

# 基于知识图谱的江南造山带金矿地质特征聚集性与找矿意义

李胜苗<sup>1,3</sup>, 贾立宁<sup>2</sup>, 王成彬<sup>2\*</sup>, 周丽芸<sup>1</sup>, 刘邦定<sup>1</sup>, 朱锦豪<sup>2</sup>, 王悦颖<sup>2</sup>, 李楠<sup>4</sup>

1 湖南省地质调查所, 湖南长沙, 410114

2 中国地质大学(武汉)资源学院, 湖北武汉, 430074

3 湖南省地质地理信息所, 湖南长沙, 410007

4 中国地质科学院矿产资源研究所, 北京, 100037

**摘要:** 为深化对江南造山带金矿成矿规律理解, 进一步评估找矿潜力。本文以江南造山带及其邻区的金矿为研究对象, 引入知识图谱技术, 采用自上而下方法构建金矿领域知识模型, 并综合利用深度学习与大语言模型(LLM)构建金矿成矿-勘查知识图谱; 基于该知识图谱开展金矿社区聚类分析与 Jaccard 相似性评价, 系统分析矿床聚集性特征。构建了含 28 类实体类型、10 种语义关系类型的金矿领域知识模型, 由此生成的知识图谱涵盖区域内 60 个代表性矿床, 包含 2212 条实体及 5497 条语义关系。社区聚类分析成功提取了“蚀变-矿物-地层”等关键控矿要素组合及成矿规律; Jaccard 系数分析显示, 水口山、黄金洞金矿与世界大型-超大型矿床具有高度相似性, 揭示了两矿床具有巨大的深部及外围的找矿潜力。

**关键词:** 知识图谱、江南造山带金矿、本体模型、社区聚类、Jaccard 系数

中图分类号: P628 收稿日期: 2025-12-07

## Clustering of Geological Characteristics and Prospecting Significance of Gold Deposits in the Jiangnan China Orogen Based on Knowledge Graphs

**Abstract** This study aims to deepen the understanding of metallogenic regularities and evaluate the prospecting potential of gold deposits in the Jiangnan Orogen. Focusing on the gold deposits within and adjacent to the Jiangnan Orogen, technologies related to the knowledge graph were introduced. A domain knowledge schema was developed using a top-down approach, and the metallogeny-exploration knowledge graph of gold deposits was constructed by integrating deep learning and Large Language Models (LLM). Community detection and Jaccard similarity evaluations were used to analyze the clustering characteristics of the gold deposits. The knowledge schema contains 28 geological entity types and 10 semantic relationship types. The resulting knowledge graph encompasses 60 representative gold deposits in the region, containing 2,212 geological entities and 5,497 semantic relationships. Community detection successfully extracted key ore-controlling factor combinations and metallogenic regularities, such as "alteration-mineral-strata." Jaccard

**基金项目:** 湖南省地质院重大项目 (HNGSTP202401)、国家重点研发计划项目 (2022YFF0801202)、地球深部探测与矿产资源勘查国家科技重大专项 (2024ZD1001205-05、2025ZD1007803) 联合资助。

**第一作者简介:** 李胜苗 (1983—), 男, 博士, 正高级工程师, 主要从事地质大数据构建、地质一张图和矿产智能预测领域研究。E-mail: [shengmiaoli@163.com](mailto:shengmiaoli@163.com), ORCID: 0009-0007-8787-2459

**通讯作者简介:** 王成彬 (1988—), 男, 博士, 副教授、博士生导师, 主要从事地学知识图谱构建与智慧应用、矿产资源智能预测方面研究工作。E-mail: [wangchb@cug.edu.cn](mailto:wangchb@cug.edu.cn), ORCID: 0000-0001-5964-1556。

similarity analysis indicates that the Shuikoushan and Huangjindong gold deposits have high similarities to global large-to-giant deposits, revealing significant prospecting potential in their deep-seated zones and peripheral areas.

**Keywords:** Knowledge Graph, Gold mine in the Jiangnan Orogenic Belt, Ontology Model, Community Clustering, Jaccard coefficient

## 0 引言

近年来,随着地学大数据与人工智能技术的迅速发展,知识图谱作为关键的知识表示与推理技术,被广泛应用于地球科学领域(邱芹军等,2023;郭飞等,2024)。作为结构化知识表示的重要方式,知识图谱为复杂地质知识的组织、推理和挖掘提供了新的技术途径(Zhang et al.,2023)。知识图谱在地质领域的应用已得到多项研究证实(Deng et al.,2024),其对地质知识的结构化呈现与深度挖掘能力也得到广泛认可(Dong et al.,2024)。在这一背景下,诸多学者开展了探索性研究:Wang et al.(2018)利用文本数据挖掘和可视化技术,实现地质文本数据中隐藏知识和关系的抽取与可视化表达;Ma et al.(2020)提出了一种新的结构,用于深时知识图谱的版本控制和跟踪;Enkhsaikhan et al.(2021)通过融合机器学习与地质文本大数据,实现地质文档报告的检索利用,提升了计算机对地学专业文本的理解能力;Qiu et al.(2023)提出了适应勘查报告的“对象-过程”双重关系模型,为地质知识表示提供了方式;Wang et al.(2023)从矿床地质特征的客观存在出发,采用用例驱动方法构建斑岩型铜矿的本体模型,构建斑岩型铜矿知识图谱。

地学知识图谱的发展推动了地学文本数据挖掘研究(Zhang et al.,2023),显著提升地学大数据二次利用效率(Peng et al.,2023;Zhou et al.,2021)。吴润泽等(2025)挖掘三峡地质资料中的地质灾害信息,构建三峡地质灾害知识图谱,服务与三峡滑坡地质灾害的检测预警;周成虎等(2021)系统构建了地学知识图谱的全域自适应表达模型、联邦式群智协同与多模态数据动态构建方法,并探讨地学知识图谱的应用方向。Shi et al.(2018)利用卷积神经网络(CNN)提取中国拉拉铜矿的实体关系;Qiu et al.(2020)利用地理字典匹配和关注机制构建字典关注层,并基于领域特定知识对地质文本进行分类;Li et al.(2022)通过远程监测,从少量标记样本中提取金矿的地质实体关系实现地质文献中金矿知识的抽取;Tian et al.(2024)将预训练的RoBERTa模型用于地质实体关系的抽取;Wang et al.(2022)提出了GeoERE-Net地质实体-语义关系联合抽取模型抽取地质文献中的三元组;Qiu et al.(2023)将ALBERT-BiLSTM-CRF模型用于地质实体关系的抽取。Yang et al.(2025)将扩展卷积神

神经网络（IDCNN）层纳入 BERT - BiLSTM - CRF 模型，抽取锂矿知识和知识图谱构建。马香菲（2025）将生成式数据增强、提示学习与 LoRA 参数高效微调相结合对大语言模型进行优化微调，用于地质灾害领域文本实体识别和关系抽取。王嘉翔（2024）以开源大语言模型（Chinese-LLaMA-2）为基础，经地学语料无监督预训练、人工标注数据集有监督微调，实现斑岩型铜矿领域三元组提取。当前研究仍以知识图谱的构建、知识表示与存储为主，如何挖掘地质知识图谱中的知识价值，驱动和服务矿产资源智能预测仍然是亟需解决的热点和难题。

本文根据江南造山带金矿地质特征构建成矿-勘查本体模型，综合利用深度学习和大语言模型方法对江南造山带及其邻区金矿地质文本进行命名实体识别和关系抽取，构建金矿知识图谱；然后对知识图谱中的实体类型和关系类型进行统计分析。在此基础上，对江南造山带金矿知识图谱进行社区聚类，基于聚类结果识别出内部关联紧密的地质实体集群，精准提取“蚀变-矿物-地层”等区域共性成矿规律，展现不同矿床的成矿模式差异。进而以水口山、黄金洞等典型矿区为研究对象，通过社区聚类分析其成矿要素特征聚集性和差异性；采用 Jaccard 相似性系数对比水口山和黄金洞金矿与已知的世界大型-超大型矿床的相似性，评价水口山和黄金洞金矿是否具有成为大型-超大型矿床的潜力。

## 1 金矿领域本体模型构建

基于江南造山带金矿成矿地质模型，结合矿床领域专家知识，采用自上而下的方法构建金矿领域本体模型（如图 1 所示），覆盖了金矿成矿系统和勘查系统。

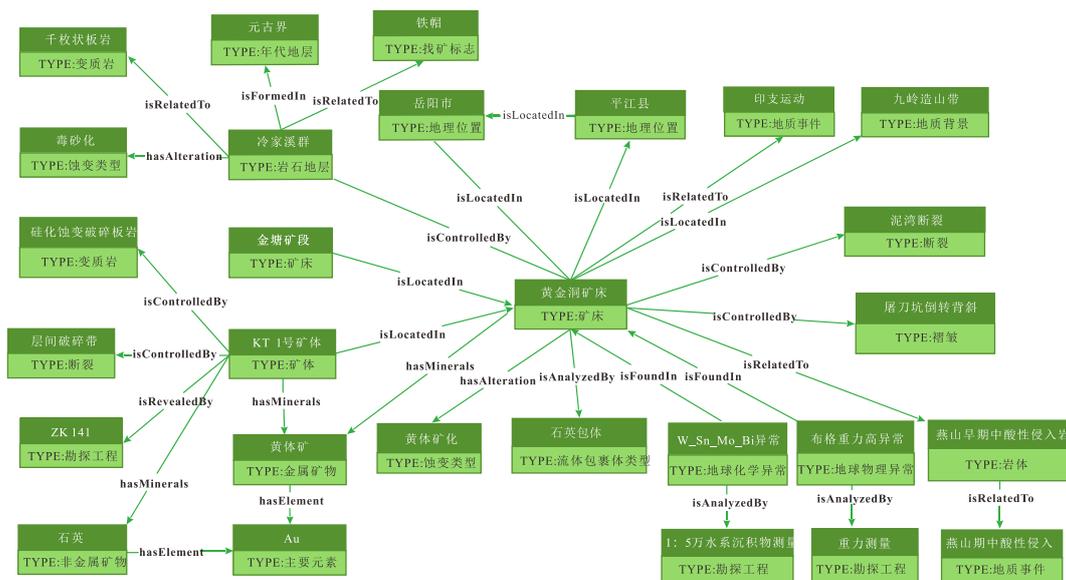


图 1 金矿成矿-勘查领域本体模型

Fig.1 Ontology model in the field of gold mineralization and exploration

## 1.1 实体类型定义

在深入理解江南造山带金矿成矿模型基础上,采用用例驱动的方法定义与江南造山带金矿相关的 28 种实体类型,即包括地理位置、地质年代、地质背景、地质事件、金属矿物、非金属矿物、断裂、褶皱、蚀变类型、沉积岩、火成岩、变质岩、岩石地层、年代地层、岩体、矿体、矿床、主要元素、微量元素、同位素、地球物理异常、地球化学异常、遥感异常、测量勘探仪器、勘探工程、找矿标志、矿化类型、流体包裹体类型。这些实体类型涵盖了矿床成矿模型和找矿模型的核心知识内容。

## 1.2 关系类型定义

除了与江南造山带金矿相关的实体类型,实体之间的语义关系是构成三元组,表征金矿知识的关键部分,它规定了各个实体之间的相互联系和交互方式。关系类型定义不仅描绘了实体之间的结构关系,还反映了它们之间的逻辑关系和动态交互。本文共构建 10 种语义关系类型,包括“hasAlteration”、“isControlledBy”、“isLocatedIn”、“isFormedIn”、“hasMinerals”、“hasElement”、“isRelatedTo”、“isFoundIn”、“isAnalyzedBy”和“isRevealedBy”,前 7 种主要表示金矿成矿系统中实体间的语义关系,后 3 种主要表示金矿勘查系统中实体间的语义关系。

语义关系定义了地质实体之间的相互作用和逻辑关系,其详细内涵如:hasAlteration 表示矿体或岩体附近出现的蚀变类型,如绿泥石化、黄铁矿化;hasMinerals 描述矿床或矿体中包含的金属或非金属矿物;hasElement 指岩石或矿物样品中的化学元素组合;isControlledBy 用于表征矿床的控矿构造,如构造破碎带;isRelatedTo 用于表征与矿床相关但非控矿的地质要素;isRevealedBy 指勘探工程(如钻孔、探槽)对矿体或岩体的揭示作用;isFoundIn 表示在矿床中发现的地质异常;isFormedIn 描述矿床或相关地层的形成年代;isLocatedIn 定义矿区的地理位置和地质背景;isAnalyzedBy 则描述某实体的勘查或分析方法,如汞气测量用于某研究区的勘查。这些关系类型系统地刻画了矿床地质中各要素的相互作用,为构建金矿的知识画像提供了实体-关系语义框架。在该实体和语义关系的本体模型中,一种语义关系的头实体和尾实体可以多种实体类型组合,从而表达复杂的地质知识;同时较少的实体和语义关系数量也兼顾了地质文本数据挖掘计算机模型训练的需求。

## 2 江南造山带金矿知识图谱构建

本文使用的地质文本数据包括水口山矿区和黄金洞矿区的 46 篇地质勘查报告和江南造山带金矿的 120 篇开放性学术文献。通过对地质勘查报告进行清洗和标注,构建 1000 多条

训练语料，利用深度学习对地质勘查报告中实体和关系进行抽取；针对开放性学术文献则采用大语言模型完成实体与关系的提取。将不同方法抽取的实体-关系三元组进行对齐和消歧校正，建立江南造山带金矿知识图谱。

## 2.1 知识图谱构建方法

### (1) 基于词增强深度学习的命名实体识别

该模型采用字-词特征融合的方法，通过 Word2Vec 模型提取词特征，同时利用 BERT 模型获取字级特征、位置特征和句级特征，并借助适配器实现特征的深度融合（Vaswani et al., 2017）。模型采用双向 Transformer 结构来提取上下文特征，相比传统 RNN 和 CNN 模型，能够更好地捕捉长距离依赖关系并避免梯度问题。在输出层引入条件随机场（CRF）模型，将命名实体识别任务转化为多分类问题，通过考虑相邻标签之间的转移概率来优化最终的标注结果。实验结果表明，该模型的精确率达到 0.88，召回率为 0.84，F1 值为 0.86，性能指标表明模型在实体识别任务中兼顾了较高的识别准确性与良好的实体覆盖范围，具备可靠的实用价值。

### (2) 基于 BERT-CasRel 语义关系抽取模型

BERT-CasRel 模型实现金矿领域语义关系提取主要经过三步（Wei et al.,2020）。第一步使用 BERT 编码器将输入文本中的每个单词转换为相应的词向量表示；第二步使用 CasRel 模型的头实体识别模块，直接解码  $n$  层 BERT 编码器生成的编码向量，识别输入句子中所有可能的头实体。第三步关系对象识别模块根据每个头实体预测对应的语义关系和尾实体。实验结果显示，该语义关系抽取模型的性能更为突出，精确率达 0.94，召回率达 0.91，F1 值为 0.93。

### (3) 基于大语言模型的实体关系抽取模型

深度学习方法对复杂语义场景、长文本处理上存在一定的局限性，并且深度学习方法需要大量的人工标注训练语料作为支撑，训练语料的构建需要大量的劳动力和经济成本。相较于深度学习模型，大语言模型不仅在长文本处理、复杂语义推理方面表现突出，更具备显著的少样本或零样本学习优势，无需依赖大规模标注语料，即可高效完成实体与语义关系抽取任务（Li et al.,2024；Zhou et al.,2024）。在金矿本体模型的引导下，编写 Prompt 指令开展金矿文本中实体和关系的抽取。在 Prompt 指令中首先明确大语言模型的服务的领域和角色，描述对金矿知识图谱相关的实体和语义关系的定义；然后对大语言模型抽取结果的存储形式进行约定。实验采用本地部署的 DeepSeek-R1-70B 模型，采用基于大语言模型的 Prompt 方

法对实体-语义关系抽取的精确率达 0.89，召回率达 0.82，F1 值为 0.85。

本文综合利用基于词增强的深度学习命名实体识别、BERT-CasRel 语义关系抽取模型和大语言模型方法对江南造山带金矿相关的地质实体及其之间的语义关系进行抽取，然后经消歧对齐处理，获得涵盖 60 个金矿床相关的 2212 个实体和 5497 条语义关系的江南造山带金矿知识图谱（如图 2 所示）。

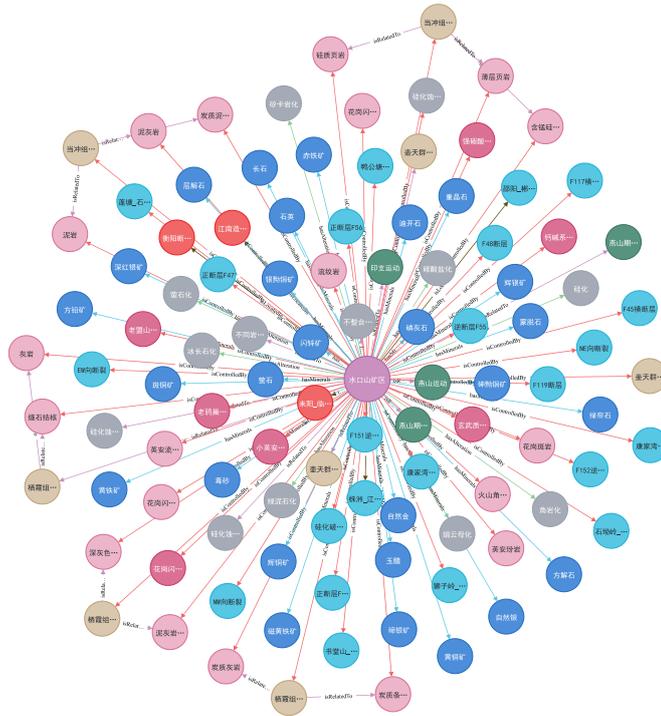


图 2 江南造山带金矿部分知识图谱

Fig.2 Partial knowledge graph of gold mines in the Jiangnan orogenic belt

## 2.2 实体类型统计分析

江南造山带金矿知识图谱中实体类型统计分布（如图 3 所示），该知识图谱的实体类型全面覆盖了矿床成矿模型和勘查模型。在矿床物质基础层面，沉积岩 118 个、变质岩 85 个、火成岩 20 个、岩体 81 个、金属矿物 77 个、非金属矿物 46 个、主要元素 30 个、微量元素 21 个、同位素 30 个，共 513 个实体，占全部实体数量的 23.19%。沉积岩和变质岩作为主要岩性为成矿提供了物质来源和赋矿空间；侵入岩体提供成矿热源与物质；其中与成矿相关的金属矿物数量较多，非金属矿物则主要反映成矿岩石的组成特点或蚀变类别。在矿床时间维度上，地质年代实体 13 个、年代地层 28 个、地质事件 8 个，共 49 个实体，占全部实体数量的 5.8%，该类实体可为地质历史演化和地质成矿时间提供时间框架约束。在矿床构造方面，断裂实体 213 个、褶皱实体 88 个，共 301 个实体，占全部实体数量的 13.61%，断裂和

褶皱体现江南造山带金矿“构造控矿”特征。在矿床的流体和蚀变作用维度，蚀变类型实体 48 个、流体包裹体类型实体 18 个，共 66 个实体，占全部实体数量的 1.2%。蚀变类型反映热液流体与围岩相互作用，流体包裹体则与成矿流体物理化学条件密切相关。在矿床空间分布上，地质背景相关的实体 113 个、地层 98 个，共 211 个实体，占全部实体数量的 9.53%，地质背景和地层相关的实体揭示了成矿的区域构造背景和地层条件。在空间位置层面，地理位置 108 个，占全部实体数量的 5.33%。在勘查方法与找矿标志方面，勘探工程 311 个、测量勘探仪器 25 个、找矿标志 44 个、地球化学异常 152 个、地球物理异常 41 个、遥感异常 1 个、重砂异常 1 个，共 575 个实体，占实体数量的 25.99%，勘探工程可直接揭露地下地质体形态，测量勘探仪器是获取勘查数据的重要手段。地球化学异常用于圈定元素富集区，也可示踪深部矿化信息、追溯成矿物质来源；地球物理异常可定位深部隐伏矿体以及构造、岩体的分布；遥感异常可作为大区域尺度靶区圈定，找矿标志能够快速指示成矿有利地段。

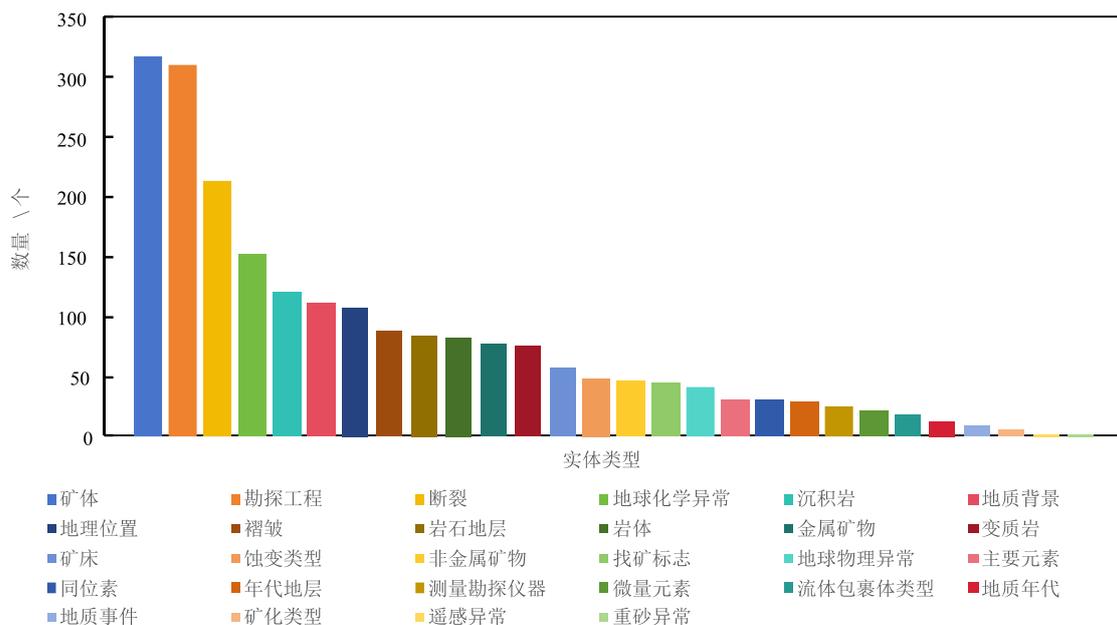


图 3 江南造山带金矿知识图谱中实体统计分布

Fig.3 Entity statistics distribution in the knowledge graph of gold deposits in the Jiangnan orogenic belt

### 2.3 关系类型统计分析

江南造山带金矿知识图谱中语义关系类型统计分布（如图 4 所示），江南造山带金矿知识图谱的 10 类语义关系共计 5497 条。其中，isControlledBy 共计 1434 条，占语义关系数量的 26.9%，该语义关系主要用于表达矿床（或矿体）与岩体、断裂、地层等地质要素之间的

控制关系。*isRelatedTo* 共计 698 条，占语义关系数量的 11.9%，该语义关系用于描述与成矿相关但起非控矿作用的岩石、地层、构造等地质要素。*hasMinerals* 共计 819 条，占语义关系数量的 14.90%，该语义关联矿床（或矿床）与金属、非金属矿物，表示矿床的矿物组合。*isLocatedIn* 共 807 条，占语义关系数量的 14.68%，实现矿床和矿体与地理位置、地质背景的空间语义描述。*hasAlteration* 共 626 条，占语义关系数量的 11.39%，该语义连接矿床和矿体与蚀变类型。*isRevealedBy* 共计 372 条，占语义关系数量的 6.77%，该语义用于表示矿床（或矿体）与勘探工程之间的连接关系，体现勘查工程对矿床和矿体的揭示作用。*hasElement* 共 295 条，占语义关系数量的 5.6%，该语义连接矿石或样本与化学元素。*isFoundIn* 共 205 条，占语义关系数量的 3.73%，该语义关联地质异常与地质单元。*isFormedIn* 共 129 条，占语义关系数量 2.35%，该语义表征矿床形成年代。*isAnalyzedBy* 共 112 条，占语义关系数量 2.04%，该语义用于表示矿床采用的分析方法。

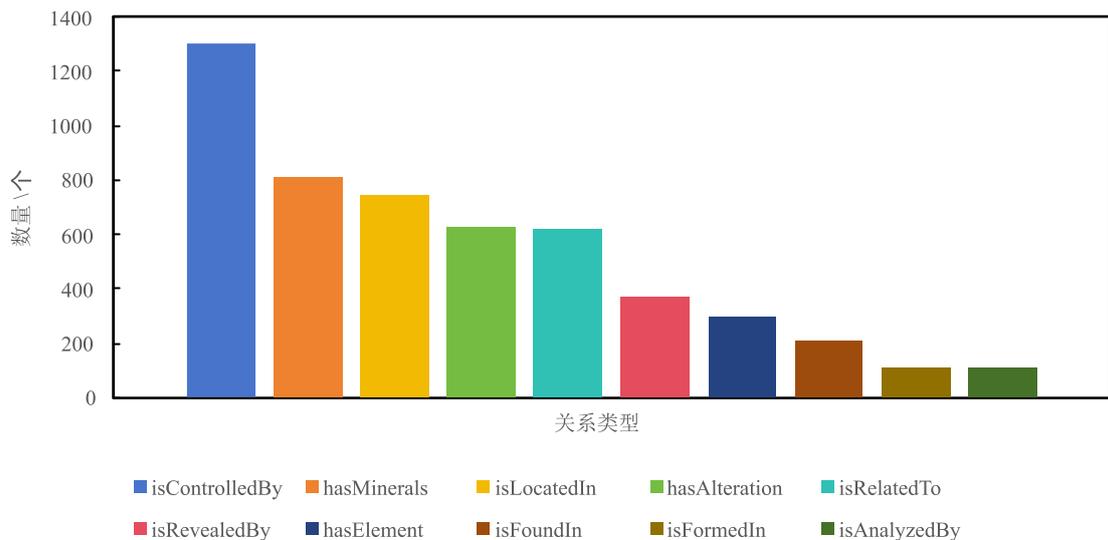


图 4 江南造山带金矿知识图谱中语义关系统计分布

Fig.4 Semantic relation statistics distribution in the knowledge graph of gold deposits in Jiangnan orogenic belt

### 3 江南造山带金矿知识图谱对比

本文采用社区检测的方法对江南造山带金矿知识图谱进行挖掘和分析。社区检测的核心是将网络划分为内部连接紧密、外部连接稀疏的社区集群。模块度 (Modularity) 通过量化实际网络连接与随机连接模型的差异衡量社区划分质量，其值越接近 1 表明社区结构越显著 (Newman et al.2006)。Louvain 算法 (Blondel et al.2008) 通过“局部优化-层次聚合”的迭代策略实现高效社区检测，其核心逻辑围绕模块度计算与增量优化展开，执行流程分为两个

核心步骤：①局部优化：将每个节点初始化为独立社区，随机遍历所有节点，计算节点迁移至各相邻社区的模块度增量  $\Delta Q$ ，仅保留使  $\Delta Q$  最大化的迁移操作。由于  $\Delta Q$  直接反映节点移动对全局  $Q$  的贡献，因此每次  $\Delta Q > 0$  的迁移都会严格提高整个网络的模块度该过程不断迭代，直至不存在任何节点的迁移能够进一步提升全局模块度，此时完成一次局部最优社区划分。②层次聚合：将局部优化后的每个社区抽象为“超级节点”，以社区间节点的边权重和作为超级节点间的连接权重，构建新的层次化网络。重复上述两个步骤直至模块度达到最大值，最终输出多层次社区结构。

其中，模块度是衡量社区划分优劣的基础，对于含  $n$  个节点、 $m$  条边的无向网络，模块度  $Q$  的计算公式如下公式 1 所示：

$$Q = \frac{1}{2m} \sum_{i,j} [A_{ij} - \frac{k_i k_j}{2m}] \delta(c_i, c_j) \quad (1)$$

其中  $A_{ij}$  表示邻接矩阵元素，表示节点  $i$  和节点  $j$  之间是否存在边。若两节点间存在边连接，则  $A_{ij} = 1$  否则  $A_{ij} = 0$ ； $k_i$  和  $k_j$  表示节点  $i$  和节点  $j$  的度数，即各自连接边的数量； $m$  表示整个网络中边的数量； $\delta(c_i, c_j)$  为 Kronecker delta 函数，当节点  $i$  和节点  $j$  属于同一个社区时取值为 1，否则为 0。

Louvain 算法的高效性源于将全局模块度优化转化为局部节点迁移决策，核心是模块度增量  $\Delta Q$  的计算。当节点  $u$  从原社区  $C$  迁移至相邻社区  $C'$  时， $\Delta Q$  可通过局部信息求解，如下公式 2 所示：

$$\Delta Q = \frac{1}{2m} [2k_{u,in} - \frac{k_u \sum_t}{m}] \quad (2)$$

其中  $k_{u,in}$  为节点  $u$  与社区  $C'$  内节点的连接边数为社区  $C'$  内所有节点的度总和； $k_u$  为节点  $u$  的总度。若  $\Delta Q > 0$ ，说明迁移可提升模块度，算法将选择  $\Delta Q$  最大的社区完成节点迁移。

### 3.1 江南造山带金矿知识图谱网络分析

模块度社区聚类算法主要涉及四个参数，分别是随机度、边权重、解析度和聚类起始标识号 (ID)。随机度是每次运行算法时随机产生种子，进行路径发现得出社区聚类结果；边权重由关系类型及其属性（如置信度、频次）的动态计算，以体现不同关系对社区结构的贡献差异；解析度通过调整社区内部与社区之间边权重的相对权重影响社区聚类结果的数量，

解析度越大，获得的聚类数越少；聚类起始标识号决定网络分析聚类的起始编号。

经社区检测聚类运算，得到了不同解析度下金矿知识图谱社区发现的模块度与聚类数量结果之间的对应关系（如图 5 所示）。在 $[0, 0.2]$ 解析度区间内的模块度集中在 0.4 到 0.52 之间，聚类数大于 60，聚类数过多，缺乏地质可解释性；在 $[0.2, 0.6]$ 解析度区间内模块度随解析度增长变缓，聚类数从 62 降到 21，但聚类数依旧过多，不利于地质解译；在 $[0.6, 1.1]$ 解析度区间内模块度出现“平台”，聚类数降至 17 个左右；解析度大于 1.1 后，模块度随解析度变大而降低，说明随着解析度变大，网络的社区聚类效果变弱。综合模块度与聚类数在不同解析度下的变化趋势，选择解析度为 1 时的结果作为社区聚类最优结果。在解析度为 1、增加边权重、随机性为真的条件下，通过多次独立运行算保证结果稳定性，最终得到金矿知识图谱的社区结构，其模块度为 0.55，共识别 19 个聚类结果。

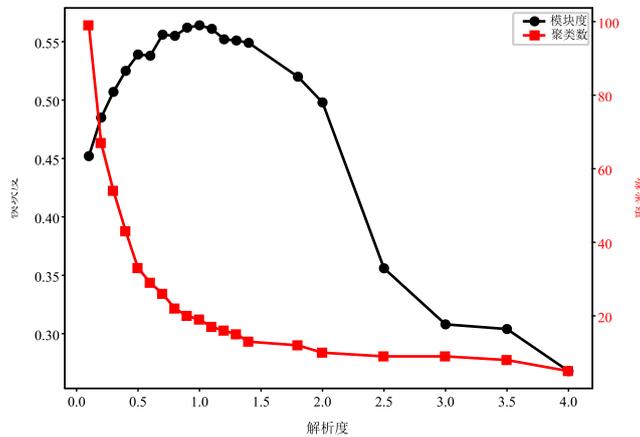


图 5 模块度、聚类数随解析度变化趋势图

Fig. 5 Trends of modularity and number of clusters changing with resolution

金矿知识图谱的知识节点在不同社区聚类中差异明显，如图 6 所示。其中 ID=8、18 的类别包含较多的节点，节点数量大于 200 个；聚类 ID 为 3、10、15、16 的 4 个聚类包含的节点数量为 100-200 个；聚类 ID 为 1、2、12、13 等 13 个聚类包含的节点数量小于 100 个。

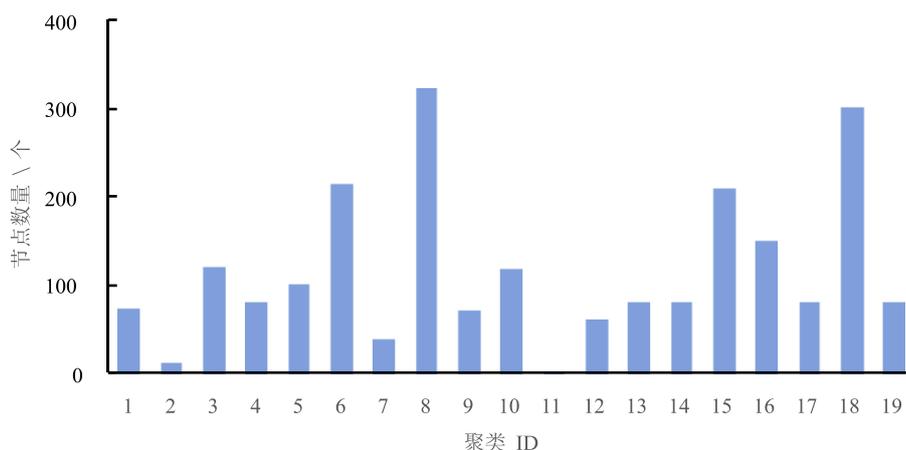


图 6 江南造山带矿床实体标签聚类节点分布

Fig. 6 Distribution of entity label clustering nodes of ore deposits in Jiangnan orogenic belt

选取类别 8、类别 15、类别 18 和类别 6 这 4 个典型类别进行详细分析，重点分析与金矿成矿模型和勘查模型密切相关的地质知识信息：

(1) 类别 8 聚类结果揭示了水口山矿区内相关矿床和成矿地质要素。该类别共包括 5 处相关矿床（段），具体包括白泥冲黄铁矿、龙王山金矿、老鸦巢、鸭公塘矿段及水口山矿田，均隶属于水口山矿区（如图 7a 所示）。这一结果充分印证了社区聚类算法的有效性与合理性，聚类边界清晰且符合实际地质规律。除上述矿床实体外，该类别中还包含大量与矿床相关的地质要素，比较集中的实体标签有岩石地层、沉积岩、断裂、褶皱等。岩石地层中，高家田组、栖霞组、当冲组与矿床或矿体之间的语义关系为 `isControlledBy`，表明这些地层是控制矿床和矿体分布的重要地层单元；沉积岩中，灰岩、砂岩和页岩与矿床或矿体之间的语义关系为 `isControlledBy`，表明这些沉积岩对水口山矿区中矿床有着重要影响；在构造地质方面，在聚类中金矿与 NE、NW 向断层和其他矿区内断层密相关，表明矿床的空间分布受这些断裂控制；褶皱构造方面，聚类结果包含康家湾倒转背斜、鸭公塘倒转背斜等褶皱实体，这一特征表明褶皱构造在该类型矿床的成矿过程中发挥了关键控制作用。以上结果表明地层、断裂、和褶皱等地质要素对水口山矿区金矿形成密切关系。

(2) 类别 15 聚类结果揭示了黄金洞矿区内相关矿床和成矿地质要素。该类别包括 11 个矿床，这些矿床均属于黄金洞金矿区或者在黄金洞矿区附近（如图 7b 所示）。除矿床实体外，该类别比较集中的实体标签有岩石地层、变质岩、断裂、蚀变类型，岩石地层中，冷家溪群、坪原组和黄浒洞组与矿床或矿体之间的语义关系为 `isControlledBy`，表明这些地层是控制矿床和矿体分布的重要地层单元；在构造地质方面，在聚类中矿床与 NE、NNW 和 NW 向断裂密切相关，表明矿床的空间分布受这些断裂控制；褶皱构造方面，在聚类中矿床与层



Fig. 7 Gold deposits and their geological characteristics present in different categories

### 3.2 江南造山带金矿知识图谱 K-core 核心关联特征

K-core 算法是一种基于图拓扑结构的子图挖掘算法，旨在挖掘满足指定核心度要求的关联子图结构 (Kong et al., 2019)。K-core 子图中，每个顶点在该子图内的度数均不小于 K，即每个顶点至少与子图中的其 K 个顶点存在直接语义关联。该算法通过剔除网络中度数未达到阈值 K 的非核心顶点，层层剥离网络外围的松散结构，从而获得核心子图。

本文拟利用 K-core 算法探索与矿床关系密切的控矿要素，对成矿模型的理解。当 K 值为 1 或 2 时，核心子图中包括地理位置、勘探工程等与成矿模型关联性较弱的实体与关系，筛选力度不足，不利于对矿物、断裂、地层等核心成矿要素的规律分析；当  $K \geq 4$  时，一些矿床、岩石地层、断裂等重要信息被排除，破坏“矿床-成矿要素”完整关联。当 K 值为 3 时，在剥离与成矿模型低价值边缘信息的同时又最大限度保留核心地质要素信息，精准达成“去冗存核”的筛选目标。

因此，选择 K 值为 3 构建江南造山带金矿知识图谱的核心子图 (如图 8 所示)。核心子图表明：金矿矿床均与黄铁矿化、毒砂化、绢云母化三类蚀变及黄铁矿、毒砂、方铅矿三种矿物紧密关联；黄铁矿与毒砂是该区域金矿典型的金属矿物组合，黄铁矿与毒砂是核心载金矿物，方铅矿指示成矿流体的多金属成矿特征，这一关联集群印证了“蚀变-矿物-金矿化”的强耦合关系，为地质找矿提供了金属矿物组合标志。岩石地层实体高度聚焦于板溪群、五强溪群、马底驿组，其中板溪群与五强溪群是江南造山带基底地层，以浅变质碎屑岩、火山碎屑岩为主，富含 Au、As 等成矿元素 (柏道远等, 2025)，为金矿形成提供物质基础与赋矿地层；马底驿组与基底地层的不整合接触带因构造活动易形成热液流体运移通道与矿体赋存空间。在断裂方面，NW 向、NE 向、EW 向断裂作为大部分矿床的控矿断裂 (肖丹等, 2025)，NE 向断裂控制成矿带宏观展布，NW 向断裂常与 NE 向断裂交叉部位是热液流体聚集与成矿物质沉淀的核心区，EW 向断裂约束前两者分布，三者共同构成“构造控矿网络”，揭示“构造控矿”成矿规律，为金矿找矿勘查提供明确的构造方向指引。

基于 K-core 筛选的核心金矿知识图谱子图通过知识图谱子图中控矿要素与金矿之间的语义逻辑关系可视化表达，有助于加深江南造山带金矿的成矿模型和找矿勘查的快速深化理解。在成矿模式层面，明确了“构造-地层-蚀变-矿物”成矿要素耦合机制：板溪群、五强溪群、马底驿组的特定地层组合为成矿提供了物质基础与容矿空间；NW 向与 NE 向断裂的交汇部位构成流体运移与成矿物质富集通道；“黄铁矿化-毒砂化-绢云母化”蚀变组合与“黄铁矿-



分析,结果显示不同社区由岩石地层、沉积岩、岩体、褶皱、断裂和金属矿物等地质要素构成。类别7中较为集中的实体标签为岩石地层和沉积岩(如图9a所示),其中高家田组地层、栖霞组地层、当冲组地层单元占比较大,地层中的岩石类型主要为灰岩、砂岩、泥岩等沉积岩类,反映出矿区基底的沉积环境特征。类别8则以岩体和蚀变类型实体标签为主(如图9b所示),岩体主要为花岗闪长斑岩、石英斑岩等岩体,以及绢云母化、绿泥石化、高岭土化、黄铁矿化等蚀变类型,岩体与矿体及蚀变的显著关联,表明矿床形成与岩浆侵入活动及其伴随的蚀变作用存在密切的成因联系。类别6的实体标签主要是褶皱和断裂(如图9c所示),其中褶皱主要为康家湾倒转背斜、盐湖式向斜、仙人岩倒转背斜为主,断裂则以NW向和NE向断裂为主,表明矿床形成与分布受特定褶皱和多方向断裂共同控制。类别9主要是矿体与金属矿物的实体标签(如图9d所示),金属矿物以闪锌矿、方铅矿、黄铁矿、银黝铜矿多种硫化物为主,表明上述金属矿物是构成矿体的主要矿物成分。综上所述,水口山矿区形成“沉积地层-岩浆活动-褶皱和断裂构造”的成矿模式:高家田组等沉积地层为矿质富集提供容矿空间与物质基础,花岗闪长斑岩等岩体侵入提供成矿热源与成矿物质,褶皱核部及NW和NE向断裂交汇带构成导矿和容矿体系。基于该成矿模式可构建以下找矿模式:以高家田组-栖霞组-当冲组地层组合、花岗闪长斑岩等侵入岩体、NW和NE向断裂与褶皱构造的区域为核心找矿靶区,将绢云母化-黄铁矿化等蚀变组合与闪锌矿-方铅矿-黄铁矿等硫化物矿物组合作为找矿标志,深部探测重点聚焦褶皱核部破碎带、断裂交汇处及岩体与地层的接触带等隐伏部位。该模式可为水口山矿区的精准勘查提供一定科学指导。

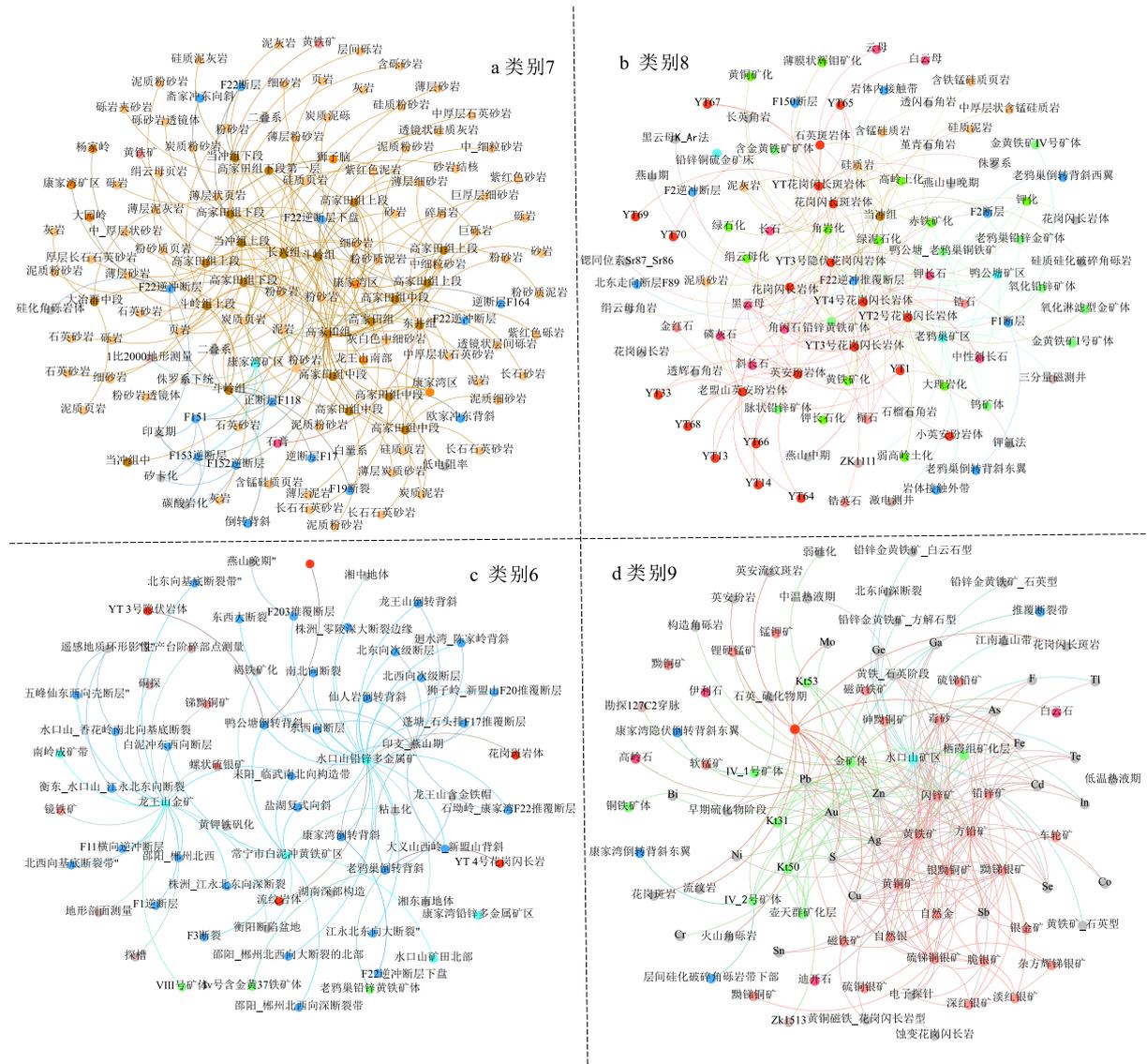


图9 水口山矿区不同社区中存在的地质要素

Fig. 9 Geological elements existing in different communities in Shuikoushan mining area

(2) 为揭示黄金洞矿区地质要素的内在联系, 对该矿区的知识图谱聚类结果进行了系统分析, 结果显示不同社区主要由岩石地层、变质岩、金属和非金属矿物、断裂等地质要素构成。类别 8 和类别 3 中较为集中的实体标签包括岩石地层、变质岩、金属矿物和非金属矿物 (如图 10a 和 10b 所示)。岩石地层以冷家溪群为主, 地层中的变质岩主要为绢云母板岩、泥质板岩等, 反映出矿区基底的变质岩属性; 金属矿物主要为黄铁矿、毒砂、闪锌矿等, 非金属矿物主要为石英、方解石、绢云母等, 其中黄铁矿和毒砂是核心载金矿物, 闪锌矿、黄铜矿等作为伴生硫化物, 其出现指示成矿流体富硫且具备金元素迁移和富集的化学环境; 而石英作为热液活动的重要产物, 为金沉淀和富集提供有利空间。类别 9 实体标签主要为蚀变类型, 主要包括黄铁矿化、毒砂化、硅化、绿泥石化等蚀变 (如图 10c 所



Fig. 10 Geological elements existing in different communities in Huangjindong mining area

### 3.4 水口山矿区和黄金洞矿区与大型-超大型矿床相似性分析

Jaccard 相似性系数是衡量两个集合相似程度的经典指标，其核心原理为通过计算两集合交集元素数量与并集元素数量的比值表征两集合的重叠程度，取值范围为 0-1，数值越接近 1 表明两集合的相似性越高(Niwattanakul et al., 2019)。为客观分析成矿规律并科学的评估成矿潜力，本文选取金属矿物、火成岩、断裂等 13 个与成矿相关的地质要素构建特征集合，基于公式 3 计算水口山和黄金洞矿区与大型-超大型矿床之间 Jaccard 相似性系数，量化矿床间地质要素的重叠程度并且规避定性对比的主观性。

$$J(A, B) = |A \cap B| / |A \cup B| \quad (3)$$

本文选取 Kalgoorlie 矿床、Muruntau 矿床、Timmins 矿床、Grasberg 矿床、Pueblo Viejo 矿床、Ladolam 矿床、Lihir 矿床 7 个典型大型-超大型矿床作为对比对象，系统计算了水口山矿区、黄金洞矿区与上述矿床在 13 个成矿相关地质要素的 Jaccard 相似性系数（图 11、12）。对相似性计算结果进行拆解与系统分析后，将两矿区与世界大型-超大型矿床的相似性特征划分为以下方面进行对比研究：

（1）两个矿区在矿物指标上表现出较高相似性但矿物类别存在差异。黄金洞矿区以金属矿物为核心优势指标，平均相似性达 0.70，与 Timmins、Muruntau 等矿床均超过 0.70，凸显金属组分与大型-超大型金矿的高度相似。水口山矿区则以非金属矿物为优势指标，平均相似性高达 0.78，与 Pueblo Viejo 矿床达 1，与 Lihir、Ladolam 等矿床均超 0.80，反映伴生非金属组分与大型-超大型矿床的广泛关联性。

（2）两个矿区在构造和地质事件指标上表现中等相似性。水口山矿区的地质事件相似性为 0.49 和断裂相似性为 0.44 指标平均相似性均处于中等水平，其中地质事件与 Lihir 矿床达 1，断裂与 Pueblo Viejo 矿床达 0.67，显示构造运动、岩浆活动期次与部分大型-超大型矿床的强关联性，并且断裂控矿作用突出。黄金洞矿区的断裂指标平均相似性 0.35 略低于水口山，与 Kalgoorlie 矿床相似性为 0.67 关联最强，而地质事件平均相似性仅 0.07，仅与 Lihir 矿床相似性为 0.5 存在微弱关联，反映其成矿地质事件的独特性。

（3）两个矿区在成矿基础条件表现低相似性。岩石地层指标均处于低水平，黄金洞矿区仅与 Pueblo Viejo 矿床相似性达到 0.50，平均相似性为 0.10，水口山矿区平均相似性仅为 0.01。地质背景指标平均相似性相近，两个矿区与 Timmins 矿床和 Ladolam 矿床相似性达到 0.5，两个矿区的平均相似性为 0.21。变质岩指标中，黄金洞矿区全部为 0，水口山相似性平

均 0.24 但无矿床达高相似; 褶皱指标均极弱, 黄金洞矿区仅 Timmins 矿床为相似性达到 0.50, 黄金洞矿区平均相似性为 0.21, 水口山矿区仅与 Kalgoorlie 矿床相似性达到 0.5, 水口山矿区平均相似性为 0.07。以上内容表明两者的岩石地层、宏观构造背景、褶皱等成矿基础条件, 均与多数超大型金矿存在根本性差异。

(4) 两个矿区在找矿标志及岩体指标上表现极低相似性并且岩体指标分异显著。黄金洞矿区仅与 Pueblo Viejo 矿床和 Kalgoorlie 矿床的相似性为 0.50, 黄金洞矿区相似性平均为 0.14, 水口山矿区与 Pueblo Viejo 矿床的相似性为 0.50, 水口山矿区平均相似性为 0.10, 表明两矿区均缺乏与世界大型-超大型矿床匹配的典型找矿线索。岩体指标则呈现显著分异, 黄金洞矿区与 Timmins 矿床在岩体指标上达 1, 与其余大型-超大型矿床均为 0, 该结果与矿区实际成矿弱依赖岩体的特征形成反差, 需重新审视岩体指标的界定标准, 或反映其与 Timmins 矿床在“非直接岩体控矿但岩体间接影响成矿流体”等隐性关联上存在共性。水口山矿区与 Ladolam 矿床达 0.50, 平均相似性 0.20, 虽然相似性较低, 但是关联范围略广, 具有一定的对比意义。

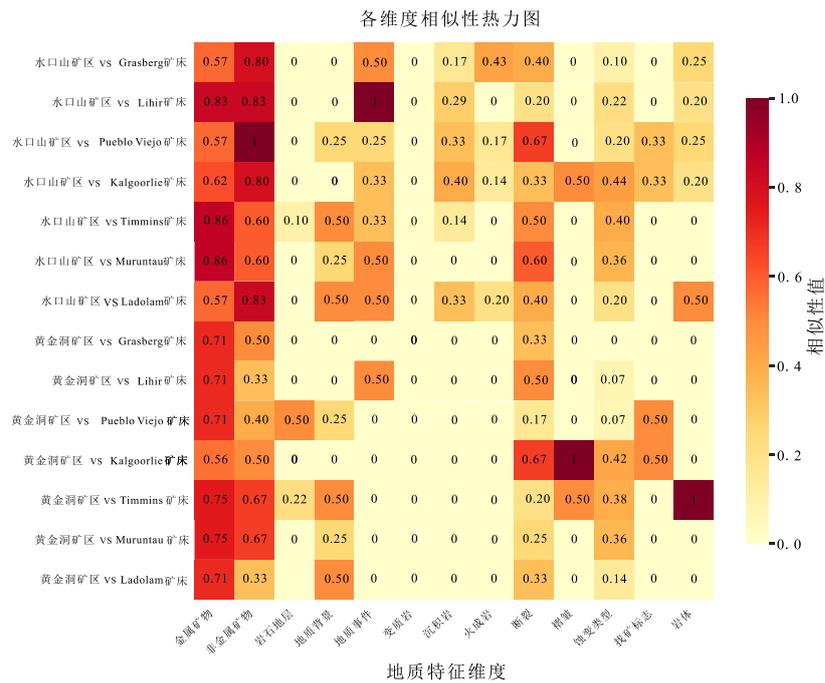


图 11 水口山矿区与黄金洞矿区与大型-超大型矿床地质要素 Jaccard 相似性热力图  
 Fig.11 Jaccard similarity heatmap of geological elements between the Shuikoushan mining area, Huangjindong mining area and large-super large deposits

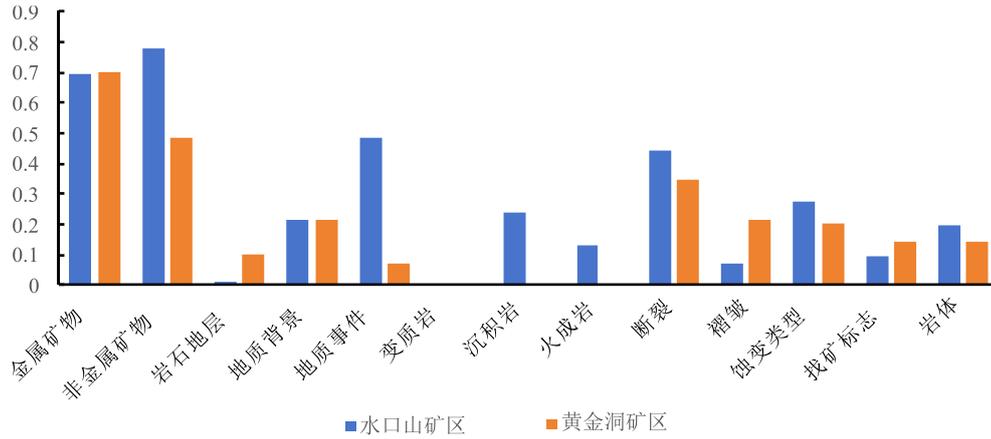


图 12 水口山矿区和黄金洞矿区与大型-超大型矿床地质要素 Jaccard 相似性平均值  
 Fig.12 Average Jaccard similarity values of geological elements in Shuikoushan mining area and Huangjindong mining area with large-super large deposits

(5) 两大矿区与 7 个大型-超大型金矿的综合相似性均存在显著差异，且整体相似性水平高相似呈现明显分异（如图 13、14 所示）。水口山矿区的综合相似性平均值为 0.45，处于中等偏低水平，与 Muruntau 矿床的匹配度最高，相似性为 0.58，接近中等相似程度，与 Grasberg 矿床的匹配度最低，相似性为 0.39，整体呈现“单矿床高匹配、其余均布中等”的格局。黄金洞矿区的综合相似性平均值为 0.32，低于水口山矿区，处于较低水平，与 Timmins 矿床的匹配度最高，相似性为 0.51，是唯一接近中等相似的矿床，与 Pueblo Viejo 矿床的匹配度最低相似性为 0.21，呈现“两极分化”的特征。两者共同特点是与 Grasberg 矿床的综合匹配度均最低，但水口山矿区整体相似性水平更高，且高匹配矿床的综合相似值略高于黄金洞矿区的最高值。

综上所述，黄金洞矿区虽在金属矿物维度与 Timmins、Muruntau 等超大型金矿具有高度相似性，但受限于地层、构造、岩石类型等成矿基础条件与多数大型-超大型金矿的根本性差异，且岩体控矿特征不显著、找矿标志匹配度低，其形成超大型矿床的潜力相对受限，该矿区可借鉴 Timmins 矿床的非直接岩浆控矿模式，将构造-地层耦合带及深部隐伏构造界面作为重点勘查靶域。水口山矿区的非金属矿物与 Pueblo Viejo、Lihir 等超大型矿床的高似性，该矿区褶皱、断裂控矿作用突出，契合超大型矿床构造控矿特征，并且与 Muruntau、Ladolam 矿床可构建多维度类比体系，具备形成大型-超大型矿床的更高潜力，水口山矿区应依托其与 Muruntau、Ladolam 矿床的多维相似性特征，聚焦断裂构造交汇处及岩性接触带的深部延展，深部矿化的探明将成为其潜力释放的关键。

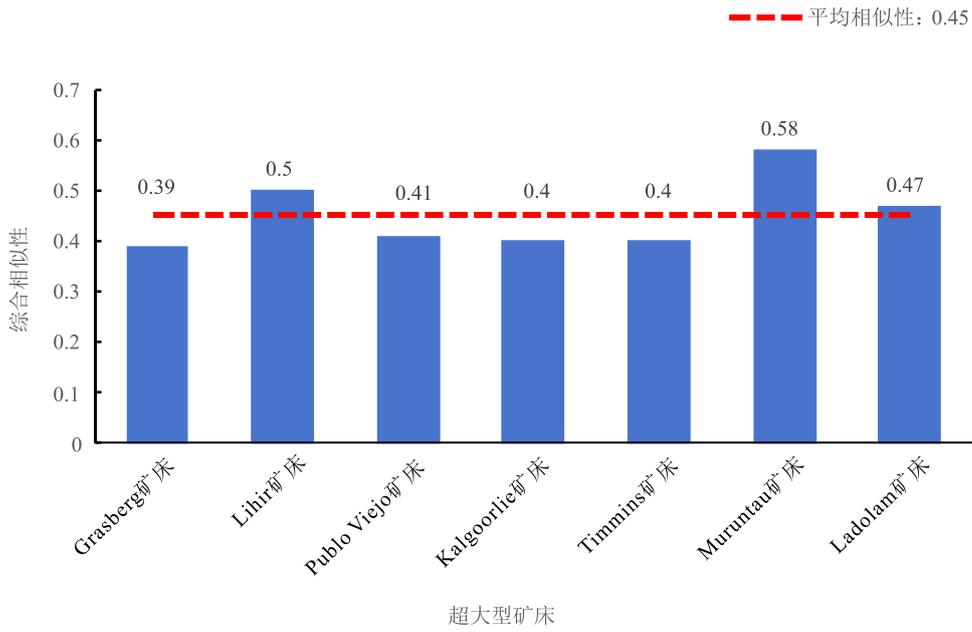


图 13 水口山矿区与大型-超大型矿床的综合相似性对比

Fig.12 Comprehensive similarity comparison between Shuikoushan mining area and large - super large deposit

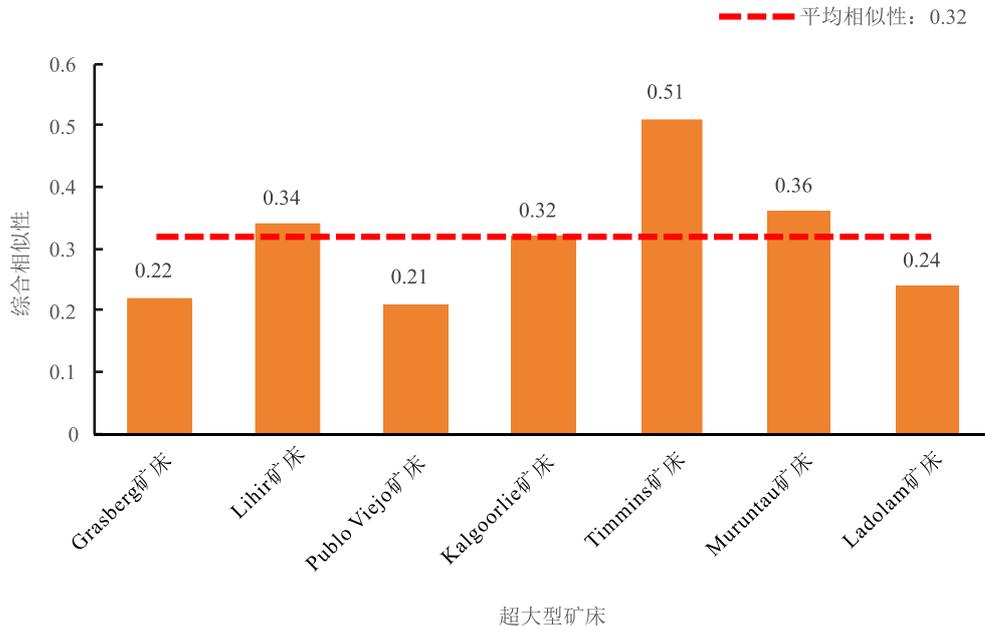


图 14 黄金洞矿区与大型-超大型矿床的综合相似性对比

Fig.13 Comprehensive similarity comparison between Huangjindong mining area and large - super large deposit

## 4 结论

本研究通过对江南造山带金矿的知识图谱的社区聚类分析以及与大型-超大型矿床进行相似性分析,为矿床知识图谱挖掘提供一定的参考性解决方案,本研究取得主要成就如下所示:

本文以江南造山带金矿为研究对象,采用自上而下构建方法,融合多维数据与专家知识,确立涵盖 28 类实体、10 种语义关系的成矿-勘查领域本体,覆盖地层、构造、岩浆岩等关键要素及成矿-勘查系统语义关联,为知识图谱构建提供基础。基于该本体,通过词增强深度学习命名实体识别、BERT-CasRel 关系抽取和大语言模型方法及消歧校正,构建包含 60 个金矿床相关 2212 个实体、5497 条语义关系的知识图谱。

通过模块度社区聚类算法,将江南造山带金矿知识图谱划分为 19 个类别,结合  $k=3$  的 K-core 分析剥离边缘信息,精准提取出蚀变组合、优势赋矿地层、构造控矿等核心控矿规律,并且对水口山矿区和黄金洞矿区知识图谱的二次聚类,结果有效区分了水口山矽卡岩型与黄金洞变质热液型成矿模式。进一步以 13 个成矿地质要素为特征集合,通过 Jaccard 相似性系数量化对比两大矿区与 7 个世界大型-超大型矿床,发现二者矿物组成具高度相似性,但成矿基础、找矿标志和岩体相似性低,其中水口山矿区因非金属矿物优势、构造匹配性强,综合相似性高于黄金洞矿区,形成大型-超大型矿床潜力更高,为区域成矿模式解析与深部勘查提供结构化知识支撑。

本文虽然对江南造山带金矿知识图谱进行一定的知识挖掘,但是知识图谱中遥感、重砂异常等实体稀缺,且数据多源于文献与报告,缺乏动态监测及野外实测属性数据。知识挖掘与分析深度有限,社区聚类语义解读依赖人工,矿床间相似性分析未考虑地质要素权重。下一步研究中需补充遥感、重砂异常等实体数据,整合动态监测数据与野外实测属性数据,提升知识图谱的数据完整性。与此同时优化矿床间相似性分析模型,引入地质要素权重赋值机制,结合成矿地质规律明确各要素的权重占比,提升相似性分析结果的科学性与精准性。

## References

- Bai, D.Y., Li, B., Zeng, G.Q., et al.,2025. Tectonic Ore-Controlling Characteristics and Dynamic Mechanism of the Hunan Segment of the Jiangnan Gold Ore Belt. *Geotectonica et Metallogenia*, 1-24(in Chinese with English abstract).
- Blondel, V.D., Guillaume, J., Lambiotte, R., et al.,2008. Fast Unfolding of Communities in Large Networks. *Journal of Statistical Mechanics: Theory and Experiment*,2008(10):P10008. <https://iopscience.iop.org/article/10.1088/1742-5468/2008/10/P10008>
- Deng, Y., Song, S., Fan, J., et al., 2024.Paleontology Knowledge Graph for Data-Driven Discovery. *Journal of Earth Science*, 35(3): 1024-1034. <https://doi.org/10.1007/s12583-023-1943-9>.
- Dong, S., Shi, Y., Ran, Y., et al., 2024. Biological Classification System Knowledge Graph and Semi-automatic Construction of its Invertebrate Fossil Branches. *Journal of Earth Science*,35(6):2119–2128. <https://doi.org/10.1007/s12583-023-1941-y>
- Enkhsaikhan, M., Liu, W., Holden, E.J. et al.,2021. Auto-Labeling Entities in Low-Resource Text: A Geological Case Study. *Knowledge and Information Systems* **63**, 695–715. <https://doi.org/10.1007/s10115-020-01532-6>
- Guo, F., Lai, P., Huang, F.M., et al.,2024. Literature Review and Research Progress of Landslide Susceptibility Evaluation Based on Knowledge Graph. *Earth Science*, 49(05): 1584-1606(in Chinese with English abstract).
- Kong, Y.X., Shi, G.Y., Wu, R.J. & Zhang, Y.C., 2019. K-Core: Theories and Applications. *Physics Reports*, 832, 1-32.
- Li, G., Wang, P., Liu, J., Guo, Y., Ji, K., Shang, Z. & Xu, Z., 2024. Meta in-Context Learning Makes Large Language Models Better Zero and Few-Shot Relation Extractors. *arXiv preprint arXiv:2404.17807*.
- Li, J., Huang, X.J., Gao, Y.T., et al.,2022. Distant Supervised Relation Extraction Based on Sentence-Level Attention with Relation Alignment. In: Sun, X., Zhang, X., Xia, Z., & Bertino, B., ed., International Conference on Adaptive and Intelligent Systems., Cham: Springer International Publishing, 142-152. [https://doi.org/10.1007/978-3-031-06794-5\\_12](https://doi.org/10.1007/978-3-031-06794-5_12)
- Li, S., Chen, J.P., & Xiang, J., 2018. Prospecting Information Extraction by Text Mining Based on Convolutional Neural Networks a Case Study of the Lala Copper Deposit, China. *IEEE access*, 6, 52286-52297. <https://ieeexplore.ieee.org/document/8466566>

- Ma, X., Ma, C., & Wang, C., 2020. A New Structure for Representing and Tracking Version Information in a Deep Time Knowledge Graph. *Computers & Geosciences*, 145,104620. <https://doi.org/10.1016/j.cageo.2020.104620>
- Ma, X.F., 2025. Research on Knowledge Extraction Method of Geological Hazards Based on Large Models. *Xining: Qinghai Normal University*, 1-69. <http://doi.org/10.27778/d.cnki.gqhzy.2025.000675>.(in Chinese with English abstract).
- Newman, M.E., 2006. Modularity and Community Structure in Networks. *Proceedings of the national academy of sciences*, 103(23), 8577-8582. <https://doi.org/10.1073/pnas.0601602103>
- Niwattanakul, S., Singthongchai, J., Naenudorn, E., & Wanapu, S., 2013. Using of Jaccard Coefficient for Keywords Similarity. In *Proceedings of the international multiconference of engineers and computer scientists*,1(6), 380-384.
- Peng, C., Xia, F., Naseriparsa, M., & Osborne, F., 2023. Knowledge Graphs: Opportunities and Challenges. *Artificial intelligence review*, 56(11), 13071-13102.
- Qiu, Q., Xie, Z., Wu, L., & Tao, L., 2020. Dictionary-Based Automated Information Extraction from Geological Documents using a Deep Learning Algorithm. *Earth and Space Science*, 7(3), e2019EA000993. <https://doi.org/10.1029/2019EA000993>
- Qiu, Q., Ma, K., Lv, H., Tao, L., & Xie, Z., 2023. Construction and Application of a Knowledge Graph for Iron Deposits using Text Mining Analytics and a Deep Learning Algorithm. *Mathematical Geosciences*, 55(3), 423-456. <https://doi.org/10.1007/s11004-023-10050-4>
- Qiu, Q., Tian, M., Xie, Z., Tan, Y., Ma, K., Wang, Q., Pan, S., & Tao, L., 2023. Extracting Named Entity using Entity Labeling in Geological Text Using Deep Learning Approach. *Journal of Earth Science*, 34(5), 1406-1417. <https://doi.org/10.1007/s12583-022-1789-8>
- Qiu, Q.J., Wu, L., Ma, K., et al.,2023. Construction Method of Geological Disaster Chain Knowledge Graph for Disaster Emergency Response. *Earth Science*, 48(05): 1875-1891(in Chinese with English abstract).
- Tian, M., Ma, K., Wu, Q., Qiu, Q., Tao, L., & Xie, Z., 2024. Joint Extraction of Entity Relations from Geological Reports Based on a Novel Relation Graph Convolutional Network. *Computers & Geosciences*, 187, 105571. <https://doi.org/10.1016/j.cageo.2024.105571>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł. & Polosukhin, I., 2017. Attention is All You Need. In: *Proceedings of the 31st International Conference on Neural*

*Information Processing Systems*, NY, USA, Curran Associates Inc,6000-1010

Wang, B., Wu, L., Xie, Z., Qiu, Q., Zhou, Y., Ma, K., & Tao, L., 2022. Understanding Geological Reports Based on Knowledge Graphs Using a Deep Learning Approach. *Computers & Geosciences*, 168, 105229. <https://doi.org/10.1016/j.cageo.2022.105229>

Wang, C., Ma, X., Chen, J., & Chen, J., 2018. Information Extraction and Knowledge Graph Construction from Geoscience Literature. *Computers & geosciences*, 112,112-120. <https://doi.org/10.1016/j.cageo.2017.12.007>

Wang, C., Li, Y., Chen, J., 2023. Text mining and knowledge graph construction from geoscience literature legacy: A review. In Ma, X., Mookerjee, M., Hsu, L., & Hills, D., Eds., *Recent Advancement in Geoinformatics and Data Science*, Geological Society of America,11–28. [https://doi.org/10.1130/2022.2558\(02\)](https://doi.org/10.1130/2022.2558(02))

Wang, J X., 2024. Intelligent Identification of Prospecting Criteria for Porphyry Copper Deposits Based on Large Language Models. *Jilin University, Changchun*. <https://doi.org/10.27162/d.cnki.gjlin.2024.008737>(in Chinese with English abstract).

Wei, Z., Su, J., Wang, Y., Tian, Y., & Chang, Y., 2020. A Novel Cascade Binary Tagging Framework for Relational Triple Extraction. In *Proceedings of the 58th annual meeting of the association for computational linguistics*, 1476-1488.

Wu, R.Z., Li, H., Mei, H.B., et al.,2025. Landslide Monitoring and Early Warning System Based on Knowledge Graph Retrieval Enhancement. *Earth Science*, 50(10):4125-4136. (in Chinese with English abstract).

Xiao, D., Song, W, G., Yan, Z. F., et al. 2025. Geological Characteristics, Metallogenic Regularity and Metallogenic Model of Gold Deposits in the Xuefeng Arc Structural Belt, Hunan Province. *Geology and Exploration*, 61(03): 450-463(in Chinese with English abstract).

Yang, X., Sun, L., Liu, M.L., Xiao, K.Y., Li, C., & Dong, X.C., 2025. Knowledge Graph Construction with BERT-BiLSTM-IDCNN-CRF and Graph Algorithms for Metallogenic Pattern Discovery: A Case Study of Pegmatite-Type Lithium Deposits in China. *Ore Geology Reviews*, 179, 106514. <https://doi.org/10.1016/j.oregeorev.2025.106514>

Zhang, L., Hou, M., Chen, A., Zhong, H., Ogg, J.G., & Zheng, D., 2023. Construction of a Fluvial Facies Knowledge Graph and its Application in Sedimentary Facies Identification. *Geoscience Frontiers*, 14(2), 101521. <https://doi.org/10.1016/j.gsf.2022.101521>

Zhang, L., Hou, Z., Shen, B., Chen, Q., Dong, S., & Fan, J., 2023. Paleobiogeographic Knowledge Graph: An Ongoing Work with Fundamental Support for Future Research. *Journal of Earth Science*, 34(5), 1339-1349. <https://doi.org/10.1007/s12583-023-1845-z>

Zhou, C., Wang, H., Wang, C., Hou, Z., Zheng, Z., Shen, S., Cheng, Q., Feng, Z., Wang, X., Lv, H., & Fan, J., 2021. Geoscience Knowledge Graph in the Big Data era. *Science China Earth Sciences*, 64(7), 1105-1114. <https://doi.org/10.1007/s11430-020-9750-4>

Zhou, C. H., Wang, H., Wang, C. S., et al. 2021. Research on Geoscience Knowledge Graph in the Big Data era. *Science China Earth Sciences*, 54(7): 1070-1079(in Chinese with English abstract).

Zhou, S., Meng, Y., Jin, B., & Han, J., 2024. Grasping the Essentials: Tailoring Large Language Models for Zero-Shot Relation Extraction. *arXiv preprint arXiv:2402.11142*.

#### 中文参考文献

柏道远, 李彬, 曾广乾, 等, 2025.江南金矿带湖南段构造控矿特征及动力机制. *大地构造与成矿学*, 1-24.

郭飞, 赖鹏, 黄发明, 等, 2024. 基于知识图谱的滑坡易发性评价文献综述及研究进展. *地球科学*, 49(05): 1584-1606.

马香菲, 2025.基于大模型的地质灾害知识抽取方法研究(硕士学位论文).西宁: 青海师范大学, 1-69. <http://doi.org/10.27778/d.cnki.gqhzy.2025.000675>.

邱芹军, 吴亮, 马凯, 等, 2023. 面向灾害应急响应的地质灾害链知识图谱构建方法. *地球科学*, 48(05): 1875-1891.

王嘉翔, 2024.基于大语言模型的斑岩型铜矿找矿标志智能识别研究. 吉林大学. <https://doi.org/10.27162/d.cnki.gjlin.2024.008737>

吴润泽, 李浩, 梅红波, 等, 2025.基于知识图谱检索增强生成的滑坡监测预警系统. *地球科学*, 50(10): 4125-4136.

肖丹, 宋维国, 严志飞, 等, 2025.湖南雪峰弧形构造带金矿地质特征、成矿规律与成矿模式. *地质与勘探*, 61(03): 450-463.

周成虎, 王华, 王成善, 等, 2021.大数据时代的地学知识图谱研究. *中国科学:地球科学*, 51(07): 1070-1079.