性化的差异（比如音色的差异，再如，作为乐器文物的出土信息、孤品、残品的特征），这种个性化的差异是有意义的，至少可以从应有勤教授编写的《中国乐器大辞典》中看到这种隐含的“垂直的”3个层面的区分，其作会将很多独立的乐器藏品作为一个乐器类目的词条。[[4]](#footnote-3)此外，本2期项目特地将中国音乐图书馆乐器博物馆里的每一件实体乐器作为单件层的条目（ctm::ItemInMusicalInstrumentMuseumOfChinaConservatoryOfMusic），从而初步搭建了这种三层框架乐器分类的示例，具体可参2.1.2.4目。

鉴于本项目借鉴了LinkedMusic项目的经验，在OWL文档中，也增加了若干来自 Wikidata的类或属性。对于类，用命名空间前缀wd标识；对于属性，用前缀wdt标识。如wd:Q7403902为Category:霍恩博斯特尔-萨克斯分类法类目（作为cidoc-crm:E55\_Type的子类），wdt:P1762为声学分类，wdt:P189为出土地等。

在本轮本体中，还特地采用了空节点技术，用以表达更特殊场景中的语义。问题的缘起在于，一些传统乐器具有别称（ctm:instrumentAlternateName），但是，其别称只是在特定民族的语境中才有效，对于这种情况，就不能简单地用一个三元组表达，而需要空节点技术。

其他两版本体的初步介绍

**2.3 “音乐地图”案例拓展**

增补地理信息数据，可用以检索我国任意一级行政单位（村、寨除外）的乐种、乐器等信息

**2.3.1 检索案例分析**

2.3.1.1 地域毗邻关系检索

在2.0版的知识库中，补充了各市级行政单位的毗邻关系数据，借助这种毗邻关系的结构，也可比较有效地以知识网络可视化的方式，对应地理区位分布关系展示特定范围的乐种信息，譬如，问：

与河北省石家庄市相邻3步长范围内的其他城市及其下属各级行政单位会有什么乐种？就这些乐种，我馆是否收藏了一些特藏资源？

如下的检索方案先回答前半部分的问题“与河北省石家庄市相邻3步长范围内的其他城市及其下属各级行政单位会有什么乐种？”

define input:inference 'urn:owl.ccmusicrules0214'

PREFIX ctm: <https://lib.ccmusic.edu.cn/ontologies/chinese\_traditional\_music#>

PREFIX places: <http://purl.org/ontology/places#>

PREFIX gn: <https://www.geonames.org/ontology#>

PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

PREFIX bf: <http://id.loc.gov/ontologies/bibframe/>

SELECT DISTINCT ?MusicType ?MusicTypeName ?adminUnit ?adminUnitName ?SpecialIndependentResource ?SIR\_Label

WHERE {

# Identify Shijiazhuang city

?shijiazhuang rdfs:label "石家庄市" .

?shijiazhuang a places:City .

# Find cities within 3 steps of Shijiazhuang

{

# Include Shijiazhuang itself

BIND(?shijiazhuang AS ?city)

} UNION {

# Direct neighbor (1 step)

?shijiazhuang gn:neighbour ?city .

} UNION {

# 2 steps away

?shijiazhuang gn:neighbour ?intermediateCity1 .

?intermediateCity1 gn:neighbour ?city .

FILTER(?city != ?shijiazhuang)

} UNION {

# 3 steps away

?shijiazhuang gn:neighbour ?intermediateCity1 .

?intermediateCity1 gn:neighbour ?intermediateCity2 .

?intermediateCity2 gn:neighbour ?city .

FILTER(?city != ?shijiazhuang && ?city != ?intermediateCity1)

}

# Find administrative units at all levels under these cities

{

# The city itself

BIND(?city AS ?adminUnit)

} UNION {

# Counties under the city

?county gn:parentADM2 ?city .

BIND(?county AS ?adminUnit)

} UNION {

# Towns under counties

?county gn:parentADM2 ?city .

?town gn:parentADM3 ?county .

BIND(?town AS ?adminUnit)

} UNION {

# Villages under towns

?county gn:parentADM2 ?city .

?town gn:parentADM3 ?county .

?village gn:parentADM4 ?town .

BIND(?village AS ?adminUnit)

}

# Find music types associated with these administrative units

?MusicType bf:place ?adminUnit .

# Get names for readability

?adminUnit rdfs:label ?adminUnitName .

?MusicType rdfs:label ?MusicTypeName .

# Filter out any results where ?MusicTypeName contains pinyin (values with @py language tag)

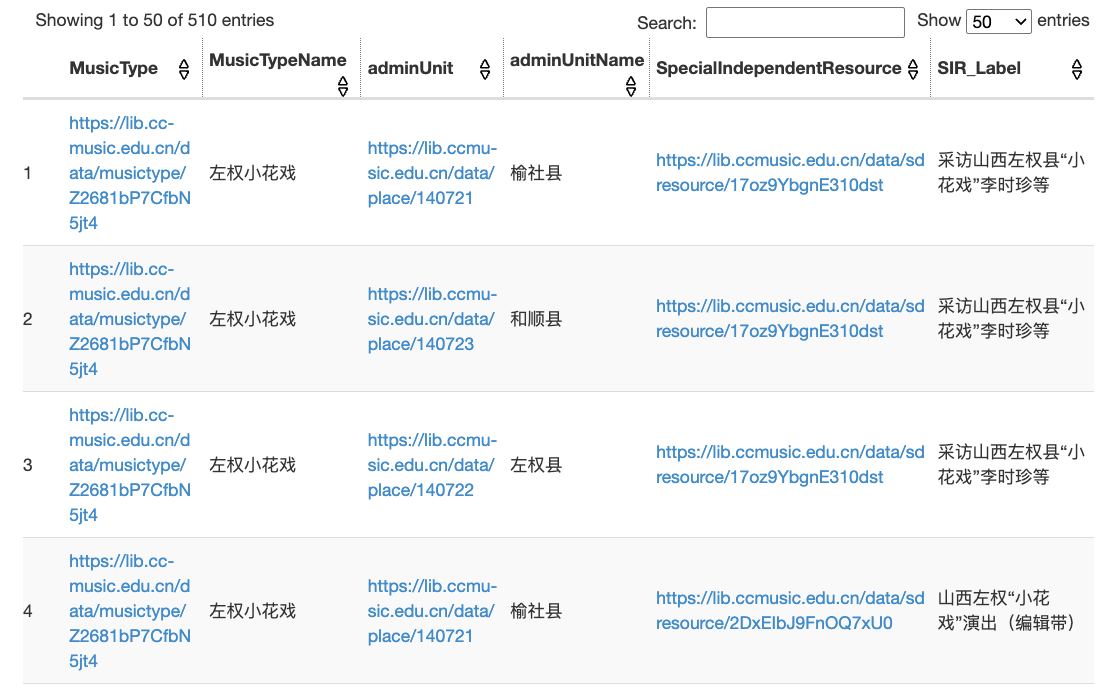
FILTER(LANG(?MusicTypeName) != "py")

# 就这些乐种，我馆是否收藏了一些特藏资源？如果有，请将其返回

optional { ?SpecialIndependentResource ctm:relatesMusicType ?MusicType ; rdfs:label ?SIR\_Label }

}

根据如上代码，返回的结果如：



图

如榆社县（属于晋中市）的左权小花戏，对应的我馆特藏资源名称为《采访山西左权县“小花戏”李时珍等》。其他（截图之外的）有祁县（属于晋中市）的祁太秧歌，对应的我馆特藏资源名称为《王西麟音乐会（中山音乐堂）》；有涞水县（属于保定市）的“河北音乐会-北乐会”，对应的我馆特藏资源名称为《河北省屈家营音乐会重建二十四周年（演奏篇）：屈家营“音乐圣会”世界音乐周专场音乐会》。另外，须注意，返回的第1-91, 178-235, 389, 408-412, 420-441条数据，相应的乐种则在我馆特藏独立资源中有收藏，且对应的行政单位为县区级；其他的对应的行政单位基本为乡镇级或村寨级的；等等。

2.3.1.1.1 地域毗邻关系检索的知识网络可视化操作示例

首先，可以展示石家庄市及与其相邻的2步长范围内的其他城市（参考如下SPARQL代码Part 1）及这些城市所有的下辖县区级行政单位（代码Part 2）构成的邻接关系网络[[5]](#footnote-4)；其二，在此基础上关联与县级行政单位相邻的乐种（代码Part 3）；其三，再关联与关联到的乐种相关联的特藏独立资源（代码Part 4）。

可以通过SPARQL语句，逐步获取数据：

define input:inference 'urn:owl.ccmusicrules0214' # 用于Open Link Virtuoso中激活推理机

PREFIX bf: <http://id.loc.gov/ontologies/bibframe/>

PREFIX ctm: <https://lib.ccmusic.edu.cn/ontologies/chinese\_traditional\_music#>

PREFIX gn: <https://www.geonames.org/ontology#>

PREFIX places: <http://purl.org/ontology/places#>

PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

# 注意：激活变量?distance，则可显示相应的边列表关系是第几步长范围内的

SELECT distinct ?source ?sourceLabel ?target ?targetLabel ?relationType #?distance

WHERE {

# Identify 石家庄市 as our starting point

?shijiazhuang rdfs:label "石家庄市" .

?shijiazhuang a places:City .

{

**# Part 1: City-to-City connections 城市间的毗邻关系**

{

# Direct neighbors (1 step)

?shijiazhuang gn:neighbour ?target .

BIND(?shijiazhuang AS ?source)

BIND(1 AS ?distance)

BIND(gn:neighbour AS ?relationType)

}

UNION

{

# 2 steps away

?shijiazhuang gn:neighbour ?intermediate1 .

?intermediate1 gn:neighbour ?target .

FILTER(?target != ?shijiazhuang) # Avoid cycles back to origin

BIND(?intermediate1 AS ?source)

BIND(2 AS ?distance)

BIND(gn:neighbour AS ?relationType)

}

# Get readable labels for cities

?source rdfs:label ?sourceLabel .

?target rdfs:label ?targetLabel .

}

UNION

**{**

**# Part 2: County-to-City relationships 县、区级行政单位到市级行政单位的信息**

{

# Direct parent-child relationship - counties of Shijiazhuang

?county gn:parentADM2 ?shijiazhuang .

?county a places:County .

?county rdfs:label ?sourceLabel .

?shijiazhuang rdfs:label ?targetLabel .

BIND(?county AS ?source)

BIND(?shijiazhuang AS ?target)

BIND(gn:parentADM2 AS ?relationType)

}

UNION

{

# Counties of neighboring cities

?shijiazhuang gn:neighbour ?neighborCity .

?neighborCity a places:City .

?county gn:parentADM2 ?neighborCity .

?county a places:County .

?county rdfs:label ?sourceLabel .

?neighborCity rdfs:label ?targetLabel .

BIND(?county AS ?source)

BIND(?neighborCity AS ?target)

BIND(gn:parentADM2 AS ?relationType)

}

UNION

{

# Counties of cities 2 steps away

?shijiazhuang gn:neighbour ?intermediate1 .

?intermediate1 gn:neighbour ?distantCity .

FILTER(?distantCity != ?shijiazhuang)

?distantCity a places:City .

?county gn:parentADM2 ?distantCity .

?county a places:County .

?county rdfs:label ?sourceLabel .

?distantCity rdfs:label ?targetLabel .

BIND(?county AS ?source)

BIND(?distantCity AS ?target)

BIND(gn:parentADM2 AS ?relationType)

}

**}**

UNION

{

**# Part 3: MusicType-to-County relationships 乐种及其分布地域（仅以县区级行政单位为例）**

{

# Music types related to counties in Shijiazhuang

?county gn:parentADM2 ?shijiazhuang .

?county a places:County .

?musicType bf:place ?county .

?musicType rdfs:label ?sourceLabel .

?county rdfs:label ?targetLabel .

BIND(?musicType AS ?source)

BIND(?county AS ?target)

BIND(bf:place AS ?relationType)

FILTER(LANG(?sourceLabel) != "py")

}

UNION

{

# Music types related to counties in neighboring cities

?shijiazhuang gn:neighbour ?neighborCity .

?neighborCity a places:City .

?county gn:parentADM2 ?neighborCity .

?county a places:County .

?musicType bf:place ?county .

?musicType rdfs:label ?sourceLabel .

?county rdfs:label ?targetLabel .

BIND(?musicType AS ?source)

BIND(?county AS ?target)

BIND(bf:place AS ?relationType)

FILTER(LANG(?sourceLabel) != "py")

}

UNION

{

# Music types related to counties in cities 2 steps away

?shijiazhuang gn:neighbour ?intermediate1 .

?intermediate1 gn:neighbour ?distantCity .

FILTER(?distantCity != ?shijiazhuang)

?distantCity a places:City .

?county gn:parentADM2 ?distantCity .

?county a places:County .

?musicType bf:place ?county .

?musicType rdfs:label ?sourceLabel .

?county rdfs:label ?targetLabel .

BIND(?musicType AS ?source)

BIND(?county AS ?target)

BIND(bf:place AS ?relationType)

FILTER(LANG(?sourceLabel) != "py")

}

}

UNION

{

**# Part 4: SpecialIndependentResource-to-MusicType relationships 特藏独立资源及其涉及的乐种**

{

# specialResources related to musicTypes in Shijiazhuang counties

?county gn:parentADM2 ?shijiazhuang .

?county a places:County .

?musicType bf:place ?county .

?specialResource a ctm:SpecialIndependentResource .

?specialResource ctm:relatesMusicType ?musicType .

?specialResource rdfs:label ?sourceLabel .

?musicType rdfs:label ?targetLabel .

BIND(?specialResource AS ?source)

BIND(?musicType AS ?target)

BIND(ctm:relatesMusicType AS ?relationType)

FILTER(LANG(?targetLabel) != "py")

}

UNION

{

# specialResources related to musicTypes in neighboring cities' counties

?shijiazhuang gn:neighbour ?neighborCity .

?neighborCity a places:City .

?county gn:parentADM2 ?neighborCity .

?county a places:County .

?musicType bf:place ?county .

?specialResource a ctm:SpecialIndependentResource .

?specialResource ctm:relatesMusicType ?musicType .

?specialResource rdfs:label ?sourceLabel .

?musicType rdfs:label ?targetLabel .

BIND(?specialResource AS ?source)

BIND(?musicType AS ?target)

BIND(ctm:relatesMusicType AS ?relationType)

FILTER(LANG(?targetLabel) != "py")

}

UNION

{

# specialResources related to musicTypes in counties of cities 2 steps away

?shijiazhuang gn:neighbour ?intermediate1 .

?intermediate1 gn:neighbour ?distantCity .

FILTER(?distantCity != ?shijiazhuang)

?distantCity a places:City .

?county gn:parentADM2 ?distantCity .

?county a places:County .

?musicType bf:place ?county .

?specialResource a ctm:SpecialIndependentResource .

?specialResource ctm:relatesMusicType ?musicType .

?specialResource rdfs:label ?sourceLabel .

?musicType rdfs:label ?targetLabel .

BIND(?specialResource AS ?source)

BIND(?musicType AS ?target)

BIND(ctm:relatesMusicType AS ?relationType)

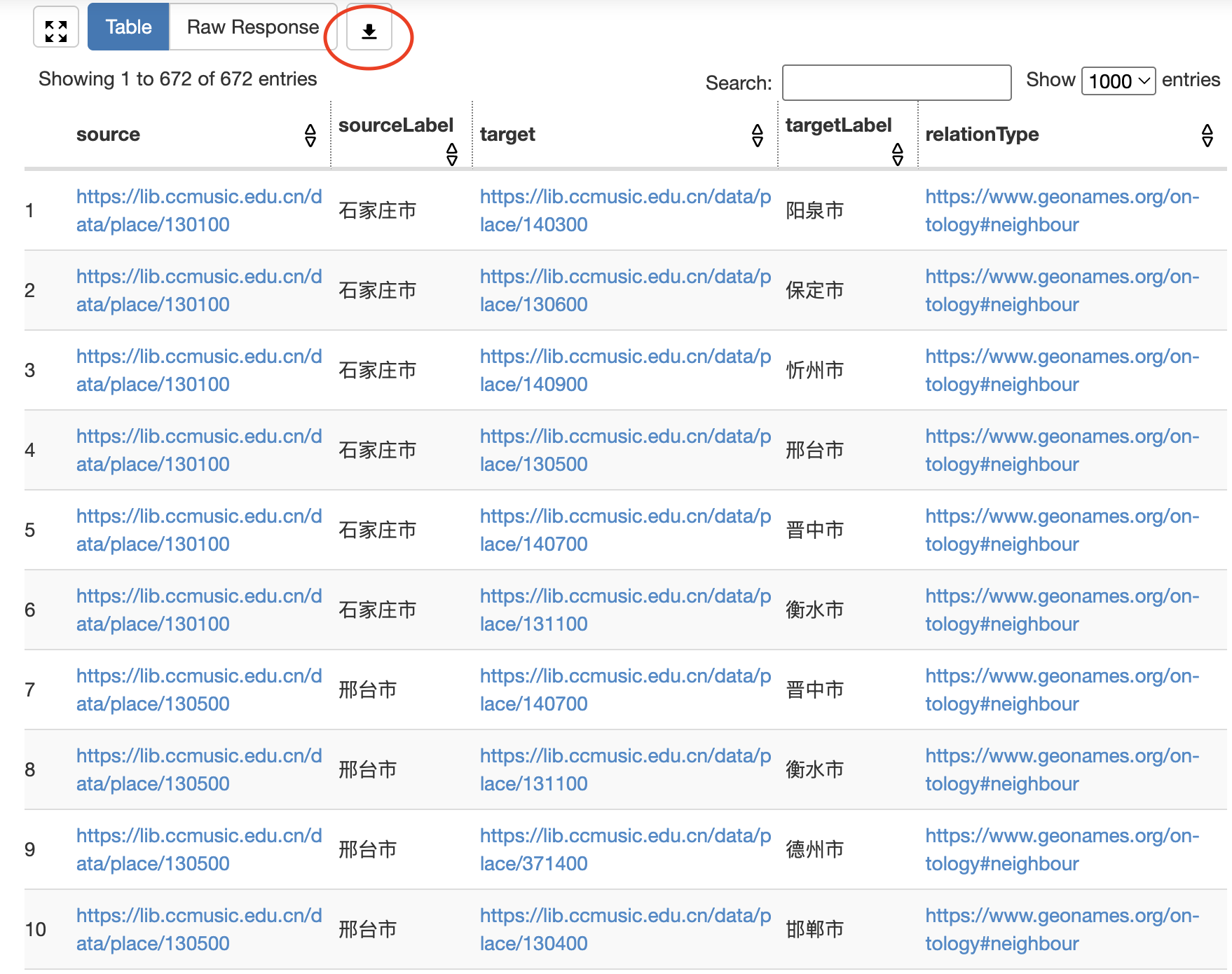
FILTER(LANG(?targetLabel) != "py")

}

}

}

根据如上，它返回的结果是一个“类型化边列表格式”的数据，即从源节点（source）指向目标节点（target）的列表：



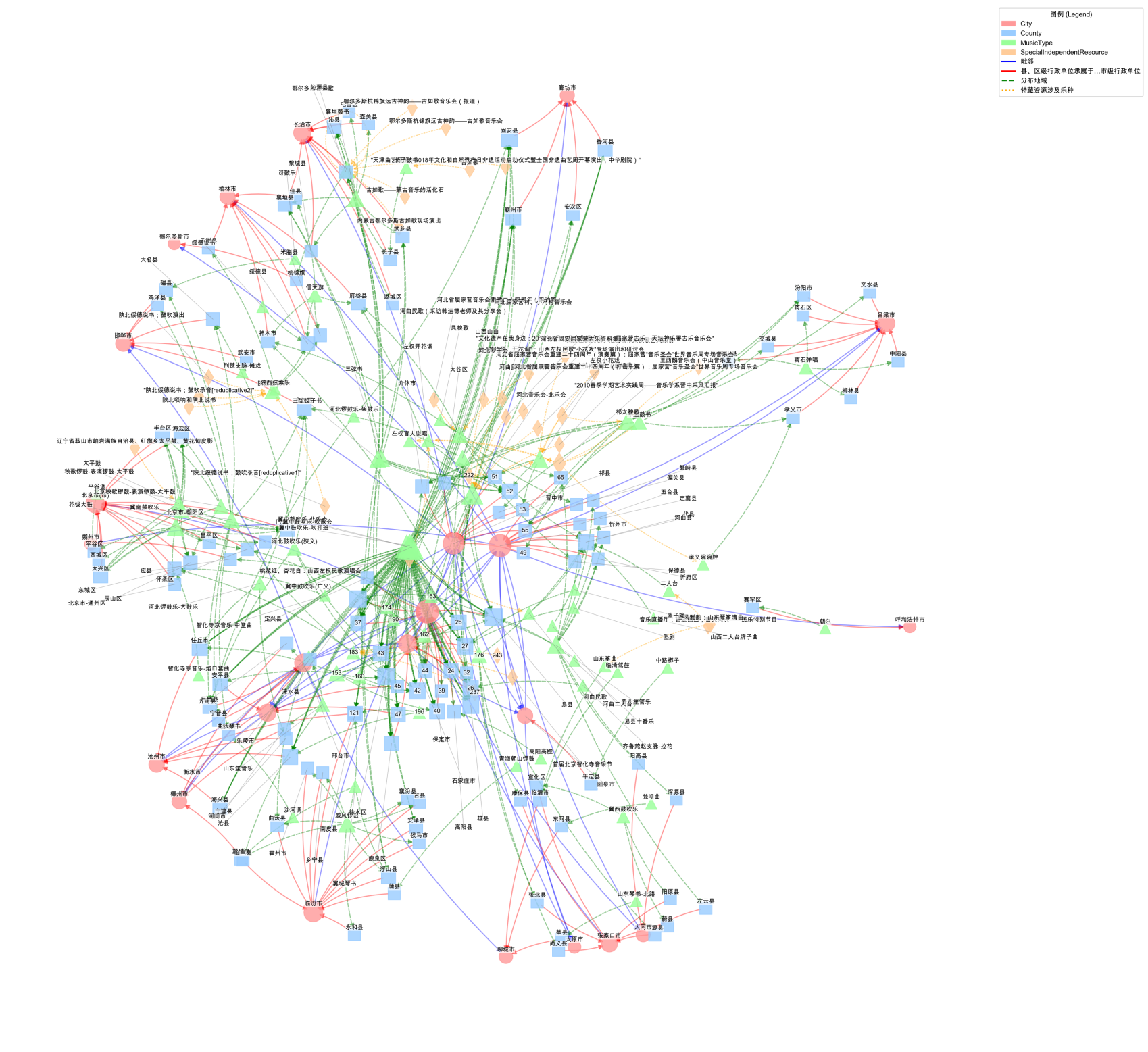
……

图

如上图所示（“Show 1000 entries”，共返回672行结果），而就所谓类型化边列表格式，可见其最右列对连边的类型做出了标注。在此返回的数据中，依次共有4种类型：<https://www.geonames.org/ontology#neighbour>(市级行政单位与另一市级行政单位毗邻), gn:parentADM2(县、区级行政单位地域隶属于), <

<http://id.loc.gov/ontologies/bibframe/place>(分布地域),> ctm:relatesMusicType。点击上图红圈所示的下载按钮，可以获得CSV格式的数据，并进行异构知识网络分析。

仅须简单的数据处理，然后，为了视图的简化，比如，筛除那些没有关联任何乐种的县、区级行政单位，作图及显示如下[[6]](#footnote-5)：



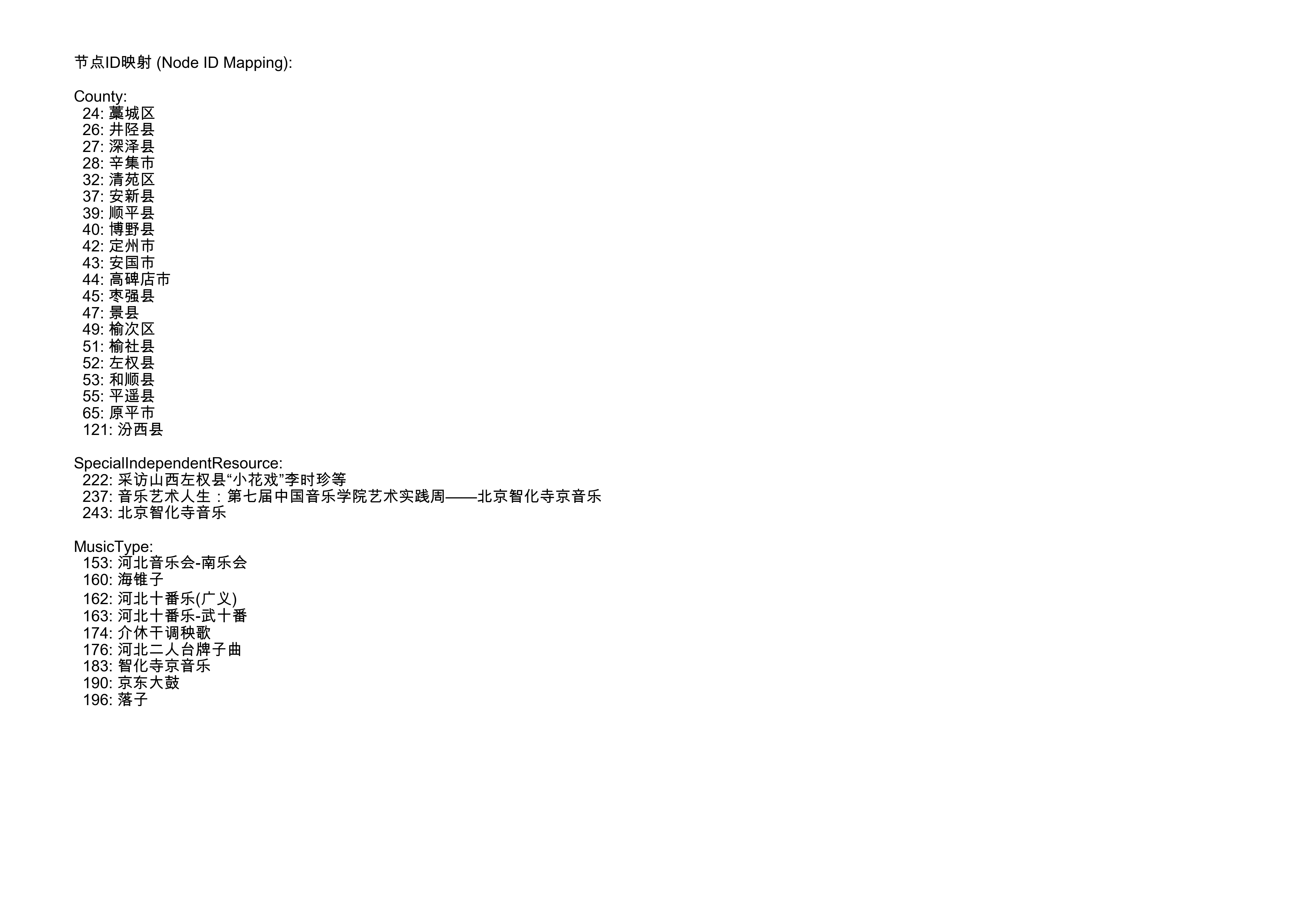
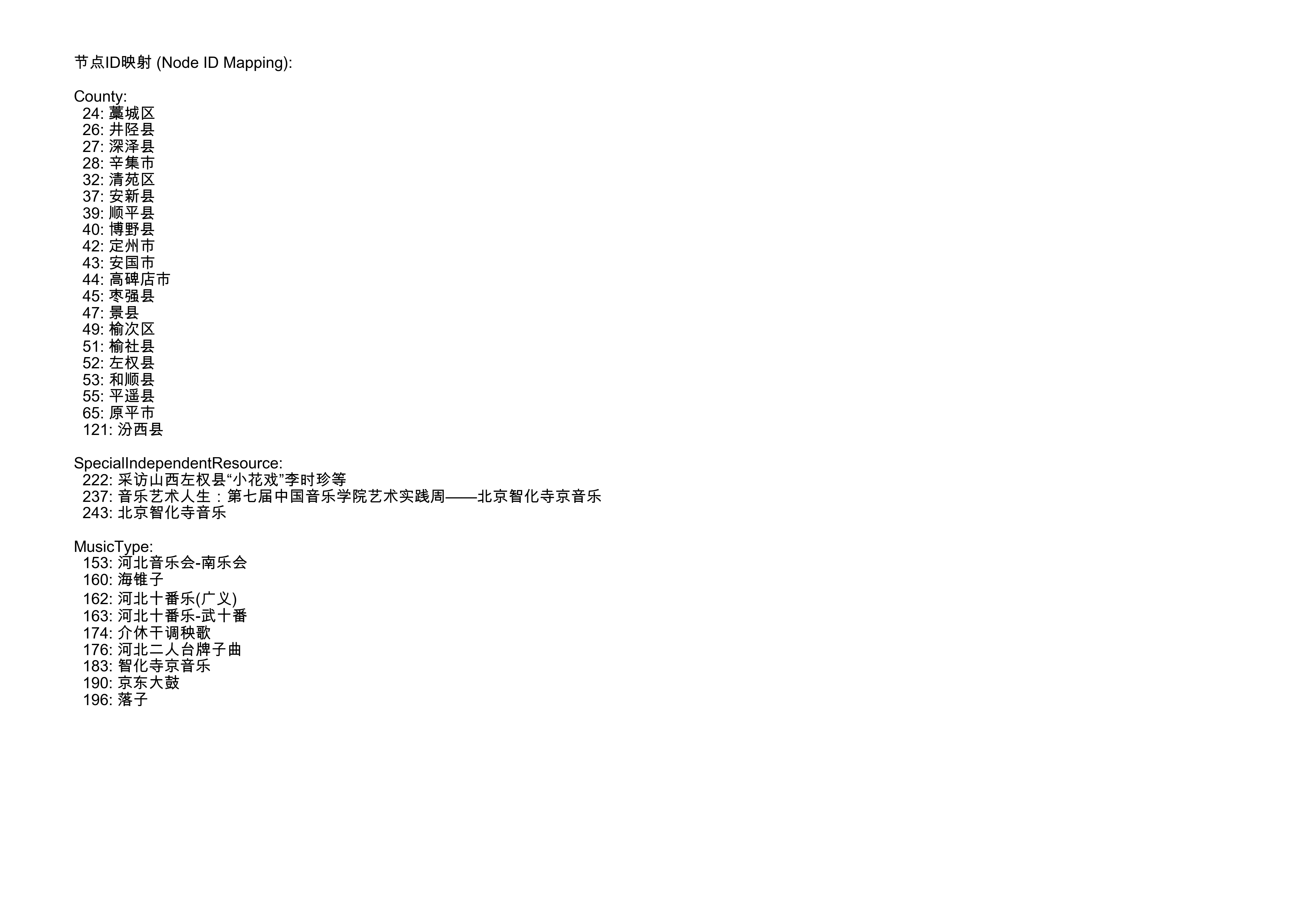
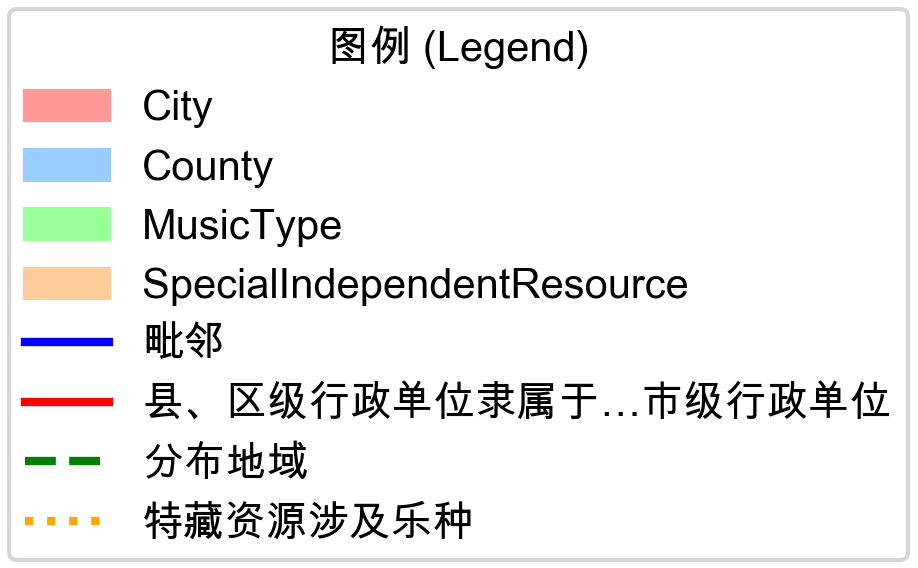
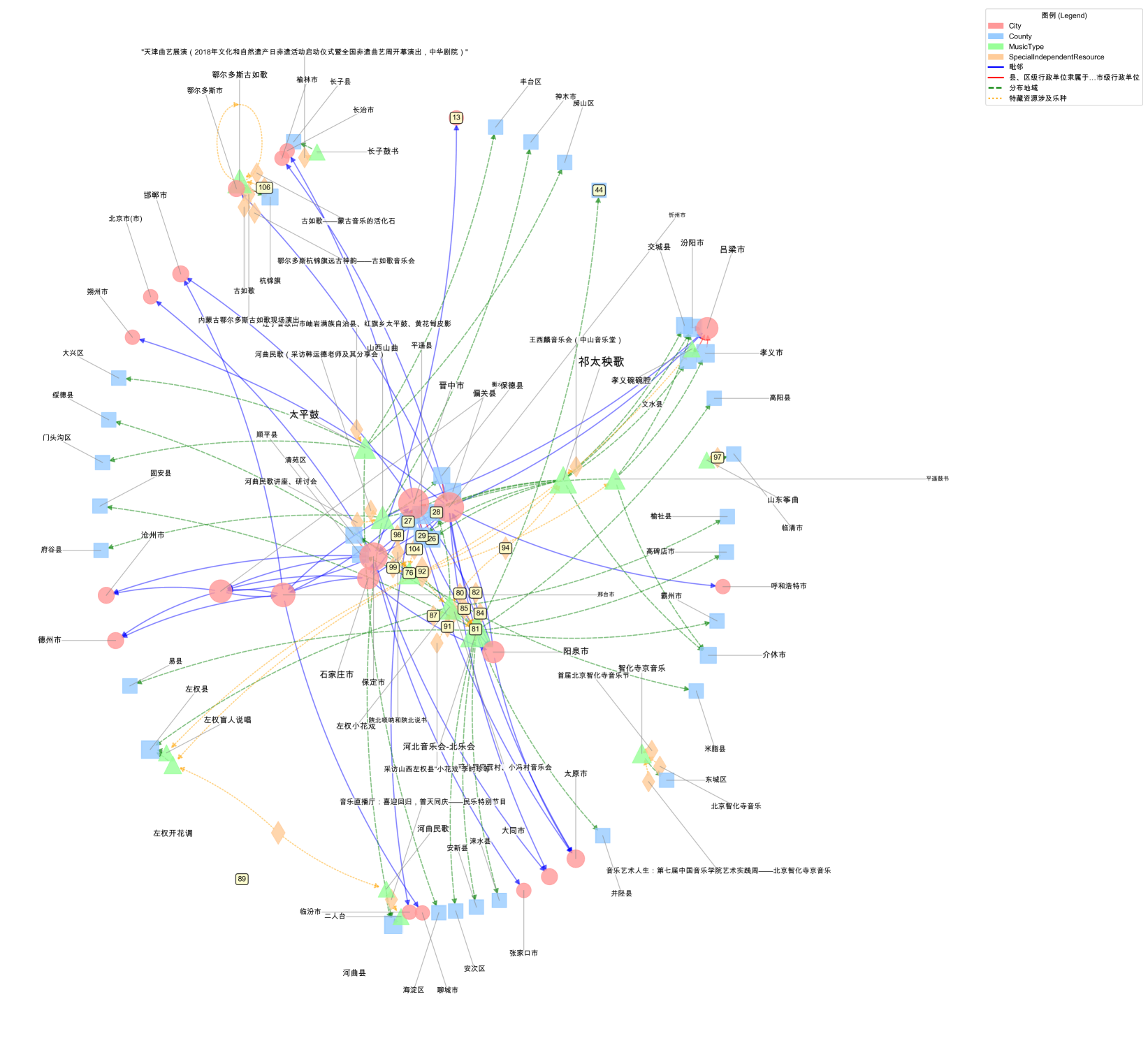


图 城市、县区、乐种、特藏资源间的异构知识网络（2024Apr20）

该图采用了弹簧模型布局，适当放大了高节点中心度的节点。图例中有关于“数字标识符—节点标签”的列表：图中部分节点的标签过长，加上展示空间有限，故对节点标签采用数字标识符替代，从而读者可在列表中查找相应标识符所匹配的标签

……我们还可以通过筛除掉“没有关联特藏资源的乐种所分布的区县”而更清楚地观察：中国音乐学院图书馆特藏资源—乐种—县—以石家庄市为中心2步长范围内的市[[7]](#footnote-6)——它们之间的语义关系：



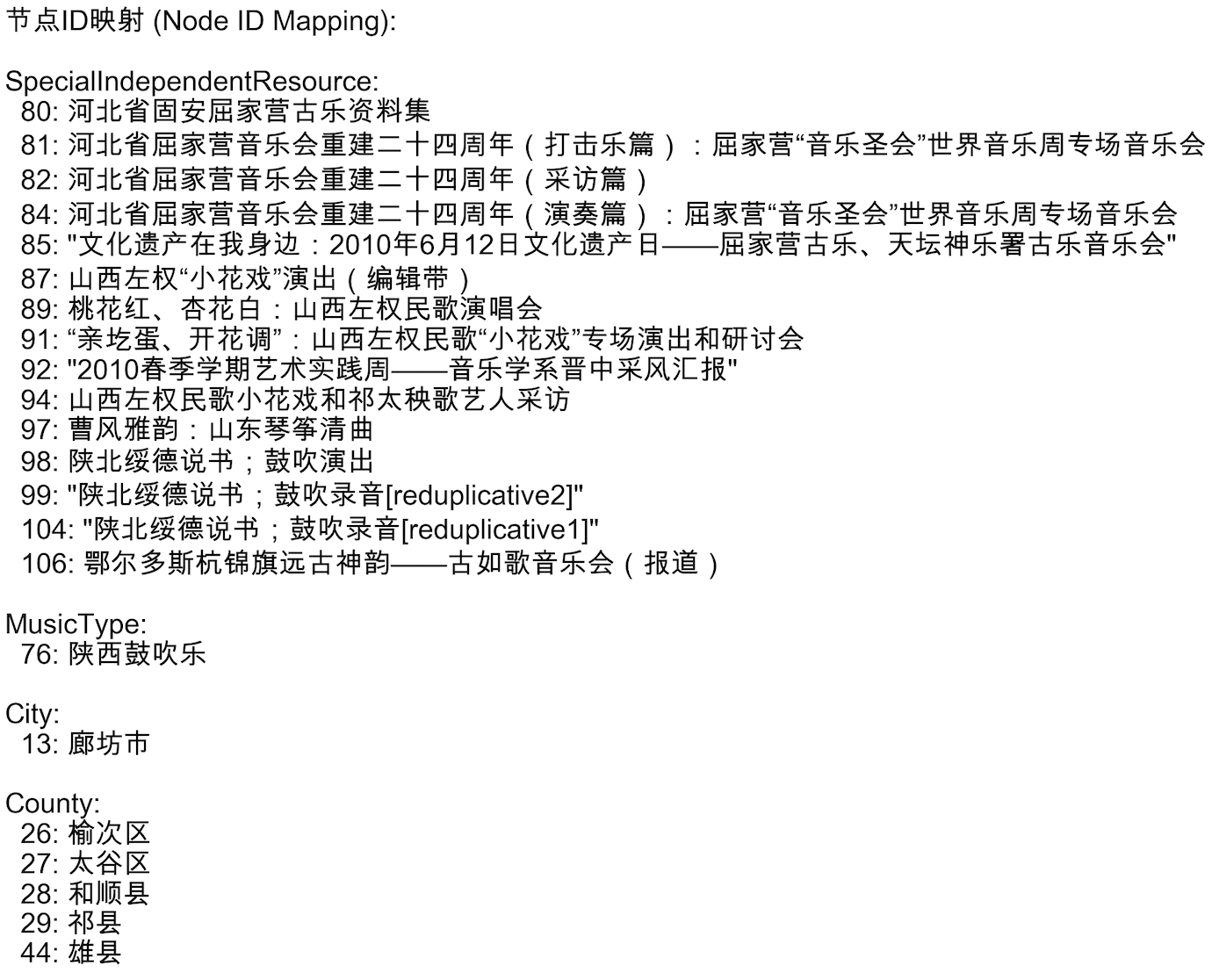


图 城市、县区、乐种、特藏资源间的异构知识网络（有删节）

2.3.1.2地理坐标距离检索

地域毗邻关系检索毕竟只是以网络拓扑结构的方式展示近似地图分布的数据信息。有时，根据毗邻关系捕获的数据是意料之外的，却缺乏了真实空间定位的参考性。譬如，返回与湖北省相邻2步长的其他省份，会发现，中国所有的省都包括在内了……要想更精确地呈现音乐地图，最好是通过地理坐标距离进行检索。

在2.0版的知识库中，补充了湖北省所有县级行政单位的地理坐标信息，例如，武汉市武昌区的坐标为东经114.316520度，北纬30.553860度。以RDF数据表示，诸如湖北省的武昌区、随县、赤壁市（县级市）的坐标信息记录如下：

@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .

@prefix wdt: <http://www.wikidata.org/prop/direct/> .

@prefix geo: <http://www.openlinksw.com/schemas/virtrdf#> .

<https://lib.ccmusic.edu.cn/data/place/420106> rdfs:label "武昌区" ;

wdt:P625 "POINT(114.31652 30.55386)"^^geo:Geometry .

<https://lib.ccmusic.edu.cn/data/place/421321> rdfs:label "随县" ;

wdt:P625 "POINT(113.29995 31.8538)"^^geo:Geometry .

……[[8]](#footnote-7)

湖北省荆门市（东宝区）方圆75公里内有哪些乐种，其中，哪些在我们图书馆有收藏？

define input:inference 'urn:owl.ccmusicrules0214' # 用于激活推理机

PREFIX gn: <https://www.geonames.org/ontology#>

PREFIX ctm: <https://lib.ccmusic.edu.cn/ontologies/chinese\_traditional\_music#>

PREFIX bf: <http://id.loc.gov/ontologies/bibframe/>

PREFIX wdt: <http://www.wikidata.org/prop/direct/>

PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

select distinct ?OtherCity ?otherCityLab ?OtherCounty ?coordin ?otherCounLab ?distan ?MTForCity ?mTFCityLab ?mTFCounty ?mTFCounLab ?Reso # ?resouL # ?MusicType2

where {

# （1）找到荆门市中心的东宝区的坐标，以及东宝区可能存在的音乐类型(乐种)：

?CenterCounty rdfs:label "东宝区" ;

wdt:P625 ?centerCoordinateLocation .

# optional { ?MusicType1 bf:place ?CenterCounty ; a ctm:MusicType } # [随后事实发现，并没有音乐类型(乐种)是直接关联东宝区的]

# （2）规定好备选的其他县区级行政单位的坐标，以及其所隶属的市级行政单位，以及隶属于该县区级行政单位的乡镇级行政单位：

?OtherCounty wdt:P625 ?coordin ; # P625为坐标属性

gn:parentADM2 ?OtherCity ; # 县、区级行政单位地域隶属于某市级行政单位?OtherCity

rdfs:label ?otherCounLab .

?OtherCity rdfs:label ?otherCityLab .

# ?OtherTown gn:parentADM3 ?OtherCounty . # 某些乡镇级行政单位?OtherTown隶属于该县级行政单位（假定一个县区级行政单位必然有下辖的乡镇级行政单位）

# （3）如果某些音乐类型(乐种)的分布地域是那些县区级或市级行政单位，进而，如果这些音乐类型涉及到我馆的特藏独立资源

optional {

?mTFCounty a ctm:MusicType ;

bf:place ?OtherCounty ;

rdfs:label ?mTFCounLab .

FILTER(LANG(?mTFCounLab) != "py")

optional { ?mTFCounty ctm:relatesWork ?Reso .

#?SpecialIndependentResource rdfs:label ?resouL

}

}

optional {

?MTForCity a ctm:MusicType ;

bf:place ?OtherCity ;

rdfs:label ?mTFCityLab .

FILTER(LANG(?mTFCityLab) != "py")

optional { ?MTForCity ctm:relatesWork ?Reso .

#?SpecialIndependentResource rdfs:label ?resouL

}

}

# optional { ?MusicType2 a ctm:MusicType ;

# bf:place ?OtherTown . # [可能并有音乐类型(乐种)是直接关联此片县级行政单位下属的乡镇级行政单位的，为了本案例研究的简约性，暂对此略去]

# optional { ?MusicType2 ctm:relatesWork ?SpecialIndependentResource . }}

#（4）计算东宝区和其他县级行政单位之间的距离，设筛选条件为75公里之内

bind (bif:st\_distance(?centerCoordinateLocation, ?coordin) AS ?distan)

filter (?distan<=75)

}

group by ?OtherCity ?otherCityLab ?OtherCounty ?coordin ?otherCounLab ?distan ?MTForCity ?mTFCityLab ?mTFCounty ?mTFCounLab ?Reso # ?resouL # ?MusicType2 # 为了避免返回的数据行有重复，故加group by

order by ?distan

为了解答如上问题，基本思路是：（1）荆门市的中心为东宝区，所以以该区的坐标代表该市的中心坐标（2）选定其他县区级行政单位及其所隶属的市级行政单位（3）找到这些县区级行政单位及其所隶属的市级行政单位所关联的（bf:place）乐种，以及这些乐种可能关联的特藏资源（4）以东宝区和其他县区级行政单位的距离小于75公里作为筛选条件，从而最终找到相应资源。所用到的属性http://www.wikidata.org/prop/direct/P625即coordinate location（坐标位置）。注意，就这个问题，它用到了特殊的SPARQL 函数“bif:st\_distance”，用于计算两个坐标间的距离。检索结果：150行数据（（在滤除修订、增补的内容的前提下)）——如果不返回乐种名，它会包含拼音名，造成统计上的重复）。例如，荆门市的府河傢业、天沔小曲，其下辖的县级市，又有梁山调，等等。检索返回的表格的字段排序如下：①城市OtherCity；②城市名otherCityLabel；③县区级行政单位OtherCounty；④县区级行政单位的坐标（经度—纬度）coordinate；⑤县区级行政单位名otherCountyLabel；⑥和荆门市的距离distance；⑦与城市相关的乐种MusicTypeForCity；⑧与城市相关的乐种名musicTypeForCityLabel；⑨与县区级行政单位相关的乐种MusicTypeForCounty；⑩与县区级行政单位相关的乐种名MusicTypeForCountyLabel；最后一列则是涉及的特藏资源Resource。第⑦⑧列乐种所涉及的城市在第①②列，第⑨⑩列乐种所涉及的县区级行政单位在第③⑤列，也可推理其对应的市级行政单位在第①②列。



图

例如从上图可见，根据本知识库的记录，钟祥市是荆门市的县级市。可以说，府河傢业的分布地域在荆门市，但不一定在其下辖的钟祥市；梁山调的分布地域在钟祥市，而推理出其分布地域在荆门市也是无妨的。最后，重要的是，发现在荆门市（东宝区）方圆75公里内，宜城市（县级市）所隶属的襄阳市有乐种襄河道坠子，我馆收藏的资源中，有3项目与此相关，即《湖北襄河道坠子系列之一：湖北“襄河道坠子”演出》《湖北襄河道坠子系列之二：“襄河道坠子”作为地域性的流派和形成（姚艺君）》《湖北襄河道坠子系列之三：“襄河道坠子”讲座（李大庆）》。详参如下：



图 SoooPa关联数据发布平台对特藏资源《湖北襄河道坠子子系列之一》的展示

最后，还需注意的是，荆门市的京山市和东宝区的距离竟然达到87.4812公里，导致京山市本身没有被统计在内。

2.3.1.2.1 地理坐标距离检索的知识网络可视化操作示例

为了充分展现基于地理坐标的可视化的功能信息表达的优势，可直接基于上例下载的数据，通过数据透视表略作整理，将其转成CSV格式数据并通过提示词工程（prompt engineering），实现python作图。回答这样的问题：在荆州市（参考东宝区坐标为市中心坐标）周围方圆75公里内的县区级行政单位有哪些代表性乐种？[[9]](#footnote-8)（此处为了清晰显示，略去市级行政单位关联的乐种）

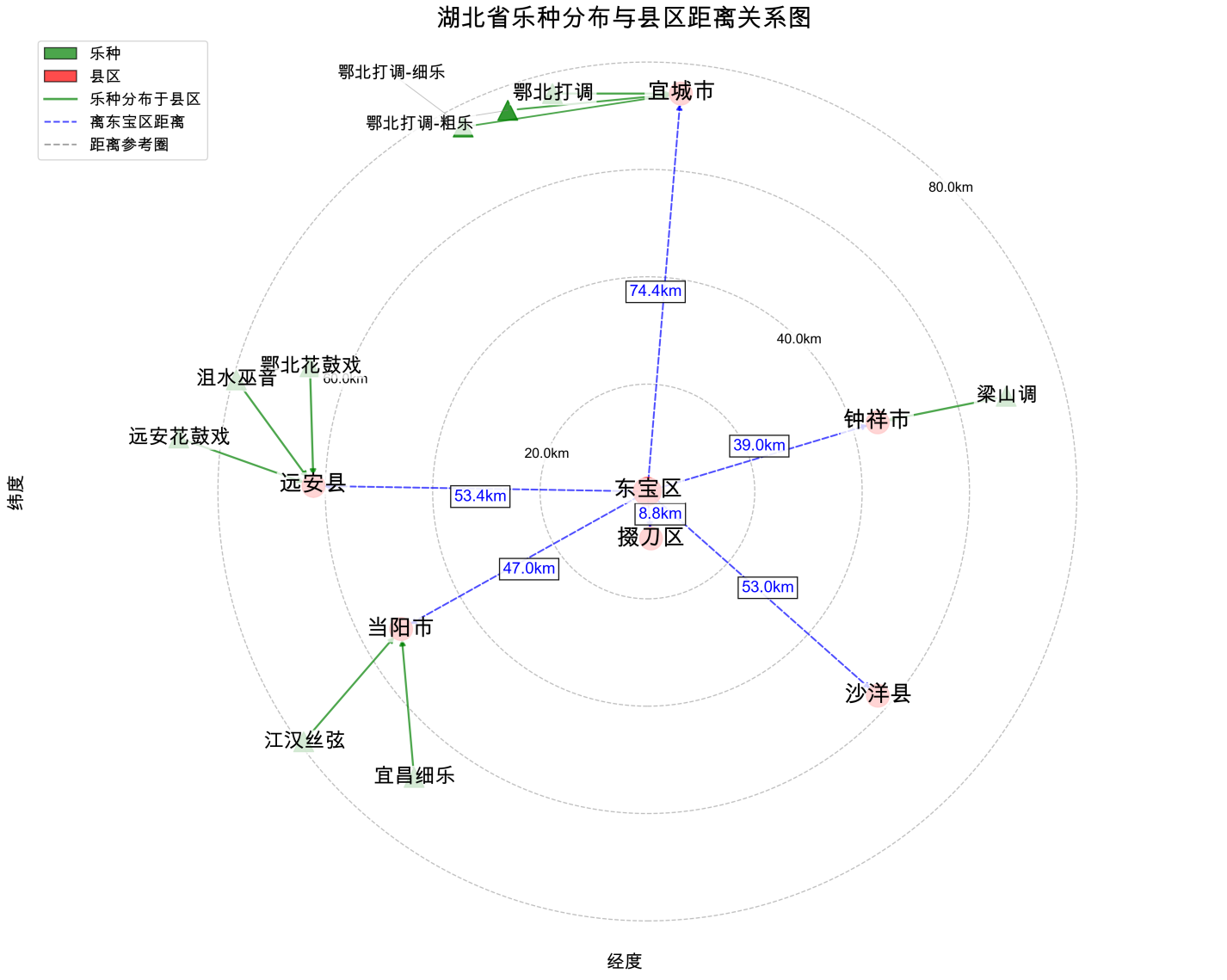


图 湖北省荆州市（东宝区）方圆75公里内其他县区级行政单位乐种信息展示

上图客观复现了🧍‍各个县级行政单位之间的真实距离与方位关系。通过该图可见，从东宝区出发，方圆75公里内，可达的最近的有明确乐种记载的县区级单位是西南方向的当阳市，距离47公里，可采集到的乐种是汉江丝弦等；距离较远的有74.4公里处的宜城市，有鄂北打调及其细乐、粗乐；等等。

**2.3.2 知识地图研究意义小结**

在这里可以补上对“城市、县区、乐种、特藏资源间的异构知识网络（2024Apr20）”一图对应的原始数据的基于提示词工程的数据分析，正好用于承上启下

**3 基于大语言模型的知识库智能检索系统（后端）开发**

**3.1 研究的问题、意义与背景综述**

对于2.3.1（1）的问题“与河北省石家庄市相邻4步长范围内的其他城市及其下属各级行政单位会有什么乐种？……”：

提示词/prompt:

Given

```

@prefix ctm: <https://lib.ccmusic.edu.cn/ontologies/chinese\_traditional\_music#> .

@prefix places: <http://purl.org/ontology/places#> .

@prefix gn: <https://www.geonames.org/ontology#> .

# 类：

ctm:MusicType rdfs:label "乐种" .

ctm:SpecialIndependentResource rdfs:label "特藏资源" .

ctm:Town rdfs:subClassOf bf:Place .

places:City rdfs:subClassOf bf:Place .

places:County rdfs:subClassOf bf:Place .

places:Village rdfs:subClassOf bf:Place .

# 属性：

bf:place rdfs:label "分布地域" .

rdfs:range bf:Place .

ctm:relatesMusicType rdfs:domain [ owl:intersectionOf ( ctm:SpecialIndependentResource [ rdf:type owl:Class ; owl:complementOf ctm:PieceWithPerformance ]) ;

rdf:type owl:Class ] ;

rdfs:range ctm:MusicType ;

rdfs:label "特藏资源涉及…音乐类型(乐种)" .

gn:neighbour rdfs:label "市级行政单位与另一市级行政单位毗邻" ;

rdfs:domain places:City ;

rdfs:range places:City .

gn:parentADM2 rdf:type owl:ObjectProperty ;

rdfs:domain places:County ;

rdfs:range places:City ;

rdfs:label "县、区级行政单位地域隶属于" .

gn:parentADM3 rdf:type owl:ObjectProperty ;

rdfs:domain ctm:Town ;

rdfs:range places:County ;

rdfs:label "乡、镇级行政单位地域隶属于" .

gn:parentADM4 rdf:type owl:ObjectProperty ;

rdfs:domain places:Village ;

rdfs:range ctm:Town ;

rdfs:label "村、社区级行政单位地域隶属于" .

```

, for the question “与河北省石家庄市（rdfs:label “石家庄市”）相邻3步长范围内的其他城市及其下属各级行政单位会有什么乐种？”, please create a corresponding SPARQL

对于2.3.1.1.1的异构网络绘制，提示词可参考文档promptFor2.3.1.1.1.md。

1. 比如，一个乐种分布在某（乡镇）级行政单位，那么可理解，该乐种也应该分布在该（乡镇）级行政单位所有的上级行政单位，通过基于规则的推理，目前已补齐这方面的数据（我记不清对于乐器，是否也已补齐此数据了，待查）。同样的，未来可考虑其他能补齐的数据：如果一个特藏资源涉及某级行政单位…… [↑](#footnote-ref-0)
2. 见中国乐器大词典excel版\_wellFormatted\_October2024\_final.xlsx；就随后提及的python程序流，参见文件夹“提示词工程\_用于从语料库中自动提取乐器知识图谱RDF 数据（python脚本）”。在本条的下文，会多次提及该文件夹中所涉及的文件，即在行文的括弧中提及，并附文件后缀名，供读者参考。该研究获得的数据结果参见https://github.com/DDMAL/linkedmusic-datalake/tree/main/data/ESEA(East-and-Southeast-Asian)TraditionalMusicInstrument，即文档ChineseMusicInstrument\_=36625Triples\_2025Jan.ttl [↑](#footnote-ref-1)
3. 详参https://github.com/DDMAL/linkedmusic-queries/tree/main/ChineseTraditionalMusicKnowledgeBase/3versionsOfOntology [↑](#footnote-ref-2)
4. 或者换句话说，从《中国乐器大辞典》对乐器品种区分的处置来看，它其实相当程度地混合了不同视角的分类，既有概念层、又有实例层、又有单件层。本研究正是发现了这个问题，将其提出并推举采用这种3层知识组织的方式重新审视乐器分类问题 [↑](#footnote-ref-3)
5. 与上例相比，该例略做了简化改动，即将原来的3步长范围改为了2步长范围，同时，将原来的“下属各级行政单位”改成了“下辖县区级行政单位” [↑](#footnote-ref-4)
6. 参考python作图代码2.3.1.1.1.py [↑](#footnote-ref-5)
7. 参考python作图代码2.3.1.1.1forSpecialIndependentResource，下图在显示中，有节略，可另参考 [↑](#footnote-ref-6)
8. 这些数据直接来自对chatGpt的询问。一般而言，行政单位不宜过大，否则面积太大，则不宜以单坐标点表示其位置；也不宜过小，否则区分度不够，大语言模型也未必能给出完备、准确的坐标数据 [↑](#footnote-ref-7)
9. 就通过上例SPARQL查询返回并下载的数据，做了数据透视表整合，对此，可参见queryResults\_for2.3.1.2\_refined.xlsx，将整合后的数据内容直接置于提示词和代码中；就提示词，可参考promptFor2.3.1.2.1.md；就根据此提示词生成并经修改后定稿的python作图代码，见2.3.1.2.1.py [↑](#footnote-ref-8)