LLM-based Multi-Level Knowledge Generation for Few-shot Knowledge Graph Completion

Qian Li[1,2] , **Zhuo Chen**[3] , **Cheng Ji**[2] , **Shiqi Jiang**[2] and **Jianxin Li**[2,4*]
[1]State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing, China
[2]School of Computer Science and Engineering, Beihang University, Beijing, China
[3]College of Computer Science and Technology, Zhejiang University, Zhejiang, China
[4]Zhongguancun Lab, Beijing, China
liqian@bupt.edu.cn, zhuo.chen@zju.edu.cn, {jicheng,lijx}@act.buaa.edu.cn, jiangsq@buaa.edu.cn

Abstract

Knowledge Graphs (KGs) are pivotal in various NLP applications but often grapple with incompleteness, especially due to the long-tail problem where infrequent, unpopular relationships drastically reduce the KG completion performance. In this paper, we focus on Few-shot Knowledge Graph Completion (FKGC), a task addressing these gaps in long-tail scenarios. Amidst the rapid evolution of Large Language Models, we propose a generation-based FKGC paradigm facilitated by LLM distillation. Our MuKDC framework employs multi-level knowledge distillation for few-shot KG completion, generating supplementary knowledge to mitigate data scarcity in few-shot environments. MuKDC comprises two primary components: Multi-level Knowledge Generation, which enriches the KG at various levels, and Consistency Assessment, to ensure the coherence and reliability of the generated knowledge. Most notably, our method achieves SOTA results in both FKGC and multi-modal FKGC benchmarks, significantly advancing KG completion and enhancing the understanding and application of LLMs in structured knowledge generation and assessment.

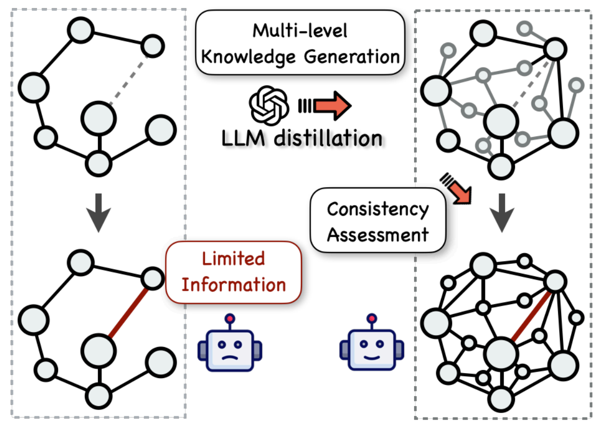


Figure 1: We break free from conventional thought patterns and introduce a generation-based FKGC paradigm through LLM distillation.

1 Introduction

Knowledge Graphs (KGs) are structured databases representing information through entities and their interrelations, typically organized as triples comprising a head entity, a relation, and a tail entity [Chen *et al.*, 2024]. KGs play a pivotal role in diverse NLP tasks, including question answering [Chen *et al.*, 2021b; Chen *et al.*, 2022], entity search [Gerritse *et al.*, 2022], and recommendation systems [Du *et al.*, 2022]. Despite their extensive utility, KGs frequently confront the challenge of incompleteness, manifested in missing or underrepresented entities and relations. Knowledge Graph Completion (KGC), synonymous with link prediction in the context of this work, focuses on identifying and inferring potential missing triples in KGs to address existing gaps. However, the long-tail problem, a common issue in real-world KGs, complicates this task. It arises from the significant variance in the frequency of relationship types and the disparity in the number of neighbors among different entity types. This results in many relations being sparsely represented, leading to inadequate training data for traditional KGC methods. To overcome this, the focus has shifted towards Few-shot Knowledge Graph Completion (FKGC).

FKGC is designed to enhance KG completion in scenarios with limited data availability. It involves predicting missing entities in triples with rare relations (the query set) using a minimal number of example triples containing the same relation (the support set). This approach aligns with few-shot learning methodologies in KGs, addressing the critical challenge of data sparsity in KG augmentation. Current methods mainly focus on model-level enhancements. This includes strategies such as utilizing paths between entities to capture complex relations and their interactions, which leverages the local sub-graph structure for for enriched entity learning [Xu *et al.*, 2021]. Additionally, employing a model-agnostic meta-learning framework is also proved to be helpful in distilling relation-specific information [Chen *et al.*, 2019; Niu *et al.*, 2021].

*Corresponding Author.

In this paper, amidst the rapid advancement of Large Language Models (LLMs), we break free from conventional thought patterns and introduce a **generation-based FKGC paradigm through LLM distillation** (Figure 1), aiming to revisit and reassess the future development and potential value of FKGC. Specifically, we propose a LLM-based Multi-level Knowledge Distillation framework for few-shot KG Completion, termed as MuKDC[1]. MuKDC generates additional knowledge for sparse entities and relations, addressing the challenges of data scarcity in few-shot multi-modal learning environments and supporting further FKGC tasks. As a pioneering study in this domain, we have endeavored to keep our approach as straightforward and effective as possible.

This framework includes two main components: **Multi-level Knowledge Generation (MKG)** and **Consistency Assessment (CA)**. ***(i)*** MKG is designed to enrich and expand the KG by generating additional knowledge at multiple levels, including Triplet Generation (TG), Attribute Generation (AG), and Decision Path Generation (DPG). These components collectively enhance the KG, offering a more comprehensive base for few-shot learning applications. TG augments the relational structure of the KG; AG enriches entities with additional descriptive attributes; and DPG creates logical pathways to facilitate inferencing and robust knowledge discovery. ***(ii)*** CA assesses the coherence and reliability of the generated knowledge, comparing existing data with the new knowledge to ensure the expanded graph's integrity and consistency.

Note that the LLM distillation method we propose can naturally serve as a means to complete and construct KGs. This method can be viewed as a data augmentation strategy for downstream FKGC, effectively distilling parametric knowledge from LLMs into structured knowledge. This assists not only in knowledge verification but also in enhancing the explainability of other downstream tasks. To validate the generality and robustness of MuKDC, we extend our analysis to multi-modal FKGC scenarios, demonstrating that our model achieves **SOTA results in both FKGC and multi-modal FKGC benchmarks**.

2 Related Work

2.1 Few-Shot Knowledge Graph Completion

Numerous studies [Niu *et al.*, 2021; Liang *et al.*, 2023] in few-shot relational modeling for KGC have concentrated on long-tail relations, broadly categorizing into three approaches: ***(i)*** **Metric-based Methods**: These methods focus on learning metrics to determine similarity between support and query triples. Initial models like GMatching [Xiong *et al.*, 2018] utilize one-hop neighbors to refine entity embeddings. Later, FSRL [Zhang *et al.*, 2020] and FAAN [Sheng *et al.*, 2020] integrate attention mechanisms for handling relation-specific neighborhoods more effectively. Advancements like REFORM [Wang *et al.*, 2021b] and YANA [Liang *et al.*, 2022b] employe GNNs for latent relation detection, while P-INT [Xu *et al.*, 2021] and MetaP [Jiang *et al.*, 2021] innovate with directed subgraphs and convolutional layers to extract relation patterns. ***(ii)*** **Optimization-based Methods**: These methods [Chen *et al.*, 2019; Niu *et al.*, 2021; Lv *et al.*, 2019] adapt swiftly to new relations using Model-Agnostic Meta-Learning (MAML). Meta-KGR further [Lv *et al.*, 2019] combines MAML with multi-hop path-finding for enhanced entity selection. ***(iii)*** **Cognitive Graph-based Methods**: An example is CogKR [Du *et al.*, 2021], which applies cognitive science principles to construct and update a cognitive graph, facilitating KGC.

Building on these foundations, our paper, set against the backdrop of the rapid advancement of LLMs [Zhang *et al.*, 2023c], introduces a generation-based FKGC paradigm through LLM distillation, as shown in Figure 1. Our approach aims to revisit and reassess the future development and potential value of FKGC, marking a departure from conventional methodologies.

2.2 Multi-modal Knowledge Graph Completion

Multi-modal Knowledge Graph Completion (MMKGC) models primarily focus on incorporating visual information to augment structural-only or text-only KGC tasks [Chen *et al.*, 2024; Fang *et al.*, 2022]. Recent MMKG-based models [Liang *et al.*, 2022a; Zhang *et al.*, 2023a] often process visual and structural data separately [Chen *et al.*, 2023b; Chen *et al.*, 2023c], employing general KG embedding (KGE) methods like TransE [Bordes *et al.*, 2013] for unified modeling. Various multimodal context embedding fusion methods, such as simple concatenation [Sergieh *et al.*, 2018], DeViSE [Frome *et al.*, 2013], and Imagined [Collell *et al.*, 2017], have been explored to integrate these different modalities.

Moreover, TransAE [Wang *et al.*, 2019] introduces an autoencoder mechanism for seamless visual and structural integration; RSME [Wang *et al.*, 2021a] focuses on evaluating different image encoders, highlighting the Visual Transformer [Dosovitskiy *et al.*, 2021] as particularly effective; VBKGC [Zhang and Zhang, 2022] and MANS [Zhang *et al.*, 2023b] suggest fine-grained visual negative sampling to better align visual embeddings with structural embeddings, representing a novel approach for fine-grained comparison training; Additionally, MACO [Zhang *et al.*, 2023d] proposes an adversarial training method to complete missing modal information. The current trend in MMKGC primarily utilizes image information as attributes for task design, facilitating the solution of FKGC and Multi-modal FKGC under a unified framework.

3 Preliminaries

Here we define the FKGC task with potential multi-modal contexts as follows:

Definition 1. ***(Multi-modal) Few-shot Knowledge Graph Completion.*** *Given an incomplete KG* $\mathcal{G} = (\mathcal{E}, \mathcal{R}, \mathcal{T}, \mathcal{A}, \mathcal{V})$ *with* $\mathcal{T} = \{\mathcal{T}_{\mathcal{A}}, \mathcal{T}_{\mathcal{R}}\}$*, where* $\mathcal{E}$*,* $\mathcal{R}$ *and* $\mathcal{T}$*,* $\mathcal{A}$ *are the sets of entities, relations, triples, attributes, and values, respectively. The FKGC task aims to complete* $\mathcal{G}$ *by identifying a set of missing triples* $\mathcal{T}' = \{(h, r, zt) | (h, r, t) \notin \mathcal{T}, h, t \in \mathcal{E}, r \in \mathcal{R}\}$*, given only few-shot entity pairs* (h, t) *and their potential multi-modal attributes for each relation* r*. To differentiate modal characteristics of attributes, we define text attributes as* v^l *and visual attributes as* v^v*. For instance, in a MMKG, the*

[1]The source code is available at https://github.com/xiaoqian19940510/MuKDC.

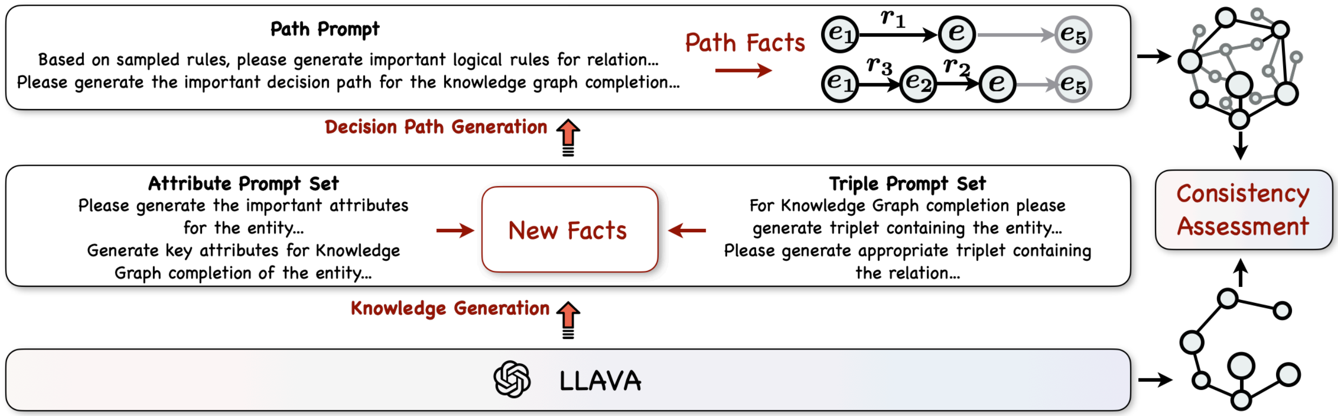


Figure 2: The framework of MuKDC. It includes two main components: Multi-level Knowledge Generation (MKG) and Consistency Assessment (CA). **MKG** enriches and expands the KG by generating additional knowledge at multiple levels; and Decision Path Generation creates logical pathways to facilitate inferencing and robust knowledge discovery. **CA** assesses the coherence and reliability of the generated knowledge, ensuring the expanded KG's integrity and consistency.

attribute triple (e, a, v^v) *in* $\mathcal{T}_{\mathcal{A}}$ *may associate an entity* e *from* $\mathcal{E}$ *with a visual value* v^v *under the attribute* a*, designated as hasImage.*

Definition 1 also refers to the K-shot KGC task when K training examples are provided for each relation. Unlike prior studies that assume abundant training triples, FKGC addresses scenarios with limited training data. Specifically, the objective is to rank the correct tail entity higher than other candidates, given only K example triples $(h'_i, r, t'_i)_{i=1}^K$ for relation r. The candidate set is formed based on entity type constraints.

For a given relation $r \in \mathcal{R}$, a dataset D_{train} consists of **support** (D_{s_r}) and **query** (D_{q_r}) sets: $D_r = D_{s_r}, D_{q_r}$. Each of the support set D_{s_r} contains K triples for K-shot tasks. The query set $D_{q_r} = \{h_i, r, t_i, C_{h_i,r}\}$ includes query triples for relation r with ground-truth tail entities t_i for each query (h_i, r), and a candidate set $C_{h_i,r} = \{t_{ij}\}$ where each t_{ij} is an entity in $\mathcal{G}$. The metric model is then evaluated on this dataset by ranking the candidate set $C_{h_i,r}$, given the test query (h_i, r) and the support triples in D_{s_r}. The datasets D_{val} and D_{test} comprise similar structures.

After sufficient training, the model predicts facts for new relations $r' \in \mathcal{R}'$. The relation label spaces in the datasets are disjoint, i.e., $\mathcal{R} \cap \mathcal{R}' = \emptyset$, to adhere to the K-shot learning assumption. Otherwise, the model would have access to more than K-shot labeled data during testing, violating the few-shot learning premise.

4 Framework

As illustrated in Figure 2, our MuKDC framework comprises two primary components: Multi-level Knowledge Generation and Consistency Assessment.

(i) **Multi-level Knowledge Generation (MKG)** aims to enrich and expand the KG by generating multi-level knowledge. This includes: **Knowledge Generation**: Generating additional triples and attributes to densify the original KG. This deepens entity characterization by enriching entities with descriptive attributes and triplets, providing a more comprehensive understanding of each entity; **Decision Path Generation**: Constructing logical pathways linking entities and relations. These paths are pivotal for understanding relationship dynamics, facilitating inference, and enhancing knowledge discovery, while also providing interpretability for graph completion.

(ii) **Consistency Assessment (CA)**: A crucial part of MuKDC, it involves comparing the newly generated knowledge against existing data and predictions. This component uses predicted tail entities and known head entities to ensure the consistency and coherence of relationships in the expanded graph.

Overall, MuKDC is designed to overcome the challenges of sparse data in few-shot multi-modal learning environments. By generating multi-level knowledge, MuKDC not only enriches the KG but also ensures its integrity and consistency, providing a holistic understanding of the graph.

4.1 Multi-level Knowledge Generation

Knowledge Generation

Knowledge Generation in our framework can be likened to a brainstorming session by the LLM on the training set D_{train}. This process involves divergent thinking based on the existing triples and their attributes in the support set, thereby transforming parameter knowledge into structured knowledge to address sparse data challenges.

Specifically, Triplet Generation (TG) focuses on creating new relational triples, leveraging known entities and relations in a sampled one-hop sub-graph. Its goal is to enhance the KG's relational structure, thus broadening and deepening the understanding of entities and their interconnections. Attribute Generation (AG) engages LLMs to generate attributes for nodes in both textual and potential visual forms. This enriches entities with descriptive attributes, leading to a more comprehensive characterization of each entity within the graph. Note that these two processes, TG and AG, operate in parallel and do not directly influence each other.

Triplet Generation. Let $\mathcal{T}_g = \{\mathcal{T}_{\mathcal{A}}, \mathcal{T}_{\mathcal{R}}\}$ represent the set of existing triplets in D_{train}, with $\mathcal{E}_g$ and $\mathcal{R}_g$ denoting the sets of existing entities and relations, respectively. We define

$\mathcal{G}_e$ as a one-hop subgraph centered around entity e, sampled from $\mathcal{T}g$, where each entity in the sub-KG includes textual and potential visual attributes. The triplet generation process can be formulated as $\mathcal{T}_{\mathcal{R}\text{new}} = \text{TG}(\mathcal{G}_e)$, where $\mathcal{T}_{\mathcal{R}\text{new}}$ is the set of newly generated triplets. The function TG, realized by the LLM, generates new triplets based on the existing sub-KG. In practice, this is achieved by defining appropriate instruction templates for input into the LLM. Specifically, we have developed an "Instruction Pool", of which only a select few examples are presented below:

```
1. For knowledge graph completion please
   generate triplet containing the [ENT]:
2. Please generate appropriate triplet
   containing the [REL]:
```

For each instruction, we replace the placeholders `[ENT]` and `[REL]` with the sets of entities and relations contained in $\mathcal{G}_e$, respectively, to form an instance $Inst$. We then serialize $\mathcal{G}_e$ as a collection of triplets (including relational and attribute triples) to construct the final Template Seq$(\mathcal{G}_e)$ $||$ $Inst$. Here, "||" denotes the concatenation of the serialized triplet set with the instantiated instruction.

This process continues for X rounds, where X equals the number of entities in $\mathcal{E}_g$, effectively iterating through the entire D_{train} for data augmentation. note that for potential visual value v^v inputs that may occur, a trainable projection is utilized for transformation.

Attribute Generation. Similar to Triplet Generation, the function AG, realized by the LLM, generates new attributes based on the existing sub-KG to generate new attribute set $\mathcal{T}_{\mathcal{A}}$new. A corresponding "Instruction Pool" has been defined, with a few selected examples shown below:

```
1. Please generate the important attributes
    for the [ENT]:
2. Generate key attributes for knowledge
   graph completion of the [ENT]:
```

Other settings for AG are consistent with those of TG.

Decision Path Generation

If Knowledge Generation represents brainstorming within a set framework (creating triples and attributes from existing sub-KG data), Decision Path Generation (DPG) expands on this by extrapolating relationships more freely. DPG analyzes existing relationships and patterns to generate decision paths, uncovering deeper insights into the KG. For instance, from the facts "Hangzhou is in Zhejiang" and "Zhejiang is in China", it deduces that "Hangzhou is in China". Similarly, if "LeBron James" and "Anthony Davis" are known to play for the "Lakers", DPG infers that they are teammates.

Note that both Knowledge Generation and DPG can involve and be assisted by images, providing additional information inputs. This process can be facilitated by multi-modal LLMs (e.g., LLava [Liu *et al.*, 2023]). Furthermore, the input for DPG is based on the output of Knowledge Generation combined with the existing D_{train}. We refer to this combined set as $\hat{D}_{train} = \mathcal{T}_{\mathcal{R}\text{new}} \cup \mathcal{T}_{\mathcal{A}\text{new}} \cup \mathcal{T}$. The one-hop subgraph sampled from $\hat{D}_{train}$ is denoted as $\hat{\mathcal{G}}_e$.

The decision path generation process can be formulated as $\mathcal{T}_{\mathcal{R}\text{dpg}} = \text{DPG}(\hat{\mathcal{G}}e)$, where $\mathcal{T}_{\mathcal{R}\text{dpg}}$ is the set of newly generated decision paths. A corresponding "Instruction Pool" has been defined, with a few selected examples shown below:

```
1. Based on sampled rules, please generate
   important logical rules for relation
   for [ENT]:
2. Please generate the important decision
   path for the knowledge graph completion
    for [REL]:
```

Other settings for DPG are consistent with TG and AG.

4.2 Consistency Assessment

MuKDC incorporates a crucial Consistency Assessment (CA) process to verify the coherence and reliability of the newly generated knowledge, including triples and attributes. This process evaluates the compatibility of predicted tail entities with known head entities against the existing KG data. By ensuring that these predictions align with established relationships, the integrity and accuracy of the expanded graph are maintained.

Specifically, we employ a Knowledge Graph Embedding (KGE) model pre-trained on $\mathcal{T}_g$ (using TransE [Bordes *et al.*, 2013]) as the base scoring model for CA. This model is used to conduct consistency checks on all triplets in the newly generated sets $\mathcal{T}_{\mathcal{R}}$new and $\mathcal{T}_{\mathcal{A}\text{new}}$. Each triplet is sequentially input into the CA model, and those scoring below a certain threshold are discarded. Considering the likelihood of encountering many unseen entities and relations in this process, as highlighted in various zero-shot KGC works [Chen *et al.*, 2023a; Chen *et al.*, 2021a; Geng *et al.*, 2021], we train the KGE model where all entity and relation representations are derived from their embeddings obtained via the LLM. This designation facilitates the inductive aspect of the CA process.

4.3 Loss Function

Given a query relation r and its associated support triples $(h'_i, r, t'_i)_{i=1}^{K}$, we adopt negative sampling to construct query triples. Specifically, we gather a set of valid positive query triples $\{(h_i, r, t_i^+) \mid (h_i, r, t_i^-) \notin \mathcal{G}\}$ and corrupt the tail entities to construct another group of negative query triples $\{(h_i, r, t_i^-) \mid (h_i, r, t_i^-) \notin \mathcal{G}\}$. In line with established few-shot learning paradigms, our model is equipped with a hinge loss function:

$$l_\theta = \max(0, \gamma + \text{score}_\theta^- - \text{score}_\theta^+) \tag{1}$$

where score_θ^+ and score_θ^- are the scalar values derived from comparing the query triple $(h_i, r, t_i^+/t_i^-)$ with the support triples $(h'_i, r, t'_i)_{i=1}^{K}$ via a metric learning model from [Zhang *et al.*, 2022], and γ represents a tunable hyperparameter margin. During each training episode, we commence by sampling D_r from the designated training set D_{train}. Subsequently, K triples are selected to act as the support set D_{s_r}, and additional triples are chosen to formulate the positive query/test set D_{q_r}, compiled from the entirety of known triples within D_r.

Models		NELL					Wiki				
		MRR	Hits@10	Hits@5	Hits@1	Δ Avg (%)	MRR	Hits@10	Hits@5	Hits@1	Δ Avg (%)
KGE methods	TransE [Bordes *et al.*, 2013]	17.4%	31.3%	23.1%	10.1%	↓26.6	13.3%	18.7%	15.7%	10.0%	↓32.3
	DistMult [Yang *et al.*, 2015]	20.0%	31.1%	25.1%	13.7%	↓24.6	7.1%	15.1%	9.9%	2.4%	↓38.1
	ComplEx [Trouillon *et al.*, 2016]	18.4%	29.7%	22.9%	11.8%	↓26.4	8.0%	18.1%	12.2%	3.2%	↓36.3
	RotatE [Sun *et al.*, 2019]	17.6%	32.9%	24.7%	10.1%	↓25.8	4.9%	9.0%	6.4%	2.6%	↓41.0
FKGC methods	GMatching [Xiong *et al.*, 2018]	17.6%	29.4%	23.3%	11.0%	↓26.8	26.3%	38.7%	33.7%	19.7%	↓17.1
	MetaR [Chen *et al.*, 2019]	20.9%	35.5%	28.0%	14.1%	↓22.5	32.3%	41.8%	38.5%	27.0%	↓11.8
	FSRL [Zhang *et al.*, 2020]	15.3%	31.9%	21.2%	7.3%	↓28.2	15.8%	28.7%	20.6%	9.7%	↓28.0
	FAAN [Sheng *et al.*, 2020]	27.9%	42.8%	36.4%	20.0%	↓15.3	34.1%	46.3%	39.5%	28.1%	↓9.7
	GANA [Niu *et al.*, 2021]	34.4%	51.7%	43.7%	24.6%	↓8.5	35.1%	44.6%	40.7%	29.9%	↓9.1
	P-INT [Xu *et al.*, 2021]	40.5%	50.6%	50.3%	31.7%	↓3.8	-	-	-	-	-
	YANA [Liang *et al.*, 2022b]	29.4%	42.1%	36.4%	23.0%	↓14.4	38.0%	52.3%	44.2%	32.7%	↓4.9
Ours	**MuKDC**	**43.4%**	**59.2%**	**52.7%**	**33.1%**	-	**43.8%**	**54.7%**	**50.3%**	**38.0%**	-

Table 1: Summary of 5-shot entity completion performance on NELL and Wiki datasets. **Bold** indicates the best results; underlined indicates the second-best. "Δ Avg" represents the average decrease in the four metrics relative to our model. "↓" denotes a decrease compared to MuKDC, while "↑" indicates an improvement over the second-best result. A "–" signifies unavailable results. Notably, P-INT results are absent for the Wiki dataset, as it is not optimized for sparse data conditions.

Models		MM-FB15K					MM-DBpedia				
		MRR	Hits@10	Hits@5	Hits@1	Δ Avg (%)	MRR	Hits@10	Hits@5	Hits@1	Δ Avg (%)
KGE methods	TransE [Bordes *et al.*, 2013]	11.6%	16.4%	13.9%	8.9%	↓34.0	10.3%	15.5%	12.0%	7.7%	↓26.2
	DistMult [Yang *et al.*, 2015]	8.3%	13.2%	9.5%	3.7%	↓38.0	9.1%	14.1%	11.8%	8.8%	↓26.6
	ComplEx [Trouillon *et al.*, 2016]	6.7%	14.7%	8.9%	5.0%	↓37.9	12.1%	17.0%	12.3%	10.9%	↓24.5
	RotatE [Sun *et al.*, 2019]	13.1%	18.9%	16.0%	10.1%	↓32.2	15.0%	24.2%	17.9%	12.0%	↓20.3
FKGC methods	GMatching [Xiong *et al.*, 2018]	26.1%	37.7%	34.0%	18.9%	↓17.5	17.6%	29.3%	23.1%	11.6%	↓17.2
	FSRL [Zhang *et al.*, 2020]	16.2%	28.9%	19.7%	8.5%	↓28.4	15.8%	30.4%	22.0%	7.1%	↓18.7
	FAAN [Sheng *et al.*, 2020]	34.1%	45.8%	38.2%	27.9%	↓10.2	19.5%	31.0%	21.7%	13.6%	↓16.1
MKGC methods	TransAE [Wang *et al.*, 2019]	13.0%	24.3%	15.5%	11.6%	↓30.6	15.6%	23.7%	18.5%	13.1%	↓19.8
	RSME [Xu *et al.*, 2021]	18.8%	30.8%	24.9%	15.2%	↓24.3	17.7%	28.0%	21.9%	14.5%	↓17.0
	MULTIFORM [Zhang *et al.*, 2022]	43.7%	55.0%	46.1%	30.5%	↓2.9	30.3%	42.5%	33.4%	27.9%	↓4.0
Ours	**MuKDC**	**46.3%**	**58.6%**	**49.2%**	**32.7%**	-	**34.2%**	**46.9%**	**37.6%**	**31.5%**	-

Table 2: Main experiments of 5-shot multi-modal FKGC results.

5 Experiments

5.1 Detailed Setting

Implementation Details Our FKGC is implemented using PyTorch and trained on a Tesla V100 GPU. In the experiments, all entity embeddings are represented using the embedding of the last token from the final layer's output of the LLava model [Liu *et al.*, 2023], a common entity encoding approach in PLMs [Chen *et al.*, 2023d; Geng *et al.*, 2023]. For MMKG-based FKGC tasks, the standard visual projection head in LLava is employed, whereas for typical KG tasks, this module is not utilized. During Knowledge Generation and Decision Path Generation, all triple sequences are simply concatenated in the order of h, r, and t, with delimiters used to separate different triples. The threshold for the TransE [Bordes *et al.*, 2013] model during the Consistency Assessment process is set to 1.0. All other experimental settings not mentioned here, including the training procedures for FKGC, are kept consistent with those reported in [Zhang *et al.*, 2022].

FKGC Datasets. We employ two public benchmark datasets for FKGC: NELL and Wiki [Mitchell *et al.*, 2018; Vrandečić and Krötzsch, 2014]. NELL is an evolving KG dataset containing a broad spectrum of knowledge, while Wiki is derived from Wikipedia content. For both, we select relations with 50 to 499 triples as few-shot tasks. NELL includes 67 few-shot relations, and Wiki comprises 183. We divide NELL into 51/5/11 and Wiki into 133/16/34 relations for training, validation, and testing, respectively.

Multi-Modal FKGC Datasets. We employ another two public benchmark datasets, MM-FB15K and MM-DBpedia, tailored for few-shot multi-modal KG completion [Zhang *et al.*, 2022]. These datasets feature entities accompanied by images and textual descriptions of at least 15 words. For a robust evaluation, we consider only those relations with 50 to 500 triples as few-shot tasks. MM-FB15K comprises 356 few-shot relations, and MM-DBpedia includes 69. Relations exceeding 500 triples serve as background information, enhancing the knowledge topology. The splits for training, validation, and testing are allocated as 267/18/71 task relations for MM-FB15K and 51/6/12 for MM-DBpedia, following the 15:1:4 ratio, in line with prior studies [Zhang *et al.*, 2022].

Evaluation Metrics. We assess our model's performance using two standard metrics: Hits@N and MRR. Hits@N quantifies the proportion of correct entities ranked within the top N predictions, for N set to 1, 5, or 10. MRR computes the mean of the reciprocal ranks assigned to the correct entities across all test triples.

Variants	NELL					Wiki				
	MRR (%)	Hits@10 (%)	Hits@5 (%)	Hits@1 (%)	Δ Avg (%)	MRR (%)	Hits@10 (%)	Hits@5 (%)	Hits@1 (%)	Δ Avg (%)
MuKDC (Ours)	**43.4 ± 0.53**	**59.2 ± 0.28**	**52.7 ± 0.74**	**33.1 ± 0.65**	-	**43.8 ± 0.39**	**54.7 ± 0.70**	**50.3 ± 0.81**	**38.0 ± 0.42**	-
w/o Multi-level Know. Gen.	41.3 ± 0.47	58.4 ± 0.26	50.9 ± 0.85	32.0 ± 0.62	↓1.5	42.2 ± 0.41	53.6 ± 1.95	48.4 ± 0.82	36.5 ± 1.10	↓1.5
w/o Triplet Generation	42.8 ± 1.28	58.6 ± 1.43	51.5 ± 1.26	32.2 ± 1.04	↓0.8	43.0 ± 1.15	53.5 ± 1.26	49.5 ± 0.90	37.3 ± 0.87	↓0.9
w/o Attribute Generation	42.5 ± 1.32	58.7 ± 0.84	51.6 ± 1.23	32.8 ± 0.85	↓0.7	43.3 ± 1.37	53.2 ± 0.89	49.6 ± 1.24	37.1 ± 0.85	↓0.9
w/o Decision Path Generation	43.1 ± 1.35	58.8 ± 1.27	52.3 ± 1.28	32.9 ± 1.71	↓0.3	43.7 ± 1.10	54.1 ± 1.24	49.3 ± 1.15	37.4 ± 1.34	↓0.6
w/o Consistency Assessment	41.9 ± 1.21	58.3 ± 1.45	50.8 ± 0.89	32.3 ± 1.36	↓1.3	42.4 ± 0.72	53.8 ± 0.53	49.1 ± 1.01	36.2 ± 1.38	↓1.3

Table 3: Variant experiments on NELL and Wiki datasets. "w/o" indicates module removal from our full model. "Δ Avg" shows the average decrease across four metrics relative to our model. "↓" denotes an average decrease compared to MuKDC.

Variants	MM-FB15K					MM-DBpedia				
	MRR (%)	Hits@10 (%)	Hits@5 (%)	Hits@1 (%)	Δ Avg (%)	MRR (%)	Hits@10 (%)	Hits@5 (%)	Hits@1 (%)	Δ Avg (%)
MuKDC (Ours)	**46.3 ± 0.37**	**58.6 ± 0.25**	**49.2 ± 0.54**	**32.7 ± 0.82**	-	**34.2 ± 0.40**	**46.9 ± 0.55**	**37.6 ± 0.39**	**31.5 ± 0.73**	-
w/o Multi-level Know. Gen.	44.7 ± 1.45	56.9 ± 1.23	47.8 ± 1.03	30.6 ± 1.46	↓1.7	32.6 ± 0.58	44.7 ± 1.61	35.5 ± 1.27	29.8 ± 1.26	↓1.9
w/o Triplet Generation	45.6 ± 1.26	57.1 ± 0.73	47.9 ± 1.34	31.4 ± 1.22	↓1.2	33.9 ± 1.26	45.5 ± 1.23	36.4 ± 1.83	30.1 ± 1.20	↓1.1
w/o Attribute Generation	45.8 ± 1.27	57.3 ± 1.04	48.4 ± 0.93	31.5 ± 0.60	↓1.0	33.7 ± 1.25	45.8 ± 0.56	37.0 ± 0.29	30.7 ± 0.71	↓0.8
w/o Decision Path Generation	46.0 ± 0.34	58.2 ± 0.56	48.6 ± 1.28	32.3 ± 0.83	↓0.4	33.1 ± 1.10	46.2 ± 0.94	37.1 ± 1.20	31.3 ± 0.54	↓0.6
w/o Consistency Assessment	45.9 ± 0.47	57.4 ± 1.04	48.3 ± 1.21	33.2 ± 1.08	↓0.5	33.5 ± 1.21	46.3 ± 1.30	36.3 ± 1.16	30.2 ± 0.93	↓1.0

Table 4: Variant experiments on multi-modal FKGC.

5.2 Comparison Methods

Knowledge Embedding Models. We benchmarked against four KGE models: TransE [Bordes *et al.*, 2013], DistMult [Yang *et al.*, 2015], ComplEx [Trouillon *et al.*, 2016], and RotatE [Sun *et al.*, 2019]. These models vectorize relations and entities with constraints reflecting relational structures in KGs, capturing essential structural features for effective representation given ample data.

FKGC. Seven FKGC models were evaluated: GMatching (MaxP) [Xiong *et al.*, 2018], MetaR [Chen *et al.*, 2019], FSRL [Zhang *et al.*, 2020], FAAN [Sheng *et al.*, 2020], GANA [Niu *et al.*, 2021], YANA [Liang *et al.*, 2022b], and P-INT [Xu *et al.*, 2021]. These approaches utilize metric or optimization-based meta-learning with pre-trained embeddings, focusing on local structures and relational semantics for enhanced relation and entity pair embeddings, demonstrating strong results on NELL and Wiki datasets.

Multi-modal FKGC. Three baseline methods were considered: TransAE [Wang *et al.*, 2019], RSME [Wang *et al.*, 2021a], and MULTIFORM [Zhang *et al.*, 2022]. TransAE employs an autoencoder for multimodal integration, RSME assesses different image encoders and endorses ViT for multi-modal KGC, and MULTIFORM presents a multimodal few-shot relation learning framework that leverages multimodal context for entity representation and learns a metric for matching queries to few-shot examples.

5.3 Main Results

We present the overall 5-shot FKGC results in Tables 1 and 2, with all comparative model results sourced from their respective original papers:

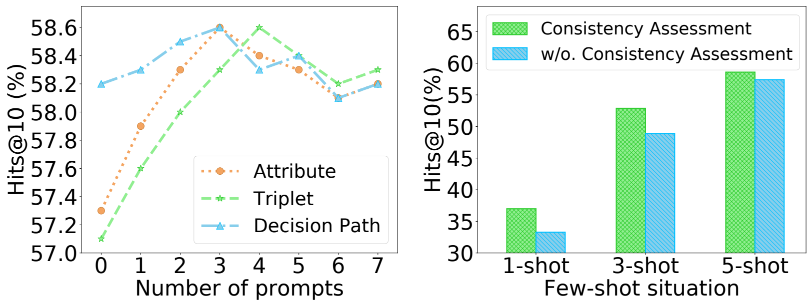
(i) Our model consistently outperforms baseline models across all four metrics on the NELL and Wiki datasets, achieving average improvements of 3.8% and 4.9%, respectively. Notably, it shows at least a 5.8% increase in MRR on Wiki and a 7.5% increase in Hits@10 on NELL. These improvements underscore the model's robust representation learning from multi-level information, affirming its capability for few-shot multi-modal KG completion tasks. ***(ii)*** Our model surpasses KG embedding baselines with substantial gains of 25.9% and 36.9% on average for the NELL and Wiki datasets, respectively. This performance showcases the benefit of generating multi-level knowledge for entities and relations, particularly apt for few-shot multi-modal KG completion. ***(iii)*** Moreover, our model outperforms few-shot KG completion baselines on all metrics, evidencing the efficacy of enriching the KG by generating additional knowledge and employing consistency assessment loss to minimize performance gaps in generated knowledge scores. ***(iv)*** Against the best-performing baselines GANA and YANA on NELL and Wiki, our model shows a minimum average improvement of 8.5% and 4.9%, respectively. This is attributed to our method's comprehensive understanding of the multi-modal KG.

In the realm of multi-modal FKGC, our model maintains a considerable lead over all baselines. Against the top-performing baseline MULTIFORM on the MM-FB15K and MM-DBpedia datasets, it registers a minimum average enhancement of 7.4% and 4.1%, respectively. Such results further confirm the model's proficiency in handling multi-modal FKGC tasks.

5.4 Discussion for Model Variants

To assess the contribution of each module in our model, we performed variant experiments on the NELL and Wiki datasets, as shown in Tables 3 and 4. These experiments compare our complete model against versions with key modules removed. The observations from these comparisons are as follows:

(i) The Multi-level Knowledge Generation module shows a notable impact, likely because it enhances the utilization of limited samples by generating entity and relation representations. ***(ii)*** Components such as Triplet Generation, Attribute Generation, and Decision Path Generation positively influence the results. They contribute to identifying valuable entity



(a) Impact on knowledge generation. (b) Impact on consistency assessment.

Path 1: **Bill Maher** -> award -> Primetime Emmy Award for Individual Performance -> ceremony -> 60th
Path 2: Santa Barbara -> **award_nominee** -> Sheri Anderson
Path 3: **Bill Maher** -> **award_nominee** -> Brad Grey -> film -> Charlie and the Chocolate Factory

(Bill Maher, award_nominee, Bernie Brillstein)

Path 1: **Bolton Wanderers F.C.** -> player-> Keith Andrews-> team -> Wolverhampton Wanderers F.C.
Path 2: **EFL Championship**-> team-> **Bolton Wanderers F.C.** -> sport -> association football
Path 3: Leicester City F.C. -> **league** -> **EFL Championship** -> award_winner-> Reading F.C.

(Bolton Wanderers F.C., league, EFL Championship)

(c) Interpretable path analysis.

Figure 3: Analysis for the few-shot multi-modal knowledge graph completion on the few-shot situations.

features with greater certainty, highlighting the importance of generating multi-dimensional knowledge at various levels. ***(iii)*** Removing the Consistency Assessment module leads to a decrease in performance, underscoring its role in evaluating the consistency and coherence of relationships within the expanded KG. These findings collectively validate the significance of each component in our model's overall effectiveness.

5.5 Discussions on Knowledge Generation

We conducted experiments to assess the impact of the multi-level knowledge generation module, focusing on the number of prompts used, as illustrated in Figure 3 (a).

The findings indicate: ***(i)*** Optimal Performance: The model performs best when using three attribute prompts, four triplet prompts, and three decision path prompts. This suggests a balance is necessary to prevent overfitting, which may occur with an excessive number of prompts. Overfitting leads to reduced generalizability as the model becomes overly reliant on specific patterns in the prompts. ***(ii)*** Effect of Excessive Prompts: Increasing the number of prompts beyond this optimal range leads to a notable decline in performance. Too many prompts can introduce confusion and noise, impairing the model's effectiveness and decision-making capacity.

These results highlight the importance of a balanced approach to the number of prompts in the knowledge generation module for effective model performance.

5.6 Discussion on Consistency Assessment

We evaluated the impact of the consistency assessment across 1-shot, 3-shot, and 5-shot scenarios on the MM-FB15K and MM-DBpedia datasets, as depicted in Figure 3 (b).

Key findings include: ***(i)*** Models lacking the consistency assessment module ("w/o. Consistency Assessment") show a notable decrease across all metrics, highlighting the critical role of evaluating generated facts in enhancing model performance. ***(ii)*** Our model demonstrates a greater sensitivity in the 1-shot multi-modal KG completion tasks compared to other baselines. This sensitivity can be attributed to our model's reliance on consistency to refine the generated KG. ***(iii)*** In various few-shot contexts, our model consistently outperforms others, underscoring its aptitude for the task by rigorously considering triplet consistency and evaluating the coherence between original and generated knowledge.

These insights confirm that our model effectively leverages multi-level prompts, making it particularly adept at consistency assessment in few-shot MMKGC tasks.

5.7 Discussion on Interpretable Path

To underscore the role of decision paths in boosting interpretability and supporting predictions, we analyzed their application using the MM-FB15K and MM-DBpedia datasets, as depicted in Figure 3 (c).

Key observations include: ***(i)*** Decision paths effectively guide predictions. For instance, consider the relation "award_nominee" between the head entity "Bill Maher" and the predicted tail entity "Bernie Brillstein". A possible decision path is: "Bill Maher $\rightarrow$ award $\rightarrow$ Primetime Emmy Award $\rightarrow$ ceremony $\rightarrow$ 60th". This path not only clarifies the prediction's rationale but also highlights Bill Maher's connection with the Primetime Emmy Award. ***(ii)*** Decision paths reveal the crucial roles of various relations and entities in the prediction process. Examining the relation "league" between "Bolton Wanderers F.C." and the predicted tail entity "EFL Championship", a plausible path is: "EFL Championship $\rightarrow$ team $\rightarrow$ Bolton Wanderers F.C. $\rightarrow$ sport $\rightarrow$ association football". This demonstrates the model's reliance on specific relationships for its predictions.

These insights confirm that decision path generation facilitates interpretable paths for KGC, offering a clearer understanding of model's reasoning and enhancing interpretability.

6 Conclusion

In this work, we introduced the MuKDC framework for FKGC, addressing data scarcity in few-shot scenarios through LLM distillation. Comprising two integral components, Multi-level Knowledge Generation and Consistency Assessment, MuKDC goes beyond merely expanding and deepening KGs; it ensures their coherence and reliability through rigorous evaluation and alignment with existing knowledge structures. This framework has demonstrated SOTA performance in both FKGC and multi-modal FKGC tasks, highlighting its efficacy in enriching KGs in long-tail scenarios. The success of MuKDC in advancing KG completion and its potential in leveraging LLMs for NLP tasks sets a new benchmark in the field and paves the way for future research in KG augmentation.

Acknowledgements

We thank the anonymous reviewers for their insightful comments and suggestions. The corresponding author is Jianxin Li. The authors of this paper were supported by the NSFC through grant No.62225202.

References

[Bordes *et al.*, 2013] Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. Translating embeddings for modeling multi-relational data. *NeurIPS*, 26, 2013.

[Chen *et al.*, 2019] Mingyang Chen, Wen Zhang, Wei Zhang, Qiang Chen, and Huajun Chen. Meta relational learning for few-shot link prediction in knowledge graphs. In *EMNLP*, pages 4217–4226, 2019.

[Chen *et al.*, 2021a] Jiaoyan Chen, Yuxia Geng, Zhuo Chen, Ian Horrocks, Jeff Z. Pan, and Huajun Chen. Knowledge-aware zero-shot learning: Survey and perspective. In *IJCAI*, pages 4366–4373, 2021.

[Chen *et al.*, 2021b] Zhuo Chen, Jiaoyan Chen, Yuxia Geng, Jeff Z. Pan, Zonggang Yuan, and Huajun Chen. Zero-shot visual question answering using knowledge graph. In *ISWC*, volume 12922 of *Lecture Notes in Computer Science*, pages 146–162, 2021.

[Chen *et al.*, 2022] Zhuo Chen, Yufeng Huang, Jiaoyan Chen, Yuxia Geng, Yin Fang, Jeff Z. Pan, Ningyu Zhang, and Wen Zhang. Lako: Knowledge-driven visual question answering via late knowledge-to-text injection. In *IJCKG*, pages 20–29, 2022.

[Chen *et al.*, 2023a] Jiaoyan Chen, Yuxia Geng, Zhuo Chen, Jeff Z. Pan, Yuan He, Wen Zhang, Ian Horrocks, and Huajun Chen. Zero-shot and few-shot learning with knowledge graphs: A comprehensive survey. *Proc. IEEE*, 111(6):653–685, 2023.

[Chen *et al.*, 2023b] Zhuo Chen, Jiaoyan Chen, Wen Zhang, Lingbing Guo, Yin Fang, Yufeng Huang, Yichi Zhang, Yuxia Geng, Jeff Z. Pan, Wenting Song, and Huajun Chen. Meaformer: Multi-modal entity alignment transformer for meta modality hybrid. In *ACM Multimedia*, pages 3317–3327, 2023.

[Chen *et al.*, 2023c] Zhuo Chen, Lingbing Guo, Yin Fang, Yichi Zhang, Jiaoyan Chen, Jeff Z. Pan, Yangning Li, Huajun Chen, and Wen Zhang. Rethinking uncertainly missing and ambiguous visual modality in multi-modal entity alignment. In *ISWC*, volume 14265 of *Lecture Notes in Computer Science*, pages 121–139, 2023.

[Chen *et al.*, 2023d] Zhuo Chen, Wen Zhang, Yufeng Huang, Mingyang Chen, Yuxia Geng, Hongtao Yu, Zhen Bi, Yichi Zhang, Zhen Yao, Wenting Song, Xinliang Wu, Yi Yang, Mingyi Chen, Zhaoyang Lian, Yingying Li, Lei Cheng, and Huajun Chen. Tele-knowledge pre-training for fault analysis. In *ICDE*, pages 3453–3466, 2023.

[Chen *et al.*, 2024] Zhuo Chen, Yichi Zhang, Yin Fang, Yuxia Geng, Lingbing Guo, Xiang Chen, Qian Li, Wen Zhang, Jiaoyan Chen, Yushan Zhu, Jiaqi Li, Xiaoze Liu, Jeff Z. Pan, Ningyu Zhang, and Huajun Chen. Knowledge graphs meet multi-modal learning: A comprehensive survey. *CoRR*, abs/2402.05391, 2024.

[Collell *et al.*, 2017] Guillem Collell, Ted Zhang, and Marie-Francine Moens. Imagined visual representations as multimodal embeddings. In *AAAI*, pages 4378–4384, 2017.

[Dosovitskiy *et al.*, 2021] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. In *ICLR*, 2021.

[Du *et al.*, 2021] Zhengxiao Du, Chang Zhou, Jiangchao Yao, Teng Tu, Letian Cheng, Hongxia Yang, Jingren Zhou, and Jie Tang. Cogkr: Cognitive graph for multi-hop knowledge reasoning. *TKDE*, 2021.

[Du *et al.*, 2022] Yuntao Du, Xinjun Zhu, Lu Chen, Baihua Zheng, and Yunjun Gao. HAKG: hierarchy-aware knowledge gated network for recommendation. In *SIGIR*, pages 1390–1400, 2022.

[Fang *et al.*, 2022] Xiang Fang, Daizong Liu, Pan Zhou, and Yuchong Hu. Multi-modal cross-domain alignment network for video moment retrieval. *IEEE Transactions on Multimedia*, 2022.

[Frome *et al.*, 2013] Andrea Frome, Gregory S. Corrado, Jonathon Shlens, Samy Bengio, Jeffrey Dean, Marc'Aurelio Ranzato, and Tomás Mikolov. Devise: A deep visual-semantic embedding model. In *NeurIPS*, pages 2121–2129, 2013.

[Geng *et al.*, 2021] Yuxia Geng, Jiaoyan Chen, Zhuo Chen, Jeff Z. Pan, Zhiquan Ye, Zonggang Yuan, Yantao Jia, and Huajun Chen. Ontozsl: Ontology-enhanced zero-shot learning. In *WWW*, pages 3325–3336, 2021.

[Geng *et al.*, 2023] Yuxia Geng, Jiaoyan Chen, Yuhang Zeng, Zhuo Chen, Wen Zhang, Jeff Z. Pan, Yuxiang Wang, and Xiaoliang Xu. Prompting disentangled embeddings for knowledge graph completion with pre-trained language model. *CoRR*, abs/2312.01837, 2023.

[Gerritse *et al.*, 2022] Emma J. Gerritse, Faegheh Hasibi, and Arjen P. de Vries. Entity-aware transformers for entity search. In *SIGIR*, pages 1455–1465, 2022.

[Jiang *et al.*, 2021] Zhiyi Jiang, Jianliang Gao, and Xinqi Lv. Metap: Meta pattern learning for one-shot knowledge graph completion. In *SIGIR*, pages 2232–2236, 2021.

[Liang *et al.*, 2022a] Ke Liang, Lingyuan Meng, Meng Liu, Yue Liu, Wenxuan Tu, Siwei Wang, Sihang Zhou, Xinwang Liu, and Fuchun Sun. Reasoning over different types of knowledge graphs: Static, temporal and multi-modal. *arXiv preprint arXiv:2212.05767*, 2022.

[Liang *et al.*, 2022b] Yi Liang, Shuai Zhao, Bo Cheng, Yuwei Yin, and Hao Yang. Tackling solitary entities for few-shot knowledge graph completion. In *KSEM*, volume 13368 of *Lecture Notes in Computer Science*, pages 227–239, 2022.

[Liang *et al.*, 2023] Ke Liang, Lingyuan Meng, Meng Liu, Yue Liu, Wenxuan Tu, Siwei Wang, Sihang Zhou, and Xinwang Liu. Learn from relational correlations and periodic events for temporal knowledge graph reasoning. In *SIGIR*, pages 1559–1568, 2023.

[Liu *et al.*, 2023] Haotian Liu, Chunyuan Li, Qingyang Wu, and Yong Jae Lee. Visual instruction tuning. *CoRR*, abs/2304.08485, 2023.

[Lv *et al.*, 2019] Xin Lv, Yuxian Gu, Xu Han, Lei Hou, Juanzi Li, and Zhiyuan Liu. Adapting meta knowledge graph information for multi-hop reasoning over few-shot relations. In *EMNLP*, pages 3374–3379, 2019.

[Mitchell *et al.*, 2018] Tom Mitchell, William Cohen, Estevam Hruschka, Partha Talukdar, Bishan Yang, Justin Betteridge, Andrew Carlson, Bhavana Dalvi, Matt Gardner, Bryan Kisiel, et al. Never-ending learning. *Communications of the ACM*, 61(5):103–115, 2018.

[Niu *et al.*, 2021] Guanglin Niu, Yang Li, Chengguang Tang, Ruiying Geng, Jian Dai, Qiao Liu, Hao Wang, Jian Sun, Fei Huang, and Luo Si. Relational learning with gated and attentive neighbor aggregator for few-shot knowledge graph completion. In *SIGIR*, pages 213–222, 2021.

[Sergieh *et al.*, 2018] Hatem Mousselly Sergieh, Teresa Botschen, Iryna Gurevych, and Stefan Roth. A multimodal translation-based approach for knowledge graph representation learning. In **SEM@NAACL-HLT*, pages 225–234, 2018.

[Sheng *et al.*, 2020] Jiawei Sheng, Shu Guo, Zhenyu Chen, Juwei Yue, Lihong Wang, Tingwen Liu, and Hongbo Xu. Adaptive attentional network for few-shot knowledge graph completion. In *EMNLP*, pages 1681–1691, 2020.

[Sun *et al.*, 2019] Zhiqing Sun, Zhi-Hong Deng, Jian-Yun Nie, and Jian Tang. Rotate: Knowledge graph embedding by relational rotation in complex space. In *ICLR*, 2019.

[Trouillon *et al.*, 2016] Théo Trouillon, Johannes Welbl, Sebastian Riedel, Éric Gaussier, and Guillaume Bouchard. In *ICML*, volume 48, pages 2071–2080, 2016.

[Vrandečić and Krötzsch, 2014] Denny Vrandečić and Markus Krötzsch. Wikidata: a free collaborative knowledgebase. *Communications of the ACM*, 57(10):78–85, 2014.

[Wang *et al.*, 2019] Zikang Wang, Linjing Li, Qiudan Li, and Daniel Zeng. Multimodal data enhanced representation learning for knowledge graphs. In *IJCNN*, pages 1–8, 2019.

[Wang *et al.*, 2021a] Meng Wang, Sen Wang, Han Yang, Zheng Zhang, Xi Chen, and Guilin Qi. Is visual context really helpful for knowledge graph? A representation learning perspective. In *ACM Multimedia*, pages 2735–2743, 2021.

[Wang *et al.*, 2021b] Song Wang, Xiao Huang, Chen Chen, Liang Wu, and Jundong Li. REFORM: error-aware few-shot knowledge graph completion. In *CIKM*, pages 1979–1988, 2021.

[Xiong *et al.*, 2018] Wenhan Xiong, Mo Yu, Shiyu Chang, Xiaoxiao Guo, and William Yang Wang. One-shot relational learning for knowledge graphs. In *EMNLP*, 2018.

[Xu *et al.*, 2021] Jingwen Xu, Jing Zhang, Xirui Ke, Yuxiao Dong, Hong Chen, Cuiping Li, and Yongbin Liu. P-INT: A path-based interaction model for few-shot knowledge graph completion. In *ACL Finding*, pages 385–394, 2021.

[Yang *et al.*, 2015] Bishan Yang, Wen-tau Yih, Xiaodong He, Jianfeng Gao, and Li Deng. Embedding entities and relations for learning and inference in knowledge bases. In *ICLR*, 2015.

[Zhang and Zhang, 2022] Yichi Zhang and Wen Zhang. Knowledge graph completion with pre-trained multimodal transformer and twins negative sampling. *CoRR*, abs/2209.07084, 2022.

[Zhang *et al.*, 2020] Chuxu Zhang, Huaxiu Yao, Chao Huang, Meng Jiang, Zhenhui Li, and Nitesh V. Chawla. Few-shot knowledge graph completion. In *AAAI*, pages 3041–3048, 2020.

[Zhang *et al.*, 2022] Xuan Zhang, Xun Liang, Xiangping Zheng, Bo Wu, and Yuhui Guo. MULTIFORM: few-shot knowledge graph completion via multi-modal contexts. In *ECML/PKDD*, volume 13714 of *Lecture Notes in Computer Science*, pages 172–187, 2022.

[Zhang *et al.*, 2023a] Jingdan Zhang, Jiaan Wang, Xiaodan Wang, Zhixu Li, and Yanghua Xiao. Aspectmmkg: A multi-modal knowledge graph with aspect-aware entities. In *CIKM*, pages 3361–3370, 2023.

[Zhang *et al.*, 2023b] Yichi Zhang, Mingyang Chen, and Wen Zhang. Modality-aware negative sampling for multi-modal knowledge graph embedding. *CoRR*, abs/2304.11618, 2023.

[Zhang *et al.*, 2023c] Yichi Zhang, Zhuo Chen, Yin Fang, Lei Cheng, Yanxi Lu, Fangming Li, Wen Zhang, and Huajun Chen. Knowledgeable preference alignment for llms in domain-specific question answering. *CoRR*, abs/2311.06503, 2023.

[Zhang *et al.*, 2023d] Yichi Zhang, Zhuo Chen, and Wen Zhang. MACO: A modality adversarial and contrastive framework for modality-missing multi-modal knowledge graph completion. In *NLPCC (1)*, volume 14302, pages 123–134, 2023.